Zhenan Sun
Jianhuang Lai
Xilin Chen
Tieniu Tan (Eds.)

LNCS 7098

Biometric Recognition

6th Chinese Conference, CCBR 2011
Beijing, China, December 2011
Proceedings

Springer

Lecture Notes in Computer Science 7098

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Zhenan Sun Jianhuang Lai Xilin Chen
Tieniu Tan (Eds.)

Biometric Recognition

6th Chinese Conference, CCBR 2011
Beijing, China, December 3-4, 2011
Proceedings

Springer

Volume Editors

Zhenan Sun
Tieniu Tan
Chinese Academy of Sciences, Institute of Automation
National Laboratory of Pattern Recognition
Center for Biometrics and Security Research
P.O. Box 2728
Beijing, 100190, China
E-mail: {znsun; tnt}@nlpr.ia.ac.cn

Jianhuang Lai
Sun Yat-Sen University
School of Information Science and Technology
Guangzhou, 510275, China
E-mail: stsljh@mail.sysu.edu.cn

Xilin Chen
Chinese Academy of Sciences
Institute of Computing Technology
Beijing, 100190, China
E-mail: xlchen@ict.ac.cn

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-25448-2 e-ISBN 978-3-642-25449-9
DOI 10.1007/978-3-642-25449-9
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011941147

CR Subject Classification (1998): I.5, I.4, I.2.10, F.2.2, I.3.5, K.6.5

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

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

Automatic biometric recognition has received unprecedented attention due to its wide applications in access control, secure bank transactions, national ID cards, welfare distribution, etc. However, the overall performance of current biometric systems in terms of usability, accuracy, robustness, scalability, and security is still unsatisfactory in real-world applications. Therefore great efforts are needed in the research and development of innovative biometric sensors and algorithms.

Biometrics has become a hot topic in academia and industry as well as for the government of China, driven by the increasing requirements of fast and reliable personal identification in a country with the largest population in the world. Biometrics researchers from China have contributed about one-third of papers presented at the International Conference on Biometrics (ICB). A large number of research projects on biometrics have been funded through the National Basic Research Program of China, the Natural Science Foundation of China, and the National Hi-Tech Research and Development Program of China, etc. Furthermore, there are more than 200 biometrics-related companies in China and the annual increase in the Chinese biometrics market is over 30%. Large-scale biometrics systems have been deployed in government, banks, telecom companies, prisons, education, etc.

The Chinese Conference on Biometric Recognition (CCBR) has been successfully held in Beijing, Hangzhou, Xi'an, and Guangzhou for five times since 2000. The 6th Chinese Conference on Biometric Recognition (CCBR 2011) was held in Beijing during December 3–4, 2011. This volume of conference proceedings includes 35 papers carefully selected from a total of 71 submissions. The papers address the problems in face, iris, hand biometrics, speaker, handwriting, gait, soft biometrics and other related topics, and they contribute new ideas to the research and development of reliable and practical solutions for biometric authentication.

We would like to express our gratitude to all the contributors, reviewers, Program Committee and Organizing Committee members who made this a very successful conference. We also wish to acknowledge the China Computer Federation, Chinese Association for Artificial Intelligence, Springer, IrisKing Co., and Chinese Academy of Sciences' Institute of Automation for sponsoring this conference. Special thanks are due to Ran He, Xiaobo Zhang, Hui Zhang, Man Zhang, and Xingguang Li for their hard work in the conference organization.

September 2011

Zhenan Sun
Jianhuang Lai
Xilin Chen
Tieniu Tan

Organization

General Chairs

Tieniu Tan	Institute of Automation, Chinese Academy of Sciences, China
Anil K. Jain	Michigan State University, USA
Jingyu Yang	Nanjing University of Science and Technology, China

Program Chairs

Xilin Chen	Institute of Computing Technology, Chinese Academy of Sciences, China
Jianhuang Lai	Sun Yat-sen University, China
Zhenan Sun	Institute of Automation, Chinese Academy of Sciences, China

Program Committee

Chengjun Liu	New Jersey Institute of Technology, USA
Gang Hua	IBM Watson Research Center, USA
Qiang Ji	Rensselaer Polytechnic Institute, USA
Wen-Yi Zhao	Sarnoff Corporation, USA
Jaihie Kim	Yonsei University, Republic of Korea
Suthep Madarasmi	KMUT, Thailand
Xudong Jiang	Nanyang Technological University, Singapore
Karthik Nandakumar	I2R, Singapore
Koichiro Niinuma	Fujitsu Labs, Japan
Xiaoqing Ding	Tsinghua University, China
Shiqi Yu	Shenzhen University, China
Shiguang Shan	Institute of Computing Technology, Chinese Academy of Sciences, China
Li Ma	Beijing Irisking Ltd. Co., China
Xiangwei Kong	Dalian University of Technology, China
Yilong Yin	Shandong University, China
Yuchun Fang	Shanghai University, China
Kuanquan Wang	Harbin Institute of Technology, China
Yunhong Wang	Beijing University of Aeronautics and Astronautics, China

Zhaoyang Lu	Xi'an University of Electronic Science and Technology, China
Qiuqi Ruan	Beijing Jiaotong University, China
Xihong Wu	Peking University, China
Changshui Zhang	Tsinghua University, China
Zhaoxiang Zhang	Beijing University of Aeronautics and Astronautics, China
Stan Z. Li	Institute of Automation, Chinese Academy of Sciences, China
Wenxin Li	Peking University, China
Jie Yang	Shanghai Jiao Tong University, China
Jian Yang	Nanjing University of Science and Technology, China
Yingchun Yang	Zhejiang University, China
Xin Yang	Institute of Automation, Chinese Academy of Sciences, China
Jinfeng Yang	Civil Aviation University of China
Zengfu Wang	Institute of Intelligent Machines, Chinese Academy of Sciences, China
Guangda Su	Tsinghua University, China
Jie Zhou	Tsinghua University, China
Zongying Ou	Dalian University of Technology, China
Wei-shi Zheng	Sun Yat-sen University, China
Jufu Feng	Peking University, China
Dewen Hu	National University of Defense Technology, China
Daoliang Tan	Beijing University of Aeronautics and Astronautics, China
Zhichun Mu	University of Science and Technology Beijing, China
Weiqi Yuan	Shenyang University of Technology, China

Organizing Committee Chair

Ran He	Institute of Automation, Chinese Academy of Sciences, China

Organizing Committee Members

Xiaobo Zhang	Institute of Automation, Chinese Academy of Sciences, China
Hui Zhang	Institute of Automation, Chinese Academy of Sciences, China
Yunqi Tang	Institute of Automation, Chinese Academy of Sciences, China

Zhenhua Chai	Institute of Automation, Chinese Academy of Sciences, China
Man Zhang	Institute of Automation, Chinese Academy of Sciences, China
Xingguang Li	Institute of Automation, Chinese Academy of Sciences, China
Lihu Xiao	Institute of Automation, Chinese Academy of Sciences, China
Haiqing Li	Institute of Automation, Chinese Academy of Sciences, China
Jing Liu	Institute of Automation, Chinese Academy of Sciences, China
Libin Wang	Institute of Automation, Chinese Academy of Sciences, China
Jianwei Yu	Institute of Automation, Chinese Academy of Sciences, China
Dong Wang	Institute of Automation, Chinese Academy of Sciences, China

Table of Contents

Face

Iris

Hand

Behavioral Biometrics

Soft Biometrics

Security

Other Biometrics

Video-Based Face Recognition: State of the Art

Zhaoxiang Zhang, Chao Wang, and Yunhong Wang

Laboratory of Intelligent Recognition and Image Processing,
Beijing Key Laboratory of Digital Media,
School of Computer Science and Engineering, Beihang University, Beijing, China
{zxzhang,yhwang}@buaa.edu.cn, ableblaze@163.com

Abstract. Face recognition in videos is a hot topic in computer vision and biometrics over many years. Compared to traditional face analysis, video based face recognition has advantages of more abundant information to improve accuracy and robustness, but also suffers from large scale variations, low quality of facial images, illumination changes, pose variations and occlusions. Related to applications, we divide the existing video based face recognition approaches into two categories: video-image based methods and video-video based methods, which are surveyed and analyzed in this paper.

Keywords: Face recognition, video, survey.

1 Introduction

In recent years, face recognition is always an active topic in the field of biometrics. Compared to traditional face recognition in still images, video based face recognition has great advantages listed as follows. Firstly, videos contain more abundant information than a single image. As a result, more robust and stable recognition can be achieved by fusing information of multi frames. Secondly, temporal information becomes available to be exploited in videos to improve the accuracy of face recognition. Finally, multi poses of faces in videos make it possible to explore shape information of face and combined into the framework of face recognition. However, video based face recognition is also a very challenging problem, which suffers from low quality facial images, illumination changes, pose variations, occlusions and so on.

Due to its importance and difficulties, many research institutes have focused on video based face recognition with all kinds of approaches proposed, such as Massachusetts Institute of Technology [1] , Carnegie Mellon University [2,3], University of Illinois at Urbana-Champaign [4,5], University of Maryland [6,7,8], University of Cambridge [9,10,11], Toshiba [12,13], Institute of Automation Chinese Academy of Sciences [14,15].

The whole procedure of video based face recognition is shown in Fig. 1. Related to applications, we can divide video based face recognition methods into two categories: video-image based methods and video-video based methods. The first category can be seen as an extension of still image based face recognition. The

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 1–9, 2011.

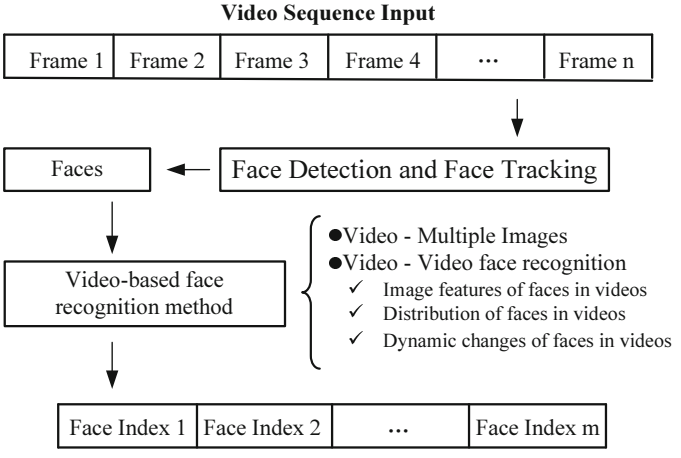


Fig. 1. Process of face recognition in video

second category is more complicated with more abundant solutions proposed. In the following, we will describe both of the categories in detail.

2 Video-Image Face Recognition

Video-image face recognition can be seen as an extension of still image based face recognition. The input of the system is videos while the database are still face images. Compared to traditional still image based face recognition, how to explore the multi-frame information of the input video is the key to enhance the performance.

One solution is based on frame selection. Faces in videos are tracked by tracking algorithms and those high quality face images of better resolution, pose and clarity are selected for matching based on still image based methods. Eigenfaces [16] and fisherfaces [17] are two of the essential basic techniques in this category. In [18], Satoh proposed a straightforward extension of the traditional eigenface and fisherface methods, by introducing a new similarity measure for matching video data. The distance between videos was calculated by considering the smallest distance between frame pairs (one from each video), in the reduced feature space. Raffaella Lanzarotti [19] extracted the 24 facial feature points from the color image, then extracted the feature of the points by Gabor, and the recognition the faces by compare the similarity of points' feature.

Another strategy is based on multi-frame fusion. Rainer Stiefelhagen et al. [20] made use of 3 different metric model to fuse recognition results of different frames. Pascal Frossard et al. [21] adopted the semi-supervised learning and graph theory to convert the problem of face recognition to an optimization task. To deal with occlusions, J. Aggarwal et al. [22] adopted patch information from multi-frames to obtain a complete face model for matching. In addition, Jae Young Choi et al. [23] made use of low resolution subspace of high resolution

image sets to deal with problem of different resolutions of input videos and databases.

In summary, image-video based methods make use of multi-frame information to improve the accuracy of face recognition, and improve the robustness to deal with pose variations, occlusions and illumination changes.

3 Video-Video Face Recognition

Compared to video-image based methods, both the system input and the database in this category are in the form of videos, which is a more difficult problem to solve. Based on the state of the arts, there are mainly three types of solutions of this problem, which are listed as follows:

1. Based on feature vector extracted from video input;
2. Based on probability density function or manifold to depicts the distribution of faces in videos;
3. Based on generative models to describe dynamic variance of face in images.

In the following, we will survey and analyze the existing methods of these three solutions.

3.1 Feature Vector Based Methods

The basic idea of this solution is to extract feature vectors from input videos, which are used to match with all the videos in the database.

Horst Eidenberger [24] proposed a Kalman filter-based method to describe invariant face features in videos based on a compact vision model, which achieves high performance in the UMIST database. Park et al. [25] created multiple face templates for each class of face in database according to the video information, dynamic fuse the multiple templates as the feature. Lapedriza et al. [26] builded PCA feature subspaces for the faces of the input and database, which achieves recognition based on the distance between the subspace measured by geometric angles.

In order to remove the effect of light, gesture and facial expressions, Fukui and Yamaguchi [12] projected the feature space to the constraint subspace. The disadvantages of this method are that they ignore the global probable distribution of each category, and parameters are based experiences or experiments. Ajmal Mian [27] proposed a unsupervised video-based method. Faces from a video sequence are automatically clustered based on the similarity of their local features and the identity is decided on the basis of best temporally cohesive cluster matches.

Since the pose, illumination and expression of face are non-linear in feature space, Wolf and Shashua [28] mapped the feature space into the kernel Hilbert space, combined with Support Vector Machine (SVM) to improve recognition performance. Fan et al. [29] extracted the geodesic distance as the feature to depict the position relations in the manifold space, and use HAC (Hierarchical

Agglomerative Clustering) to obtain K samples. Similar to [30], dual subspace probabilistic model is obtained as similarity score to recognize face.

This kind of solution makes use of multi frames of the input video to obtain a discriminant feature representation. However, the spatial information of input videos is neglected, which limits the performance of feature vector based approaches.

3.2 Distribution of Faces in Videos

The main idea of this category is to treat faces in videos as random variables of certain probability. The similarity of faces are measured by similarity of corresponding probability density distributions.

In [9], each faces in the database and input video are modeled with GMM, and Kullback-Leibler divergence is measured as the similarity measurement to achieve recognition. Arandjelovi'c et al. [10] made use of kernel-based methods to map low-dimensional space to high dimensional space, and then use low-dimensional space of linear methods (such as PCA) to solve complex nonlinear problems in the high-dimensional space. Zhou et al. [31] mapped the vector space into RKHS (Rep reducing Kernel Hilbert Space) by kernel based methods to calculate the distance between the probability distribution.

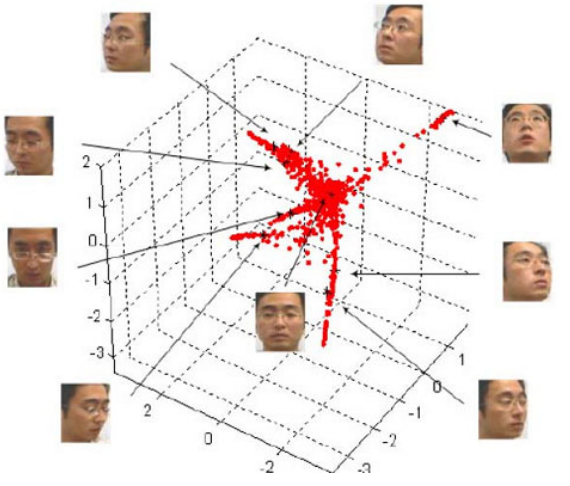


Fig. 2. LLE applied to a sequence of face images corresponding to a single person arbitrarily rotating his head [32]

In [33,34,35], a multi-view dynamic face model is built to achieve face recognition. Firstly, dynamic face model are constructed including a 3D model, a texture model and an affine change model. Then a Kalman filter is adopted to obtain the shape and texture, which builds a segmented linear manifold for

each single person with the face texture reduced by KDA (Kernel Discriminate Analysis). Face recognition is achieved in the following by trajectory matching. However, the 3D model estimation requires a lot of multi-angle images and a larger complexity computational.

Wang et al. [36,32] proposed a incremental online learning model for face recognition in videos. This method use a defined face model to learn new face contour model online, then use a linear model and feature space transformation matrix to generate the face manifold, as shown in Fig. 2. Minyoung Kim et al. [37] integrated face tracking and recognition and employed some priori knowledge. They obtain 70% recognition rate in YouTube and 100% in Honda/UCSD database [4,5].

This kind of solution is much better than the feature vector based solution, which makes use of probability theory to enhance the performance. However, the dynamic change information of faces in videos is neglected, which has potential to improve the video based face recognition.

3.3 Dynamic Changes of Faces in Videos

The temporal information in video sequences enables the analysis of facial dynamic changes and its application as a biometric identifier for person recognition [38].

Matta et al. proposed a multi-modal recognition system [39,40]. They successfully integrated the facial motion information with mouth motion and facial appearance by taking advantage of a unified probabilistic framework.

Some strategies have been developed to integrate tracking and recognition into a single framework. Lee et al. [5] developed the probabilistic appearance manifold approach for tracking and recognition using video sequences. Bayesian inference was employed to include the temporal coherence of human motion in the distance calculation. And they replaced the conditional probability by using the joint conditional probabilities, which were recursively estimated using the transitions between sub-manifolds. In [41], Matta and Dugelay presented a person recognition system that exploited the unconstrained head motion information extracted by tracking a few facial landmarks in the image plane. In [42], Huang and Trivedi developed a face recognition system by employing HMMs for facial dynamic information modeling in videos. Each covariance matrix was gradually adapted from a global diagonal one by using its class-dependent data in training algorithms. Afterwards, Liu and Cheng [3] successfully applied HMMs for temporal video recognition (as illustrated in Fig. 3) by improving the basic implementation of Huang and Trivedi. Each test sequence was used to update the model parameters of the client in question by applying a maximum a posteriori (MAP) adaptation technique.

In summary, the three main solutions of video-video based face recognition is surveyed and analyzed in this section. We think how to fuse all these solutions should be the potential direction of video based face recognition.

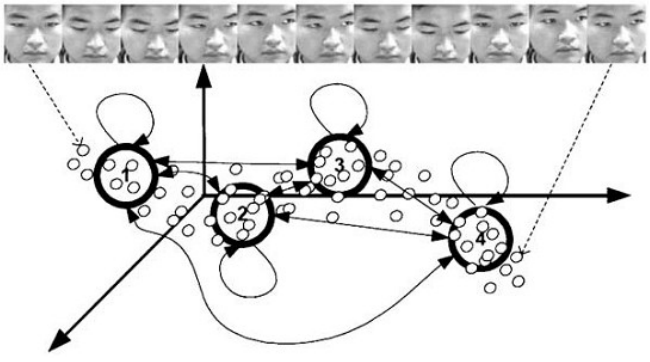


Fig. 3. Example of a hidden Markov model temporally applied to video sequences [3]

4 Video Face Database

It is recommended to use a standard test dataset to benchmark an algorithm. The Honda/UCSD [4,5] database includes 20 individuals moving their heads in different combinations of 2-D and 3-D rotation, expression and speed. The CMU Motion of Body (MoBo) [43] database contains 25 individuals walking on a treadmill in the CMU 3D room. The subjects perform four different walk patterns: slow walk, fast walk, incline walk and walking with a ball. All subjects are captured using six high resolution color cameras distributed evenly around the treadmill. The video based face recognition databases have common drawbacks that they are always fixed in the same scene. Most of the changes are pose changes while lighting and facial expressions are less considered. Tab. 1 shows the experimental results of some typical methods of video-based face recognition on the database.

Table 1. Comparison against similar system in the literature

Literature	Approaches	Accuracy
Zhou et al. [31]	PCA and majority voting	87.1%(MoBo) 89.6%(Honda/ UCSD)
Zhou et al. [31]	LDA and majority voting	90.8%(MoBo) 86.5%(Honda/ UCSD)
Liu et al. [3]	Video level matching using HMM	98.8%(Honda/ UCSD)
Zhou et al. [6]	SIS base on Bayesian framework	92%(MoBo)
Zhou et al. [5]	Frame and video level matching using statistical models	88% $\sim$ 100% (Honda/ UCSD)
Lee et al. [4]	Matching frames using appearance manifolds	93.4% (Honda/ UCSD)
Aggarwal et al. [8]	Video level matching using ARAM	92.1% (Honda/ UCSD)
Ajmal Mian et al. [27]	Local feature clutering	99.5% (Honda/ UCSD)

5 Conclusion

In this article we proposed a detailed state of the art on video-based face recognition. Two categories of video based face recognition methods are surveyed and analyzed. We can see that video-image based methods only exploit physiological information of the face while the video-video based methods have more information to be exploited. It is evident that video based face recognition has great potential to make progress and be adopted in real application.

Acknowledgement. This work is funded by the National Basic Research Program of China (No. 2010CB327902), the National Natural Science Foundation of China (No. 60873158, No. 61005016, No. 61061130560) and the Fundamental Research Funds for the Central Universities.

References

1. Shakhnarovich, G., Fisher, J.W., Darrell, T.: Face Recognition from Long-Term Observations. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part III. LNCS, vol. 2352, pp. 851–865. Springer, Heidelberg (2002)
2. Liu, X., Chen, T., Thornton, S.M.: Eigenspace updating for non-stationary process and its application to face recognition. Pattern Recognition 36(9), 1945–1959 (2003)
3. Liu, X., Cheng, T.: Video-based face recognition using adaptive hidden markov models. In: Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. I–340. IEEE (2003)
4. Lee, K.C., Ho, J., Yang, M.H., Kriegman, D.: Video-based face recognition using probabilistic appearance manifolds. In: Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. I–313. IEEE (2003)
5. Lee, K.C., Ho, J., Yang, M.H., Kriegman, D.: Visual tracking and recognition using probabilistic appearance manifolds. Computer Vision and Image Understanding 99, 303–331 (2005)
6. Zhou, S., Krueger, V., Chellappa, R.: Probabilistic recognition of human faces from video q. Computer Vision and Image Understanding 91, 214–245 (2003)
7. Zhou, S.K., Chellappa, R., Moghaddam, B.: Visual tracking and recognition using appearance-adaptive models in particle filters. IEEE Transactions on Image Processing 13(11), 1491–1506 (2004)
8. Aggarwal, G., Chowdhury, A.K.R., Chellappa, R.: A system identification approach for video-based face recognition. In: Proceedings of 17th International Conference on Pattern Recognition, ICPR 2004, vol. 4, pp. 175–178. IEEE Computer Society (2004)
9. Arandjelović, O., Cipolla, R.: Face recognition from face motion manifolds using robust kernel resistor-average distance. In: Conference on Computer Vision and Pattern Recognition Workshop, CVPRW 2004, p. 88. IEEE (2004)
10. Arandjelović, O., Shakhnarovich, G., Fisher, J., Cipolla, R., Darrell, T.: Face recognition with image sets using manifold density divergence (2005)
11. Cipolla, R., Arandjelović, O.: A pose-wise linear illumination manifold model for face recognition using video. Computer Vision and Image Understanding 113(1), 113–125 (2009)

12. Fukui, K., Yamaguchi, O.: Face recognition using multi-viewpoint patterns for robot vision. In: Robotics Research, pp. 192–201 (2005)
13. Nishiyama, M., Yamaguchi, O., Fukui, K.: Face Recognition with the Multiple Constrained Mutual Subspace Method. In: Kanade, T., Jain, A., Ratha, N.K. (eds.) AVBPA 2005. LNCS, vol. 3546, pp. 71–80. Springer, Heidelberg (2005)
14. Li, J., Wang, Y., Tan, T.: Video-Based Face Recognition Using A Metric of Average Euclidean Distance. In: Li, S.Z., Lai, J.-H., Tan, T., Feng, G.-C., Wang, Y. (eds.) SINOBIOMETRICS 2004. LNCS, vol. 3338, pp. 224–232. Springer, Heidelberg (2004)
15. Li, J., Wang, Y., Tan, T.: Video-Based Face Recognition Using Earth Mover's Distance. In: Kanade, T., Jain, A., Ratha, N.K. (eds.) AVBPA 2005. LNCS, vol. 3546, pp. 229–238. Springer, Heidelberg (2005)
16. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 1991, pp. 586–591. IEEE (1991)
17. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 711–720 (1997)
18. Satoh, S.: Comparative evaluation of face sequence matching for content-based video access. In: Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000, pp. 163–168. IEEE (2000)
19. Arca, S., Campadelli, P., Lanzarotti, R.: A face recognition system based on automatically determined facial fiducial points. Pattern Recognition 39(3), 432–443 (2006)
20. Stallkamp, J., Ekenel, H.K., Stiefelhagen, R.: Video-based face recognition on real-world data. In: IEEE 11th International Conference on Computer Vision, ICCV 2007, pp. 1–8. IEEE (2007)
21. Kokiopoulou, E., Frossard, P.: Video face recognition with graph-based semi-supervised learning. In: IEEE International Conference on Multimedia and Expo, ICME 2009, pp. 1564–1565. IEEE (2009)
22. Hu, C., Harguess, J., Aggarwal, J.K.: Patch-based face recognition from video. In: 2009 16th IEEE International Conference on Image Processing (ICIP), pp. 3321–3324. IEEE (2009)
23. Choi, J.Y., Ro, Y.M., Plataniotis, K.N.: Feature subspace determination in video-based mismatched face recognition. In: 8th IEEE International Conference on Automatic Face & Gesture Recognition, FG 2008, pp. 1–6. IEEE (2008)
24. Eidenberger, H.: Kalman filtering for pose-invariant face recognition. In: 2006 IEEE International Conference on Image Processing, pp. 2037–2040. IEEE (2006)
25. Park, U., Jain, A.K., Ross, A.: Face recognition in video: Adaptive fusion of multiple matchers. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE (2007)
26. Lapedriza, A., Masip, D., Vitria, J.: Face verification using external features. In: 7th International Conference on Automatic Face and Gesture Recognition, FGR 2006, pp. 132–137. IEEE (2006)
27. Mian, A.: Unsupervised learning from local features for video-based face recognition. In: 8th IEEE International Conference on Automatic Face & Gesture Recognition, FG 2008, pp. 1–6. IEEE (2008)
28. Wolf, L., Shashua, A.: Kernel principal angles for classification machines with applications to image sequence interpretation (2003)
29. Fan, W., Yeung, D.Y.: Locally linear models on face appearance manifolds with application to dual-subspace based classification (2006)

30. Moghaddam, B., Jebara, T., Pentland, A.: Bayesian face recognition. Pattern Recognition 33(11), 1771–1782 (2000)
31. Zhou, S.K., Chellappa, R.: From sample similarity to ensemble similarity: Probabilistic distance measures in reproducing kernel hilbert space. IEEE transactions on pattern analysis and machine intelligence, 917–929 (2006)
32. Fan, W., Wang, Y., Tan, T.: Video-Based Face Recognition Using Bayesian Inference Model. In: Kanade, T., Jain, A., Ratha, N.K. (eds.) AVBPA 2005. LNCS, vol. 3546, pp. 122–130. Springer, Heidelberg (2005)
33. Li, Y., Gong, S., Liddell, H.: Modelling faces dynamically across views and over time. In: Proceedings of Eighth IEEE International Conference on Computer Vision, ICCV 2001, vol. 1, pp. 554–559. IEEE (2001)
34. Li, Y., Gong, S., Liddell, H.: Video-based online face recognition using identity surfaces. ratfg-rts, 40 (2001)
35. Li, Y., Gong, S., Liddell, H.: Constructing facial identity surfaces in a nonlinear discriminating space (2001)
36. Liu, L., Wang, Y., Tan, T.: Online appearance model learning for video-based face recognition. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–7. IEEE (2007)
37. Kim, M., Kumar, S., Pavlovic, V., Rowley, H.: Face tracking and recognition with visual constraints in real-world videos. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–8. IEEE (2008)
38. Matta, F., Dugelay, J.L.: Person recognition using facial video information: A state of the art. Journal of Visual Languages & Computing 20(3), 180–187 (2009)
39. Saeed, U., Matta, F., Dugelay, J.L.: Person recognition based on head and mouth dynamics. In: 2006 IEEE 8th Workshop on Multimedia Signal Processing, pp. 29–32. IEEE (2006)
40. Matta, F., Dugelay, J.L.: Video face recognition: a physiological and behavioural multimodal approach. In: IEEE International Conference on Image Processing, 2007. ICIP 2007, vol. 4, pp. IV–497. IEEE (2007)
41. Matta, F., Dugelay, J.-L.: Person recognition using human head motion information. In: Perales, F.J., Fisher, R.B. (eds.) AMDO 2006. LNCS, vol. 4069, pp. 326–335. Springer, Heidelberg (2006)
42. Huang, K.S., Trivedi, M.M.: Streaming face recognition using multicamera video arrays. In: Proceedings of 16th International Conference on Pattern Recognition 2002, vol. 4, pp. 213–216. IEEE (2002)
43. Gross, R., Shi, J.: The cmu motion of body (mobo) database (2001)

Face Recognition with Directional Local Binary Patterns

Linlin Shen and Jinwen He

College of Computer Science & Software Engineering, Shenzhen University, Guangdong, Shenzhen, 518060
llshen@szu.edu.cn

Abstract. A novel local feature descriptor, namely Directional Local Binary Patterns (DLBP), was proposed in this paper and applied for face recognition. The descriptor first extracts directional edge information, then codes these information using Local Binary Patterns (LBP). When applied for face recognition, a face image is divided into a number of small sub-windows, DLBP histogram extracted from each sub-window are then concatenated to form a global description of the face. The proposed method was extensively evaluated on two publicly available databases, i.e. the FERET face database and the PolyU-NIRFD near-infrared face database. Experimental results show advantages of DLBP over LBP and Directional Binary Code (DBC).

Keywords: Directional Local Binary Patterns, Local Binary Patterns, Directional Binary Code, Face Recognition.

1 Introduction

As a primary modality for biometric authentication, face recognition has become a very active topic in pattern recognition and computer vision in recent years. There are a large number of commercial and secure applications requiring the use of face recognition technologies, such as access control and video surveillance.

When appearance based methods like Principal Component Analysis (PCA) [1] and Linear Discriminant Analysis (LDA) [2] extract features from the whole face area, local feature descriptors like Local Binary Patterns (LBP) [3], Directional Binary Code (DBC) [4], and Gabor features [10] etc. extract local features and are believed to be more robust against illumination and pose variations. LBP is firstly applied to face recognition in [5] and experimental results show that it can yield very good performance, due to its invariance to monotonic gray-level changes and computational efficiency. In recent years, LBP have been widely used in face analysis [6 7 8]. Inspired by the idea of LBP, DBC is a new efficient descriptor proposed in [4] to capture the directional edge information. Both texture information and edge one are important for face recognition. Nevertheless, LBP can obtain more rich texture information than DBC and DBC can acquire more edge information than LBP.

In this work, we propose a simple yet efficient descriptor, namely Directional Local Binary Patterns (DLBP), to capture both texture and directional edge information for face recognition. For DLBP, directional edge information is first extracted by a similar approach with DBC, which is followed by LBP operator for

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 10–16, 2011.

texture representation. To encode spatial information, face image is divided into a number of small sub-windows and DLBP histogram features of each sub-window are then concatenated to form a global description for face representation. Experimental results were reported using two publicly available databases, i.e. the FERET visible light face database [9] and the PolyU-NIRFD near-infrared face database [4]. The comparative results with LBP and DBC prove advantages of our proposed method.

The rest of this paper is organized as follows. In section 2, the DLBP method is proposed and introduced, which is followed by the description of classification in section 3. Section 4 presents experiments for evaluating the performances of different descriptors like LBP, DBC and DLBP for face recognition. Finally, conclusions are given in section 5.

2 Directional Local Binary Patterns (DLBP)

The LBP operator assigns a binary label to every pixel of an image by thresholding its 3×3 neighborhood with it, and represents the results with an 8-bits integer. As shown in Fig. 1, the local information around pixel with value five can be represented by a bit string '00111100'.

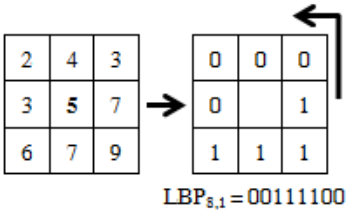


Fig. 1. The basic LBP operator

Similar to LBP, DBC encode local edge information by two steps as follows. Firstly, the center point and its eight neighbors are compared one by one with corresponding points located with certain distance along a direction, i.e. 0, 45, 90 or 145 degree direction. Secondly, according to the signs, the nine differences are thresholded to a binary label (0/1). Normally, a positive difference is set as 1, and a negative one is set as 0. The thresholded results are then concatenated as a nine-bits code and converted into an integer for feature representation. Fig.2 shows the calculation of DBC with distance 1 along 0 degree direction. Finally, the local edge information along horizontal direction is encoded with a bit string '111111011'.

To include both textures and edge information, the proposed DLBP operator combines DBC and LBP for feature extraction. When directional difference is first applied to extract directional edge information, following LBP operator codes the differences between central pixel and its eight neighbors. The directional edge information can thus be coded as a texture pattern. Fig.3 shows the calculation of DLBP with distance 1 along 0 degree direction. Based on the horizontal differences, the edge strength computed at each pixel is compared with its central pixel and coded as a bit string '00101101'. One can observe that there are two bits difference between DLBP code and LBP code, as edge information, instead of gray levels, are used.

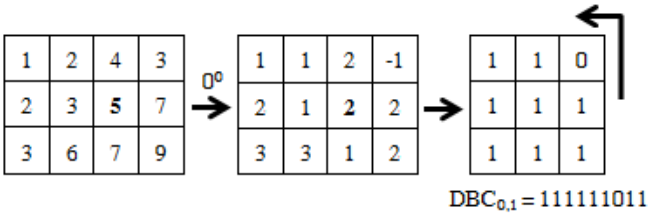


Fig. 2. Calculation of DBC with distance 1 along 0 degree direction

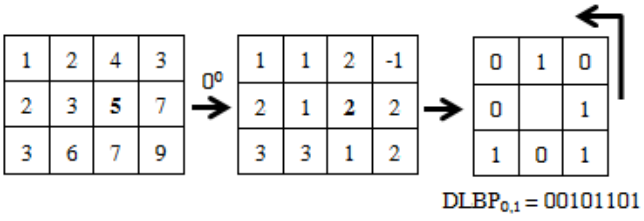


Fig. 3. Calculation of DLBP with distance 1 along 0 degree direction

Fig.4 shows the LBP, DBC and DLBP feature images calculated from an example face image. It can be seen that DLBP can extract more edge information than LBP and capture more texture information than DBC.

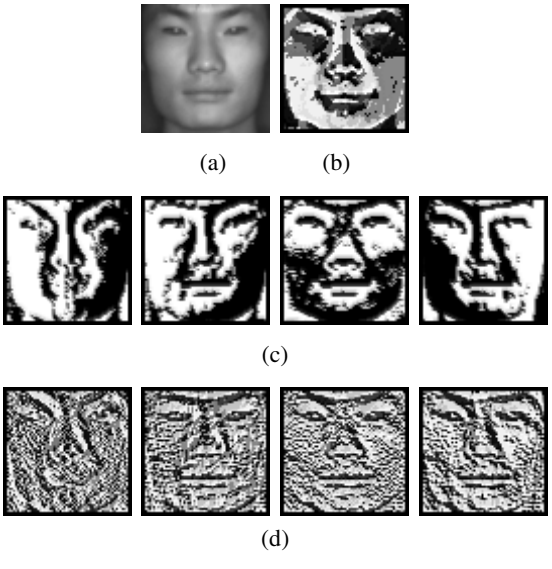


Fig. 4. Example of LBP, DBC and DLBP feature maps. (a) is the original image;(b) is the LBP feature map; (c) is the DBC feature maps along 0, 45, 90, 145 degree directions; (d) shows the DLBP feature maps along 0, 45, 90, 145 degree directions.

When four directions like 0, 45, 90 or 145 degree are used, the DLBP feature extraction process produces four DLBP feature maps. To include spatial information, each of the feature maps is first divided into a number of sub-windows, histograms are then calculated for each window and concatenated to form a global description. Fig. 5 shows the process of histogram extraction and concatenation. In implementation, a window is shifted with a preset step size to sample the face regions and normally windows with different sizes are used. Given a face image with size H × W, when window with S different sizes are used to sample the face image, the total number of windows can be calculated as below:

$$N = \sum_{i=1}^{S} \frac{H - h_i + \Delta h_i}{\Delta h_i} \cdot \frac{W - w_i + \Delta w_i}{\Delta w_i} \tag{1}$$

where h_i, w_i is the size of sub-window, and Δh_i, Δw_i is the step size in *i*th scale.

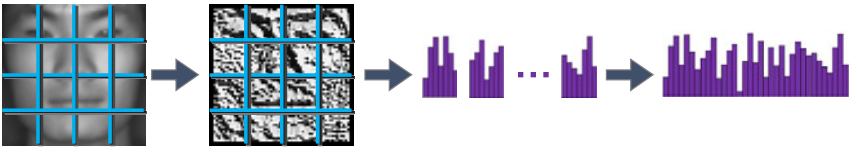


Fig. 5. The process of extracting features from a face image

3 Classification

Once two histograms H1 and H2 are extracted from two face images, Chi square statistic is adopted in this paper to measure the similarity between them:

$$\chi^2(H1, H2) = \sum_{i=1}^{n} \frac{(H1_i - H2_i)^2}{(H1_i + H2_i)} \tag{2}$$

where $H1_i$ and $H2_i$ are the value for *i*th bin of histograms H1 and H2, respectively, n is the bin number.

The nearest neighbor classifier is then used to classify a presented face image to the people whose face images gives the closest match.

4 Experiments

4.1 Results on the FERET Database

We first validate the effectiveness of the proposed DLBP method on the widely used FERET face database [9]. According to the evaluation protocol, a gallery of 1196 frontal face images and 4 different probe sets should be used for testing. The numbers of images in different probe sets are listed at Table 1, with example images shown in

Fig 6. Fb and Fc probe sets are used for assessing the effects of facial expression and illumination changes respectively. DupI and DupII consist of images taken on different days from their gallery images, particularly, there is at least a one year gap between the acquisition of the probe image in Dup II and the corresponding gallery image.

Table 1. List of different probe sets for FERET

Probe	Gallery	Probe size	Gallery size	Variations
Fb	Fa	1195	1196	Expression
Fc	Fa	194	1196	Illumination and Camera
Dup I	Fa	722	1196	Time gap < 1 week
Dup II	Fa	234	1196	Time gap > 1 year

Fig. 6. Example images of FERET database

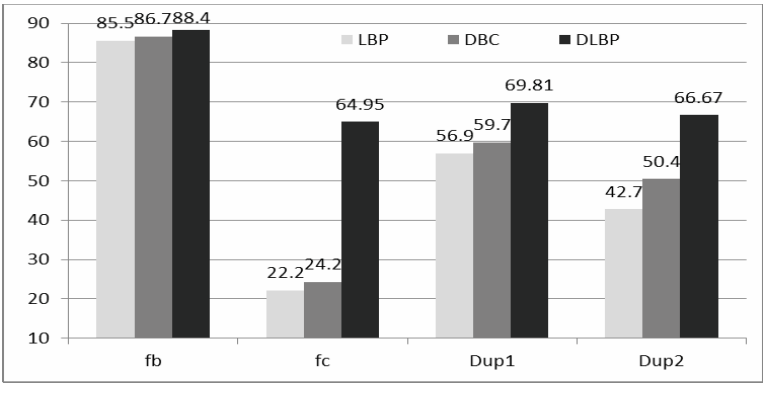


Fig. 7. Recognition rates of LBP, DBC and DLBP on FERET database

The facial parts of the images are first cropped based on the positions of the eyes and then normalized to size 64×64, then followed by histogram equalization for illumination correction. The circle (P, R) of LBP is assigned to (8, 2), and the distance of DBC and DLBP is assigned to 2. 225 small sub-windows with size 8×8 were generated by shifting the sub-window at step size 4×4, the number of histogram bins for LBP, DBC and DLBP are 59, 256 and 256, respectively. The parameters were tuned for best performances of different approaches. Fig.7. shows the performances of the three methods for different test sets. One can observe that DLBP achieves much

higher accuracy than LBP and DBC for all four test sets. DLBP captures both directional edge information and spatial relationship in a local region. While DBC captures the directional edge information only, LBP considers the relationship between a given pixel and its surrounding neighbors. Moreover, DLBP is more robust to illumination and time changes than LBP and DBC, especially to monotonic gray-level changes and gradual ones. The performance of LBP seems not satisfying, which as well was reported in [11] and [12].

4.2 Results on the PolyU-NIRFD Database

We are now using PolyU-NIRFD near-infrared face database [4] to test performance of the proposed DLBP method. This database contains near-infrared face images of 350 subjects, which include frontal face images as well as images with expression and pose variations, scale changes, and time difference, etc. Similar to the protocols used in [4], three experiments are designed here. When Exp#1 emphasizes expression variations and scale changes, Exp#2 focuses more on uncontrolled environment conditions and Exp#3 includes large pose variations. The number of images in these three experiments is shown in Table 2.

Table 2. Number of images in the three experiments

	Exp#1	Exp#2	Exp#3
Training set	419	1876	1876
Gallery set	574	1159	951
Probe set	2763	4747	3648

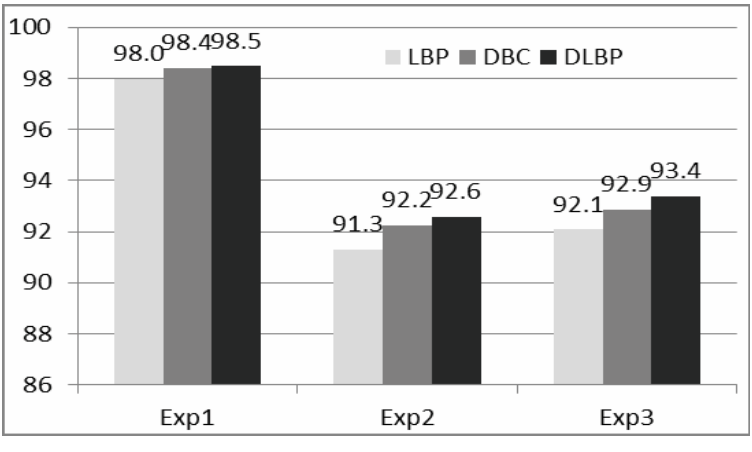


Fig. 8. Recognition rates of LBP, DBC and DLBP for PolyU database

Similarly, the facial parts of all these images are cropped according to the location of the eyes and then normalized to 64×64 pixels. The circle (P, R) of LBP was set as (8, 2), and the distance of DBC and DLBP was set to 2. Each of the facial parts is divided into 81 small sub-windows with size 16×16 and move at step size 6×6. Fig.8

shows the performances of LBP, DBC and the proposed DLBP. As shown in the figure, DLBP significantly outperforms LBP and DBC in all three experiments, which proves advantages of DLBP in near infrared face recognition.

5 Conclusion

A novel local feature descriptor, DLBP, has been proposed in this paper. By extracting both edge and texture information, DLBP is believed to be more representative than LBP and DBC in coding local patterns. The proposed method was successfully applied to face recognition and extensively tested using two publicaly available databases, i.e. FERET and PolyU- NIRFD databases. The experimental results on FERET database shows that DLBP is more robust against illumination and time differences than DBC and LBP, while the results on PolyU-NIRFD near-infrared database show that DLBP significantly outperforms LBP and DBC in all three experiments.

Acknowledgments. The work is supported by National Natural Science Foundation of China (60903112 and 60873168) and the Science Foundation of Shenzhen City (JC200903120074A).

References

1. Turk, M., Pentland, A.: Eigenfaces for Recognition. J. Cognitive Neuroscience 3(1), 71–86 (1991)
2. Etemad, K., Chellappa, R.: Discriminant Analysis for Recognition of Human Face Images. J. Optical Soc. Am. 14, 1724–1733 (1997)
3. Ojala, T., Pietikäinen, M., Harwood, D.: A Comparative Study of Texture Measures with Classification Based on Feature Distributions. Pattern Recognition 29(1), 51–59 (1996)
4. Zhang, B.C., Zhang, L., Zhang, D., Shen, L.L.: Directional Binary Code with Application to PolyU Near-Infrared Face Database. Pattern Recognition Letters 31, 2337–2344 (2010)
5. Ahonen, T., Hadid, A., Pietikainen, M.: Face Recognition with Local Binary Patterns. Proc. European Conf. Computer Vision, 469–481 (2004)
6. Hadid, A., Pietikäinen, M., Ahonen, T.: A Discriminative Feature Space for Detecting and Recognizing Faces. Proc. IEEE CS Conf. Computer Vision and Pattern Recognition 2, 797–804 (2004)
7. Feng, X., Pietikäinen, M., Hadid, A.: Facial Expression Recognition with Local Binary Patterns and Linear Programming. Pattern Recognition and Image Analysis 15(2), 546–548 (2005)
8. Li, S.Z., Chu, R., Ao, M., Zhang, L., He, R.: Highly Accurate and Fast Face Recognition Using Near Infrared Images. Proc. Int'l Conf. Advances in Biometrics, 151–158 (2006)
9. Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.: The FERET Database and Evaluation Procedure for Face Recognition Algorithms. Image and Vision Computing 16(10), 295–306 (1998)
10. Shen, L., Bai, L.: A review on Gabor wavelets for face recognition. Pattern Analysis and Applications 9(2), 273–292 (2006)
11. Liu, Z., Yang, J., Liu, C.: Extracting Multiple Features in the CID Color Space for Face Recognition. IEEE Transactions on Image Processing 19(9), 2502–2509 (2010)
12. Liu, Z., Liu, C.: Fusion of Color, Local Spatial and Global Frequency Information for Face Recognition. Pattern Recognition 43(8), 2882–2890 (2010)

A Novel Feature Extraction Method for Face Recognition under Different Lighting Conditions

Jianjun Qian and Jian Yang

School of Computer Science and Technology,
Nanjing University of Science and Technology,
Nanjing, P.R. China
qjjtx@126.com, csjyang@mail.njust.edu.cn

Abstract. This paper develops a novel method named image decomposition based on locally adaptive regression kernels (ID-LARK) for feature extraction. ID-LARK is robust to variations of illumination, since it decomposes the local features into different sub-images. And they describe the structure information hidden in the unobserved space. More specially, ID-LARK first exploits local structure information by measuring geodesic distance between the central pixel and its neighbors in the local window with locally adaptive regression kernels. So, one image can be decomposed into several sub-images (structure images) according to the local feature vector of each pixel. We thus downsample every structure images and concatenate them to obtain the augmented feature vector. Finally, fisher linear discriminant analysis is used to provide powerful discriminative ID-LARK feature vector. The proposed method ID-LARK is evaluated using the Extended Yale B and CMU PIE face image databases. Experimental results show the significant advantages of our method over the state-of-art ones.

Keywords: feature extraction, locally adaptive regression kernels, image decomposition, face recognition.

1 Introduction

Face recognition has attracted significant attention due to its wide range of applications in information security, law enforcement and surveillance [1]. In the past several decades, numerous face representation approaches and much progress have been made, including subspace based global features and local appearance features. Among them, principal component analysis (PCA) and fisher linear discriminant analysis (FLDA) are two well-known linear subspace learning methods which have been widely used in pattern recognition and computer vision areas and have become the most popular techniques for face recognition [2]. Recently, Yan et al. [3] proposed a general framework called graph embedding, and various methods such as PCA, ISOMAP, LLE and LPP etc, can be reformulated as a unified model in this framework.

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 17–24, 2011.

Compared with the global features like PCA and FLDA, local appearance features are more robust to handle the local changes such as illumination, expression and pose. Local binary pattern (LBP) operator, which captures spatial structure of local image texture, is one of the best texture descriptors. And it not only has been widely used in texture image classification, but also successfully applied in face recognition and verification [5]. Gabor wavelet used in face recognition has attracted the attention of many scholars in the past few years [4]. The Gabor feature extracts characteristic visual properties such as local direction character at multi-scale. Therefore, it is used for feature extraction can effectively improve the performance of recognition tasks. In addition, some works combined LBP and Gabor feature to improve the face recognition performance effectively in contrast with individual representation [6, 7].

Recently, H. J. Seo et.al proposed a novel visual descriptor called locally adaptive regression kernels (LARK) [8] motivated by the early work [9]. LARK essentially describes the local structure information by measuring a self-similarity between a central pixel and its surrounding pixels. Meanwhile, they also compared LARK with the sate-of-the-art local descriptors evaluated in the paper [10]. Experiment verified that their LARK descriptors give more discriminative power than others. And LARK descriptors were also developed in other areas such as face verification, action recognition and static and space-time saliency detection [11, 12].

The LARK method, however, may fail make full use of entire local information of different orientations. It's only preserves the characteristic features of each pixel. In this paper, we present a method (ID-LARK) which can decompose an image according to all of the local information including significant features and weak features to describe image structure information in different orientations. The overview of the proposed method is illustrated in Fig 1.

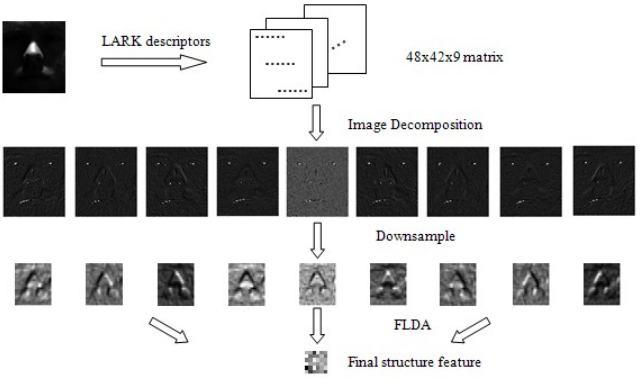


Fig. 1. The overview of ID-LARK

2 Outline of Locally Adaptive Regression Kernels Descriptor

LARK effectively captures local structure information of images by analyzing the radiometric differences based on estimated gradient, which also determine the shape

and size of canonical kernels. The locally adaptive regression kernel is modeled as follows [8]

$$K(x_l - x_i) = \frac{\sqrt{\det(C_l)}}{2\pi h^2} \exp\{-\frac{(x_l - x_i)^T C_l (x_l - x_i)}{2h^2}\} \tag{1}$$

where h is a global smooth parameter, and $l \in (1 \cdots p)$, p is the number of pixel in the local window. The covariance matrix C_l is computed from a collection of spatial gradient vectors with the local analysis window around the center position. Moreover, this matrix can be first estimated as $M_l^T M_l$ with:

$$M_l = \begin{bmatrix} Z_{x1}(x_1)\ Z_{x2}(x_1) \\ \vdots \quad \vdots \\ Z_{x1}(x_P) Z_{x2}(x_P) \end{bmatrix} \tag{2}$$

where $Z_{x1}(\cdot)$ and $Z_{x2}(\cdot)$ are the first derivatives along the vertical and horizon respectively. The more stable matrix C_l can be estimated by invoking singular value decomposition (SVD) of M_l for the sake of robustness [11].

$$C_l = \gamma \sum_{q=1}^{2} a_q^2 v_q v_q^T \in R^{2\times 2} \tag{3}$$

with

$$a_1 = \frac{s_1 + \lambda'}{s_2 + \lambda'}, a_2 = \frac{s_2 + \lambda'}{s_2 + \lambda'}, \gamma = (\frac{s_1 s_2 + \lambda''}{P})^{\alpha} \tag{4}$$

where λ', λ'' are set to 1 and 10^{-7} respectively in our experiments. And α is set to 0.5 (fixed for all experiments in this paper) that restricts γ. The singular values (s_1, s_2) and the singular vectors (v_1, v_2) are achieved by evaluating the SVD of M_l [11].

Subsequently, we will essentially apply the formula (1) as a function of x_l to represent local structure around the position x_l. Specially, local features of each patch is densely calculated and normalized as follows [8]:

$$W_j(x_l - x) = \frac{K_j(x_l - x)}{\sum_{i=1}^{p} K_j(x_i - x)}, \begin{cases} j = 1, \ldots n \\ i = 1, \ldots p \end{cases} \tag{5}$$

where n is the number of patches in the image, and p is the number of pixels in the local window.

For the sake of organizing dense LARK features evaluated from the image. Suppose W is a matrix whose columns are column stacked version of $W_j(x_l - x)$.

$$W = [w_1, \cdots w_n] \in IR^{p\times n} \tag{6}$$

The next step, characteristics of LARKs can be obtained by employing PCA for dimensionality reduction. The top d principal components form the transformation matrix A. Then, the lower dimension feature F computed as follows:

$$F = [f^1, \cdots, f^n] = A^T W \in IR^{d \times n} \tag{7}$$

The whole procedure of LARK is shown in Fig 2. For extensive details on this section, we refer the reader to [8, 9].

3 Image Decomposition Based on Locally Adaptive Regression Kernels Descriptors (ID-LARK)

3.1 Extracting Local Structure Features

LARK is successfully applied in object detection and action recognition as visual descriptors owning to it describes the local structure information well. It measures geodesic distance between central pixel and its neighbors around it with local adaptive regression kernel. Thus, local structure feature of each pixel is obtained by normalizing it as shown in formula (5). The whole feature vector W is achieved by using formula (6). The difference of this paper, we won't use PCA to reduce the feature vector dimension of each pixel so as to preserve entire local structure information for image decomposition. In other words, we not only pay more attention to the characteristic features of each pixel, but also concern the weak structure information in this way. Therefore, one can get the completed local structure features of an image.

3.2 Image Decomposition and Feature Representation

After obtaining the local structure information, we can decompose an image into several sub-images according to the corresponding local feature vectors of each pixel. Specially, the local structure features $W = (W_1, W_2 \cdots W_N)$ of one image are achieved by using the procedure described in section 3.1. There is a Q-D (the number of dimension is Q) vector in one pixel. Therefore, one image is dissembled to Q new sub-images as shown in Fig 1 (experiment on the Extend Yale B database). Each of them depicts structure information from different orientations. Since the local window size of LARK is 3×3 in our experiments, so there is only one sub-image in one direction. In this way, we believe one sub-image (also called structure image) accumulate almost all the structure information in one direction.

In order to encompass different structure images, we concatenate all these images and derive an augmented feature vector g . Before the concatenation, we first downsample each structure images by factor λ (In our experiments λ is set 2) to reduce feature dimension and normalize it to zero mean and unit variance. We thus concatenate rows (or columns) of each structure image to form a feature vector v_q ($q = (1, \cdots Q)$). Subsequently, let v_q^{λ} represents the normalized vector constructed from v_q , the augmented feature g^{λ} is defined as follows:

$$g^{\lambda} = (v_1^{\lambda}, v_2^{\lambda}, \cdots, v_Q^{\lambda}) \tag{8}$$

Nevertheless, the augmented feature is still in a space of very high dimensionality. Numerous researches in recent years indicate that dimensionality reduction techniques seek to find a meaningful low dimensional subspace in a higher input data space. The subspace can provide a compact representation of higher dimensional data. So, the final feature vector G is obtained by using the fisher linear discriminant analysis to reduce the dimension of augmented feature vector. Actually, this point is similar with Gabor feature [8].

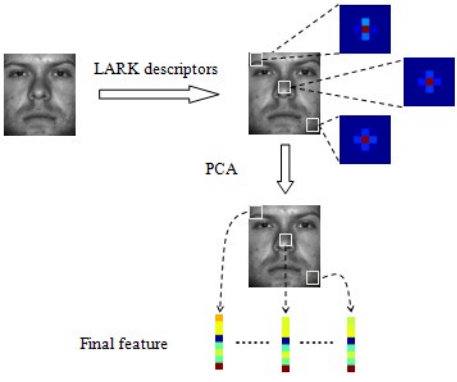


Fig. 2. The overview of the LARK

3.3 Relationship to LARK

Compared with LARK, ID-LARK decomposes an image into structure images according to the complete local structure information of each pixel. Further, low dimensional and compact ID-LARK features are obtained by using FLDA. The difference between ID-LARK and LARK are also shown in Fig 1 and Fig 2. In addition, we know that LARK describes dense structure information in their local window. In general, the local window size is also set to 5×5, 7×7 and 9×9. However, why we set the local window size to 3×3? On account of some significant structure information is distributed to various sub-images when using the larger local window size. As a result, each of them perhaps has little structure information. In contrast, ID-LARK not only weaken the lighting effect, but also ensures that one sub-image accumulate almost all the structure information in one direction at least, on the basis of the local window size is 3×3. It will help to enhance the discriminating power of each structure image to some extent.

4 Experiments

4.1 Experiment Using the Extend Yale B Database

The extended Yale B face database [13] contains 38 human subjects under 9 poses and 64 illumination conditions. The 64 images of a subject in a particular pose are

Fig. 3. Samples of a person under different illuminations in the extended Yale B face database

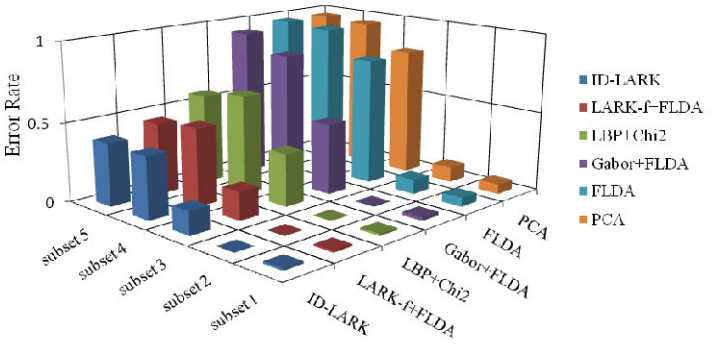


Fig. 4. Breakdown of error rates on five Extended Yale B subsets for PCA, FLDA, Gabor+FLDA, LBP+Chi2, LARK+FLDA and the proposed method ID-LARK

acquired at camera frame rate of 30 frames / second, so there is only small change in head pose and facial expression for those 64 images. However, its extreme lighting conditions still make it a challenging task for most face recognition methods. All frontal-face images marked with P00 are used in our experiment. Each image is resized to 42×48 pixels. Some sample images of one person are shown in Fig 3.

In our experiment, five subsets of Extended Yale B database are obtained according to the angle between the light source direction and central camera axis (12°, 25°, 50°, 77°, 90°) and used in our experiments. Here, the images with the most natural lighting sources ('A+000E+00') are used for training, and all frontal face images of each standard subset for testing (in all, 2394 face images of 38 individuals). Subsequently, PCA, FLDA, Gabor+FLDA, LBP+Chi2 and the proposed method ID-LARK are, respectively, used for feature extraction. Moreover, we also compute LARK features (M×N×D) of each image and end up with features by reducing dimensionality from M×N×D to M×N×d, where, M, N are row and column of image respectively. D is vector dimension of each pixel. And d is salient vector dimension that is obtained by using principal component analysis. D and d are 25, 5 in our experiments, respectively. LARK features consisting of each pixel vector in a serial manner. Here, LARK-f stands for this approach. Therefore, LARK-f plus FLDA also used for feature extraction. Before implementing FLDA, we use PCA to reduce dimension to be 200 with respect to different number of training samples per class. At last, NN classifier is applied to face image classification. The performances of these methods in different subsets are shown in Fig 4.

Fig 4 shows how the various feature extraction methods descend with increasingly extreme illumination from subset 1 to subset 5 of Extended Yale B database. ID-LARK performs quite well under the lighting changes of subsets 1-3. Moreover,

ID-LARK also gives significant better result than others under the most difficult subset 5. In contrast, PCA, FLDA and Gabor+FLDA do well on the earlier subsets, however, they have trouble to address the difficult subsets. This also demonstrates our claims that ID-LARK is more robust to difficult illumination than LBP+Chi2 and other approaches.

4.2 Experiment Using the PIE Database

The CMU PIE face database contains 68 subjects with 41368 face images as a whole [14]. Images of each person were taken across 13 different poses, under 43 different illumination conditions, and with 4 different expressions. We choose the subset (Pose 09) of PIE database containing 24 face images (a nearly frontal pose) of each person [15]. All images are cropped and resized to be 64×64 pixels in our experiment. Some sample images of one person are shown in Fig 5.

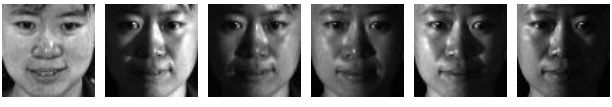


Fig. 5. Sample images of a person under different illuminations in the PIE database (Pose 09)

In this experiment, L (L varies from 4 to 12 with 2 interval) images are randomly selected from the gallery images of each individual to form the training set, and the remaining images are used for testing. We run the experiment 10 times for each L. The average recognition rate and standard deviation (*std*) across 10 runs of tests of each method are illustrated in Table 1.

Table 1. The average rates (%) and *std* of PCA, FLDA, LBP+Chi2, Gabor+FLDA, LARK+FLDA and ID-LARK under NN classifier using 10 run test on the PIE (pose 09) database

Methods/ No. training sample	4	6	8	10	12
PCA	60.3 ± 6.76	7.58 ± 7.73	82.4 ± 10.94	84.9 ± 11.46	94.8 ± 3.26
FLDA	88.1 ± 9.05	92.5 ± 4.56	92.8 ± 4.01	93.6 ± 3.97	94.0 ± 4.91
LBP+Chi2	90.5 ± 2.45	91.6 ± 4.07	95.6 ± 3.37	96.0 ± 3.95	96.0 ± 2.57
Gabor+FLDA	83.6 ± 17.6	92.9 ± 7.37	95.1 ± 4.84	95.2 ± 3.02	96.3 ± 4.00
LARK-f+FLDA	87.1 ± 3.71	89.7 ± 6.60	91.8 ± 5.67	94.3 ± 4.18	95.9 ± 3.37
ID-LARK	92.8 ± 2.84	94.3 ± 4.44	94.7 ± 1.62	96.7 ± 2.36	96.8 ± 3.29

5 Conclusions

In this paper, we develop a novel method image decomposition based on locally adaptive regression kernels for image feature extraction. We first compute dense local structure features of each pixel via LARK. Structure images, are achieved by

decomposing images according to local feature vector, provide structure information in different orientations. We thus downsample all the structure images of one image, however, they are still in a high dimension feature space. FLDA is used to obtain the low-dimension, compact and discriminative feature vector. Finally, NN classifier is employed for classification. Our experimental results on two popular face image datasets under different illumination (Extended Yale B and PIE) demonstrate that ID-LARK outperforms PCA, FLDA, LBP, Gabor and LARK-f. In future, we will further investigate the local structure information to improve the performance when encountering the variations of facial expression, pose and other challenges.

References

1. Zhao, W., Chellappa, R., Phillips, P.J., et al.: Face recognition: A literature survey. Acm Computing Surveys 35(4), 399–459 (2003)
2. Hespanha, J.P., Belhumeur, P.N., Kriengman, D.J.: Eigenfaces versus Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Trans. Pattern Analysis and Machine Intelligence 19(7), 10 (1997)
3. Yan, S.C., Xu, D., Zhang, B.Y., et al.: Graph embedding and extensions: A general framework for dimensionality reduction. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(1), 40–51 (2007)
4. Liu, C.J., Wechsler, H.: Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition. IEEE Transactions on Image Processing 11(4), 467–476 (2002)
5. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(12), 2037–2041 (2006)
6. Zhang, W.C., Shan, S.G., Gao, W., et al.: Local gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In: Proceedings of Tenth IEEE International Conference on Computer Vision, vol. 1, pp. 786–791 (2005)
7. Lei, Z., Liao, S.C., Pietikainen, M., et al.: Face Recognition by Exploring Information Jointly in Space, Scale and Orientation. IEEE Transactions on Image Processing 20(1), 247–256 (2011)
8. Seo, H.J., Milanfar, P.: Training-Free, Generic Object Detection Using Locally Adaptive Regression Kernels. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(9), 1688–1704 (2010)
9. Takeda, H., Farsiu, S., Milanfar, P.: Kernel regression for image processing and reconstruction. IEEE Transactions on Image Processing 16(2), 349–366 (2007)
10. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(10), 1615–1630 (2005)
11. Seo, H.J., Milanfar, P.: Static and space-time visual saliency detection by self-resemblance. Journal of Vision 9(12) (2009)
12. Seo, H.J., Milanfar, P.: Action Recognition from One Example. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(5), 867–882 (2011)
13. Lee, K.C., Ho, J., Kriegman, D.J.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(5), 684–698 (2005)
14. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression database. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(12), 1615–1618 (2003)
15. Cai, D., He, X.F., Han, J.W., et al.: Spectral regression for efficient regularized subspace learning. In: 2007 IEEE 11th International Conference on Computer Vision, vol. 1-6, pp. 214–221 (2007)

Large Scale Identity Deduplication Using Face Recognition Based on Facial Feature Points

Xiaoli Yang, Guangda Su, Jiansheng Chen, Nan Su, and Xiaolong Ren

Department of Electronic Engineering, Tsinghua University, Beijing, China

Abstract. The algorithm of 105 facial feature points localization has been proposed in [1]. In this paper, we studied the stability of these feature points in different photos of the same person, and then we presented an improved face recognition system using these facial feature points to perform face recognition and check duplicate entries in database. All of these analyses and experiments are performed on identity photographs. Experimental results show that our recognition algorithm has obvious improvement in normal face recognition application and also performances satisfactorily in finding out duplicate entries in huge face image database of more than 60,000 items.

Keywords: Facial feature points, Face recognition, Duplicate entries, Deduplication.

1 Introduction

With the development of Chinese Second-generation ID card database system, it is highly required to find out the identity of one person by querying in the ID card database using face images.

Many face recognition algorithms have been proposed, for example, methods based on PCA such as Eigenface [2][3], Fisherface [4], Hierarchical graph matching [5], Sparse Representation Classification [6], and so on. However, these methods are supposed to deal with normal face images but not photos in ID card database, which have very strict demand for light condition and facial pose. Therefore in the database we can use facial feature points for measuring the difference between two photos.

Our lab has a long-term research on face recognition. We constructed a THFR (Tsinghua Face Recognition) system in 2006 based on modified PCA [7], aiming at querying facial images in huge ID card database. Then we developed it with facial feature points location algorithm [8], MMP-PCA [9], similarity normalization algorithm [10]. Based on these works, we applied 105 facial feature points localization algorithm to THFR system to improve the query results.

Moreover, there might be some person who owes two or more different ID cards and has at least two entries in the database. Therefore we designed an algorithm to help our system to find out these duplicate entries.

Experimental results showed that facial feature points are reliable in ID card database and can help the system to performance better.

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 25–32, 2011.

The rest of this paper is organized as follows: Section 2 analyzes the stability of the facial feature points. In Section 3 an improved face recognition system based on THFR is presented, and the algorithm of finding out duplicate entries in database is proposed. Section 4 presents the experimental results on both normal query application and duplicate entries use. Section 5 is conclusion.

2 Stability of 105 Facial Feature Points

Our lab presented an improved facial feature points localization algorithm combining Active Shape Models (ASM) and Active Appearance Models (AAM) [1], which can achieve higher accuracy than the traditional ASM and AAM method. It is also called 105 facial feature points localization algorithm.

Our algorithm locates on a facial image 105 points, of which the locations are shown in Fig. 1.

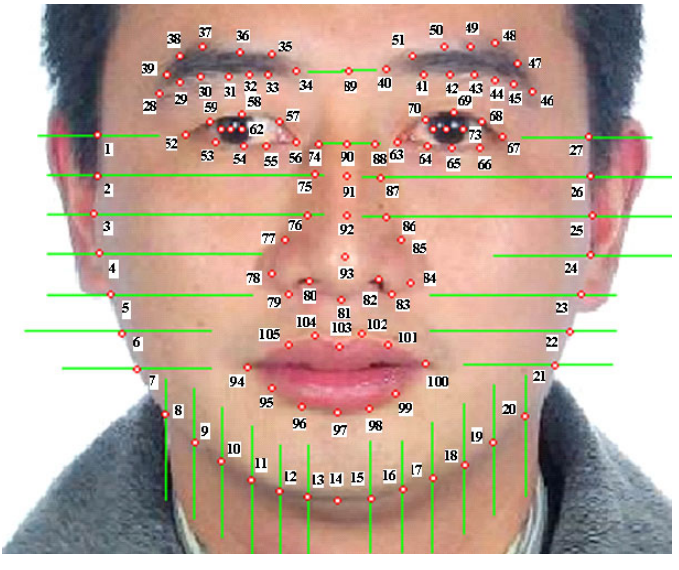


Fig. 1. An example of facial feature points localization (105 points in total)

We took 200 facial photos from 100 individuals with the same light condition and facial pose. That means two different photos of each person. All of these photos are normalized to the same size (320*480 pixels), then are located with 105 facial feature points localization algorithm. So we can get the coordinates of 105 points of every photo, and we will analyze the overall Euclidean distance between 105 points from different photos.

Fig.2 and Fig.3 shows the average offsets of all the feature points under the condition that the test photos come from the same person (in Fig. 2) or different persons (in Fig. 3).

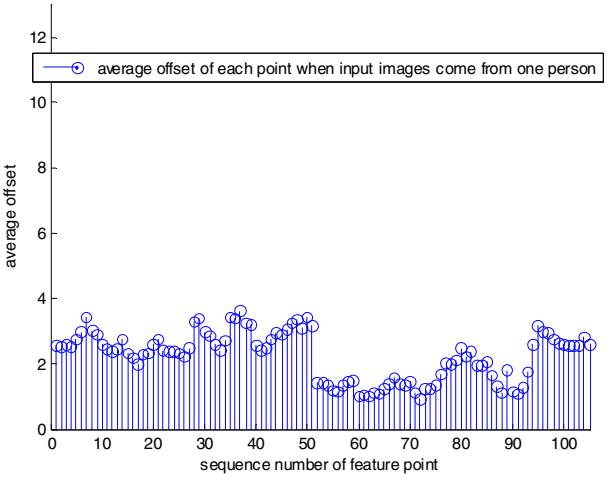


Fig. 2. Average offset of each feature when input images come from the same person

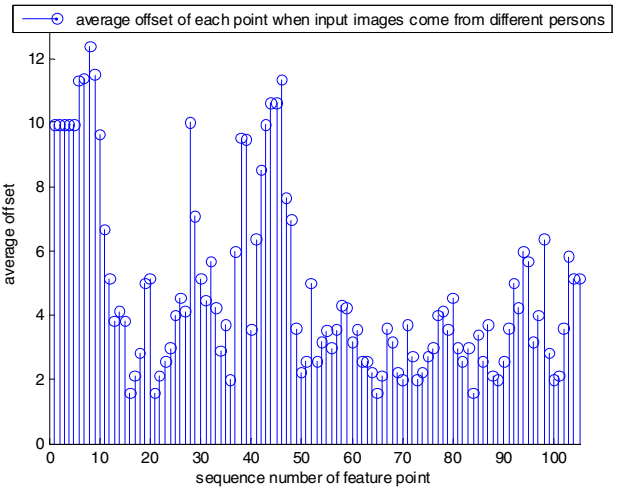


Fig. 3. Average offset of each feature when input images come from different persons

The average offsets of 105 points in the test above are listed in Table 1.

Table 1. The average offset of 105 points for the test in Fig. 2 and Fig. 3

Type of input image	Average offset of 105 feature points
From same person	2.1253
From different persons	4.3520

We can infer from Fig.2, Fig. 3 and Table 1 that the distance of 105 feature points from two photos of one same person is probably smaller than that of different persons. Therefore the 105 feature points are fairly steady in ID card database, and the distance of feature points can be regarded as an auxiliary information in querying.

3 Improved Face Recognition System

In this Section, we will introduce THFR system and the fusion algorithm of former THFR recognition algorithm and 105 feature points.

3.1 Introduction of THFR System

THFR is a face recognition system. It consists of 6 parts – import data, main server, sub server, client, database management and client management. Fig.4 shows how the system works.

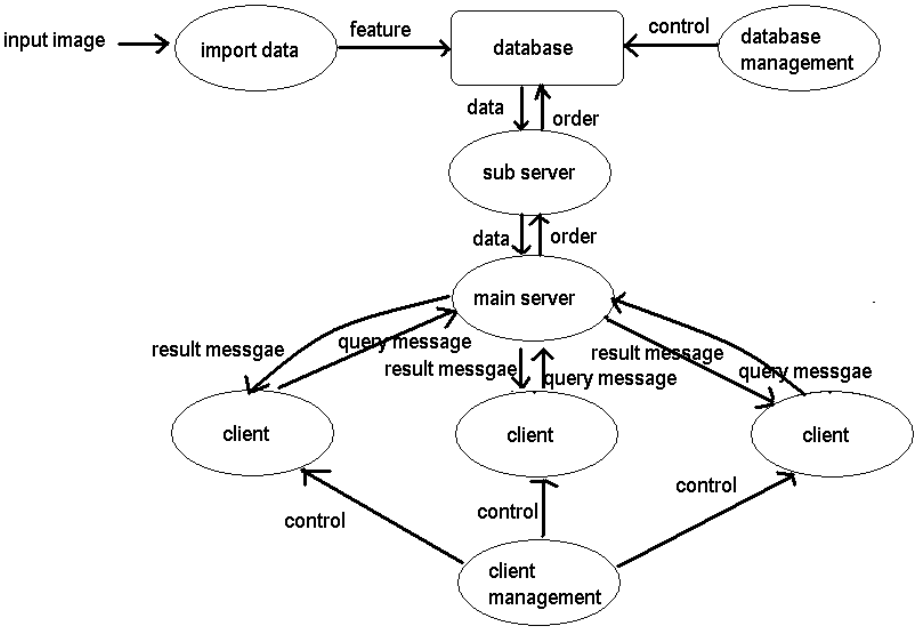


Fig. 4. Architecture of THFR system. Program "import data" receives input images, locates facial areas, extract facial feature, and send the extracted feature to the database to store. When a client logged in and sends a query message to the main server, the main server response and order the sub server to execute. The sub server query in the database and return the query result to the main server. Finally the main server send the result message to the client. There are other two management program to supervise the database and all the clients.

3.2 Improved Algorithm of Face Recognition

We applied the algorithm of 105 facial feature points to THFR, and developed a fusion algorithm to modify the final similarity.

Our fusion algorithm goes as follows:

1. Suppose that the query image is I, and the client want to select top k entries with the similarity score > x.
2. Query message is: select top 2*k entries where similarity > x-0.1 order by similarity descend. Client send query message to the main server, and get result {Ri}, the similarity of {Ri} is {Si}, i = 1,2…m, m≤k, 0≤Si≤1.
3. for (i = 1;i<m;i++)
   {

$$dis_k = \frac{1}{105}\sum_{i=1}^{105}\sqrt{(Rx_i - Ix_i)^2 + (Ry_i - Iy_i)^2}$$

   // dis = average distance between the facial feature of Ri and I;
   if(Si >0.9)
   Si = Si + 0.1 – dis_k /50;
   else if(Si >0.8)
   Si = Si + 0.1 – dis_k /75;
   else if(Si >0.75)
   Si = Si + 0.1 – dis_k /100;
   else if(Si >0.7)
   Si = Si + 0.1 – dis_k /125;
   else if(Si >0.68)
   Si = Si + 0.1 – dis_k /150;
   If(Si>1)
   Si = 1;
   }
   Sort {Ri} by {Si} descend.
4. Select top k entries in {Ri} where Si > x.

3.3 Algorithm of Deduplication

Finding out duplicate entries in the database is a new application of face recognition system. Based on the improved face recognition algorithm in THFR proposed in section 3.2, it is very easy to present the algorithm for finding out duplicate entries:

1. Suppose there are L entries in all in the database {Di}, i = 1,2…L.
2. Result Queue { Qi }
3. For(i=1;i≤L;i++)
   {
   the query image is Di;

```
        k = 5; x = 0.7;
        Execute step 2 and 3 in section 3.2;
        If(S1≥S0)
        {
          Select top k entries in {Ri} where Si > S0;
          Store the selected entries and Di in { Qi };
        }
      }
```

4. All the entries in { Q_i } are duplicated.

4 Experimental Results

In this section, we will carry out experiments both on normal query application and duplication application. The database is Chinese Second-generation ID card database containing over 60,000 entries, and 200 of them are duplicated.

4.1 Normal Query Application

We carried out 200 experiments with every duplicated photo as the input image, and it is expected to find out the other photo of this person. The experimental result of query accuracy is shown in Table 2 and Table 3 presents the distiguishment of query score between the right query result and the nearest wrong query result.

Table 2. The experimental result of accuracy on normal query application

	Former algorithm	New algorithm
Rank 1 identification rate	93.5%	95.5%
Rank 3 identification rate	95.5%	97.0%
Rank 10 identification rate	96.5%	98.0%
Rank 100 identification rate	98.0%	100.0%

Rank N identification rate means the rate of that the correct choice is in the first N choices of the query result queue.

Table 3. The experimental result of query score distiguishment on normal query application

	Former algorithm	New algorithm
Average of query score distiguishment	5.3621%	9.2708%

Average of query score distiguishment means when the input image is a duplicated photo, average of the distiguishment of query score between the correct query result and the incorrect query result which scores highest.

We can infer from Table 2 and Table 3 that compared with former algorithm, our system performance much better both in Rand N identification rate and distiguishment of query scores.

4.2 Deduplication Application

This test aims at finding out all the duplicate entries in the database. Since there are few articles in this field published, we just carry out the experiment and show our results.

There are 100 pairs of duplicate entries in the database and our improved THFR system found out 108 pairs, among which 100 pairs are correct and 8 pairs are wrong, with 0 pair missed. A typical result of this test is shown in Fig. 5.

Deduplication is a new function of our system, and obviously it is of great importance in household management. The new version of THFR system equipped with deduplication function has been put into application of household management in some provinces of China already, and it played a significant role in management of the ID card system.

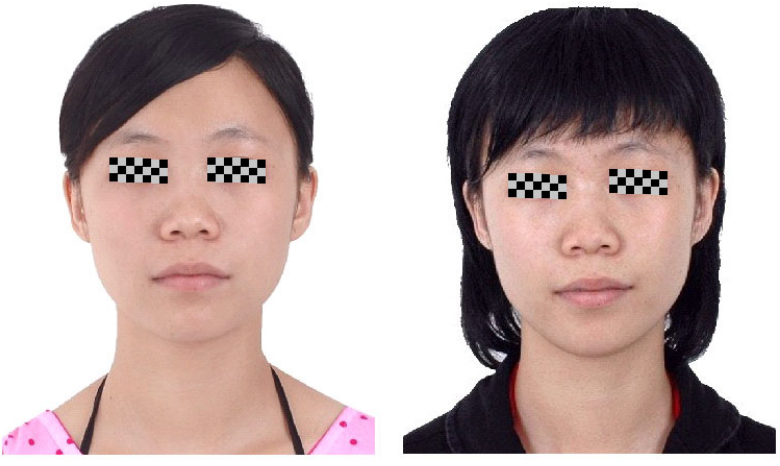


Fig. 5. A typical result of deduplication application – a pair of duplicated entries in the database

5 Conclusion

In this paper, we studied the stability of facial feature points and proposed two algorithms to fuse the feature points in THFR system. In addition, we exploit a new application of face recognition system – to find out duplicate entries in database. Experimental results show that in the database of identity photos, the improved THFR system performances much better in face recognition, and also works very well deduplication application.

References

1. Wang, J., Su, G., Liu, J., Ren, X.: Facial Feature Points Localization Combining ASM and AAM. Journal of Optoelectronics·Laser (2010)
2. Wang, X.G., Tang, X.: Hallucinating face by eigentransformation. IEEE Trans. Syst. Man Cybern. 35(3), 425–434 (2005)
3. Turk, M., Pentland, A.: Face recognition using eigenfaces. In: Int. C. Computer Vision Pattern Recognition, pp. 586–591. IEEE Comput. Sco. Press, Los Alamitos (1991)
4. Bellhumer, P.N., Hespanha, J., Kriegman, D.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, Special Issue on Face Recognition 17(7), 711–720 (1997)
5. Buhmann, J., Lades, M., von der Malsburg, C.: Size and distortion invariant object recognition by hierarchical graph matching. In: Proceedings of IEEE Intl. Joint Conference on Neural Networks, San Diego, pp. 411–416 (1990)
6. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Trans. Pattern Analysis and Machine Intelligence 31(2), 210–227 (2009)
7. Meng, K., Su, G., Li, C., Fu, B., Zhou, J.: A high performance face recognition system based on a huge face database. In: IEEE The International Conference on Machine Learning and Cybernetics (ICMLC), Guangzhou, China, pp. 5159–5164 (2005)
8. Gu, H., Su, G., Du, C.: Automatic locating of facial feature points. Journal of Optoelectronics·Laser 15(8), 975–979 (2004)
9. Xiang, Y., Su, G.: Multi-parts and Multi-feature Fusion in Face Verification. In: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR (2008)
10. Face recognition based on scale normalization, China patent (2007)
11. Chan, H., Bledsoe, W.W.: A man-rmachine facial recognition system: some preliminary results, Technical report. Panoramic Research Inc., Cal (1965)
12. Phillips, P.J., Moon, H., Rizvi, S., Rauss, P.: The FERET evaluation methodology for face recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(10), 1090–1104 (2000)
13. Yang, M.H.: Kernel Eigenfaces vs Kernel Fisherfaces: Face recognition using kernel methods. In: Proceedings of International Conference on Automatic Face and Gesture Recognition, USA, Washington DC, pp. 215–220 (2002)
14. Shan, S., Yang, P., Chen, X., Gao, W.: AdaBoost Gabor Fisher Classifier for Face Recognition. In: Zhao, W., Gong, S., Tang, X. (eds.) AMFG 2005. LNCS, vol. 3723, pp. 279–292. Springer, Heidelberg (2005)

A Sparse Local Feature Descriptor for Robust Face Recognition

Na Liu[1], Jianhuang Lai[2], and Wei-Shi Zheng[2]

[1] School of Maths and Computing Science, Sun Yat-Sen University, Guangzhou, China
[2] School of Information and Technology, Sun Yat-Sen University, Guangzhou, China
lln45@126.com, stsljh@mail.sysu.edu.cn, wszheng@ieee.org

Abstract. A good face recognition algorithm should be robust against variations caused by occlusion, expression or aging changes etc. However, the performance of holistic feature based methods would drop dramatically as holistic features are easily distorted by those variations. SIFT, a classical sparse local feature descriptor, was proposed for object matching between different views and scales and has its potential advantages for face recognition. However, face recognition is different from the matching of general objects. This paper investigates the weakness of SIFT used for face recognition and proposes a novel method based on it. The contributions of our work are two-fold: first, we give a comprehensive analysis of SIFT and study its deficiencies when applied to face recognition. Second, based on the analysis of SIFT, a new sparse local feature descriptor, namely SLFD, is proposed. Experimental results on AR database validates our analysis of SIFT. Comparison experiments on both AR and FERET database show that SLFD outperforms the SIFT, LBP based methods and also some other existing face recognition algorithms in terms of recognition accuracy.

Keywords: face recognition, SIFT, local feature descriptor.

1 Introduction

Face recognition has been extensively used in a wide range of video surveillance, criminal identification etc. A key issue of face recognition is to find good feature descriptor for face representation. There are two types of features extracted from face images, i.e. holistic feature and local feature, and the face recognition methods can be divided into twofold accordingly: holistic feature based and local feature based methods. Different holistic feature extraction methods such as PCA and LDA have been widely studied for face recognition. Recently local descriptors have received many interests due to their robustness to variations caused by occlusion, expression or pose changes etc.

According to the way of using local features, local invariant feature descriptors can be divided into two classes [1]: sparse local descriptors and dense local descriptors.

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 33–41, 2011.

Sparse local descriptors first detect interest points in a given image and then describe invariant features of a local patch, such as SIFT [2]. While dense local descriptor extracts local features pixel by pixel over input image, such as HoG and LBP [3][4], and these dense local descriptors have been broadly applied to face recognition. However, these dense local descriptor based algorithms will suffer performance drop when face images are distorted by some variations. So it is necessary to study the characteristic of sparse local descriptor for robust face recognition.

SIFT as a classical sparse local descriptor was proposed for matching object between different views or scenes [2]. It transforms an image into a large collection of local feature vectors, each of these features is invariant to image scale, rotation and translation. Recently there are some attempts to use SIFT for face recognition. Aly [5] first attempted to use the original SIFT features for face recognition Bicego et al. [6] proposed three different matching schemes for face authentication by using SIFT features. Majundar et al. [7] proposed a discriminative ranking of SIFT features based on Fisher's Discriminate Analysis for face recognition. Geng and Jiang [8] proposed Volume-SIFT and Partial-descriptor-SIFT which kept more large scale keypoints.

From the above analysis we can know that most of the algorithms just simply apply the original SIFT features and focus on the improvement of matching methods for face recognition. As we know, SIFT was proposed for reliable matching objects between different views or scenes [2]. However, face recognition is different from the matching of general objects. First, general objects are rigid and have sharp transitions between different sides. However, faces are non-rigid and smooth. There are much less structures with high contrast or high edge responses. Then, the general objects to be matched are under different scales and viewpoints, but face images used for recognition are more or less aligned. Hence it could not be a good idea to use original SIFT directly for face recognition. In this paper, we first give a systematic analysis about SIFT on face recognition and show that the processing of SIFT feature need to be improved either for accuracy or efficiency. Furthermore, to overcome the shortcomings of SIFT features for face recognition, we propose a simplified sparse local feature descriptor based on SIFT, namely SLFD.

This paper is organized as follows. Section 2 gives the review and analysis of SIFT. Section 3 describes the new sparse local feature descriptor, SLFD. In section 4 extensive experiments are performed to examine the validity of the analysis, comparison experiments with other algorithms are also conducted to show the effectiveness and advantages of the proposed method. Section 5 concludes the paper.

2 Overview and Analysis of SIFT Features

2.1 Overview of SIFT Features

According to [2], the computation of SIFT feature consists of four major stages: (1) Scale-space extrema detection; (2) Keypoints localization: unreliable keypoints removal; (3) Orientation assignment; (4) Keypoints description.

The first stage is to identify the locations and scales that can be repeatedly assigned under differing views of the same object. It is implemented efficiently by using a Difference-of-Gaussian function to identify potential interest points. Given an input image, $I(x, y)$, the scale space $L(x, y, \sigma)$ is produced from the convolution of a variable-scale Gaussian, $G(x, y, \sigma)$

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{1}$$

where $G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$. Then the difference-of-Gaussian (DoG) images $D(x, y, \sigma)$ (see Fig. 1) are produced by computing the difference of two adjacent scale images separated by a constant multiplicative factor k:

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \tag{2}$$

If the number of scale images in each octave required for feature points extraction is S. Then S+3 scale images needed to be computed in each octave. After each octave, the Gaussian images are down-sampled by a factor of 2, and the process repeated. Potential interest points are identified by detect the extreme points of $D(x, y, \sigma)$.

In the second stage, the candidate keypoints obtained in the first stage are selected for stability by rejecting keypoints with low contrast and eliminating the edge response keypoints.

For rotation variance, the third stage assigns one or more orientations to the stable keypoints based on the local gradient directions in its local neighborhood.

One kind of keypoint descriptor is created in the last stage based on the image gradient, first, compute the gradient orientations and magnitudes of the image in its 16×16 local neighborhood, then an orientation histograms is created with 8 orientation bins over sub-regions of size 4×4, therefore a 128-element vector is formed from a 4×4 array of histograms.

For integrity, here we also give the original SIFT feature matching strategy: each feature vector extracted from the keypoint in the test image is compared to all those feature vectors in the template image by distance. If the ratio of the closest distance to the second-closest distance is below some threshold T_{ratio}, then the match is found.

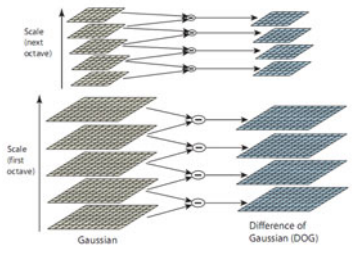


Fig. 1. Illustration of scale space and DOG (from [2])

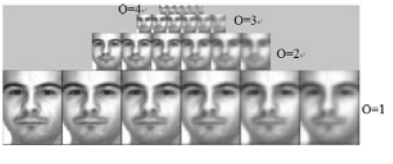


Fig. 2. Illustration of scale images with four octaves

2.2 Analysis of SIFT Features

In the following, we analyze the characteristics of SIFT feature and study its deficiencies when applied to face recognition.

First, in order to achieve scale invariance, i.e. matching the same object images under different scale, the SIFT features are computed in the scale space with a large scale range. However, alignment is a particularly important problem in face recognition and the face images used for recognition are more or less aligned. So for face recognition the important thing is which scale feature is good for recognition not the scale itself. From the above review, we know that the scale images in large octave which are computed by down-sampling of images in the last octave together with Gaussian smoothing would discard many important information (see Fig. 2). So for face recognition, it is reasonable for us to neglect the large scale spaces and use only some scale spaces in the first octave.

The second stage of SIFT is to process stable keypoints selection by removing keypoints with low contrast and eliminating the edge response keypoints. This is done under the assumption that the low contrast keypoints are sensitive to noises and the edge response keypoints are corresponding to unstable feature points for object matching. But this is not the case for face recognition. First, the SIFT features are computed in the Gaussian scale images and the noises have been smoothed after Gaussian smoothing. So the effect of noises can be ignored here. What's more, the edge responses keypoints such as eye boundary and mouth boundary contain important discriminative information for face recognition. As a result, the process of this stage is actually degrading the performance of face recognition.

In the third stage of SIFT, one or more dominant orientations are assigned to every keypoint and the feature descriptor is represented relatively to this orientation. But the face images are simply aligned, so a descriptor which is represented relative to the dominant orientation may lead to false matching correspondences for face recognition [9]. In other words, the process of this stage should be avoided in face recognition.

Finally, one problem of the original SIFT feature matching strategy is that it is very time consuming. It requires P^2 computations for matching between two images, where P is the average number of SIFT features extracted in each image. Another is false matched keypoints, since two matched keypoints satisfying the distance criterion could be not related to the same face part. So the matching criterion should take the specialty of face images into account for face recognition.

Through the above analysis of the SIFT feature descriptor, we know that there are some weakness of the original SIFT features for face recognition. It is the sparsity characteristic in the first stage that makes SIFT features suitable for robust face recognition, especially when feature distortion problem exists.

3 Proposed Sparse Local Feature Descriptor

3.1 Sparse Local Feature Descriptor（SLFD）

From the analysis of the SIFT feature descriptor, we have known that there are some deficiencies when applying SIFT to face recognition and the sparsity is what we need for robust face recognition. In this section, we propose a new simplified sparse local

feature descriptor based on SIFT, denoted as SLFD, by detecting the interest feature points primitively without keypoints refinement, no orientation assignment followed and then by applying the feature descriptor on the keypoints.

The procedure of the construction of SLFD is illustrated at Procedure 1. First, we search the locations of feature points in the scale space which is similar to the first stage of SIFT, and the difference is that we only need one octave of scale-space instead of the Gaussian pyramid used in SIFT. Then we describe the local features of the keypoints by using the SIFT feature descriptor.

Procedure 1. Computation of SLFD

1) Feature points location.
   a) Given an input image $I(x, y)$, number of the scale space S and σ_0.
   b) Construct the S+3 scale spaces $L(x, y, \sigma)$ by equation (1) in the first octave, where $\sigma = k^s \sigma_0, s = 0, 1, \ldots, S + 2.$
   c) Construct the Difference-of-Gaussian (DoG) space by equation (2) where $k = 2^{1/s}$.
   d) Feature points location: Feature locations are obtained by detecting the extremas of $D(x, y, \sigma)$. those extremas are computed based on comparing each sample points to its eight neighbors in current image and nine neighbors in the scale above and below.
2) Computation of the SLFD.
   a) Take a 16×16 size block centered at the feature coordinates.
   b) Compute the gradient magnitudes and orientations in the block, then sample the magnitudes and orientations in each 4×4 sub-block and an orientation histogram is created with 8 orientation bins.
   c) Finally, a normalized 128-element vector (4×4×8) is formed in the block.

From Procedure 1, we can see that the time consumption of SLFD can be reduced greatly compared to SIFT. More important, the feature descriptors obtained using SLFD are more discriminative than SIFT descriptors for face recognition. First, by keeping edge response keypoints, the important edge information which is critical for face recognition can be used in SLFD. Second, without orientation alignment, the coordinates of the descriptors and the gradient orientations do not need to rotate relatively to the keypoint dominant orientation, so all the SLFD are computed under the same coordinate, and the false matching caused by the orientation assignment can be avoided since face images are extracted and aligned in advance for recognition.

3.2 Matching Criterion

To avoid the problems in the original matching criterion as stated in the end of section 2.2, we here exploit the local matching distance criterion [10] for recognition. Given a probe image X^m and its extracted feature vectors $V_{X^m} = \{x_1^m, \ldots x_P^m\}$, template images

$Y^n, n = 1,...N$ and their corresponding feature vectors $V_{Y^n} = \{y_1^n, ...y_Q^n\}$, $n = 1,...N$. For every feature vector in V_{X^m}, we only compute the distance between it and the feature vectors within its corresponding r-neighborhood in template image instead of computing the distance with all the feature vectors. For convenience, we formulate the distance vector in ascending order $D(x_p^m, Y^n) = (d_1^m, d_2^m, ...d_r^m)$, $where\ d_1^m < d_2^m < ... < d_r^m$. Different from the original matching criterion, a distance constraint is imposed to reject the matching if the smallest distance is larger than a threshold $T_{dist.}$. For accuracy, we do not use the number but a weighted average distance of the matched feature vectors for recognition as follows:

$$Label(X^m) = \arg\min_n((\sum_{p=1}^{P} \delta_{lc}(x_p^m, Y^n) * d_1^m) / LCnum(m,n) \wedge 2) \tag{3}$$

$$LCnum = \sum_{p=1}^{P} \delta_{lc}(x_p^m, Y^n) \tag{4}$$

$$\delta_{lc}(x_p^m, Y^n) = \begin{cases} 1, & if\ d_1^m / d_2^m < T_{ratio}\ \&\&\ d_1^m < T_{dist}; \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

This matching criterion can be much more efficient and accurate than the original matching method. If r-neighborhood is adopted and the image size is M*M, the computation is reduced by an r*r/M*M factor. This method can also avoid mismatched keypoints which are not related to the same face par.

4 Experiments

4.1 Verification Experiments and Discussion

To investigate the analysis of using SIFT for face recognition presented in section 2.2, we have conducted several experiments on AR dataset. The AR database consists of over 4000 images of 126 individuals. There are 26 frontal view faces per person captured in two different sessions with different expressions (smiling, angry and screaming), illumination conditions (left light, right light and uniform light) and occlusions (sun glasses and scarf), each session consists of 13 images. The neutral expression images from the first session are selected as gallery set. Two expression faces (smiling, angry) and two occlusion images with uniform light from the first session are selected as probe set. Through the experiments, the parameter ratio T_{ratio} and the constraint distance T_{dist} are set to 0.7 and 0.32 respectively.

First, we discuss the influence of the octave number and scale number using SLFD features and local matching distance criterion. The recognition rates which are presented separately according to the variation patterns, i.e. occlusion and expression, are shown in Table 1.

From the results we can see that when the scale number S of each octave is fixed the recognition rates increase only slightly or even decrease as the octave number

increases. This confirms our previous hypothesis that we could neglect the large scale spaces and use only some scale spaces of the first octave. So in our latter experiments, we only use one octave with 6 scales.

We conducted the second experiment to investigate the influence of the keypoints localization and the orientation assignment using SIFT. Here we used the number of matched keypoints for recognition to show the influence of the two factors directly. The performances are shown in Table 2.

From the results in Table 2 we know that both the keypoints location and orientation assignment can lead to false matching correspondences. So we could avoid these two processes in SIFT feature computation procedure when SIFT is used for face recognition.

Table 1. Results of SLFD with different number of octaves and scales

(a) Performance under occlusion

	O=1	O=2	O=3	O=4
S=3	97.48%	98.74%	98.74%	98.74%
S=4	98.74%	99.16%	99.16%	99.16%
S=5	99.16%	98.74%	99.16%	99.16%
S=6	100%	99.58%	99.58%	99.58%
S=7	99.58%	99.16%	99.16%	99.16%

(b) Performance under expression

	O=1	O=2	O=3	O=4
S=3	96.22%	99.58%	99.58%	99.58%
S=4	97.48%	99.16%	99.58%	99.58%
S=5	98.32%	99.58%	99.58%	99.58%
S=6	98.74%	99.58%	99.58%	99.58%
S=7	99.16%	99.58%	99.58%	99.58%

Table 2. Performance comparison of whether using keypoints localization or orientation assignment in the procedure of SIFT. KL denotes keypoints localization and OA denotes orientation assignment.

	KL and OA	No KL	No OA	No KL and No OA
Ocl	93.28%	97.48%	96.22%	97.90%
Exp	93.21%	94.96%	95.80%	97.48%

4.2 Performance of the Proposed Method

To verify the performance of the proposed SLFD, we have conducted various experiments on both the AR database and FERET database.

On AR database, besides the image set selected in section 4.1, three expression faces (neural, smiling, angry) and two occlusion images with uniform light from the second session are also added to the probe set.

The FERET database is one of the largest publicly available databases. In this experiment, the training set contains 731 images. In test phase, the gallery set contains 1196 images from 1196 subjects. Four probe sets (fb, fc, dup1 and dup2) including expression, illumination and aging variations are used to compare the performance of different methods. The comparison results on AR and FERET dataset are shown in Table 3. On AR dataset, we only use one sample per person in the gallery set, so we do not give the results of using FLDA.

From Table 3, one can find that the proposed simplified sparse local feature descriptor, SLFT, consistently obtains much better recognition rates than the original SIFT based method, one dense local feature descriptor, LBP and two holistic feature based methods. The results indicates that our proposed method can performs well under most of the variants such as occlusion, expression etc.

Table 3. Performance comparison on AR and FERET databases. Exp denotes expression and Ocl denotes occlusion.

AR	Session1		Session 2	
	Exp	Ocl	Exp	Ocl
SIFT	94.54%	94.96%	90.19%	83.61%
PCA	91.18%	55.18%	10.5%	4.2%
LBP	97.89%	89.92%	92.44%	71.85%
SLFD	98.74%	100%	96.92%	96.22%

FETET	Fb	Fc	Dup1	Dup2
SIFT	90.80%	61.86%	54.43%	49.57%
PCA	78.91%	9.79%	33.66%	11.97%
FLDA	87.78%	47.42%	44.32%	20.09%
LBP	97%	79%	66%	64%
SLFD	97.49%	93.82%	69.95%	74.79%

5 Conclusion

For robust face recognition, this paper investigates the characteristics of SIFT and proposes a novel method based on sparse local feature to solve it. First, we give a systematic analysis of SIFT and study its deficiency when applied to face recognition. Second, based on the analysis of SIFT, a new simplified sparse local feature descriptor, namely SLFD, is proposed. It's very attractive to find that the proposed methods perform consistently superior to the original SIFT, LBP based methods and also several other existing face recognition algorithms in terms of recognition accuracy.

Acknowledgment. This project was supported by the NSFC-GuangDong (U0835005) and the 985 project in Sun Yat-sen University with grant no. 35000-3181305.

References

1. Chen, J., Shan, J., He, C., Zhao, G., Chen, X., Gao, W.: WLD: A Robust Local Image Descriptor. IEEE TPAMI 32(9), 1705–1720 (2010)
2. Lowe, D.: Distinctive image features from scale invariant keypoints. IJCV 60(2), 91–110 (2004)
3. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for human Detection. In: CVPR (2005)
4. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution Gray Scale and Rotation Invariant Texture Analysis with Local Binary Patterns. IEEE TPAMI 24(7), 971–987 (2002)
5. Aly, M.: Face Recognition using SIFT Features. CNS/Bi/EE report 186 (2006)
6. Bicego, M., Lagorio, A., Grosso, E., Tistarelli, M.: On the Use of SIFT Features for Face Authentication. In: CVPR Workshop (2006)

7. Majumdar, A., Ward, R.K.: Discriminative SIFT Features for Face Recognition. In: Canadian Conference on Electrical and Computer Engineering, pp. 27–30 (2009)
8. Geng, C., Jiang, X.D.: Face recognition using sift features. In: ICIP, pp. 3313–3316 (2009)
9. Dreuw, P., Sterngrube, P., Hanselmann, H., Ney, H.: SURF-Face: Face Recognition Under Viewpoint Consistency Constraints. In: BMVC (2009)
10. Liu, N., Lai, J.H., Qiu, H.N.: Robust Face Recognition by Sparse Local Feature from a Single Image under Occlusion. In: ICIG (2011)

Asymmetric Facial Shape Based on Symmetry Assumption

Jianfang Hu[1], Guocan Feng[1], Jianhuang Lai[2], and Wei-Shi Zheng[2]

[1] School of Maths and Computing Science, Sun Yat-Sen University, Guangzhou, China
[2] School of Information and Technology, Sun Yat-Sen University, Guangzhou, China
hujianfang21@126.com, mcfgc@mail.sysu.edu.cn
stsljh@mail.sysu.edu.cn, wszheng@ieee.org

Abstract. It has been known that it is hard to capture the high-frequency components (shadows and specularities) during the modeling of illumination effects. In this paper, we propose a reflectance model to simulate the interaction of light and the facial surface under the assumption that face is strictly axial symmetry. This model works well not only in fitting the intensities of pixel but also in processing the DC component contained in the image. To compute a facial 3D shape, we first augment the input images to get a symmetric facial normal field, then propose a method to obtain a more accurate normal field, and finally compute an integrable shape using the field. Experimental results for face relighting, facial shape recovery demonstrate the effectiveness of our method.

Keywords: shape recovery, face relighting, Non-Lambertian model, symmetry.

1 Introduction

For computing a facial 3D shape, it is useful to make an assumption about the interaction of light and the facial surface i.e. the reflectance model. Most of the proposed models can be classified into two classes, Lambertian and Non-Lambertian model. The Lambertian model, which is popular for capturing the intensity of the facial image effectively, has been widely used in 3D shape recovery. There are different methods to handle the model even using the same reflectance model. One of the most straightforward methods is linear subspace [1, 2], which makes out some basis images that can be used to represent the facial image with variable lighting and pose linearly, and then the 3D facial shape is obtained by an iterative algorithm. Another method is to express the facial surface (referred to as the reflectance function) in terms of spherical harmonics, through which the author draws the conclusion that the facial surface can act as a low-pass filter to the light signal, and it can capture more than 87.5 percent of the energy even when only the first order approximation is used. With the spherical harmonics representation, [3] proposes a reconstruction algorithm from a single image using a single reference face shape. As discussed above, the facial surface acts as a low-pass filter under the Lambertian assumption, it behaves unsatisfactorily when handling the high-frequency component such as cast shadows and specularities. Therefore some shape reconstruction methods exploit No-Lambertian model to capture

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 42–49, 2011.

the rapidly changing section of image, [4] proposes a Non-Lambertian reflectance model using tensor-splines, through which they can simulate the high-frequency composition well. However, it takes the intensity of shadow as 0 and doesn't make full use of facial symmetrical structure. The intensity of shadow is a small positive integral due to the DC component added in the image. So the estimated error may be very great. In this paper, we propose a reflectance model which not only can approximate the shadows and specularities well but also can simulate the DC composition well. We also use the symmetry of face to increase the number of the input images. Another mentionable method used in this paper is a mean to obtain height from the normal field in [5, 8], by which we can get a facial 3D shape satisfying a certain integrability and smoothness.

This paper is organized as follows: In section 2, we will propose a reflectance model and illustrate the existence of DC component in image, we will then show the capability of our model to capture specularities and DC component. In section 3, we will propose a reconstruction algorithm based on the assumption of facial symmetry. In section 4, we will apply our method to face relighting and shape recovery. And finally the section 5 is the sum-up about this paper.

2 Reflectance Model

2.1 Reflectance Model

If we consider the intensity of the pixel I as a function of the light direction, we can write it as $I = f(s_1, s_2, s_3)$ where $s = (s_1, s_2, s_3)$ is a unit vector, now we can expand it with Taylor's series as follows:

$$E=f(0,0,0)+f_x(0,0,0)s_1+f_y(0,0,0)s_2+f_z(0,0,0)s_3+\cdots \tag{1}$$

where $f(0,0,0)$, $f_x(0,0,0)$, $f_y(0,0,0)$, $f_z(0,0,0)$, $\cdots$ are associated with the object's normal, the incident light intensity, and the albedo to the corresponding point. If the incident light intensity is fixed, all of these coefficients are considered as constants.

The Lambertian reflectance Model can be seen as a special case of Eq.1, where only the linear terms are considered.

Image usually can be added with an extra constant intensity due to the diffuse reflection and imaging system itself, i.e. a picture signal often contains a DC component. As illustrated in Fig.1, the face in the white block are in the shadow region, according to the Lambertian rule, their intensities should be 0, but the values are not 0 but relatively small integer as shown in (b). In other words, the image intensity is added with a constant value i.e. DC component we called above. In fact face can not only give diffuse reflection but also specular reflection. So it is essential to have the nonlinear terms in Eq.1 been kept to approximate the specular reflection. In this paper, the second order terms are kept. So we can get the reflectance model as follows:

$$E=\sum_{k+l+m\le 2} T_{klm}(s_1)^k(s_2)^l(s_3)^m \tag{2}$$

(a) (b)

Fig. 1. Facial image. (a) facial image with white block (b) intensities of pixels in white block.

However, the vector $\vec{s}$ is unit and the second order part can contain $f(0,0,0)$, we can write the above formula as

$$E = \sum_{1 \le k+l+m \le 2} T_{klm} (s_1)^k (s_2)^l (s_3)^m \tag{3}$$

Now, we have obtained Eq.3 as our reflectance model. From the following section we can find it can capture the DC component and the specular reflection well.

2.2 Fitting Compared

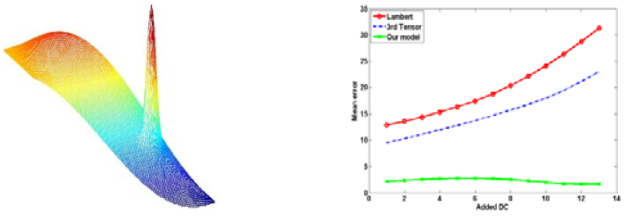
From the previous section we know that image can be added with a DC component, but its magnitude which is related with the digitalization system and the imaging ambient conditions is uncertain. In this section, we mainly illustrate the effectiveness of our model to capture specularities and DC component. As shown in the Fig.2, the standard error estimated by Lambertian model and the third order tensor [4] increases as the DC component increase, while stays relatively constant at a small value when our model is used. Typically, the Lambertian model has the largest error, the third tensor takes the 2nd place, and our model has the least error.

3 Shape Recovery

Zhao introduced symmetric constraint method to reconstruct the 3D shape of the symmetrical objects [7] and he can obtain object's 3D shape from single image. However, the method can't work properly when the object has cast shadow or specularity. In this paper, strictly symmetric face is assumed at first to obtain its normal field, then an image is picked up to modify the normal field, at last the height field(3D shape) is obtained from the modified normal.

3.1 Symmetry

Generally speaking, the two types of geometrical symmetry discussed mostly are central symmetry and axial symmetry. In this section, we mainly focus on axial symmetry in R^3. For any axisymmetric objects, we can make the axis pass the original through a coordinate transformation, such that the lines between symmetric points are perpendicular to the axial plane, i.e. y-z plane. Then we can get $z(x, y) = z(-x, y)$,



(a) (b)

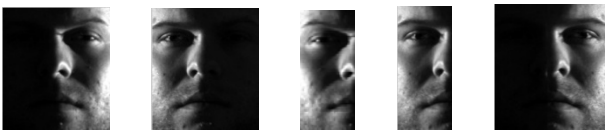
Fig. 2. Synthetic example (a) synthesis example (b) fitting error

where $z(x, y)$ represents the height of the object at (x, y), so $n_1(x, y) = -n_1(-x, y)$, where $n_1(x, y)$ is the first component of normal at (x, y). In this section, we assume face is symmetry in both geometrical shape and surface reflectance property. Then we can draw the conclusion that the intensity at (x, y) with (s_1, s_2, s_3) as its incident light direction is equal to the intensity at $(-x, y)$ with light direction $(-s_1, s_2, s_3)$. In Fig.3, (a) and (b) are two images from Yale Face Database B, The light direction of (a) is$(-95, 0)$, and (b) is$(95, 0)$ separately, (c) is right half of (a) with a mirror left-right transformation. (d) is left half of (b), (e) is a synthetic image whose left half is from (a) and right half from (b). (c) and (d) are almost the same except in the regions where specular reflection is happened. We attribute this difference to the nonlinear terms in the Eq.3. The mean Euclidean distance between (c) and (d) is 0.3649, and their correlation coefficient is 0.9223. This means that if the direction of the light turns to the symmetrical direction, the new image can be obtained just from the old one. The high correlation between (c) and (d) also illustrate that the linear terms play a major role in Eq.3. This consists with the conclusion that the linear estimators of reflective function can capture more than 87.5 percent of the energy [2].

Based on the above analysis, the symmetry of face in this paper can be used to augment the gallery (facial) set with meaningful images. For any input image with a light direction (s_1, s_2, s_3), we can safely draw a method to synthesize the image with light direction $(-s_1, s_2, s_3)$: Interchanging the left half with right half of the original image. It means that we can double the input images on the premise of $s_1 \neq 0$.

3.2 Shape Recovery

Based on the above symmetry assumption, we can just reconstruct the 3D shape of the half-face, and this can reduce the computational time to a certain degree. However, face may be some departing from symmetry, it is necessary to make some amendment after getting the symmetric normal. There are 9 coefficients in the reflectance model determined by Eq.3. As discussed in the previous section, we can derive another image from each input image, so if we want to solve the coefficients we need 4 input images at least. In this paper, more than 5 images are used to gain a reliable result. Note that the application of symmetry is to synthesis images with symmetric light direction, so we require the light directions of the input images should not be symmetry, i.e. all the first



(a) (b) (c) (d) (e)

Fig. 3. Symmetry (a) (-95,0) (b) (95,0) (c) right half of (a) with mirror transform (d) left half of (b) (e) Synthetic image

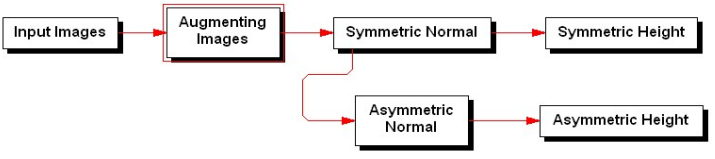


Fig. 4. Algorithm flow chart

component of the light directions should not be opposite number, only in this case can the symmetry assumption be fully used. The reconstruction algorithm in this paper mainly consists of three acts, (1) global normal computation, (2) normal field modification, (3) shape (height) recovery.

It is a simple process to solve the normal vector field from the input images. Similar to the method in [4], 10 linear equations can be gotten on the basis of Eq.3 at any pixel, and then the coefficients in Eq.3 can be obtained by solving these equations, at last the direction that makes Eq.3 reach its maximum can be taken as its normal direction.

Right now, the normal field we've gotten is a strictly symmetric field. However, to a certain degree, face may be departing from symmetry, so it is essential to make some adjustment in the normal field. So we pick out one image that without shadows and specularities firstly, then we make an amendment based on the image. The amendment is achieved by solving the following optimization problem:

$$\min_{\vec{n}} \int_{\Omega} (\left\| \vec{n} - \overrightarrow{n_1} \right\|_{L^2}^2 + \lambda (I - \rho \vec{n} \vec{s})^2) dxdy \tag{4}$$

Where $\overrightarrow{n_1}$ is the normal before modifying, ρ satisfies:

$$\rho \overrightarrow{n_1} \vec{s} = I(x, y) \tag{5}$$

In order to get a height field that satisfies some conditions as smoothness and integrability from the modified normal field, we adopt the method proposed in [5], although another reliable method is also proposed in [8]. Firstly, from the normal vector field, we can obtain the gradient field of the facial height as follows:

$$z_x = -\frac{n_1}{n_3}, z_y = -\frac{n_2}{n_3}, \vec{n} = (n_1, n_2, n_3) \tag{6}$$

Let:

$$\tilde{c}(w) = \frac{-jw_1 c_x(w) - jw_2 c_y(w)}{w_1^2 + w_2^2} \tag{7}$$

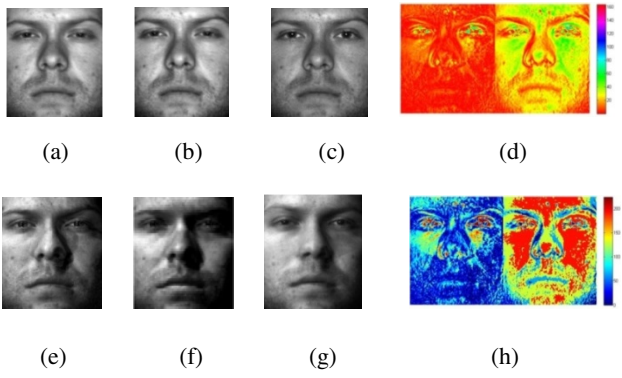


Fig. 5. Relighting comparison. (a) our method (b) 3rd tensor (c) ground truth (d) the errors (e) without symmetry hypothesis (f)3rd tensor (g)symmetry hypothesis. (h) SMQT of (d).

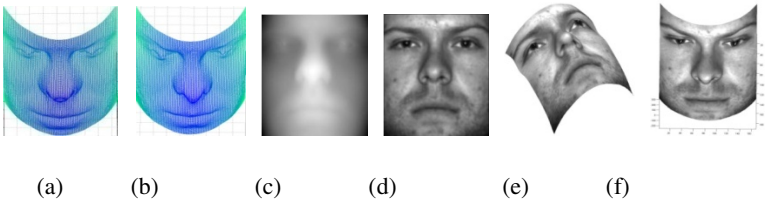


Fig. 6. Shape recovery. (a) symmetric shape (b) asymmetric shape (c) facial height (d) one input image (e) shape with texture (f) shape with texture.

where $\{c_x\}$, $\{c_y\}$ are the Fourier coefficient of z_x and z_y , w=(w_1,w_2) is a two-dimensional index. Then we can get the integrable height field only by performing the inverse 2D Fourier Transform on the coefficients $\tilde{c}(w)$.

Finally, we make a summary for our reconstruction algorithm. As shown in Fig.4, firstly, we augment the number of input images with symmetry assumption, then we obtain a facial symmetric normal field with the proposed reflectance model, in this step, we assumed that Eq.3 reaches maximal value at its normal. With the obtained field we can figure out facial symmetric height field, and the result will be presented in the following section. Because face may be not strictly symmetry, we have to adjust the symmetric normal to get a more accurate facial height, i.e. facial 3D shape.

4 Experimental Results

4.1 Face Relighting

In this section, we apply our method to face relighting. Firstly, we obtain face images with illumination direction (0,0,1) by the 3rd tensor and the our model without symmetry postulate, they are presented in Fig.5. We can see that (a) and (b) are much similar to the ground truth except in some specular regions. The left half of (d) is the

Fig. 7. Shape from varied reflectance models. Top-bottom: Lambertian, 3rd tensor, our model.

visualization of deviation between (a) and (c), while the right half is that between (b) and (c). (h) is the regularized results of (d) with the method called successive mean quantization transform (SMQT) [6]. From (d) and (h) we can find that images obtained by our method are closer to the ground truth than 3rd tensor. Also from (h), we can see that the DC component contained in image signal should not be neglected. (e), (f), (g) is the relighting results taking (40,30) as the direction of illumination. Unlike (a), (b), (c), the direction of (e), (f), (g) is not consisted in the Yale B Database, so we can't draw a parallel between the experimental results and the ground truth, but we can also see that the results are quite according with the truth. However, the relighting images are not very good in the shadows due to lack of input images. Comparing (e) with (g), we can find that the symmetry hypothesis can marginally reduce the vagaries of shadows, and it is useful to reconstruct 3D facial shape.

4.2 Shape Recovery

In Fig.6, a 3D symmetric facial shape reconstructed by the method in this paper is shown in (a) and a asymmetric shape from symmetry hypothesis is presented in (b). Regarding them as a whole, quite good agreement between experimental results and the real facial shape is obtained. However, some deficiencies are also existed in the reconstructed shape. Take nose for example, because shadows and specularities often happen near the nose, the reconstructed shape would probably deviate from the proper value. From (b) we can hardly find that the 3D facial shape is not symmetrical. And the correlation coefficient of left half and right half is 0.9901 which implies the retouched part is quite small. (c) is the height field reconstructed from input images, and (d) is one of the five input images. (e) and (f) are 3D facial shapes overlaid with an input image at different poses.

In order to illustrate the effectiveness of our reflectance model in the reconstruction algorithm, the model is replaced with Lambertian and 3rd tensor model and the corresponding experiments are performed. Some results are presented in Fig.7. We can find that the performance of our model to simulate the interaction of light and facial surface is better than 3rd tensor model but more or less the same with Lambertian model.

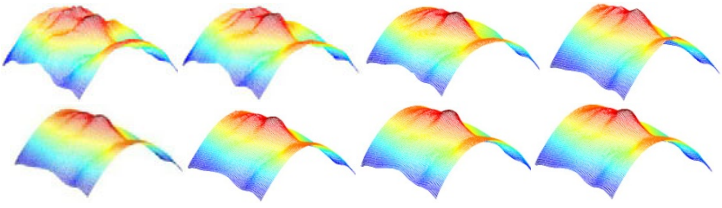


Fig. 8. Shapes from varied numbers of images. Left-right(up-down) 6,11,16,21,26,31,36,41.

As shown in Fig.8, Our reconstructed algorithm can obtain a quite good result (except in the region near the lips) even when only 6 input images are used, and we attributed this inaccuracy to the lack of input images. We can find that the facial shape becomes more and more accurate as the number of input images increases.

5 Conclusion

We proposed a reflectance model to simulate the common photo-effects (e.g. shadow, specularity) and process the DC component contained in image signal. Different from previous work, the symmetric assumption was used to augment the input images rather than being used as a constraint for modeling. Our experiments showed that the proposed method could reduce the effect of shadow to some extent, and could achieve good performance for shape reconstruction.

References

1. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 643–660 (2001)
2. Basri, R., Jacobs, D.W.: Lambertian Reflectance and Linear Subspaces. IEEE Transactions on Pattern Analysis and Machine Intelligence 25, 218–233 (2003)
3. Kemelmacher-Shlizerman, I., Basri, R.: 3D Face Reconstruction from a Single Image Using a Single Reference Face Shape. IEEE Transactions on Pattern Analysis and Machine Intelligence 33, 394–405 (2011)
4. Kumar, R., Barmpoutis, A., Banerjee, A., Vemuri, B.C.: Non-Lambertian Reflectance Modeling and Shape Recovery of Faces Using Tensor Splines. IEEE Transactions on Pattern Analysis and Machine Intelligence 33, 553–566 (2011)
5. Frankot, R.T., Chellappa, R.: A Method for Enforcing Integrability in Shape from Shading Algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 10, 439–451 (1988)
6. Nilsson, M., Dahl, M., Claesson, I.: The Successive Mean Quantization Transform. In: IEEE International Conference on Acoustics, Speech and Signal Processing, 2005. Proceedings(ICASSP 2005), vol. 4, pp. 429–432 (2005)
7. Zhao, W.Y., Chellappa, R.: Symmetric Shape-from-Shading Using Self-ratio Image. International Journal of Computer Vision 45, 55–75 (2001)
8. Kovesi, P.: Shapelets Correlated with Surface Normals Produce Surfaces. In: Tenth IEEE International Conference on Computer Vision (ICCV 2005), vol. 2, pp. 994–1001 (2005)

Fuzzy Cyclic Random Mapping for Face Recognition Based on MD-RiuLBP Feature

Wu Yichen, Fang Yuchun, and Tan Ying

School of Computer Engineering and Science, Shanghai University, Shanghai, China

Abstract. In this paper, we propose a novel face-hashing algorithm named Fuzzy Cyclic Random Mapping (FCRM) and utilize it with our previously proposed Multi-directional Rotation Invariant Uniform Local Binary Pattern (MD-RiuLBP) feature for both face authentication and identification. The kernel part of the FCRM is a cyclic random mapping process. The fault-tolerant technology is also introduced in the FCRM to reduce the impact of random noise existing in the face features. Several popular face features are compared to verify the effectiveness of FCRM. Experiments show that the FCRM performs the best when using the MD-RiuLBP feature. Experimental results prove that the proposed FCRM takes into consideration both the security and the fault tolerance and can prevent the imposters while keeping high accuracy.

Keywords: Face Recognition, Biometric Encryption, Bio-hashing.

1 Introduction

In recent years, biometric technologies have been applied in various domains and show a huge market potential. Biometric features are inherent and could not be changed. Once the biometric information saved in the servers was stolen, it would bring serious security problems. Hence, to put the biometric recognition technologies into large-scale usage, adequate security and confidentiality should be bound with the biometric feature. Biometric Encryption, as the combination of biometrics and cryptography technology, is proposed to ensure the safe usage of biometrics.

Biometric Hashing is one of the most popular strategies of Biometric Encryption. It binds biometric with a secret key and save it as a cipher text. Hence, the system does not need to save the raw biometric information of a particular user and thus the attackers could not steal the biometric information easily. If the biometrics provided at the authentication stage is close enough to the samples provided at registration, the system could generate the same key as created at the registration stage. In this way, more safety and convenience are introduced in comparison with the traditional password authentication.

During the last decade, a lot contribution has been made to solve the biometric encryption problems. Juel [1] firstly proposed the idea of Fuzzy Commitment Scheme (FCS) base on the coding theory, which has become one of the standard frameworks in biometric encryption. Fu et al [2] realized the FCS and did experiments on a small number of samples. Zhao [3-4] used feature fusion technology at the feature extraction

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 50–57, 2011.

stage. It fuses two different facial features at the feature level. Experimental results show that when the authentication face is the same as the registered face, the two feature vectors generated by the algorithm will be very close to each other. Zhang et al [5] adopted Neural network to alternative Hamming code as error correction and obtain FRR (False Rejection Rate) =7.5% when FAR (False Acceptance Rate) =2.46%. One disadvantage of FCS is that the vector should be of specific order, hence Juel proposed to use the Reed-Solomon (RS) code instead of the hamming code in the FCS algorithm [6]. Currently, some FCS-based algorithms are designed for multiple-template applications which lowered its practicality in real applications. Another drawback is that the FAR is relatively high for these algorithms.

In face biometrics hashing, Teoh et al [7] proposed the benchmark methods named Random Multi-space Quantization (RMQ) [7] that could result in a very high level recognition rate and the error rate could be controlled in a very low level when the user's token is correct. Chen [8] also implemented a face hashing algorithm and applied it to an encryption system based on RMQ and Local Binary Pattern (LBP) feature. Since the security of RMQ is based on the token or user key, the invaders would have a great chance to get authenticated only if he gets the user's token even without the face feature [10].

To handle these problems, we propose a novel face-hashing algorithm FCRM that takes into consideration both the security and feasibility. Instead of utilizing the token or user key, the FCRM adopts a cyclic random mapping process. We perform both face authentication and identification test with the proposed FCRM and several popular LBP features such as the most classical Uniform LBP feature (ULBP), the Rotation Invariant Uniform LBP feature (RiuLBP) and Multi-directional RiuLBP feature (MD-RiuLBP). Experiments show that the FCRM performs exceptionally well when using the MD-RiuLBP feature.

2 The Realization of FCRM

The FCRM has two major features. First, the template is bound with a set of random keys and stored in a changeable and cipher text way to protect the security of the user's biometric. Second, the key of FCRM has multiple functions.

2.1 The Framework of FCRM

The core of the FCRM algorithm is a cyclic random mapping process as shown in Fig. 1. In each cycle, the encryption model utilizes the key of the previous cycle as a random seed to generate a mapping matrix and makes random mapping of the face features. The cyclic random mapping uses the avalanche effect to enhance the security of the algorithm. A little difference in the face hash will result in prominent change of the released key.

As any other biometrics system, the proposed FCRM contains registration stage and recognition stage. In the registration stage, the FCRM uses the key of the last circle as a seed to generate a random matrix. After random mapping the face feature

with this matrix, the binarization is performed to output the face hash. Then a random key with error-correcting code is bound with the face hash as the output of each circle. The output of each circle and the hash value of the key will be stored in the server as the template. The recognition stage is the reverse to the registration stage. Use the face hash of the previous circle to extract the key from the template and use the key to do the random mapping in current circle.

We also adopt the error correction coding and decoding as fault-tolerant technology in FCRM to reduce the impact of random noise existing in features.

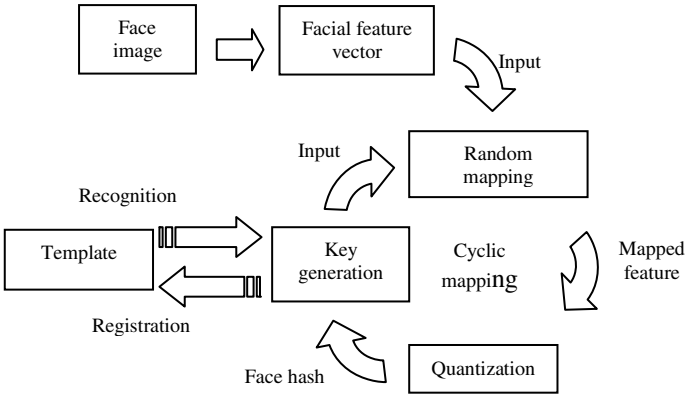


Fig. 1. Framework of the FCRM

2.2 Face Feature Extraction

In previous study, we found that some features will have great deformation after random mapping attributing to the inherent over descriptiveness of the high dimensional feature and results in essentially random noise in the high order projective components. In this paper, several low dimensional LBP features are compared.

The LBP feature is one of the dominant methods in face recognition. Firstly, the original face image is transformed to an LBP image with the LBP operators determined by the radius of neighborhoods R and the sampling density P. And then the image is blocked into M-by-N blocks to reserve the space structure of face. The LBP histograms of all blocks are concatenated into the feature. Most of the face recognition algorithms use ULBP proposed by Ojala [9]. The RiuLBP is another popular LBP feature neglecting the direction information in LBP coding. The ULBP and RiuLBP features are denoted as $LBP^{u2}_{(P,R)}(M,N)$ and $LBP^{Riu2}_{(P,R)}(M,N)$ respectively.

In our previous study, we found that with the feature-level fusion of multi-directional RiuLBP features, the feature dimension is drastically decreased while the precision is comparable or even better than the widely adopted ULBP features [11]. By splitting the P neighbors of a pixel into several uniformly distributed

non-overlapped subsets as, several neighbors are created with the same sampling density but different initial angels $\theta_i (i \in \{1,2,...,k\})$, each subset is denoted as $S(\theta_i, P_i)$. The regular RiuLBP feature for each $S(\theta_i, P_i)$ $(i \in \{1,2,...k\})$ is denoted as $LBP^{Riu2}_{(P_i,\theta_i,R)}(M,N)$. A feature level fusion of these feature denoted as $\bigcup_{i=1}^{k} LBP^{Riu2}_{(P_i,\theta_i,R)}(M,N)$ is defined as MD-RiuLBP feature, which is with better precision in face recognition than the other types of LBP feature utilizing the same neighbors, but of much lower dimension.

2.3 Random Mapping

Random mapping is an effective dimension reduction method as shown in Eqn. (1).

$$v = \kappa R \omega \tag{1}$$

where $R \in \Re^{p \times m} (m \le p)$ denotes the random matrix, p and m are respectively the dimension before and after random mapping, and κ is a constant. According to the Johnson-Lindenstrauss (J-L) lemma, the distance difference of any set of N points in p -dimensional Euclidean space can be reserved after being embedded into a $O(\frac{\ln(N)}{\varepsilon^2})$ dimensional space with random mapping [7].

The randomness of the mapping is aroused by the mapping matrix. In FCRM, the user's key is used as the random seed s_n to create the random matrix $R(s_n, m)$, and the process of random mapping is shown in Eqn.(2)

$$v_n = R(s_n, m) \cdot \omega \tag{2}$$

where n represents the current cycle number, $v_n \in \Re^m$ is the m-dimensional vector after random mapping of the feature ω, $s_n = \{k_0, \cdots, k_{n-1}\}$ (k_i presents the random key generated at the i-th cycle, $0 \le i \le n-1$, k_0 is a predefined constant). The size of m can be selected to adapt to actual safety requirement.

2.4 Feature Binarization

We first adopt zero-mean normalization in the FCRM as shown in Eqn. (3),

$$v'_n = v_n - \overline{v}_n \tag{3}$$

where $\overline{v}_n$ is the mean of feature vector v_n. The face hash h_n is calculated as Eqn. (4)

$$h_n^{(k)} = \begin{cases} 0 & v_n'^{(k)} \ge 0 \\ 1 & v_n'^{(k)} < 0 \end{cases} \tag{4}$$

2.5 Template Generation in Registration Stage

In the registration stage, a random key k_n generated in the n-th cycle will be firstly encoded into c_n with the error correction algorithm (Hamming code in this paper) as shown in Eqn. (5).

$$c_n = E(k_n) \tag{5}$$

Then the binding of c_n and the face hash h_n is calculated as Eqn. (6)

$$b_n = c_n \oplus h_n \tag{6}$$

where the symbol $\oplus$ represents the binary space bitwise XOR.

The binding result and the hash value of the key will be saved as a template $T_n = \{b_n, H(k_n)\}$ in the server. $H(x)$ denotes hash function such as the Secure Hash Algorithm 1(SHA-1) or the Message-Digest Algorithm (MD5) algorithms. In order to protect the security of keys, the hash value of the key is saved in the template instead of the clear-text key.

2.6 Key Extraction and Judgment in Authentication and Identification

In the authentication stage, we first perform random mapping to the face feature vector to obtain face hash h'_n, and then recover the key with error correction code extracted from the template as shown in Eqn. (7)

$$c'_n = b_n \oplus h'_n \tag{7}$$

Lastly, we decode c'_n according to Eqn. (8) to obtain the key.

$$k'_n = D(c'_n) \tag{8}$$

The decoding process helps to ensure the correct acceptance rate of the legitimate users due to its error correction ability. In template generation stage, k_n has an error-correcting code in it. So in the authentication stage, if the face hash is close enough to that got in the registration stage, the random noise in the feature can be ignored by the error-correcting code, and extract a key the same as the template. The correct acceptance rate could be controlled to adapt to different security levels by adjusting the error correction rate. If the hash value of the decoded key is the same as that in the template, the user is considered as legal otherwise the authentication fails.

FCRM can also be used for face identification. The difference is that in authentication the face hash is not stored for safety consideration, but in identification, we save the face hash of every user to calculate the distance between the login image and all images of the registered users.

3 Experimental Analysis

Our experiments are performed on the frontal subset of FERET database, which contains 2398 frontal-face images from 1199 subjects and 2 for each subject. 1199 is the maximum number of persons in FERET with at least 2 frontal images in the database. The 2 images are randomly picked from all frontal images of the person. Hence, the database involves multiple variations in expression, hairstyle, glasses and age. All face images are preprocessed to align the face region geometrically to size 160*140, mask the background and balance the lighting around face area. The key length is set to the most popular 128 bits in the experiments.

3.1 Validation of Binarization and Random Mapping

To validate the effect of random mapping, we show in Fig. 2 the separability of raw feature, feature after 1-cycle RMQ and feature after 3-cycle RMQ with the histogram of distance for MD-RiuLBP feature. It can be observed that after only 1-cycle random mapping, the seperability keeps the same as the raw feature with L_1 distance, which shows that random mapping is effective in dimension reduction. However, the overlapping part of the genuine and imposter is still large. While with the proposed 3-cycle random mapping, the seperability is prominently enlarged.

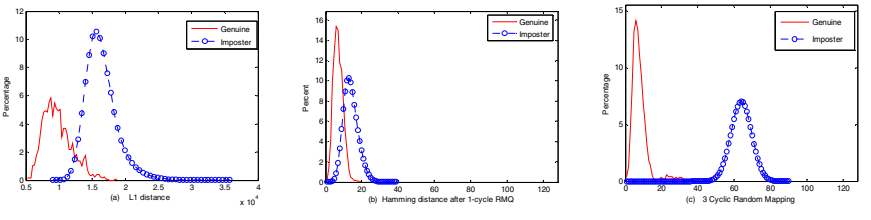


Fig. 2. Comparison of seperability after random mapping for MD-RiuLBP feature

3.2 Performance of Authentication

We compared the proposed FCRM with other authentication algorithms as shown in Table 1, in which the FCRM using $k_1 = 9, k_2 = 9, k_3 = 110$. Due to the randomness of FCRM, we do 10 times of experiments for each group of parameter settings of face features, and the result in Table 1 is based on the median precision among the 10 tests. The results show that FCRM can result in similar performance as other algorithms. However, the number of subjects in our experiments is much larger than that of the others. Another feature of FCRM is when the FAR is near 0, the FRR is as low as 3.42%.

We also explore the influence of key length to the recognition rate in Table 2. Normally too short key will be conquered easily, while too long key will decline the recognition rate. According to the results, we can find that with the growth of key length in each cycle, FRR increases while FAR decreases. So we can adjust the key length according to the actual needs of the applications.

Table 1. Performance of Authentication of FCRM

Algorithm	Database	#subject	#Images	FRR (%)	FAR (%)
FCRM+MD-RiuLBP	FERET	1199	2398	2.25	1.36
FCRM+MD-RiuLBP	FERET	1199	2398	3.34	0.025
PCA+NN [5]	ORL	40	400	7.5	2.46
PCA+RMQ-90 [7]	FERET	300	1200	1.56	1.31
RHLBPQ-90 [8]	ORL	40	400	0.6837	0.021

Table 2. Analysis of the key length

k1 length	k2 length	RiuLBP FRR (%)	RiuLBP FAR (%)	MD-RiuLBP FRR (%)	MD-RiuLBP FAR (%)
6	6	5.30	0.00176	4.59	0.00179
9	9	10.26	6.96×10^{-6}	7.38	$<1\times10^{-6}$
12	12	15.38	$<1\times10^{-6}$	9.01	$<1\times10^{-6}$

3.3 Performance of Identification

We evaluate the identification performance of FCRM with the statistical values of correct recognition rate from the 10 tests as summarized in Table 3. It can be observed that the FCRM in combination with the MD-RiuLBP feature could results in greatly improved performance in identification compared with RiuLBP feature. While for ULBP, the performance is poor. This is due to the high dimension of the feature.

Table 3. Identification results of FCRM

Feature	Dimension	L_1	FCRM+Hamming Max(%)	Min(%)	Mean(%)	Median(%)
$LBP^{Riu2}_{(16,2)}(7,8)$	1008	75.813	94.245	79.483	89.024	92.494
$LBP^{u2}_{(8,2)}(7,8)$	3304	81.401	84.571	73.311	78.257	80.901
$\bigcup_{i=1}^{4} LBP^{Riu2}_{(4,\theta_i,2)}(7,8)\ \theta_i=\{0,\frac{\pi}{8},\frac{\pi}{4},\frac{3\pi}{8}\}$	1344	80.234	97.415	83.236	92.377	96.664

4 Conclusions

In this paper, we propose a novel biometric encryption algorithm FCRM for face recognition. Through 3 times of recycling random mapping on the face biometric, the FCRM could prominently improve the security and performance of face recognition especially when taking the MD-RiuLBP feature. In FCRM, we use error-correcting code in the key to avoid the impact of random noise, and use the template scheme to protect the user's biometric and the security of keys. Experimental results demonstrate that the FCRM can achieve very low FAR while reserves an acceptable FRR

compared with other available algorithms. Moreover, the proposed FCRM is tested with only one template per subject in registration, which has practical merits for storage and computation in some applications. In addition, the level of security could be adjusted with the FCRM to meet the requirements of different applications through changing the key length.

Acknowledgement. This paper is supported by the National Natural Science Foundation of China (60605012); the Natural Science Foundation of Shanghai (08ZR1408200); the Shanghai Leading Academic Discipline Project (J50103); the Open Project Program of the National Laboratory of Pattern Recognition of China.

References

1. Juels, A., Wattenberg, M.: A Fuzzy Commitment Scheme. In: Proceedings of the 6th ACM Conference on Computer and Communications Security, pp. 28–36. ACM Press (1999)
2. Fu, B., Li, J.P.: Error-tolerant generation of biometric key from face features (in Chinese). J. Application Research of Computers. 25(1), 260–262 (2008)
3. Zhao, Z., Paul, W.: A face hashing algorithm using mutual information and feature fusion. In: Proceedings of the 2007 IEEE International Conference on Networking, Sensing and Control, pp. 386–391. IEEE, London (2007)
4. Zhao, Z., Paul, W.: A Novel Face hashing Method with Feature Fusion for Biometric Cryptosystems. In: Proceedings of the Fourth European Conference on Universal Multiservice Networks, pp. 439–444. IEEE, Toulouse (2007)
5. Zhang, D.X., Tang, Q.S., Lu, X.J., Zhu, H.G.: Cipher Key Management Based on Neural Networks and Facial Biometrics Feature. Journal of Northeastern University (Natural Science) 30(6), 817–820 (2009) (in Chinese)
6. Juels, A., Sudan, M.: A Fuzzy Vault Scheme. Designs, Codes and Cryptography 38(6), 237–257 (2006)
7. Teoh, J., Goh, A., Ngo, L.: Random Multispace Quantization As an Analytic Mechanism for Biohashing of Biometric and Random Identity Inputs. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(12), 1892–1901 (2006)
8. Chen, N.N.: Research on LBP-Based Face Crypto and Its Application for the Crypto System. Nanjing University of Aeronautics and Astronautics (2008) (in Chinese)
9. 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)
10. Kong, B., Cheung, K., Zhang, D., et al.: An Analysis of Biohashing and Its Variants. Pattern Recognition 39(7), 1359–1368 (2006)
11. Fang, Y., Luo, J., Lou, C.: Fusion of multi-directional rotation invariant uniform LBP features for face recognition. In: IITA 2009, Nanchang, China, vol. 2, pp. 332–335 (2009)

Color Face Recognition Based on Statistically Orthogonal Analysis of Projection Transforms

Jiangyue Man[1], Xiaoyuan Jing[1,2,3,*], Qian Liu[1], Yongfang Yao[1], Kun Li[1], and Jingyu Yang[4]

[1] College of Automation, Nanjing University of Posts and Telecommunications, China
[2] State Key Laboratory of Software Engineering, Wuhan University, China
[3] State Key Laboratory for Novel Software Technology, Nanjing University, China
[4] College of Computer Science, Nanjing University of Science and Technology, China
jingxy_2000@126.com

Abstract. In this paper, we propose a novel color face feature extraction approach named statistically orthogonal analysis (SOA). It in turn calculates the projection transforms of the red, green and blue color component image sets by using the Fisher criterion, and simultaneously makes the obtained transforms mutually statistically orthogonal. SOA can enhance the complementation and remove the correlation between discriminant features separately extracted from three color component image sets. Experimental results on the AR and FRGC version 2 color face image databases demonstrate that SOA achieves better recognition results than several related color face recognition methods.

Keywords: Color face recognition, feature extraction, statistically orthogonal analysis (SOA), projection transform.

1 Introduction

Color images are increasingly used in the field of face recognition because they can afford more useful information than grayscale images [1] [2]. The RGB color space is a basic and widely used color space, and other color spaces (or color models) are usually defined by linear or nonlinear transformations of the RGB color space. The key of color image recognition technique is to effectively utilize the complementary information between the red (R), green (G) and blue (B) components. But these color components are usually highly correlated, and this correlation leads to little complementary information. Therefore, reducing the correlation should contribute to enhance the complementation and further improve the recognition performance. Some methods have tried to reduce the correlation. Shih and Liu [2] showed that the color configuration YQC_r, where Y and Q components are from the YIQ color space and C_r is from the YC_bC_r color space, is effective for face recognition by using the enhanced Fisher linear discriminant model (EFM) [3]. Yang and Liu [4] presented an extended general color image discriminant (EGCID) algorithm that acquires three uncorrelated

* Corresponding author.

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 58–65, 2011.

groups of combination coefficients to separately produce three new color component images, and then performed PCA+EFM on the concatenated new color component images. Liu [5] presented the uncorrelated color space (UCS), the independent color space (ICS) and the discriminating color space (DCS) to form effective color image representations and used EFM to do classification by concatenating their color component images.

The linear discriminant analysis technique is a very important research topic in feature extraction. Typical linear discriminant analysis methods include linear discriminant analysis (LDA) [6], Foley-Sammon LDA [7], EFM [3], Direct LDA [8], uncorrelated LDA (ULDA) [9], semi-supervised discriminant analysis [10] and Perturbation LDA [11]. In particular, ULDA makes the acquired projective vectors satisfy both the Fisher criterion and the statistically orthogonal constraints.

Current color face feature extraction methods (including UCS, ICS, DCS, EGCID, etc.) first transform the original RGB space to new color spaces, and then employ the commonly used linear discriminant analysis technique to extract features and do classification. These methods reduce the correlation between three color components of the original images. However, this decorrelation is not directly connected with feature extraction. ULDA removes the statistical correlation between extracted features by making the acquired projective vectors mutually statistically orthogonal. But it is only designed for a single sample set and is quite time-consuming because it calculates the projective vectors one by one to form a projection transform. Inspired by ULDA, we propose a statistically orthogonal analysis (SOA) for color face feature extraction, which can remove the statistical correlation between discriminant feature sets of three color components. SOA in turn calculates the projection transforms of R, G and B color component image sets with the Fisher criterion [6], and simultaneously makes the obtained transforms mutually statistically orthogonal.

We use the AR [12] and face recognition grand challenge (FRGC) version 2 [13] color face image databases to evaluate the proposed SOA approach. Experimental results demonstrate that SOA acquires better recognition results than several related color face feature extraction methods.

2 Statistically Orthogonal Analysis (SOA)

Motivated by ULDA, we think that the discriminant features of R, G and B color component image sets will be uncorrelated if these features separately lie in three mutually statistically orthogonal feature spaces. Based on this, SOA obtains three mutually statistically orthogonal projection transforms of the R, G and B color component image sets in a serial manner. Without loss of generality, we calculate the projection transforms in order of R, G and B components. (There are six calculation orders for the projection transforms of R, G and B components, i.e., the RGB, RBG, GRB, GBR, BRG and BGR orders. We have already validated that the experimental results of these six calculation orders are very close.)

In this paper, SOA is used in the RGB color space, while the realizations of SOA in other color spaces are not discussed. The reason is that we have already experimented on the HSV, YC_bC_r and YIQ color spaces and the obtained results are similar to the experimental results obtained from the RGB color spaces.

First we show the notations used in our approach:

X_R, X_G, X_B :	R, G and B color component image sets of the color image samples, respectively;
S_{bR}, S_{bG}, S_{bB} :	Between-class scatter matrices of X_R, X_G, X_B, respectively;
S_{wR}, S_{wG}, S_{wB} :	Within-class scatter matrices of X_R, X_G, X_B, respectively;
S_{tR}, S_{tG}, S_{tB} :	Total scatter matrices of X_R, X_G, X_B, respectively;
W_R, W_G, W_B :	Projection transforms consisting of projective vectors for X_R, X_G, X_B, respectively;
$\lvert\bullet\rvert$:	Determinant of a square matrix;
$\sqrt{M}$:	A matrix satisfying $\sqrt{M}\left(\sqrt{M}\right)^T = M$.

2.1 Projection Transform of R Component

Based on the Fisher criterion, we calculate W_R by

$$\max\ J(W_R) = \frac{\left|W_R^T S_{bR} W_R\right|}{\left|W_R^T S_{wR} W_R\right|}. \tag{1}$$

As proved by LDA, we can achieve W_R by solving the following eigenequation:

$$S_{wR}^{-1} S_{bR} W_R = \lambda W_R, \tag{2}$$

Hence W_R is a matrix that consists of the eigenvectors corresponding to the nonzero eigenvalues of $S_{wR}^{-1} S_{bR}$.

2.2 Projection Transform of G Component

For two samples $x_1 \in X_R$ and $x_2 \in X_G$, let $y_1 = W_R^T x_1$ and $y_2 = W_G^T x_2$ separately denote the projected features of x_1 and x_2. The covariance between y_1 and y_2 is

$$\begin{aligned}\mathrm{Cov}(y_2, y_1) &= E\left[y_2 - E(y_2)\right]\left[y_1 - E(y_1)\right]^T \\ &= W_G^T \left\{E\left[x_2 - E(x_2)\right]\left[x_1 - E(x_1)\right]^T\right\} W_R = W_G^T \sqrt{S_{tG}} \left(\sqrt{S_{tR}}\right)^T W_R,\end{aligned} \tag{3}$$

where $\sqrt{S_{tG}} = E\left[x_2 - E(x_2)\right]$ and $\sqrt{S_{tR}} = E\left[x_1 - E(x_1)\right]$. The auto variances of y_1 and y_2 separately are

$$\mathrm{Var}(y_1, y_1) = E\left[y_1 - E(y_1)\right]\left[y_1 - E(y_1)\right]^T = W_R^T S_{tR} W_R \tag{4}$$

and

$$\mathrm{Var}(y_2, y_2) = E\left[y_2 - E(y_2)\right]\left[y_2 - E(y_2)\right]^T = W_G^T S_{tG} W_G. \tag{5}$$

The correlation between y_1 and y_2 can be defined as

$$\mathrm{Corr}(y_2, y_1) = \frac{\mathrm{Cov}(y_2, y_1)}{\sqrt{\mathrm{Var}(y_2, y_2)}\sqrt{\mathrm{Var}(y_1, y_1)}} = \frac{W_G^T \sqrt{S_{tG}} \left(\sqrt{S_{tR}}\right)^T W_R}{\sqrt{W_G^T S_{tG} W_R}\sqrt{W_R^T S_{tR} W_R}} . \tag{6}$$

The correlation measured by Formula (6) contains the statistical information of original samples, which is provided by S_{tG} and S_{tR}. To remove the statistical correlation, we set $\mathrm{Corr}(y_2, y_1) = 0$, which is equivalent to $W_G^T \sqrt{S_{tG}} \left(\sqrt{S_{tR}}\right)^T W_R = 0$. We consider that W_R and W_G are statistically orthogonal if $\mathrm{Corr}(y_2, y_1) = 0$.

We can obtain W_G by solving the following problem:

$$\begin{aligned} &\max \ J(W_G) = \frac{\left|W_G^T S_{bG} W_G\right|}{\left|W_G^T S_{wG} W_G\right|} \\ &s.t. \quad W_G^T \sqrt{S_{tG}} \left(\sqrt{S_{tR}}\right)^T W_R = 0 . \end{aligned} \tag{7}$$

For solving Formula (7), we present a theorem as follows:

Theorem 1. W_G in Formula (7) can be achieved by solving the eigenequation:

$$S_{wG}^{-1}\left(I - W(W^T S_{wG}^{-1} W)^{-1} W^T S_{wG}^{-1}\right) S_{bG} W_G = \lambda W_G , \tag{8}$$

where $W = \sqrt{S_{tG}}(\sqrt{S_{tR}})^T W_R$, and I is an identity matrix. W_G consists of the eigenvectors associated with the nonzero eigenvalues of $S_{wG}^{-1}\left(I - W(W^T S_{wG}^{-1} W)^{-1} W^T S_{wG}^{-1}\right) S_{bG}$.

The proof is given in Appendix A.

2.3 Projection Transform of B Component

Like W_G, the projection transform W_B is required to be statistically orthogonal to both W_R and W_G. We can calculate W_B by

$$\begin{aligned} &\max \ J(W_B) = \frac{\left|W_B^T S_{bB} W_B\right|}{\left|W_B^T S_{wB} W_B\right|} \\ &s.t. \quad W_B^T \sqrt{S_{tB}} (\sqrt{S_{tR}})^T W_R = 0 \\ &\qquad W_B^T \sqrt{S_{tB}} (\sqrt{S_{tG}})^T W_G = 0 \quad . \end{aligned} \tag{9}$$

For solving Formula (9), we present a theorem as:

Theorem 2. W_B in Formula (9) can be achieved by solving the eigenequation:

$$S_{wB}^{-1}\left(I - W(W^T S_{wB}^{-1} W)^{-1} W^T S_{wB}^{-1}\right) S_{bB} W_B = \lambda W_B , \tag{10}$$

where $W = \left[\sqrt{S_{tB}} \left(\sqrt{S_{tR}} \right)^T W_R,\ \sqrt{S_{tB}} \left(\sqrt{S_{tG}} \right)^T W_G \right]$, and I is an identity matrix. W_B is a matrix that consists of the eigenvectors associated with the nonzero eigenvalues of $S_{wB}^{-1} \left(I - W(W^T S_{wB}^{-1} W)^{-1} W^T S_{wB}^{-1} \right) S_{bB}$.

Formula (7) can be transformed into Formula (9) via separately replacing W_R, W_G, S_{bG} and S_{wG} by W, W_B, S_{bB} and S_{wB}. Thus the proof of Theorem 2 is similar to that of Theorem 1.

2.4 Realization Algorithm of SOA

Step 1. Calculate W_R according to Formula (2).

Step 2. Calculate W_G according to Formula (8).

Step 3. Calculate W_B according to Formula (10).

Step 4. Project the R, G and B color component images by $Y_R = W_R^T X_R$, $Y_G = W_G^T X_G$ and $Y_B = W_B^T X_B$, and obtain the new sample set $Y = \left[Y_R^T, Y_G^T, Y_B^T \right]^T$.

Step 5. Use the nearest neighbor classifier with cosine distance to do classification.

3 Experiments

We use the AR [12] and face recognition grand challenge version 2 (FRGC-v2) [13] color face image databases to evaluate the proposed SOA approach.

The AR database contains over 4000 images of 126 people. We select 2652 (=102×26) images from 102 individuals for our experiments. All images are cropped to the size of 60×60. Fig. 1 illustrates all cropped images of one subject.

Fig. 1. Demo images of one individual on AR database

The FRGC-v2 database contains 12776 training images, 16028 controlled target images and 8014 uncontrolled query images. The controlled images have good image quality, while the uncontrolled images display poor image quality. The training image set, which consists of both controlled and uncontrolled images, contains 222 individuals, each 36-64 images. We choose 36 images of each person to comprise the experimental sample set, and crop every image to 60×60. All cropped images of one subject are shown in Fig. 2.

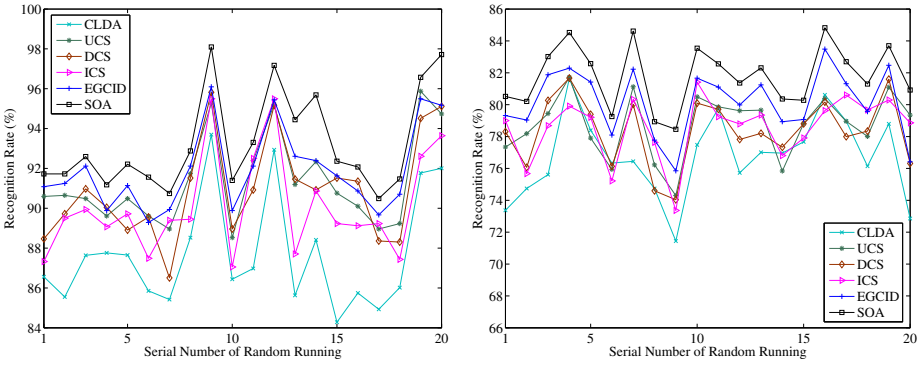
Fig. 2. Demo images of one individual on FRGC-v2 database

In our experimens, we apply LDA and ULDA to color images, which are called color LDA (CLDA) and color ULDA (CULDA), respectively. For each color image sample, they combine the R, G and B components into a vector. Then, they extract discriminant features from the vectors corresponding to color image samples.

We compare the proposed SOA approach with the related methods including CLDA, UCS, ICS, DCS and EGCID. Here, we do not compare CULDA since it obtains different eigenvalues in experiments, and thus CULDA acquires the same recognition results as CLDA due to the related theory in [9].

On the AR and FRGC-v2 databases, we randomly select 8 images per person for training, use the remainder for testing, and run all methods for 20 times. With regard to all compared methods, we perform PCA transform to reduce the sample dimensionality before feature extraction, which can also avoid the singularity of the within-class scatter matrix in the discriminant analysis technique. For all compared methods, we search the appropriate number of principal components, which can yield the best recognition result.

Fig. 3 (a) and (b) show the recognition rates of all compared methods for 20 random running on the AR and FRGC-v2 databases, respectively. Table 1 compares the recognition rates of all compared methods on the AR and FRGC-v2 databases. Compared with CLDA, UCS, ICS, DCS and EGCID, SOA separately improves the average recognition rates at least by 1.33% (=93.27%-91.94%) and 1.64% (=81.80%-80.16%) on the AR and FRGC-v2 databases.



(a) on AR database (b) on FRGC database

Fig. 3. Recognition rates of all compared methods

Table 1. Comparison of recognition results

Method	Recognition rate (%) (Mean and standard deviation)	
	AR database	FRGC-v2 database
CLDA	87.69±2.71	76.73±2.52
UCS	91.34±2.22	78.71±1.94
ICS	90.91±2.50	78.34±2.07
DCS	90.11±2.50	78.58±1.94
EGCID	91.94±2.03	80.16±2.04
SOA	93.27±2.40	81.80±1.88

4 Conclusions

In this paper, we propose a novel color face feature extraction and recognition approach named statistically orthogonal analysis (SOA). SOA in turn performs discriminant analysis on the red, green and blue color component image sets, and simultaneously removes the statistical correlation between discriminant features separately extracted from these three image sets. Experimental results on the AR and FRGC-v2 color face image databases demonstrate that SOA outperforms several representative color face recognition methods.

Appendix A: Proof of Theorem 1

Given $W = \sqrt{S_{tG}}\left(\sqrt{S_{tR}}\right)^T W_R$, the constraint can be rewritten as $W_G^T W = 0$. We construct the Lagrange function:

$$L(W_G) = W_G^T S_{bG} W_G - \lambda\left(W_G^T S_{wG} W_G - C_1\right) - \mu\left(W_G^T W - C_2\right), \tag{11}$$

where λ and μ are the Lagrange multipliers, and C_1 and C_2 are two constant matrices. We set the derivative of $L(W_G)$ on W_G to be zero:

$$\frac{\partial L(W_G)}{\partial W_G} = 2S_{bG}W_G - 2\lambda S_{wG}W_G - \mu W = 0. \tag{12}$$

Multiplying Eq. (12) by $W^T S_{wG}^{-1}$, we have

$$2W^T S_{wG}^{-1} S_{bG} W_G - \mu W^T S_{wG}^{-1} W = 0. \tag{13}$$

Thus μ may be expressed as

$$\mu = 2(W^T S_{wG}^{-1} W)^{-1} W^T S_{wG}^{-1} S_{bG} W_G. \tag{14}$$

Due to Eqs. (12) and (14), we have

$$S_{bG}W_G - \lambda S_{wG}W_G - W(W^T S_{wG}^{-1} W)^{-1} W^T S_{wG}^{-1} S_{bG} W_G = 0, \tag{15}$$

that is,

$$S_{wG}^{-1}\left(I - W(W^T S_{wG}^{-1} W)^{-1} W^T S_{wG}^{-1}\right) S_{bG} W_G = \lambda W_G , \tag{16}$$

where I is an identity matrix. Eq. (16) is equivalent to Formula (8). Proof is over.

Acknowledgement. The work described in this paper was fully supported by the NSFC under Project No.61073113, the New Century Excellent Talents of Education Ministry under Project No. NCET-09-0162, the Qing-Lan Engineering Academic Leader of Jiangsu Province, the Foundation of Jiangsu Province Universities under Project No.09KJB510011, the Research Fund for the Doctoral Program of Higher Education of China under Project No. 20093223110001.

References

1. Rajapakse, M., Tan, J., Rajapakse, J.: Color Channel Encoding with NMF for Face Recognition. In: Int. Conf. Image Processing, vol. 3, pp. 2007–2010 (2004)
2. Shih, P., Liu, C.: Improving The Face Recognition Grand Challenge Baseline Performance Using Color Configurations across Color Spaces. In: Int. Conf. Image Processing, pp. 1001–1004 (2006)
3. Liu, C., Wechsler, H.: Robust Coding Schemes for Indexing and Retrieval from Large Face Databases. IEEE Trans. Image Processing 9(1), 132–137 (2000)
4. Yang, J., Liu, C.: Color Image Discriminant Models and Algorithms for Face Recognition. IEEE Trans. Neural Network 19(12), 2088–2098 (2008)
5. Liu, C.: Learning The Uncorrelated, Independent, and Discriminating Color Spaces for Face Recognition. IEEE Trans. Information Forensics and Security 3(2), 213–222 (2008)
6. Martinez, A.M., Kak, A.C.: PCA versus LDA. IEEE Trans. Pattern Anal. Mach. Intell. 23(2), 228–233 (2001)
7. Foley, D.H., Sammon, J.W.: An Optimal Set of Discriminant Vectors. IEEE Trans. Computer 24(3), 281–289 (1975)
8. Yu, H., Yang, J.: A Direct LDA Algorithm for High-dimensional Data with Application to Face Recognition. Pattern Recognition 34(10), 2067–2070 (2001)
9. Jing, X.Y., Zhang, D., Jin, Z.: UODV: Improved Algorithm and Generalized Theory. Pattern Recognition 36(11), 2593–2602 (2003)
10. Zhang, Y., Yeung, D.Y.: Semi-supervised Discriminant Analysis Using Robust Path-based Similarity. In: IEEE Conf. Computer Vision and Pattern Recognition, pp. 1–8 (2008)
11. Zheng, W.S., Lai, J.H., Yuen, P.C., Li, S.Z.: Perturbation LDA: Learning The Difference between The Class Empirical Mean and Its Expectation. Pattern Recognition 42(5), 764–779 (2009)
12. Martinez, A.M., Benavente, R.: The AR Face Database. CVC Technical Report #24 (1998)
13. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of The Face Recognition Grand Challenge. In: IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 947–954 (2005)

Real-Time Head Pose Estimation Using Random Regression Forests

Yunqi Tang, Zhenan Sun, and Tieniu Tan

National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
{yqtang,znsun,tnt}@nlpr.ia.ac.cn

Abstract. Automatic head pose estimation is useful in human computer interaction and biometric recognition. However, it is a very challenging problem. To achieve robust for head pose estimation, a novel method based on depth images is proposed in this paper. The bilateral symmetry of face is utilized to design a discriminative integral slice feature, which is presented as a 3D vector from the geometric center of a slice to nose tip. Random regression forests are employed to map discriminative integral slice features to continuous head poses, given the advantage that they can maintain accuracy when a large proportion of the data is missing. Experimental results on the ETH database demonstrate that the proposed method is more accurate than state-of-the-art methods for head pose estimation.

Keywords: Head pose estimation, Random regression forests, Integral slice features.

1 Introduction

Automatic estimation of head pose from range images has become an active topic in recent years, inspired by the availability of affordable and reliable depth camera, and the increasing demands of many applications such as face recognition, human computer interaction and facial feature analysis. Compared with 2D-based methods [2,3,4,5], estimation of head pose with depth images has several advantages. Firstly, pixel value of depth image has physical meaning, indicating the distance from objects to the viewpoint. Secondly, it is easy to detect and segment head area from depth images. Thirdly, depth images are illumination invariant. With these advantages, we proposed a robust method to estimate head pose from depth images in [13], which achieves encouraging performance.

In this paper, a more accurate method is proposed based on our previous work [13] to estimate head pose from depth images. In this method, one more dimension, namely difference of depth value, is included in the atom vector of Integral Slice Features (ISF) [13]. It is a more discriminative feature for indicating different head poses. Furthermore, a more powerful regressor, random forests [11], is employed to map the discriminative integral slice features to real-valued parameters of head poses. The performance of this method is evaluated on ETH [9] and our results are superior to the state-of-the-arts.

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 66–73, 2011.

The rest of this paper is organized as follows. Section 2 presents the existing work of head pose estimation. In section 3, we describe the technical details of proposed head pose estimation method. Section 4 discusses the experimental results on the public database, and section 5 concludes the paper.

2 Related Work

Head pose estimation has been a hot research topic over the last decade. There is a large amount of literature on this topic [1]. Generally, existing methods can be classified into three categories: appearance template based methods [9,4], geometry based methods [6,12] and learning based methods [2,3,7,5,8,10].

Appearance template based methods [9,4] compare directly a new head image with a set of exemplars, which are labeled with discrete poses, in order to find the most similar image and take its angle as the estimation result. For instance, the work of [9] firstly used geometric features to get some hypothesis nose tips, then compared the new image with a set of template images and finally took the most similar template's orientation as the estimation result. Although these methods do not require negative training samples and facial feature points, they are time consuming and cannot estimate a continuous head pose.

Geometry based methods [6,12] take facial features points, such as the inner and outer corners of eyes, nose tip and left and right corners of mouth, to calculate head pose directly. The advantage of these methods is that they do not need training process, but the performance of these methods highly depends on the accuracy of facial feature point detection.

Learning based methods [7,5,8,10] use the tools of machine learning to map the input images or feature data to discrete or continuous head poses. For example, Fanelli et al. [7] synthesized a large head pose database to train a mapping from single depth features to real-valued parameters. However, these methods may suffer from over fitting.

This paper is a tradeoff between geometry based method and learning based method. It takes the position of nose tip as precondition to design a new efficient feature and uses the tools of machine learning to estimate the parameters of head poses. Nose tip is one of the most significant points of face and can be detected easily compared with other feature points. The feature based on nose tip has physical meaning and can accurately indicate different head poses. The experimental results on public dataset show that the performance of proposed method achieves state-of-the-arts [9,7,13] performance.

3 Head Pose Estimation with Random Regression Forests

In this section, we detail the discriminative integral slice features and describe how random regression forests are used to map depth features to head poses.

3.1 Discriminative Integral Slice Features

A depth image is an image or image channel that contains information related to the distance from the surfaces of scene objects to the camera. Therefore a head's depth image can be regarded as a 3D surface. Then a vertical slice of head's depth image can be defined as a set of pixels whose values belong to $[l, h]$. We formulate a slice of image I as

$$S_\omega(I) = \{(x, y) | l \leq d_I(x, y) \leq h\} \tag{1}$$

where $S_\omega(I)$ is a vertical slice of depth image I, $\omega = (l, h)$ denotes the slicing parameters (low threshold l and high threshold h), and $d_I(x, y)$ denotes the depth value of pixel (x, y). Thus the integral center of this slice can be defined as

$$C_\omega(I) = (\sum_{(x_i,y_i)\in S_\omega(I)} x_i/N(S_\omega(I)), \sum_{(x_i,y_i)\in S_\omega(I)} y_i/N(S_\omega(I))) \tag{2}$$

where $C_\omega(I)$ is the coordinate of the integral center of slice S_ω, $N(S_\omega(I))$ denotes the number of pixels within a slice, and (x_i, y_i) is the coordinate of pixel i in depth image I. For a given slice S_ω, the feature is defined as a 3-dimentional vector from the geometric center of slice to a reference point, which can be formulated as

$$F_\omega(I) = (F_\omega(I)|_x, F_\omega(I)|_y, F_\omega(I)|_z) \tag{3}$$

where $F_\omega(I)|_x = C_\omega(I)|_x - R_x$ is x-component, $F_\omega(I)|_y = C_\omega(I)|_y - R_y$ is y-component, $F_\omega(I)|_z = d_I(C_\omega(I)) - d_I(R)$ is z-component, and R is the coordinate of reference point which can be the nose tip and optionally a slice center.

It is well known that face is bilateral symmetric about nose tip. Thus the 3-dimension vector from the geometric center of a slice of head surface to nose tip has significant physical meaning. $F_\omega(I)|_x$ and $F_\omega(I)|_y$ indicate roll information of a head pose, $F_\omega(I)|_x$ and $F_\omega(I)|_z$ provide yaw information of a head pose, and $F_\omega(I)|_y$ and $F_\omega(I)|_z$ provide pitch information of a head pose.

Figure 1 illustrates five features crossing different poses. We can see that if somebody turns his or her head to right, $F_\omega(I)|_x$ would have a positive response, and negative response for left; if somebody raises his or her head up, $F_\omega(I)|_y$ would have a negative response when the slices with low threshold l and h, and positive response for down.

3.2 Random Regression Forests

As a bagging based classifier, random forest have lots of advantages. It is able to yielding good results with common datasets; It have a comparable performance with Adaboost, while more robust than Adaboost; and due to its parallel structure, it can be implemented efficiently on GPU. A forest consists of a number of parallel decision trees. Each tree can be trained or used to predicate separately. There are two types of nodes within each tree : split node and leaf node. A split

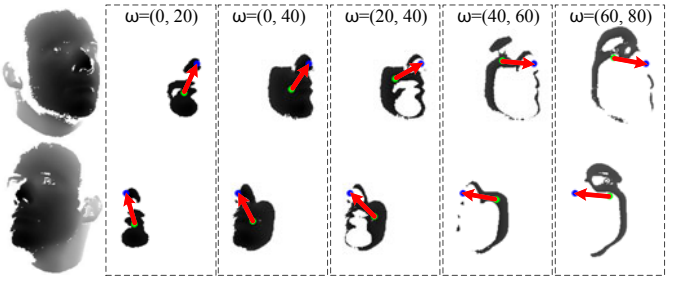


Fig. 1. Examples of discriminative integral slice features crossing different poses. There are five slices for each head depth image with the different parameter ω. The blue points are the position of nose tip. The green points are the geometric center of slices. The red arrows are the vectors from geometric center of slices to nose tips.

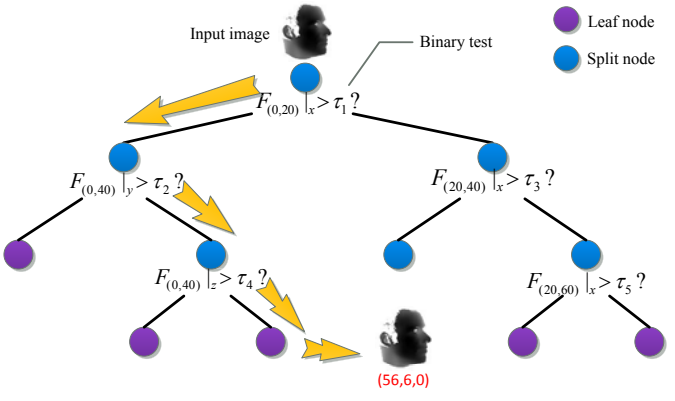


Fig. 2. How head pose is estimated with a decision tree. The purple circle represents leaf node. The blue circle represents split node. The orange arrow denotes the path of the input image going down from the root to a leaf.

node is always corresponding to a binary test, which directs samples towards the left or right child. A leaf node can be regarded as a cluster of samples that can be described by a simple model. During training process, the samples falling in a leaf node are averaged to get the model's parameters, and the parameters are stored in this leaf node. During testing process, the new sample goes down from the root of decision trees till leaf nodes are reached. Then the average model of these leaf nodes is used to predicate its result. Figure 2 simply illustrates how head pose is estimated with a decision tree.

Training. Usually it is impossible to accurately describe a large dataset using a uniform distribution. Thus it is natural to divide the large dataset into dozens of small datasets, and employ simple distributions to describe these small datasets. This is the main idea of decision trees, namely divide-and-conquer. Each split node divides the sample set reached it into two subsets according the result of a binary test:

$$F_\omega(I)|_{\theta=\{x,y,z\}} > \tau \tag{4}$$

where τ is the threshold of the binary test; and the leaf node is related to the small set that can be described with a simple distribution. We model the pose of a head as a 3-dimention variation $P = (\alpha, \beta, \gamma)$, which follows the distribution of multivariate Gaussian at each leaf node. We build a decision tree recursively starting from the root node with the following steps:

Firstly, randomly select a subset of training samples which are composed of discriminative integral slice features and the annotated poses of heads.

Secondly, select a subset of features by randomly generating a set of binary tests $T = (\omega, \theta, \tau)$ according to Equation (4).

Thirdly, for the samples (S) reaching the node, if the number of these samples is bigger than a fixed threshold, then split the samples into left and right subsets (S_L and S_R) using each test t in T:

$$S_L = \{I|F_\omega(I)|_{\theta=\{x,y,z\}} < \tau, t(\omega, \theta, \tau) \in T, I \in S\} \tag{5}$$

$$S_R = \{I|F_\omega(I)|_{\theta=\{x,y,z\}} \geq \tau, t(\omega, \theta, \tau) \in T, I \in S\} \tag{6}$$

and compute its information gain, which is defined as:

$$IG(t) = H(S) - (w_L H(S_L) + w_R H(S_R)) \tag{7}$$

where H denotes differential entropy. w_L is the ratio between the number of samples in S_L and S, and w_R is the ratio between the number of samples in S_R and S.

Fourthly, compute the largest information gain

$$IG(t^*) = \arg\max_{t \in T}(IG(t)) \tag{8}$$

If $IG(t^*)$ is below a fixed threshold, then keep it as a split node and assign the parameters of corresponding test $t^*$to the node. Or keep it as a leaf node and store the mean and covariance of P in this node.

Finally, if the depth of the tree is below a maximum value, then go to the third step for recursively building the left and right children.

Testing. When we get a new depth image of head, discriminative integral slice features are extracted and sent to the trained forest. In each tree, the new sample goes down from the root, and is recursively directed to left or right child according the binary test stored in split nodes until a leaf node is reached. The distributions of all reached leaf nodes are averaged to generate the final result:

$$p(d|l) = \frac{1}{N} \sum_{i=1}^{N} p_i(d|l) \tag{9}$$

where N denotes the number of trees within a forest, $p_i(d|l)$ means the distribution of the leaf node reached by the given sample in the ith tree.

4 Experiments

To evaluate the performance of proposed method, we conducted an experiment on ETH Face Pose Range Image Data Set [9] and compared the experimental results with the state of the art methods [9,7,13] whose results are also based on the dataset of ETH. ETH is a large and public database with about 10k range images covering almost all possible rotations information of head including yaw from $-90°$ to $+90°$ and pitch from $-45°$ to $+45°$. This paper focuses on the problem of head pose estimation, thus assumes that head and nose tip are detected in the image. And noise of depth images is removed with the method of [13].

Our method is mainly controlled by 3 parameters which are the l, h and the number of trees n. l and h contribute to the production of feature slices. Each pair of l and h can result in a slice which is related with a feature of images. For a depth image, there would be $\sum_{i=0}^{255}(255-i) = 32640$ slices at most. In present setup, we set l and $d = h - l$ as increasing arithmetic sequences of $[0, 240]$ with a common difference of $\Delta d = 10$. Thus there will be $3 * \sum_{i=0}^{24}(24-i) = 900$ features for each image.

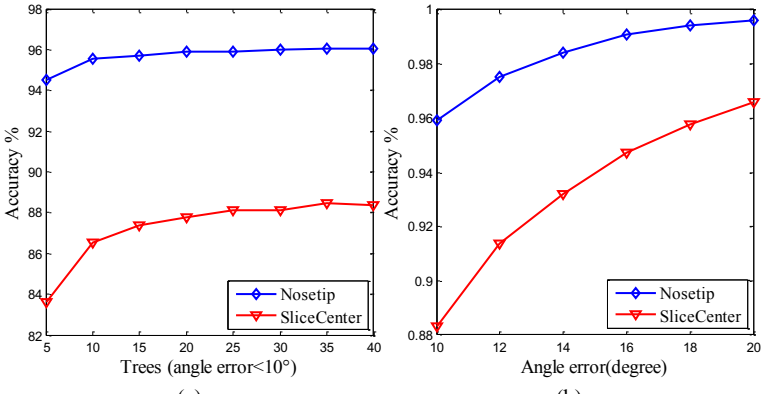


Fig. 3. (a) Estimation accuracy as a function of the number of trees within a forest, with the threshold of angle error is set to 10°. (b) Estimation accuracy as a function of the angle error threshold.

The plots in Fig.3 show the estimation accuracy of our method as a function of the number of trees and angle error threshold. The blue plots represent that nose tip is used as reference point, and the red plots represent that slice center of $C_{(0,220)}$ is used as reference point. As we can see that the accuracy of using nose tip as reference point is greatly better than using slice center as reference point. The reason is that nose tip is one of significant points of face. Thus the feature based on the vector from a slice center to nose tip is powerful to describe the changes of head pose.

Table 1. Comparison of proposed method with state-of-the-arts. The first two columns show mean and standard deviation of yaw error and pitch error. The third column shows the estimation accuracy with different error thresholds of 10°, 15° and 20°.

	Yaw error(°)	Pitch error(°)	Estimation accuracy(%)
Proposed method	1.5/1.9	2.5/3.8	95.9/98.5/99.6
Tang et al.	1.5/3.0	2.1/5.5	93.6/97.2/99.0
Breitenstein et al.	6.1/10.3	4.2/3.9	80.8/97.8/98.4
Fanelli et al.	5.7/15.2	5.1/4.9	90.4/95.4/95.9

The comparison of results is show in Table 1, including mean and standard deviation of estimation error and the estimation accuracy with different angle error thresholds. The experimental results of Breitenstein [9] and Fanelli [7] are equoted from their papers. The results presented in the second row are the results our previous work [13] which is based on general regression neural networks. Compared with [13], the improvements of this paper are two-fold: Firstly, the feature vector from slice center to reference point is extended from 2 dimensions to 3 dimensions. Secondly, a more powerful classifier, random forests, is employed to replace general regression neural networks. From the table, we can see that the method presented by this paper achieves the lowest error and highest accuracy.

Table 2. Computation time for different parts of proposed algorithm

Preprocessing	16 ms
Feature extraction	9 ms
Random forests regression	1 ms
Total time	26 ms(38.5fps)

According the sequence of processing, we divide the algorithm into three parts: preprocessing, feature extraction and regression, and use the preprocessing algorithm of [13] to remove noise. Finally, we implemented this algorithm using C++ and tested it on a computer with an Intel Core 2 CPU @2.83GHZ*2 and 4GB RAM. The averaged computation time of each part of this method for processing a frame is presented in Table 2. On average, our algorithm could achieve a speed of 38.5 fps, which can meet the requirement of real-time system.

5 Conclusion

In this paper, we propose a robust and accurate method for head pose estimation. In this method, the atom vector of Integral Slice Features (ISF) is extended from 2D to 3D, which is more powerful to describe the changes of head poses. In our experiment, a quantitative evaluation is performed on a public dataset. The experimental results show that our method achieves state-of-the-art performance. Furthermore, the proposed method is easy to implement and does not

rely on special hardware, thus the system based on our method could have a wide application.

Acknowledgments. This work is funded by National Natural Science Foundation of China (Grant No. 60736018, 61075024) and International S&T Cooperation Program of China (Grant NO.2010DFB14110).

References

1. Murphy-Chutorian, E., Trivedi, M.M.: Head pose estimation in computer vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 607–626 (2009)
2. Balasubramanian, V.N., Ye, J., Panchanathan, S.: Biased manifold embedding: A framework for person-independent head pose estimation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–7 (2007)
3. Huang, C., Ding, X., Fang, C.: Head pose estimation based on random forests for multiclass classification. In: International Conference on Pattern Recognition, pp. 934–937 (2010)
4. Ng, J., Gong, S.: Composite support vector machines for detection of faces across views and pose estimation. Image and Vision Computing 20, 359–368 (2002)
5. Huang, D., Storer, M., De la Torre, F., Bischof, H.: Supervised Local Subspace Learning for Continuous Head Pose Estimation. In: IEEE Conference on Computer Vision and Pattern Recognition (2011)
6. Malassiotis, S., Strintzis, M.G.: Robust real-time 3D head pose estimation from range data. Pattern Recognition 38, 1153–1165 (2005)
7. Fanelli, G., Gall, J., Van Gool, L.: Real Time Head Pose Estimation with Random Regression Forests. In: IEEE Conference on Computer Vision and Pattern Recognition (2011)
8. Seemann, E., Nickel, K., Stiefelhagen, R.: Head pose estimation using stereo vision for human-robot interaction. In: Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 626–631 (2004)
9. Breitenstein, M.D., Kuettel, D., Weise, T., Van Gool, L., Pfister, H.: Real-time face pose estimation from single range images. In: IEEE Conference on Computer Vision and Pattern Recognition, pp.1–8 (2008)
10. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., Blake, A.: Real-time human pose recognition in parts from single depth images. In: IEEE Conference on Computer Vision and Pattern Recognition (2011)
11. Breiman, L.: Random forests. Machine Learning 45, 5–32 (2001)
12. Wang, J.G., Sung, E.: EM enhancement of 3D head pose estimated by point at infinity. Image and Vision Computing 25, 1864–1874 (2007)
13. Tang, Y., Sun, Z., Tan, T.: Face Pose Estimation based on Integral Slice Features of Single Depth Images. In: Asian Conference on Pattern Recognition (2011)

Head Pose Estimation Using Simple Local Gabor Binary Pattern

Weijun Hu[1], Bingpeng Ma[1], and Xiujuan Chai[2]

[1] School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China
[2] Key Lab of Intelligent Information Processing of Chinese Academy of Sciences(CAS), Institute of Computing Technology, CAS, Beijing 100190, China

Abstract. In this paper, a novel method named simple Local Gabor Binary Pattern method (sLGBP) is presented to improve the accuracy of head pose estimation. The motivation of sLGBP comes from the great success of Local Gabor Binary Pattern (LGBP) in many areas. Considering the relationship between the symmetry of face image and the head pose, the Gabor filters and the LBP operators in sLGBP are all based on the one-dimension. By this way, feature extracted by sLGBP is more related to the head pose while the compute efficiency is improved greatly. To show the effectiveness of sLGBP, we compared them with other methods under two different databases. The results of the experiments show that the proposed methods can improve the accuracy of head pose estimation.

Keywords: Head Pose Estimation, Local Binary Pattern, Gabor filters; Local Gabor Binary Pattern.

1 Introduction

Statistics indicate that approximately 75% of the faces in photographs are non-frontal [1]. However, the best known face perception systems can only deal with near-frontal faces reliably, and the performances of these systems degrade dramatically on non-frontal faces. Therefore, pose-invariant face perception has been an active research topic for several years. To achieve the expected robustness to pose variation, one may expect to process face images differently according to their pose parameters. In this case, the pose of the input faces must be estimated as a prerequisite for sequent processes.

The methods of head pose estimation can be categorized into two main groups [2]: model-based methods [3] [4] [5] and appearance-based methods [6] [7] [8] [9]. Typically, the model-based methods build 3D models for human faces and attempt to match the facial features such as the face contour and the facial components of the 3D face model with their 2D projections. Since these methods generally run very fast, they can be used in video tracking and multi-camera surveillance. However, these methods are sensitive to the misalignment of the facial feature points, while the accurate and robust localization of facial landmarks remains an open problem.

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 74–81, 2011.

The appearance-based methods typically assume that there exists a certain relationship between the 3D face pose and some properties of the 2D facial image and use a large number of training images to infer the relationship by using statistical learning techniques. The unified projection is calculated in a closed-form solution based on the graph embedding linearization, and then project new data into the embedded low-dimensional subspace with the identical projection.

In this paper, we proposed a novel discriminative feature named simple Local Gabor Binary Pattern(sLGBP). On one side, Local Gabor Binary Pattern (LGBP) is applied in face recognition and attained the impressive result on FERET database[10]. Then, in the paper [11], LGBP is applied in head pose estimation and improve the accuracy of head pose estimation greatly. In LGBP, a face image is modeled as a "histogram sequence" by concatenating the histograms of the local Gabor magnitude binary pattern maps. And it is impressively insensitive to appearance variations due to lighting, expression, and misalignment. The effectiveness of LGBP benefits from several aspects including the multi-resolution and multi-orientation Gabor decomposition, the LBP operator, and the local spatial histogram modeling. Though LGBP can arrive the good performance in head pose estimation, it is difficult to be applied in the real-time system for its complexity.

On other side, the asymmetry of 2D facial appearance has been applied in head pose estimation[12]. B. Ma et al. has argued that the asymmetry in the intensities of each row of the face image is closely relevant to the yaw rotation of head, and at the same time evidently insensitive to the identity of the input face. Specifically, to extract the asymmetry information, 1D Gabor filters and Fourier transform are exploited.

Finally, based on the success of LGBP and the asymmetry of 2D facial appearance, in this paper, sLGBP is proposed to reduce the computational complexity of LGBP while keep the relationship between the feature and the head pose variations. In sLGBP, the Gabor filters and the LBP operators are all based on the one-dimension. By this way, on one hand, compared with the 2D-dimensional Gabor filters and 2D-dimension-LBP operators, the compute efficiency of sLGBP is improved greatly; on other hand, the features extracted by sLGBP is more related to the pose variations. To show the effectiveness, we test sLGBP on the two different databases. The results of the experiments show that sLGBP can improve the accuracy of head pose estimation.

The remaining part of this paper is organized as follows: in Section 2, we introduced how to use the simple LGBP to extract the feature of the face image; experiments are given in Section 3. Conclusions are drawn in Section 4.

2 Feature Description

In this second, we first introduced the LGBP method in briefly, and then proposed the sLGBP method.

2.1 Local Gabor Binary Pattern

Recently, LGBP is used for face recognition and other areas. And it is impressively insensitive to appearance variations due to lighting, expression, and misalignment. The effectiveness of the LGBP benefits from several aspects including the multi-resolution and multi-orientation Gabor decomposition, the LBP operator, and the local spatial histogram modeling. The first step of LGBP is to convolute a face image with the Gabor filters $G(\mu, \nu)$. The processing of facial images by Gabor filters is chosen for its biological relevance and technical properties. Generally, Gabor filters are employ a discrete set of 5 different scales, with $\nu = 0, \cdots, 4$, and 8 orientations, with $\mu = 0, \cdots, 7$. And then 40 Gabor Magnitude Pictures(GMPs) can be calculated.

In the second step, LBP operator operates on each GMP. The original LBP operator labels the pixels of an image by threshold the pixels $f_p(p = 0, \cdots, 7)$ of 3×3 neighborhood with the center value f_c and considering the result as a binary number $S(f_p - f_c)$. Then, by assigning a binomial factor 2^p for each $S(f_p - f_c)$, the LBP pattern at the pixel can be achieved, which characterizes the spatial structure of the local image texture. In LGBP, the transform result of $G(\mu, \nu)$ is $\mathbf{LG}(\mu, \nu)$. For the 40 GMPs, there are 40 **LG**s for each image.

In the third step, to enhance the representation of LGBP, some operations are applied. First, to keep the spatial information of the multi-view face images, LGBP is operated on many sub-regions of the images. In addition, the histogram information is extracted and concatenated into a single histogram sequence **LH**. **LH** is taken as our final feature description in LGBP. More details about LGBP can be find from paper [10].

2.2 Simple Local Gabor Binary Pattern

Though LGBP can improve the accuracy of head pose estimation, it is difficult to be applied in the real-time system, considering the computational consume and the memory requirement. For example, in the stage of Gabor feature extraction, the time consume of 40 filters with 5 scales and 8 orientations is very large. Extracting the histogram information from the sub-region is consume a lot of times. In addition to the time consume, the memory requirement of LGBP is other bottleneck. For the image with 32×32, the dimension of 40 Gabor features is $40,960 = 32 \times 32 \times 5 \times 8$. The 40 Gabor features increase the time consume of the stage of histogram sequence, the time consume is also large since for each Gabor feature.

In this paper, we propose sLGBP and argue that sLGBP can be applied in head pose estimation. In sLGBP, the Gabor filters and LBP operators are both based on the one-dimension. Obviously, compared with the 2D-dimensional Gabor filters and 2D-dimension-LBP operators of LGBP, the compute efficiency of sLGBP is improved greatly.

Besides the advantage of the efficiency, the features extracted by sLGBP is more related to the pose varies. For the appearance based method of head pose estimation, the features used by the traditional methods are extracted from the

entire face, which is generally vectorized as 1D vector which lose the face structure in some sense; therefore, these features contain not only pose information, but also information about identity, lighting, expression, etc. Understandably, given a representation of the same dimension, its discriminative ability will be inevitably lower if more non-pose information is preserved. In paper [12], in order to eliminate the influence of lighting and noise, multiple-scale 1D Gabor filters are further exploited to filter the images before Fourier transform. This paper also pointed out that the pose information of the head can computed from the row of the 2D intensity image. So in sLGBP, 1D Gabor filters are used on the row of the images and the sLGBP feature is relevant to pose but less relevant to other properties (e.g. the identity) of the face image.

The multi-scale Gabor filters have similar shapes as the receptive fields of simple cells in the primary visual cortex [13]. 1D Gabor filters can be defined as:

$$g_\mu(r) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-r^2}{2\sigma^2}} e^{i(2\pi\mu r)} \tag{1}$$

where μ is the modulation frequency and σ is the scale parameter which determines the width of the Gaussian envelope. The Gabor representation of a signal is the convolution of the signal with a family of Gabor filters. For the gray row signal $s(r)$ of an image, the convolution result $O_\mu(r)$ corresponding to the Gabor filter at frequency μ can be defined as follows:

$$O_\mu(r) = s(r) * g_\mu(r) \tag{2}$$

where $*$ denotes the convolution operator.

Considering sLGBP is applied to extract the features of the row signal of the image, in sLGBP, the neighborhood is also based on the 1D LBP. In Fig. 1, we introduce the neighborhood of the 1D LBP. Unlike the original 2D LBP operator labels the pixels of an image by thresholding the 3×3 neighborhood, in our 1D LBP operator, the neighborhood is set to 1×5 and the center pixel is f_c, and the neighborhood is $\{f_p, p = 0, 1, 2, 3\}$. To reduce the dimension of the sLGBP feature, in sLGBP, f_3 is redefined as the combination of f_0 and f_3: $f_3 = (f_0 + f_3)/2$. Then, the result as a binary number can be defined as follows:

$$S(f_p - f_c) = \begin{cases} 0, (f_p < f_c, p = 1, 2, 3) or (p = 0) \\ 1, (f_p \geq f_c, p = 1, 2, 3) \end{cases} \tag{3}$$

By assigning a binomial factor 2^p for each $S(f_p - f_c)$, the 1D LBP pattern at the pixel f_c is achieved as:

$$LBP = \sum_{p=1}^{3} S(f_p - f_c) 2^{(p-1)} \tag{4}$$

which characterizes the spatial structure of the local image texture.

To enhance the representation ability, sLGBP is operated on many sub-regions of the images. In our experiment, for a 32×32 image, the region is set as 2×2.

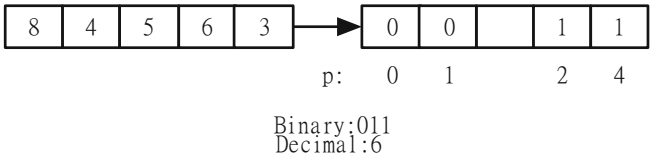


Fig. 1. The neighborhood of sLGBP

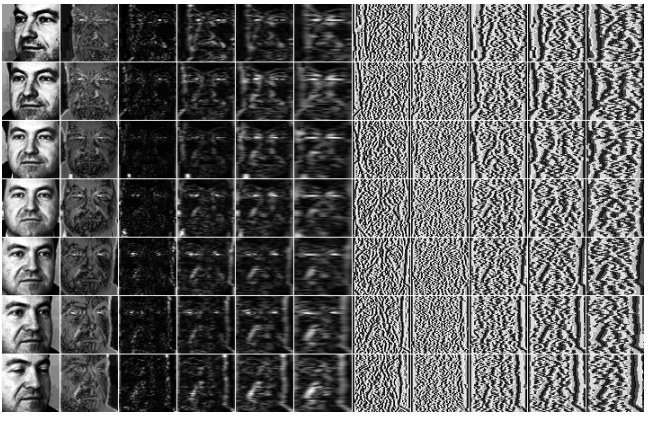


Fig. 2. The face images and its feature images of sLGBP

And for the number of the bins of each region is 1, the dimension of sLGBP is $10,240 (= 5 \times 256 \times 8)$.

In Fig. 2, we show the feature images of sLGBP. In the image, the face images in the first column are the original images; the images from the second column to the six column are the 1D Gabor features with 5 scales; the images from the seven column to the eleven column are the features of sLGBP.

3 Experiments

In this section, the proposed sLGBP are evaluated on the face database by comparing with other feature extraction methods.

There are two databases applied in the experiments. One is the public CAS-PEAL database [14], the other is the private Multi-Poses database. In practical applications, the imaging conditions of the testing images might be not homogeneous with those of the training ones. In other words, system developers may hardly know what kind of data will be presented to the system in practical applications. Therefore, it is very significant to evaluate a system by using some testing data heterogeneous with the training data. Essentially, heterogeneous testing is highly related to the generalizability problem in pattern recognition.

The Multi-Pose database consists of $3,030$ images of 102 subjects taken under normal indoor lighting conditions and fixed background with a Sony EVI-D31

camera. The yaw angles and the pitch angles range within [−50°, +50°] and [−50°, +50°] with intervals of 1°, respectively. The sample number is 30 for each class(i.e. yaw angle).

The CAS-PEAL database contains twenty-one poses combining seven yaw angles (−45°, −30°, −15°, 0°, 15°, 30° and 45°) and three pitch angles (30°, 0° and −30°). We use a subset containing totally 4200 images of 200 subjects whose IDs range from 401 through 600.

In the experiments, the Multi-Pose database is taken as the training dataset while the CAS-PEAL database is taken as the testing dataset. The images in the Multi-Pose database are captured by Sony EVI-D31 camera, while the images in the CAS-PEAL database are captured by a simple USB camera. Inevitably, these two databases are constructed with different population under different imaging conditions, such as different lighting conditions, background, which are very appropriate for our purpose.

For all the images, the face detection method [15] is applied to locate the face region from the input images, and then all the face regions are normalized to the same size of 32×32. Finally, histogram equalization is used to reduce the influence of lighting variations.

We compare the performance of sLGBP with the following methods: PCA, LDA, Gabor Fisher Classifier (GFC), Gabor Fourier Feature (GaFour), Gabor Fourier Fisher feature (GFFF) and LGBP. As one of the baseline methods in face recognition, PCA can be seen as the baseline method in appearance-based pose estimation. The LDA-based baseline algorithm, similar to the Fisherfaces method, applies first PCA for dimensionality reduction and then LDA for discriminant analysis. Since GaFour and GFFF arrived the best accuracy on the heterogeneous database [12], they are taken as the compared method in this paper. As sLGBP can be the simple of LGBP, LGBP is also taken as the compared method. In the experiments, for all the methods, PCA is used after feature extraction to reduce the dimension of features and 95% of total energy of eigenvalues is kept.

In Fig. 3, we shown the error mean of yaw estimation with the different features on the NC classifier while the center number changes from 1 to 7. In the figure, the x-axis is the centroid number of each angle, and the y-axis is the error mean. In Tab. 1, we shown the error mean (unit °) of yaw estimation with the different features on the NC classifiers and the SVM classifier. For the NC classifier, we only list the error mean while the center number of each class is 7.

From the table and figure, several points can be drawn. First, compared with the accuracy of LGBP, the accuracy of sLGBP is much better. This can be attributed to the asymmetry of sLGBP, which means that the feature of sLGBP is related to the head pose by using the 1D Gabor filters and 1D LBP operators. Considering the advantage of computational efficiency, the advantage of sLGBP is more obviously.

Second, under SVM classifier, the accuracy 9.25° of sLGBP is much less than 12.65° of GaFour and 12.17° of GFFF. Under NC classifier, the error mean of

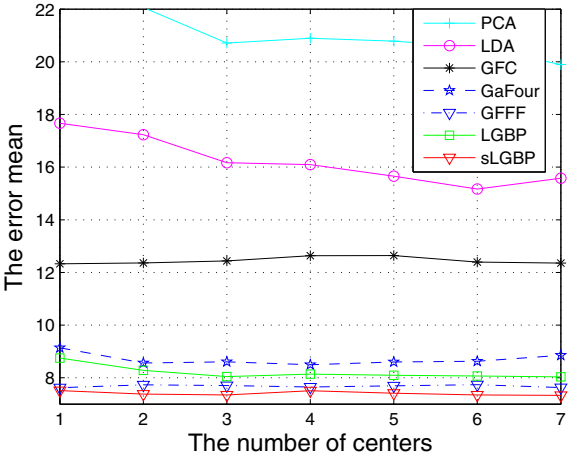


Fig. 3. Error mean of yaw estimation on the NC classifier

Table 1. Error mean of yaw estimation. The training and the testing sets are heterogeneous.

Method	PCA	LDA	GFC	GaFour	GFFF	LGBP	sLGBP
NC(k=7)	19.90	15.58	12.36	8.85	7.63	8.04	7.33
SVM	20.71	20.00	10.99	12.43	12.65	12.17	9.25

sLGBP is near those of GFFF for different center number. We argue that LBP characterizes the spatial structure of the local image texture.

Finally, the accuracy of sLGBP is the best of all the results under both NC classifier and SVM classifier, which means sLGBP can improve the accuracy of head pose estimation.

4 Conclusion

Based on that the feature should be related to the head pose, this paper proposed the sLGBP method. In sLGBP, 1D Gabor filters and 1D LBP operators are combined to extract the pose information while the computational consume is decreased greatly. The experiments show the effectiveness of sLGBP.

Acknowledgment. This paper is partially supported by National Natural Science Foundation of China under contract No. 61003103 and 61173065, Research Fund for the Doctoral Program of Higher Education under contract No. 20100142120029 and Fundamental Research Funds for the Central Universities under No. 2011QN048.

References

1. Kuchinsky, A., Pering, C., Creech, M.L., Freeze, D., Sera, B., Gwizdka, J.: Fotofile: A consumer multimedia organization and retrieval system. In: Proc. the SIGCHI Conference on Human Factors in Computing Systems: the CHI is the Limit, pp. 496–503 (1999)
2. Ji, Q., Hu, R.: 3D face pose estimation and tracking from a monocular camera. Image and Vision Computing 20(7), 499–511 (2002)
3. Cootes, T.F., Walker, K., Taylor, C.J.: View-based active appearance model. In: Proc. IEEE International Conference on Automatic Face and Gesture Recognition, pp. 227–232 (2000)
4. Krüger, N., Pötzsch, M., Malsburg, C.V.D.: Determination of face position and pose with a learned representation based on labeled graphs. Image and Vision Computing 15(8), 665–673 (1997)
5. Yan, S., Zhang, Z., Fu, Y., Hu, Y., Tu, J., Huang, T.S.: Learning a person-independent representation for precise 3D pose estimation. In: First International Evaluation Workshop on Classification of Events, Activities and Relationships, pp. 297–306 (2007)
6. Chen, L., Zhang, L., Hu, Y., Li, M., Zhang, H.: Head pose estimation using fisher manifold learning. In: Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp. 203–207 (2003)
7. Li, S.Z., Fu, Q., Gu, L., Scholkopf, B., Cheng, Y., Zhang, H.: Kernel machine based learning for multi-view face detection and pose estimation. In: Proc. IEEE International Conference on Computer Vision, pp. 674–679 (2001)
8. Krüger, V., Bruns, S., Sommer, G.: Efficient head pose estimation with Gabor wavelet networks. In: Proc. British Machine Vision Conference (2000)
9. Fu, Y., Huang, T.S.: Graph embedded analysis for head pose estimation. In: Proc. IEEE International Conference on Automatic Face and Gesture Recognition, pp. 3–8 (2006)
10. Zhang, W., Shan, S., Gao, W., Chen, X., Zhang, H.: Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition. In: International Conference on Computer Vision, vol. 1, pp. 786–791 (2005)
11. Ma, B., Yang, F., Gao, W., Zhang, B.: The Application of Extended Geodesic Distance in Head Poses Estimation. In: Zhang, D., Jain, A.K. (eds.) ICB 2005. LNCS, vol. 3832, pp. 192–198. Springer, Heidelberg (2005)
12. Ma, B., Shan, S., Chen, X., Gao, W.: Head yaw estimation from asymmetry of facial appearance. IEEE Transactions on Systems, Man, and Cybernetics–Part B: Cybernetics 38(6), 1501–1512 (2008)
13. Wiskott, L., Fellous, J.M., Krüger, N., Malsburg, C.V.D.: Face recognition by elastic bunch graph matching. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 775–779 (1997)
14. Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., Zhao, D.: The CAS-PEAL large-scale chinese face database and baseline evaluations. IEEE Transactions on Systems, Man and Cybernetics, Part A 38(1), 149–161 (2008)
15. Yan, S., Shan, S., Chen, X., Gao, W., Chen, J.: Matrix-structural learning (MSL) of cascaded classifier from enormous training set. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (2007)

Ethnic Classification Based on Iris Images

Hui Zhang[1,2], Zhenan Sun[2], Tieniu Tan[2], and Jianyu Wang[1]

[1] Shanghai Institute of Technical Physics, Chinese Academy of Sciences
[2] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
{zhanghui,znsun,tnt}@nlpr.ia.ac.cn, jywang@mail.sitp.ac.cn

Abstract. Iris texture is commonly thought to be highly discriminative between eyes and stable over individual lifetime, which makes iris particularly suitable for personal identification. However, iris texture also contains more information related to genes, which has been demonstrated by successful use of ethnic and gender classification based on iris. In this paper, we propose a novel ethnic classification method based on supervised codebook optimizing and Locality-constrained Linear Coding (LLC). The optimized codebook is composed of codes which are distinctive or mutual. Iris images from Asian and non-Asian are classified into two classes in experiments. Extensive experimental results show that the proposed method achieves encouraging classification rate and largely improves the ethnic classification performance comparing to existing algorithms.

Keywords: Iris recognition, Ethnic classification, Bag-of-Words model, Codebook optimizing.

1 Introduction

The iris is the annular part between the pupil and white sclera of human eye, whose texture is extremely rich. Iris is commonly regarded as a kind of biometrics with high unique patterns. Even genetically identical irises, right and left iris from one person have different textural appearance. After Daugman first introduced an iris recognition algorithm [1], a lot of work about iris recognition was conducted [2] [3] [4] [5]. However, iris patterns include more information, e.g., ethnic and gender. Some of previous work argued that the iris texture is race related [6] [7]. While focusing on small scales, the local features of iris are unique to each subject, whereas on certain large scales, the global feature distributions of irises are similar for a specific race. The texture distribution differences of irises from different races seem to dependent on the genes. Some example iris images from different races are shown in Fig. 1, including iris images of Asian and Non-Asian.

There are different kinds of applications of ethnic classification based on different systems or applications. For our iris recognition system, irises of Asian are the majority of subjects. The related applications are mainly about distinguishing Asian and non-Asian. This paper focuses on the Asian and non-Asian

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 82–90, 2011.

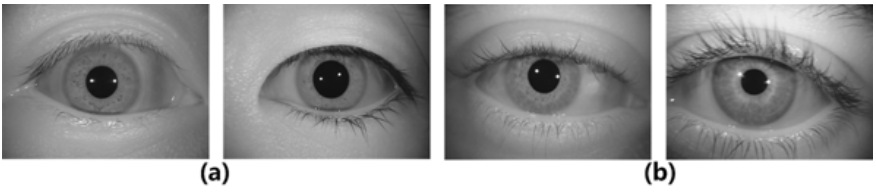


Fig. 1. Some examples of iris image from different races: (a) Asia; (b) Non-Asia

iris classification. For this task, Qiu et al. [7] has used the Iris-Texton, which is commonly thought to share the same concept with Bag-of-Words model (BoW). The basic idea is to approximate feature vectors by using vector quantization algorithm with a set of prototypes, and the prototypes (called codes or Textons) constitute a vocabulary. We use the term "code" for unification in this paper. Iris images from different races exhibit different patterns, while sharing some same patterns. However, the distribution of these patterns is different for different kinds of images. Therefore, the BoW model is suitable for ethnic classification.

In this paper, we propose a novel algorithm that combines supervised codebook optimization and Locality-constrained Linear Coding (LLC) for ethnic classification based on iris images. The proposed method takes into account the class label information for codebook optimization, which makes the codebook contain both distinctive and mutual codes for different classes. The distinctive code has different probability of occurrence frequencies in different classes, which is used to represent the difference of different classes. The mutual code is used for representing the iris image with little quantization error. The LLC is a simple but effective coding scheme. It applies locality constraint to select similar bases from the codebook and learns a linear combination weight of these bases to reconstruct each descriptor, and its speed-up version has high computation speed [8].

The remainder of this paper is organized as follows. Section 2 presents related work. Section 3 introduces the proposed method. Section 4 presents and discusses experiments. Section 5 concludes the paper.

2 Related Work

Racial and ethnic classification is an old topic in social science. It is often assumed to be a fixed trait based on ancestry. Some attempts have been made to perform automatic ethnic classification based on biometric images. Most of the early work is based on face images. Gutta et al. [9] used hybrid RBF/decision-trees which has similar architecture with Quinlans C4.5 algorithm for ethnic classification based on human face images. Gregory et al. [10] used a variant of AdaBoost for ethnic classification based on face images. They considered the well-defined binary categorization: Asian and non-Asian. Lu et al. [11] proposed a Linear Discriminant Analysis based scheme for Asian and non-Asian classification from face images.

Although iris textures are commonly thought to be highly discriminative between eyes, they still present several desirable common properties that could

be used for coarse iris classification. Within an iris, scales of the iris microstructures vary a lot along the radius, and these varies of different eyes are different. However, iris texture patterns from a same ethnic display a certain degree of consistence and correlation. There are some related work has been done. Ethnic classification methods based on iris and periocular images are proposed. Qiu et al. [6] used a bank of multichannel 2D Gabor filters to extract the global texture information from iris images and adopted AdaBoost to learn a discriminate classifier for ethnic classification. Later, Qiu et al. [7] proposed the Iris-Texton and SVM to classify the Asian and non-Asian based on iris images. In [12], authors used periocular biometrics for gender and ethnic classification. Local Binary Patterns are used as feature and SVM is used as the classifier.

3 Proposed Method

Fig. 2 shows the flowchart of the proposed approach for ethnic classification based on iris images. There are six main steps: (1)iris image preprocessing; (2)feature extraction; (3)codebook learning; (4)codes selection (codebook optimization); (5)image representing; (6)classifier training and testing using the linear SVM [13]. We will introduce the proposed approach in detail in this section.

3.1 Iris Image Preprocessing and Feature Extraction

The iris image preprocessing mainly includes iris detection, localization, segmentation, and normalization. We use the iris image preprocessing method proposed by He et al. [14] (see Fig. 3). The size of normalized unwrapped iris image used is 512×80 pixels.

The SIFT descriptor [15] has been well applied in computer vision in recent years. The densely extracted SIFT descriptor shows advantages in the object classification tasks when it cooperating with the BoW model. We use the SIFT descriptor as low level visual features to represent iris images. For the second step in the flowchart, SIFT descriptors [15] are extracted at regular pixel intervals in normalized iris images. We use six pixels interval which results in 913 descriptors for each iris image.

3.2 Codebook Optimization and Iris Image Representing

After low level visual feature extraction, we use the Locality-constrained Linear Coding (LLC) method [8] combined with code selection (codebook optimization) for iris image representation.

As shown in Fig. 2, the third step of the flowchart is codebook learning. We use the K-means method to learn a initial codebook based on the extracted feature pool. The Iris-Texton method introduced in [7] uses cluster centers learned by K-means as codes directly. The codebook optimizing method in [8] focuses on decrease the quantization error of feature reconstruction. These methods do

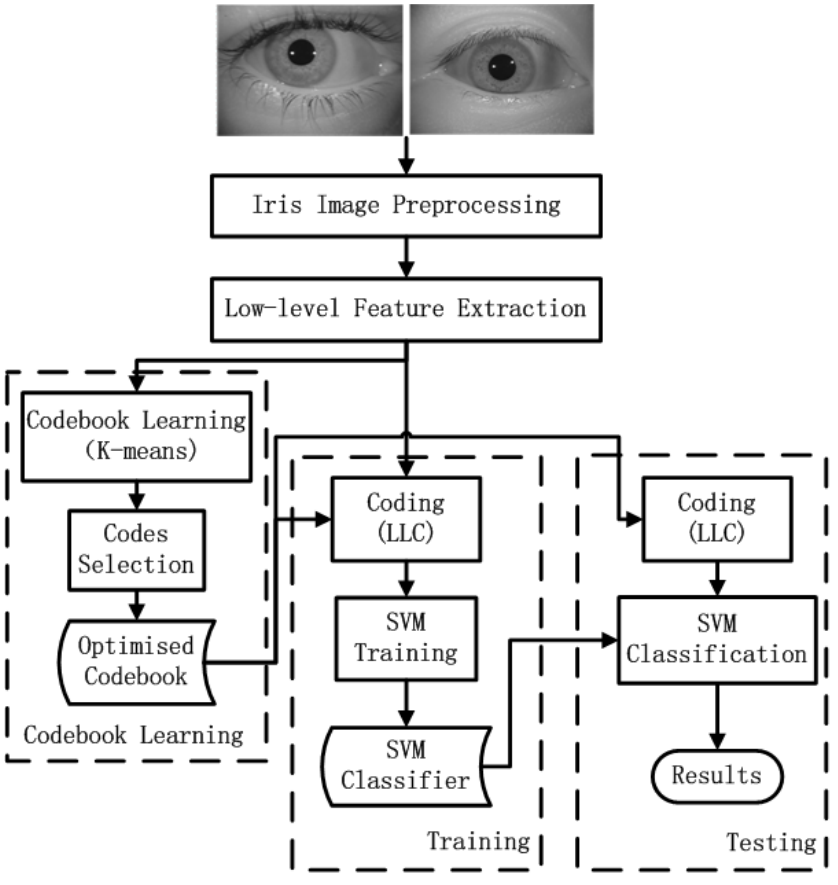


Fig. 2. Flowchart of the proposed approach for ethnic classification based on iris images

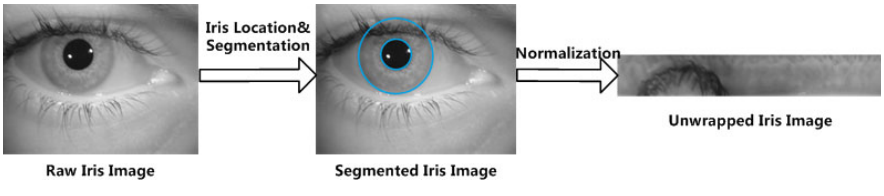


Fig. 3. Illustration of image preprocessing

not consider the label information of two classes. We utilize the label information during codebook optimization to select distinguishing codes for particular classification task, as while as mutual codes for keeping little quantization error.

The original codebook learnt by K-means is denoted as B, the final codebook used for coding is denoted as $\bar{B}$. The size of B (M) is larger than the size of $\bar{B}$ (N), and $M = 5120, N = 1024$ in our experiments. For codebook optimization,

statistic histograms of B are calculated through training images. For each feature extracted from training iris images, k nearest neighbor codes in B is found, then the corresponding bins in the codebook histogram increase. The normalization histogram indicates appearing frequencies of codes in a collection of iris images.

The first rule for code selection is the stability of codes. The stability is evaluated by histogram similarities that are calculated from different collection of iris images. The histograms of codebook B are calculated on 500 iris images (250 images from each class) and 1000 iris images (500 images from each class) respectively. If the frequencies of one code in these two histograms are quite different, this code is an unstable code which should be abandoned. In our experiments, we abandoned codes with K_n largest different frequencies, and $K_n = 1024$.

The second rule for code selection is the discrimination of codes. For two classes problem, by traversing training iris images from each class, two different histograms of two classes are calculated respectively, denoted as H_1 and H_2. If the appearing frequencies of a code $c_i, (i = 1, 2, ..., M)$ in H_1 and H_2 are very different, it indicates that this code is distinct to two classes, otherwise this code is a common minute local features in all iris images. For iris classification, we need to extract distinguishing features from different class iris images. The difference of H_1 and H_2 is calculated: $\Delta H = \|H_1 - H_2\|$. The codes corresponding to the K_d largest bins of ΔH are kept to form the final $\bar{B}$, and these codes are called distinctive codes.

The third rule for code selection is to keep codes with high frequency of occurrence in iris images, which aims to represent iris images with little quantization error. We keep the codes with the K_c largest appearing probabilities in both H_1 and H_2 except the selected K_d codes according to the second rule. First, codes corresponding to the $K_c{}'(K_c{}' > K_c)$ largest $\min(H_1, H_2)$ are selected. Then, codes corresponding to the K_c largest $\sum(H_1, H_2)$ from the selected $K_c{}'$ codes are kept to form the final $\bar{B}$, and these codes are called mutual codes. In our experiments, $K_d = 512, K_c{}' = 1024, K_c = 512$, and the final $\bar{B}$ includes 1024 codes.

After codebook optimization, the LLC method [8] is adopted to represent iris images. The LLC is an effective coding scheme which utilizes the locality constraints to project each descriptor into its local-coordinate system with low computational complexity. It emphases to reconstruct features with locality constraint instead of emphasizing the sparsity constraint by using the following criteria:

$$\begin{aligned} &\min_{C} \sum_{i=1}^{N} \|x_i - Bc_i\|^2 + \lambda \|d_i \cdot c_i\|^2 \\ &s.t. 1^T c_i = 1, \forall i \end{aligned} \tag{1}$$

where B is codebook, $\cdot$ is the element-wise multiplication, d_i is the locality adaptor that gives different freedom for each basis vector which is proportional to its similarity to the input descriptor x_i, $C = [c_1, c_2, ..., c_N]$ is the set of codes for x, λ is constant.

Besides the efficient coding strategy, to include spatial information of iris images, the spatial pooling [16] is employed to obtain statistic summary of codes.

We use spatial pyramid matching (SPM) and max pooling [17] in this paper, and the $1 \times 1, 2 \times 2, 4 \times 4$ sub-regions are used for SPM.

4 Experimental Results

4.1 Experiment Setup

We evaluate our proposed method on two iris image databases. (1) The first iris image database is captured by hand-held optical sensor, denoted as DB1. The iris images from non-Asian include 1320 images from 66 different eyes, the same as the database used in [7]. The iris images from Asian includes 10000 images from 2000 eyes. (2) The second database is combined iris image database, denoted as DB2. The Asia iris image dataset includes the CASIA-Iris-Lamp database [18] and the same images of Asian in DB1. The non-Asian iris image dataset is composed by UPOL [19] iris image database and the same iris images of non-Asian in DB1. For each database, we use 500 iris images from each class for codebook learning, codebook optimization and SVM classifier training, and the other images are used for testing.

4.2 Results and Discussions

Experiments are performed on two datasets, and accuracies of the proposed algorithm, global feature method [6] and Iris-Texton method [7] are compared. Correct Classification Rate (CCR) and Equal Error Rate (EER) of the algorithms are examined. CCR is considered while using the median of decision values as threshold for SVM. The ethnic classification results are shown in Table 1.

Table 1. Comparison of classification accuracy between Iris-Texton [7] and the proposed method while using the whole iris image as ROI

	DB1		DB2	
Method	CCR(%)	EER(%)	CCR(%)	EER(%)
Iris-Texton [7]	90.05	11.54	88.57	13.16
Proposed	96.70	3.40	96.67	3.35

In upper experiments, the unwrapped iris image is used as ROI for feature extraction. Most of iris images include eyelid and eyelash occlusion, which means that some eyelids and eyelashes are included in the ROI and involved in the classification. To analysis the effect of the eyelid and eyelash, we do experiments by using the ROI used in [7], as shown in Fig. 4, which includes little eyelid and eyelash parts. The size of ROI is 60×256 pixels. The ethnic classification results based on small ROIs are shown in Table 2.

Fig. 4. Small ROI used for experiments

Table 2. Comparison of classification accuracy between Iris-Texton [7] and the proposed method while using small ROIs

	DB1		DB2	
Method	CCR(%)	EER(%)	CCR(%)	EER(%)
Iris-Texton [7]	85.11	16.58	82.58	19.48
Proposed	94.28	6.36	94.00	6.48

The ROC curve(DET curve) of experiments on DB1 and DB2 by using different methods and different ROIs are shown in Fig. 5.

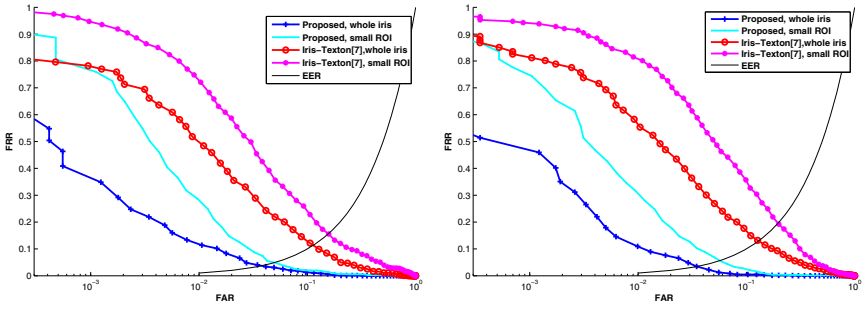


Fig. 5. ROC curve(DET curve) of experiments by using different methods and different ROI. (a) Results of DB1; (b) results of DB2

Compared to previous method in [7], the proposed method has largely improved the ethnic classification accuracy. There are some possible reasons that cause the improvement: the SIFT descriptor is suitable for representing local iris pattern; class label related codebook optimization can generate more effective codebook for extracting distinct features from images of different classes; the LLC and SPM strategy represent iris images with little quantization error. As shown in the experiment, using the whole unwrapped iris image for ethnic classification achieves better results than using a part of unwrapped iris which includes little eyelash and eyelid occlusion. It indicates that the eyelash and eyelid also can provide clues for ethnic classification. This conclusion meets the usage of periocular region in [12] for ethnic classification.

5 Conclusions

In this paper, we present a novel method for ethnic classification based on iris images. The proposed method adopts codebook optimization and robust coding strategy to represent iris images with the purpose of ethnic classification. By using the optimized codebook including both distinctive and mutual codes and LLC coding strategy, the proposed method achieves encouraging correct classification rate. Our future work includes following issues: ethnic classification related to more races; work about supervised codebook learning and optimization; the ethnic classification co-works with the iris recognition system as index information to improve the recognition accuracy.

Acknowledgments. This work is funded by National Natural Science Foundation of China (Grant No. 60736018, 61075024) and International S&T Cooperation Program of China (Grant NO.2010DFB14110).

References

1. Daugman, J.: High confidence visual recognition of persons by a test of statistical independence. IEEE TRANS. on PAMI 15(11), 1148–1161 (1993)
2. Daugman, J.: Iris recognition. American Scientist 89, 326–333 (2001)
3. Wildes, R.: Iris recognition: an emerging biometric technology. Proc. of the IEEE 85(9), 1348–1363 (2002)
4. Ma, L., Tan, T., Wang, Y., Zhang, D.: Personal identification based on iris texture analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1519–1533 (2003)
5. Sun, Z., Tan, T.: Ordinal measures for iris recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2211–2226 (2008)
6. Qiu, X., Sun, Z., Tan, T.: Global texture analysis of iris images for ethnic classification. Advances in Biometrics, 411–418 (2006)
7. Qiu, X., Sun, Z., Tan, T.: Learning appearance primitives of iris images for ethnic classification. In: IEEE Int'l Conference on Image Processing, vol. II, pp. 405–408 (2007)
8. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: Proc. of CVPR, pp. 3360–3367 (2010)
9. Gutta, S., Wechsler, H., Phillips, P.J.: Gender and ethnic classification. In: Int'l Conference on Automatic Face and Gesture Reconition, pp. 194–199 (1998)
10. Shakhnarovich, G., Viola, P.A., Moghaddam, B.: A unified learning framework for real time face detection and classification. In: Int'l Conference on Automatic Face and Gesture Reconition, p. 16 (2002)
11. Lu, X., Jain, A.K.: Ethnicity identification from face images. In: Proc. of SPIE Defense and Security Symposium., p. 16 (2004)
12. Lyle, J., Miller, P., Pundlik, S., Woodard, D.: Soft biometric classification using periocular region features. In: 2010 Fourth IEEE Int'l Conference on Biometrics: Theory Applications and Systems (BTAS), pp. 1–7. IEEE (2010)
13. Fan, R., Chang, K., Hsieh, C., Wang, X., Lin, C.: Liblinear: A library for large linear classification. Jour. of Machine Learning Research 9, 1871–1874 (2008)

14. He, Z., Tan, T., Sun, Z., Qiu, X.: Towards accurate and fast iris segmentation for iris biometrics. IEEE Trans. on Pattern Analysis and Machine Intelligence 31(9), 1617–1632 (2009)
15. Lowe, D.: Distinctive image features from scale-invariant keypoints. Int'l Jour. of Computer Vision 60(2), 91–110 (2004)
16. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proc. of CVPR, vol. 2, pp. 2169–2178 (2006)
17. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: Proc. of CVPR, pp. 1794–1801 (2009)
18. CASIA Iris Database., http://biometrics.idealtest.org
19. Dobes, M., Machala, L.: Upol iris database., http://www.inf.upol.cz/iris/

Iterative Directional Ray-Based Iris Segmentation for Challenging Periocular Images*

Xiaofei Hu, V. Paúl Pauca, and Robert Plemmons

Departments of Mathematics and Computer Science
127 Manchester Hall, Winston-Salem, NC27109, United States
{hux,paucavp,plemmons}@wfu.edu
http://www.math.wfu.edu

Abstract. The face region immediately surrounding one, or both, eyes is called the periocular region. This paper presents an iris segmentation algorithm for challenging periocular images based on a novel iterative ray detection segmentation scheme. Our goal is to convey some of the difficulties in extracting the iris structure in images of the eye characterized by variations in illumination, eye-lid and eye-lash occlusion, de-focus blur, motion blur, and low resolution. Experiments on the Face and Ocular Challenge Series (FOCS) database from the U.S. National Institute of Standards and Technology (NIST) emphasize the pros and cons of the proposed segmentation algorithm.

Keywords: Iris segmentation, ray detection, periocular images.

1 Introduction

Iris recognition is generally considered to be the most reliable biometric identification method. But, it most often requires a compliant subject and an ideal iris image. Non-ideal images are quite challenging, especially for segmenting and extracting the iris region. Iris segmentation for biometrics refers to the process of automatically detecting the pupillary (inner) and limbus (outer) boundaries of an iris in a given image. Here we apply iris segmentation within the periocular region of the face. This process helps in extracting features from the discriminative texture of the iris for personnel identification/verification, while excluding the surrounding periocular regions. Several existing iris segmentation approaches for challenging images, along with evaluations, can be found in [3,5,6].

This paper presents an iris segmentation technique for non-ideal periocular images that is based on a novel directional ray detection segmentation scheme.

* This work was sponsored under IARPA BAA 09-02 through the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF10-2-0013. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of IARPA, the Army Research Laboratory, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 91–99, 2011.

The method employs calculus of variations and directional gradients to better approximate the boundaries of the pupillary and limbus boundaries of the iris. Variational segmentation methods are known to be robust in the presence of noise [1,11] and can be combined with shape-fitting schemes when some information about the object shape is known a priori. Quite commonly, circle-fitting is used to approximate the boundaries of the iris, but this assumption may not necessarily hold for non-circular boundaries or off-axis iris data. For computational purposes, our technique uses directional gradients and circle-fitting schemes, but other shapes can also be easily considered. This technique extends the work by Ryan et al. [9], who approached the iris segmentation problem by adapting the Starburst algorithm to locate pupillary and limbus feature pixels used to fit a pair of ellipses. The Starburst algorithm was introduced by Li, et al. [7], for the purpose of eye tracking.

In this paper, experiments are performed on a challenging periocular database to show the robustness of the proposed segmentation method over other classic segmentation methods, such as Masek's Method [8] and Daugman's Integro-Differential Operator method [2]. The paper is organized as follows. In Section 2, we briefly introduce a periocular dataset representative of the challenging data considered in our tests. In Section 3, we touch upon the pre-processing procedures used to improve the accuracy of iris segmentation in the presence of challenging data. In Section 4, iterative directional ray detection with curve fitting is proposed. Section 5 describes the experimental results obtained with directional ray detection and conclusions are presented in Section 6.

2 Face and Ocular Challenge Series (FOCS) Dataset

To obtain accurate iris recognition from periocular images, the iris region has to be segmented successfully. Our test database is from the National Institute of Standards and Technology (NIST) and is called the Face and Ocular Challenge Series database[1]. Performing iris segmentation on this database can be very challenging (see Fig. 1), due to the fact that the FOCS database was collected from subjects walking through a portal in an unconstrained environment [6]. Some of the challenges observed in the images include: poor illumination, out-of-focus blur, specular reflections, partially or completely occluded iris, off-angled iris, small size of the iris region compared to the size of the image, smudged iris boundaries, and sensor noise.

3 Image Pre-processing

For suppressing the illumination variation in FOCS images and thus improving the quality of iris segmentation, the following pre-processing scheme was used: (i) illumination normalization, (ii) eye center detection, and (iii) inpainting.

[1] See `http://www.nist.gov/itl/iad/ig/focs.cfm`

(a) Poor illumination (b) Blur (c) Occlusion (d) Off-angle

Fig. 1. Periocular FOCS imagery exhibiting non-ideal attributes

3.1 Illumination Normalization

Illumination variation makes it challenging to accurately and reliably determine the location of the iris boundaries. In general, the image contrast is very low and the iris boundaries are somewhat obscured. To increase the contrast and highlight the intensity variation across the iris boundaries, illumination normalization was performed by using the *imadjust* command in MATLAB. This normalization helps make intensity variation of different images more uniform, improving the stability of the proposed segmentation algorithm. Sample images obtained of before and after illumination normalization are shown in Fig. 2.

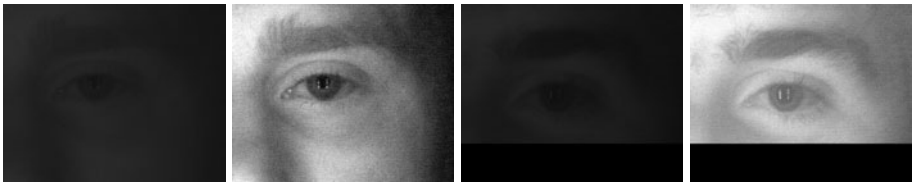
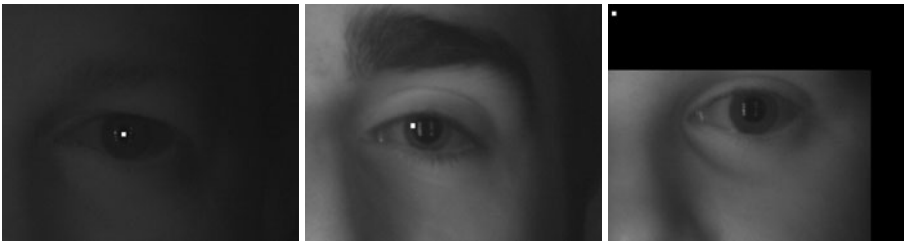


Fig. 2. Sample FOCS imagery before and after illumination normalization

3.2 Eye Center Detection

Reliably determining the location of the iris in the image is challenging due to noise and non-uniform illumination. We use an eye center detector method based on correlation filters. In this method, a specially designed correlation filter is applied to an ocular image, resulting in a peak at the location of the eye center. The correlation filter for the eye center detector was trained on 1,000 images, in which the eye centers were manually labeled. The correlation filter approach for detecting the eye centers, when applied on the full FOCS dataset, yielded a success rate of over 95%. Fig. 3(a) shows a typical image after eye center detection, where the eye center is shown with a small white dot. Fig. 3(b-c) show some of the rare cases where the eye center was not accurately determined. The accuracy of our iris segmentation method is crucially related to the correctness of eye center detection.



(a) Correct (b) Slightly offset (c) Completely offset

Fig. 3. Eye center detection for sample FOCS imagery (marked by white dots)

3.3 Inpainting

Specular reflections located within the iris region are additional factors affecting the success of iris segmentation. We use inpainting for alleviating this effect. In particular, an inpainting method [10] is applied to each image to compensate for the information loss due to over or under exposure. However, because of the non-uniform illumination distribution on some images, this step cannot completely remove specular reflections.

4 Iterative Directional Ray-Based Iris Segmentation

The proposed method involves multiple stages of a directional ray detection scheme, initialized at points radiating out from multiple positions around the eye region, to detect the pupil, iris, and eyelid boundaries. Specifically, it consists of three main sequential steps: (i) Identification of a circular pupil boundary by directional ray detection, key point classification and Hough transform [4]; (ii) Identification of the limbic boundary by iterative directional ray detection; and (iii) Identification of the eyelid boundary by the directional ray detection applied at multiple points.

4.1 Directional Ray Detection

Directional ray detection aims to identify the local edge features of an image around a reference point (x_s, y_s) or a set of reference points $\mathcal{S}$ along selective directions. Assume that the reference point (x_s, y_s) is an interior point of a pre-processed image $\mathcal{I}(x, y)$. Let r be the searching radius and a finite set $\Theta = \{\theta_i\}_{i=1}^{M} \subset [0, 2\pi]$ be the selective searching direction set. Directional ray detection results in a set of key points (x_i, y_i) along Θ in a circular neighborhood of radius r around (x_s, y_s) maximizing,

$$(x_i, y_i) = argmax_{(x,y)\in\Gamma(x_s,y_s,\theta_i,L)} \left|\omega_H * \mathcal{I}(x, y)\right|, \qquad i = 1, \ldots, M, \tag{1}$$

where $\Gamma(x_s, y_s, \theta_i, L)$ denotes the θ_i direction rays of length of L radiating from (x_s, y_s), and ω_H is a high-frequency filter for computing gradients along rays.

In the implementation, we use $\omega_H = [-1, 0, 1]$. Locations for which the gradient difference reaches an absolute maximum are then selected as key points.

The main advantages of directional ray detection are flexibility and computational efficiency. With our method, we can easily control the reference points, searching radius, and searching directions, while exploring local information instead of the entire image.

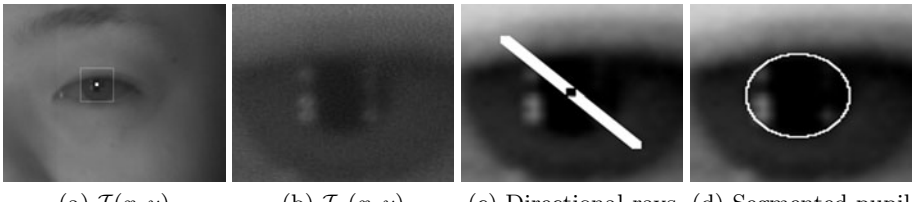
4.2 Pupil Segmentation

Assume that the eye center (x_c, y_c) of a pre-processed image $\mathcal{I}(x, y)$ is correctly located within the pupil, shown as a white dot in Fig. 4(a). Let $\mathcal{I}_p(x, y)$ denote a cropped region of size $2(r_p^{(l)} + r_p^{(u)})$ centered at (x_c, y_c), where $0 < r_p^{(l)} < r_p^{(u)}$ are pupil radius bounds determined experimentally from the dataset. A sample region $\mathcal{I}_p(x, y)$ containing a pupil is shown in Fig. 4(b). For computational efficiency, pupil segmentation is performed on $\mathcal{I}_p(x, y)$. Pupil segmentation is an iterative algorithm. Each iteration includes ray detection to extract structural key points, classification to eliminate noisy key points, and the Hough transform to estimate the circular pupil boundary.

In each iteration, key points of the pupil are efficiently extracted using Eq. (1), setting (x_c, y_c) as the reference point. For the FOCS dataset, we set

$$\Theta = \{\frac{\pi}{4} + \frac{i\pi}{M}\}_{i=1}^{\frac{M}{2}} \cup \{\frac{5\pi}{4} + \frac{i\pi}{M}\}_{i=1}^{\frac{M}{2}},$$

assuming M is an even number, to avoid the specular reflection in the horizontal direction as shown in Fig. 4(c). To remove outliers, we apply k-means intensity classification for grouping all the key points into two groups, pupillary edge points and non-pupillary edge points. The process depends on the fact that the intensity of pixels inside the pupil is lower than in other portions of the eye, such as the iris and eyelids. The Hough transform is then applied to selected pupillary edge points, the estimated pupil center (x_p, y_p), and the radius of the pupil r_p as shown in Fig. 4(d). We use (x_p, y_p) and r_p as inputs of the next iteration.



(a) $\mathcal{I}(x, y)$ (b) $\mathcal{I}_p(x, y)$ (c) Directional rays (d) Segmented pupil

Fig. 4. One iteration of Iterative Directional Ray Segmentation for the pupil

4.3 Iris Segmentation

Similarly to pupil segmentation, we determine key points along the irs based on our adaptive directional ray detection and apply the Hough transform to find the iris boundary. Let $\mathcal{I}_I(x, y)$ be a cropped image region of size $2(r_p^{(l)} + r_I^{(u)})$ centered at the estimated pupil center (x_p, y_p), where $r_I^{(u)}$ is a bound parameter of the iris radius determined experimentally from the data. In this step, directional ray detection in Eq. (1) is performed on $\mathcal{I}_I(x, y)$ with a set of reference points $\mathcal{S}_{\theta_i}$ along directions $\theta_i \in \Theta$ for $i = 1, \ldots, M$. Here, given $0 < \alpha < 1$ and $N_s > 0$, we have

$$\mathcal{S}_{\theta_i} = \{(x, y) | x = x_p + (1 + \alpha) r_p \cos t; y = y_p + (1 + \alpha) r_p \sin t,\ t \in \mathcal{T}_{\Theta_i}\}$$

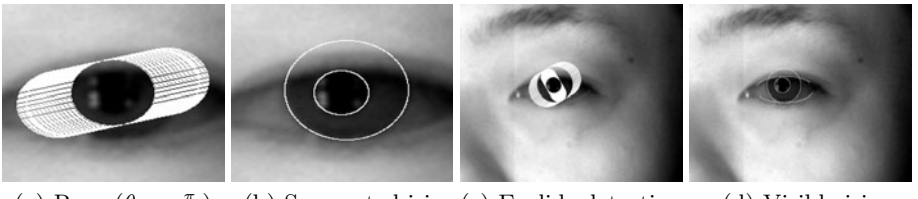
with

$$\mathcal{T}_{\Theta_i} = \{\theta_i - \frac{\pi}{2} + \frac{j\pi}{Ns}\}_{j=0}^{N_s}$$

and

$$\Theta = \{-\frac{\pi}{6} + \frac{i2\pi}{3M}\}_{i=1}^{\frac{M}{2}} \cup \{\frac{5\pi}{6} + + \frac{i2\pi}{3M}\}_{i=1}^{\frac{M}{2}}.$$

As shown in Fig. 5(a), the direction θ_i of test rays is close to the horizontal axis to avoid the upper and lower eyelid regions. After applying the Hough transform on the iris key points, we obtain the limbus boundary, as shown in Fig. 5(b).



(a) Rays ($\theta_i = \frac{\pi}{12}$) (b) Segmented iris (c) Eyelids detection (d) Visible iris

Fig. 5. Iterative directional ray detection results for iris and eyelids

4.4 Eyelid Boundary Detection

Detecting the eyelid boundaries is an important step in the accurate segmentation of the visible iris region as well as for the determination of iris quality metrics. The eyelid boundaries extraction is accomplish by a two-step directional ray detection approach from multiple starting points. As shown in Fig. 5(c), we set the rays to radiate from the pixels outside the boundaries of the pupil and iris regions along certain directions. Two groups of rays are chosen roughly along the vertical and horizontal directions. Eq. (1) is applied to these two groups of rays resulting in edge points along the eyelid. A least squares curve fitting model is then applied, resulting in a first fitting estimation to the eyelid boundary. To improve accuracy of the estimation in the previous step, we apply directional ray detection in the interior regions of the estimated eyelids. A new set of key points is obtained. The least squares fitting model is then applied again, resulting in an improved estimation of eyelid boundary, as illustrated in Fig. 5(d).

5 Experimental Evaluation

To evaluate our approach, experiments were performed on a sample subset of 404 images chosen from the full FOCS dataset, where comparisons are made with Masek's [8] and Daugman's [2] methods. The sample dataset was assured to be representative of the full FOCS database in terms of the challenges possessed. We also applied the proposed method to 3481 FOCS challenging periocular images. Fig. 6(a-c) provides examples of where the proposed method provided good iris segmentation, while Fig. 6 (d-f) shows images where the segmentation failed.

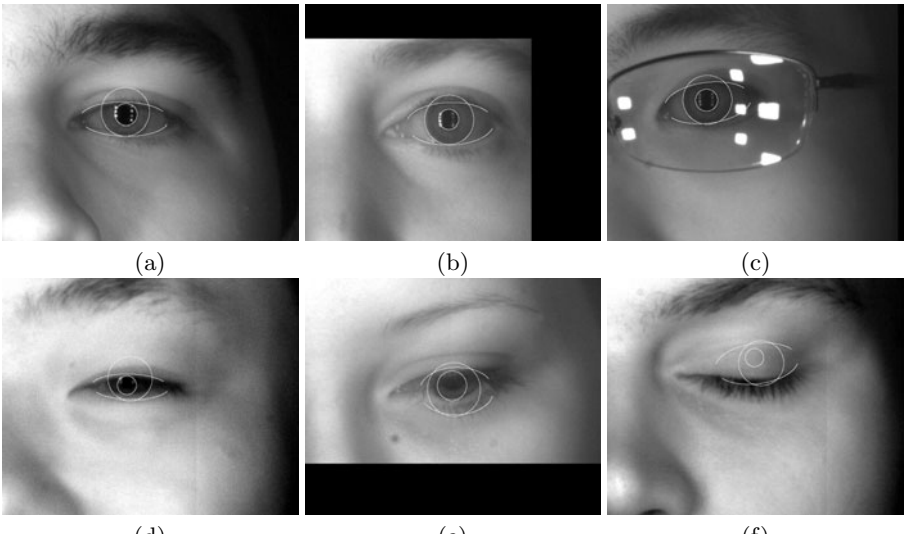


Fig. 6. Successful (top) and unsuccessful (bottom) segmentation examples obtained by Directional Ray Detection

The overall accuracy of the iris segmentation technique on the FOCS dataset was measured as follows:

$$\text{Segmentation Accuracy} = \frac{\text{Number of successfully segmented images}}{\text{Total number of images}} \cdot 100\%,$$

where successfully segmented images means that the segmented iris is exactly the visible iris or is extremely close to it with about $\pm 2\%$ tolerance error in terms of area. A comparison of the proposed method with Masek's segmentation [8] and Daugman's Integro-Differential Operator method (IDO) [2] is shown in Table 1. Although Masek's segmentation and the Integro-Differential Operator were observed to use less computation time, the quality of segmented output is not acceptable (see Table 1). The new algorithm provides better performance at the expense of higher computational cost. The average computational time

Table 1. Segmentation accuracy of the Masek, IDO, and our proposed method

Methods	Total number of images	Number of successfully segmented images	Segmentation accuracy
Masek [8]	404	210	51.98%
IDOperator [2]	404	207	51.23%
Proposed method	404	343	**83.9%**

per image of the proposed method is about 15.5 seconds on a PC with a core i5vpro CPU.

Based on the comparisons in Table 1, promising performance of the proposed method in terms of accuracy is observed. Furthermore, we evaluate the ray-based method on a larger dataset, containing 3841 FOCS periocular images. As illustrated in Table 2, the segmentation accuracy is 84.4%, which highlights the overall performance and robustness of the proposed method in processing challenging periocular images.

Table 2. Segmentation accuracy of the proposed method on 3481 challenging FOCS periocular images

	Total number of images	Number of successfully segmented images	Segmentation accuracy
Proposed method	3481	2884	**84.4%**

6 Final Remarks

Iris recognition of low-quality images is a challenging task. A great deal of research is currently being conducted towards improving the recognition performance of irides under unconstrained conditions. Iris segmentation of challenging images is one of the crucial and important components of this work, whose accuracy can dramatically influence biometric recognition results. The main contributions of our paper are: (1) we provide a new segmentation method which is shown to be robust in processing low-quality periocular imagery, and (2) test our method a challenging dataset with poorly controlled iris image acquisition.

References

1. Chan, T.F., Vese, L.A.: Active Contours Without Edges. IEEE Transactionson Image Processing 10(2), 266–277 (2001)
2. Daugman, J.: How iris recognition works. In: Proceedings of the International Conference on Image Processing, vol. 1 (2002)
3. He, Z., Tan, T., Sun, Z., Qiu, X.: Towards accurate and fast iris segmentation for iris biometrics. IEEE Trans. On Pattern Analysis and Machine Intelligence 31(9), 1617–1632 (2009)
4. Illingworth, J., Kittler, J.: A survey of the Hough transform. Computer Vision, Graph. Image Processing 44, 87–116 (1988)

5. Proenca, H.: Iris recognition: on the segmentation of degraded images acquired in the visible wavelength. IEEE Trans. on Pattern Analysis and Machine Intelligence 32(8), 1502–1516 (2010)
6. Jillela, R., Ross, A., Boddeti, N., Vijaya Kumar, B., Hu, X., Plemmons, R., Pauca, P.: An Evaluation of Iris Segmentation Algorithms in Challenging Periocular Images. In: Burge, M., Bowyer, K. (eds.) Handbook of Iris Recognition, Springer, Heidelberg (to appear, 2012)
7. Li, D., Babcock, J., Parkhurst, D.J.: Openeyes: A Low-Cost Head-Mounted Eye-Tracking Solution. In: Proceedings of the 2006 Symposium on Eye Tracking Research and Applications. ACM Press, San Diego (2006)
8. Masek, L.: Recognition of human iris patterns for biometric identification. Thesis (2003)
9. Ryan, W., Woodard, D., Duchowski, A., Birchfield, S.: Adapting Starburst for Elliptical Iris Segmentation. In: Proceeding IEEE Second International Conference on Biometrics: Theory, Applications and Systems (BTAS). IEEE Press, Washington (2008)
10. Roth, S., Black, M.J.: Fields of Experts. International Journal of Computer Vision 82(2), 205–229 (2009)
11. Vese, L.A., Chan, T.F.: A Multiphase Level Set Framework for Image Segmentation Using the Mumford and Shah Model. International Journal of Computer Vision 50(3), 271–293 (2002)
12. Vijaya Kumar, B.V.K., Hassebrook, L.: Performance Measures for Correlation Filters. Applied Optics 29, 2997–3006 (1990)
13. Vijaya Kumar, B.V.K., Savvides, M., Venkataramani, K., Xie, C., Thornton, J., Mahalanobis, A.: Biometric Verification Using Advanced Correlation Filters. Applied Optics 43, 391–402 (2004)
14. Wildes, R., Asmuth, J., Green, G., Hsu, S., Kolczynski, R., Matey, J., McBride, S.: A System for Automated Iris Recognition. In: Proceedings of the Second IEEE Workshop on Applications of Computer Vision, pp. 121–128 (December 1994)

Iris Plaque Detection Method Based on Level Set

Xiao Nan Liu and Wei Qi Yuan

Computer Vision Group, Shenyang University of Technology
No. 111, Shenliao West Road, Economic & Technological Development Zone,
Shenyang, 110870, P.R. China

Abstract. According to Iridology, the shape and position of iris plaques can reflect the health condition of human organs. In this paper, we propose a method to detect iris plaques automatically with computer vision technology. Firstly, we find the alternative regions of the plaques by searching the iris image with a window. In these regions, the mean of the central pixels' gray level is less than the minimum of the marginal pixels' gray level. Secondly, with Level set method, we obtain the edge information of these alternative regions. Finally, we determine whether the contour of the zero level sets meets the conditions of closed criterion and dimension criterion or not. If met, treat the pixels in the contour as the edge of the plaques. By comparing with the other algorithms, we found that our method is more efficient and suitable to complex background as iris images and it's less blind.

Keywords: Iris diagnosis, Iris plaque detection, Level set.

1 Introduction

Iridology, also known as iris diagnosis, is a subject which is the study of determining the body organ's health through the inspection of iris texture [1]. There are six pathological signs in iridology: crypt, lacunae, plaques, lines, color and density, as in Fig.1. Now the most popular instrument of iris diagnosis is iridology scope. Doctors capture the image of iris by the scope and observe the texture by their eyes. Finally they gave the health condition of the human organs according to their experience. Obviously, the examination and diagnosis are very subjective. In order to deal with these problems, we study an automatic detection system which uses the computer visual technology, which is called computer-aided iris diagnosis system. This system can make iris diagnosis became more objective and accurate, more convenient and easier, especially for ordinary communities. In this paper, we study the detection method of plaques on the irises based on computer vision.

Currently, there is no much literature about the detection of iris plaques. Du[2] proposed a method based on K-S distance to extract the plaque regions out from the

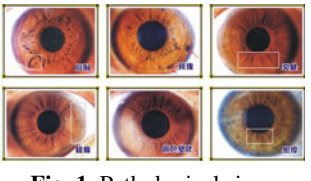


Fig. 1. Pathological signs

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 100–107, 2011.

image. However, the methods of object extracting are always a research focus of image processing. These methods can be divided into two categories. One is edge extraction. With this method, we could firstly get the edge curves of the whole image, and then find the object's edge information out from the curves. The other is image segmentation. These methods classify the targets and their background by a threshold. However, because of the special application of the detection of iris plaques, the two kinds of methods above can hardly get the accurate results. The method discussed in this paper can extract the plaques' edge information with Level set method. It can overcome the lack of the methods above, and obtain more accurate results.

2 Iris Plaques Detection Method

2.1 Characters of Plaques

By observation of a large number of iris images with plaques, we find the characteristics of the plaques as follows:

Position Unfixed. They can appear in any area outside the collarette of the iris.

Number Unfixed. There can be one or more plaques in different location in the iris.

Size Unfixed. The size of the plaques doesn't have uniform standard. But the biggest width can't exceed the width of iris ring.

Shape Unfixed. The edge of the plaques is irregular and has no uniform appearance.

Gray Level Unfixed. For the plaques in the different irises or the different plaques in the same iris, the range of their gray level is different.

Others. Moreover, we find that background region of the plaques are also inhomogeneous. The distribution of their gray level is very complicated too. So we can't use a unique gray level to recognize them. All the characters above make the detection of the plaques become more difficult. It is impossible to obtain the edge of the plaques by edge extraction of the whole iris image. Also it is hard to find a suitable threshold to obtain a correct segmentation of the plaques and their background. However, although the plaques in the iris differ from each other in thousands of ways, they also have some common characters in favor of detection. For example, the gray level of the plaque region must be smaller than that of its background region. And the edge of the plaque must be closed.

2.2 Algorithm

According to all the characters above, we propose a method to detect the plaques on the iris. Our method searches the entire image with a small window first, and gains the alternative region of the plaques. And then we use Level set method to find the zero level set contours of in these regions. Finally, we determine whether the textures are plaques by judging the shape and the dimension of the contours. We'd like to elaborate our method in this section. The procedure of our method is shown below.

Input and Preprocess. We must input the image data array first. If the image is collected by CCD camera in an RGB format, we must transform it into gray image. In this paper, we take the images collected in the ophthalmology hospital for instance, such as the images shown in Fig.2. Its original size is 2048*1536. To reduce the

computation, under the premise of accurate detection, we decimate the image and make the image size to be 512*384.

Search and Locate. According to the character that the gray level of the plaques is less than their neighbor regions', we find the alternative regions in the way of window searching method. By searching the entire image, we look for the region where the mean of central area's grey level is less than the minimum of marginal pixels' grey level. Through the observation of a large number of images and experimental analyses, we find that the size of the searching window and the central region's dimension will affect the accuracy of searching. When the central area is too small, there are so many regions being checked out. Many of them are not the regions with plaques. This will increase the burden of the later procedures. Oppositely, when the central area is too big, some region with small plaques will be missed. False identifying can be tolerated, because they can be excluded by the later procedures. But missing is fatal, it can cause the plaques being undetected. We do several experiments with the image in our database by the searching window, whose outer dimension is 70*70, and the central areas are 16*16, 14*14, 12*12, 10*10 and 8*8 respectively. For the images in Fig.2, we showed the results of the experiments in Table.1. We find that the results of the experiments, when using the window whose size is 70\10, are the best ones. Therefore, we choose the window size as 70\10 in our method for the images in our database.

Table 1. Comparision of the different window dimension

image	outer dimension\central dimension				
	70\16	70\14	70\12	70\10	70\8
1	correct	correct	false	false	false
2	correct	correct	false	false	false
3	false	false	false	false	false
4	false	false	false	false	false
5	false	false	false	false	false
6	missing	missing	missing	correct	false
7	correct	correct	false	false	false
8	false	false	false	false	false

Initialize. Initialize set $I = \{I_1, I_2, \cdots I_n\}$, where I is the alternative regions' data array, and n the number of the regions.

Read data. Read out the data array I_i, which is the alternative region.

Level Set Evolve. Evolve I_i by Level set method. First, we initialize the zero level set ϕ_0, here we use a circle to be the initial set, whose diameter is the moiety of the searching window's size and center is the image central pixel. And then we evolve the image data as the following formulas.

$$\phi_{n+1} = \phi_n + 0.1 * \delta_h * [0.01 * 255 * 255 * K - (I_i - C1)^2 + (I_i - C2)^2] \quad (1)$$

Where δ_h the Dirac function, K is the curvature, H is the Heaveside function, C1 is the constant of the target region, and C2 is the constant of the background region.

$$\delta_h = \left(\frac{\varepsilon}{\pi}\right) \Big/ \left(\varepsilon^2 + \phi_n{}^2\right) \tag{2}$$

$$K = \frac{\partial N_x}{\partial x} + \frac{\partial N_y}{\partial y} \tag{3}$$

$$H = \frac{1}{2}\left[1 + \frac{2}{\pi}\arctan\left(\frac{\phi_n}{\varepsilon}\right)\right] \tag{4}$$

$$C1 = \frac{sum(H.*I_i)}{sum(H)} \tag{5}$$

$$C2 = \frac{sum((1-H).*I_i)}{sum(1-H)} \tag{6}$$

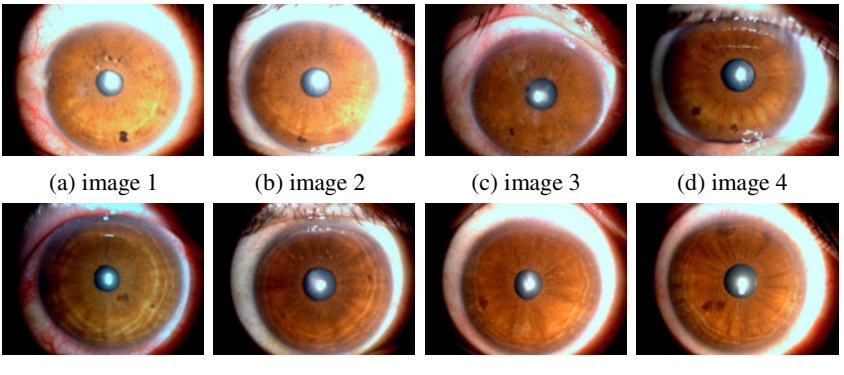
Shape Criterion. Judge the closed quality of the coordinates of the zero level sets. If the contour of the zero level set is closed, go to the next step. If not, terminate and abandon.

Dimension Criterion. Judge the dimension of the zero level sets. If the dimensions' are big enough, then output the coordinates of the zero level sets. If not, abandon them.

Output. Output the final results of the detection of the plaques in the iris.

3 Experiment and Analysis

We randomly selected eight images from the image database, each of which has plaques in it, and then analyzed the results of our experiments and discussed the advantages of our method. The eight original images are shown in Fig.2.

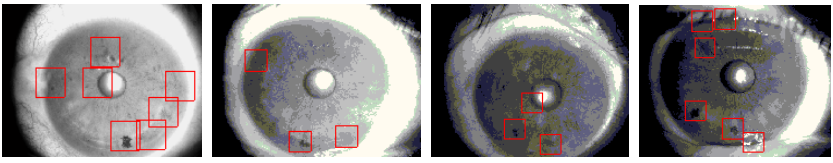


(a) image 1 (b) image 2 (c) image 3 (d) image 4

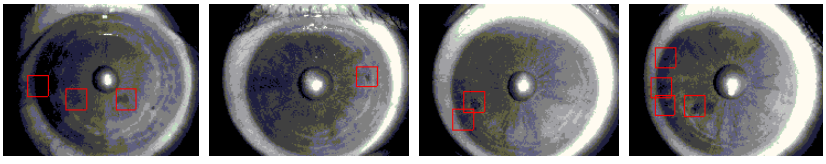
(e) image 5 (f) image 6 (g) image 7 (h) image 8

Fig. 2. Original images

First, we search the entire image by searching window to find out the regions where the plaques possibly exist. The results of this process are shown in Fig.3. We mark the alternative regions by rectangles.



(a)image 1 (b)image 2 (c)image 3 (d) image 4

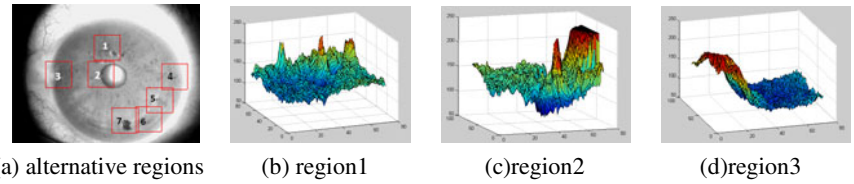


(e) image 5 (f) image 6 (g) image 7 (h) image 8

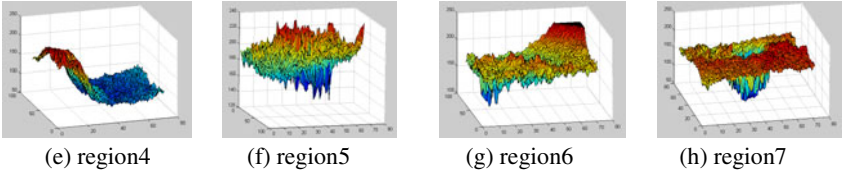
Fig. 3. The result of searching

With Fig.3 we can discover that the regions found by the searching progress are not actually the ones who have plaques. Some of them are false recognition, but there is no missing. Because of the effect of the sclera, pupil, light dots, eyelashes or the inhomogeneous texture in the iris, some of the regions searched out are not those who have plaques, although they meet the searching condition.

To illustrate this situation we take image 1 for example. In Fig.4 we show the searching result of image 1. We can find that there are seven alternative regions which meet the searching condition. Here we mark them from region 1 to region 7. Each region's 3D chart of gray level distribution are shown in Fig.4(b) to figure Fig.4 (h). We can discover that region 5 and region 7 maybe the plaque regions. Their 3D gray level distributions are like caldron basin and depressed in the center. Although region 5 is center depressed, but its round part is inhomogeneous. This will make it difficult to determine a threshold of the image segmentation. In region 1, the central region's gray level is less than the marginal pixel's gray level because of the inhomogeneous texture there. This region can match the searching condition, but it's a false recognition. Its gray level distribution doesn't match the character of plaque. Region 2, region 3, region 4 and region 6 are false recognitions too. Those are all caused by the sclera and the light spots in the pupil. Their 3D distributions are sloping and can't be matched. In conclusion, the alternative regions being found out do not contain plaques entirely, and their textures are various. We need to do further analysis to get the final results.



(a) alternative regions (b) region1 (c)region2 (d)region3



(e) region4 (f) region5 (g) region6 (h) region7

Fig. 4. Searching result of image 1

In order to get the accurate texture information of the alternative regions, we evolve the region's data array with Level set method. The results are shown in Fig.5.

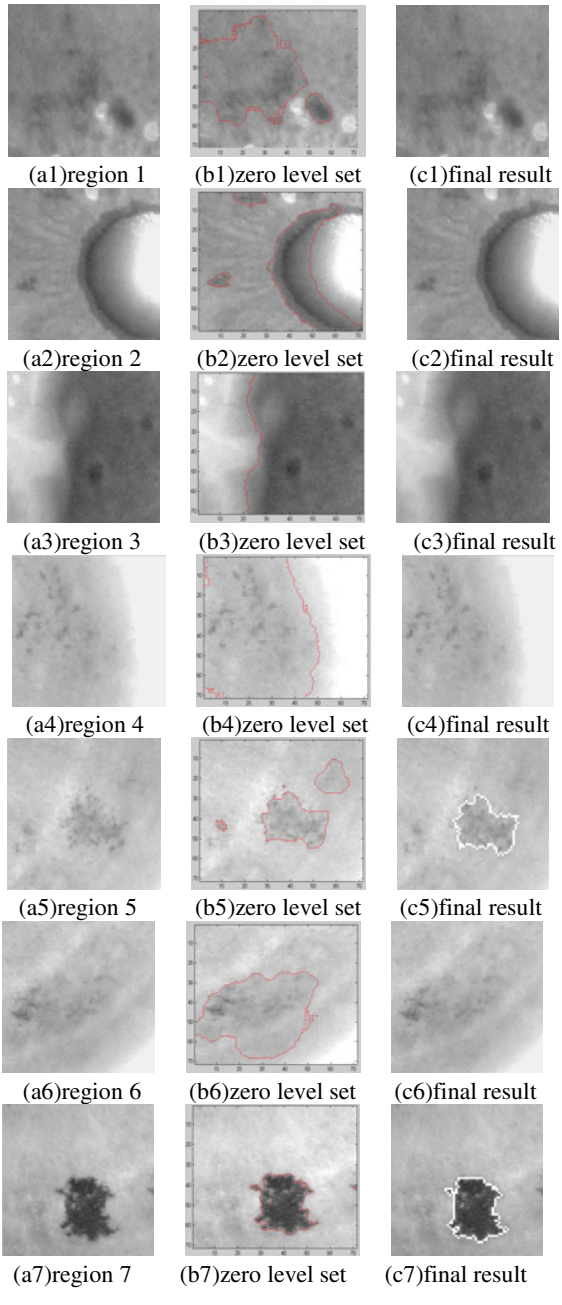
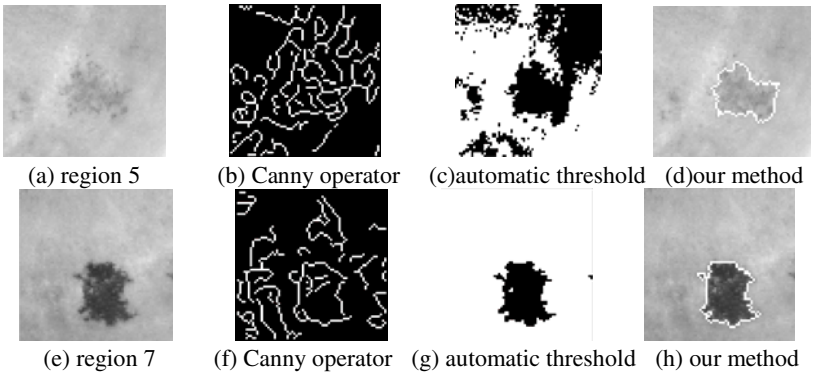


Fig. 5. The results of our method

In Fig.5, (a1) to (a7) are the original gray images of the seven regions shown in Fig.4. (b1) to (b7) are the results of the images which has been evolved for 200 times. We show the convergent curves in black lines. (c1) to (c7) are the final results of our method. All the contours marked in white lines are the final extracted texture information of plaques. They are closed and their dimensions are large enough.

Level set method evolves according to gray level distribution of the image, and the result finally converges to the edges where the gray levels of the pixels around are different. Level set method functions totally based on the gray gradient, and it doesn't need to determine the threshold of the judgment. However, the edges obtained by the Level set method are not specific to plaques. So we need to identify the shape of the edges and the distributions of the regions' gray levels in compliance with the requirements of the iris plaques character. According to Fig.5, we found that all the images have convergent curves being evolved and each of them has more than one zero level set. They are caused by the mutation and the inhomogeneous texture in this region. For region 2, region3, region 4 and region 6, the gray gradients are caused by the sclera and pupil, and the coordinates of the zero level set are not closed. So these regions' textures are not suitable to plaque's characters and there is no curves detected. For region 1, some of the zero level sets are caused by the inhomogeneous texture rather than the plaques of the iris, and their dimensions are too small to match the criterion. Moreover, the texture is adhered with the surrounding area, so the coordinates of the zero level set are not closed too. Also there is no curve being detected. For region5 and region 7, whose zero level sets are not only closed, but also suitable to the dimensional criterion, our method detects the accurate edge information.

Focus on region 5 and region 7, we compare our method with edge extracting with Canny operator and the automatic threshold image segmentation.



(a) region 5 (b) Canny operator (c)automatic threshold (d)our method

(e) region 7 (f) Canny operator (g) automatic threshold (h) our method

Fig. 6. Comparison of the algorithms

As shown in Fig.6, because of the characters of the iris plaques and their back ground, the edges extracted by Canny operator are blind and not continuous. The result is not the accurate information of the plaque. The automatic threshold image segmentation method can't classify the plaque correctly either. Such as region 5, the contrast of this region is too small to get a correct segmentation. However, we use Level set method to gain the edge information. We use the information of the gray gradient instead of an absolute threshold of gray level, and what we use is the relative feature of the gray level

between different pixels. In addition, by the judgments of the closed feature and dimension of the zero level set, we can avoid the erroneous judgment caused by other texture and noise and get the accurate edge information of the iris plaques.

4 Conclusion

In this paper, we present a new auto-detection method which is suitable to iris plaques for iris diagnosis. With this method, we can obtain the edge information of the iris plaques effectively. By searching the regions where the central pixels' gray levels are smaller than the marginal ones', we find the alternative regions of the plaques. Thereby, we can reduce the computation of the evolving procedure, and make the algorithm be more efficient, and less storage space is required. With the application of Level set method, our method used the relative feature of the gray level instead of the absolute gray level of the pixels. This can make the algorithm suitable to the complicated gray level distribution of the iris. This can also avoid the problem that fixed threshold cannot attain accurate segmentation of the image. After evolving, we can obtain the accurate edge information of the region's texture. In addition, our method got the plaques contour by closed criterion and dimensional criterion. The contour of the zero level set, which is satisfactory to the criteria, will be considered as the plaque's edge information. Thus, we can detect the plaques of the iris and obtain the accurate information of the plaques' edge. Our method is more efficient and less blind. It uses less computation and storage.

Aknowledgement. This work is supported by Science and Technology Plan of Shenyang, grant #1091180-1-00 and grant #F10-213-1-00.

References

1. Fragnany, P.: Introduction of Iridology. Yunnan Publishing Company, Yunnan (1982)
2. Du, W.: Research on Iris Pathological Feature Extraction and Diagnosis Model. Harbin Institute of Technology (2009)
3. Xin, G., Wang, W.: Study on Feature Extraction in Computer-aided Iridology. Computer Engineering and Design 27, 3322–3323, 3376 (2006)
4. Shen, B., Xu, Y., Lu, G., Zhang, D.: Detecting Iris Lacunae Based on Gaussian Filter. In: 3rd International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Taiwan, pp. 233–236 (2007)
5. Lodin, A., Demea, S.: Design of an Iris-Based Medical Diagnosis System. In: International Symposium on Signals, Circuits and Systems, Romania, pp. 129–132 (2009)
6. Osher, S., Sethian, J.A.: Fronts Propagating with Curvature Dependent Speed: Algorithms based on Hamilton-Jacobi Formulation. Journal of Computational Physics 79, 12–49 (1988)
7. Tsai, R., Osher, S.: Level Set Methods and Their Applications in Image Science. In: IEEE International Conference on Image Processing, Barcelona, pp. 631–634 (2003)
8. Cremers, D., Rousson, M., Deriche, R.: A Review of Statistical to Level Set Segmentation: Integrating Color, Texture, Motion and Shape. International Journal of Computer Vision 72(2), 195–215 (2007)
9. Taheri, S., Ong, S.H., Chong, V.F.H.: Level-set Segmentation of Brain Tumors Using a Threshold-based Speed Function. Image and Vision Computing 28, 26–37 (2010)

Invariant Hand Biometrics Feature Extraction

Alberto de Santos Sierra, Carmen Sánchez Ávila, Javier Guerra Casanova, and Gonzalo Bailador del Pozo

Group of Biometrics, Biosignals and Security (GB2S)
Centro de Domótica Integral (CeDInt-UPM)
Universidad Politécnica de Madrid
Campus de Montegancedo, 28223 Pozuelo de Alarcón, Madrid, Spain
{alberto,csa,jguerra,gbailador}@cedint.upm.es
http://www.gb2s.es

Abstract. Hand biometrics relies strongly on a proper hand segmentation and a feature extraction method to obtain accurate results in individual identification. Former operations must be carried out involving as less user collaboration as possible, in order to avoid intrusive or invasive actions on individuals.

This document presents an approach for hand segmentation and feature extraction on scenarios where users can place the hand on a flat surface freely, without no constraint on hand openness, rotation and pressure.

The performance of the algorithm highlights the fact that in less than 4 seconds, the method can detect properly finger tips and valleys with a global accuracy of 97% on a database of 300 users, achieving the second position in the International Hand Geometric Competition HGC 2011.

Keywords: Hand segmentation, invariant feature extraction, mathematical morphology, biometrics, hand geometry.

1 Introduction

In recent years, hand-based biometric systems are evolving from constrained, contact-based approaches [4,16] to completely non-collaborative, unconstrained and contact-less scenarios, where almost no collaboration from user is required.

The main aim of this trend is focus on providing hand biometrics with more comfortability and usability characteristics, increasing the acceptabiblity of final user.

However, as system development attempts to adapt hand biometrics to daily non-collaborative, non-intrusive and non-invasive scenarios, the operations of segmentation and feature extraction increases their current strain.

Therefore, this document presents a segmentation algorithm and feature extraction method to provide accurate results for hand biometrics in a scenario with a flat surface where hand is placed, but with a considerable margin of freedom, providing samples with a wide range of rotation and hand openness.

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 108–115, 2011.

The evaluation of the proposed method considers a publicly available database, leading to accurate results, which made this algorithm to achieve the second position in International Hand Geometric Points Detection Competition HGC2011 [12].

The layout of the paper is as follows: First of all, previous approaches are studied under the literature review (Section 2). Afterwards, both the segmentation method (Section 3) and the feature extraction strategy (Section 4) are described, together with the corresponding results (Section 5). Finally, conclusions an future work are presented in Section 6.

2 Literature Review

Several approaches have been proposed to solve the problem of hand segmentation. As an overview, segmentation methods are more complicated, as the background increases in difficulty. Early works attempted to isolate hand from background, given a monochromatic or apriori known background [11,16]. However, more challenging backgrounds were demanded when hand biometrics required no constraint on background, providing solutions for contact-less approaches [3,8,13,5].

At present, hand biometrics are oriented to mobile applications in order to provide more security to mobile devices [15,1]. Therefore, more complicated algorithms are proposed given both the complexity of the segmentation aim and that mobile devices are increasing their capability to carry out complex operation. These algorithm are based on multi-aggregation strategies [14,2]. Moreover, the inclusion of small objects like rings, bracelets and watches has received also attention in previous work [18,14].

In addition to this, a wide range of methods are proposed to detect tips and valleys, in order to provide a starting point for a posterior feature extraction procedure. Most common approaches consider to work on the hand contour [8,10] extracting maximum and minimum values from several transformation on such a contour. In contrast, other methods proposed to separate fingers separately [18] but not to detect tips or valleys. Instead, this finger isolation is proposed to correct the effect of rings on fingers.

However, these approaches frequently lack of precision in peg-free or contact-less environments, since hand can be rotated or presented with different poses, having effect on hand contour. Furthermore, hand contour works properly as long as fingers are separated one from each other, otherwise, the contour-based approach provides imprecise and fuzzy results [6,7,19].

Therefore, there exist justification to explore new approaches on both segmentation and finger tip and valley detection, with the aim of extending feature extraction to more challenging and non-collaborative acquisition environments.

3 Segmentation

Before extracting features it is required to isolate hand completely from background. In this case, hands were obtained with a scanner, so that background is under control, although there exist difficulty in segmentation due to the flat surface to be streamed up.

The proposed segmentation algorithm is based on multiscale aggregation [2,14]. Concretely, the method considers image I as a graph $G = (V, E, W)$, where nodes $v_i \in V$ correspond to pixels in image; edges $e_{i,j} \in E$ represent the union between two nodes v_i and v_j; weights $w_{i,j} \in W$ describe the similarity between two nodes v_i and v_j associated by an edge $e_{i,j}$.

The main contribution of this algorithm is to describe each node as a similarity function based on a specific neighbourhood. In other words, each node v_i is described as a function ϕ_{v_i}, assuming a normal distribution $\mathcal{N}(\mu, \sigma)$ in terms of intensities within the 4-neighbour structure. Parameters μ and σ make reference to the average and standard deviation of the intensity in the proposed neighbour structure.

Therefore, the weight $w_{i,j}$ is defined in terms of functions ϕ_{v_i} and ϕ_{v_j} as in Equation 1:

$$w_{i,j} = \int_{\alpha} \sqrt{\phi_{v_i} \phi_{v_j}} d\alpha \tag{1}$$

where α represents the color space. In this case, this color space corresponds to CIELAB, concretely the a layer.

The method carries out the following procedure until only two segment remains:

- Obtain set graph G for image I
- Order pair of nodes according to weights W
- Aggregate nodes in descendent order, based on previous ordering in W
- Calculate function ϕ for each aggregated segment.
- Provide neighbour structure applying Delaunay triangulation

Finally, this method comes out with precise and accurate results for hand segmentation [14], in comoparison to other more demanding approaches [11,17].

4 Feature Extraction

This section describe an approach to detect properly tips and valleys given a hand image. This approach is opposite to those provided within the literature, not being entirely based on hand contour [16,8]. Therefore, the section will be divided into two parts describing both tips and valleys detection. In addition, the proposed scheme is able to classify fingers and thus associate each tip to their corresponding finger.

4.1 Finger Classification

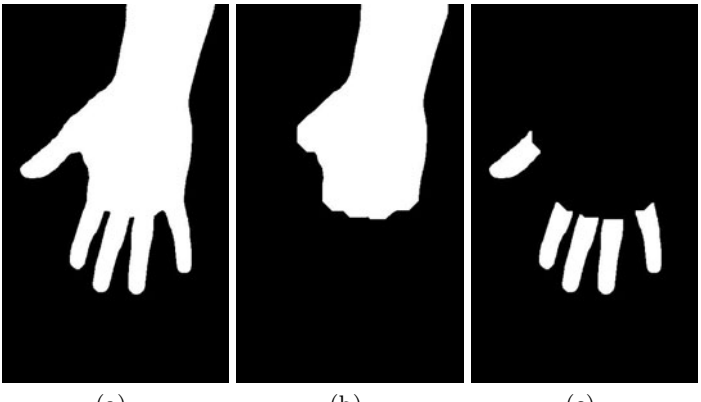
First of all, fingers are splitted from the segmented hand in order to facilitate their classification. Let H be the result provided by segmentation procedure. Applying an opening morphological operator [9] with a disk structural element of size 40 will cause fingers to dissapear, remaining only the part corresponding to palm. This image is named H_p, since it represent those pixels corresponding to palm. Although this operation is very severe, it allows conserve those region blobs which are very dense in terms of pixels, being suitable for deleting prominent blobs like fingers from hand.

Given H and H_p, it is straightforward to calculate H_f which represents the region blobs corresponding to fingers (five fingers), by the following relation (Equation 2)

$$H_f = H \cdot \bar{H}_p \tag{2}$$

being $\cdot$ an operator indicating a logical AND operation between H and the complementary of H_p.

Figure 1 provides a visual example of the fingers isolation method.



(a) (b) (c)

Fig. 1. Fingers isolation steps: (a) represents the original segmented image, H; (b) the result after applying morphological operator (opening, disk 40), H_p; (c) H_f represents fingers after subtracting H_p to H

Afterwards, five blobs are contained in H_f (Figure 1)one of each corresponding to each finger. In case more than five blobs are obtained, an opening morphological operator based on a small disk structural element (size 5) will erase those small and undesired region blobs, with lack of interest for a finger classification.

In order to distinguish among fingers, all of them are classified according to two criteria: relation between blob length and width and area (number of pixels within blob).

The blob which verifies to have the lowest values in both criteria is the little finger. The next finger with lower area is thumb, and ring, middle and indez are classified according to the distance between their centroids to previous calculated fingers. In other words, that blob whose centroid is closer to little is classified as ring finger, for instance.

4.2 Tip Detection

Having the finger blobs calculated, tip detection consists of calculating the finger extrema. In other words, obtain the furthest pixel in each blob in relation to a reference point, which coincides with the hand centroid, due to their geometric properties of being located in the middle of the palm.

The furthest pixel position is obtained then in relation to hand centroid and based on euclidean distance between two different pixels within image.

Since there are five fingers, this method would lead to five tips.

4.3 Valley Detection

In contrast to tip detection, obtaining valleys requires more effort. Let c be the hand contour obtained from the edge blob in H. Let t_k be a finger tip corresponding to finger k, with $k = \{t, i, m, r, l\}$ meaning thumb, index, middle, ring and little respectively. In addition, $\zeta = c(t_k, t_{k+1})$ is the edge portion from tip t_k and t_{k+1}. Valley points verify to be the closest point to hand centroid h_c, opposite to tip points. However, only little-ring, ring-middle and middle-index valleys support this criterion. The valley corresponding to index-thumb will be treated separately.

Then, the former valleys are calculated according to Equation 3

$$v_k = argmin(||\zeta - h_c||) \tag{3}$$

Finally, index-thumb valley is calculated as the point which provides the biggest area between candidates in $c(t_i, t_t)$, t_t and t_i.

Notice that valley detection is a considerable challenging task, given that some fingers could be together one to each other, difficulting the valley point calculation.

4.4 Left/Right Hand Classification

Hand can be classified as right or left by using three points: t_t, t_l and h_c. Two vectors are considered, joining h_c to each point tip t_t and t_l, which are represented by v_T and v_L respectively. These former vectors are on the same plane, so that their cross-vector product will be normal to that plane.

There exist a direct relation between right-left hand classification and vector $v_T \times v_L$. The sign of the z component of $v_T \times v_L$ is associated with right hand, in case the sign is positive and left hand, otherwise.

This left/right classification can be used to ensure the correct identification of little and thumb, given that images in the training and testing databases correspond to right hands.

5 Results

The results provided under this section regards the evaluation in terms of accuracy of the tip and valley detection. The requirements of the algorithm were to detect properly points (tips and valleys) in comparison to a ground-truth set of points. The result tolerance should be less than 20 pixels, being in that case well detected, otherwise uncorrectly detected.

Figure 2 is provided to show the histogram of scores for the training database, which contained a total of 300 hand images [12]. Scores represent the distance between groud-truth (provided within the training database) and the proposed algorithm result.

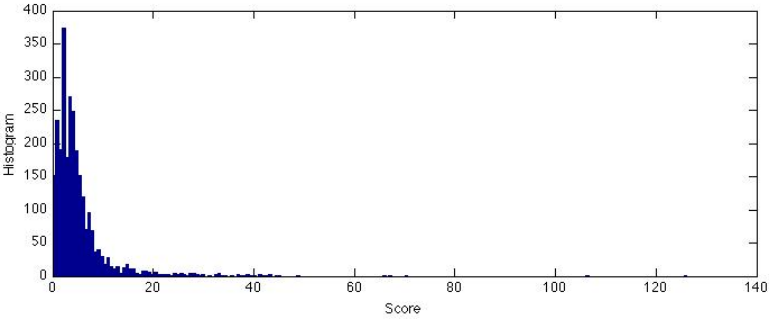


Fig. 2. Histogram of scores for the training datase (300 images)

Most values are lower than the threshold provided by the 20 pixels. In fact, the score average value is 5.13±6.66, with a median of 3.61, which is considerably far from the provided threshold.

Most errors (55.5% of total errors) were due to the index-thumb valley, which authors found very difficult to detect. In addition, the majority of errors were based on an uncorrect valley detection (96.3% of total errors). In total, the algorithm detect unproperly 81 of 2700 points (3%) of the training database.

Finally, the time performance of this approach is described in Table 1. Time results have been measured based on a MATLAB implementation to be run in a PC computer @2.4 GHz Intel Core 2 Duo with 4GB 1067 MHz DDR3 of memory.

Table 1. Time performance of the different operations from segmentation to tip and valley detection. Times are measured in seconds.

Operation	Time (seconds)
Segmentation	0.23±0.02
Fingers Extraction	1.59±0.02
Fingers Classification	< 0.1
Tip Detection	0.20±0.01
Valley Detection	1.07±0.01

Furthermore, points are calculated in less than 4 seconds, given previous implementation and the computer.

6 Conclusions

A hand tip and valley detector algorithm has been presented within this paper. The method is able to locate these points independently from orientation, hand openness and hand colour.

The acquisitions were collected with a scanner, providing the same background for each hand image, being a proper approach to segment image from background a method based on adaptive threshold, combining both low computational cost and efficiency in segmentation.

In addition, a classification is provided to associate each tip to their corresponding finger, so that a posterior left-right classification can be done. This is essential in applications where users can provide any of both hand, and the system must distinguish between them.

Tip and valley detection is described based on mathematical morphology and simple geometrical operations. The results (a score average value of 5.13 ± 6.66 and a success rate of 97%) highlight the fact that the overall method (segmentation and feature extraction) is able to detect properly and precisely the required points. However, more efforts are required in valley detection, since they cause the 96% of the total errors. Moreover, its low computational cost makes this algorithm remarkably suitable and appropriate for contact-less hand biometric applications. In addition, this algorithm achieved the second position in International Hand Geometric Points Detection Competition HGC2011 [12].

Finally, as future work, authors would like to explore the applications of the proposed feature extraction to a biometric system, and evaluate to what extent the proposed method improves the current feature extraction approaches in the literature.

References

1. Alhussain, T., Drew, S., Alfarraj, O.: Biometric authentication for mobile government security. In: 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS), vol. 2, pp. 114–118 (2010) iD: 1
2. Alpert, S., Galun, M., Basri, R., Brandt, A.: Image segmentation by probabilistic bottom-up aggregation and cue integration. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2007, pp. 1–8 (June 2007)
3. Arif, M., Brouard, T., Vincent, N.: Personal identification and verification by hand recognition. In: 2006 IEEE International Conference on Engineering of Intelligent Systems, pp. 1–6 (2006)
4. Ashbourn, J.: Practical implementation of biometrics based on hand geometry. In: IEE Colloquium on Image Processing for Biometric Measurement, pp. 5/1–5/6 (1994) iD: 1

5. de Santos Sierra, A., Guerra Casanova, J., Sánchez Ávila, C., Jara Vera, V.: Silhouette-based hand recognition on mobile devices. In: 43rd Annual 2009 International Carnahan Conference on Security Technology, pp. 160–166 (October 2009)
6. Doublet, J., Lepetit, O., Revenu, M.: Contact less hand recognition using shape and texture features. In: 2006 8th International Conference on Signal Processing, vol. 3 (2006) iD: 1
7. Doublet, J., Lepetit, O., Revenu, M.J.: Contactless hand recognition based on distribution estimation. In: Biometrics Symposium, pp. 1–6 (2007) iD:1
8. Ferrer, M., Fabregas, J., Faundez, M., Alonso, J., Travieso, C.: Hand geometry identification system performance. In: 43rd Annual 2009 International Carnahan Conference on Security Technology, 2009, pp. 167–171 (5-8, 2009)
9. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Addison-Wesley Longman Publishing Co., Inc., Boston (1992)
10. Kanhangad, V., Kumar, A., Zhang, D.: Combining 2d and 3d hand geometry features for biometric verification. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009, pp. 39–44 (20-25, 2009)
11. Lew, Y.P., Ramli, A.R., Koay, S.Y., Ali, R., Prakash, V.: A hand segmentation scheme using clustering technique in homogeneous background. In: Student Conference on Research and Development, SCOReD 2002, pp. 305–308 (2002) iD: 1
12. Magalhaes, F., Oliveira, H.P., Matos, H., Campilho, A.: hGC2011 - Hand Geometric Points Detection Competition Database (2010), http://www.fe.up.pt/~hgc2011/
13. Morales, A., Ferrer, M., Alonso, J., Travieso, C.: Comparing infrared and visible illumination for contactless hand based biometric scheme. In: 42nd Annual IEEE International Carnahan Conference on Security Technology, ICCST 2008, pp. 191–197 (2008)
14. García-Casarrubios Muñoz, Á., Sánchez Ávila, C., de Santos Sierra, A., Guerra Casanova, J.: A Mobile-Oriented Hand Segmentation Algorithm Based on Fuzzy Multiscale Aggregation. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Chung, R., Hammoud, R., Hussain, M., Kar-Han, T., Crawfis, R., Thalmann, D., Kao, D., Avila, L. (eds.) ISVC 2010, Part I. LNCS, vol. 6453, pp. 479–488. Springer, Heidelberg (2010)
15. Munoz, A.C., de Santos Sierra, A., Ávila, C., Casanova, J., del Pozo, G., Vera, V.: Hand biometric segmentation by means of fuzzy multiscale aggregation for mobile devices. In: 2010 International Workshop on Emerging Techniques and Challenges for Hand-Based Biometrics (ETCHB), pp. 1–6 (2010)
16. Sanchez-Reillo, R., Sanchez-Avila, C., Gonzalez-Marcos, A.: Biometric identification through hand geometry measurements. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(10), 1168–1171 (2000)
17. Spruyt, V., Ledda, A., Geerts, S.: Real-time multi-colourspace hand segmentation. In: 2010 17th IEEE International Conference on Image Processing (ICIP), pp. 3117–3120 (2010) iD: 1
18. Yoruk, E., Konukoglu, E., Sankur, B., Darbon, J.: Shape-based hand recognition. IEEE Transactions on Image Processing 15(7), 1803–1815 (2006)
19. Zheng, G., Wang, C.J., Boult, T.E.: Application of projective invariants in hand geometry biometrics. IEEE Transactions on Information Forensics and Security 2(4), 758–768 (2007) iD: 1

Palm Vein Recognition Based on Three Local Invariant Feature Extraction Algorithms*

Mi Pan[1,2] and Wenxiong Kang[2,**]

[1] Department of Control Science and Engineering, Zhejiang University, China
[2] College of Automation Science and Engineering,
South China University of Technology, China
auwxkang@scut.edu.cn

Abstract. In contrast to minutiae features, local invariant features extracted from infrared palm vein have properties of scale, translation and rotation invariance. To determine how they can be best used for palm vein recognition system, this paper conducted a comprehensive comparative study of three local invariant feature extraction algorithms: Scale Invariant Feature Transform (SIFT), Speeded-Up Robust Features (SURF) and Affine-SIFT (ASIFT) for palm vein recognition. First, the images were preprocessed through histogram equalization, then three algorithms were used to extract local features, and finally the results were obtained by comparing the Euclidean distance. Experiments show that they achieve good performances on our own database and PolyU multispectral palmprint database.

Keywords: Palm vein, local invariant feature, pattern matching.

1 Introduction

Hand vein recognition technology, a burgeoning biometrics mainly used for identification, has been proposed in recent years. Compared to fingerprint or other biological features, hand vein, including palm vein, finger vein and back vein, is much more reliable because it is highly distinctive and difficult to be changed by operation, as well as more feasible and easier for collecting data with non-contact acquisition system. These advantages make hand vein recognition technology more accurate and promising, which has attracted an increasing amount of attentions from research communities and industries all over the world [1].

As one of the most important contents of vein recognition, feature extraction for vein images is still in the preliminary study stage. Unlike other images, vein image is particular with specific structure and simple feature, unsuited to normal methods of

* This work was supported by National Natural Science Foundation of China (No. 61105019), Natural Science Foundation of Guangdong Province (S2011040002474), Science and Technology Planning Project of Guangdong Province (2009B030803032, 2011B010200023) and the Fundamental Research Funds for the Central Universities, SCUT (2009ZM0070).

** Corresponding author.

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 116–124, 2011.

feature extraction and image matching. A majority of the existing algorithms analyze the entire image and extract vein structure features, using the similarity of the locations of features in two different vein images for matching and authentication [2]. Thus, the results are sensitive to the location of hand and some other situations, which impact the recognition accuracy. In terms of informed research, Cross et al. [3] extracted the vein skeleton and used constrained sequential correlation for matching; Wang et al. [4] utilized multi-resolution wavelet analysis for vein image feature extraction. There is still lack of simple but efficient method for vein recognition.

In this paper, we propose a kind of novel way for recognition that implements the palm vein image matching through extracting local invariant features of images, which are invariant to change in scale, translation, and rotation, even illumination and affine. These approaches have already been widely used in solving plenty of problems in generic pattern recognition and object detection field, but seldom in vein recognition area.

This paper proposes using three local feature algorithms, SIFT [5, 6], SURF [7] and ASIFT [8, 9] on palm vein recognition and gets a performance comparison through experiments on Matlab 7.0. Only SIFT has been used in vein image matching in some informed research [4, 10], and other two are new ones in vein authentication related field. Therefore, our research on the comparison of three approaches concerning palm vein recognition is meaningful and promising. Such kind of comparative studies have already been reported in literature, and we believe this study can generate a clear framework for the local invariant feature algorithms on palm vein recognition and inspire some new ideas of the related researchers.

In our work, we use the NIR sub-database of PolyU multispectral palmprint Database from the Biometric Research Centre (UGC/CRC) at The Hong Kong Polytechnic University [11]. There are 500 groups of palm vein images (352x288pixels) in this database, and each group is collected from the same hand, containing 8 samples. Additionally, we use our own database for experiments. It was collected by our own equipment, including 100 groups with 2 samples for each.

Following are the main stages of the proposed approach for the recognition of the vein images from the database:

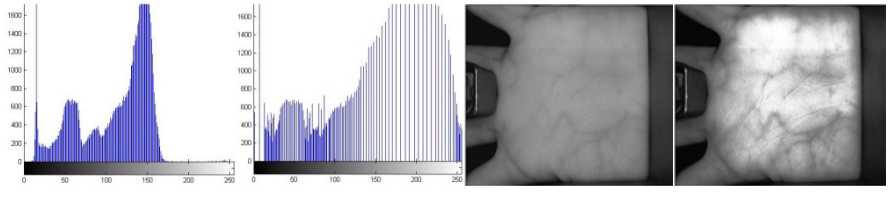
1. Image preprocessing: After extracting the Region of Interest (ROI) from the palm vein images, which is done by the collecting step of the database, we utilize histogram equalization to enhance the resolution and contrast of the images. The preprocessed images will be used for the local features extraction in the next step.
2. Local invariant features extraction: In our work, three algorithms are implemented to extract SIFT features, SURF features and ASIFT features in several.
3. Image matching and recognition: Local feature points in two images are matched based on the similarity of Euclidean distance, refined by ratio method.

According to the main steps, the following paper is organized as follows. In section 2 we present the way of image preprocessing. In section 3 we introduce three local invariant feature extraction algorithms, followed by matching and recognition method in section 4. Section 5 shows the experimental results and performance evaluation. The conclusions are given in section 6.

2 Image Preprocessing

The palm vein images in database are acquired in near Infrared light and have already extracted the ROIs, but it is still difficult for feature extraction because of the concentrated gray value. Thus, image processing should be implemented before feature extraction.

In order to enhance the image contrast, we try a lot of algorithms, including adjusting image intensity values, homomorphic filter, gamma correction and histogram equalization. Considering the effect and process-simplicity, we choose the last one, a method for histogram uniform distribution, in our preprocessing work. Although this approach is relatively coarse, it can remain the basic image feature for the later work. We use three local invariant feature extraction algorithms on original and image preprocessed by several methods for a lot of image samples, and find that the image preprocessed by histogram equalization can obtain the maximum quantity of local features. Therefore, we choose it. As can be seen in Fig.1, histogram equalization can readjust the grayscale distribution, as illustrated in Fig.1 (a) and (b), and clarify the veins as illustrated in Fig.1 (c) and (d).



(a) (b) (c) (d)

Fig. 1. An example of histogram equalization. (a) histogram for original image (b) histogram for preprocessed image (c) original image (d) preprocessed image.

3 Local Invariant Feature Extraction

3.1 SIFT on Palm Vein

The Scale Invariant Feature Transform (SIFT) was proposed by Lowe in 1999 [5] and perfected in 2004 [6]. It is based on scale space [12], having a robust performance, invariant to image scaling, translation and rotation.

More detailed illustrations can be found in the original text [5, 6], and we only give a brief description in this paper as follows. Firstly, keypoints are detected in the images through searching the scale-space extrema in difference-of-Gaussian function D. Secondly, Taylor expansion of function D is used to exactly localize the local features (keypoints). Thirdly, an orientation with its gradient magnitude is assigned to every image sample, and an orientation histogram is formed for each keypoint region to get the orientation assignment for keypoints. Finally, the gradient information is computed for a 16*16 sample array near each keypoint, and 4*4 keypoint descriptors with each containing the information in 8 orientations are generated. Then, all local features have transformed to 128-dimensional feature vectors, ready for image matching.

Using SIFT algorithm on preprocessed images for feature extraction, we can obtain the results like examples in Fig.2. Approximately 100 feature points can be obtained.

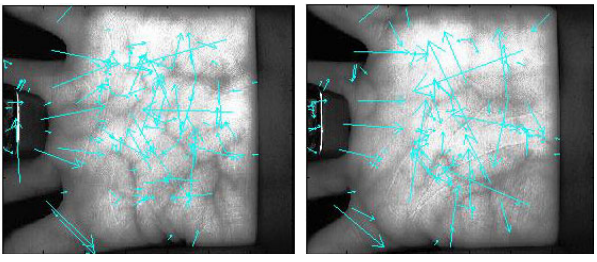


Fig. 2. Examples of SIFT features extraction

3.2 SURF on Palm Vein

Bay et al discovered that the convolution between integral image [13] and box filter has the approximation of the Gaussian convolution, thus he proposed the Speeded-Up Robust Features (SURF) [7], best known for the excellent computing speed.

SURF algorithm has the different theoretical foundation but same ideas as the SIFT. Firstly, the integral images of original input images are computed for image convolutions with box filters, and the keypoints are detected based on approximated Hessian matrix. Secondly, scale space interpolation is used for exactly keypoints localization. Thirdly, the Haar wavelet responses of all the samples are calculated for the orientation assignment. Finally, the Haar wavelet responses for every keypoint are summed up and descriptor vectors of length 64 for each are generated.

Fig.3 shows two examples of extracting SURF features, of which the number are similar to SIFT approach with less extracting time.

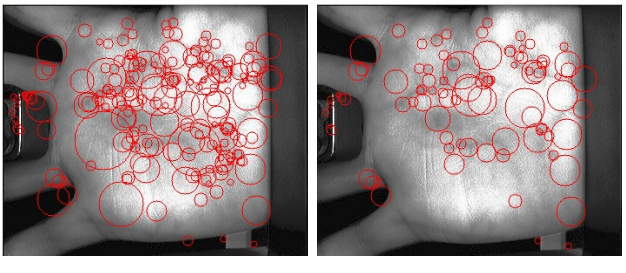


Fig. 3. Examples of SURF features extraction

3.3 ASIFT on Palm Vein

In order to extract features which are fully invariant to affine transformation, Jean-Michel Morel, presented Affine-SIFT (ASIFT) [8] in 2009. By means of two camera axis orientation parameters, the latitude and the longitude angles, ASIFT can simulate all possible affine distortions of the original images. With fixed sampling steps for the latitude and the longitude angles, we can get simulated images with different

orientation parameters on sampling points. After these steps, SIFT method is employed on the simulated images for keypoints extraction. The author points out that SIFT algorithm can be replaced by other similar ones. The website [9] also offers online demonstration and releases the source code, allowing others to do relevant research.

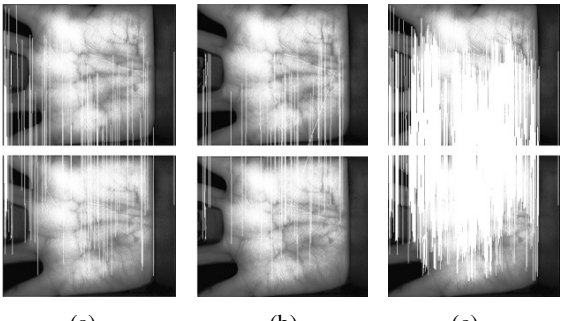
Compared to the former SIFT and SURF, ASIFT algorithm can gain quite a remarkable number of local features, more than about 100 times. Additionally, the computing time is relatively longer. Due to the tremendous number, it is ill-suited and confused to show those feature points on the images. Feature vectors are stored in text for further matching and recognition.

4 Image Matching and Recognition

After the above stages, we have already obtained local feature points from the image. Then, in view of matching and recognition, we choose a good similarity judgment, Euclidean distance between two corresponding local features in different palm vein images. During the matching of two images, the Euclidean distance of all the features in different images are computed and compared individually. A feature point in the first image is matched by the nearest point in the second one only if the ratio of the nearest neighbor and the second nearest one is less than a given threshold, normally 0.4~0.8. Namely satisfy:

$$\frac{\text{the nearest neighbour distance}}{\text{the second nearest neighbour distance}} < threshold$$

The bigger the threshold is, the more the matching pairs can be obtained, at the expense of accuracy. We select 0.6 as the threshold, and utilize the above matching approach respectively on features extracted by SIFT, SURF and ASIFT algorithm. Then, for samples from the same hand, we can have the matching results as follows.

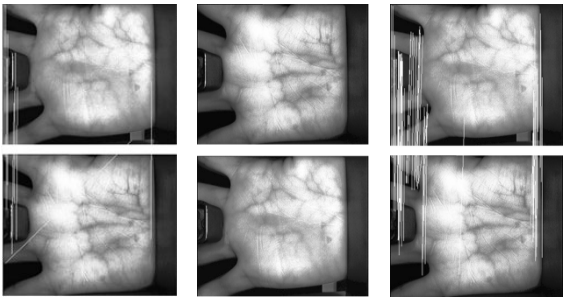


(a) (b) (c)

Fig. 4. Examples of matching results for same hand images. (a) SIFT result (b) SURF result (c) ASIFT result.

As shown in Fig.4, even though the whole calculating time is much longer, the amount of matching pairs that ASIFT algorithm obtains completely beats the other two approaches.

For samples from the different hand, we can get the matching results as shown in Fig.5. There are only few matching pairs in two matched images, especially in vein area. Therefore, this kind of approach is reasonable and accurate.



(a) (b) (c)

Fig. 5. Examples of matching results for different palm vein images. (a) SIFT result (b) SURF result (c) ASIFT result.

The recognition performance of each method depends on the correct authentication ability on the images in database. The amount of matching pairs for recognition judgment, related to authentication accuracy, varies on the basis of different algorithms.

5 Experimental Results

To evaluate the recognition performance of the proposed methods, False Rejection Rate (FRR) and False Accept Rate (FAR) should be used. FRR is the rate of false rejection among the same hand vein images, while FAR is the rate of false acceptation among different hand vein images.

When defining a threshold of the number of matching pairs to judge whether two vein images are recognized, we have the following definition. If the matching pairs of two images are greater than the threshold, we call it like couple, otherwise, unlike couple. And if the matching experiment is implemented between the same hand images, we call it inner experiment, otherwise, outer experiment. Then, we get the following calculating equations.

$$FRR = \left(N_1 / N_{ti} \right) \times 100\% \tag{1}$$

$$FAR = \left(N_2 / N_{to} \right) \times 100\% \tag{2}$$

Where N_1 is unlike couples in inner experiments and N_2 is like couples in outer experiments. N_{ti} and N_{to} are the total times of inner experiments and outer ones.

Ideally, we desire that the value of FRR and FAR be as low as possible. In real-world palm vein recognition systems the FAR and FRR can typically be traded off against each other by changing some parameters. One of the most common measures is the rate at which both accept and reject errors are equal, which is called the Equal Error Rate (EER).

For NIR database, we randomly select 100 groups of palm vein images in the database for matching and recognition experiments. For SIFT and SURF algorithm, we carry out 1500 times inner experiment and 19800 times outer one, for each plotting the ROC curve, a plot of FRR and FAR on the grounds of the varying thresholds of feature matching pairs.

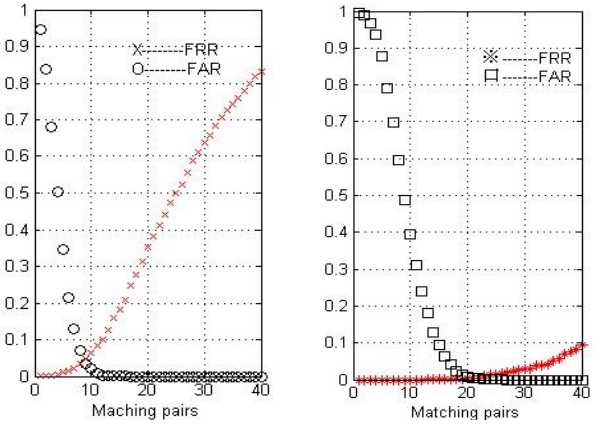


Fig. 6. Recognition ROC curves of SIFT (left) and SURF (right) on NIR database

Fig.6 (a) shows the recognition results of SIFT algorithm, the value of EER is about 2%, which indicates a good performance of SIFT on palm vein recognition. Fig.6 (b) shows the recognition results of SURF algorithm, the EER is about 0.4%, less than that of SIFT. Therefore, the recognition performance of SURF is much better than the SIFT. Besides, the value of FRR stays relatively low even the FAR arrives at 0.

Because of the low calculating efficiency executed in single-thread manner, it is unreasonable and complicated to conduct such a large number of experiments as above on ASIFT algorithm. After carrying out experiments of 100 groups, each select 3 samples, we find that the number of matching pairs are more than 400 for inner experiments, while less than 50 for outer ones. Therefore, the EER is 0 for ASIFT approach, meaning that ASIFT has a quite accurate recognition performance.

In terms of extracting time and matching time, the former is related to algorithms while the latter, the amount of local features. We compare the extracting time of each algorithm for evaluation, which is the average time cost for feature extraction in two images in one experiment.

The comparison of recognition performance of three algorithms can be seen in Table 1, which illustrates that SURF has the best synthetic performance in terms of both computation speed and recognition accuracy. SIFT is a little inferior in our experiments. ASIFT is highly accurate but quite time-consuming, which can be solved by multi-thread operation, as the online demo shows[9].

Table 1. Comparison of recognition performance of three algorithms

Algorithm	EER/%	Extracting time/s
SIFT	2.2	2.79
SURF	0.4	1.78
ASIFT	0	37

For our own database, we use all 100 groups palm vein samples, and conduct 200 times of inner experiment and 19800 times outer one. Then, we have the FRR and FAR curves for SIFT and SURF as shown in Fig.7.

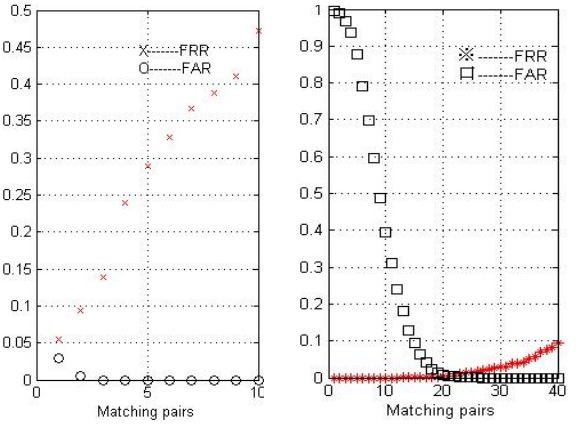


Fig. 7. Recognition ROC curves of SIFT (left) and SURF (right) on our own database

The EER of both SIFT and SURF algorithm is about 4%, which once again prove the feasibility and applicability of using SIFT and SURF on palm vein image recognition.

In the meantime, ASIFT algorithm still generates obvious different results for inner experiments and outer ones. Thus, the EER is 0 and ASIFT remains the high accuracy as the former recognition experiments.

6 Conclusion

In this paper, we applied three local invariant feature algorithms, SIFT, SURF and ASIFT, on palm vein recognition, and obtained satisfactory results. We used histogram equalization method to preprocess the collected images and then extracted local features respectively by three algorithms. Experimental results indicate that all three can extract enough invariant features from vein image of the same hand and can get correct matching and recognition. For images of different hand, they can hardly be identified. Moreover, SURF algorithm has the best synthetic performance and ASIFT has the highest accuracy.

Further researches need to be explored in several aspects, including more accurate method for image preprocessing, similar experiments on more databases, and other local feature algorithms on vein recognition.

References

1. Li, Q., Zeng, Y., Peng, X., Yan, K.: Curvelet-based palm vein biometric recognition. Chin. Opt. Lett. 8(6), 577–579 (2010)
2. Wang, Y., Liu, T., Jiang, J., Zhang, Z., Zhou, S.: Hand vein recognition using local SIFT feature analysis. Journal of Opto-electronics·Laser 20(5), 681–684 (2009)
3. Cross, J.M., Smith, C.L.: Thermo graphic Imaging of the Subcutaneous Vascular Network of the Back of the Hand for Biometric Identification. In: Proceedings of 29th International Carnahan Conference on Security Technology 20 (1995)
4. Wang, Y., Liu, T., Jiang, J.: A multi-resolution wavelet algorithm for hand vein pattern recognition. Chin. Opt. Lett. 6(9), 657–660 (2008)
5. Lowe, D.G.: Object Recognition from Local Scale-Invariant Features. In: Proceeding of the International Conference on Computer Vision, pp. 1150–1157 (1999)
6. Lowe, D.G.: Distinctive image features from scale- invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
7. Bay, H., Tuytelaars, T., Gool, L.: SURF: Speeded Up Robust Features. In: Proceeding of the 9th European Conference on Computer Vision, pp. 404–417 (2006)
8. Morel, J.M., Yu, G.: ASIFT: A New Framework for Fully Affine Invariant Image Comparison. SIAM Journal on Imaging Sciences (2009)
9. Morel, J.M., Yu, G.: http://www.ipol.im/pub/algo/my_affine_sift/
10. Ladoux, P.-O., Rosenberger, C., Dorizzi, B.: Palm Vein Verification System Based on SIFT Matching. In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 1290–1298. Springer, Heidelberg (2009)
11. PolyU multispectral palmprint Database., http://www.comp.polyu.edu.hk/~biometrics/MultispectralPalmprint/MSP.htm/
12. Lindeberg, T.: Scale-space theory: A basic tool for analyzing structures at different scales. Journal of Applied Statistics 21(2), 224–270 (1994)
13. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2001)

Finger Knuckleprint Based Recognition System Using Feature Tracking

Aditya Nigam and Phalguni Gupta

Indian Institute of Technology Kanpur, Kanpur - 208016, India
{naditya,pg}@cse.iitk.ac.in
http://www.cse.iitk.ac.in

Abstract. This paper makes use of finger knuckleprints to propose an efficient biometrics system. Edge based local binary pattern (ELBP) is used to enhance the knuckleprint images. Highly distinctive texture patterns from the enhanced knuckleprint images are extracted for better classification. It has proposed a distance measure between two knuckleprint images. This system has been tested on the largest publicly available Hong Kong Polytechnic University (PolyU) finger knuckleprint database consisting 7920 knuckleprint images of 165 distinct subjects. It has achieved CRR of more than 99.1% for the top best match, in case of identification and ERR of 3.6%, in case of verification.

Keywords: Biometrics,Knuckleprint, Local Binary Pattern, Lukas Kanade Tracking, Edge-map.

1 Introduction

Biometrics authentication is extensively applied in law enforcement, computer security, banking etc. Exponential increase in the computational power and the requirement of the society lead researchers to develop fast, efficient and low cost authentication systems to meet the real time challenges.

Recently a significant amount of research has been carried out to design efficient biometric systems. Several traits such as face, fingerprint, iris, palmprint, ear, signature *etc* are investigated exhaustively. Every trait has its own pros and cons. There exist various challenges depending on the trait such as pose and illumination for face, occlusion and cooperative acquisition for iris *etc*.

Biometric recognition systems based on hand (*e.g.* palm print and fingerprint) have gathered attraction over past few years because of their good performance and inexpensive sophisticated acquisition sensors. Pattern formation at finger knuckle bending are unique [1,2,3] and hence can be considered as a discriminative biometrics trait. Factors favoring knuckleprint include higher user acceptance and less expected user cooperation.

Patterns extracted from finger knuckle surface have high discriminative power [4] and can be useful for personal identification. Surface curvatures of knuckleprints have been considered for matching. But it has not performed well because of its large size and costly as well as time consuming acquisition system. 2D finger knuckle surface has been used in [5] for authentication by combining several global feature extraction methods

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 125–132, 2011.

such as ICA, PCA and LDA. But these methods do not extract features based on knuckle lines well. Hence, these methods have achieved limited performance. Local features such as robust line orientation code (RLOC) [6] and modified finite radon transform (MFRAT) [3] have been proposed to extract the local pixel orientation and stored as *knucklecode*.

Zhang *et. al* [7] have designed a low cost CCD based finger knuckle data acquisition system and extracted the region of interest using convex direction coding for knuckleprint verification. It calculates the correlation between two knuckleprint images using band limited phase only correlation (BLPOC) where high frequencies are not considered as they are prone to noise. In [8] bank of gabor filters has been used to extract the features. It considers pixels that have varying gabor response and fused orientation and magnitude information to get much better performance. In [2] local and global features are fused to get better performance.

In [9], SIFT features which are invariant to rotation and scaling are considered as key-points. Real part of orthogonal gabor filter with contrast limited adaptive histogram equalization (CLAHE) has been used to enhance knuckleprint images to compensate non-uniform reflection. Knuckleprint based recognition system using local gabor binary patterns (LGBP) has been proposed in [10]. Eight gabor filters are applied on a knuckleprint image and histogram features are extracted by 8-neighborhood LBP within blocks. Classification of test samples is done using chi-squared distance statistics.

This paper proposes a measure to compare finger knuckleprint images which is robust against slight amount of local non-rigid distortions. Its performance has been studied on the largest publicly available Hong Kong Polytechnic University (PolyU) finger knuckleprint database [11].

The paper is organized as follows. Section 2 discusses mathematical basis of the proposed system. Section 3 proposes an efficient knuckleprint based recognition system. The system has been analyzed in Section 4. Conclusions are given in the last section.

2 Mathematical Basis

2.1 LBP Based Image Enhancement

In [12] a transformation, very similar to LBP [13] has been introduced to preserve the distribution of gray level intensities in iris images. It helps to address the problems like robustness against illumination variation and local non-rigid distortions. It is observed that a pixel's relative gray value with respect to its 8-neighborhood pixels can be more stable than its own gray value. But this transformation fails when gray values of 8-neighbors are very similar to each other. In [14,15] gt-transformation providing more tolerance to variations in illumination and local non-rigid distortions is proposed. It is observed that gray level intensities are indistinguishable within a small range.

2.2 Lukas Kanade Tracking

LK tracking algorithm [16] estimates the sparse optical flow between two frames. Let there be a feature at location (x, y) at time instant t with intensity $I(x, y, t)$ and this

feature has moved to the location $(x + \delta x, y + \delta y)$ at time instant $t + \delta t$. Three basic assumptions used by LK Tracking [16] are:

- **Brightness Consistency:** Features on a frame do not change much fo small value of δt, *i.e*

$$I(x, y, t) \approx I(x + \delta x, y + \delta y, t + \delta t) \tag{1}$$

- **Temporal Persistence:** Features on a frame moves only within a small neighborhood. It is assumed that features have only small movement for small value of δt. Using the Taylor series and neglecting the high order terms, one can estimate $I(x + \delta x, y + \delta y, t + \delta t)$ as

$$\frac{\delta I}{\delta x}\delta x + \frac{\delta I}{\delta y}\delta y + \frac{\delta I}{\delta t}\delta t = 0 \tag{2}$$

  Dividing both sides of Eq 2 by δt one gets

$$I_x V_x + I_y V_y = -I_t \tag{3}$$

  where V_x, V_y are the respective components of the optical flow velocity for pixel $I(x, y, t)$ and I_x, I_y and I_t are the derivatives in the corresponding directions.

- **Spatial Coherency:** In Eq 3, there are two unknown variables for every feature point (*i.e* V_x and V_y). Hence finding unique V_x and V_y for every feature point is an ill-posed problem. Spatial coherency assumption is used to solve this problem. It assumes that a local mask of pixels moves coherently. Hence one can estimate the motion of central pixel by assuming the local constant flow. LK gives a non-iterative method by considering flow vector (V_x, V_y) as constant within 5×5 neighborhood (*i.e* 25 neighboring pixels, $P_1, P_2 \ldots P_{25}$) around the current feature point (center pixel) to estimate its optical flow. The above assumption is reasonable and fair as all pixels on a mask of 5×5 can have coherent movement. Hence, one can obtain an overdetermined linear system of 25 equations which can be solved using least square method as

$$\underbrace{\begin{pmatrix} I_x(P_1) & I_y(P_1) \\ \vdots & \vdots \\ I_x(P_{25}) & I_y(P_{25}) \end{pmatrix}}_{\text{C}} \times \underbrace{\begin{pmatrix} V_x \\ V_y \end{pmatrix}}_{\text{V}} = -\underbrace{\begin{pmatrix} I_t(P_1) \\ \vdots \\ I_t(P_{25}) \end{pmatrix}}_{\text{D}} \tag{4}$$

  where rows of the matrix C represent the derivatives of image I in x, y directions and those of D are the temporal derivative at 25 neighboring pixels. The 2×1 matrix $\widehat{\mathbf{V}}$ is the estimated flow of the current feature point determined as

$$\widehat{\mathbf{V}} = (C^T C)^{-1} C^T (-D) \tag{5}$$

The final location $\widehat{\mathbf{F}}$ of any feature point can be estimated using its initial position vector $\widehat{\mathbf{I}}$ and estimated flow vector $\widehat{\mathbf{V}}$ as

$$\widehat{\mathbf{F}} = \widehat{\mathbf{I}} + \widehat{\mathbf{V}} \tag{6}$$

3 Proposed System

The proposed system consists of three components: image enhancement, feature extraction and matching. Image enhancement is performed using edge based local binary pattern (ELBP). Corner features are extracted using the method proposed in [17] while LK tracking [16] is used for matching. Details of each task are discussed in the following subsections.

3.1 Image Enhancement

Knuckleprint images have strong vertical edges that can be useful for recognition purposes. Proposed transformation calculates *edge based local binary pattern* (ELBP) for each pixel in the image. A knuckleprint image is transformed into an *edgecode* (as shown in Fig. 1) that is robust to illumination and local non-rigid distortions. Knuckleprint image A is preprocessed by applying the sobel edge operator in horizontal direction to obtain vertical edge map. To obtain the *edgecode*, ELBP value for every pixel $A_{j,k}$ in the vertical edge map is defined as a 8 bit binary number S whose i^{th} bit is

$$S_i = \begin{cases} 0 & \textbf{if } (Neigh[i] < threshold) \\ 1 & \textbf{otherwise} \end{cases} \tag{7}$$

where $Neigh[i], i = 1, 2, ...8$ are the horizontal gradient of 8 neighboring pixels centered at pixel $A_{j,k}$. The value of *threshold* is evaluated experimentally.

In *edgecode* (as shown in Fig. 1), every pixel is represented by its $ELBP$ value which is an encoding of strong edge pixels in its 8-neighborhood. It can be noted that any change caused due to sudden change in the illumination can affect the gray values but $ELBP$ value is not affected much because the strong edge pattern near the pixel remains to be more or less same. This property has been used in knuckleprint images as it contains lot of illumination variation.

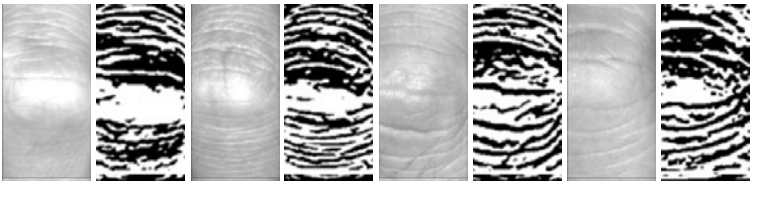


Fig. 1. Original and Transformed (*edgecodes*) knuckleprint Images

3.2 Feature Extraction

Strong derivative points except corner ones in the $edgecode$ cannot be considered as features because they look alike along the edge. But corners have strong derivative in two orthogonal directions and can provide enough information for tracking. The autocorrelation matrix M can be used to calculate good features from $edgecode$ having

strong orthogonal derivatives. Matrix M can be defined for any pixel at i^{th} row and j^{th} column of $edgecode$ as:

$$M(i,j) = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{8}$$

such that

$$\begin{aligned} A &= \sum_{-K \le a,b \le K} w(a,b).I_x^2(i+a, j+b) \\ B &= \sum_{-K \le a,b \le K} w(a,b).I_x(i+a, j+b).I_y(i+a, j+b) \\ C &= \sum_{-K \le a,b \le K} w(a,b).I_y(i+a, j+b).I_x(i+a, j+b) \\ D &= \sum_{-K \le a,b \le K} w(a,b).I_y^2(i+a, j+b) \end{aligned}$$

where $w(a,b)$ is the weight given to the neighborhood, $I_x(i+a, j+b)$ and $I_y(i+a, j+b)$ are the partial derivatives sampled within the $(2K+1) \times (2K+1)$ window centered at each selected pixel.

The matrix M can have two eigen values λ_1 and λ_2 such that $\lambda_1 \ge \lambda_2$ with e_1 and e_2 as the corresponding eigenvectors. Like [17], all pixels having $\lambda_2 \ge T$ (smaller eigen value greater than a threshold) are considered as corner feature points. Let $a = \{i, j\}$ be a 2-tuple array to indicate that $(i,j)^{th}$ pixel of the knuckleprint image A, $A_{i,j}$ is a corner point.

3.3 Matching

Let A and B be two knuckleprint images that are to be compared. Let a and b be the 2-tuple arrays containing the corner information of knuckleprint images A and B respectively. In order to make the decision on matching between A and B, LK Tracking, discussed in Section 2, has been used to determine the average number of features tracked successfully in one knuckleprint image against all corner points of another image. Let $a(i,j)$ be a corner point of knuckleprint image A. LK Tracking calculates its estimated location in $edgecode$ of B, say $edgecode_B(k,l)$. For every $a(i,j)$ of a, we tell that a pixel $a(i,j)$ is tracked successfully if the euclidean distance between $a(i,j)$ and $edgecode_B(k,l)$ is less than or equal to a preassigned threshold, TH_d and the sum of the absolute difference between every neighboring pixel of $a(i,j)$ and $edgecode_B(k,l)$, termed as tracking error, is less than or equal to a preassigned threshold, TH_e. Thus,we can define $Tracked(a(i,j), edgecode_B)$ for successful/unsuccessful tracking as

$$Tracked(a(i,j), edgecode_B) = \begin{cases} 1 \textbf{ if } ||a(i,j), b(k,l)|| \le TH_d \\ \quad \textbf{and } T_{Error} \le TH_e \\ \\ 0 \textbf{ otherwise} \end{cases} \tag{9}$$

where T_{Error} is the tracking error. For every point in a, one can determine whether it can successfully tracks a pixel in $edgecode_B$. Features Tracked Successfully (fts) for a to $edgecode_B$ can be defined by

$$fts(a, edgecode_B) = \sum_{\forall a(i,j) \in a} Tracked(a(i,j), edgecode_B)) \tag{10}$$

Thus,the average number of features tracked successfully (FTS) for a to $edgecode_B$ and b to $edgecode_A$ is defined by

$$FTS(A,B) = \frac{1}{2} \times [fts(a, edgecode_B) + fts(b, edgecode_A)] \tag{11}$$

4 Experimental Results

This section analyses the performance of the proposed system. It has been evaluated on the publicly available largest FKP database from the Hong Kong Polytechnic University (PolyU) [11]. This database contains 7920 FKP images obtained from 165 subjects. Images are acquired in two sessions. At each session, 6 images of 4 fingers (distinct index and middle fingers of both hands) are collected. Subjects comprise of 125 males and 40 females. The age distribution of users are as follows: 143 subjects are having age lying between 20 and 30 while remaining are between 30 and 50. Like [1,2,3,7,8,9,10] images collected in first session are considered for training and those collected in the second session are used for query.

Performance of the system is measured using correct recognition rate (CRR) in case of identification and equal error rate (EER) for verification. CRR of the system is defined by

$$CRR = \frac{N_1}{N_2} \tag{12}$$

where N_1 denotes the number of correct (Non-False) top best match of FKP images and N_2 is the total number of FKP images in the query set.

At a given threshold, the probability of accepting the impostor, known as false acceptance rate (FAR) and probability of rejecting the genuine user known as false rejection rate (FRR) are obtained. Equal error rate (EER) is the value of FAR for which FAR and FRR are equal.

$$EER = \{FAR | FAR = FRR\} \tag{13}$$

The proposed FTS measure is parametrized by two parameters TH_d and TH_e. TH_d depends on the amount of expected motion and TH_e is the pixel-wise patch absolute difference around the initial and estimated feature. Both of these parameters are calculated empirically. These values are chosen in such a way that CRR is maximum.

The proposed system has been compared with all well known knuckleprint based systems reported in [10]. It is found that the CRR of the proposed system which is more than 99.1% for all 4 fingers is better. $CRRs$ of various systems obtained from various fingers of the PolyU database are shown in Table 1. Further, EER of the proposed verification system is 3.6%. For each finger, Receiver Operating Characteristics

Table 1. Identification Performance (compared as reported by [10])

	CRR % Left Index	CRR % Left Middle	CRR % Right Index	CRR % Right Middle
PCA [10]	0.5638	0.5364	0.6051	0.6010
LDA[10]	0.7283	0.7030	0.7606	0.7525
Gabor+PCA[10]	0.9253	0.9101	0.9586	0.9293
Gabor+LDA[10]	0.9485	0.9263	0.9626	0.9323
LBP[10]	0.9010	0.8909	0.9556	0.9121
LGBP[10]	0.9414	0.9424	0.9727	0.9475
Proposed	**0.9910**	**0.9926**	**0.9936**	**0.9922**

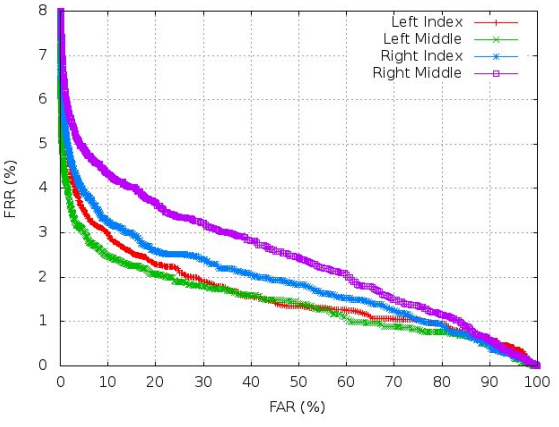


Fig. 2. Fingerwise ROC Curves

(ROC) curves which plots FAR against FRR is shown in Fig. 2. It is observed that the performance of left hand fingers is better than right hand fingers. It can be noted here that previously known systems have not reported their respective EER and hence the proposed system could not be compared.

5 Conclusion

This paper has proposed a measure termed as Features Tracked Successfully (FTS) to compare shapes in gray scale images. Further, this measure is used to design a finger knuckleprint based biometric system. This system works on edge-maps to compensate the effect of illumination variations.

It works on the features obtained from gray images and uses FTS measure to achieve the appearance based comparison on knuckleprint images. FTS measure has experimentally shown tolerance to slight variation in translation, illumination and rotation. The system has been tested on publicly available PolyU database of knuckleprint images. It has considered knuckleprints of 4 fingers (index and middle fingers of both

hands) of 165 subjects to measure its performance. It has achieved CRR of more than 99.1% for the top best match, in case of identification and EER of 3.6% in case of verification. The proposed system has been compared with all well known knuckleprint based systems and is found to perform better.

References

1. Zhang, L., Zhang, L., Zhang, D.: Finger-knuckle-print: A new biometric identifier. In: International Conference on Image Processing, ICIP, pp. 1981–1984 (2009)
2. Zhang, L., Zhang, L., Zhang, D., Zhu, H.: Ensemble of local and global information for finger-knuckle-print recognition. Pattern Recognition 44(9), 1990–1998 (2011)
3. Kumar, A., Zhou, Y.: Personal identification using finger knuckle orientation features. Electronics Letters 45(20), 1023–1025 (2009)
4. Woodard, D.L., Flynn, P.J.: Finger surface as a biometric identifier. Computer Vision and Image Understanding 100(3), 357–384 (2005)
5. Kumar, A., Ravikanth, C.: Personal authentication using finger knuckle surface. IEEE Transactions on Information Forensics and Security 4(1), 98–110 (2009)
6. Kumar, A., Zhou, Y.: Human identification using knucklecodes. In: 3rd IEEE International Conference on Biometrics: Theory, Applications and Systems. BTAS (2009), pp. 147–152 (2009)
7. Zhang, L., Zhang, L., Zhang, D.: Finger-knuckle-print verification based on band-limited phase-only correlation. In: Jiang, X., Petkov, N. (eds.) CAIP 2009. LNCS, vol. 5702, pp. 141–148. Springer, Heidelberg (2009)
8. Zhang, L., Zhang, L., Zhang, D., Zhu, H.: Online finger-knuckle-print verification for personal authentication. Pattern Recognition 43(7), 2560–2571 (2010)
9. Morales, A., Travieso, C., Ferrer, M., Alonso, J.: Improved finger-knuckle-print authentication based on orientation enhancement. Electronics Letters 47(6), 380–381 (2011)
10. Xiong, M., Yang, W., Sun, C.: Finger-knuckle-print recognition using lgbp. In: Liu, D., Zhang, H., Polycarpou, M., Alippi, C., He, H. (eds.) ISNN 2011, Part II. LNCS, vol. 6676, pp. 270–277. Springer, Heidelberg (2011)
11. Zhang, D.: Polyu finger-knuckle-print database., http://www4.comp.polyu.edu.hk/~biometrics/FKP.htm
12. Sudha, N., Wong, Y.: Hausdorff distance for iris recognition. 22nd IEEE International Symposium Intelligent Control 1(1), 614–619 (2007)
13. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis Machine Intelligence 24(7), 971–987 (2002)
14. Nigam, A., Gupta, P.: A new distance measure for face recognition system. In: International Conference on Image and Graphics, ICIG (2009), pp. 696–701 (2009)
15. Nigam, A., Gupta, P.: Comparing human faces using edge weighted dissimilarity measure. In: International Conference on Control, Automation, Robotics and Vision, ICARCV (2010), pp. 1831–1836 (2010)
16. Lucas, B.D., Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision. In: International Joint Conference on Artificial Intelligence. IJCAI (1981), pp. 674–679 (1981)
17. Shi, J.: Tomasi: Good features to track. In: Computer Vision and Pattern Recognition. CVPR (1994), pp. 593–600 (1994)

A Preliminary Study of Handprint Synthesis

Jianjiang Feng, Huapeng Zhou, and Jie Zhou

Department of Automation
Tsinghua University, Beijing, China
{jfeng,jzhou}@tsinghua.edu.cn

Abstract. Handprint, the friction ridge pattern on human hand, is of vital importance for recognizing repeat offenders and suspects in forensics. In this paper, we proposed a set of statistical models for the main features in whole handprints, including hand contour, major creases, and ridge patterns. Based on these statistical models, a handprint synthesis algorithm is proposed, which is useful for technology evaluation due to the lack of high resolution handprint databases in the public domain.

Keywords: Texture synthesis, fingerprint, palmprint, singularity.

1 Introduction

For over one hundred years, handprints (see Fig. 1a) have been routinely used by law enforcement agencies throughout the world to identify suspects and victims [1]. Without question, among all the regions of hands, ridge pattern on fingertips has the highest practical value and has received the most attention. In fact, only fingerprints have found wide application in non-forensic areas. However, recognition solely based on fingertips has some inherent problems when fingerprints are of poor quality, unreliable or unavailable. For example, fingerprints of elderly people and manual workers are often flattened and contain many creases or cuts, criminals may surgically alter their fingerprints to avoid being recognized by fingerprint recognition systems, and latent prints lifted from crime scenes may have been left by lower finger segments or palms.

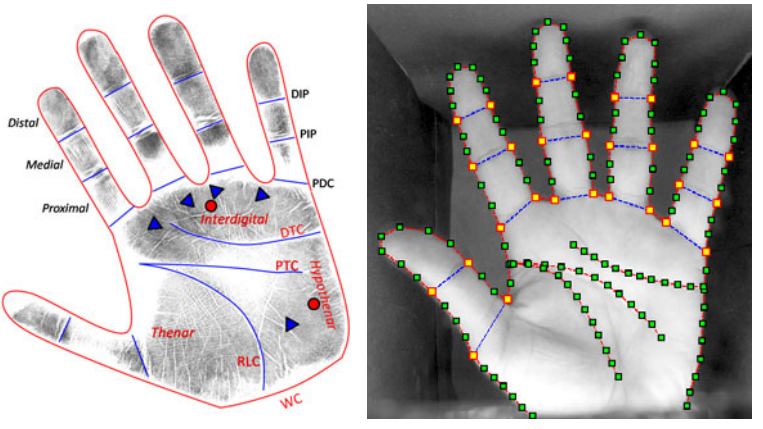
To deal with these problems, a very nature solution is to extend ridge pattern recognition from fingerprints to whole handprints. However, simply applying fingerprint analysis and matching techniques to the whole handprints is not very effective, due to the differences between different regions of handprints. Significant research efforts should be devoted to lower finger segments and the palm, which have received very little attention in the past.

In this paper, we study the probability distributions of main features in handprints, including hand contour, major creases, and singular points. Statistical models of handprints have potential values for a number of problems in handprint recognition, including feature extraction, matching, indexing, and synthesis. Although several researchers have studied the probability distributions of some of the features in handprints, such as hand contour [2] and singularity in fingerprints [3], our statistical model is more comprehensive, covering three main

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 133–140, 2011.

features in all regions of handprints and considering the correlation between different features and between different regions.

We also applied the proposed statistical model to handprint synthesis. Handprint synthesis technique is very desirable because the lack of public domain high resolution handprint databases has hindered the research on handprint recognition. Although there exists some researches that can synthesize some of features or regions of handprints, such as hand shape synthesis [2], low resolution palmprint synthesis [4], and fingerprint synthesis [5], synthesis of high resolution full handprint images has not yet been addressed.



(a) (b)

Fig. 1. (a) A high resolution handprint obtained using ink-on-paper method. Hand contour, major creases, and singular points (loop and delta) are marked. (b) A low resolution handprint and manually marked lankmark points.

2 Statistical Modeling

2.1 Terms

The human hand consists of the palm and five fingers. Both the finger and the palm are divided into regions by permanent flexion creases, which are wide white lines in the handprint image in Fig. 1a. The finger consists of three parts: the distal segment, the medial segment, and the proximal segment, which are separated by the distal interphalangeal crease (DIP), the proximal interphalangeal crease (PIP), and the proximal digital crease (PDC). The palm can be further divided into three areas: interdigital, thenar, and hypothenar, by three major palm flexion creases: radial longitudinal crease (RLC), proximal transverse crease (PTC), and distal transverse crease (DTC). The wrist crease (WC) defines the boundary between the palm and the arm.

2.2 Contour and Creases

Since both hand contour and major palmar creases have regular structure, we employ a statistical shape model [2] to model these two features. By modeling contours and major palmar creases together, we can take into account the correlation between these features and avoid generating invalid handprints. Phalangeal creases are assumed to be straight lines between corresponding landmark points on hand contour.

The first step in statistical shape modeling is to label the training set. As we do not have a large database of high resolution handprint images like the image in Fig. 1a, IIT Delhi Touchless Palmprint Database (`http://web.iitd.ac.in/~ajaykr/Database_Palm.htm`), was used to study the interclass variations of hand contours and major palmar creases. Landmark points of 100 different left hands were manually marked by the authors. The hand contour is represented by 118 landmark points, while each of the three major palmal creases is defined by 12 points uniformly distributed on the crease. Figure 1b shows the annotated landmark points on a low resolution handprint image. The algorithm described in [2] is then used to learn the statistics of these points in the training set.

2.3 Ridge Pattern

Handprint identification in forensics is exclusively based on matching ridge patterns. Thus modeling the ridge patterns on hands is of crucial importance in modeling handprints. Inspired by the method in [5], a ridge pattern $\mathbf{r}$ is viewed as the result of performing iterative contextual filtering, governed by ridge orientation field $\mathbf{o}$ and ridge frequency map $\mathbf{f}$, on a random noise image $\mathbf{n}$. It is defined as $\mathbf{r} = \mathit{Filtering}(\mathbf{n}, \mathbf{o}, \mathbf{f})$. In other words, a ridge pattern $\mathbf{r}$ is a deterministic transformation of three random vectors: noise image $\mathbf{n}$, orientation field $\mathbf{o}$, and frequency map $\mathbf{f}$. The function $\mathit{Filtering}()$ represents the iterative Gabor filtering. The role of the noise image $\mathbf{n}$ is to support the variability of ridge patterns with the same orientation field and frequency map. We determined that an i.i.d. uniform distribution model in the range $[0, 1]$ is suitable for the noise image. Since the variability of ridge frequency map $\mathbf{f}$ is not large, we assume a fixed ridge frequency (0.1 ridges per pixel) for ridge patterns and focus on studying the probability distribution of orientation field $\mathbf{o}$ in this work.

2.4 Orientation Field

We have observed that ridge orientation field on the hand is quite consistent with major creases except for the singularity region. Thus, we model an orientation field $\mathbf{o}$ as $\mathbf{o} = \mathit{Reconstruct}(\mathbf{c}, \mathbf{s})$, namely a deterministic transformation of two random vectors: creases $\mathbf{c}$ and singular points $\mathbf{s}$. Statistical modeling of major creases has already been covered in Sec. 2.2. Statistical modeling of singularity will be discussed in the next subsection. Here we describe how orientation field is reconstructed based on creases and singularity.

Orientation field is reconstructed separately for the palm and each segment of each finger. When reconstructing orientation field for a specific region, only the boundary creases of that region and the singular points in that region are used. The boundaries of the distal segment of finger include the hand contour around the fingertip and the distal interphalangeal crease. The boundaries of the medial segment and the proximal segment of finger includes the three phalangeal creases. The boundaries of the palm contain the proximal interphalangeal creases of five fingers, the wrist crease, and three major palm creases.

Given the singular points (if present), the boundary creases (each represented as an ordered set of sampling points), and the region of interest, the proposed orientation field reconstruction algorithm consists of five steps (see the flowchart in Fig. 2):

1. For each line segment between two consecutive sampling points on each boundary crease, the orientation of the line segment is used as the local ridge orientation at all the blocks passed by the line segment. A sparse orientation field, called boundary orientation field, is generated after this step.
2. The singular orientation field is computed by the Zero-Pole model [6].
3. The singular orientation field is subtracted from the boundary orientation field to obtain the boundary continuous orientation field.
4. A bivariate polynomial model is fitted to the boundary continuous orientation field and used to predict the whole continuous orientation field.
5. The final orientation field is obtained by adding the singular orientation field to the continuous orientation field.

Steps 3 and 5 can be omitted if there are no singular points, for example, in arch-type fingerprints. Note that the same algorithm is used for each region of the hand. The order of polynomial model is three for the palm and two for the finger region.

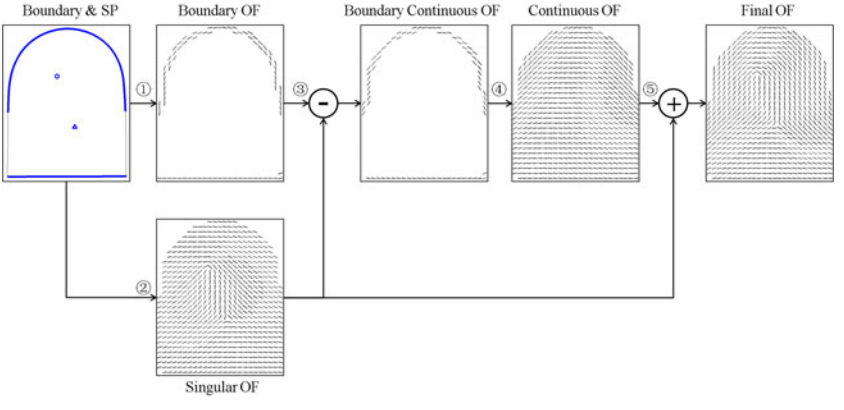


Fig. 2. Orientation field (OF) reconstruction from finger boundary and a pair of singular points (SP)

2.5 Singularity Configuration

A singularity configuration is a set of singularities $s = \{(x_1, y_1, t_1), \cdots, (x_n, y_n, t_n)\}$, where n is the number of singularities, x_i and y_i are the location of the ith singularity, and t_i is the type of the ith singularity. For a loop-type singularity, $t_i = 1$. For a delta-type singularity, $t_i = -1$. Both the number n and the location of singularities are random variables. As the statistical modeling of a random set of points is not convenient, we categorize all possible singularity configurations on the hand into several types according to the number of loops and deltas. In each type of configuration, a fixed order is defined for the singularities. In this way, we only need to study the probability distributions of the locations of singularity of each type of configuration, which is a random vector, $s_t = (x_1, y_1, \cdots, x_{t_n}, y_{t_n})$, where t_n denotes the number of singular points in the configuration of type t.

To reduce the dimensionality of the singularity vector, we chose to study the distribution of singularity configurations in different regions separately. We have observed that singular points generally appear only on the distal segment of the fingers, and the interdigital and hypothenar region of the palm. Since the singular points in these regions are far from each other, we assume that they are independent of each other and study the probability distributions of singularity vectors in each region separately. The number of singular points in a simply-connected region satisfies the Penrose formula [7]: $(N_L - N_D) \cdot \pi = \theta$, where θ denotes the cumulative orientation change along the boundary of the region, and N_L and N_D denote the numbers of loops and deltas inside the region, respectively. For this reason, there are only a small number of feasible singularity configurations in fingers and palms. The details of singularity configurations in three regions are given below.

Distal Segment of Finger. Since the cumulative orientation change along the boundary of a complete fingerprint is always zero, $N_L = N_D$ holds for all fingerprints. Since the probability that a fingerprint contains more than two loops is very rare, we can categorize singularity configurations on fingerprints into three major types: no singularity, one pair of loop and delta points, and two pairs of loop and delta points. The order of singular points is fixed for each of the two types of configurations. In the case of one pair of loop and delta points, the order of singular points is loop and then delta. In the case of two pairs of loop and delta points, the order is top loop, bottom loop, left delta, and then right delta.

The distribution of singularity on fingerprints has been quantitatively studied by Cappelli and Maltoni [3]. But the method used here has two differences from [3]. The first difference is that we use the distal interphalangeal crease as the x axis and the center point of the crease as the origin. This coordinate system is more robust than the coordinate system in [3], which is based on the left and right boundaries of fingerprints and the centroid of fingerprints. The second difference is that we unify ulnar loop, radial loop, and tented arch into one type, since a finer classification is ambiguous in some situations. Singular points and

creases of 200 plain fingerprints in NIST SD29 were manually marked and the p.d.f. of two types of singularity vectors are estimated with a Gaussian mixture model.

Interdigital Region. The number of loops, N_L, and the number of deltas, N_D, in the interdigital region satisfy the formula $N_D - N_L = 3$. We noticed that two types of singularity configurations are very common in our palmprint database, which is provided by a law enforcement agency. The most common type consists of four deltas, each associated with a finger, and one loop. The second most common type consists of five deltas and two loops. Other types of configurations are not considered now since sufficient data is not available for estimating their probability distribution. The order of singular points is fixed for each of the two types of configurations. In the first type of configuration, the order is the loop, and then the four delta points sorted along the ulnar direction (from little to ring finger). In the second type of configuration, the order is the two loop points sorted along the ulnar direction, four delta points associated with fingers sorted along the ulnar direction, followed by the additional delta.

The top endpoint of crease PTC and the bottom endpoint of crease DTC are used as reference points to register palmprints. Singular points and reference points of 2,000 palmprints in our database were manually marked for studying the distribution of singularity on the palm. Gaussian mixture models are used to estimate the p.d.f. of the two types of singularity configurations.

Hypothenar Region. The number of loops, N_L, and the number of deltas, N_D, in the interdigital region satisfy the formula $N_D - N_L = 1$. The most common singularity configuration consists of only one delta, the so-called carpal delta, which is generally located between thenar and hypothenar and just above the wrist crease. The second most common type consists of two deltas and one loop. Other types of configurations are not considered since sufficient data is not available for estimating their probability distribution. The order of singular points is fixed for each of the two types of configurations. In the second type of configuration, the order is the loop followed by the two delta points sorted from top to bottom. Gaussian mixture models are used to estimate the p.d.f. of the two types of singularity configurations.

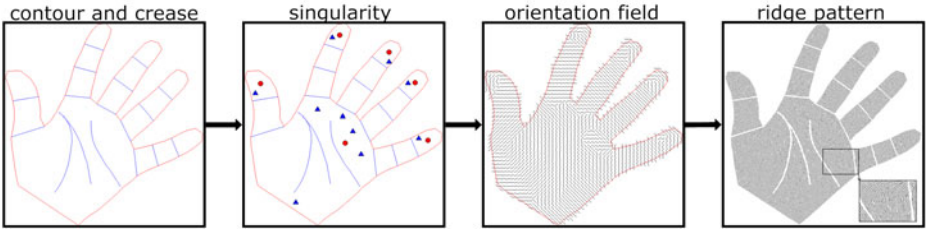


Fig. 3. Flowchart of the proposed handprint synthesis algorithm

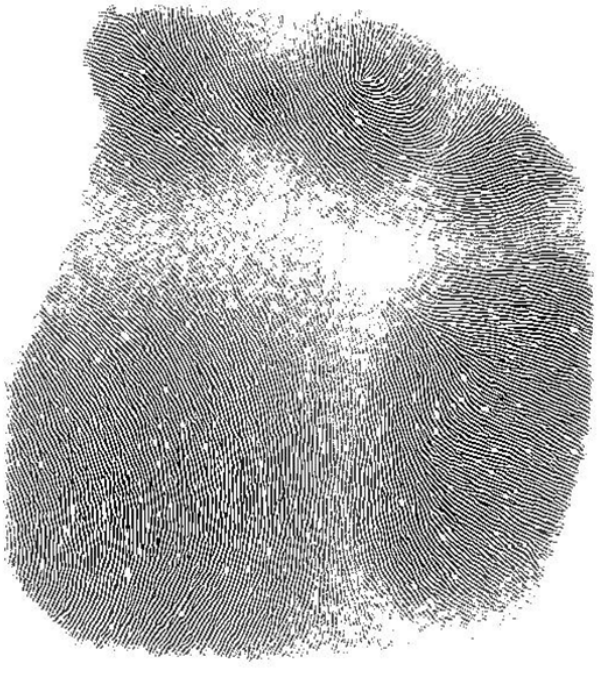


Fig. 4. Palm region of a synthesized handprint with noise added

3 Synthesis

With the statistical models described in the preceding sections, we generate a handprint image through the following four steps (see the flowchart in Fig. 3):

1. Contour and major crease generation. Landmark points on hand contour and major palmar creases are generated by sampling their statistical model. Hand contour and palmar creases are obtained by connecting landmark points. Phalangeal creases are generated by connecting the corresponding landmark points on the hand contour.
2. Singular point generation. Singular points are generated for each finger and the palm region separately. The first step is to sample the singularity configuration. The second step is to generate a random vector according to the statistical model of the selected configuration. The last step is to transform the singularities using the finger or palm coordinate system.
3. Orientation field reconstruction. For each finger segment and the palm, orientation field is reconstructed using the singular points in the region and the boundary.
4. Ridge pattern generation. Ridge pattern is generated by performing iterative Gabor filtering (5 iterations) on a randomly seeded image. Finally, major creases are drawn on the ridge pattern as wide white lines.

The synthesized handprint can be made more realistic by adding noise using the method of SFinGe [5]. See Fig. 4 for the palm area of a synthesized handprint with noise added.

4 Summary and Future Directions

Fingerprint recognition systems do not function well when fingerprints are of poor quality or not available. A natural solution to overcome these problems is to develop recognition systems to utilize the entire handprint. To understand the handprints, we have conducted a comprehensive study on the probability distribution of the main features in handprints, including hand contour, major creases, and ridge pattern. For different features, appropriate statistical models are proposed and estimated using large databases of training images. Based on these statistical models, we have proposed a handprint synthesis algorithm that can generate high resolution synthetic handprint images.

To generate more realistic handprint images, we need to model the probability distribution of minor creases and the intraclass variations in several aspects, such as contact area, distortion, brightness, contrast, and ridge width. We will also conduct experiments to see whether the performance of handprint recognition algorithms on synthesized handprint datasets is consistent with their performance on real handprint datasets.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under Grants 61020106004, 60875017, 61005023, and 61021063.

References

1. Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: Handbook of Fingerprint Recognition. Springer, Heidelberg (2003)
2. Cootes, T.F., Taylor, C.J., Cooper, D.H.: Active Shape Models—Their Training and Application. CVIU 61(1), 38–59 (1995)
3. Cappelli, R., Maltoni, D.: On the Spatial Distribution of Fingerprint Singularities. IEEE Trans. on PAMI. 31(4), 742–748 (2009)
4. Wei, Z., Han, Y., Sun, Z., Tan, T.: Palmprint Image Synthesis: A Preliminary Study. In: ICIP, pp. 285–288 (2008)
5. Cappelli, R., Maio, D., Maltoni, D.: Synthetic Fingerprint-Database Generation. In: ICPR, pp. 744–747 (2002)
6. Sherlock, B.G., Monro, D.M.: A Model for Interpreting Fingerprint Topology. Pattern Recognition 26(7), 1047–1055 (1993)
7. Penrose, R.: The Topology of Ridge Systems. Annals of Human Genetics 42(4), 435–444 (1979)

A Survey of On-line Signature Verification

Zhaoxiang Zhang, Kaiyue Wang, and Yunhong Wang

Laboratory of Intelligent Recognition and Image Processing,
Beijing Key Laboratory of Digital Media,
School of Computer Science and Engineering, Beihang University, Beijing, China
{zxzhang,yhwang}@buaa.edu.cn, wangkkyy@163.com

Abstract. Signature is a commonly accepted biometric feature for individual identification. On-line signature has begun to prevail in the last decades for it exploits dynamic features which traditional off-line signature fails to preserve. This paper presents a review of researches on on-line signature verification during recent years and lists some of the works that provide promising results.

Keywords: Signature Recognition, Biometrics, Online.

1 Introduction

Handwritten signature has been long served as one of the most widespread biometric traits for human identity verification. It is a personal stylized depiction of someones name, nickname or some other characters that a person writes down as a proof of identity and intent, which is frequently used in daily activities, such as check validation, physical entry to protected areas, and financial and legal documents. The formation of signatures is influenced by physical condition, family environment, education, and many other factors, thus signatures are assumed to be unique, even though there has been no study proving this point [1], and are difficult to imitate.

A traditional signature is written on a document and verified be experts or other experienced verifiers. With the development of computer science and technology, signatures written on document can be scanned into computers and verified by specialized verification systems, thus to achieve higher accuracy and efficiency. However, this type of signatures, usually referred as off-line signatures, only take into account their static features, which can be easily impersonated if the forger is accessed to some samples of genuine signatures. Technology of on-line signature has emerged during late 1900s to cope with the drawbacks of off-line signature. On-line signatures are is composed of a series of sample points, which contains of several extracted features.

On-line signatures are considered more reliable than off-line signatures, for they take into account dynamic features, such as pressure and velocity of pen-tip, which are more complex and more difficult to imitate. A typical on-line signature verification system is made up by consecutive phases of data acquisition, preprocessing, feature extraction, training and verification, as shown in Fig. 2. In the following, we will describe each of the modular in detail.

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 141–149, 2011.

Fig. 1. A traditional signature and an on-line signature

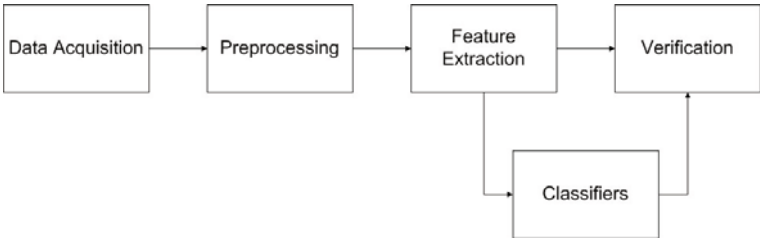


Fig. 2. Framework of on-line signature verification system

2 Data Acquisition

The most commonly used data acquisition device is digitizing tablets (Fig.3). D. J. Hamilton et al proposed one of the earliest viable low cost digitizing tablet in 1995 [2]. However, the signing process using digitizing tablets is far from natural, for the tablet is quite different from paper, which people usually sign on, and more importantly, the signer cannot see what he has written immediately, which may cause inconvenience. Touch-sensitive screen turns out to be a solution to this problem. This technology is extensively used in devices such as tablet PC and personal digital assistant (PDA), and its performance in on-line signature verification has been studied [3]. A different type of device is electronic pen (Fig.3), which detects pen motion, velocity, inclination, and other properties with the electronic components built in it. One of the earliest study of signature verification using electronic pen may be presented by H. D. Crane and J. S. Ostrem in 1983 [4]. Many adjustments and improvements have been made to exploit electronic pen's use in signature verification, such as in[5]. Recently, N. S. Kamel et al proposed a glove-based signature verification system [6]. This system employed data gloves (Fig.3) to record the change of hand shape during the signing process, which provided more dynamic information than traditional systems using digitizing tablets or electronic pens.

3 Preprocessing

Using different data acquisition devices, the data structure of the raw data may be various. However, with certain transformations, an on-line signature can

Fig. 3. Digitizing tablet, electronic pen, and data gloves

always be represented as a sequence $s_1, s_2, ..., s_n$, where s_i is the signal value sampled at the i-th sample point. Furthermore, s_i may contain values of different types of features, according to the acquisition device employed, such as pen-tip position, pressure, and pen inclination. Preprocessing of these data influences the successive phases of signature verification. Typically, there are three steps in the preprocessing phase; however, not all of them are required.

Firstly, signatures should be smoothed and resampled. In some works, smoothing is referred as filtering or noise reduction; however, we do not strictly distinguish them as it is in context of signal processing. One of the common methods of smoothing is based on Gaussian filters, which eliminates time dependencies and preserves spatial features of signatures [1]. Meanwhile, resampling could be done through interpolation.

Secondly, signatures are normalized to eliminate the variations in both position and scale. This can be achieved simply by transforming the absolute values of position into relative values. To normalize the orientation of signatures, Fourier transform and its inverse can be employed [7].

Finally, signatures could be partitioned into segments. Although this step has heavy impact on a signature verification system, it is not necessary, and in fact, many works do not include the partitioning step. Still, some researches have been done in signature segmentation, such as component-oriented segmentation [8], variable length segmentation [9], geometric extrema segmentation [10], and segmentation based on perceptually important points [11].

4 Feature Extraction

Original features extracted by acquisition devices (Table 1) can be calculated into more specific features. These features are usually classified into two categories: local features, where one feature is extracted for each sample point in the input domain, global features, where one feature is extracted for a whole signature, based on all sample points in the input domain [12]. Generally, local features are assumed to have better performances as well as higher computational cost, because they exploit specific details of signatures.

J. Richiardi et al looked into 46 global features and 39 local features [12]. They used a modified Fisher ratio to measure the class separation ability of differet feature sets and selected 12 best global features and 12 best local features. H. Lei et al studied the consistency of 22 features, including both global and local

Table 1. Dynamic features captured by a typical digitizing tablet

X-coordinate	Pen-tip position along the x-axis of tablet
Y-coordinate	Pen-tip position along the y-axis of tablet
Time stamp	System time at which the event was posted
Button status	Current button status (pen-up or pen-down)
Azimuth	Clockwise rotation of cursor about the z-axis
Altitude	Angle upward toward the positive z-axis
Pressure	Writing pressure of the pen-tip

ones [13]. Based on their observations, they reached conclusions such as speed of X-coordinates and speed of Y-coordinates are two of the most consistent features; azimuth and altitude have relatively high consistency, yet have high standard deviations; some other features, like acceleration and pressure, are of low consistency. Julian F.-Aguilar et al studied 100 global features and sorted them by individual discriminative power [14].

More complex features can be computed from simple features and new representations of on-line signatures can be obtained. Y. Chen et al used a sequence, each item of which describes the moving direction of the pen between two consecutive sample points, to represent a signature [15]. D. Z. Lejtman et al looked into the wavelet features of on-line signatures [16] and H.-W. Ji et al further studied the use of zero-crossing representation of wavelets [17]. H. Deng et al studied the spatio-temporal correlation of sample points, and converted the spatio-temporal matrix into grey-level image to represent the signature [18].

5 Training and Verification

Having obtained certain types of features, either simple features or complex ones, classifiers need to be trained in order to verify given signatures. Most methods of pattern recognition can be applied here.

Dynamic time warping (DTW) is one of the most commonly used and best performing approaches in on-line signature verification. DTW compares two sequence of different lengths using dynamic programming, giving the minimum of a given distance value. Since two signatures usually vary in length, DTW turns out to be quite suitable for the task of signature verification. DTW is always combined with other methods to improve the performance. In the First International Signature Verification Competition [19], a DTW-based principal component analysis (PCA) method won the first place [20]. Aside of PCA, minor component analysis (MCA) is also combined with DTW for on-line signature verification [21]. Y. Chen et al represented signatures as sequences and used DTW for sequence matching [15]. T. A. Osman et al first employed DTW to partition signatures, and then adopted multivariate autoregressive model to extract features of signatures [22]. Besides combination with other techniques,

there has been some researches in improve DTW to fit the task of on-line signature verification. Instead of warping the whole signature, H. Feng et al tried to warp a selected set of extreme points to be more adaptive [23]. L. Hu proposed enhanced DTW, which enhanced the separability between genuine and forged signatures [24].

Hidden Markov model (HMM) is another pervasively used pattern recognition approach. HMM is a double stochastic approach in which an unobserved process is estimated through a sequence of observed states. It seems to be suitable for on-line signature verification since it is highly adaptable to personal variability [25]. The topologies of HMM frequently resorted to include left-to-right, ergodic and ring. Left-to-right is the most commonly adopted topology in signature verification, such as in [7] [26]. The study of ergodic topology can be found in [27]. Same as DTW, HMM is also combined with other techniques to improve the performance. N. Mohankrishnan et al introduced the combination of HMM and autoregressive models [28]. M. Fuentes et al adopted both HMM and multi-layer perceptron (MLP) neural network, and used support vector machine for fusion [29].

Support vector machine (SVM) is another effective approach for data separation. SVM maps vectors in a low dimensional space where they cannot be directly separated into a higher dimensional space where they can. The separating hyperplane is then mapped back into the original space as the decision surface. A. Kholmatove et al studied the use of SVM, comparing to PCA and Bayesian decision method [20]. M. Fuentes et al used SVM to fuse HMM and MLP [29]. C. Gruber proposed a new kernel for SVM based on longest common subsequence for on-line signature verification [30].

6 Signature Databases and Experiment Protocols

To compare the results in different works, a large public signature database and commonly accepted experiment protocols are required. Many efforts have been made on these issues.

One of the most widely used signature databases is MCYT database, provided by Biometric Recognition Group - ATVS at Escuela Politecnica Superior of the Universidad Autonoma de Madrid [31]. MCYT databse contains both fingerprints and signatures. The MCYT-Signature subcorpus consists of online signature samples of 330 individuals. For each individual, there are 25 genuine signatures and 25 are forgeries, which were all enrolled using WACOM pen tablet, model Intuos A6 USB. In total, the database contains 8,250 genuine and 8,250 forged on-line signatures. Each signature contains information of position in x-axis, position in y-axis, pressure, azimuth angle and altitude angle.

Another frequently employed database is SVC2004 on-line signature database, which was designed for the First International Signature Verification Competition [19]. It contains 20 genuine signatures and 20 forgeries for each of the 100 signers, that is 4,000 signatures in sum, collected using WACOM Intuos tablet. However, only 40 sets of genuine and forged signatures are publicly available.

Table 2. Part of results from the Competition, against skilled forgeries

Team ID	Avg EER	SD EER	Max EER	Min EER
219b	6.90%	9.45%	50.00%	0.00%
219c	6.91%	9.42%	50.00%	0.00%
206	6.96%	11.76%	65.00%	0.00%
229	7.64%	12.62%	60.00%	0.00%
219a	8.90%	11.72%	71.00%	0.00%
214	11.29%	13.47%	70.00%	0.00%
218	15.36%	13.88%	60.00%	0.00%
217	19.00%	14.43%	70.00%	0.00%
203	20.01%	18.44%	76.19%	0.00%
204	21.89%	17.05%	73.33%	0.00%

Signatures in this database are not real names of the signers, but a new signature, either in Chinese or English, specifically practiced for the enrollment. It is claimed that "although most of the data contributors are Chinese, many of them actually use English signatures frequently in daily applications." The features of each signature is the same as in MCYT.

SUSIG database is established by Biometrics Research Group, Sabanci University. It consists of two parts, namely Visual and Blind sub-corpora. Visual sub-corpus was collected using Interlink Elec. ePad Ink signature tablet with built-in LCD screen. For each subject there are 20 genuine and 10 forgery signatures. Genuine signatures were collected in two different sessions. Blind subcorpus was collected using Wacom Graphire2 pressure sensitive tablet. For each subject there are 10 genuine and 10 forgery signatures. Genuine signatures were collected in a single s session. Signature data consists of x and y coordinates, time stamp, pressure level and a pen up or down indicator [32]. It is worth noting that the acquisition of Visual sub-corpus is different from the databases above, for the signers sees what they have written immediately on the LCD screen of the tablet.

Apart from a public database, commonly accepted test protocols is also needed. The protocols in the First International Signature Verification Competition have been adopted in many other works. For each user, all genuine signatures are divided into two subsets. Five genuine signatures are randomly selected from the first subset to be the training set. The second subset of genuine signatures and all forgeries of this user are used as testing samples. Receiver operating characteristic (ROC) curves can be drawn by changing a certain threshold, and equal error rates (EER) can be subsequently computed (Table 2). Should random forgeries be taken into account, they can be randomly selected from signatures of other users.

7 Conclusions

This paper reviews the recent developments of on-line signature verification and summarizes representative works in this field. Digitizing tables, electronic pens, and more recently, data gloves are employed to acquire data, which are later smoothed, normalized and segmented. Dynamic features, such as pressure, azimuth and altitude can be extracted from the preprocessed signatures. More complex features such as wavelet features and correlation images can be computed from simple ones. Based on these features, pattern recognition, like DTW, HMM and SVM can be adopted to fulfil the task of on-line signature verification. Public databases, such as MCYT, SVC2004 and SUSIG, and commonly accepted protocols are introduced. Signatures are pervasively used in daily life, so on-line signature verification has been and will still be a hot topic in the filed of pattern recognition. Since signatures are produced by a complex process, which is sensitive to the psychological state and external conditions, yet the training samples are quite limited, signature verification will remain a challenging problem in the near future.

Acknowledgement. This work is funded by the National Basic Research Program of China (No. 2010CB327902), the National Natural Science Foundation of China (No. 60873158, No. 61005016, No. 61061130560) and the Fundamental Research Funds for the Central Universities.

References

1. Jain, A.K., Griess, F.D., Connell, S.D.: On-line signature verification. Pattern Recognition 35(12), 2963–2972 (2002)
2. Hamilton, D.J., Whelan, J., McLaren, A., Macintyre, I., Tizzard, A.: Low cost dynamic signature verification system. In: Proc. Eur. Convention Secur. Detection
3. Alonso-Fernandez, F., Fierrez-Aguilar, J., Ortega-Garcia, J.: Sensor interoperability and fusion in signature verification: A case study using table pc. In: Proc. Int. Workshop Biometric Recognit. Syst.
4. Crane, H.D., Ostrem, J.S.: Automatic signature verification using a three-axis force-sensitive pen. IEEE Trans. Syst. Man, Cybern. 13(3), 329–337 (1983)
5. Zhukov, A., Vaqquez, M., Garcia-Beneytez: Magnetoelastic sensor for signature identification based on mechanomagnetic effect in amorphous wires. J. Phys. 4(8), 763–766 (1998)
6. Kamel, N.S., Sayeed, S., Ellis, G.A.: Glove-based approach to online signature verification. IEEE Trans. Pattern Anal. Mach. Intell 30(6), 1109–1113 (2008)
7. Kashi, R.S., Hu, J., Nelson, W.L.: On-line handwritten signature verification using hidden markov model features. In: Proc. 4th Int. Conf. Doc. Anal. Recognit., vol. 1, pp. 253–257 (1997)
8. Dimauro, G., Impedovo, S., Pirlo, G.: Component-oriented algorithms for signature verification. Int. J. Pattern Recognit. Artif. Intell. 8(3), 771–794 (1994)
9. Shafiei, M.M., Rabiee, H.R.:
10. Lee, J., Yoon, H.-S., Soh, J., Chun, B.T., Chung, Y.K.: Using geometric extrema for segment-to-segment characteristics comparison in online signature verification

11. Brault, J.-J., Plamondon, R.: Segmenting handwritten signatures at their perceptually important points. IEEE Trans. Pattern Anal. Mach. Intell. 15(9), 953–957 (1993)
12. Richiardi, J., Ketabdar, H.: Local and global feature selection for on-line signature verification. In: Proceeding of the 8th International Conference on Document Analysis and Recognition, Seoul, Korea, pp. 625–629 (2005)
13. Lei, H., Govindaraju, V.: A comparative study on the consistency of features in on-line signature verification. Pattern Recognit. Lett. 26, 2483–2489 (2005)
14. Fierrez-Aguilar, J., Nanni, L., Lopez-Penalba, J., Ortega-Garcia, J., Maltoni, D.: An on-line signature verification system based on fusion of local and global information. In: Audio- and Video-based Biometric Person Authentication (2005)
15. Chen, Y., Ding, X.: On-line signature verification using direction sequence string matching. In: Proc. SPIE, vol. 4875, pp. 744–749 (2002)
16. Lejtman, D.Z., George, S.E.: On-line handwritten signature verification using wavelets and back-propagation neural networks. In: Proc. 6th Int. Conf. Doc. Anal. Recognit., pp. 992–996 (2001)
17. Ji, H.-W., Quan, Z.-H.: Signature verification using wavelet transform and support vector machine. In: Proc. Int. Conf. Intell. Comput. (2005)
18. Deng, H.R., Wang, Y.H.: On-line signature verification based on spatio-temporal correlation. In: Huang, D.-S., Jo, K.-H., Lee, H.-H., Kang, H.-J., Bevilacqua, V. (eds.) ICIC 2009. LNCS, vol. 5754, pp. 75–81. Springer, Heidelberg (2009)
19. Yeung, D.-Y., Chang, H., Xiong, Y., George, S.E., Kashi, R.S., Matsumoto, T., Rigoll, G.: SVC2004: First International Signature Verification Competition. In: Zhang, D., Jain, A.K. (eds.) ICBA 2004. LNCS, vol. 3072, pp. 16–22. Springer, Heidelberg (2004)
20. Kholmatov, A., Yanikoglu, B.: Identity authentication using improved online signature verification method. Pattern Recognition Letters, pp. 2400–2408 (2005)
21. Li, B., Wang, K., Zhang, D.: On-line signature verification based on pca (principal component analysis) and mca (minor component analysis). In: Zhang, D., Jain, A.K. (eds.) ICBA 2004. LNCS, vol. 3072, pp. 540–546. Springer, Heidelberg (2004)
22. Osman, T.A., Paulik, M.j., Krishnan, M.: An online signature verification system based on multivariate autoregressive modeling and dtw segmentation. In: IEEE Workshop on Signal Processing Applications for Public Security and Forensics (2007)
23. Feng, H., Wah, C.C.: Online signature verification using a new extreme points warping technique. Pattern Recognit. Lett. 24
24. Hu, L., Wang, Y.-H.: On-line signature verification based on fusion of global and local information. In: Proceedings of International Conference on Wavelet Analysis and Pattern Recognition (2007)
25. Plamondon, D., Pirlo, G.: Automatic signature verification: The state of the art. Systems, Man, and Cybernetics,Part C: Applications and Reviews 38(5), 609–635 (2008)
26. Igarza, J.J., Goirizelaia, I., Espinosa, K., Hernaez, I., Mendez, R., Sanchez, J.: Online handwritten signature verification using hidden markov models. In: Sanfeliu, A., Ruiz-Shulcloper, J. (eds.) CIARP 2003. LNCS, vol. 2905, pp. 391–399. Springer, Heidelberg (2003)
27. Quan, Z.-H., Liu, K.-H.: Online signature verification based on the hybrid hmm/ann model. Int. J. Comput. Sci. Netw. Secur. 7(3), 313–321 (2007)
28. Mohankrishnan, N., Paulik, M.J., Khalil, M.: On-line signature verification using a nonstationary autoregressive model representation. In: Proc. IEEE Int. Symp. Circuits Syst., pp. 2303–2306 (1993)

29. Fuentes, M., Garcia-Salicetti, S., Dorizzi, B.: On-line signature verification: Fusion of a hidden markov model and a neural network via a support vector machine. In: Proc. 8th Int. Workshop Front. Handwriting Recognit., pp. 253–258 (2002)
30. Gruber, C., Gruber, T., Krinninger, S., Sick, B.: Online signature verification with support vector machines based on lcss kernel functions. In: IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, pp. 1088–1100 (2010)
31. Ortega-Garcia, J., Fierrez-Aguilar, J., Simon, D., Gonzalez, J., Faundez, M., Espinosa, V., Satue, A., Hernaez, I., Igarza, J.-J., Vivaracho, C., Escudero, D., Moro, Q.-I.: Mcyt baseline corpus: A bimodal biometric database. Inst. Elect. Eng. Proc. Vis., Image Signal Process., Special Issue Biometrics Internet 150(6), 395–401 (2003)
32. Kholmatov, A., Yanikoglu, B.: Susig: an on-line signature database, associated protocols and benchmark results. Pattern Anal Applic. 12, 227–236 (2009)

A Survey of Advances in Biometric Gait Recognition

Zhaoxiang Zhang, Maodi Hu, and Yunhong Wang

Laboratory of Intelligent Recognition and Image Processing,
Beijing Key Laboratory of Digital Media,
School of Computer Science and Engineering, Beihang University, Beijing, China
{zxzhang,yhwang}@buaa.edu.cn, Londeehu@gmail.com

Abstract. Biometric gait analysis is to acquire biometric information such as identity, gender, ethnicity and age from people walking patterns. In the walking process, the human body shows regular periodic motion, especially upper and lower limbs, which reflects the individual's unique movement pattern. Compared to other biometrics, gait can be obtained from distance and is difficult to hide and camouflage. During the past ten years, gait has been a hot topic in computer vision with great progress achieved. In this paper, we give a general review and a simple survey of recent gait progresses.

Keywords: Gait Analysis, Biometric, Recognition.

1 Introduction

In the past ten years, terrorist attacks such as London subway bombings occur frequently, which make people clearly understand the importance of security monitoring and control in national defense and public safety. A large number of surveillance cameras have been installed in public places, but these security-sensitive early warning systems require intelligent approaches. Ideal intelligent monitoring system should be able to automatically analyze the collected video data, give out a early warning before the adverse event happens, and reduce injury and economic loss. For example, when the system detects abnormal behavior, it can immediately determine the identities of all persons in the scene, rapidly investigate their previous activities, and track the suspects across the regions. It requires the monitoring system can not only estimate the quantity, location and behavior, but also obtain the identity information.

Gait is the most suitable biometrics in the case of intelligent visual surveillance. In monitoring scenes, people are usually distant from cameras, which makes most of biometric features no longer available. Most of existing systems use face for identification. The shortcomings are obvious, for example, unexpected view angle and occlusion cause full faces can not be photographed, distance brings about low-resolution face image. Therefore, face can not always achieve acceptable results in practical. In contrast, gait is a behavioral biometric, including not only individual appearance, such as height, leg length, shoulder

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 150–158, 2011.

width, but also the dynamics of individual walking. Compared with other biometrics, gait is remote accessed and difficult to imitate or camouflage. Moreover, the capturing process does not require cooperation, contact with special equipment, or high image resolution.

Because of the urgent need for intelligent monitoring and the development of computer-related fields, video-based gait recognition research has gradually emerged since 1990s. Back in the 1960s, M. P. Murray et al. [1,2] proved gait is a recognizable pattern of cyclical movement in medical experiments, and did a preliminary analysis of the impact of height, age and other factors on identity. H. J. Ralston et al. [3] decomposed the gait cycle into detailed synthetic movement of multiple joints and muscle, in which the parameters of factors include body weight, limb length, joint velocity, and bone structure. They pointed out the uniqueness of gait pattern, and did a performance evaluation of biometric gait. In the early psychological research on gait recognition, most work is based on the observation experiments. G. Johansson et al. [4] and C. D. Barclay et al. [5] equipped reflectors and moving lights in several joints, and made observations can not directly see pedestrians but only see these light points. Results are consistent with physiological measurements, the majority of observers are able to recognize their familiar friends with limited points of light. These experiments above proved that the gait pattern is personally unique, and can be used for biometric recognition.

2 Overview

By data source, the gait studies can be divided into two main categories, which are sensor-based and video-based.

In the studies of sensor-based gait, common used sensor devices are mainly tactile ones and wearable ones. Tactile sensors generally refers to multi-degree of freedom (angle) pressure sensor [6], usually installed in a particular road, to collect the pressure signal generated when walking. And wearable sensors [7] are attached to the key points of different body parts, selective collect the speed, acceleration, position and other information. Commonly used devices include light senses (such as reflectors, moving lights), acceleration sensors, magnetic sensors, gyroscopes, etc. Sensors can directly access to the motion information of specified parts. Although the sensor-based data can easily assure accuracy, but requires more complex equipment in collection. P. Vanitchatchavan [8] analyzed the angle pattern between joints. Three goniometers pasted at the pelvis, knee, and ankle joints, recording the joint change information during walking and stoping. C. D. Barclay et al. [9] and JECutting et al. [10] proposed the use of shoulder width, hip width, and other body parameters in identification and gender classification, and analyzed the recognition accuracy and feasibility. They become the cornerstone of a large number of of follow-up studies. However, most applications of these methods are limited in medical researches. They have been frequently used in the health status of diagnosis, such as Parkinson's disease diagnosis and feedback of treatment [11].

Video-based gait recognition research generally refers to shot through the optical camera to get the video and identify biometric information, but not rely on special equipment. Currently, widely used large gait databases in academic research include CASIA Gait Database (Dataset B) [12], collected by National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, CMU Motion of Body Database [13], collected by Robotics Institute, Carnegie Mellon University, Southampton Human ID at a Distance Database [14], collected by Information: Signals, Images and Systems Research Group, School of Electronics and Computer Science, University of Southampton, and USF HumanID Gait Baseline Database [15], collected by Computer Vision and Pattern Recognition Group, Department of Computer Science and Engineering, University of South Florida. Even though some covariances such as viewing angle change, shoe type change, and carrying condition change etc. have been considered in these large databases, they were built for human identification. The obvious quantity difference between males and females can impact the performance of gender classification. Therefore, we have collected the BUAA-IRIP Gait Database [16]. The camera layout is shown as Figure 1. During the course of data collection, every participant was asked to walk along the straight line between camera C_1 and C_8, which are denoted by two black points in Figure 1, from left to right and then return, repeating five times. Thus, every camera recorded five left-to-right and five right-to-left walking video sequences for each person. Meanwhile, we label camera C_1 with the 0° view, C_2 with the 30° view, till C_8 with the 180° view. Camera C_4 and C_5 have the same view angle. Camera C9 records human face. Based on these databases, many related researches have been published, and most of them can accurately recognize the identity, gender, age, race, and other information in controlled environment and walking style. Related to the challenging factors, such as view angle, clothing, and walking speed, there are also some preliminary work. However, existing approaches are far from perfect. For example, it is difficult to track the pedestrian and extract gait sequences in the crowd, and gait feature may be extremely inaccurate if camera shakes or weather dramatic changes.

3 Recent Progresses

Given several image sequences capturing human walking, the main researches of gait recognition lie in the analysis of spatial feature and temporal feature.

3.1 Spatial Feature

This section describes the extraction of the body profile, joint position, or other information from videos, which has been used to represent the static state of body movement in each frame.

Two-Dimensional Model. Yang Ran et al. [17] Combined edge detection and Hough Transform to extract the main leg angle, and categorized each frame into

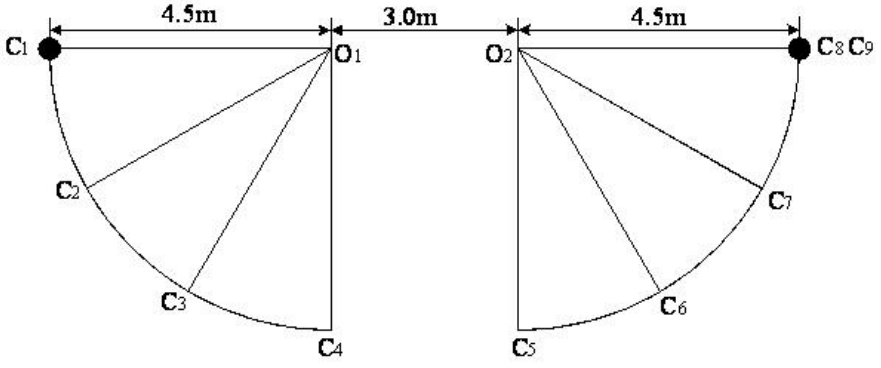


Fig. 1. Camera layout

positive and negative detection of Principle Gait Angle with Bayesian classifier. Frederic Jean et al. [18] proposed an efficient and promising features, which is the trajectories of the head and feet. The difficulties lie in the occlusion and other related issues caused by foot movement. They estimated the separation between feet in each frame, and determined whether the front foot changes. And then they keep on tracking in each half gait cycle by optical flow.

Three-Dimensional Model. Junxia Gu et al. [19] proposed a method to automatically extract key points and pose parameters from the sequence of label-free three-dimensional volume data, and then estimated the multiple configuration (combination of joints) and movement features (position, orientation, and height of the body). A Hidden Markov Model (HMM) and an Exemplar-based HMM are then used to model the movement features and configuration features respectively. Based on these two features, actions are classified by a hierarchical classifier with sum and MAP (Maximum A Posteriori) rules. And identities are recognized from their gait sequences with the configuration features.

Model Free. Human body silhouette is the most frequently used initial feature, which can be easily obtained from background substraction. N.V.Boulgouris et al. [20] applied Radon transform on silhouette to extract the feature of each frame, and employed Linear Discriminant Analysis (LDA) to reduce the dimension of accumulated feature vector of one period. Shiqi Yu et al. [21] conducted psychology experiments on gait-based gender classification, and proposed a automatic gender classification approach based on the weighted sum of block features. Experiments show that human can recognize the gender by using the human gait, and upper and lower halves of the body have different contributions. The averaged body silhouette is divided into five parts, which are head, chest, back, hip, and leg. Support Vector Machine (SVM) is employed to train the classification weights of all the parts. The prior knowledge is combined with the automatic weighting method to improve the classification accuracy of psychology experiments. Maodi Hu et al. [22] took clothing and carrying conditions

into account. They adopted Gabor filters to decompose body shape into local orientations and scales, and obtain low dimensional discriminative representation through the agency of Principal Components Analysis (PCA) and Maximization of Mutual Information (MMI). Gender related Gaussian Mixture Model-Hidden Markov Models are trained for classification and achieve the state of the art accuracy. Ibrahim Venkat et al. [23] also divided the averaged silhouette into several overlapped parts, including upper, middle, and lower parts as well as left and right parts. They trained a Bayesian network to evaluate the impact of these parts on identification, and achieved promising accuracy with backpack pedestrians. Outer contour is the outer boundary of human body silhouette. Kyung Su Kwon et al. [24] combined geodesic active contour models (GACMs) with mean-shift algorithm to extract and track human shape for gait recognition. Optical flow is the velocity field associated with image changes. Khalid Bashir et al. [25] divided flow filed into four parts in accordance with the direction and symbol, and use the weighted sum of these parts for gait recognition.

Fig. 2. Contrast enhanced images in gait sequences of CASIA Gait Database (Dataset B) [12]. Three samples from left to right show the gait patterns of normal walking, clothing and carrying condition changes.

3.2 Temporal Feature

Spatial feature can be used in conjunction with time series analysis methods, such as HMM, Autoregressive Moving Average Model (ARMA), etc. for dynamics modeling and recognition.

Periodic Feature. Yang Ran et al. [17] presented two methods of period estimation and used them for pedestrian detection. The first employ Fourier transform and periodogram to efficiently estimate gait frequency. The other marked cyclic pattern as a binary sequence by using Maximal Principal Gait Angle (MPGA) fitting. And cycle characteristics is expressed by a Phase-Locked Loop (PLL), whose operation is based on the detection of the phase difference between the input and output signals of a Voltage Controlled Oscillator (VCO).

Meng-Fen Ho et al. [26] estimated gait cycle by using the cyclical swing. For each gait cycle, static information is extracted by analyzing the motion vector histogram, and dynamic information is extracted by Fourier descriptors. They used PCA and multiple discriminant analysis (MDA) projection to represent individual characteristics in a low-dimensional space, and trained a nearest neighbor classifier for identification. Changyou Chen et al. [27] proposed a tensor-based Riemannian manifold distance-approximating projection (TRIMAP), which is a two-stage projection method. It can quickly compute an approximately optimal projection for a given data set. In the first stage, they constructed a graph from data, whose distance preserves pairwise geodesic distances. In the second stage, they enhanced the discrimination ability. Finally, they extracted Gabor features of GEI (Gait Energy Image) to generate a third-order tensor data, and conducted gait identification experiments with TRIMAP projection. They reached better recognition rates than LDA methods.

Temporal Projection. Vili Kellokumpu et al. [28] assume time as the third dimension other than XY axes in the image plane, so that consider the accumulation of gait sequence as XYT three-dimensional space. Three-dimensional local binary features (LBP) is used for XYT histogram extraction. To simulate the effect of multi-resolution analysis, they changed the radius of local binary features, and finally reached a better recognition rate. Yang Ran et al. [29] proposed a periodic pattern referred as double helical signature (DHS), which decomposes a video sequence into XT slices and generates DHS by iterative local curve embedding algorithm. It is used for segmentation and labeling of body parts in cluttered scenes and load-carrying conditions.

Temporal Modeling. Alessandro Bissacco et al. [30] directly collected key points by special equipments, and proposed a hybrid dynamical model of human motion. Different from traditional autoregressive model, it uses a collection of self-regression function to represent the entire gait cycle. Between the autoregressive parameters and observation, an intermediate filter layer is embedded to estimate the most possible states. The weighted sum of these filters derives the posterior probability. They also discussed the problem of the distance between the autoregressive models, and calculated the similarities for identification. Xin Zhang et al. [31] presented a approach include two generative models, called the kinematic gait generative model (KGGM) and the visual gait generative model (VGGM), which represent the kinematics and appearances of a gait by a few latent variables, respectively. And a new particle filtering algorithm is proposed for dynamic gait estimation. Gracian Trivinoa et al. [32] divided a gait cycle into four approximately equal phases. Based on Computational Theory of Perceptions (CTP), the relationship among horizontal acceleration, vertical acceleration, and other indicators are learned by rule-based approach. Homogeneity, symmetry, and the four root model are extracted to represent the individual characteristics.

3.3 Other Issues

It is worth notice that, there is also a number of interesting work related to gait recognition under view changes, speed changes, and other intractable conditions. For example, Maodi Hu et al. [33] proposed a multi-view multi-stance gait identification method using unified multi-view population HMMs (pHMMs). Hence, the gait dynamics in each view can be normalized into fixed-length stances by Viterbi decoding, whose results are exemplified in Figure 3. To optimize the view-independent and stance-independent identity vector, a multi-linear projection model is learned from tensor decomposition.

Fig. 3. Normalized dynamics of 18°, 54°, 90° and 126° views, which are extracted from CASIA Gait Database (Dataset B) [12]

4 Conclusions

This paper briefly reviews the main approaches in gait recognition, and recent progress in spatial and temporal modeling. The further trends of gait biometrics should be more robust features extracted; more accurate modeling of spatial and temporal information and improve the practicability of gait in real surveillance systems.

Acknowledgement. This work is funded by the National Basic Research Program of China (No. 2010CB327902), the National Natural Science Foundation of China (No. 60873158, No. 61005016, No. 61061130560) and the Fundamental Research Funds for the Central Universities.

References

1. Murray, M.P., Drought, A.B., Kory, R.C.: Walking patterns of normal men. Journal of Bone Join Surgery 46(2), 335–360 (1964)
2. Murray, M.P.: Gait as a total pattern of movement. American Journal of Physical Medicine 46(1), 290–333 (1967)
3. Ralston, H.J., Inman, V., Todd, E.: Human walking. Williams and Wilkins (1981)

4. Johansson, G.: Visual perception of biological motion and a model for its analysis. Perception and Psychophysics 14(2), 201–211 (1977)
5. Cutting, J.E., Kozlowski, L.T.: Recognizing friends by their walk: gait perception without familiarity cues. Bulletin of the Psychonomic Society 9(5), 353–356 (1977)
6. van Doornikc, J., Sinkjaer, T.: Robotic platform for human gait analysis. IEEE Trans. Biomed. Eng. 54(9), 1696–1702 (2007)
7. Lee, S.W., Mase, K., Kogure, K.: Detection of spatio-temporal gait parameters by using wearable motion sensors. In: Proc. IEEE Conf. on Eng. Med. Biol. Soc., pp. 6836–6839 (2005)
8. Vanitchatchavan, P.: Patterns of joint angles during termination of human gait. In: Proc. IEEE Conf. on Syst., Man, Cybern., pp. 1226–1230 (2000)
9. Barclay, C.D., Cutting, J.E., Kozlowski, L.T.: Temporal and spatial factors in gait perception that influence gender recognition. Perception and Psychophysics 23(2), 145–152 (1978)
10. Cutting, J.E., Proffitt, D.R., Kozlowski, L.T.: A biochemical invariant for gait perception. Journal of Experimental Psychology: Human Perception and Performance 4, 357–372 (1978)
11. Field, M., Stirling, D., Naghdy, F., Pan, Z.: Mixture model segmentation for gait recognition. In: ECSIS Symposium on Learning and Adaptive Behaviors for Robotic Systems, pp. 3–8 (2008)
12. Yu, S., Tan, D., Tan, T.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: Proc. IEEE/IAPR Int. Conf. Pattern Recog., vol. 4, pp. 441–444 (2006)
13. Gross, R., Shi, J.: The cmu motion of body (mobo) database. Robotics Institute, Pittsburgh, PA, Tech. Rep. CMU-RI-TR-01-18 (June 2001)
14. Shutler, J.D., Grant, M.G., Nixon, M.S., Carter, J.N.: On a large sequence-based human gait database. In: Proc. Int. Conf. Recent Advances Soft Comput., pp. 66–72 (2002)
15. Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P., Bowyer, K.W.: The human id gait challenge problem: Data sets, performance, and analysis. IEEE Trans. Pattern Anal. Mach. Intell. 27(2), 162–177 (2005)
16. Zhang, D., Wang, Y.: Investigating the separability of features from different views for gait based gender classification. In: Proc. IEEE/IAPR Int. Conf. Pattern Recog., pp. 1–4 (2008)
17. Ran, Y., Weiss, I., Zheng, Q., Davis, L.S.: Pedestrian detection via periodic motion analysis. Int. J. Comput. Vis. 2(71), 143–160 (2007)
18. Jean, F., Albu, A.B., Bergevin, R.: Towards view-invariant gait modeling: Computing view-normalized body part trajectories. Pattern Recog. 42(11), 2936–2949 (2009)
19. Gu, J., Ding, X., Wang, S., Wu, Y.: Action and gait recognition from recovered 3-d human joints. IEEE Trans. Syst., Man, Cybern. B 40(4), 1021–1033 (2010)
20. Boulgouris, N.V., Chi, Z.X.: Gait recognition using radon transform and linear discriminant analysis. IEEE Trans. Image Process. 16(3), 857–860 (2007)
21. Yu, S., Tan, T., Huang, K., Jia, K., Wu, X.: A study on gait-based gender classification. IEEE Trans. Image Process. 18(8), 1905–1910 (2009)
22. Hu, M., Wang, Y., Zhang, Z., Wang, Y.: Combining spatial and temporal information for gait based gender classification. In: Proc. IEEE/IAPR Int. Conf. Pattern Recog., pp. 3679–3682 (August 2010)
23. Venkat, I., DeWilde, P.: Robust gait recognition by learning and exploiting sub-gait characteristics. Int. J. Comput. Vis. 91(1), 7–23 (2011)

24. Kwon, K.S., Park, S.H., Kim, E.Y., Kim, H.J.: Human shape tracking for gait recognition using active contours with mean shift. In: Proc. Int. Conf. Human-Comput. Interaction, pp. 690–699 (2007)
25. Bashir, K., Xiang, T., Gong, S.: Gait representation using flow fields. In: Proc. British Mach. Vis. Conf. (2009)
26. Ho, M.-F., Chen, K.-Z., Huang, C.-L.: Gait analysis for human walking paths and identities recognition. In: Proc. Int. Conf. Multimedia Expo., pp. 1054–1057 (2009)
27. Chen, C., Zhang, J., Fleischer, R.: Distance approximating dimension reduction of riemannian manifolds. IEEE Trans. Syst., Man, Cybern. B 40(1), 208–217 (2010)
28. Kellokumpu, V., Zhao, G., Li, S.Z., Pietikainen, M.: Dynamic texture based gait recognition. In: Proc. IAPR/IEEE Int. Conf. Biometrics, pp. 1000–1009 (2009)
29. Ran, Y., Zheng, Q., Chellappa, R., Strat, T.M.: Applications of a simple characterization of human gait in surveillance. IEEE Trans. Syst., Man, Cybern. B 40(4), 1009–1020 (2010)
30. Bissacco, A., Soatto, S.: Hybrid dynamical models of human motion for the recognition of human gaits. Int. J. Comput. Vis. 85(1), 101–114 (2009)
31. Zhang, X., Fan, G.: Dual gait generative models for human motion estimation from a single camera. IEEE Trans. Syst., Man, Cybern. B 40(4), 1034–1049 (2010)
32. Trivinoa, G., Alvarez-Alvareza, A., Bailadorb, G.: Application of the computational theory of perceptions to human gait pattern recognition. Pattern Recog. 43(7), 2572–2581 (2010)
33. Hu, M., Wang, Y., Zhang, Z., Zhang, D.: Multi-view multi-stance gait identification. In: Proc. IEEE Int. Conf. Image Process. (2011)

Significance of Dynamic Content of Gait Present in the Lower Silhouette Region

Shreyas Saxena

Punjab Engineering College University of Technology, Chandigarh, India
shreyassaxena.ee08@pec.edu.in

Abstract. In the present scenario, Gait descriptors are required to extract the dynamic and static information of the gait. The static and dynamic descriptors are formed from the entire region of the body. We know that majority of dynamic information is in the lower silhouette whereas majority of static information is in the upper silhouette. In our work we have evaluated the significance of dynamic information extracted from the lower silhouette. State of the art feature descriptors are used along with a feature selection mask to form the final signature templates for classification. Our results indicate a significant dynamic content in the lower silhouette which itself is able to give decent recognition rate. Future work can improve performance by using dynamic information from lower silhouette in conjunction with static information derived from upper silhouette.

Keywords: Gait Recognition, Dynamic Content, Gait Entropy, Feature Selection.

1 Introduction

In the recent past, human gait has gained significant interest in the field of human identification.This is mainly due to the fact that gait does not require subject cooperation, is non-invasive and tough to replicate or hide.

There are two aspects involved in Gait Recognition, a robust feature extractor and selection of appropriate features from the feature space. The need for a robust feature extractor increased after the introduction of challenging datasets like CASIA Set B [17], Gait Human ID challenge [13], SOTON Database [15]. Although with time many robust shape descriptors were developed, they all had an inherent problem of high computational load and feature space. To tackle the problem of shape co-variates like bag and clothes, the community started investigating the content of the gait. Broadly speaking, the feature extractors either try to extract the dynamic part [1] or static part [12] of the gait. Some even try using both in conjunction to enhance recognition [3]. Although using the static and dynamic information gives appreciable results but it leads to increase in dimensionality of the data. Earlier, significant work has been done to perform feature selection on the high dimensional database the objectives being, either to extract information from dynamic areas of gait [9], [2] or to select relevant features for classification [6].

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 159–166, 2011.

The shortcomings of these feature selectors can be removed by knowing the static and dynamic content of different regions, so that we know where to look for dynamic and static information. We all know that majority of the dynamic content of the gait comes from lower silhouette but we do not know how much it accounts for in total. Once we have a fair idea of the significance of dynamic content in lower silhouette, we can confidently extract the static part of gait from the upper silhouette and use it along with the dynamic content of the lower silhouette. This motivates our research to evaluate the actual dynamic content of lower silhouette.

2 Related Works

In the past people have tried to evaluate the relative importance of different regions of a human body. The debate always exists about the significance of upper or lower part of the silhouette. From any region of the body one can get both dynamic as well as static information. It is a general perception that the upper part of the body gives more of static information whereas majority of the dynamic information is reflected in the lower part of silhouette. Foster et al. [5] constructed four different masks to evaluate the importance of corresponding regions. Apart from these, various combinations were also tested to see if different regions of body reinforce each other. The results in [5] were preliminary and not encouraging. This was partly due to the fact that many robust shape descriptors were not developed at that time.

Until now, there exists a plethora of features available for doing gait recognition. They can be classified according to various criteria, such as temporal versus static, area based versus point based etc. For our investigation we want features which are able to efficiently extract the dynamic information to evaluate its significance in the lower silhouette. Inspired by the psychological evidence provided by Johansson MLD experiment [8] we explore the features in the model free domain. Significant work can be seen in model free domain [2], [7], [4], [16] in the past. The problem with most of these approaches is that the signature is formed to represent the dynamic or static content by using the entire silhouette. If we know where to look at and for what, then we can extract static and dynamic properties from their corresponding appropriate regions.

One of the obvious choices for extracting dynamic information is a feature extractor which use optical flow as raw data. Methods using optical flow/temporal difference have a common advantage that irrespective of whether they are making a spatial template or a temporal template series, they inherently capture the dynamic aspect of human motion. The disadvantage is that the calculation of optical flow consumes some of the available computational power. Also, if the direction of optical flow is taken into account then the recognition results deteriorate; especially, if the video is having a poor resolution. In the beginning, features introduced extracted local properties of an image and hence had a high computational overhead [14]. Recently, the scientific community has started looking for global features and some note worthy work can be seen on SVB frieze patterns [11], CHILAC features [10], Flow fields [3].

3 Feature Extraction

To perform our investigation about dynamic content, we are required to select two descriptors. One which effectively represents the gait dynamics and another which selects and extracts dynamic information. Out all the descriptors, we select the feature extractor proposed in [3] as not only it uses optical flow which by default is a good descriptor for dynamics, but also quantizes it in four directions. In this way we are able to use the information about direction of flow, which is generally avoided due to the presence of noise in a low-resolution video. For extracting dynamic information, we use the feature selection mask proposed in [2]. Since the mask is formed to select the dynamic areas of the entire silhouette, we evaluate its selectivity for dynamic areas. The section below gives a description about the construction of the descriptors and the feature selection mask:

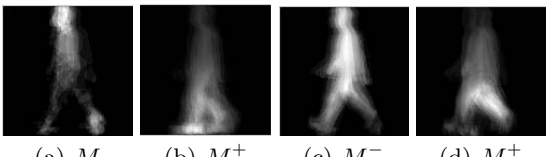
1. Optical Flow [3]

In 2009, Bashir proposed the use of discretized optical flow for gait recognition. The optical flow has two important aspects: magnitude and direction. Bashir discretized the optical flow in four directions and formed five signatures for each person, $M_x^+, M_y^+, M_x^-, M_y^-$ representing the dynamic aspect of motion and M representing the static aspect. The benefit of using this feature is that since it deals with temporal data, we have features robust against segmentation errors and shape covariates.More so, discretized optical flow makes the method robust to the noise induced because of poor resolution.

The whole image sequence is divided into individual gait cycles, and optical flow F is computed for each cycle. F is discretized in four directions, $F_x^+, F_y^+, F_x^-, F_y^-$ and the four bins, $M_x^+, M_y^+, M_x^-, M_y^-$ are computed in a similar way as done in [3]. The value for bin M at a pixel location is incremented if the magnitude of optical flow at that pixel is zero. The value for bin M_x^+ is incremented if there is a non zero value for flow field vectorF_x^+ and becomes

$$M_x^+(i,j) = \sum_{1}^{T} F_x^+(i,j) \tag{1}$$

Values for other bins M_y^+, M_y^- and M_x^- and are computed in a similar fashion. The bins are normalized so that they can be treated as a feature. The signature template M_y^- is neglected due to its low discriminative power [3] .



(a) M (b) M_x^+ (c) M_x^- (d) M_y^+

Fig. 1. Gait Descriptors using Optical Flow

2. Masked Gait Energy Image [2]

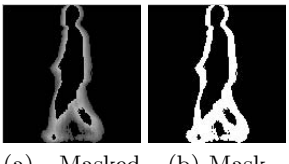
In 2010, Bashir proposed the use of gait energy image masked with gait entropy image. The benefit of using the masking scheme is that it makes the signature template more robust to shape covariates. In addition the computational requirement to compute the mask is low. We compute the gait energy image by computing the average along the entire gait cycle:

$$GEI = G(x, y) = \frac{1}{T} \sum_{t=1}^{T} I(x, y, t) \tag{2}$$

where T is the number of frames in the gait cycle, I is the silhouette image. This signature is not yet robust to shape covariates. Therefore to make it robust we follow the approach as done by author in original paper. We compute the Gait Entropy Image using Shannon Entropy operator:

$$GEnI = H(x, y) = -\sum_{k=1}^{K} p_k(x, y) \log_2(p_k(x, y)) \tag{3}$$

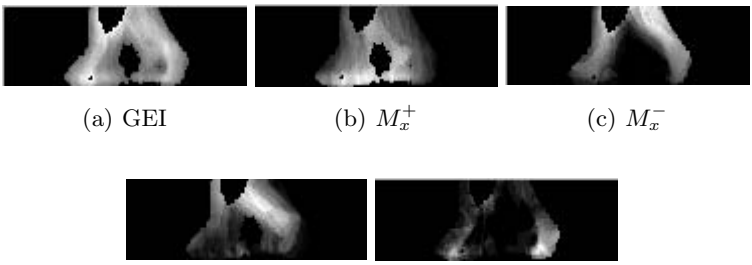
where $p_k(x, y)$ is the probability that the pixel takes the k^{th} value.Once, we have the normalized gait entropy image, we form the mask using this image. Specifically, the areas in the gait entropy image having value higher than threshold 0.75 are assigned a value equal to one. Whereas the areas having a value lower than 0.75 are made equal to zero. In this way, we obtain a mask which selects the dynamic areas of the silhouette.



(a) Masked GEI (b) Mask

Fig. 2. Gait Descriptors using Optical Flow

Now we construct the masked signature templates to evaluate their capability to represent the dynamic content of gait. Apart from the masked gait energy image, M_x^+, M_y^+, M_x^-, M are masked with the mask as proposed in [2]. To form the final templates only the lower 1/3 rd portion of silhouette is used, as generally legs reside within this region.



(a) GEI (b) M_x^+ (c) M_x^-

(d) M_y^+ (e) M

Fig. 3. Corresponding masked signature templates

4 Experiments

In the preliminary stage of our evaluation, we wanted the silhouettes to be free from any distortion possible so that, a neat conclusion can be achieved regarding the dynamic content of the lower silhouette. Hence, we use CASIA database Set B [17] to evaluate our framework and refrain from using HumanID gait challenge database [13] as the silhouettes suffer from shadow co-variates. The CASIA database comprises of 124 subjects from which we have used 63 for evaluation of our framework. For each subject there are 10 walking sequences consisting of 6 normal walking sequences (Set A), 2 carrying bag sequences (Set B) and 2 wearing-coat sequences (Set C). The gallery set used for the experiment consists of the last 4 sequences of each subject in Set A (Set A1). The probe set is the rest of the sequences in Set A (Set A2), Set B and Set C.

4.1 Results and Analysis

It is evident from the past work that the mask proposed in [2] selects the dynamic areas of the silhouette. Although, we do not doubt its capability in selecting areas in top silhouette region, we are uncertain about it's selection capability for lower silhouette. Therefore, we perform the evaluation of dynamic content with and without the mask to see its usefulness in removing the static areas. To reduce the affect of noise and curse of dimensionality, the signatures were projected in a low dimensional eigenspace. To classify the signatures, we use a simple k-NN classifier for k= 1 and 3 which uses Euclidean distance in eigen subspace as a metric. For the probe set B and C, in addition to the use of k-NN on reduced

Table 1. Recognition results for Probe set A2

		M_x^+	M_x^-	M_y^+	M	GEI
K-NN, K=1	With mask	81.74	70.63	80.15	39.68	84.57
	Without mask	75.39	64.29	65.07	25.39	73.04

Table 2. Recognition results for Probe set B

		M_x^+	M_x^-	M_y^+	M	GEI
K-NN, K=1	With mask	38.10	40.48	31.75	34.13	53.97
	Without mask	24.60	39.68	22.22	33.33	39.68
Direct Template Matching	With mask	38.09	34.92	28.57	38.09	38.90
	Without mask	23.02	41.27	21.43	31.75	58.73

Table 3. Recognition results for Probe set C

		M_x^+	M_x^-	M_y^+	M	GEI
K-NN, K=1	With mask	72.22	77.78	42.86	61.90	84.13
	Without mask	41.27	60.32	17.46	42.86	68.25
Direct Template Matching	With mask	63.49	74.60	39.68	73.02	74.60
	Without mask	41.27	63.49	23.81	37.30	88.89

subspace, we also evaluate the performance of direct template matching when used on the original space.

From Table 1, 2 and 3, it can be seen that that all the signature templates have a decent performance except for M. This does not come as a surprise because in the original paper [3], M was supposed to represent the static part of the gait. Therefore instead of using M to extract properties of entire silhouette, one should use this to derive static information from the upper silhouette region. Also, it can be seen that in general, the proposed mask helps in selecting the dynamic areas and as a result, increases the overall correct classification rate. The decrease in the recognition rate of M_y^+ for probe set B and C is quite a lot. This can be explained by the fact that a long coat primarily restricts the motion of legs in vertical direction. In the case of briefcase the vertical oscillations tamper with optical flow in vertical direction. It is observed that the results for direct template matching are far better than results of k-NN classifier on a reduced subspace. This might be due to the fact that a distortion is produced in the probe B and probe C templates while projecting them on eigen space which is not the case in the normal carrying sequences.

By analyzing the individual performances of the descriptors from the tables, it can be seen that the most important ones in order of significance are: M_x^+, M_x^-, GEI. As a side research, the three masked signatures M_x^+, M_x^-, GEI were merged together in order to evaluate their performance. The dissimilarity score of each individual signature was fused with a corresponding weight of 0.6, 0.5 and 0.3 to form the final score for the 1-NN classifier. This weight was decided in accordance with the individual capability of each signature for classification. The recognition rate achieved with the fusion was 88.89% for probe set A.

5 Conclusion

We have evaluated the dynamic content of the lower part of the silhouette, which was never investigated with state of the art features. The individual performance of motion descriptors proposed in [3] and [2] with and without feature selection mask has been analyzed. The mask has a significant impact on the overall recognition rate and is hence found out to be useful for selecting dynamic areas of the lower silhouette. We combine the dissimilarity scores of important masked descriptors M_x^+, M_x^-, GEI and achieve a decent recognition rate with a simple 1-NN classifier. Thus the significance of the dynamic information present in the lower silhouette is highlighted in this work.

6 Future Works

From our work it is clear that significant dynamic information exists in the lower region of silhouette. With the help of dynamic information, we are able to achieve a CCR(Correct Classification Rate) of 88.89% .To increase the CCR, we can extract static information from the upper part of silhouette and use it in conjuction with the dynamic information from the lower part of silhouette. Since now we know what we are looking for in the upper silhouette, we can use robust feature extractors to extract static information of the gait. Also, once we have the static signatures from upper silhouette and dynamic signatures from lower silhouette, factoral analysis can be performed to see if the dynamic and static information are the underlying factors responsible for recognition.

References

1. Bashir, K., Xiang, T., Gong, S.: Gait recognition using gait entropy image. In: 3rd International Conference on Crime Detection and Prevention (ICDP 2009), pp. 1–6. IET (2009)
2. Bashir, K., Xiang, T., Gong, S.: Gait recognition without subject cooperation (2010)
3. Bashir, K., Xiang, T., Gong, S., Mary, Q.: Gait representation using flow fields. In: British Machine Vision Conference (2009)
4. BenAbdelkader, C., Cutler, R.G., Davis, L.S.: Gait recognition using image self-similarity. Journal on Applied Signal Processing, 572–585 (2004)
5. Foster, J.P., Nixon, M.S., Prugel-Bennett, A.: Automatic gait recognition using area-based metrics. Pattern Recognition Letters 24(14), 2489–2497 (2003)
6. Guo, B., Nixon, M.S.: Gait feature subset selection by mutual information. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 39(1), 36–46 (2009)
7. Hayfron-Acquah, J.B., Nixon, M.S., Carter, J.N.: Automatic gait recognition by symmetry analysis. Pattern Recognition Letters 24(13), 2175–2183 (2003)
8. Johansson, G.: Visual perception of biological motion and a model for its analysis. Attention, Perception, & Psychophysics 14(2), 201–211 (1973)
9. Kellokumpu, V., Zhao, G., Li, S., Pietikäinen, M.: Dynamic texture based gait recognition. Advances in Biometrics, 1000–1009 (2009)

10. Kobayashi, T., Otsu, N.: Action and Simultaneous Multiple-Person Identification Using Cubic Higher-Order Local Auto-Correlation. In: International Conference on Pattern Recognition, vol. 4, pp. 741–744 (2004)
11. Lee, S., Liu, Y., Collins, R.: Shape variation-based frieze pattern for robust gait recognition. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE (2007)
12. Liu, Z., Sarkar, S.: Simplest representation yet for gait recognition: Averaged silhouette. Pattern Recognition 4, 211–214 (2004)
13. Phillips, P.J., Sarkar, S., Robledo, I., Grother, P., Bowyer, K.: The gait identification challenge problem: Data sets and baseline algorithm. In: Proceedings of 16th International Conference on Pattern Recognition, vol. 1, pp. 385–388. IEEE (2002)
14. Polana, R., Nelson, R.C.: Detection and Recognition of Periodic, Nonrigid Motion. International Journal of Computer Vision 23, 261–282 (1997)
15. Shutler, J., Grant, M., Nixon, M.S., Carter, J.N.: On a Large Sequence-Based Human Gait Database (2002)
16. Tao, D., Li, X., Wu, X., Maybank, S.J.: General tensor discriminant analysis and gabor features for gait recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1700–1715 (2007)
17. Yu, S., Tan, D., Tan, T.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. Pattern Recognition 4, 441–444 (2006)

Emotional Speaker Identification by Humans and Machines*

Yingchun Yang**, Li Chen, and Wenyi Wang

College of Computer Science & Technology, Zhejiang University, Hangzhou, China
{yyc,stchenli}@zju.edu.cn, wangwy1105@gmail.com

Abstract. This paper concerns the problem of the effect of emotional change on humans and machines for speaker identification. A contrasting experiment is carried out between Automatic Speaker Identification (ASI) system (applying GMM-UBM and Emotional Factor Analysis (EFA) algorithm) and aural system on emotional speech corpus MASC. The experimental result is similar to that in channel-mismatched condition, i.e. the ASI system is much better than the single listener, especially when emotion compensation algorithm EFA is applied. Meanwhile,fusion of multiple listeners can significantly improve the aural system performance by 23.86% and make it outperform the ASI system.

Keywords: emotional speaker recognition, human listeners, EFA, emotional speaker identification.

1 Introduction

Human ears were always regarded as a perfect instrument for speaker recognition tasks under various situations, and the goal of Automatic Speaker Recognition (ASR) is to approach human-like performance. To find some cues that human use to discriminate different people and to give insights for improving ASR algorithm, a couple of experiments have been carried out to compare the performance of human listeners with the machine algorithms. For instance, Astrid [1] conducted an experiment on NIST 1998 Speaker Recognition Evaluation to distinguish accuracy between human listeners and the machine verification algorithms. The result showed that the performance of the human listeners was comparable with the computer algorithms in the same handset condition. In the different handsets condition, human and machine performance both degraded, but the former was more robust than the latter. Almost the same result was obtained in [2] on another dataset, which also observed the main characteristics that human depend on to identify different speakers.

Recent studies have shown significant improvement in ASR performance. The famous GMM-UBM (Gaussian Mixture Model-Universal Background Model) has

* This paper is supported by NSFC60970080 and the Special Funds for Key Program of the China No. 2009ZX01039-002-001-04.

** Corresponding Author.

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 167–173, 2011.

high accuracy for speaker recognition task on high-quality data under controlled conditions. Unfortunately, in actual environment, many disruptive factors from outside and inside could result in dramatic degradation in the ASR system performance. Many efforts have been made to deal with the inter-session variability (especially the channel variability), e.g. JFA (Joint Factor Analysis), and the results are very promising. Ville et al. [3] set up an experiment to compare human listeners with the state-of-the-art JFA automatic system on NIST SRE 08 core set. It proved that JFA was a better performer, and combining human decision with it could not bring positive effect on improving system performance.

However, intra-speaker variability hasn't attracted enough attention despite being another important factor which prevents the ASI system from stepping into the real world. NIST Speaker Recognition Evaluation [4] first took the effect of vocal effort into consideration in 2010. SRI [5] has studied systematically the effects of vocal effort and speaking style on speaker verification performance, which demonstrated that intrinsic variation caused the degradation of the speaker verification performance. Nevertheless, the past research mainly focused on Automatic Speaker Verification (ASV) task and little effort has been made to investigate the ASI task. Therefore, this paper makes performance comparison between ASI system and aural system in speaker-emotion variability situation. An experiment on MASC [6] (an emotional speech corpus) has been designed to compare the performance of the aural system and the ASI system. We mainly try to find the answers to the following questions:

(1) Does the ASI system outperform the aural system facing speaker-emotion variability just like channel variability?
(2) Can fusion of multiple human listeners contribute to a better performance?

The rest of the paper is organized as follows. Section 2 introduces the experiment dataset. Section 3 presents the aural system together with the setup of the listening protocol. Section 4 elaborates on the ASI system, including two algorithms: GMMUBM and EFA. Section 5 evaluates the validity of our dataset and compares the performance of ASI system with that of aural system. Finally, the conclusion is drawn in Section 6.

2 Dataset

Mandarin Affective Speech Corpus (MASC), an emotional speech database, is adopted in our experiments. The corpus contains recordings of 68 Chinese speakers (23 female and 45 male). There are five kinds of emotion: neutral, anger, elation, panic and sadness. 20 utterances are spoken for three times under each emotional state, and 2 extra neutral paragraphs. Utterance is short, lasting between 5s and 10s, while the paragraph is longer, with a length of about 20s. All utterances and paragraphs are recorded by an OLYMPUS DM-20 digital voice recorder.

Our experimental dataset is the subset of the MASC, containing the data of the last fifty persons. Utterances under each of the five emotional states are selected as the test samples. Each trial is composed of the test sample and its corresponding five candidate speakers represented by their neutral paragraphs. Human listeners should

choose one from these five speakers. The number of the candidate speakers is decided to be five with the consideration of human memory limitation. For each test sample, the five candidate speakers are chosen as follows: the target speaker is selected first and then the other four are determined by likelihood scores based on GMM-UBM, choosing the top four speakers.

To investigate whether the fusion result of multiple listeners can improve the performance, the neutral utterances are duplicated for eight times, and utterances under anger, elation, panic and sadness are duplicated for four times. All these trials are divided into sixty segments at random to ensure different test samples are in each segment, and each segment containing 400 trials is assigned to one listener, thus there were 60 human listeners.

3 Aural System

3.1 Task Definition

ASI system aims to identify an input test sample by selecting one speaker from a set of candidate speakers. The performance of ASI system can be measured in terms of Identification Rate (IR), which can be calculated through the equation(1).

$$IR = \frac{right\ number\ of\ trials}{number\ of\ trials} \tag{1}$$

3.2 Listening Protocol

To make the aural system comparable to the ASI system, we try our best to design the human listener test to parallel the ASI as closely as possible.

(1) None of the listeners possesses any formal training in phonetics and does not personally know the speakers whose voices appear in the experiment.
(2) The machine evaluates each speaker with likelihood scores in ASI system, whereas subjects just need to select the most likely speaker and give their confidence level in aural system. There are three levels of confidence: sure (certain about his/her choice), uncertain (doubt about the decision. For example, he/she hesitated between two possible speakers), and no idea (doubt among more than 3 speakers).
(3) Listeners are ensured that there must be a target speaker for each test sample.
(4) Listeners are not allowed to know the emotion to which the utterance belongs.
(5) Listeners are not provided with the scripts of each utterance.

3.3 Test Procedure

A software tool is built for human listeners, and a screenshot is shown by Fig. 1.

The listeners need to make clicks to listen to the testing utterances or the candidate speakers. They can listen to these sentences in a trial for any times and in any order until making a decision and choose the corresponding confidence level. The listeners are asked to have a rest after finishing 50 trails to prevent them from getting tired.

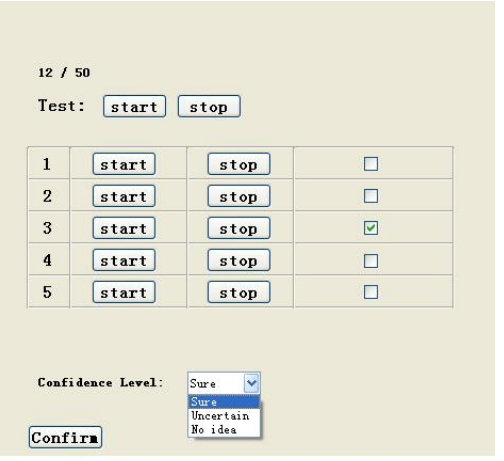


Fig. 1. The screenshot of the software tool used for human listening experiments

4 ASI System

The ASI system adopts GMM-UBM and EFA (Emotional Factor Analysis) algorithm, using 13 Mel Frequency Cepstral Coefficients (MFCC) with 32ms window length and 16ms frame rate.

For GMM-UBM (Gaussian Mixture Models-Universal Background Model) algorithm, the gender-independent UBM, with 512 Gaussian mixtures, is trained using the first 18 speakers' neutral utterances and paragraphs. The total duration of the training data is approximately 1 hour. Target GMMs are adapted from the UBM with the speaker's neutral paragraph by maximum a posterior (MAP) adaptation. Only the means of the Gaussian components are adapted, and no score normalization technique is used.

JFA algorithm proposed in [7] is successfully applied in Automatic Speaker Verification (ASV) to cope with session-variability problem, which models each utterance with the sum of speaker space, channel space and variability. Each model can be represented by(2):

$$M(s) = m + ux(s) + vy(s) + dz(s) \tag{2}$$

Where u is the eigenvoice matrix, v is the eigenchannel matrix, and d is the residual loading matrix. When the eigenchannel matrix v is replaced by the eigenemotion matrix, JFA would develop into EFA algorithm, and can be applied as the emotion compensation technique [8].

In MASC, the eigenemotion matrix is composed of the four smaller eigenemotion matrixes corresponding to the four emotional states respectively. Each smaller eigenemotion matrix is trained independently with 50 emotion factors. The training data come from the first 18 speakers' utterances under all emotional states in development dataset.

The EFA algorithm is implemented in matlab on the basis of "Joint Factor Analysis Matlab Demo" [9], which doesn't contain the second-order statistics or update the

covariance matrix. We revise it and apply it into our experiment. The covariance is updated in solving the eigenvoice matrix and residual matrix, but in order to the keep the consistency of the combined eigenemotion matrix, we do not update it in solving the eigenemotion matrix. Finally, when calculating the likelihood score, we select the linear scoring method.

5 Experiment Results

5.1 IR (Human Listeners vs. Machine Algorithm)

The performance comparison between ASI system and aural system is shown in Fig. 2. The performance of human listeners is poorer than the machine algorithms. The average IR is 50.1%, 55.52% and 61.64% for human listeners, GMM-UBM and EFA respectively. Through comparing the result of aural system and GMM-UBM algorithm, we can see that the performance gap is mainly caused by the IR under the emotion of neutral because the IR of aural system is 20.61% lower than that of the GMM-UBM algorithm. For the other four emotions, the performances of human and GMM-UBM are similar, and the human is more robust when the emotional states change. This phenomenon indicates that the aural system can't compete with GMM-UBM algorithm in IR, but it still possesses substantial advantages over GMM-UBM in dealing with speaker-emotion variability. Utilizing emotion compensation, the EFA algorithm is much better than the GMM-UBM algorithm and the aural system. The result is 93.07%, 50.13%, 51.33, 45.97% and 65.2% under neutral, anger, elation, panic and sadness respectively. The better result is reflected not only by the higher average IR, but also by the increase of IR under each emotional state, which indicates that EFA algorithm can alleviate the effect exerted by the change of emotional states.

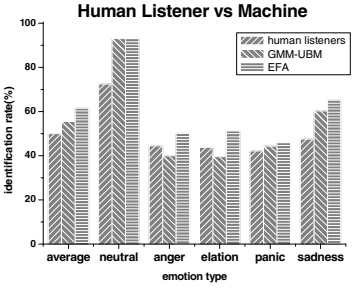


Fig. 2. Comparison between the GMM-UBM, EFA algorithm and the aural system

5.2 IR(Multiple Human Listeners vs. Machine Algorithm)

In our experiments, the choice among five candidate speakers together with the confidence level is given to each trial. Fusion method of these choices is that each choice is

assigned a score according to the confidence level, which is 5, 3, 1 for sure, uncertain and no idea respectively. Then, averaging strategy is employed to fuse these scores by (3):

$$score_s = \frac{\sum_{i=1}^{n} s_i * c_i}{n} \tag{3}$$

where n means the number of listeners. $s_i = 1$ means the $i-th$ listener selects the $s-th$ speaker, while $s_i = 0$ means he/she doesn't. c_i is the value of the confidence level of the choice. The largest score is selected as the answer. The IR of fusion result is shown in Table 1.

Table 1. Comparison of IR between single listener system and multiple listeners system

emotion	Single Listener	Multiple Listeners	EFA
neutral	72.0%	92.3%	93.07%
Anger	44.0%	69.6%	50.13%
elation	43.0%	68.7%	51.33%
Panic	41.0%	69.1%	45.97%
sadness	47.1%	70.1%	65.20%
average	50.1%	73.96%	61.64%

As mentioned in [3], the Equal Error Rate (EER) of multiple listeners decreases dramatically from 23% to approximately 12.5%. From Table 1, it is obvious that the performance of multiple listeners is much better than that of the single listener. The fusion result brings 23.86% improvement of IR to aural system. For each emotional state, the improvement is basically the same. When compared with the performance of EFA algorithm, the performance of group decision is comparable in the neutral condition, yet the performance of multiple listeners is better and more robust in the other emotional states.

6 Conclusion

The contrasting experiment conducted here shows the ASI system is better than the aural system in the aspect of IR. At the same time, ASI system is still advantageous over the aural system under different emotional states. The GMM-UBM and EFA algorithms are better than the aural system by 5.42% and 10.14% respectively. Meanwhile, the ASI system is not as robust to emotional changes as the aural system, for the performance of ASI system under different emotional states varies more. Therefore, the different advantages of the two systems suggest a possibility to fuse these two systems in the future.

References

1. Schmidt-Nielsen, A., Crystal, T.H.: Speaker verification by human listeners: experiments comparing human and machine performance using th2e NIST 1998 speaker verification data. Digital Signal Processing 10, 249–266 (2000)

2. Kajarekar, S.S., Bratt, H., Shriberg, E., de Leon, R.: A study of intentional voice modifications for evading automatic speaker recognition. In: Speaker Odessy (2006)
3. Hautamaki, V., Kinnunen, T., Nosratighods, M., Lee, K.-A., Ma, B., Li, H.: Approaching human listener accuracy with modern speaker verification. In: Interspeech 2010, pp. 1473–1476 (2010)
4. The NIST year, speaker recognition evaluation plan (2010)
5. Shriberg, E., Graciarena, M., Bratt, H., Kathol, A., Kajarekar, S., Jameel, H., Richey, C., Goodman, F.: Effects of vocal effort and speaking style on text-independent speaker verification. In: Interspeech 2007, Antwerp, pp. 950–954 (2007)
6. Wu, T., Yang, Y., Wu, Z., Li, D.: MASC: A Speech Corpus in Mandarin for Emotion Analysis and Affective Speaker Recognition. In: ODYSSEY 2006, pp. 1–5 (June 2006)
7. Kenny, P., Boulianne, G., Ouellet, P., Dumouchel, P.: Joint factor analysis versus eigenchannels in speaker recognition. IEEE Transaction on Audio Speech and Language Processing 15(4), 1435–1447 (2007)
8. Chen, L., Yang, Y.: Applying Emotional Factor Analysis and I-Vector to Emotional Speaker Recognition. Submitted to CCBR (2011)
9. `http://speech.fit.vutbr.cz/en/software/joint-factor-analysis-matlab-demo`

Applying Emotional Factor Analysis and I-Vector to Emotional Speaker Recognition*

Li Chen and Yingchun Yang**

College of Computer Science & Technology, Zhejiang University, Hangzhou, China
{stchenli,yyc}@zju.edu.cn

Abstract. Emotion variability is an important factor that degrades the performce of speaker recognition system. This paper borrows ideas from Joint Factor Analysis (JFA) algorithm based on the similarity between emotion effect and channel effect and develops Emotional Factor Analysis (EFA) into solving the emotion variability problem. I-Vector is appiled also. The experiment carried on MASC (Madarin Affective Speech Corpus) shows that EFA and I-Vector method can bring an IR increase of 7%~10% and an EER reduction of 3%~4% compared with the GMM-UBM system.

Keywords: EFA, I-Vector, emotional speaker recognition.

1 Introduction

The most common approach to speaker recognition today is the Gaussian Mixture Model (GMM). It can achieve promising performance under controlled conditions. However, the performance of the traditional algorithm deteriorate dramatically with many varying factors (such as different channels, different emotional states). NIST SRE has spent many efforts to solving the channel variability problem and developed two effective methods --- JFA and I-Vector.

JFA (Joint Factor Analysis) [1] [2] has been the state-of-the-art technique in solving session-variability problem in speaker recognition task. It supposes that most of variance in the session-dependent GMM supervector is accounted for by a small number of hidden variables named speaker and channel factors. The general model combines the priors underlying classical MAP, eigenvoice MAP and eigenchannel MAP.

I-Vector [3] [4] can be seen as a simplified JFA because it only trains one space named “total variability space” containing both the speaker and channel variabilities.

However, little effort has been made to solve the emotion-variability problem. Based on the pervious analysis [5], the problem of emotion effect is somewhat similar to that of channel effect. Meanwhile, SRI [6] found out that the technique modeling the extrinsic variability could also model the intrinsic variability (including speaking

* This paper is supported by NSFC60970080 and the Special Funds for Key Program of the China No. 2009ZX01039-002-001-04.

** Corresponding Author.

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 174–179, 2011.

style, emotion, level of vocal effort etc.) well. The emotion variability can be considered as a special kind of channel variability that the different channel is due to the different shape of our vocal organs, and this variability is predictable under each emotional state as under each channel. Thus, borrowing ideas from JFA and I-Vector to deal with the emotion variability problem is reasonable. EFA is proposed based on the theory of JFA and I-Vector is applied to deal with the session variability problem.

The paper is organized as follows. Section 2 and Section 3 introduce the principle of EFA and I-Vector respectively. Section 4 gives details on our experiment procedure and results. Finally, Section 5 concludes the paper.

2 Emotional Factor Analysis

EFA (Emotional Factor Analysis) supposes that each speaker- and emotion- dependent supervector $M(s)$ can be decomposed into a sum of two supervectors: a speaker supervector s and an emotion supervector c:

$$M = s + c \tag{1}$$

The speaker supervector contains the speaker and emotion independent supervector, the speaker space and the residual space:

$$s = m + Vy + Dz \tag{2}$$

V is the eigenvoice matrix and D is the diagonal matrix means the residual space. y and z are the speaker and residual factors.

The emotion-dependent supervector is supposed to be distributed as equation (3):

$$c = Ux \tag{3}$$

where U is the eigenemotion matrix, and x is the emotion factor.

To apply EFA into speaker recognition task, the subspace (U, V, D) should be estimated on the labeled development corpora and the speaker model (x, y, z) should be estimated for a given training utterance. The score is calculated by computing the likelihood of the test feature on the session-compensated model $(M - Ux)$.

2.1 Subspace Estimating

The subspace estimating process contains three phases: estimating the eigenvoice matrix, estimating the eigenemotion matrix, and estimating the residual matrix.

All these three phases all use EM algorithm similarly. The process of training eigenvoice matrix is demonstrated as an example:

Step 1: Estimate the vector $y(s)$ using equation(4):

$$\begin{aligned} L(s) &= I + V^T E^{-1} N(s) V \\ E[y(s)] &= L^{-1}(s) V^T \Sigma^{-1} S_x(s) \end{aligned} \tag{4}$$

where Σ is the covariance matrix and $S_x(s)$ is the first-order statics of speaker s.

Step 2: Updating the eigenvoice matrix V as equation(5):

$$\sum_s N(s)VE[y(s)y^T(s)] = \sum_s S_x(s)E[y^T(s)] \tag{5}$$

Step 3: Updating the covariance matrix Σ using equation(6):

$$\Sigma = N^{-1}S_{XX^T}(s) - N^{-1}diag\{\sum_s S_x(s)E[y^T(s)]V^T\} \tag{6}$$

$S_{XX^T}(s)$ is the second-order statistics of speaker s.

3 I-Vector

The EFA algorithm models speaker space and emotion space separately, while I-Vector method supposes there is only a single space named total variability space, containing both speaker and emotion space variability. Thus, the new model is represented by

$$M = m + Tw \tag{7}$$

T is a low-rank matrix meaning the total variability space.

The training process is similar to estimating the eigenvoice matrix in EFA. The main difference is that all recordings of a given speaker are considered to belong to the same person in EFA, while they are regarded as having been produced by a different speaker in I-Vector.

Another important difference is that a simplified cosine similarity scoring method is adopted in the test phase of I-Vector.

$$score(w_{target}, w_{test}) = \frac{w_{target}^t \cdot w_{test}}{\| w_{target} \| \| w_{test} \|} \begin{matrix} \geq \\ < \end{matrix} \theta \tag{8}$$

where w_{target} and w_{test} are the I-Vectors of target speaker and test sample. θ is the decision threshold.

Two emotion compensation methods are commonly used in I-Vector to remove the nuisance effects in total factor space: WCCN and LDA.

3.1 WCCN

WCCN (Within-Class Covariance Normalization) is proposed in SVM in one-versus-all decision. The main idea is to apply kernel as illustrated in equation(9):

$$k(w_1, w_2) = w_1 W^{-1} w_2 \tag{9}$$

W is the within-class covariance matrix computed over all the speakers in the training corpus.

$$W = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{n_s}\sum_{i=1}^{n_s}(w_i^s - \overline{w_s})(w_i^s - \overline{w_s})^t \tag{10}$$

S is the number of the speakers in training corpus. $\overline{w_s} = \frac{1}{n_s}\sum_{i=1}^{n_s} w_i^s$ is the mean of the $s-th$ speaker. Through Cholesky decomposition, B is obtained by $W^{-1} = BB^t$. Then the cosine scoring equation is changed into:

$$score(w_{t\arg et}, w_{test}) = \frac{(B^t w_{t\arg et})^t \cdot (B^t w_{test}) \geq}{\| B^t w_{target} \| \| B^t w_{test} \| <} \theta \tag{11}$$

3.2 LDA

LDA (Linear Discriminant Analysis) is a classical method to seek new space to discriminate between different classes better. The new space is found by maximizing the between-class variance and minimizing the within-class variance. It is calculated by the general eigenvalue equation in(14):

$$S_b = \sum_{s=1}^{S} (w_s - \overline{w})(w_s - \overline{w})^t \tag{12}$$

$$S_w = \sum_{s=1}^{S} \frac{1}{n_s} \sum_{i=1}^{n_s} (w_i^s - \overline{w_s})(w_i^s - \overline{w_s})^t \tag{13}$$

$$S_b v = \lambda S_w v \tag{14}$$

S_b is the within-class variance and S_w is the between-class variance. The project matrix from original space to the new space is composed by the best eigenvectors of(14). The new cosine distance is computed as(15):

$$score(w_{t\arg et}, w_{test}) = \frac{(A^t w_{t\arg et})^t \cdot (A^t w_{test}) \geq}{\| A^t w_{target} \| \| A^t w_{test} \| <} \theta \tag{15}$$

4 Emotional Speaker Recognition System

4.1 Corpus

Mandarin Affective Speech Corpus (MASC) [8], an emotional speech database, is used in our experiments. The corpus contains recordings of 68 Chinese speakers (23 female and 45 male). All speech is expressed in five kinds of emotion: neutral, anger, elation, panic and sadness. 20 utterances are spoken for three times under each emotional state, and 2 extra paragraphs for neutral. Utterance is short and the length is between 5s and 10s, while paragraph is longer relatively, with lasting about 40s length. All speech is recorded with the same microphone.

The development dataset is composed of the first 18 speakers' all paragraphs and utterances. The UBM model is trained using the paragraphs of all. This system uses 13 Mel Frequency Cepstral Coefficients (MFCC) with 32ms window length and 16ms frame rate. We use 512 Gaussian components for the background model.

In all our methods, no norm technique is applied.

4.2 Training EFA Model

The data for training eigenvoice matrix is the 18 speakers' all utterances. The eigenemotion matrix is composed of the four smaller eigenemotion matrixes under each emotional state. Each smaller eigenemotion matrix is trained independently containing 50 emotion factors.

The EFA algorithm is implemented by matlab on the basis of "Joint Factor Analysis Matlab Demo" [9]. This demo doesn't contain the second-order statistics or update the covariance matrix. We revised them and applied them into our experiment. The covariance is updated in solving the eigenvoice matrix and residual matrix, but not in solving the eigenemotion matrix in order to the keep the consistency of the combined eigenemotion matrix. Finally, when calculating the likelihood score, the linear scoring method is selected.

4.3 Training I-Vector Model

Training the I-Vector model is the same as training the eigenvoice matrix in EFA except that the utterance of the same speaker is treated as different one. The rank of the I-Vector is chosen to 300.In LDA emotion compensation method, the projection matrix is composed of the first 150 best eigenvectors.

4.4 Experiment Results

The IR result of each algorithm is shown in Table 1. It can be seemed that IR of EFA and I-Vector algorithm increases dramatically compared with the GMM-UBM system. The increase is relatively higher under anger and elation, and the increase in panic and sadness is smaller relatively. The performance under neutral is worse than the GMM-UBM, but the decrease is acceptable. The average increase of IR is 10.09%, 8.33% and 7.53% for EFA, I-Vector (WCCN) and I-Vector (LDA) respectively.

Table 1. Comparison of IR between these algorithms

emotion	GMM-UBM	EFA	I-Vector(WCCN)	I-Vector(LDA)
neutral	96.23%	93.07%	93.47%	93.8%
anger	31.50%	50.13%	48.83%	45.43%
elation	33.57%	51.33%	49.4%	48.67%
panic	35.00%	45.97%	42.77%	42.5%
sadness	61.43%	65.20%	64.93%	65.0%
average	51.55%	61.64%	59.88%	59.08%

EER (Equal Error Rate) is powerful in measuring the performance of ASV (Automatic Speaker Verification) task. The EER plot of our algorithm is show in Fig. 1. EER for each algorithm is 18.83%, 14.44%, 14.69%, and 15.28% for GMM-UBM, EFA, I-Vector (WCCN) and I-Vector (LDA) respectively. EFA has the best performance of all algorithms, and reduces the EER with 4.39%.

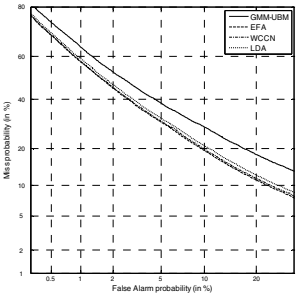


Fig. 1. DET plot of each algorithm

5 Conclusion

Emotion variability problem is a big challenge to the speaker recognition system. This paper borrows the ideas from the technique dealing the channel variability and proposes the EFA algorithm into solving the emotion-variability problem. Meanwhile, the session variability algorithm I-Vector is applied also. The experiment carried on MASC shows that these algorithms can bring a promising result compared with the GMM-UBM algorithm. The IR grows 7%~10%, and the EER reduces 3% ~ 5%. More effective session compensation algorithm will be applied in the future.

References

1. P. Kenny, G. Boulianne, P. Ouellet, P. Dumouchel. Joint factor analysis versus eigenchannels in speaker recognition. IEEE Transaction on Audio Speech and Language Processing, vol 15(4):1435-1447. May 2007.
2. P. Kenny, G. Boulianne, P. Ouellet, P. Dumouchel. Speaker and session variability in GMM-based speaker verification. IEEE Transactions on Audio, Speech and Language Processing, vol 15(4):1448-1460. May 2007.
3. Najim Dehak, Patrick Kenny, Reda Dehak, Pierre Dumouchel, Pierre Ouellet.Front-End Factor Analysis for Speaker Verification. IEEE Transaction on Audio, Speech and Language Processing. Vol 19(4):788-798. May 2011.
4. Najim Dehak, Reda Dehak, Patrick Kenny, Niko Brummer, Pierre Ouellet, Pierre Dumouchel. Support vector machine versus fast scoring in the low-dimensional total variability space for speaker verification. Interspeech, 1559-1562, 2009, Brighton.
5. H Bao, M Xu, F Zheng. Emotion Attribute Projection for Speaker Recognition on Emotional Speech[C].interspeech, pp.758-761, 2007.
6. E. Shriberg, S. Kajarekar, and N. Scheffer. Does Session Variability Compensation in Speaker Recognition Model Intrinsic Variation Under Mismatched Conditions?" Interspeech 2009, Brighton, United Kingdom, pp. 1551-1554
7. “The NIST year 2010 speaker recognition evaluation plan”.
8. T. Wu, Y. Yang, Z. Wu, D. Li, “MASC: A Speech Corpus in Mandarin for Emotion Analysis and Affective Speaker Recognition.” ODYSSEY 2006, pp.1-5, June 2006.
9. `http://speech.fit.vutbr.cz/en/software/joint-factor-analysis-matlab-demo`

Sub-band Main Peak Frequency Application for Speaker Identification

Limin Hou, Juanmin Xie, and Su Xie

School of Communication and Information Engineering, Shanghai University,
149 Yanchang Road, Shanghai, China
lmhou@staff.shu.edu.cn

Abstract. The paper proposes the sub-band main peak frequencies(SMPF) for speaker identification (SI). The SMPF could be derived from the sub-band first formant frequencies by all-pole model of speech signal. Compared with MFCC features for SI based on a Gaussian mixture model (GMM), only SMPF features for SI is better than only the MFCC, with one of improved relative rate up to 15%. Experimental utterances are Chinese mandarin under clean background recording circumstances.

Keywords: speaker identification, phase spectrum, all-pole model, MFCC.

1 Introduction

MFCC (Mel frequency cepstral coefficients) is a classical parameter used in speaker identification (SI)[1], that is calculated based on the amplitude of speech signal's spectrum. But with the development of speech signal application, researchers have realized that the ignoring of phase spectrum will cause great distortion, so the phase information's influence on the perception and comprehension of speech signal has draw more and more researcher's attention[2].

Prof. Kuldip K. P. [3, 4]has done much work in this field, which revealed that the speech signal cannot be understood when reconstructed only from the amplitude of frequency spectrum, but can be partly understood from the phase of frequency spectrum. But speech signal is reconstructed by group delay function (GDF) and instantaneous frequency (IF) distribution together[5]. Prof. Aarabi P. et al wrote a monograph[2] on the state of art of the phase spectrums of speech signal, that is reported research results about the important functions of the phase information in sound source location and speech recognition. Vibha Tiwari[6] evaluated speech recognition using phase spectrum is better than amplitude spectrum in the noise robust. Limin H. etal et al[7,8,9]have worked on IF of speech formants for SI and speech recognition. Marco G.[10] utilized Hilbert transformation and extracted IF in the whole frequency range as the discriminating speaker parameter, which performance of SI only using phase features is better than only using MFCC. Based on speech AM-FM (Amplitude Modulation, Frequency Modulation) model, Thiruvaran T.[11,12] put forward the frequency offset of FM in sub-band signal of

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 180–185, 2011.

speech are used as parameters for speaker verification, that sub-band IF obtained from all-pole model of the speech signal with equidistance filters in Mel domain and combined with MFCC.

Owing to speaker's characteristic should be consisted of the frequency offset and the central frequency in the sub-band frequency modulation. The frequency offset is the stable part of sub-band IF, which reflect the property of sound source such as glottal waveform; the central frequency of FM is the rapid episode of sub-band IF, which described the vocal tract' property. Both glottal and vocal tract map speaker individualities. This paper proposes the sub-band instantaneous frequency of the FM on sub-band's main peak (SMPF) is for differentiating features in SI. With GMM for speaker model, Our experiments show only the SMPF features for SI is better than only the MFCC.

2 SMPF Extracted by All-Pole Model

Given speech signal $x(n)$, its modulate relation with sub-band signal is described as equation (1) and (2):

$$\begin{aligned} x(n) &= \sum_{k=1}^{K} s_k(n) \\ &= \sum_{k=1}^{K} a_k(n)\cos(\phi_k(n)) \end{aligned} \tag{1}$$

$$s_k(n) = a_k(n)\cos\left[\frac{2\pi f_{ck} n}{f_s} + \frac{2\pi}{f_s}\sum_{m=1}^{n} q_k(m)\right] \tag{2}$$

Where K is the number of all sub-bands, k is the sub-band's order, $s_k(n)$ is the kth sub-band signal, $a_k(n)$, $\phi_k(n)$ and $q_k(n)$ respectively means its amplitude modulation, angle modulation and frequency modulation , f_s is sample frequency, f_{ck} is the central frequency of the kth sub-band.

In speech signal processing, in order to evaluate the frequency spectrum, the parameters of all pole model are calculated by linear prediction, and the order p of linear prediction is often chosen between 8 and 20. But it is not suitable for extracting SMPF, because generally speaking, the frequency spectrum only has one peak value in a sub-band, so it is reasonable to set p equals to 2. Therefore, sub-band signal $s_k[n]$ can be written as equation (3):

$$\begin{aligned} H_{sk}(z) &= \frac{G_k}{1 - \sum_{i=1}^{p} a_i z^{-i}} \\ &\overset{p=2}{=} \frac{G_k}{(1 - r_k e^{j\theta_k} z^{-1})(1 - r_k e^{-j\theta_k} z^{-1})} \end{aligned} \tag{3}$$

where, $\pm\theta_k$ and r_k are the phase angle and radius of one pole. The time domain response of $H_{sk}(Z)$ is shown as below:

$$h_{sk}(n) = \frac{G_k (r_k)^n \sin[(n+1)\theta_k]}{\sin\theta_k}, \quad n \geq 0 \tag{4}$$

equation (2) can be rewritten as:

$$s_k(n) \approx \frac{G_k (r_k)^n}{\sin\theta_k} \cos\left[(n+1)\theta_k - \frac{\pi}{2}\right] \tag{5}$$

compare equation (2) with equation (5), we may noticed:

$$a_k(n) = \frac{G_k (r_k)^n}{\sin\theta_k} \tag{6}$$

$$(n+1)\theta_k - \frac{\pi}{2} = \frac{2\pi f_{ck} n}{f_s} + \frac{2\pi}{f_s} \sum_{m=1}^{n} q_k(m) \tag{7}$$

as to equation (7), resort first order differential to *n*, then connect it with equation (2), the SMPF of FM of sub-band signal can be determined by f_k:

$$f_k(n) = \theta_k \frac{f_s}{2\pi} = q_k(n) + f_{ck} \tag{8}$$

Where the order of all pole model is set to *p*=2, The SMPF evaluated from equation (8) is the IF corresponding to sub-band main peak.

The extraction of SMPF contains two part, one is getting sub-band signal from band pass filters, the other is calculate SMPF from sub-band signal based on all pole model.

The extraction process of SMPF can be divided into three steps, the flow chart is shown in figure 1.

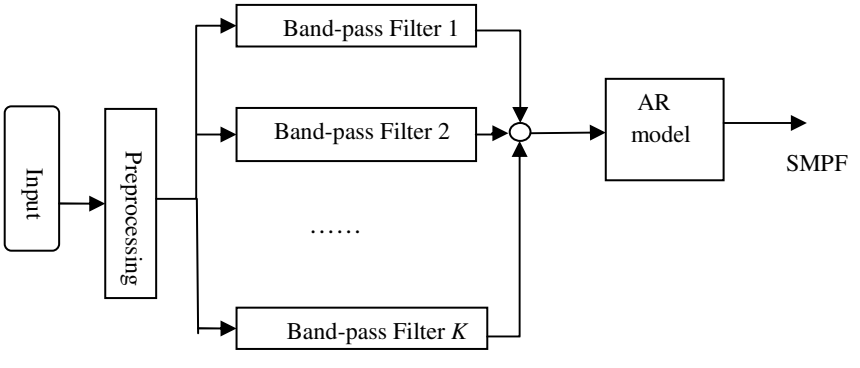


Fig. 1. Flow chart of SMPF extraction

Step 1: Preprocessing: remove silence, divided into frames and windowing.
Step 2: Passing the speech signal through k different band pass filters to get the sub-band signals. The band pass filters are Gabor filter with half bandwidth overlap, and its central frequency equally distribute along the whole frequency band.
Step 3: Calculate SMPF in each sub-band based on all pole model.

3 Speaker Identification by SMPF

3.1 Database

All the material comes from PKU-SRSC speech database. The 48 persons used for training for each person contains a short essay and a short free speech last 150 seconds. The test data are ten different digital series, each of them last 1 second. The speaker identification is text-independent. The speech signals were recorded by 8 kHz of sample frequency and 16bits quantization.

3.2 Experiment Result

The speaker models are got with a 64 order GMM system with frame length is 20ms, frame shift is 10ms, 48 persons participated in the experiment.

Under our speaker identification experiment platform, the accurate rate for MFCC is 0.692. The accurate rates for SMPF with different sub-band number are listed in table 1. The accurate rate only by SMPF is higher than MFCC on the whole, especially for the 24 sub-bands, the accurate rate improved by 15%.

Table 1. Accurate rate for SMPF and MFCC

Parameters	Accurate rate
MFCC	69.2%
SMPF-12(12 sub-band)	74.6%
SMPF-16(16 sub-band)	80.2%
SMPF-20(20 sub-band)	82.1%
SMPF-24(24 sub-band)	84.2%

Showed as table 1, the accurate rate of speaker identification can be improved with the increasing of sub-band number from 12 to 24 sub-band for SMPF.

Thiruvaran T.[7] utilized the frequency offset of FM in sub-band corresponding $q_k(n)$ in equation (8) based on all pole model, which called SFMO. Now, we compare the performance of SMPF and SFMO when the sub-band number is12, 16, 20 and 24, the experiment was carried out in Hz domain. The result is shown in table 2.

Table 2. Accurate rate of SFMO and SMPF

Parameters	12 sub-bands	16sub-bands	20sub-bands	24sub-bands
SFMO	74.6%	80.2%	82.1%	84.2%
SMPF	74.6%	80.2%	82.1%	84.2%

Table 2 illustrated that in Hz domain, for SFMO and SMPF, the accurate rate also has some relationship with sub-band number, and the same for those two parameters. Due to the frequency offset $q_k(n)$ taken by central frequency subtracting from SMPF in equation (8), the central frequency f_{ck} is constant in sub band.

4 Analysis and Conclusion

1) For speaker identification, SMPF features get better performance than MFCC. This proved that speaker's personality features are contained in voice signal's subtle phase spectrum. Such as vocal track eddy response in high frequency section, the information of vocal track length contains in middle frequency section, the modulation of the low frequency section close relationship with glottal wave and its differential, and so on. All of those have a close connection with speaker's characteristic. Sub-band IF described these information in detail, so it has a higher accurate rate than MFCC. Due to test speech duration very short, about 1second, the accurate rate for MFCC is only 69.2%.

2) SMPF gets the highest accurate rate when the sub-band number chosen between 20 and 24. Meanwhile, there is 1 to 3 harmonics exist in single sub-band. If the sub-band with AM-FM method, we will find it consistent with the mechanism of cochlear receiving voices. The optimum number of cochlear filters is 20 to 24, and the filter bank for MFCC also has 20 to 24 sub-bands, they all coincidence to human's sense of hearing.

3) SMPF is originated from the IF of the main peak in a sub-band, this related to the "phase synchronization" mechanism in hearing system. For low frequency band filter (lower than 1000Hz), the impulse of nerve fiber synchronize with the peak of sine wave. All pole model captured the main peak in each sub-band, which is the first resonant frequency of sub-band and the hearing system keep "phase synchronization" with it. So it is reasonable use the main peak value as the central frequency to extract corresponding SMPF.

4) Comparing experiment of SMPF and SFMO, we get the same conclusion. While the SFMO is $q(n)$ extracted by equation (8), that sub-band central frequencies are fixed, so both SFMO and SMPF are actually not different. The SMPF contains overall the frequency offset and central frequency, that is in accordance with modulation phenomenon of sound source and vocal track.

References

1. Reynolds, D.A., Rose, R.C.: Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models. IEEE Transactions Speech Audio Processing 3(1), 72–83 (1995)
2. Parham, A., Guangji, S., Maryam, M.S., Seyed, A.B.: Phase-based Speech Processing. World Scientific, USA (2006)

3. Kuldip, K.P., Leigh, D.A.: On the Usefulness of STFT phase spectrum in human listening test. Speech Communication 45, 153–170 (2005)
4. Leigh, D.A., Kuldip, K.P.: Iterative reconstruction of speech from short-time Fourier transform phase and magnitude spectra. Computer Speech and Language 21(1), 174–186 (2007)
5. Leigh, D.A., Kuldip, K.P.: Short-time phase spectrum in speech processing: A review and some experimental results. Digital Signal Processing: A Review Journal 17(3), 578–616 (2007)
6. Vibha, T., Jyoti, S.: AM-FM Features and Their Application to Noise Robust Speech Recognition: A Review. The IUP Journal of Telecommunications 2(1), 7–19 (2010)
7. Limin, H., Juanmin, X.: A New Approach to Extract Formant Instantaneous Characteristics for Speaker Identification. International Journal of Computer Information System and Management Applications 1, 295–302 (2009)
8. Limin, H., Juanmin, X.: Compensating function of Formant Instantaneous Characteristics in Speaker Identification. In: Fifth International Conference on Information Assurance and Security, IAS 2009, pp. 744–750 (2009)
9. Limin, H., Xiaoning, H., Juanmin, X.: Formant Instantaneous Characteristics application to Speech Recognition and Speaker Identification. Journal of Shanghai University 15(2), 123–127 (2011)
10. Marco, G., Fred, C.: Speaker Identification Using Instantaneous Frequencies. IEEE Trans. Speech and Language Processing 16(6), 1097–1111 (2008)
11. Thiruvaran, T., Ambikairajah, E., Epps, J.: Extraction of FM components from speech signals using all-pole model. Electronics Letters 44(6), 449–450 (2008)
12. Thiruvaran, T., Nosratighods, M., Ambikairajah, E., Epps, J.: Computationally efficient frame-averaged FM feature extraction for speaker recognition. Electronics Letters 45(6), 335–337 (2009)

Main Dialect Identification in Mainland China, Hong Kong and Taiwan

Dunxiao Wei[1], Jun-Yong Zhu[1], Wei-Shi Zheng[2], and Jianhuang Lai[2]

[1] School of Mathematics and Computational Science, Sun Yat-Sen University
Guangzhou, P.R. China
[2] School of Information Science and Technology, Sun Yat-Sen University
Guangzhou, P.R. China
dunxiao1985@163.com, jonesjunyong@gmail.com,
wszheng@ieee.org, stsljh@mail.sysu.edu.cn

Abstract. As an emerging field of speech recognition, dialect identification plays an important role for promoting applications of speech recognition technology. Since the communications among Mainland China, Hong Kong and Taiwan are becoming frequently, it is particularly necessary to identify their dialects. This paper makes contributions to this issue in the following three-folds: 1) we build a speech corpus for main dialects of the three areas; 2) we use the popular GMM based method to extensively evaluate the main dialects between Mainland China and Hong Kong and the ones between Mainland China and Taiwan, and we find the differences between Mainland China Mandarin and Taiwan Mandarin are much smaller than those between Mandarin and Cantonese, resulting in unsatisfactory results in the latter case; 3) we propose an improved method based on the analysis of GMM, namely, *maximum KL distance based Gaussian component selection* (MKLD-GCS) in order to improve the performance of dialect identification between Mainland China Mandarin and Taiwan Mandarin. Experimental results show that our proposed method obtains better identification performance than related methods.

Keywords: Dialect Identification, Speech Corpus, Gaussian Mixture Model.

1 Introduction

As an emerging field of speech recognition, dialect identification is important to advanced developments of speech recognition system and speaker recognition system. What's more, it is widely used in information service, public safety, criminal investigation, national defense and language engineering.

In this paper, we are interested in the Chinese dialect identification. Chinese dialects are also known as regional variations of Chinese language. Chinese dialect identification is a technology that automatically identifies the type of dialect of a unknown speaker and then distinguish which area the speaker belongs to. Since Mainland China, Hong Kong and Taiwan hold very important positions in China and

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 186–194, 2011.

communication among these three regions is more and more frequent, to identify their primary languages, i.e. Mainland China Mandarin, Cantonese and Taiwan Mandarin, is therefore very important.

Related works on Chinese dialect identification have been investigated for years, including works using vector quantization (VQ), hidden Markov model (HMM), Gaussian mixture model (GMM), support vector machine (SVM), artificial neural network (ANN), etc. Recently, Gu et al. set up a GMM-based dialect identification system for three different Chinese dialects [1], i.e. dialects of Wusi, Nantong, Muyang in Jiangsu province. A novel Chinese dialect identification method using clustered SVM was presented in [2].

In this paper, a corpus containing speeches of main dialects in Mainland China, Hong Kong and Taiwan is established. We will investigate main dialect identification in the three areas on the basis of this speech corpus. The remainder of this paper is organized as follows. Section 2 provides a detailed introduction to the speech corpus. Section 3 discusses the dialect identification in the three areas based on GMM and proposes an improved method. Section 4 presents a series of experimental results. Finally, Section 5 summarizes research findings.

2 Speech Corpus for Mainland China, Hong Kong and Taiwan

The main foreign corpora for dialect identification are MIAMI corpus [3], Arabic corpus, etc. However, there are limited to the domestic corpora for dialect identification, e.g., 863 regional accent speech corpus. As far as we know, there is no specific corpus for dialect identification among Mainland China, Hong Kong and Taiwan. So the first contribution of this paper is to build a speech corpus for the three areas above. In the following section, we are going to introduce the detail of this corpus.

2.1 Speech Acquisition Scheme

Speech Source: All the speech utterances in the corpus are collected from the people whose birthplaces and long-term living areas have been strictly checked. Each speech utterance is extracted from videos containing movies and teleplays of Mainland China, Hong Kong and Taiwan in recent ten years. For every movie or teleplay, we cut out speech data for several persons simultaneously. Note that, speech utterances of the same person obtained from different movies or teleplays are totally different. Each person has several speech utterances whose noises are relatively small.

Speech Content: Since all the speech utterances are from movies or teleplays, their content is basically concern about daily expressions, commercial languages, official languages, i.e., involving daily life, economy, politics, sports and other fields. We intend to make the speech content of this corpus have a more extensive coverage.

Table 1. The composition of the corpus (M: moive, T: teleplay)

	Mainland China Mandarin corpus		Hong Kong Cantonese corpus		Taiwan Mandarin corpus	
	Training set	Testing set	Training set	Testing set	Training set	Testing set
# of videos	7M, 6T	8M, 5T	7M, 5T	4M, 4T	9M, 10T	4M, 7T
# of speakers	30	32	30	34	30	32
# of speech utterances	181	267	171	251	169	274
Total duration (s)	1091	1414	1024	1347	1008	1425

Speech Naming Rule: Each speech utterance is named by four consecutive numbers. The first two numbers represent the person number and the last two numbers represent the number of speech utterances of a person, as 0507 represents the 8th speech utterance of the 6th person.

2.2 Composition of the Corpus

This corpus consists of three subsets, including corporas of Mainland China Mandarin, Hong Kong Cantonese and Taiwan Mandarin. The composition is listed as in Table 1.

The total training set of this corpus includes 90 speakers and its total duration is 0.87 hours, thus each person has an average of 35 seconds speech. The testing set contains another 90 speakers, and the amount of speech utterances and the total duration are 792 and 1.16 hours respectively. Each speaker has 8 speech utterances averagely, and the average length of each speech utterance is 5.29 seconds. In summary, the total length of the corpus is 2.03 hours.

As shown above, from the comparison of training sets or testing sets, we can find that the corpus is relatively balanced. Its sufficiency and balance guarantee the accuracy and validity of the experimental results in this paper.

3 Main Dialect Identification in Mainland China, Hong Kong and Taiwan

3.1 Gaussian Mixture Model (GMM) for Dialect Identification

Statistically speaking, the speech features of each dialect form a specific distribution that describes the characteristics of the dialect. Gaussian mixture model, the linear combination of several Gaussian distributions, can be used to approximately describe complicated probability distributions with well studied theoretical guarantee. Thus GMM is often adopted to effectively model the feature distribution of a dialect.

M-order GMM uses the linear combination of *M* Gaussian distributions to describe the distribution of speech features in feature space [4]

$$p(x \mid \lambda) = \sum_{i=1}^{M} w_i p_i(x) \tag{1}$$

where

$$p_i(x) = N(x, \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{P/2} \left|\Sigma_i\right|^{1/2}} \times \exp\{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)\} \tag{2}$$

and *M* is the order of GMM, w_i is the parameter weight, which satisfies $\sum_{i=1}^{M} w_i = 1$, *P* is the feature order. The kernel functionis $p_i(x)$ is Gaussian distribution function whose mean vector and covariance matrix are μ_i and Σ_i respectively. To deal with it more simply and more conformably, we assume that the forms of Gaussian distributions in GMMs are all the same and the task for dialect identification is to learn a set of parameters $\lambda = \{w_i, \mu_i, \Sigma_i \mid i = 1, 2, \cdots, M\}$ for each dialect through expectation maximization (EM) algorithm [5].

According to our observation, when directly using GMM, the accuracy of main dialect identification between Mainland China and Taiwan is much lower than that of main dialect identification between Mainland China and Hong Kong. Since Mandarin differs from Cantonese greatly, GMM can identify them effectively. Nevertheless, for the former situation, which is the identification between two different variations of Mandarin, since there is only little distinction between these two dialects, it is ineffective to distinguish them by GMM directly. Experimental proofs can be refered in the next section. Thus, to provide improvement for the second kind of dialect identification between Mainland China and Taiwan, we propose a new method.

3.2 MKLD-GCS Method for Dialect Identification

We assume that a GMM represents the speech feature space of a dialect, and each Gaussian component represents a region of the speech feature space [6]. Since both mainland Mandarin and Taiwan Mandarin are belong to the sub-langues of Mandarin, they are very similar, which implies that their speech feature spaces have a lot of overlapping regions. We can category these Gaussian components into two types. We call the components distributed in the overlapping region of the feature spaces of two dialects as the shared components, while the rest are called discriminant components. As we explored in the experiments, the shared components contribute little to dialect discrimination or even decrease the dialect identification performance, so identifying dicriminant components is very important for dialect identification.

Generally, we wish to find a distance measure between components such that the distance between any two shared components is small, while that between any two discriminant components is large. So in order to distinguish them, we need to introduce the measurement for calculating the distance between two components. To our best knowledge there are several popular techniques to do so, including Kullback-Leibler distance (KL distance), Pearson-χ^2 distance and total variance distance [7].

KL distance, for its effectiveness, is widely used to describe the distance between two distributions. Thus, we adopt it to evaluate the distance between two mixtures.

Assuming $f(x)$ and $g(x)$ are two probability density functions, KL distance from $f(x)$ to $g(x)$ is defined as [8]

$$D(f \| g) = \int f(x) \log \frac{f(x)}{g(x)} dx \tag{3}$$

KL distance satisfies many properties except symmetry. In order to overcome this issue, this paper utilizes maximum KL distance to describe the distance between two distributions. Maximum KL distance of $f(x)$ and $g(x)$ is defined as

$$D_m(f, g) = \max\{D(f \| g), D(g \| f)\} \tag{4}$$

Maximum KL distance is very effective but satisfies symmetry. For obtaining the maximum KL distance between two Gaussian components, we firstly compute the KL distances between them. The KL distance between two components $f(x)$ and $g(x)$ is as follows:

$$D(f \| g) = \frac{1}{2}\left[\log \frac{|\Sigma_g|}{|\Sigma_f|} + tr\left[\Sigma_g^{-1}\Sigma_f\right] - P + \left(\mu_f - \mu_g\right)^T \Sigma_g^{-1}\left(\mu_f - \mu_g\right) \right] \tag{5}$$

Since $\sum_i$ in a Gaussian model is generally assumed to be a diagonal matrix, the KL distance is simplified as

$$D(f \| g) = \frac{1}{2}\left[\sum_{i=1}^{n} (\log \sigma_{gi}^2 - \log \sigma_{fi}^2) + \sum_{i=1}^{n} \frac{\sigma_{fi}^2}{\sigma_{gi}^2} + \sum_{i=1}^{n} \frac{(\mu_{fi} - \mu_{gi})^2}{\sigma_{gi}^2} - P \right] \tag{6}$$

where $\{\sigma_{f1}^2, \sigma_{f2}^2, \cdots, \sigma_{fn}^2\}$ and $\{\sigma_{g1}^2, \sigma_{g2}^2, \cdots, \sigma_{gn}^2\}$ are variance vectors of $f(x)$ and $g(x)$. Using the above KL distance, we can calculate the maximum KL distance by (4).

By utilizing the maximum KL distance, we select discriminant Gaussian components from dialects which represented by GMMs according to distance matrix. First, a $M * M$ maximum KL distance matrix D is calculated, where D_{ij} denotes the maximum KL distance between the i^{th} component of GMM_A and the j^{th} component of GMM_B, besides, GMM_A and GMM_B here present two dialect models respectively. Since there could be no clear boundary between two GMMs, we need to partition different components by thresholding D. Generally, assume that values in D are sorted in ascending order and those within [0, r] are discarded. Note that, the threshold r is determined according to overall values in D. More intuitionally, since r would be vary with D, we adopted the ratio of small values in D to be discarded to be our threshold, denoted as $M_r (0 \le M_r < 1)$. In order to identify discriminant Gaussian components, the tail of D is first considered. Assume $D_{i_0 j_0}$ is the minimum value in D, we then remove the i_0^{th} component from GMM_A and the j_0^{th} component from GMM_B, the same to the second smallest value $D_{i_1 j_1}$. This process will be repeated until the proportion of small values removed from D comes up to M_r. The remaining

components make up the improved GMMs. We call this maximum KL distance based Gaussian component selection as MKLD_GCS.

Table 2. Experiment parameters

Parameters	values
Pre-emphasis	$H(z)=1-0.95\ z^{-1}$
Window type	Hamming window
Frame length	30 ms
Frame shift	50%
Feature type	MFCC+F0+△(MFCC+F0)
Feature order	26
GMM order	8、16、32、64、128、256

4 Experimental Results and Analysis

4.1 Comparison between Two Kinds of Dialect Identification

In this part, we performed experimental comparisons between two kinds of dialect identification and verified that the overlap between Mainland China Mandarin and Taiwan Mandarin in feature space would lead to performance degeneration. Experiments were conducted using GMM based on two pairs of corpus setup, Mainland Chain vs. Hong Kong and Mainland China vs. Taiwan. Parameters used in the experiments are listed in Table 2 and the results are illustrated in Fig.1.

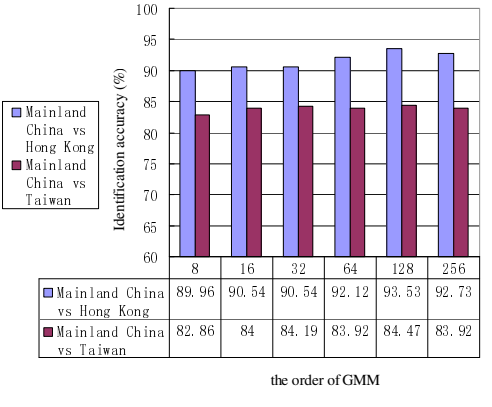


Fig. 1. Two kinds of dialect identification

As shown above, no matter how much the order of GMM we use, the accuracy of main dialect identification between Mainland China and Taiwan is much lower than that between Mainland China and Hong Kong.

4.2 The Second Kind of Dialect Identification Based on MKLD-GCS

For dialect identification based on MKLD-GCS, it is crucial to determine the value of M_r. So we discuss how the accuracy of the dialect identification between Mainland China and Taiwan based on MKLD-GCS changes with M_r.

From Fig.2, MKLD-GCS method can effectively improve the identification accuracy based on GMM. At the beginning as the ratio increases, more and more shared components are abandoned and the accuracy rises. After the ratio increases to certain value, the accuracy declines as the ratio increases, because at the moment more and more discriminative components are discarded. When M=128, the highest accuracy of MKLD-GCS method is 86.88% for M_r=12‰ with 2.41% improvement and 15.52% relative error reduction over GMM. In the case of M=256, we obtain the best performance when M_r=8‰ with 2.03% improvement and 12.62% relative error reduction over GMM. It suggests that MKLD-GCS method is effective and stable for main dialect identification between Mainland China and Taiwan.

4.3 Comparison for the Second Kind of Dialect Identification

In this part, we conducted comparisons between MKLD-GCS method and GMM method using different GMM orders and compared VQ method and HMM method to our proposed MKLD-GCS method under different feature types.

From Fig.3, Compared with traditional GMM method, identification accuracy based on MKLD-GCS method achieves 1.5%-2.5% improvement with the relative error reducing 8.5%-15.5%. It shows that our method is effective.

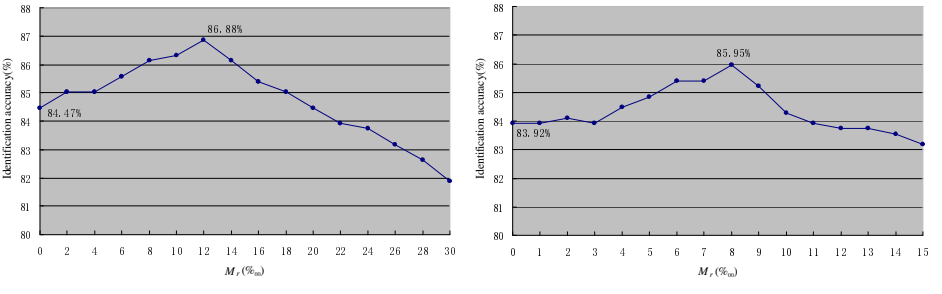


Fig. 2. Dialect identification based on MKLD-GCS (Left: M=128, Right: M=256)

In Fig.4, identification accuracy of three methods is obviously different. Regardless of which feature type we used, the accuracy based on MKLD-GCS method is higher than that based on the other two methods. Compared to VQ method, when using MFCC as feature, MKLD-GCS method has a 8.23% improvement with

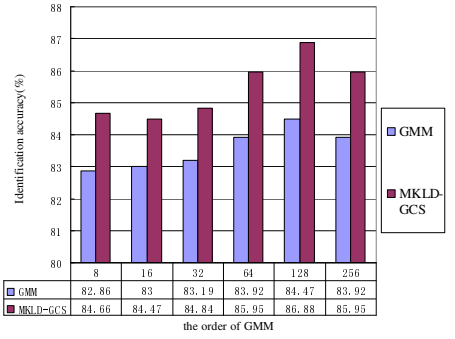


Fig. 3. Comparison of GMM and MKLD-GCS

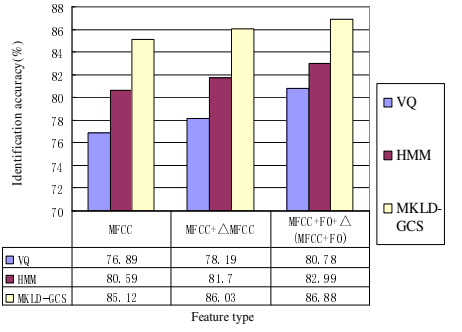


Fig. 4. Comparison of VQ, HMM and MKLD-GCS

a relative error reduction of 35.61%. While MFCC+F0+Δ(MFCC+F0) is applied, MKLD-GCS method has a 6.1% improvement and a relative error reduction of 31.74%. Obviously, MKLD-GCS performs much better than VQ method. Although HMM method is harder to implement, the accuracy based on it is not very high. For both types of feature, MKLD-GCS method has about 4% improvement and the relative error reduction is about 23% compared to the HMM method. This suggests that identification effect of MKLD-GCS method is some better than that of HMM method. The comparison of MKLD-GCS method and traditional method validates the superiority of MKLD-GCS method in dialect identification between Mainland China and Taiwan again.

5 Conclusion

In this paper, we build a speech corpus for main dialects of Mainland China, Hong Kong and Taiwan. And then we use the popular GMM based method to extensively evaluate the main dialects between Mainland China and Hong Kong and the ones between Mainland China and Taiwan. We find that unsatisfactory results occur in the latter case due to overlapping in feature space. Thus, we propose an improved method based on the analysis of GMM, namely, maximum KL distance based Gaussian components selection method (MKLD-GCS), in order to improve the performance of dialect identification between Mandarin in Mainland China and Taiwan.

Acknowledgments. This project was supported by the NSFC-GuangDong (U0835005) and the 985 project in Sun Yat-sen University with grant No. 35000-3181305.

References

1. Gu, M., Ma, Y.: GMM-based Chinese Dialect Identification System. J. Computer Engineering and Applications 3(43), 204–206 (2007)
2. Gu, M., Xia, Y.: Chinese Dialect Identification using Clustered Support Vector Machine. In: IEEE Int. Conference Neural Networks & Signal Processing, Zhenjiang, China (June 2008)

3. Zissman, M.A., Gleason, T.P.: Automatic Dialect Identification of Extemporaneous Conversational, Latin American Spanish speech. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 777–780 (1996)
4. Wu, Z., Yang, Y.: Models and Methods for Speaker Recognition, pp. 27–28. Tsinghua University Press, Beijing (2009)
5. Jelinek, F.: Statistical Methods for Speech Recognition. MIT Press, Massachusetts (1999)
6. Lei, Y., Hansen, J.H.L.: Dialect Classification via Text-Independent Training and Testing for Arabic, Spanish, and Chinese. IEEE Transactions on Audio, Speech, and Language Processing 19(1), 85–96 (2011)
7. Reiss, R.D.: Approximate Distributions of Order Statistics. Springer, New York (1980)
8. Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley & Sons, New York (1991)

Human Identification and Gender Recognition from Boxing

Jian Wang[1], Wuzhenni Hu[2], Zhiling Wang[1], and Zonghai Chen[1]

[1] Department of Automation
[2] Department of Computer Science,
University of Science and Technology of China, 230027, China
{wj0910,janehwzn}@mail.ustc.edu.cn, {zlwang3,chenzh}@ustc.edu.cn

Abstract. We describe an approach of human identification and gender classification based on boxing action. A period detection approach based on time-involved-cutting-plane is first applied and then a boxing sequence of a period is represented by an averaged silhouette. A Nearest Neighbor classifier based on Euclidian distance is used for human identification. The experiments were carried out on the KTH boxing dataset on which the accuracy can reach 80% or higher. After dimensionality reduction by PCA, a SVM is used for gender classification. The experimental results on a dataset containing 20 males and 20 females demonstrate that by applying the proposed algorithm the gender recognition can reach the accuracy of 80% or higher. We also present a numerical analysis of the contributions of different human components. Experimental results show that the head has a positive impact on system performance with the basis of the arm while the buttocks and the leg have not.

Keywords: biometrics, boxing, human identification, gender recognition, time involved cutting plane.

1 Introduction

Biometrics recognition is a technology to identify humans which takes advantage of individual physical or behavioral features. Most frequently-used behaviors in Biometrics are gait, keystroke/mouse behavior, and waving behavior [1], etc. We propose utilizing the boxing action to authenticate the identities of people. This method can be used in analyzing the fights in crime scenes, identifying criminals through their fighting behaviors. Mostly, if the identification database is very large, it is difficult to identify every single person merely using behavior information. Even though in this case, yet identification from boxing can still be used to narrow down the search scale of the target, for example, by criminal's sex and age.

In fact, one can identify a person through point-light motion of her/his boxing action. Point-light motion, proposed by Sweden psychologist Gunnar Johansson, is the movements of bulbs attached in a moving person's joints. The point-light motion is often used to analyze people's motion perception ability [2]. In 2005, Loula [3] did a psychological experiment: observers identified people (themselves, their acquaintances, and strangers) through the point-light motions of ten actions (dancing,

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 195–203, 2011.

boxing, hitting bags, playing table tennis on the wall, prancing in place, shaking and waving hands, hugging, whole body laughing, walking, and running). Fig.1A illustrates the identification accuracies vary across the ten actions from the self-, friend, stranger trials while Fig.1B illustrates the results with the stranger trials excluded. Observers' identification performance is best for dancing and boxing and relatively poor with walking and running. From Fig.1 it can be concluded that boxing action includes abundant information for identification.

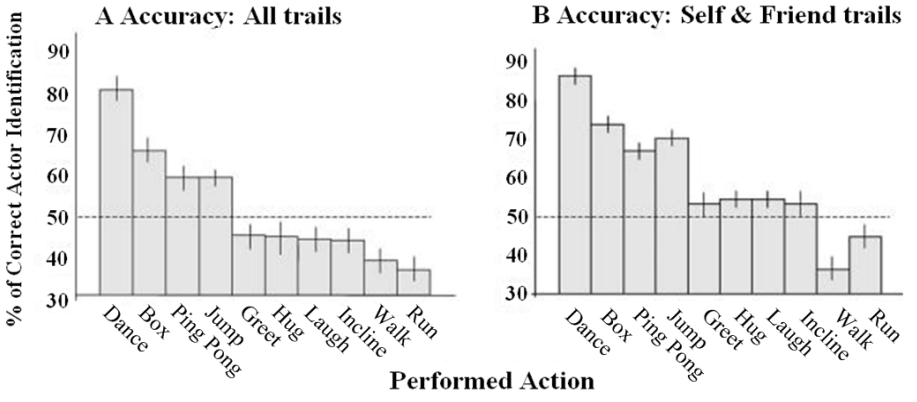


Fig. 1. Performance accuracy across ten actions. A: Performance of all trials including the self-, friend and stranger trials. B: Performance of self- and friend trials (reprinted from [3]).

From the view of biomechanics, human actions are synthesis of hundreds of muscles' and joints' motions. Lacquaniti et al. [4] found that the arm and wrist trajectories of a person during a task prove to be consistent from trial to trial, independent of movement velocity. And trajectories of different people are unique, such as in the aspects of magnitude, location and relative chronology. These trajectories are functions of muscles, skeleton [5, 6], and brain planning [7].

Anthropometry and psychology also give evidences motivating gender recognition from boxing. In anthropometry, the body size is classified as structure size and function size [8]. Structure size refers to the static body size. After height normalization, females' heads and trunks take a higher proportion of the body than males'. Also, female shoulders are narrower, arms and legs are shorter, and hips are broader in relation to the male body. The function size refers to how much room human body can reach. In most cases, females have less powerful arms than males. As a result of this, female function size of arms is smaller than males. According to psychology experiments, not only can people identify genders of individuals through point-light motion of gaits, but also through point-light motion of facial movement [9] and hand movement (including knocking, waving, lifting) [10]. Mather et al. [11] concluded from their experiments that humans identify genders from gait by the two facts that (1) males tend to swing their shoulders from side to side more than their hips and (2) females tend to swing their hips more than their shoulders.

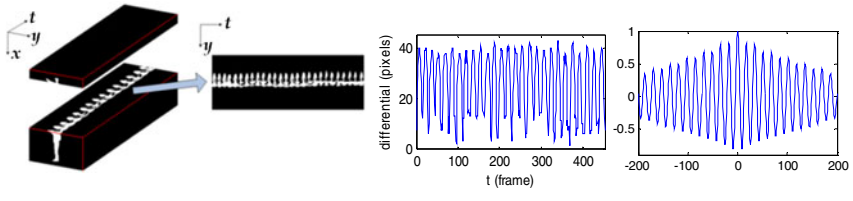
Research results in the aforementioned fields motivate utilizing boxing for behavioral biometrics. Though boxing actions may not be periodic motions at some

times, people will be asked to perform boxing multiple times in a repetitive manner in the experiments. We consider the boxing action as a periodic motion in this study like [12]. First we detect the period using a time-involved-cutting-plane based method. Then we use the Average Silhouette (AS) [13] to represent a boxing action sequence of a period. In the following, in order to analyze functions of different components of the body in identity and gender recognition, we partition the body as head, the part below the head and above hip (which is called the arm part in this paper), and hip and leg as in [14, 15]. For identification, we use Nearest Neighbor classifier based on Euclidian distance. For gender recognition, we use PCA to first reduce the dimensions, then apply SVM for classification. The rest of the paper is organized as follows. Section 2 introduces the period detection method. In Section 3 and 4 human identification and gender recognition from boxing action are studied, respectively. In Section 5, conclusion is given.

2 Period Detection

Assuming that the background of a video is known, and people box in place, we can get silhouettes through background subtraction. Many methods of period detection have been carried out in gait recognition, such as a method based on the number of foreground pixels from the bottom half of the silhouettes [13], the width/height ratio based method [16], etc. These methods are simple and efficient, but they cannot handle the occlusion problem. For that matter, we propose a period detection method which is based on time-involved-cutting-plane. As long as the boxing action itself is not occluded, its period can be detected.

Let H_1 be the maximum height of silhouettes in a boxing sequence and H_2 be the height where the maximum width occurs in the frame which has the maximum width. Considering the video as a cuboid, cutting it up at the height of H_2-αH_1 (α can take the value from the range 0.01~0.15; In Fig.2a it takes 0.07) along with the t-axis, we can find the texture of the boxing in the time-involved-cutting-plane and the period of the texture is twice the period of boxing. In order to detect the period of the texture, in this paper we 1) calculate the differentials between the uppermost and lowest nonzero pixels' y coordinate values (Fig.2b); 2) calculate the autocorrelation signal of the differential signal (Fig.2c); 3) calculate the first-order derivative to find the peak position of the autocorrelation by seeking the positive-to-negative zero-crossing points; and finally 4) estimate the real period as the average distance between each pair of consecutive peaks.



(a) (b) (c)

Fig. 2. Boxing period detection: (a) time-involved-cutting-plane, (b) differential signal, (c) autocorrelation signal

3 Human Identification

3.1 Design of Algorithm

Let S={S(1),…,S(*M*)}be the given silhouette sequence and *T* be its period. We partition S into subsequences of period length and for each subsequence compute the AS. Finally we can get AS(*i*), *i*=1,…, $\lfloor M / T \rfloor$, where $\mathrm{AS}(i) = \sum_{k=iT}^{(i+1)T-1} \mathrm{S}(k) / T$. The distance between two sequences S_P and S_G is defined by

$$\mathrm{Dist}(\mathrm{S_P}, \mathrm{S_G}) = \mathrm{Median}_{i=1}^{N_P} (\min_{j=1}^{N_G} \| \mathrm{AS_P}(i) - \mathrm{AS_G}(j) \|) \tag{1}$$

where $\mathrm{AS_P}(i)$, *i*=1,…, N_P is the *i*th AS of S_P and $\mathrm{AS_G}(j)$, *j*=1,…, N_G is the *j*th AS of S_G.

Note that when measuring the contributions of different human components we chose a template to pop out only the relevant components for measuring the distance by (1). As shown in Fig.3, we partition the human body into four parts: head (part A), arm (part B), hip (part C) and leg (part D). All ASes are normalized to the same height *H* by bilinearity interpolation. Note *a*~*e* in Fig.3 are parameters. Finally, the Nearest Neighbor classifier is used to do the classifying job.

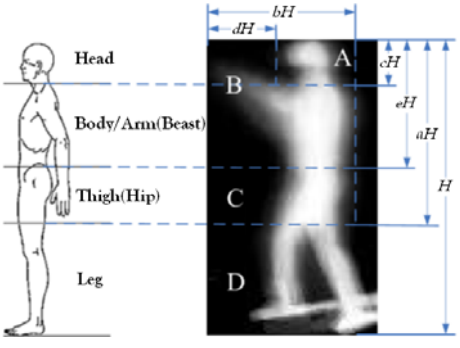


Fig. 3. Human body partition model

3.2 Identification Experiments

Experiments were done on the boxing part of the KTH human action dataset [17]. It contains six actions including boxing performed by 25 subjects at four different scenes including outdoors (Scene1) and indoors (Scene4). The frame rate is 25fps and the resolution is 120×160. Though the dataset does not provide a background model, yet in most scenes, people wore clothes with darker colors than backgrounds, so we extracted the foreground of a video by setting a threshold. There existed some slight tremble in camera when the KTH dataset was established. We applied a trivial translation and rotation to make up for this problem. The compensation values were exactly chosen for each frame to reach the maximum similarity to the former one in

the Euclidean distance sense. Finally, morphological operators such as erosion and dilation were used to eliminate the noise and fill the small holes. We chose 20 people (person 05,08,20,22,23 were excluded) in Scene1 and Scene4 for experiments. Some sample images in the dataset are illustrated in Fig.4. In some situations, there existed strong shadows in the outdoor scenes. These were wiped out manually after the ASes were generated. Fig.5 illustrates some ASes of the subjects. We can find out that there is much similarity between the ASes of the same subject in different scenes, and there are relatively many differences among different subjects.

ASes were partitioned in the way mentioned in Section 3.1 and parameters *a~e* were set to 0.63,0.5,0.15,0.23,0.43, respectively. The accuracies of using different combinations of components for identification are shown in Tabel 1. There are two accuracies in one combination: the first one is the InIn accuracy like in [1] (indoor-indoor, the indoor videos were divided into two parts, one part as the gallery data and the other as the probe data), and the second one is the InOut accuracy (indoor-outdoor).

If the recognition rate is significantly increased after adding a component with the basis of the arm component, the component is deemed to have a positive effect and is marked as "+". On the contrary, if the recognition rate is significantly reduced after adding a component, the component is deemed to have a negative effect and is marked as "-". Otherwise, the component is deemed to have little effect and is remarked as "N" like [14]. From Tabel 1, we can conclude that contributions of "1", "3", "4" are "+", "N", "+" (InIn) and "+", "+", "-" (InOut) respectively. The best combinations of three components used for identification are (2,1,4) and (2,3,4) in the case of InIn, and (2,1,3) in the case of InOut. We can see that the contribution of leg is "+" in some situations, but "-" in some other cases. For that matter, we suggest to wipe out the leg component or give it a small weight like [15].

Fig. 4. Sample images from the KTH dataset. The left three images were in Scene1 while the right three images were the same subjects in Scene4.

Fig. 5. ASes of boxing. The left nine images were in Scene1 while the right nine images were the same subjects in Scene4.

Table 1. Accuracies of identification using different combinations of components

Experiment No.	Combination type	Accuracy	Experiment No.	Combination type	Accuracy
1	(B)	90%, 80%	5	(B,A,C)	95%, 90%
2	(B,A)	95%, 90%	6	(B,A,D)	100%, 70%
3	(B,C)	90%, 85%	7	(B,C,D)	100%, 75%
4	(B,D)	100%, 60%	8	(B,A,C,D)	100%, 75%

4 Gender Recognition

4.1 Design of Algorithm

Familiar with identity recognition, we perform period at first, and then we calculate the ASes. We use SVM for the classification. Before classification, we chose the component combination from ASes and utilized PCA to reduce dimensions.

4.2 Gender Recognition Experiments

There is not much data concerning females in the KTH boxing dataset. So we collected data of boxing actions of 20 males and 20 females as the experiment dataset[1]. A digital camera (Sony W300) fixed on a tripod was used to capture boxing sequences indoors. The subjects boxed freely for more than seven periods in the lateral view with respect to the image plane. These boxing sequences were captured at a rate of 25fps and the resolution is 240×320. Some sample images are shown in Fig.6.

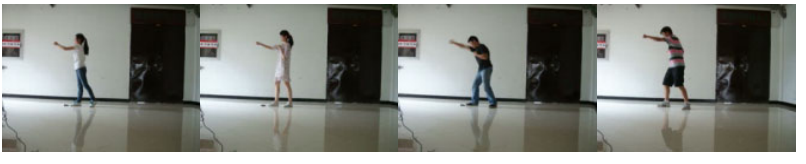


Fig. 6. Sample images from the boxing dataset

We first calculated the ASes of normalized height for every sequence (Fig.7a). Then we calculated the mean image of ASes of all females (Fig.7b) and males (Fig.7c), respectively. Fig.7b and Fig.7c show that there exists following major differences for males and females: 1) males are more inclined to lean forward than females; 2) males have a wider motion area than females do; 3) females' boxing actions are more horizontal than males'; 4) females' boxing action trajectories are more consistent than males'; 5) males have smaller heads than females do; 6) males have thicker sides than females do.

(a) (b) (c)

Fig. 7. (a) The left four images are female ASes while the right four images are male ASes. (b) The mean image of all female ASes. (c) The mean image of all male ASes.

[1] http://home.ustc.edu.cn/~wj0910/Boxing.htm

The human body was partitioned into four components as parameters *a*~*e* in Section 3.1 were set to 0.63,0.5,0.17,0.23,0.39 respectively. The experiment design was shown in Table 1.

We divided the dataset into four disjoint sets and the data in each set were from five males and five females. Three of the sets were taken as training data and the remaining one set as testing data. The correct classification rate was the average of all these four combinations. PCA retained 90% of the information. The performance of different combinations of body parts and different kernel functions (Linear kernel function, Polynomial kernel function and Sigmoid kernel function took default parameters in LIBSVM [18], while the RBF kernel function was configured through grid-search) is shown in Fig.8. We can see in Fig.8 that, except for the Polynomial kernel function, the kernel functions share the same tendency of accuracy while the combination of body parts is changed. So we use the tendency of these three kernel functions for analysis. When we recognize gender only by using the arm component, the accuracy can reach 80%. If the RBF kernel function is used, the accuracy can reach 90%. If the combination has two components, the head-arm combination gets the highest precision which reaches to 83%, and 93% if the RBF kernel function is used. The lowest precision is observed in the arm-hip combination which achieved only 75% or lower. There is not much difference between three-components combinations in recognizing genders. The accuracies of RBF kernel function are equal to each other and lead to the highest accuracy in all the experiments. According to the experimental results, to identify genders only by the arms data is feasible, and it will be better if information of heads is added. However, hips serve little use in recognizing genders. Matter of fact, the hips data can even reduce system performance in some situations. When it comes to legs, they effect little in performance in our experiments.

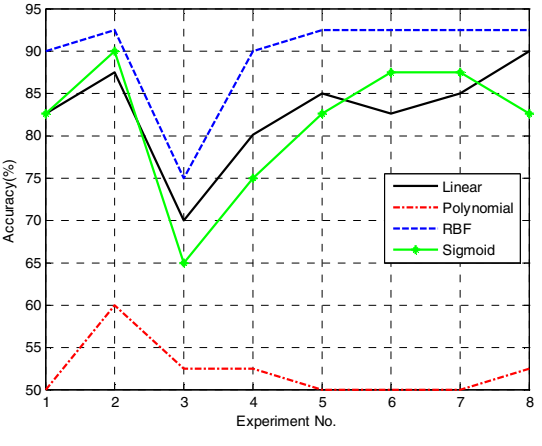


Fig. 8. Performance of gender recognition across different combinations of body components and different kernel functions

5 Conclusions

The most frequently used recognition method in smart video surveillance nowadays is video facial recognition. However, if faces cannot be obtained, for example, when robbers were robbing a bank their faces were covered in most cases. Their facial features are not available. Instead, identification of people and gender may be done based on their movement. Boxing is one example for movements and may be considered as a benchmark case for general biometrics/motion-based identification methods. We use this benchmark case for our experiments. The results show that boxing can be used for identification and gender recognition. In such general cases, it can be a crucial complement to traditional biometric schemes. We also analyzed the contributions of different components of the body. After combining experimental results on two datasets, we can conclude that: it is feasible to do recognizing work with the use of only arms motion; head data is helpful for recognition, and information from hips and legs is of little use. Though the method based on AS has high accuracy, yet AS misses temporal information such as speed information. What we still need to spare efforts focuses on developing feature extraction algorithms which have more judging abilities in identities and genders than what we have now, and testing algorithm based on larger scale datasets including datasets of different views. What we also need to consider is combining different kinds of existing biological features to improve system performance.

Acknowledgments. The authors would like to thank those who have contributed to the establishment of the boxing database for gender recognition and the National Natural Science Foundation of China (Grant No. 61005091) and the Specialized Research Fund for the Doctoral Program of Higher Education (Grant No. 20093402110014) for supporting this study.

References

1. Pratheepan, Y., Condell, J., Prasad, G.: Individual Identification from Video Based on "Behavioural Biometrics". In: Wang, L., Geng, X. (eds.) Behavioral Biometrics for Human Identification: Intelligent Applications 2010, pp. 75–100. Medical Information Science Reference, New York (2010)
2. Jiang, Y., Wang, L.: Biological Motion Perception: The Roles of Global Configuration and Local Motion. Advances in Psychological Science 19, 301–311 (2011) (in Chinese)
3. Loula, F., Prasad, S., Harber, K., Shiffrar, M.: Recognizing people from their movement. Journal of Experimental Psychology: Human Perception and Performance 31, 210–220 (2005)
4. Lacquaniti, F., Soechting, J.: Coordination of arm and wrist motion during a reaching task. The Journal of Neuroscience 2, 399–408 (1982)
5. Atkeson, C.G., Hollerbach, J.M.: Kinematic features of unrestrained vertical arm movements. The Journal of Neuroscience 5, 2318–2330 (1985)
6. Sergio, L.E., Ostry, D.J.: Coordination of multiple muscles in two degree of freedom elbow movements. Experimental Brain Research 105, 123–137 (1995)

7. Desmurget, M., Pélisson, D., Rossetti, Y., Prablanc, C.: From eye to hand: planning goal-directed movements. Neuroscience & Biobehavioral Reviews 22, 761–788 (1998)
8. Zhang, Y.: Interior ergonomics, pp. 8–30. China Architecture and Building Press, Beijing (1999) (in Chinese)
9. Hill, H., Johnston, A.: Categorizing sex and identity from the biological motion of faces. Current Biology 11, 880–885 (2001)
10. Pollick, F.E., Lestou, V., Ryu, J., Cho, S.B.: Estimating the efficiency of recognizing gender and affect from biological motion. Vision Research 42, 2345–2355 (2002)
11. Mather, G., Murdoch, L.: Gender discrimination in biological motion displays based on dynamic cues. In: Proceedings: Biological Sciences, pp. 273–279. The Royal Society, London (1994)
12. Wang, L., Suter, D.: Learning and matching of dynamic shape manifolds for human action recognition. IEEE Transactions on Image Processing 16, 1646–1661 (2007)
13. Liu, Z., Sarkar, S.: Simplest representation yet for gait recognition: Averaged silhouette. In: 17th International Conference on Pattern Recognition, pp. 211–214. IEEE Press, New York (2004)
14. Li, X., Maybank, S.J., Yan, S., Tao, D., Xu, D.: Gait components and their application to gender recognition. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 38, 145–155 (2008)
15. Yu, S., Tan, T., Huang, K., Jia, K., Wu, X.: A study on gait-based gender classification. IEEE Transactions on Image Processing 18, 1905–1910 (2009)
16. Wang, L., Tan, T., Ning, H., Hu, W.: Silhouette analysis based gait recognition for human identification. IEEE Transactions on Pattern Analysis and Machine Intelligence 25, 1505–1518 (2003)
17. Schuldt, C., Laptev, I., Caputo, B.: Recognizing human actions: A local SVM approach. In: International Conference on Pattern Recognition, pp. 32–36. IEEE Press, New York (2004)
18. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 27, 27:1–27:27 (2011)

Learning Gabor Features for Facial Age Estimation

Cuixian Chen[1], Wankou Yang[1], Yishi Wang[1], Shiguang Shan[2], and Karl Ricanek[1]

[1] University of North Carolina Wilmington, USA
[2] Institute of Computing Technology, Chinese Academy of Sciences, China
{chenc,wangy,ricanekk}@uncw.edu, wankou_yang@yahoo.com.cn,
sgshan@ict.ac.cn

Abstract. In this work we aim to study rigorously the facial age estimation in a multiethnic environment with 39 possible combination of four feature normalization methods, two simple feature fusion methods, two feature selection methods, and three face representation methods as Gabor, AAM and LBP. First, Gabor feature is extracted as facial representation for age estimation. Inspired by [3], we further fuse the global Active Appearance Model (AAM) and the local Gabor features as the representation of faces. Combining with feature selection schemes such as Least Angle Regression (LAR) and sequential selection, an advanced age estimation system is proposed on the fused features. Systematic comparative of 39 experiments demonstrate that (1) As a single facial representation, Gabor features surprisedly outperform LBP features or even AAM features. (2) With global/local feature fusion scheme, fused Gabor and AAM or fused LBP and AAM features can achieve significant improvement in age estimation over single feature representation alone.

1 Introduction

A lot of useful information can be revealed from human faces including gender, race, mood, and age [1], and these information, especially age, could benefit the study of human computer interaction (HCI), surveillance monitoring, and biometrics. Face age estimation is still a challenging topic because of the different aging patterns among individuals due to genetic difference, behavior and environmental factors, etc. From the perspective of quantitative analysis, one of the difficulties in face age estimation is the representations and dimensionality of human faces. More specifically, the question is may we find an effective and efficient representations of human faces for the age estimation task.

An age estimation system usually consists of two modules [6]: feature representation and quantitative estimation. Existing facial feature representation methods for human age estimation mainly include anthropometric model [12], active appearance model (AAM) [14], aging pattern subspace [9], aging manifold analysis [7], part-based [19] and patch based appearance model [22], local binary patterns (LBPs) [23], Gabor wavelets [8] and bio-inspired features (BIFs) [11]. Given a feature representation, quantitative age estimation can be implemented by using a multiclass classification [13], regression method [7], or hybrid of both approaches [10]. One of the key challenges of face recognition is to find an efficient and discriminative facial appearance descriptor that can counteract large variations in illumination, pose, facial expression, ageing, partial occlusions and other changes [24].

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 204–213, 2011.

1.1 Prior Work

The Active Appearance Model (AAM) is proposed in [4] that described a statistical model of face shape and texture. It is a popular facial descriptor which makes use of Principle Components Analysis (PCA) in a multi-factored way for dimension reduction while maintaining important structure (shape) and texture elements of face images. As pointed by Mark [16], shapes are accounted for the major changes during ones younger years, while wrinkles and other textural pattern variations are more prominent during ones older years . Since AAM extracts both shape and texture facial features, it is appropriate to use AAM in the age estimation system for feature acquisition. However, the adoption of PCA's in AAM can muddle important features because it attempts to maintain the greatest variance while creating orthogonal-projection vectors.

It has been shown that local features can be more robust against small misalignment, variation in pose and lightings [10]. On the other hand, the Gabor operator is a popular local feature-based descriptor due to its robustness against variation in pose or illumination than holistic methods. It has been widely applied in many computer vision applications such as age estimation task [8]. Therefore, applying Gabor on the shape-normalized patch may take both advantages of shape model and local features.

Each feature representation has its advantages and disadvantages, so does the facial representation from either AAM or from Gabor, which has its inherent strengths, and also its limitation and weakness. The feature fusion scheme typically achieves boosted system performance or robustness [5]. Therefore, fusing two feature representation with model selection could be a better way to get an effective age estimation system [3]. Hence, the fusion of global and local facial features are investigated in this study. LBP encodes the fine details of facial appearance and texture while the Gabor features encode facial shape and appearance information over a range of coarser scales [21]. Both representations are rich in information and computationally efficient. Their complementary nature makes them good candidates for fusion [20]. [21] fused the normalized Gabor and LBP features at the feature level for age categorization and showed its effectiveness and robustness. It is proposed in [3] to fuse the global facial feature extracted from Active Appearance Model (AAM) and the local facial features extracted from Local Binary Pattern (LBP), as the representation of faces. Model selection schemes such as Least Angle Regression (LAR) and sequential approaches for age estimation were applied afterwards. In addition, they compare multiple normalization schemes for both facial features under the consideration of fact that different facial feature representations may come with various types of measurement scales. They demonstrated that the feature fusion with model selection can achieve significant improvement in age estimation over single feature representation alone. It is considered in [5] the multiple feature fusion as a general subspace learning problem. The objective of the framework is to find a general linear subspace in which the cumulative pairwise canonical correlation between every pair of feature sets is maximized after the dimension normalization and subspace projection. The learned subspace couples dimensionality reduction and feature fusion together, which can be applied to both unsupervised and supervised learning cases. Gao and Ai [8] considered Gabor feature as face representation and used fuzzy LDA for age group classification on a large consumer image dataset. Comparative

experiments showed that Gabor feature outperforms other features like pixel intensity and LBP, and the fuzzy LDA can further improve the classification accuracy.

1.2 Contribution of Work

In this work, we examine the effectiveness of Gabor feature for age regression under the schemes of single feature and feature fusion with AAM, in order to improve age estimation performance. We systematically compare single face representation methods such as AAM, LBP, and Gabor. We will also compare all possible fusion combination of AAM and LBP, AAM and Gabor, and, LBP and Gabor. However, detailed results for fused LBP and Gabor did not report in this paper due to under performance than other two fusion methods. Our experiment results suggest that as single face representation for age estimation task, Gabor feature outperforms LBP and even AAM. Furthermore, feature fusion based on local feature of Gabor or LBP with global feature AAM achieves better accuracy compared with single feature representation consistently, which demonstrates the need to perform feature selection for the fused features.

In this work, we propose to use the feature selection: LAR and Sequential selection methods on the fused features, which produces an effective and computational efficient age estimation system. This work also investigates different normalization methods on single/fused facial feature representation to further improve the performances. We evaluate our approaches for age estimation with the multi-ethnicity PAL image database. The organization of this paper is laid out as follows: Section 2 presents the Gabor feature extraction. The experiment results on proposed approaches are presented in Section 3; and conclusions are drawn in final section of this paper.

2 Gabor Feature

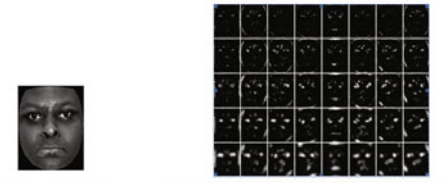
Gabor feature is a popular face representation due to its effectiveness in many research fields such as face recognition [15] and face age classification [8]. The Gabor wavelets, whose kernels are similar to the 2D receptive field profiles of the mammalian cortical simple cells, exhibit desirable characteristics of spatial locality and orientation selectivity, and are optimally localized in the space and frequency domains. The Gabor wavelet can be defined as follows:

$$\psi_{\mu,\nu}(z) = \frac{||k_{\mu,\nu}||^2}{\sigma^2} e^{\frac{-||k_{\mu,\nu}||^2||z||^2}{2\sigma^2}} [e^{ik_{\mu,\nu}z} - e^{-\sigma^2/2}], \tag{1}$$

where μ and ν define the orientation and scale of the Gabor kernels, and $z = (x, y)$, $||\cdot||$ denotes the norm operator. The wave vector $k_{\mu,\nu}$ is defined as follows:

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu}, \tag{2}$$

where $k_\nu = k_{\max}/f^\nu$, and $\phi_\mu = \mu(\pi/8)$. $k_{\max}$ is the maximum frequency, and $f = \sqrt{2}$ is the spacing factor between kernels in the frequency domain. In the most cases one would use Gabor wavelet of five different scales, $\nu = \{0, \cdots, 4\}$, and eight orientations, $\mu = \{0, \cdots, 7\}$.



(a) A sample image (b) The corresponding Gabor feature representation

Fig. 1. Examples of Gabor Feature Representation

Table 1. MAEs of different normalization methods on single feature representation on the PAL database. Note: Type 1 means no scaling; Type 2 means to use MinMax to map each covariate into range[0, 1]; Type 3 means to use MinMax to map each covariate into range[-1, 1]; Type 4 means to standardize each covariate into a vector with mean 0 and unit variance; Type 5 means to normalize each covariate into a vector with mean 0 and unit length.

MAEs (year) of normalized Gabor features on PAL				
	$Gabor^1$	$Gabor^{2,3}$	$Gabor^4$	$Gabor^5$
MAE.	**6.38**	7.54	6.98	7.05
Std.	**0.81**	0.49	0.71	0.75
#-Var	**41**	40	48	47
Total-Var	**539**	539	539	539

3 Experiment

In this section we shall systematically evaluate the effectiveness of applying Gabor features or fused Gabor and AAM features with feature selection methods.

3.1 Face Aging Databases

The UIUC Productivity Aging Laboratory (PAL) face database [17] is selected for this experiment due to its quality of images and diversity of ancestry. Only the frontal images with neutral facial expression are selected for our age estimation algorithm. It contains 540 images with ages ranging from 18 to 93 years old. It is worth mentioning that PAL contains adult face images of African-American, Asian, Caucasian, Hispanic and Indian.

3.2 Performance Measure

The performance of age estimation is measured by the mean absolute error (MAE) and the cumulative score (CS). The MAE is defined as the average of the absolute errors between the estimated ages and the observed ages, i.e., $MAE = \sum_{i=1}^{N} |\hat{y}_i - y_i|/N$, where $\hat{y}_i$ is the estimated age for the $i-$th test image, y_i is the corresponding observed age, and N is the total number of test images. The cumulative score is defined as the proportion of test images such that the absolute error is not higher than an integer j, $CS(j) = \sum_{i=1}^{N} I(|\hat{y}_i - y_i| \leq j)/N$ where $I(A)$ is an indicator function such that when an event A is true, $I(A) = 1$; and 0 otherwise.

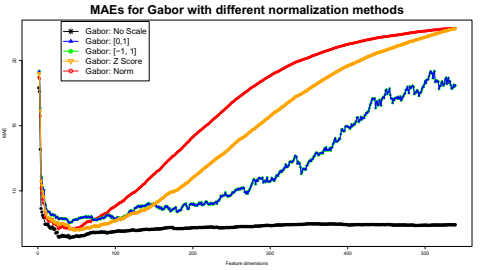


Fig. 2. MAE curves for Gabor feature on PAL. Note that curves of Gabor: [0,1] and Gabor: [-1, 1] are overlapped.

Table 2. MAEs (year) of different algorithms on the PAL database. Note: a represents AAM with no scaling, and g^i represents $Gabor^i$ respectively (see definitions and details in Table 1). Furthermore, ag^iL means to use fusion of AAM and Gabor with LAR algorithm for model selection; ag^iS means to concatenate AAM and Gabor features into a vector and then use sequential selection; g^iaS means to concatenate Gabor and AAM features into a vector and then use sequential selection.

MAEs (year) of different feature fusion and model selection algorithms on PAL												
	ag^2L	ag^3L	ag^4L	ag^5L	ag^2S	ag^3S	ag^4S	ag^5S	g^2aS	g^3aS	g^4aS	g^5aS
MAE.	**6.11**	6.46	9.04	6.51	**5.69**	5.96	7.71	6.17	**6.82**	7.53	6.98	7.00
SE.	**0.72**	0.68	0.89	0.77	**0.66**	0.72	0.62	0.50	**0.73**	0.48	0.70	0.94
#-Var	**32**	31	35	477	**257**	255	270	239	**48**	40	48	759
Total-Var	**480**	480	480	480	**769**	769	769	769	**769**	769	769	769

3.3 Experiment Setups

In PAL database, each image is annotated with 161 landmarks as shown in [18]. The annotated faces with shape and texture information are presented to the AAM system to obtain the encoded appearance features, a set of transformed features with dimension size 230. Meanwhile the shape-free patch is also extracted from the annotated faces via the Active Shape Model provided by the AAM-Library tool. Next, Gabor wavelet transformation is applied on each shape-free image with 5 scales and 8 directions. In the end, PCA is applied to the Gabor wavelet features to obtain a Gabor feature vector with dimension size 539.

Due to the fact that different facial feature representations may come with different types of measurement scales, we need to consider how to find a proper way to normalize global/local feature to build an effective age estimation system, for either single representation or fusion of both representation. We consider four different mappings as [3]: Min-Max-[0, 1], Min-Max-[-1,1], Z-score standardization, and Normalization methods.

If two feature representations extracted from the face images are (somewhat) independent to each other, it is reasonable to simply concatenate the two vectors into a single new vector, provided both global/local features are in the same type of measurement

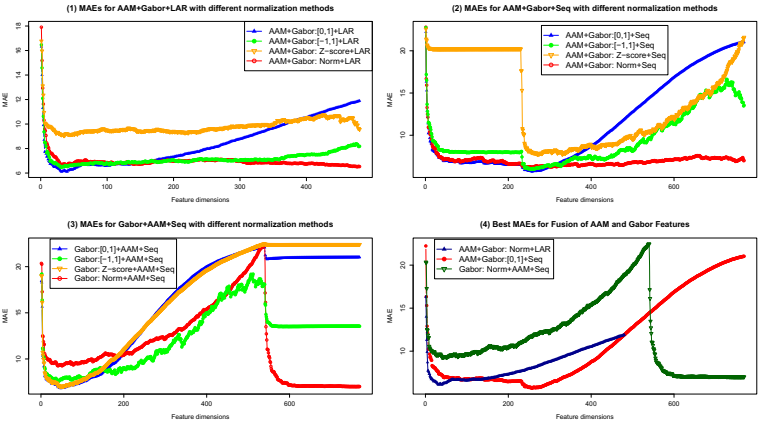


Fig. 3. MAE curves verse number of parameters used in the regression models on PAL database

Table 3. MAEs (year) of different algorithms on the PAL database. Note: a represents AAM with no scaling, and g^i represents $Gabor^i$ respectively (see definitions and details in Table 2). Furthermore, ag^iL means to use fusion of AAM and Gabor with LAR algorithm for model selection; ag^iS means to concatenate AAM and Gabor features into a vector and then use sequential selection; g^iaS means to concatenate Gabor and AAM features into a vector and then use sequential selection. Similarly, b^i represents LBP^i (see [3] for details).

Summary to MAEs (year) of normalized global/local feature and feature fusion on PAL database									
	$Gabor^1$	AAM^1 [3]	$LBP^{2,3}$ [3]	ag^2L	ag^2S	g^2aS	ab^2L[3]	ab^2S[3]	b^5aS[3]
MAE.	**6.38**	6.47	7.70	6.11	**5.69**	6.82	6.18	**5.65**	6.17
Std.	**0.81**	0.69	0.61	0.72	**0.66**	0.73	0.72	**0.68**	0.97
#-Var	**41**	200	80	32	**257**	48	116	**265**	348
Total-Var	**539**	230	150	480	**769**	769	380	**380**	380

scale. However, due to the fact that AAM features and Gabor features are representations to the same face, both feature vectors may have correlation at certain level. It becomes prominent to adopt a proper model selection technique which can be employed to extract a reasonable number of salient features from the larger set of candidates, and partially solve the correlation problem.

Inspired by [3], LAR is selected as one of two model selection techniques in this work. For all approaches, we use SVR as the age estimation regressor. We perform a standard 10-fold cross validation to evaluate the prediction error of the proposed normalization, fusion and model selection approaches. In order to draw a reliable conclusion for model comparison, we set up a fixed random seed in order to get the consistent partition for 10-fold cross-validation. We use the contributed package "Libsvm" [2] in Matlab for the computation of SVR. We use default parameters from Libsvm unless otherwise mentioned.

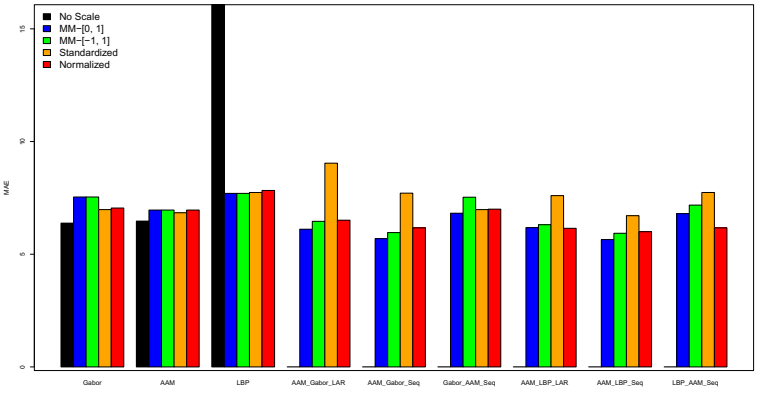


Fig. 4. Systematic performance comparison on age estimation for PAL Database

3.4 Experiment Results

In this work we systematically evaluate the performances of a total 39 different combinations of four feature normalization methods, with two simple feature fusion methods, and with face representation of Gabor, AAM, and LBP.

First, we compare four normalization methods with no-scaling on Gabor feature alone for age estimation with sequential selection. The experiment results are shown in Table 1 and Figure 2. One interesting finding is that for both AAM features and Gabor features, no-scaling turns out to achieve the best MAE, comparing to the rest normalization methods. Note that for both AAM and Gabor single features, the Min-Max-[0,1] and Min-Max-[-1,1] share exactly the same results, with distinct hyper-parameters for SVR. Consequently, the age estimation results for Gabor features with Min-Max-[0,1] and Min-Max-[-1,1] overlaps in Figure 2. It is surprised to find out Gabor features with no-scaling even outperforms both AAM and LBP features. It suggests that with single facial feature representation, Gabor can be potential facial feature representations for age estimation task. However, preliminary study shows that feature fusion with no-scaling AAM and no-scaling Gabor features produces under performance results comparing to the rest four normalization methods. Hereafter, we only adopt the *original* no-scaling AAM feature for further feature fusion studies.

Next, we systematically compare three possible combinations of two feature fusion methods, two feature selection methods, and three face representation of Gabor, AAM, and LBP. The experiment results are shown in Table 2 and Figure 3(1)-(3). In the first approach, we concatenate the AAM features with Gabor features, and use LAR as the feature selection method, which is denoted as ag^iL. It turns out that AAM+Gabor-[0,1]+LAR gives the best MAE=6.11 with 32 selected variables. In the second approach, we concatenate the AAM features with Gabor features, and use sequential feature selection method, which is denoted as ag^iS. It turns out that AAM+Gabor-[0,1]+Seq gives the best MAE=5.69 with first 257 variables. In the third approach, we concatenate the Gabor features with the AAM features, and use sequential model

Table 4. MAEs (year) per Decade of different algorithms on the PAL database

Summary to MAEs per Decade of normalized global/local feature and feature fusion on PAL database										
Range	#Im	$Gabor^1$	AAM^1 [3]	$LBP^{2,3}$ [3]	ag^2L	ag^2S	g^2aS	ab^2L[3]	ab^2S[3]	b^5aS[3]
18-19	31	7.67	6.79	9.14	6.50	6.29	10.12	6.37	**4.99**	7.28
20-29	170	6.36	6.60	7.34	6.08	5.56	6.15	5.57	**5.51**	6.19
30-39	38	5.27	5.66	8.66	5.63	5.42	**5.18**	6.24	6.33	5.37
40-49	29	6.21	5.58	8.31	4.53	**4.35**	6.31	5.24	4.46	5.18
50-59	26	**5.63**	7.78	8.03	6.36	6.63	6.40	6.65	6.95	7.36
60-69	94	5.31	6.33	7.23	5.75	5.45	5.77	5.81	**5.19**	5.53
70-79	98	6.59	5.72	7.19	6.17	**5.39**	7.45	6.36	5.60	5.75
80-89	49	8.56	8.26	8.54	7.40	6.93	9.24	8.63	**6.77**	7.90
90-93	5	6.51	**4.63**	9.76	8.97	6.13	10.62	7.05	7.80	6.71
Ave.	_	**6.38**	6.47	7.70	6.11	**5.69**	6.82	6.18	**5.65**	6.17
#Var	_	**41**	200	80	32	**257**	48	116	**265**	348
Total-Var	_	**539**	230	150	480	**769**	769	380	**380**	380

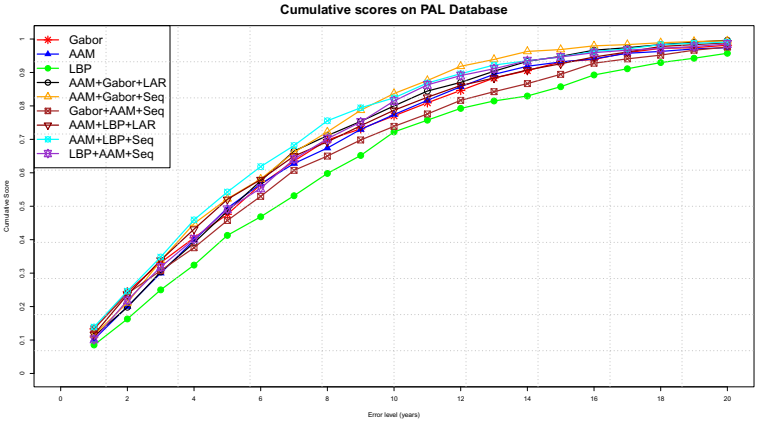


Fig. 5. Systematic performance comparison on age estimation for PAL Database

selection method, which is denoted as g^iaS. It turns out that Gabor-Norm+AAM+Seq gives the best MAE=6.82 with first 48 variables.

Finally, we compare the best MAEs among all these 39 combinations on age estimation in Table 3 and Figure 4. We can find out that the fusion of global/local features, either Gabor+AAM or LBP+AAM work better than the single feature representation of Gabor, AAM, or LBP consistently. Under the feature fusion framework, both ag^2S and ab^2S achieves the better MAE than the rest approaches.

We further study the MAE per Decade and Cumulative Scores for rigorous system comparison in Table 4 and Figure 5. It shows the fused Gabor and AAM performs similarly to fused LBP and AAM for MAE per Decade. As to the Cumulative Scores, fused LBP and AAM outperforms the fused Gabor and AAM with error level of 9 years or less. However, the performance on Cumulative Scores reverses for error level of 10 years or above. Due to it generally used error level of 10 years as a performance

comparison protocol, fused Gabor and AAM can be a potential face representation for age estimation.

4 Conclusion

In this work we evaluate the performances of a total 39 different combinations of four feature normalization methods, two simple feature fusion methods, two model selection methods and three face representation methods as Gabor, AAM and LBP. Our experiment results suggest that with single face representation, Gabor outperforms AAM and LBP. Moreover, fusion of global/local facial features achieve better results over single facial feature. It is interesting to find out that for both Gabor and AAM features, the original feature without any scaling works the best for age estimation task. For feature fusion and feature selection methods, combination of $AAM + Gabor^2 + Seq$ and $AAM + LBP^2 + Seq$ are the top two choices.

Further research on this work include: 1) use canonical correlation analysis or Bayesian analysis to attack the dependence problem; 2) From Figure 3, we can further improve the performance by selecting part of AAM features and part of Gabor features, rather than a simply fusion of concatenation.

Acknowledgment. This work is supported by the Intelligence Advanced Research Projects Activity, Federal Bureau of Investigation, and the Biometrics Task Force. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of our sponsors.

References

1. Albert, A.M., Ricanek, K., Patterson, E.: A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. Forensic Science International 172, 1–9 (2007)
2. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines (2001) Software, http://www.csie.ntu.edu.tw/cjlin/libsvm
3. Chen, C., Yang, W., Wang, Y., Ricanek, K., Luu, K.: Facial feature fusion and model selection for age estimation. In: 9th International Conference on Automatic Face and Gesture Recognition (2011)
4. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active Appearance Models. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, pp. 484–498. Springer, Heidelberg (1998)
5. Fu, Y., Cao, L., Guo, G., Huang, T.S.: Multiple feature fusion by subspace learning. In: CIVR, pp. 127–134 (2008)
6. Fu, Y., Guo, G., Huang, T.: Age synthesis and estimation via faces: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(11), 1955–1976 (2010)
7. Fu, Y., Huang, T.: Human age estimation with regression on discriminative aging manifold. IEEE Transactions on Multimedia 10(4), 578–584 (2008)
8. Gao, F., Ai, H.: Face Age Classification on Consumer Images with Gabor Feature and Fuzzy Lda Method. In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 132–141. Springer, Heidelberg (2009)
9. Geng, X., Zhou, Z., Miles, K.S.: Automatic age estimation based on facial aging patterns. IEEE Trans. on PAMI 29(12), 2234–2240 (2007)

10. Guo, G., Fu, Y., Dyer, C., Huang, T.: Image-based human age estimation by manifold learning and locally adjusted robust regression. IEEE Transactions on Image Processing 17(7), 1178–1188 (2008)
11. Guo, G., Mu, G., Fu, Y., Dyer, C., Huang, T.: A study on automatic age estimation using a large database. In: 2009 IEEE 12th International Conference on Computer Vision, September 29-October 2, pp. 1986–1991 (2009)
12. Kwon, Y.H., da Vitoria Lobo, N.: Age classification from facial images. In: Proceedings CVPR 1994, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 762–767 (June 1994)
13. Lanitis, A., Draganova, C., Christodoulou, C.: Comparing different classifiers for automatic age estimation. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 34(1), 621–628 (February 2004)
14. Lanitis, A., Taylor, C.J., Cootes, T.F.: Toward automatic simulation of aging effects on face images. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(4) (2002)
15. Liu, C., Wechsler, H.: Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Transactions on Image Processing 11(4), 467–476 (2002)
16. Mark, L.S., Pittenger, J.B., Hines, H., Carello, C., Shaw, R.E., Todd, J.T.: Wrinkling and head shape as coordinated sources of age level information. Journal Perception and Psychophysics 124, 117–124 (1980)
17. Minear, M., Park, D.C.: A lifespan database of adult facial stimuli. Behavior Research Methods, Instruments, & Computers 36, 630–633 (2004)
18. Patterson, E., Sethuram, A., Albert, M., Ricanek, K.: Comparison of synthetic face aging to age progression by forensic sketch artist. In: IASTED International Conference on Visualization, Imaging, and Image Processing, Palma de Mallorca,Spain (2007)
19. Suo, J., Zhu, S., Shan, S., Chen, X.: A Compositional and Dynamic Model for Face Aging. 32(3), 385–401 (2010)
20. Tan, X., Triggs, B.: Fusing Gabor and Lbp Feature Sets for Kernel-Based Face Recognition. In: Zhou, S.K., Zhao, W., Tang, X., Gong, S. (eds.) AMFG 2007. LNCS, vol. 4778, pp. 235–249. Springer, Heidelberg (2007)
21. Wang, J.-G., Yau, W.-Y., Wang, H.L.: Age categorization via ecoc with fused gabor and lbp features. In: Workshop on Applications of Computer Vision (WACV 2009), pp. 1–6 (December 2009)
22. Yan, S., Zhou, X., Liu, M., Hasegawa-Johnson, M., Huang, T.: Regression from patch-kernel. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–8 (June 2008)
23. Yang, Z., Ai, H.: Demographic Classification with Local Binary Patterns. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 464–473. Springer, Heidelberg (2007)
24. Zhao, W.-Y., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. ACM Comput. Surv. 35(4), 399–458 (2003)

Gender Classification via Global-Local Features Fusion

Wankou Yang[1], Cuixian Chen[1], Karl Ricanek[1], and Changyin Sun[2]

[1] Face Aging Group, Dept. Of Computer Science, UNCW, USA
[2] School of Automation, Southeast University, Nanjing 210096, China
wankou_yang@yahoo.com.cn, {chenc,ricanekk}@uncw.edu,
cysun@seu.edu.cn

Abstract. Computer vision based gender classification is an interesting and challenging research topic in visual surveillance and human-computer interaction systems. In this paper, based on the results of psychophysics and neurophysiology studies that both local and global information is crucial for the image perception, we present an effective global-local features fusion (GLFF) method for gender classification. First, the global features are extracted based on active appearance models (AAM) and the local features are extracted by LBP operator. Second, the global features and local features are fused by sequent selection for gender classification. Third, gender is predicted based on the selected features via support vector machines (SVM). The experimental results show that the proposed local-global information combination scheme could significantly improve the gender classification accuracy obtained by either local or global features, leading to promising performance.

Keywords: Gender Classification, AAM, LBP, Feature Fusion.

1 Introduction

Human faces contain important information, such as gender, race, mood, and age. In the area of human-computer interaction, there are both commercial and security interests to develop a reliable gender classification system from a good or low quality images [1-3]. Gender perception and classification have been studied extensively from psychological prospective [4,5], which show that gender has close relationships with both 2D information and 3D shape [6,7].

Wild et al. showed that gender classification achieved much lower recognition rate for children's face than the ones of adults [5]. Moghaddam and Yang [8] developed a robust gender classification system based on RBF-kernel SVM on a set of FERET raw images, and concluded that the nonlinear SVM outperformed the traditional pattern classifiers for gender classification problems. There are many publications on gender classification using FERET database [24], such as [2,9] etc. Yang et al. investigated three gender classification algorithms (SVM, FLD and Real Adaboost) with three different preprocessing methods on a large Chinese database with a good accuracy [10]. Baluja and Rowley [3] studied a method based on an Adaboost for classification from low resolution grayscale face images. Gao and Ai [11] proposed using Active Shape Model (ASM) for face representation and using probabilistic boosting trees

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 214–220, 2011.

approach for gender classification on a set of multiethnic faces. Guo et al. [12] studied the aging effect on gender classification and showed that the gender classification accuracy on young and senior faces can be much lower than the one on adults' faces, and hence concluded that the age of a person affected the gender recognition significantly. This conclusion validated the earlier work of Wild [5]. The active appearance model was first proposed by Cootes et al. [13]. AAM decouples and models shape and pixel intensities of an object. AAM model has been successfully applied in Age estimation and gender classification [14-16]. Wang et al. [14] gave a robust gender classification system via model selection, which performed well across all ages, infants to seniors. Makinen and Raisamo [2] gave a systematic study on gender classification with face detection and face alignment.

On the other hand, feature descriptor based methods have successfully applied in face recognition. This kind of methods becomes much popular due to their robustness in environmental changes and also it independent from the location of facial components. Following this, Ojala and Pietikäinen [17] proposed the local binary patterns (LBP) which is widely used in gender classification using face image [18,19]. Jabid et al. [20] used Local Directional Pattern (LDP) to represent face images and used the extracted LDP features for classification. Gayathri Mahalingam et al. [21] used Enhanced Local Gabor Binary Patterns (Enhanced LGBP) for gender classification and age estimation.

Few papers have yet discussed to classify gender by fusing global and local features. In the literature of psychophysics and neurophysiology, many studies have shown that both local and global information is crucial for the image perception and recognition of human beings [23] and they play different but complementary roles. A global feature reflects the holistic characteristics of the image and is suitable for coarse representation, while a local feature encodes more detailed information within a specific local region and is appropriate for finer representation. Hence, better recognition accuracy can be expected if local and global information can be appropriately combined. Such an idea has already been explored in iris recognition, face recognition and finger-knuckle-print recognition [22,23,24].

In this paper, we propose a novel global-local feature fusion scheme for gender classification. Specifically, we take the AAM information as the global feature and take the local texture information as the local feature. Finally, the two kinds of features are fused by sequent selection algorithms and used to classify gender via SVM. Fig. 1 shows the framework of the proposed method.

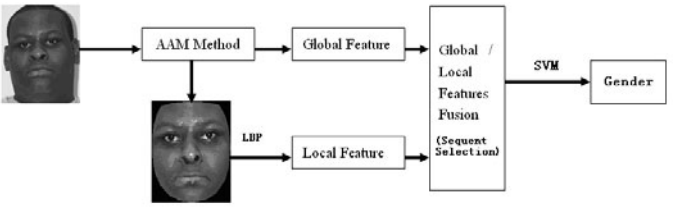


Fig. 1. The scheme of the proposed method for gender classification

2 Active Appearance Model

The active appearance model (AAM) was first proposed by Cootes et al. [8]. AAM decouples and models shape and pixel intensities of an object. The latter is usually referred to as texture. A very important step in building an AAM model is identifying a set of landmarks and obtaining a training set of images with the corresponding annotation points either by hand, or by partial- to completely automated methods. As described in [8], the AAM model can be generated in three main steps: (1) A statistical shape model is constructed to model the shape variations of an object using a set of annotated training images. (2) A texture model is then built to model the texture variations, which is represented by intensities of the pixels. (3) A final appearance model is then built by combining the shape and the texture models.

3 Local Binary Patterns

Ojala and Pietikäinen [17] propose the local binary patterns (LBP) which is widely used in texture classification. It encodes the difference between center pixel and its surrounding ones in a circular sequence manner. It characterizes the local spatial structure of image in equations (8).

$$f_{R,N} = \sum_{i=0}^{N-1} s(p_i - p_c)2^i , \quad s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \tag{8}$$

where p_i is one of the N neighbor pixels around the center pixel p_c, on a circle or square of radius R. An illustration of the basic LBP is shown in Fig. 2. LBP favors its usage as a feature descriptor is its tolerance against illumination changes and computational simplicity. LBP code can be regarded as a micro-texton. Local primitives codified by these bins include different types of curved edges, spots, flat areas etc.

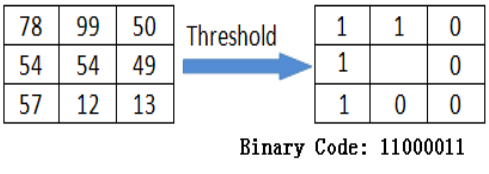


Fig. 2. The basic LBP operator

We use LBP histogram (Uniform Patterns with 59 Bins) to describe the images. First, we get the shape free image by using AAM; Second, we divide the image into m*n sub-region; Third, we calculate the LBP histogram of each sub-region and concatenate the LBP histograms to get a global description of the image; Four, since the dimensionality of the concatenated LBP histogram is m*n*59 and very large, we use PCA to reduce the dimensionality of the concatenated LBP histogram to 150 by preserving energy about 95%. Fig.5 shows the framework of LBP feature extraction.

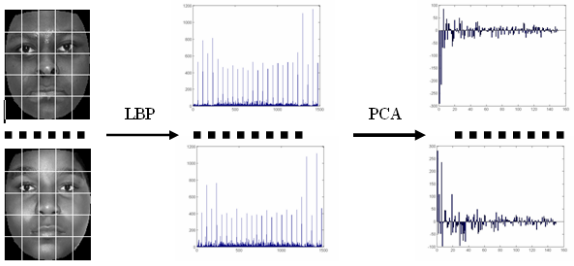


Fig. 3. LBP feature extraction

4 Support Vector Machines

Support Vector Machines (SVM) [25] is a supervised learning technique from the field of machine learning and is applicable to both classification and regression. The basic training principal behind SVM is finding the optimal separating hyperplane that separates the positive and negative samples with maximal margin. Based on this principal, a linear SVM use a systematic approach to find a linear function with the lowest VC dimension. For linearly non-separable data, SVM can map the input to a high dimensional feature space where a linear hyperplane can be found. SVM has been successfully used in gender classification and age estimation [8].

5 Experiments

The proposed method is evaluated on the FG-Net database. The FG-Net longitudinal face database is a public available image database. It contains 1002 color or grey images from 82 subjects, with age ranging from 0 to 69 years. (See Fig. 6 for sample images). Histogram of the age distribution of FG-Net database is shown in Fig. 7. The histogram is right-skewed with its majority less than 30 years old, and only eight images aged more than 60. The FG-Net face database is widely used in the areas of age estimation and age progression. Now, it has been used for gender classification, although the face this database contains 70.86% of faces with ages 20 or below, which makes it a challenging data set for gender classification. Table I shows that there are 68.56% of subjects whose ages are 18 or below, and there are 41.02% of subjects that are 10 or below.

Fig. 4. Sample images of FG-Net database

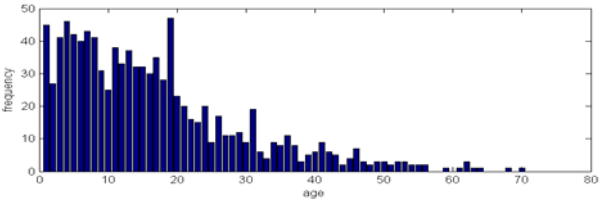


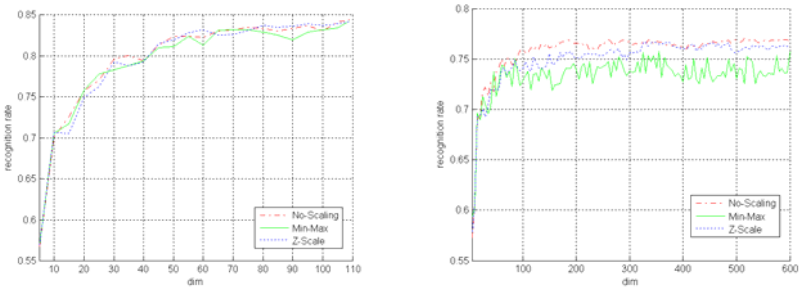
Fig. 5. Age Histogram of FGNet database

In FG-Net database, each image is annotated with 68 landmarks as shown in [13]. The annotated faces with shape and texture information are presented to the AAM system to obtain the encoded appearance features, a set of transformed features with dimension size 109. Here the AAM-Library tool [14] is utilized to implement the AAM system. Meanwhile the shape-free patch is also extracted from the annotated faces via the Active Shape Model provided by the AAM-Library tool. Next, LBP transformation is used to extract local features, respectively. Here, the shape-free image is divided into 5*5 subimage. Third, PCA transformation is performed on the LBP feature histogram to get final 600-dimensional LBP features.

In the first experiments, we compare the gender classification performance of AAM features and LBP features. Here we compare two normalization methods: Min-Max [0 1] and Z-score. The experimental results are shown in Table 1. Fig. 8 shows the classification curves with different feature dimensionality. The superscript 1 means no scaling, superscript 2 denotes Min-Max-[0 1], superscript 3 denotes Z-score to normalize the features.

Table 1. Average recognition rates of different features

	AAM^1	AAM^2	AAM^3	LBP^1	LBP^2	LBP^3
Mean(%)	**84.4337**	84.3297	84.2327	**77.0535**	75.7515	76.7416
Std(%)	**2.3849**	2.2927	1.7024	**2.4196**	2.2096	1.6425
Dim	**109**	109	109	**515**	355	370



(a) AAM (b) LBP

Fig. 6. Classification results vs. feature dimensionality

In the second experiments, we show the performance of the proposed GLFF method. We perform a standard 5-fold cross validation to evaluate the performance of the proposed method. We use the contributed package "Libsvm" [15] in Matlab for the computation of SVR (svm_type: C_SVC, kernel_type: RBF Kernel, gamma and C are set by grid search). The experimental results are shown in Table 2. From Table II, we can find that feature fusion can improve the classification results and it is valuable to research more efficient fusion strategy.

Table 2. Average recognition rates of global-local features

	A^1B^1	A^1B^2	A^1B^3	B^1A^2	B^1A^3
Mean(%)	**85.0277**	83.3356	76.9495	83.5357	**84.6317**
Std(%)	**1.9321**	0.9408	1.3972	1.6848	**2.5264**

6 Conclusions

In this paper, a novel global-local features fusion (GLFF) based gender classification method is proposed. It is based on the fact that both local and global features are crucial for the image recognition and perception and they play different and complementary roles in such a process. In GLFC, the global features extracted via AAM and the local features extracted on the shape free images via LBP. The global and local features are fused by sequent selection. The gender is predicted based on fused global-local features via SVM. The experimental results on FGNet database demonstrate that GLFF significantly improves the gender classification results. In the future, we will research feature fusion via random forest.

Acknowledgments. This work is supported by the Intelligence Advanced Research Projects Activity, Federal Bureau of Investigation, and the Biometrics Task Force. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of our sponsors.

This work is also supported by NSF of China (90820009, 61005008, 60803049, 60875010), Program for New Century Excellent Talents in University Of China (NCET-08-0106), China Postdoctoral Science Foundation (20100471000) and the Fundamental Research Funds for the Central Universities (2010B10014), and China Postdoctoral Special Science Foundation (201104505).

References

1. Albert, A.M., Ricanek, K., Patterson, E.: A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. Forensic Science International 172, 1–9 (2007)
2. Makinen, E., Raisamo, R.: Evaluation of gender classification methods with automatically detected and aligned face. IEEE Trans. Pattern Anal. Mach. Intell. 30(3), 541–547 (2008)
3. Bekios-Calfa, J., Buenaposada, J.M., Baumela, L.: Revisiting linear discriminant techniques in gender recognition. IEEE Trans. Pattern Anal. Mach. Intell. 33(4), 858–864 (2011)

4. Baluja, S., Rowley, H.: Boosting sex identification performance. IJCV 71(1), 111–119 (2007)
5. Wild, H.A., Barrett, S.E., Spence, M.J., et al.: Recognition and sex categorization of adults' and children's face: examining performance in the absence of sexstereotyped cues. J. off Exp. Child Psychology 77, 269–291 (2000)
6. Brucelf, V., et al.: Sex discrimination: how do we tell the difference between male and female face? Perception 22, 131–152 (1993)
7. Otoole, A., et al.: Sex classification is better with three dimensional head structure than with image intensity information. Perception 26, 75–84 (1997)
8. Yang, M.H., Moghaddam, B.: Gender classification using support vector machines. In: ICIP, vol. 2, pp. 471–474 (2000)
9. Gutta, S., Wechsler, H.: Gender and ethnic classifications of human faces using hybrid classifiers. In: Proceedings 1999 International Joint Conference on Neural Networks, pp. 4084–4089 (1999)
10. Yang, Z., Li, M., Ai, H.: An experimental study on automatic face gender classification. In: ICPR, pp. 1099–1102 (2006)
11. Gao, W., Ai, H.: Face Gender Classification on Consumer Images in a Multiethnic Environment. In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 169–178. Springer, Heidelberg (2009)
12. Guo, G.D., Dyer, C., Fu, Y., Huang, T.S.: Is gender recognition influenced by age? In: IEEE International Workshop on Human-Computer Interaction (HCI 2009), in conjunction with ICCV 2009 (2009)
13. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active Appearance Models. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, pp. 484–498. Springer, Heidelberg (1998)
14. Wang, Y., Ricanek, K., Chen, C., Chang, Y.: Gender classification from infants to seniors. In: Proceedings of the IEEE Conference on Biometrics, Theory, Application and Systems (2010)
15. Ricanek, K., Wang, Y., Chen, C., Simmons, S.J.: Generalized multi-ethnic face age-estimation. In: IEEE Conf. on Biometrics: Theory, Applications and Systems (2009)
16. Chen, C., Chang, Y., Ricanek, K., Wang, Y.: Face age estimation using model selection. In: CVPRW (2010)
17. 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)
18. Lian, H.-C., Lu, B.-L.: Multi-View Gender Classification Using Local Binary Patterns and Support Vector Machines. In: Wang, J., Yi, Z., Żurada, J.M., Lu, B.-L., Yin, H. (eds.) ISNN 2006. LNCS, vol. 3972, pp. 202–209. Springer, Heidelberg (2006)
19. Fang, Y., Wang, Z.: Improving lbp features for gender classification. In: ICWAPR (2008)
20. Jabid, T., Kabir, M.H., Chae, O.: Gender classification using local directional descriptor. In: ICPR (2010)
21. Mahalingam, G., Kambhamettu, C.: Can discriminative cues aid face recognition across age? In: FG (2011)
22. Sun, Z., Wang, Y., Tan, T., Cui, J.: Cascading statistical and structural classifiers for iris recognition. In: ICPR (2004)
23. Su, Y., Shan, S., Chen, X., Gao, W.: Hierarchical ensemble of global and local classifiers for face recognition. IEEE Transactions Image Processing 18(8), 1885–1896 (2009)
24. Zhang, L., Zhang, L., Zhang, D., Zhu, H.: Ensemble of local and global information for finger-knuckle-print recognition. Pattern Recognition (in Press)
25. V. N. Vapnik, The nature of statistical learning theory (Spring 2000)

Age Estimation Using Multi-Label Learning

Xiaoyu Luo, Xiumei Pang, Bingpeng Ma, and Fang Liu

School of Computer Science and Technology,
Huazhong University of Science and Technology, Wuhan, 430074, China
luo.xiaoyu@smail.hust.edu.cn,
{pangxm,bpma,fang.liu}@hust.edu.cn

Abstract. Generally, age estimation is formulated as a single-label based problem. However, since aging is a gradual process and people are always in transition period between ages, labeling a facial example with an exact age is a difficult problem. Meanwhile, sufficient training data is lack for many ages. In this paper, to improve the accuracy of age estimation, we propose a novel approach by applying Multi-Label Learning to the age features. In the proposed approach, each facial image is treated as an example associated with the origin label as well as its neighboring ages, which makes the data more reliable and sufficient. The motivation comes from the observation that, with age changes slowly and smoothly, people would look quite like themselves before and after several years. Experiments show that the proposed approach outperforms the traditional age estimation approaches.

Keywords: Age Estimation, Multi-Label Learning, facial image features.

1 Introduction

A human face contains huge amount of information, which is very useful in many applications in computer vision. Facial age estimation has attracted more and more attentions since it plays a crucial role in the real-world, including security control and surveillance [1], soft biometrics [2], etc. For example, with its help, it is easy to prevent underage people to access adult page on the Internet or get alcohol or cigarettes from vending machines.

Age estimation is also challenging because the aging variation is determined by the person's gene. And as a result, people with different gender and ethnic origins would be totally different in the aging process. Moreover, many external factors such as living environment, health condition, making up and living style would make the aging variation more specific for everyone, which makes the age estimation more difficult [3].

Many efforts have been devoted in the past few years. Since the age label is always approximated by an exact integer, the previous research can be roughly divided into two categories: a multi-class classification problem and a regression problem. In [4], Lanitis et al. regarded age estimation as a multi-class classification problem and using nearest neighbor classifier and ANN (Artificial Neural Networks) classifier to get the final predicted ages. The SVM (Support Vector Machine) classifier was applied to

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 221–228, 2011.

age estimation by Guo et al. [5]. And in many age classification works, age is partitioned into several categories, such as baby (0 to 1), child (2 to 16), adult (17 to 50), and old (after 50) [6], which would get better results for specific application. As for a regression problem, Lanitis et al. [7] investigated linear, quadratic and cubic formulations for the aging function. Guo et al. [5] applied the SVR (Support Vector Regression) method on the OLPP (orthogonal locality preserving projections) age manifold for age estimation.

Recently, more and more studies focus on the special characteristics of facial age and aging process. As age labels are not independent and have strong interrelationships to each other, regression approaches have attracted more interest and achieved a better performance since it utilizes the ordering information of age. And some novel improvements have been made on age estimation. As different people have different aging processes, Zhang et al. [8] proposed a method for personalized age estimation. In [9], multi-view facial examples got from video contexts are used to learning universal multi-view age estimation. Geng et al. [10] used IIS-LLD (Improved Iterative Scaling-learning from label distributions), and studied the age estimation problem by learning from label distributions, which could overcome the difficulty of lacking sufficient training data for many ages in age estimation. Chang et al. had mentioned in [11] that the exact age values are not stable.

As mentioned in [10], [12], one of the main challenges of facial age estimation is the lack of sufficient training data, for it requires great efforts to collect multiple images, which covers a wide age range, from the same person. It is hard to search for the image taken years ago, and future images are thoroughly unavailable. In fact, the special properties of age and its transition process make age estimation so different from other classification or regression problems. Fortunately, aging is a continuous and slow process of change, and practically facial photos of one person taken in the adjacent years look very similar, which means the facial photos taken in neighboring years can share their age labels. Besides, nowadays, a facial example is always labeled with an exact value for the convenience of storage and computing. However, the age value in fact is a fuzzy property, for each person is in the period of transition from on exact value to another in most of time. The labeling of one facial example with an exact value can reduce the reliability of data. Moreover, there may exist lying about one's age which further decreases the reliability of data. Inspired by these observations, we make use of the examples at the adjacent ages, and fuzz the age of the examples to overcome the unreliable problem of age labeling.

In this paper, we labeled each example with its origin exact age value as well as several ones neighbor it. By this way, suitable range of a new example's labels will make it contribute to the learning of more ages. Thus, the training data sets for ages are enriched. Meanwhile, as examples are no longer labeled with only one age value, the unreliable problem can be resolved to some extent. Accordingly, a simple yet effective method Multi-Label Learning (MLL) is introduced for the new examples. Based on each example associated with multi-label, Geng et al. [10] proposed an algorithm. And their experimental results demonstrate that the accuracy of age estimation can be improved with the help of multi-label data. In order to further improve the accuracy of age estimation, we exploit a different form to represent the multi-label data as well as another suitable learning method. Moreover, more advantages of multi-label for age estimation are discussed. MLL studies the

ambiguity in the label space and works well with the subjects that relevant to multiple labels simultaneously [13], [14], [15], [16], such as text categorization, where a can be regarded as belonging to different categories. As a result, MLL is introduced when the facial examples are changed into multi-label form. The results and comparisons shown demonstrate that the proposed method achieves better performances than the traditional Single-Label Learning as well as Geng's method on FG-NET databases.

The rest of the paper is organized as follows. Section 2 introduces Multi-Label Learning and its application on age estimation. In Section 3, systematic comparative experiment results on FG-NET databases are given. Finally, this paper concludes in Section 4 with a summary.

2 Multi-Label Learning on Age Estimation

In this section, we first introduced MLL and then shown its application in age estimation problem.

2.1 Multi-Label Learning

Traditional single-label learning method concerned for learning from a set of examples that are associated with a single label l from a set of disjoint labels L, $|L| > 1$. If $|L| = 2$, then the learning problem is called a binary classification problem, while if $|L| > 2$, then it is called a multi-class classification problem. On the other hand, in Multi-Label Learning [17], the examples are associated with a set of labels $Y \subseteq L$. It studies the problem where a real-world object described by one example is associated with a number of class labels, and the task is to predict the label sets of unseen examples through analyzing training examples with known label sets. Formally, the task is to learn a function $f_{MLL}: \mathcal{X} \to 2^{\mathcal{Y}}$ from a given data set $\{(x_1, Y_1), (x_2, Y_2), \cdots, (x_m, Y_m)\}$, where $x_i \in \mathcal{X}$ is an example and $Y_i \subseteq \mathcal{Y}$ is a set of labels $\left\{y_1^{(i)}, y_2^{(i)}, \cdots, y_{l_i}^{(i)}\right\}$, $y_k^{(i)} \in \mathcal{Y}(k = 1, 2, \cdots, l_i)$. And Multi-Label Learning techniques have also been successfully applied to scene classification [13].

The multi-learning frameworks result from the ambiguity in representing real-world objects. For example, in scene classification, a photograph can belong to more than one conceptual class, such as 'sunsets' and 'beaches' at the same time.

There are also some methods to solving this Multi-Label Learning problem, such as multi-label text categorization algorithms [16], multi-label decision trees [14], multi-label kernel methods [13] and multi-label neural networks [15]. In this paper, We simply transform this Multi-Label Learning task further into a traditional supervised learning task, i.e. to learn a function $f_{SLL}: \mathcal{X} \times \mathcal{Y} \to \{-1, +1\}$. For any $y \in \mathcal{Y}$, $f_{SLL}(x_i, y) = +1$ if $y \in Y_i$, and $f_{SLL}(x_i, y) = -1$ otherwise. Then the Multi-Label Learning function turns into $f_{MLL}(x_i) = \{y | \arg_{y \in \mathcal{Y}}[f_{SLL}(x_i, y) = +1]\}$.

2.2 Multi-Label Learning on Age Estimation

As mentioned above, age and its changing process have specific characteristics, which are so special that make the age estimation problem very different from other

classification problems or regression problems. Scholars have dig out those properties and made use of some of them to achieve better results. We notice that because of the slowly and smoothly aging process, images of same person taken in the consecutive adjacent years look very similar, and it is hard to know which image is taken earlier.

Besides, reliability of data is another important property. However, those special characteristics also reduce the reliability of example data if the age labels are presented by exact value, which in fact happened in most of the previous age estimation methods. In the real world, people are always with the age label of non-integer values. For example, a person is actually 12.5 years old half a year after his 12-year birthday. Meanwhile, there may also exist some data errors during the data collection such as negligence or lying about one's age, which would further weaken the reliability of data.

Motivated by those, instead of labeling one example with only one value, we assign every training example with the labels including not only its original value but also several one around it. The changes and differences between the new examples for Multi-Label example Learning and the original ones are shown Figure 1.

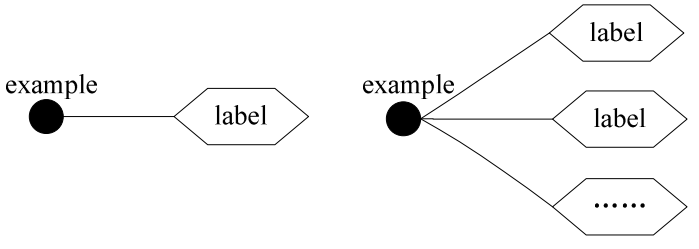


Fig. 1. Two different representations (left for traditional single-label example, right for multi-label example)

The change of traditional single-label examples to the novel multi-label examples in age estimation is a good use of the special age characters. At the same time, as the age labels are fuzzed, the problem of the unreliability is avoided to a certain extent.

Furthermore, MLL is introduced to solve our multi-label data age estimation problem. It performed well in scene classification and some other areas to solve data that with multi-label, just like our innovated form of examples for age.

In this paper, we apply ML-SVM algorithm [13], an algorithm of MLL to solve age estimation problem. In the framework of MLL, the predicted label result for each testing sample is a set of labels. However, the purpose of age estimation is to predict an exact age value for every testing sample. As a result, a decision should be made to obtain a single label result for every testing data after the prediction. We use the arithmetic mean value of all the labels' values in a testing sample's predicted set as the testing sample's final predicted value. And exploiting proper and more efficient decision methods is our future work.

3 Experiments

In this section, to show the effectiveness of MLL, we repeated experiments on the publicly available FG-NET age database [18] with different facial feature representation methods.

3.1 Experimental Setting

The FG-NET age database contains 1,002 color or gray facial images of 82 multiple-race individuals, whose age values are ranging from 0 to 69. The sample images in the database are shown in Figure 2. And all the sample images are cut into a 64x64 resolution according to the manually marked eye positions.

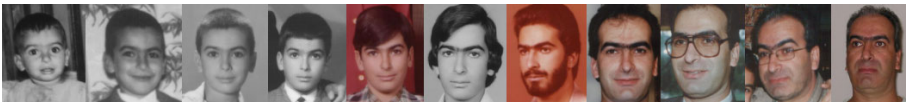


Fig. 2. Sample images of a person in the FG-NET database

We use PCA (Principal Components Analysis) [19], LGBP (Local Gabor Binary Patterns) [20] and BIF (Bio-inspired Features) [3] to extract facial image feature, and associate each example x_i with a label set L_i including its original label l_i, $l_i \pm 1$, $l_i \pm 2$,$\cdots$, $l_i \pm$ n. And we accomplished our experimental code referring to some work of the MIMLBOOST & MIMLSVM package [21]. In the experiments of MLL age estimation by using a threefold cross-validation test on the FG-NET, results are compared on the ranging of n from 0 to 10.

Meanwhile, we also compare the performances of MLL method MLSVM with standard and traditional age estimation method SVR and SVM to show the feasible and effective of MLL on age estimation. Besides, the comparison with IIS-LLD method is shown to demonstrate ours is better.

3.2 Experimental Results

The performance of the age estimation is evaluated by MAE (Mean Absolute Error), i.e. the average of the absolute errors between the estimated age and the real age. We compared the results from different values of n, which stand for the differences of the extension scopes of every example's labels. In Table 1, MAE results of different features' performances on age estimation through MLSVM are shown. And the MAE curves are also shown in Figure 3. We can see that for the PCA features, the best result is got when the value of n is 9, while two other kinds of features, LGBP and BIF get their best results when n equals 7 and 6 respectively.

Best results are got on different n according to various feature representations. As the value of n increased, we can see that the MAE reduces to a lowest value, and then begins to increase. Since the larger n means the wilder range of neighboring ages

labeled for each example, in general, before the MAE gets its lowest value, the multi-label sets decided by n are suitable to make face images contribute to the learning of age labels in the sets. However, as the n grows, the range becomes too wider to achieve better performance, for each example can contribute little to and even make bad affect on the learning of most age labels in its multi-label set.

Table 1. MAE (Year) of Age Estimation by Multi-Label Learning with Different n

Representation	n=0	n=1	n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9	n=10
PCA	7.55	6.69	6.67	6.40	6.16	5.98	5.92	5.80	5.81	**5.73**	5.88
LGBP	6.29	6.08	5.90	5.74	5.60	5.53	5.41	**5.37**	5.45	5.51	5.61
BIF	6.42	5.90	5.98	5.59	5.37	5.17	**5.04**	5.14	5.24	5.27	5.88

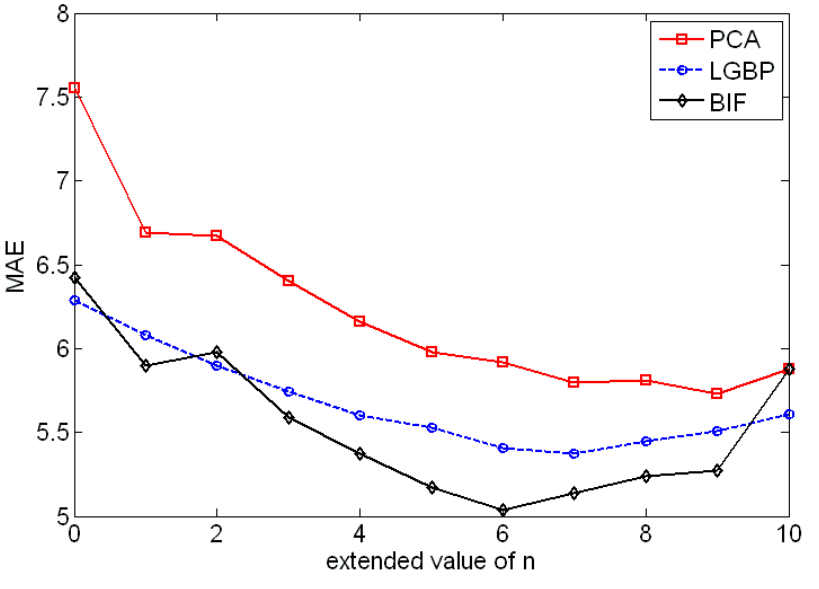


Fig. 3. MAE curves of Age Estimation by Multi-Label Learning with Different n

The contrasts of results of age estimation by MLSVM with traditional SVR and SVM are also shown in Table 2. The results on MLSVM method are the best ones among different n values. As can be seen, the performances of MLSVM method are significantly better than those traditional single-label based algorithms.

Table 2. Mean Absolute Error (Year) of Age Estimation on the FG -NET Ageing Database

Representation	ML-SVM	SVM	SVR
PCA	**5.73**	8.54	5.90
LGBP	**5.37**	7.02	5.80
BIF	**5.04**	7.29	5.30

Besides, the best MAE result got by IIS-LLD method used in [10] is 5.77. With the same feature and database, we get the result of 5.73. Moreover, IIS-LLD is tested through the LOPO (Leave-One-Person-Out) mode [22], while we use a threefold cross-validation mode. And LOPO mode would get a better result than a threefold cross-validation one theoretically, for it uses more examples for training. As a result, our method performs better than IIS-LLD.

4 Conclusion and Future Work

This paper proposes a novel method for facial age estimation based on Multi-Label Learning. By extending the single label of an example to several labels surrounding its original value, one example can contribute to more age classes' learning and overcome the unreliability of aging data. Experimental results on facial age estimation validate the advantage of new formed multi-label examples, and they also demonstrate that the proposed method outperforms other traditional single-label based algorithms, as well as the other multi-label based method. Since the predicted result of MLL for each testing sample is a vector, a way to get the optimal predicted label value from the vector is expected in the future work.

Acknowledgements. This paper is partially supported by Hubei Provincial Natural Science Foundation under contract No. 2010CDB02305 and Fundamental Research Funds for the Central Universities under contract No. 2011QN049.

References

1. Fu, Y., Guo, G., Huang, T.S.: Age synthesis and estimation via faces: A Survey. IEEE Trans. Pattern Anal. Mach. Intell. 32, 1955–1976 (2010)
2. Jain, A.K., Dass, S.C., Nandakumar, K.: Soft Biometric Traits for Personal Recognition Systems. In: International Conf. on Biometric Authentication, pp. 731–738 (2004)
3. Guo, G., Mu, G., Fu, Y., Huang, T.S.: Human Age Estimation Using Bio-inspired Features. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 112–119 (2009)
4. Lanitis, A., Draganova, C., Christodoulou, C.: Comparing different classifiers for automatic age estimation. TSMC–Part B 34, 621–628 (2004)
5. Guo, G., Fu, Y., Huang, T.S., Dyer, C.: Locally Adjusted Robust Regression for Human Age Estimation. In: IEEE Workshop Applications of Computer Vision, pp.1–6 (2008)
6. Gao, F., Ai, H.: Face Age Classification on Consumer Images with Gabor Feature and Fuzzy LDA Method. In: Int'l Conf. Advances in Biometrics, pp. 132–141 (2009)
7. Lanitis, A., Taylor, C., Cootes, T.: Toward Automatic Simulation of Aging Effects on Face Images. IEEE Trans. Pattern Anal. Mach. Intell. 24(4), 442–455 (2002)
8. Zhang, Y., Yeung, D.: Multi-Task Warped Gaussian Process for Personalized Age Estimation. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2622–2629 (2010)
9. Song, Z., Ni, B., Guo, D., Sim, T., Yan, S.: Learning Universal Multi-view Age Estimator by Video Contexts. In: 13th International Conf. on Computer Vision (2011)

10. Geng, X., Smith-Miles, K.A., Zhou, Z.-H.: Facial Age Estimation by Learning from Label Distributions. In: 24th AAAI Conf. on Artificial Intelligence (2010)
11. Chang, K.-Y., Chen, C.-S., Hung, Y.-P.: Ordinal Hyperplanes Ranker with Cost Sensitivities for Age Estimation. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 585–592 (2011)
12. Geng, X., Zhou, Z.-H., Smith-Miles, K.: Automatic age estimation based on facial aging patterns. IEEE Trans. Pattern Anal. Mach. Intell. 29(12), 2234–2240 (2007)
13. Boutell, M.R., Luo, J., Shen, X., Brown, C.M.: Learning multi-label scene classification. Pattern Recognition 37(9), 1757–1771 (2004)
14. Comité, F.D., Gilleron, R., Tommasi, M.: Learning Multi-label Alternating Decision Trees from Texts and Data. In: Perner, P., Rosenfeld, A. (eds.) MLDM 2003. LNCS, vol. 2734, pp. 35–49. Springer, Heidelberg (2003)
15. Zhang, M.-L., Zhou, Z.-H.: Multilabel Neural Networks with Applications to Functional Genomics and Text Categorization. IEEE Transactions on Knowledge and Data Engineering 18(10), 1338–1351 (2006)
16. Kazawa, H., Izumitani, T., Taira, H., Maeda, E.: Maximal margin labeling for multi-topic text categorization. In: Saul, L.K., Weiss, Y., Bottou, L. (eds.) Advances in Neural Information Processing Systems, vol. 17, pp. 649–656. MIT Press, Cambridge (2005)
17. Schapire, R.E., Singer, Y.: BoosTexter: A boosting-based system for text categorization. Machine Learning 39(2-3), 135–168 (2000)
18. The FG-NET Aging Database, `http://www.fgnet.rsunit.com/`, `http://www-prima.inrialpes.fr/FGnet/`
19. Turk, M.A., Pentland, A.P.: Face Recognition Using Eigenfaces. In: IEEE Conf. Computer Vision and Pattern Recognition, pp. 586–591 (1991)
20. Zhang, W., Shan, S., Gao, W., Chen, X., Zhang, H.: Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A Novel Non-Statistical Model for Face Representation and Recognition. In: 10th IEEE Int'l Conf. on Computer Vision, pp. 786–791 (2005)
21. The code package of algorithms MIMLBOOST & MIMLSVM, `http://lamda.nju.edu.cn/code_MIMLBoost%20and%20MIMLSVM.ashx`
22. Geng, X., Zhou, Z.-H., Zhang, Y., Li, G., Dai, H.: Learning from Facial Aging Patterns for Automatic Age Estimation. In: ACM Conf. Multimedia, pp. 307-316 (2006)

Selecting Distinctive Features to Improve Performances of Multidimensional Fuzzy Vault Scheme

Hailun Liu[1,3], Dongmei Sun[1,3], Ke Xiong[1,2,3], and Zhengding Qiu[1,3]

[1] Institute of Information Science, Beijing Jiaotong University, Beijing, China
[2] Department of Electronic Engineering, Tsinghua University, Beijing, China
[3] Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing, China
{06120393,dmsun,zdqiu}@bjtu.edu.cn,
kxiong@mail.tsinghua.edu.cn

Abstract. Fuzzy vault scheme is one of the most popular biometric cryptosystems. However, the scheme is designed for set differences while Euclidean distance is often used in biometric techniques. Multidimensional fuzzy vault scheme (MDFVS) is a modified version that can be easily implemented based on biometric feature data. In MDFVS, every point is a vector, and Euclidean distance measure is used for genuine points filtering. To get better performances, the step of feature selection in the MDFVS algorithms is very important and should be well designed. In this paper we propose applying recognition rate to measure discrimination of features and selecting strong distinctive features into genuine points. Some principles of selecting strong distinctive features to compose genuine points are discussed. An implementation of MDFVS with feature selection is also presented. Experimental results based on palmprint show that the proposed feature selection approach improves the performances of MDFVS.

Keywords: Biometric cryptosystems, Multidimensional fuzzy vault, Feature selection, Recognition rate.

1 Introduction

Biometric authentication that taking use of physiological or behavioral characteristics of an individual, such as fingerprint, palmprint, face, iris etc. for personal identification is gaining widespread concern. Since physiological or behavioral traits are innate to a person and cannot be lost or forgotten, biometrics based authentication is more reliable and convenient than password or token based authentication. With the continuously improved performances of biometrics based authentication, many biometric systems are designed and deployed successfully in both government (such as US-VISIT, J-VIS) and commercial applications (such as company sign in system).

With the widely application of biometrics based systems, the concern of privacy and security of biometric data is growing. For a biometric authentication system, during the enrollment phase, the biometric template is required to be stored in central database or smart card; during the authentication phase, the stored biometric template is retrieved for template matching. As we know, biometric data is uniquely associated with a

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 229–236, 2011.

person, unlike password or token, once it is compromised, we will never have new one to replace it. Some biometric samples may contain information about the health status of the users. Users with special disease would reject the biometric system. No one can make sure that the stored biometric data is only used for intended purpose, will the stored biometric data is abused for other purpose? E.g., the iris image database for access control would be abused for health statistics or criminal hunting.

Therefore, for widespread acceptance of biometric systems, it is essential to protect the biometric template in a practical biometric system.

For a biometric template protection algorithm, security, discriminability, cancelability are suggested to be satisfied. Security refers to the attacker cannot extract valid information from stored biometric template; discriminability refers to the system performances (False Accept Rate (FAR) and False Rejection Rate (FRR)) should not degrade after template protection mechanism is embedded; cancelability refers to revocability and diversity, once the stored template is compromised, a new one should be issued easily, and, different system store different template for the same user, one system is compromised will not affect the others.

To protect the biometric template and at the same time maintain the security, discriminability, cancelability in template protection scheme, many algorithms have been proposed. Currently, those algorithms can be classified two categories: template transformation approach and biometric cryptosystem [1, 2].

In biometric cryptosystem, not only the biometric template could be protected, but also there are extra merits: a bit string could be bound with the biometric template or generated directly from biometric template; both biometric template and the bit string are protected. The protected bit string can then be applied to cryptosystems as key or seed of random number generators etc.

Fuzzy vault scheme [3] proposed by Juels and Sudan is one kind of most famous biometric cryptosystems. The main idea of fuzzy vault scheme is concealing genuine points by adding large number of randomly generated chaff points to the vault. Biometric template or bit string are contained in genuine points. The security of the scheme depends on the number of chaff points in the vault. Since the concept of fuzzy vault scheme is proposed, many concrete implementations based on fingerprint, iris, face etc. have been designed by some researchers.

However, as some researchers pointed out, fuzzy vault scheme is not very effective for biometric feature template. Much extra works such as alignment, feature transformations etc. in some implementations [4, 5, 6] also prove our points. In our opinion, the root of the problem is that the fuzzy vault scheme is designed for set differences, while Euclidian distance metric is often employed in biometric systems. Our earlier study has proposed a modified scheme named multidimensional fuzzy vault scheme (MDFVS) [7, 8] to overcome those problems. Every points in MDFVS is vector, data in genuine points are selected from biometric template.

Since genuine points in MDFVS are vectors and each vector contains many features, the problems are: In the biometric feature vector, does each feature have same contribution to the performance of biometric authentication system? Which features are strong distinctive and which ones are weak distinctive? Which features are selected into one genuine point, and which features are selected into another genuine point? How to forming all genuine points by combining different features to get better performances? To overcome those problems, in this paper we proposed a feature

selection approach. Firstly, the discriminating ability of each feature in feature vector is measured by the recognition rate; secondly, features are selected and combined to form genuine points according to their recognition rates. Some principles of feature selection and feature combination are discussed.

The rest of this paper is arranged as follows. Traditional fuzzy vault and MDFVS are introduced in Section 2. Section 3 presents some methods measuring the discrimination of features in biometric feature vector. Principles of feature selection and combination are discussed in Section 4. An implementation of MDFVS with feature selection and experimental results based on palmprint are given in section 5. At last, we summarize our work in Section 6.

2 Fuzzy Vault Scheme and Multidimensional Fuzzy Vault Scheme

Fuzzy vault scheme is a secret sharing mechanism, protecting real data by adding noisy data is the primary idea of the scheme. The secret is encoded as the coefficients of a polynomial, and then genuine points are produced by evaluating the polynomial on locking data set A. To prevent illegal users obtaining genuine points, chaff points which do not depend on the polynomial are added to the vault. If another data set B substantially overlap A, i.e. $|A-B|<\varepsilon$, most of genuine points could be identified out from the vault and the polynomial could be reconstructed base on those identified genuine points, and then the shared secret could be retrieved.

The fuzzy vault scheme is designed based on finite field $\mathcal{F}$ of cardinality q [3], the similarity metric for two set A and B is set differences [9], i.e. $|A\Delta B|$, where $A\Delta B=\{x\in A\cup B \mid x\notin A\cap B\}$. $|A\Delta B|$ is the number of elements in set $A\Delta B$, which also represents the distance between two sets A,B. However, feature data extracted from biometric samples using numerous feature extraction methods such as principle components analysis (PCA) [10], fisher linear discrimination (FLD) [10] etc. are floating numbers which belong to real number field, and similarity metric for two biometric features f_1 and f_2 is Euclidean Distance:

$$d=\sqrt{\sum_{i=1}^{n}(f_{1i}-f_{2i})^2} \tag{1}$$

To improve the finite field based fuzzy vault scheme to adapt to real number based biometric feature vector we proposed a modified scheme named multidimensional fuzzy vault scheme in our earlier research.

The lock and unlock algorithms are as follows [8]:

2.1 Lock Algorithm

Input: Parameter k,t and r such that $k\le t\le r$. A secret or bit string or cryptographic key $\kappa\in GF(2)$ .Biometric feature vector $F_A\in R$.

Output: A set V of vectors $\{(x_i,x_v,y_i)\}_{i=1}^{r}$, $x_i,y_i\in R$

Algorithm LOCK:

$X, V \leftarrow \phi$;

$s \leftarrow \kappa$;

$p \leftarrow s$;

$F_t, F_v \xleftarrow{\text{feature selection}} F_A$

$f_i \in F_t, f_v \in F_v$

for $i = 1$ to t do

$(x_i, y_i) \leftarrow (f_i, p(f_i))$;

$V \leftarrow V \bigcup (x_i, f_v, y_i)$;

for $i = t+1$ to r do

$x_i \in_U R$;

$x_v \in_U R$;

$y_i \in_U R$;

$V \leftarrow V \bigcup (x_i, x_v, y_i)$

reorder *vectors* in V

output V;

2.2 Unlock Algorithm

Input: query feature vector F_B and V_A, which contains reordered vector set $\{(x_i, f_v, y_i)\}_{i=1}^{r}$ and parameter triple (k, t, r).

Output: bit string $\kappa' \in GF(2) \bigcup \phi$

Algorithm UNLOCK:

$Q \leftarrow \phi$;

$F_t, F_v \xleftarrow{\text{feature division}} F_B$

$\{(f_{i,1} \cdots f_{i,n-1})\}_{i=1}^{t} \xleftarrow{\text{feature selection}} F_t, F_v$

for $i = 1$ to t do

$(x_{l,1}, \cdots x_{l,n-1}, y_{l,1}) =$

$\underset{l=1 \cdots r-i+1}{\arg\min}\ \text{distance}((f_{i,1} \cdots f_{i,n-1}), (x_{l,1}, \cdots x_{l,n-1}))$

$Q \leftarrow Q \bigcup (x_{l,1}, y_{l,1})$;

$s' \leftarrow \text{polynomial reconstruction}(k, Q)$

$s' \rightarrow \kappa'$

output κ';

In order to obtain better performances, the step of feature selecting in the MDFVS scheme is very important and should be well designed. In feature selection, strong distinctive features are preferred, weak distinctive features have low priority to be selected or even be ignored. So, before feature selection, the distinguishing ability of features should be measured. Following we will present how to measure the distinguishing ability of different features in feature vector and some principles of selecting strong distinctive feature to compose genuine points.

3 Measuring the Distinguishing Ability of Features in Different Dimensions of the Feature Vector

In a typical biometric authentication system, original biometric samples are often transformed to feature vectors by feature extraction methods. There are many feature extraction methods such as PCA, FLD etc. The features in different dimensions of feature vector contribute differently to the system performances (FAR and FRR), some features are strong distinctive, some feature are weak distinctive.

Duo to the variances of samples, variances also exist among feature vectors extracted from different samples of the same user. For a single feature data in specific dimension, the interclass distribution is closely related to FAR, and the intraclass distribution is closely related to FRR. In our view, interclass variances and intraclass variances could reflect the distinguishing ability of features.

Many methods could be used to measure the distinguishing ability of features in feature vectors. The recognition rate is also one of them. Each feature in the feature vector is used for recognition test, if higher recognition rate is achieved, and then the feature is considered to be more distinctive and is preferred for generating genuine points.

Following we will present the two discrimination measuring methods.

3.1 Interclass Variances and Intraclass Variances of a Single Feature

Suppose the number of classes is C, there are M training feature vectors in each class, and the length of feature vector is n, taking FLD feature extraction algorithm for reference, we calculate the intra-class variances S_B and inter-class variances S_W as follows:

$$S_B = \sum_{k=1}^{C} (m_k - m) \tag{2}$$

$$S_W = \sum_{k=1}^{C} \sum_{i=1}^{M} (f_i - m_k) \tag{3}$$

where $m_k = \frac{1}{M} \sum_{i=1}^{M} f_i$, $m = \frac{1}{C} \sum_{k=1}^{C} m_k$.

3.2 Recognition Rate

In biometric authentication system, a whole feature vector is used for Euclidean distance calculation. When taking recognition rate (of course, the mentioned recognition rate is obtained from biometric feature set for training) for measuring the distinguishing ability of a single feature in specific dimension, the Euclidean distance calculation is modified as follow [11]:

$$d_i = \sqrt{(f_{1i} - f_{2i})^2} \tag{4}$$

Recognition rate of single feature based on training biometric feature vector set will be used for measuring discrimination of features in this paper.

4 Principles of Genuine Points Generation by Feature Selection

From section 3, we know distinguishing ability of features in feature vector could be measured by recognition rate, and then the features in feature vector will be sorted and

grouped according to the recognition rate. Some principles of generating genuine points by feature selection and combination are given below for discussion.

1. Because strong distinctive features in genuine points will make genuine points more distinguishable from chaff points in vault, thus reduce FRR. So strong distinctive features in the feature vector are preferred for genuine points generation. Weak distinctive features have low priority to be selected.
2. According to polynomial reconstruction theory, if only one chaff point is wrongly recognized as genuine point, the coefficients of polynomial reconstructed by those recognized points contained wrong genuine point will completely not be identical to original coefficients, so make the sum of recognition rates of features in different genuine points as equal as possible.
3. Obviously, to get higher performances, more strong distinctive features are preferred to be selected into genuine points. However, since the length of biometric feature vector is finite, and the vault with high dimension will require more storage space, a compromise between the system storage space and system performances should be considered.
4. The index mapping from feature vector to genuine points should be recorded for the Unlock algorithm.

5 MDFVS Implementation with Proposed Feature Selection Approach

The effectiveness of proposed feature selection approach for MDFVS is evaluated based on handmetric authentication Beijing Jiao tong university database (HA-BJTU). There are 1973 hand images captured from 98 persons in the database. Region of interest (ROI) which is named as palmprint is extracted from the hand image and is sampled to 128*128.

PCA is one of the most popular feature extraction methods and is proved highly effective for palmprint authentication [10]. In the experiment, 5 images are used for training in PCA feature extraction, 8 images are used for recognition test.

5.1 Framework of Implementation of MDFVS with Feature Selection

The feature mapping table in the framework is determined before the beginning of an actual system, so the introduce of feature selection will not occupy extra time in running the system.

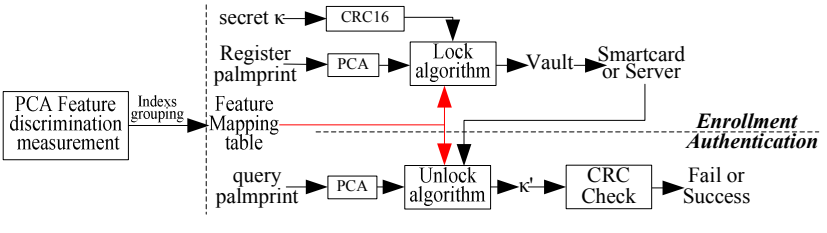


Fig. 1. Framework of implementation of MDFVS with feature selection

5.2 PCA Feature Discrimination Measurement by Recognition Rate

From the recognition histogram in Fig.1, we know that some features in the feature vector are strong distinctive, some of them are weak distinctive, which supplies the possibility of feature selection and combination for genuine points generation.

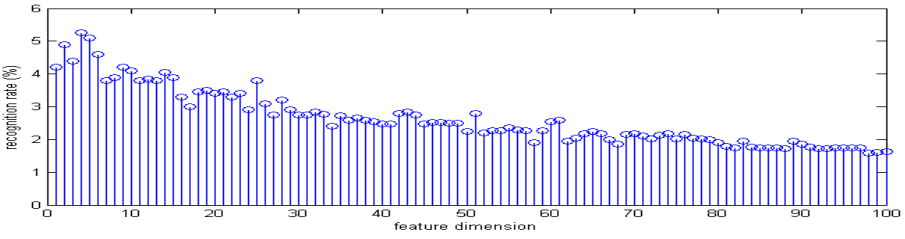


Fig. 2. Recognition rate of PCA features in different dimensions

5.3 Experimental Results

Parameters used in the experiment are shown in table 1. The performances of implementation of MDFVS are evaluated in terms of FAR and FRR.

Table 1. Parameters used in experiments

Parameters	Size
Degree	6
Number of genuine points	8
Number of chaff points	200

Table 2. FRR and FAR

Dimension of fuzzy vault	FAR without feature selection	FAR with feature selection	FRR without feature selection	FRR with feature selection
6	0.0052	0.0048	0.1054	0.0978
7	0.0048	0.0029	0.1123	0.1097
8	0.0029	0.0025	0.1023	0.0965
9	0.0029	0.0025	0.0945	0.0868

Form the table 2, we could find that both FAR and FRR are decreased in the implementation of MDFVS with proposed feature selection approach, which prove the effectiveness of proposed method for feature discrimination measurement, feature selection and combination. A special case, when the dimension of vault is 7, the FRR is a little higher than the others, because chaff points are generated randomly, and could affect the identification of genuine points.

6 Conclusions and Future Work

Given some features in the biometric feature vector are strong distinctive, some of them are weak distinctive, we have proposed selecting strong distinctive features to compose genuine points in MDFVS. There are many methods to measure the discrimination of features in different dimensions, such as interclass and intraclass variances of single feature, recognition rate of single feature, which we have already presented in section3.Some other methods, such as relative entropy measurement [12], could also be used for feature selection and improve the performances of MDFVS. Comparing the effectiveness of different feature selecting methods and applying the most effective one to improve performances of MDFVS are future works.

Acknowledgments. This work is supported by Beijing Municipal Natural Science Foundation (No.4102051), the Fundamental Research Funds for the Central Universities (2009JBZ006).

References

1. Uludag, U., Pankanti, S., Prabhakar, S., Jain, A.K.: Biometric Cryptosystems: Issues and Challenges. Proc. of the IEEE 92(6), 948–960 (2004)
2. Jain, A.K., Nandakumar, K., Nagar, A.: Biometric Template Security. EURASIP Journal on Advances in Signal Processing, 1–17 (2008)
3. Juels, A., Sudan, M.: A fuzzy vault scheme. In: Proc. of IEEE Int. Symp. on Info. Theory, p. 408. IEEE Press, New York (2002)
4. Nandakumar, K., Jain, A.K., Pankanti, A.: Fingerprint-based fuzzy vault: Implementation and performance. IEEE Trans. Inf. Forensics Secur. 2(4), 744–757 (2007)
5. Wang, Y.J., Plataniotis, K.N.: Fuzzy vault for face based cryptographic key generation. In: Proc. Biometrics Symposium, pp. 1–6. IEEE Press, New York (2007)
6. Lee, Y.J., Park, K.R., Lee, S.J., Bae, K., Kim, J.: A new method for generating an invariant iris private key based on the fuzzy vault system. IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics 38(5), 1302–1313 (2008)
7. Liu, H.L., Sun, D.M., Xiong, K., Qiu, Z.D.: 3D Fuzzy Vault Based on Palmprint. In: 2010 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC 2010), pp. 230–234. IEEE Press, New York (2010)
8. Liu, H.L., Sun, D.M., Xiong, K., Qiu, Z.D.: Is Fuzzy Vault Scheme very Effective for Key Binding in Biometric Cryptosystems? In: CyberC 2011: International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (2011)
9. Dodis, Y., Ostrovsky, R., Reyzin, L., Smith, A.: Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data. SIAM Journal of Computing 38(1), 97–139 (2008)
10. Li, Q.: Research on Handmetric Recognition and Feature Level Fusion Method, PhD thesis, BeiJing JiaoTong University, Beijing (2006)
11. Zhang, Y.Q., Qiu, Z.D., Sun, D.M.: Palmprint Identification using Weighted PCA Feature. In: IEEE 9th International Conference on Signal Processing Proceedings, pp. 2113–2116. IEEE Press, New York (2008)
12. Youmaran, R., Adler, A.: Measuring Biometric Sample Quality in terms of Biometric Information. In: Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference (2006)

A Fuzzy Vault Scheme for Feature Fusion*

Lifang Wu[1], Peng Xiao[1], Siyuan Jiang[1], and Xin Yang[2]

[1] School of Electronic Information and Control Engineering,
Beijing University of Technology, Beijing, China
lfwu@bjut.edu.cn, {xiaopeng9092,jsy}@emails.bjut.edu.cn
[2] Institute of Automation, Chinese Academy of Science (CSA), Beijing, China
xin.yang@ia.ac.cn

Abstract. Widespread application of biometric authentication brings about new problem of privacy. Biometric template protection is becoming a hot research. Efficient feature fusion is deemed to have good performance possibly. In this paper we proposed a fuzzy vault scheme for feature fusion. In our scheme, two facial features Multi-Block Local Binary Pattern (MB-LBP) and Principal Component Analysis (PCA) coefficients are extracted. A key is split into two overlapped subkeys. One is utilized to generate a set of helper data from MB-LBP. The other is utilized to generate another set of helper data from PCA coefficients. Two sets of helper data are submerged into the chaff points set and the final fuzzy vault is generated. In the fuzzy vault decoding, the MB-LBP and PCA coefficients of the query face image are utilized to recover two subkeys from the fuzzy vault. The final key is obtained from two subkeys. Because two subkeys are overlapped and complementary to each other, our scheme can obtain good authentication performance. It is confirmed by the experimental results.

Keywords: Biometric template protection, Feature fusion, Fuzzy vault, MB-LBP, PCA.

1 Introduction

Biometric authentication system has been increasingly deployed recently because they are more secure compared with traditional authentication mechanisms based on ID card or password. But the widespread application of biometric brought about new problem of privacy. If the biometric of a user is compromised, the user's privacy will be lost and it is hard to reissue a new one. Furthermore, the compromised biometric can be used for cross authentication in other database using the same biometric.

Biometric template protection combines biometric with cryptography, so that the biometric template can be utilized for authentication and its security can be protected. An ideal biometric template protection scheme should have both good performance and high security [1].

* This paper is partially supported by the Beijing municipal Nature Science Foundation under Grant No 4091004 and program of Beijing Municipality excellent under Grant No 2009D005015000010.

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 237–243, 2011.

In traditional biometric authentication research, every kind of biometric trait has strengths and weaknesses. Therefore, it is possible to obtain good performance fusing more than one biometric trait. Lin et al [2] proved in theory that multi-biometric based authentication system can obtain better performance than single biometric based system. It is also possible to obtain good performance by fusing more than one biometric features. In this paper we study the problem of biometric template protection for feature fusion.

The fuzzy vault is a popular scheme in biometric template protection. It was proposed by Juels and Sudan [3] and based on the secret sharing scheme. It combines a secret key with someone's biometric features. The secret key can be utilized to generate a n^{th} degree polynomial. The polynomial is evaluated by the values of biometric features, and it is represented as a series of points. The information of the secret key and the biometric feature are hidden into these points that are called helper data. Then the helper data is submerged into a large number of chaff points that are randomly generated. By now, the fuzzy vault is constituted. The polynomial can be reconstructed and the secret key can be recovered correctly only if we get n+1 true points. Hence, only the biometric of same person can be utilized to recover the secret key correctly. The security of authentication system can be protected.

The previous fuzzy vault schemes are usually based on single or fused biometric features. Nandakumar et al [4] fused more than one biometric into a new multi-biometric template. The template was protected by fuzzy vault. In Nandakumar's scheme the fused features were utilized in the traditional fuzzy vault scheme. There was not any improvement on the framework of the fuzzy vault. In this paper we propose a fuzzy vault scheme for feature fusion.

In our scheme, two kinds of facial features, Multi-Block Local Binary Pattern (MB-LBP) and Principal Component Analysis (PCA) coefficients are used. The key is split into two overlapped subkeys. One subkey is used to generate a set of helper data from MB-LBP. The other is used to generate another set of helper data from PCA coefficients. Two sets of helper data are submerged into chaff points set to constitute the final fuzzy vault. In the fuzzy vault decoding, the MB-LBP and PCA coefficients of the query face image are utilized to recover two subkeys from the fuzzy vault. The final key is obtained from two subkeys. Two subkeys are overlapped and complementary to each other. Therefore, our scheme can obtain good authentication performance.

The remaining parts of this paper are organized as follows: In Section 2, we describe the proposed scheme. We introduce the fuzzy vault encoding and decoding in detail. In our scheme, two face features are efficiently utilized together. In Section 3, we illustrate and analyze the experimental results. Finally, Section 4 concludes this paper with a summary.

2 The Proposed Scheme

Our scheme includes registration stage and authentication stage. The authentication includes three participants: the client, the database server and the verification server. In our scheme, two face features MB-LBP and PCA coefficients are extracted from the face image respectively. The corresponding feature extraction approaches in Ref [5-7] are referred and implemented.

In the registration stage, the user provide his or her username, password and face image to the client PC. The client PC extracts MB-LBP features $A=\{a_1,a_2,\cdots a_N\}$ and PCA coefficients $B=\{b_1,b_2,\cdots b_N\}$. In the meantime, a key of 64 bit is generated from password. The fuzzy vault encoding is designed to generate fuzzy vault $V=\{v_1,v_2,\cdots v_M\}$ from the key and two features. Then the username and the fuzzy vault are sent to the database server. The username and the key are sent to the verification server.

In the authentication stage, the users also provide his or her username, password and face image to the client PC. The client PC extracts MB-LBP features $A'=\{a'_1,a'_2,\cdots a'_{N_A}\}$ and PCA coefficients $B'=\{b'_1,b'_2,\cdots b'_{N_B}\}$ respectively. Then it sends username and two sets of features to the database server and sends the username to the verification server. The database server chooses the corresponding fuzzy vault in the database by username. Then a fuzzy vault decoding scheme is designed to recover the key from two features and the fuzzy vault. The recovered key is sent to the verification server. The verification server chooses the stored key by the username and compares two keys bit by bit. If two keys are identical, the user passes the authentication successfully; otherwise, the user is deemed invalid and rejected. The framework of this scheme is shown in Fig. 1.

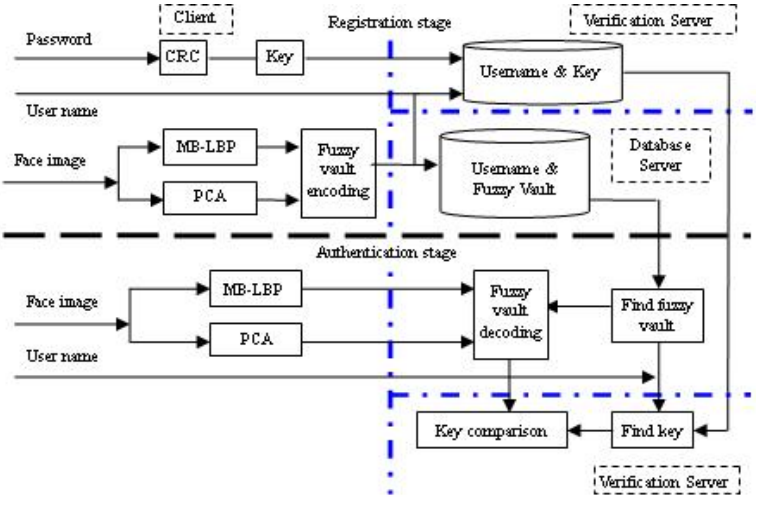


Fig. 1. The framework of proposed scheme

Let's assume that the MB-LBP features $A=\{a_1,a_2,\cdots a_{N_A}\}$ and PCA coefficients $B=\{b_1,b_2,\cdots b_{N_B}\}$ have been extracted. We will introduce the fuzzy vault encoding and decoding as follows.

2.1 Fuzzy Vault Encoding

The flowchart of the fuzzy vault encoding is shown in Fig.2. First the key is split into two subkeys SK1 and SK2. Then two polynomials $f_1(x)$ and $f_2(x)$ are generated from

the corresponding subkeys respectively. Next two polynomials $f_1(x)$ and $f_2(x)$ are evaluated using the corresponding features and two helper data sets are obtained. Finally the helper data are submerged into the chaff points to generate the final fuzzy vault V.

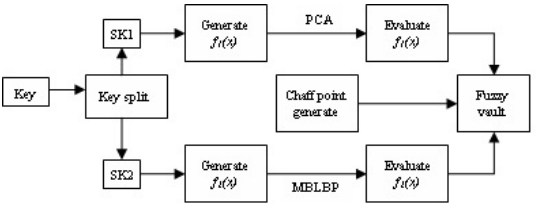


Fig. 2. The framework of fuzzy vault encoding

Step 1: Splitting the key and generating two polynomials
The key is separated into 8 coefficients (c0, c1, c2, …c7) of 8 bits. The SK1 involves the first five coefficients c0, c1, c2, c3 and c4. The SK2 involves the last five coefficients c3, c4, c5, c6 and c7. They overlap at the coefficients c3 and c4. Two polynomials are generated as follows.

$$f_1(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4 \tag{1}$$

$$f_2(x) = c_3 + c_4 x + c_5 x^2 + c_6 x^3 + c_7 x^4 \tag{2}$$

Step 2: Obtaining the helper data
We evaluate the polynomial $f_1(x)$ and $f_2(x)$ using components of MB-LBP and PCA coefficients respectively. We get helper data set G_1 and G_2.

$$\begin{cases} G_1 = \{(a_i, f_1(a_i)), i = 1, 2, \cdots N_A\} \\ G_2 = \{(b_i, f_2(b_i)), i = 1, 2, \cdots N_B\} \end{cases} \tag{3}$$

Step 3: Generating a set of chaff points in random.

$$C = \{(s_j, w_j), j = 1, 2, \cdots N_C\} \tag{4}$$

The components in the chaff point set should be different from the helper data G_1 and G_2. Therefore, we have the following criterion.

$$\begin{cases} s_j \neq a_i, s_j \neq b_i \\ w_j \neq f_1(s_j), w_j \neq f_2(s_j) \end{cases} \tag{5}$$

Step 4: Generating fuzzy vault
The components of helper data set G_1 and G_2 are submerged into the chaff points set. The fuzzy vault V is generated.

$$V = C \bigcup G_1 \bigcup G_2 = \{(u_i, v_i), i = 1, 2, \cdots N_C + N_A + N_B\} \tag{6}$$

2.2 Fuzzy Vault Decoding

Assume that the MB-LBP feature and PCA coefficients of the query face image are $A' = \{a'_1, a'_2, \cdots a'_M\}$ and $B' = \{b'_1, b'_2, \cdots b'_M\}$ respectively. A' and B' are used to find the possible points in the fuzzy vault V. then two sets of possible points are utilized to recover the key.

Step1: Finding the possible points.

Let's take A' for example. For each component a'_i, the Euclidean distance between a'_i and each component of V is computed. The minimum distance Dis_{Amin_i} and the corresponding component u_{Amin_i} of V are found. Then the minimum distances of all the components are ranked according to increasing order of corresponding minimum distance Dis_{Amin_i}. The first five components are chosen as the possible points. Without loss generality, we assume that the possible points for the set A' are TP_A, and the possible points for the set B' are TP_B.

$$TP_A = \{u_{Amin\_i}, i = 1, 2, ...5\} \tag{7}$$

$$TP_B = \{u_{Bmin\_i}, i = 1, 2, ...5\} \tag{8}$$

Step2: Recovering the key from possible points.

There are three cases in the procedure of recovering key.

Case1: We can recover the SK1 and SK2 from TP_A and TP_B correctly. Then SK1 and SK2 are cross verified. If the overlapped part of SK1 and SK2 are identical, it means that the two subkeys are true. We can get the key by union of SK1 and SK2.

Case2: SK1 (or SK2) can be correctly recovered from TP_A (or TP_B), while the other subkey can not correctly recovered. Let's assume that SK1 can be recovered from TP_A.

We first recover SK1 from TP_A. Then we get two coefficients (c_3 and c_4) of polynomial $f_2(x)$ from SK1. By now, only three coefficients of the polynomial $f_2(x)$ are unknown. We choose three points in TP_B to recover $f_2(x)$. if three of five points are true points, the polynomial $f_2(x)$ can be reconstructed correctly, and the SK2 can be obtained. The key can be obtained by union of SK1 and SK2.

Case3: Neither $f_1(x)$ nor $f_2(x)$ can be recovered correctly from its corresponding points set. The key can not be obtained correctly. We set the key as a series of '0'.

In our scheme, two subkeys SK1 and SK2 overlap 16 bits, so that they complement to each other. They can crossly verify each other. Therefore, our proposed scheme can obtain good performance.

3 Experimental Results

The proposed scheme is tested using the ORL face database [8], which contains 400 face images from 40 subjects with 10 face images of each subject. Images of some

subjects were taken at different time instances. Some vary with illumination, expression and pose variation.

The first 128 PCA coefficients and 320 MB-LBP features are selected for face features. We compare our scheme with the fuzzy vault scheme using only PCA, MB-LBP or the connected PCA and MB-LBP.

The first experiment is to test the authentication performance. The first 5 images of each subject are used as the templates for training, and the rest 5 face images are used for testing. The Receiver Operating Characteristic (ROC) curves are shown in Fig.3.

In our second experiment, we use cross verification to compare our algorithm with the others. ROC curves obtained in this experiment are shown in Fig. 4.

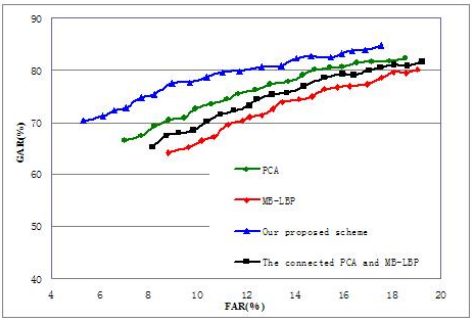


Fig. 3. Comparison of ROC curves

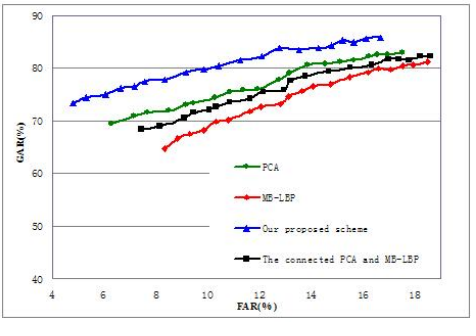


Fig. 4. ROC curves in cross verification

From Fig.3 and Fig.4, it is easy to see that the performance of PCA is much better than that of MB-LBP. The connected PCA and MB-LBP is better than MB-LBP but is worst than PCA. Our proposed scheme is much better than all of them.

4 Conclusion

In this paper we propose a fuzzy vault scheme for feature fusion. In our scheme, two kinds of facial features: Multi-Block Local Binary Pattern (MB-LBP) and Principal

Component Analysis (PCA) are utilized. The key is split into two overlapped subkeys and two polynomials are generated. Two helper data sets are generated from the corresponding polynomials and features respectively. The experimental results show that the proposed scheme has better performance than the fuzzy vault using only PCA, MB-LBP or connected PCA and MB-LBP.

References

1. Jain, A.K., Nandakumar, K., Nagar, A.: Biometric Template Security. EURASIP Journal on Advances in Signal Processing, Special Issue on Biometrics, 1–20 (January 2008)
2. Hong, L., Jain, A.: Integrating Faces and Fingerprints for Personal Identification. IEEE Trans. PAMI 20, 1295–1307 (1998)
3. Juels, A., Sudan, M.: A Fuzzy Vault Scheme. In: IEEE International Symposium on Information Theory, pp. 408–426 (2002)
4. Nandakumar, K., Jain, A.K.: Muitibiometric Template Security Using Fuzzy Vault. In: 2nd IEEE International Conference on Biometrics Theory, Applications and Systems, BTAS 2008, pp. 1–6 (2008)
5. Zhang, L., Chu, R., Xiang, S., Liao, S., Li, S.Z.: Face Detection Based on Multi-Block LBP Representation. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 11–18. Springer, Heidelberg (2007)
6. Zhang, S.C., Wu, L.F., Wang, Y.: Cascade MR-ASM for location facial feature points. In: Biometrics International Conference, pp. 683–691 (2007)
7. Turk, M.A., Pentland, A.P.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)
8. ORL face database, `http://www.cl.cam.ac.uk/research/dtg/attarchive/facedata-base.html`

Sparse Reconstruction Based Watermarking for Secure Biometric Authentication

Bin Ma[1], Chunlei Li[1,2], Zhaoxiang Zhang[1], and Yunhong Wang[1]

[1] Laboratory of Intelligent Recognition and Image Processing, Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing, China
[2] School of Electronic and Information Engineering, Zhongyuan University of Technology, Zhengzhou, China
{mabin,lichunlei1979}@cse.buaa.edu.cn, {zxzhang,yhwang}@buaa.edu.cn

Abstract. This paper presents a robust watermarking method to enhance the security of multimodal biometric authentication system. A compact representation of face is first generated by downsampling the raw image with a high ratio. Then, the face feature is employed as watermark and embedded into fingerprint image with a blind SS-QIM scheme. Under the framework of sparse reconstruction, the credibility of fingerprint data can be verified by checking the validity of extracted pattern in the authentication stage. Furthermore, if the extracted face watermark is valid, it can provide additional identity information for multimodal biometric authentication. Experimental results demonstrate that the face watermark can effectively verify the credibility of fingerprint images while not affecting their recognition performance. Plus, fusing valid face watermark with host fingerprint can also increase the reliability of authentication.

Keywords: biometric security, digital watermarking, QIM.

1 Introduction

Along with the widespread application of biometric authentication system, ensuring the security and integrity of biometric data has become a critical issue [1]. To address such problem, many template protection techniques have been proposed, and archived preliminary success [2]. However, the template protection mechanism works well only if the original raw data are from a legitimate source. Malicious attacks which replace or tamper biometric data before template generation module can hardly be resisted. Additionally, in some application scenarios, the biometric data have to be stored in original image form (e.g., face image on smart card, fingerprint image as judicial evidence) instead of encrypted templates.

Digital watermarking, which embeds secret message (watermark) in the host content, can verify data credibility through checking the presence and integrity of the hidden information, and thus provides an additional security level. Owing

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 244–251, 2011.

to such superiority, many researches have been performed to enhance biometric security with digital watermarking [1] [3]. Jain and Uludag [1] embed the Eigenface coefficients, by which can approximately reconstruct the original face image, in fingerprint image for content authentication and multimodal recognition. However, the assumption that the watermark embedder should maintain the Eigenface basis of training set greatly limits its practical applicability. Besides, the large capacity of Eigenface coefficients also confines watermark robustness. Kim and Lee [3] embed a 10×10 thumbnail of face image into fingerprint image to verify the integrity of biometric data. However, since the identity information provided by the thumbnail is limited, original face image is still required for face recognition. Previous researches have demonstrated that, due to the limitation of data payload, employing a compact and discriminative feature as watermark is a critical issue for biometric watermarking.

In this paper, we apply a sparse representation based classification method [4] to address the above mentioned problem. Different from Sheikh's work [5] that directly introduces compressive sensing to compress watermark information, we generate the watermark in a simple way and adopt sparse reconstruction method for subsequent verification issues. Under the proposed framework, the compact face feature embedded in the fingerprint image, can be used to validate the credibility of host image and serves as supplement identity information for posterior biometric authentication. Besides, we argue that the fidelity restriction which greatly confines the payload and robustness of watermark, is not crucial in biometric watermarking scenario. Namely, a perceptual distortion is acceptable provided that the intrinsic features utilized by recognition algorithm are preserved. And that motivates us to apply a watermarking strategy with large embedding strength to guarantee the robustness of hidden feature.

2 Framework

As Fig. 1 demonstrates, the proposed watermarking based multimodal authentication system can be generally divided into acquisition and authentication stage.

On the acquisition side, the original face image is downsampled to an 8×7 thumbnail with each pixel quantized to 2^5 gray scales. Subsequently, the 280-bit ($8 \times 7 \times 5$) face watermark is embedded into fingerprint image of the same individual by the proposed SS-QIM method.

During transportation (or storage), the watermarked fingerprint image may suffer incidental distortion (e.g., channel noise, lossy compression) or malicious attacks (e.g., replacing, tampering). Therefore, in the authentication stage, the following steps are performed to guarantee the reliability of the system:

(1) *Watermark extraction*: With respect to the application background, a pattern is extracted without performing watermark detection, since the bit stream extracted from a non-watermarked image is meaningless and could be easily verified in the following step.

(2) *Credibility verification*: The credibility of fingerprint image is verified by determining *whether the extracted pattern is valid face feature* with a sparse

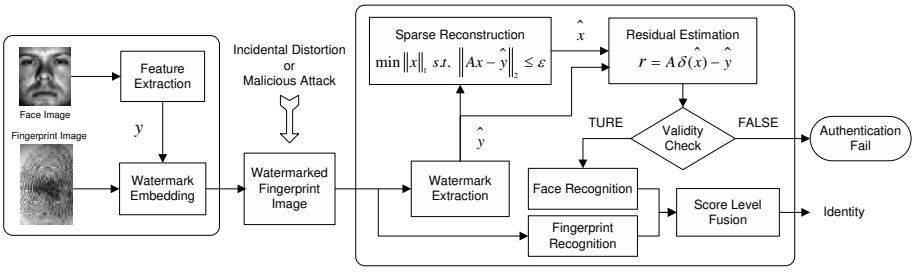


Fig. 1. Diagram of the watermarking based multimodal authentication system

reconstruction based method described in Section 4. Forge image that contains no watermark could be verified, and therefore, the security of authentication system is increased.

(3) *Multimodal biometric authentication*: If the data are credible, the extracted face feature is applied as additional identity information to increase recognition performance.

3 Watermarking Method

From the perspective of application, there are many differences between biometric watermarking and conventional watermarking: 1. Conventional watermark is a binary stream without bit priority (e.g., binary logo, hash sequence) while biometric watermark is numeric feature, whose bits have different importance; 2. The most important quality metric of host biometric data is discriminability rather than visual quality, thus the imperceptibility of biometric watermark is less important compared with robustness.

In the consideration of above mentioned matters, we adopt a large capacity robust watermarking method in DCT domain [6], and make improvements for embedding numeric biometric feature. The progress is described as follows:

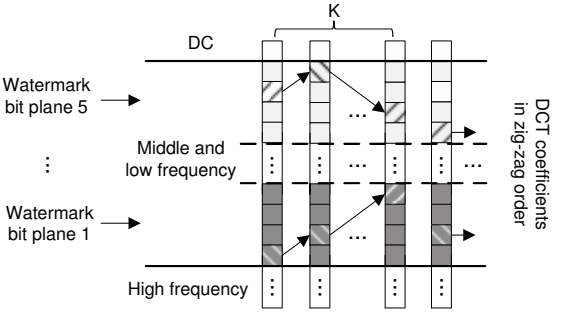


Fig. 2. Layered embedding in DCT domain

First, the DCT coefficients of 8×8 image blocks in zig-zag order are randomly arranged according to the secret key. Considering that high priority bits should be embedded in perceptually significant components to achieve good robustness, we divide the low and middle frequency bands into 5 layers, and embed most significant watermark bitplane into the lowest band layer as Fig. 2 illustrates.

To embed the ith bit b_i of one bitplane into the Lth layer, a spread spectrum sequence $s_i(k) \in \{0, 1\}$ $(1 \leq k \leq K)$ with the length of K, and a corresponding position sequence in layer L are generated according to the secret key. Then, each chosen coefficient is quantized to an even multiple of Q if $s_i(k) = 0$, or odd multiple if $s_i(k) = 1$ as Equation (1).

$$c_w(k) = \left\lfloor \frac{c(k) + s_i(k) \cdot Q}{2Q} + 0.5 \right\rfloor \times 2Q - s_i(k) \cdot Q \tag{1}$$

where Q denotes the overall quantization step in layer L.

The extraction stage is basically the inverse process of embedding: First, use the secret key to regenerate the spread spectrum sequence $s_i(k)$ and relocate the corresponding DCT coefficients. Then, since Q is assumed to be shared by watermark embedder and extractor, the embedded sequence $s'_i(k)$ could be extracted by Equation (2).

$$s'_i(k) = \mod \left\{ \left\lfloor \frac{c'_w(k)}{2Q} + 0.5 \right\rfloor, 2 \right\} \tag{2}$$

By mapping $s \in \{0, 1\} \rightarrow \{-1, 1\}$, the watermark bit can be finally decoded as: $\hat{b}_i = 1$, if $s_i \cdot s'_i > 0$; otherwise, $\hat{b}_i = 0$.

4 Sparse Reconstruction Based Authentication

Sparse reconstruction can be described as the following ill-posed inverse problem: given the m-dimension measurement $y = Ax$, and a matrix $A \in \mathbb{R}^{m \times n} (m \ll n)$, estimate the high dimensional vector $x \in \mathbb{R}^n$. Recent theory of compressive sensing has proven that, if the vector x is k-sparse (with k non-zero entries out of n dimensions), and the matrix A satisfies a restricted isometry property (RIP) [7], then x can be precisely reconstructed with a large probability by solving the ℓ_1 minimization problem of (3), provided that $m \geq Ck \log(n/k)$.

$$\min \|x\|_1 \quad s.t. \quad y = Ax \tag{3}$$

Since the real data are often noisy (e.g., the extracted watermark with slight distortion), the measurements can be rewritten as $y = Ax + z$, where z is a noise term with bounded norm $\|z\|_2 \leq \varepsilon$. The sparse solution x can be approximated by solving the modified problem:

$$\min \|x\|_1 \quad s.t. \quad \|Ax - y\|_2 \leq \varepsilon \tag{4}$$

Based on such theoretical foundation, Wright et al. propose a sparse representation based face recognition scheme in [4]. Regarding *test samples as linear*

combination of training samples, they concatenate all training samples v_{k,n_k} from k class subjects to construct a dictionary:

$$A = [A_1, A_2, ..., A_k] = [v_{1,1}, v_{1,2}, ..., v_{k,n_k}]$$

Then, the sparse representation of a test sample y which belongs to class i, can be written as the linear combination of all training samples as: $y = A\delta_i(x)$, where $\delta_i(x)$ denotes a vector whose nonzero entries are the entries in x that are associated with class i. Finally, given a test sample y, the classification method (see [4] for details) could be generally summarized as follows: First, calculate the sparse representation $\hat{x}$ by solving the problem of (1) ([8] is applied in our experiments). Second, find the class with least reconstruction residual by solving:

$$\hat{i} = \arg\min_i \ \|A\delta_i(\hat{x}) - y\|_2 \tag{5}$$

Algorithm 1. Face recognition via SRC [4]

Input: Matrix of training samples A $A = [A_1, A_2, ..., A_k] \in \mathbb{R}^{m \times n}$

1. Normalize the columns of A to have unit ℓ_2 norm.
2. Solve the ℓ_1 minimization problem:

$$\min \|x\|_1 \quad s.t. \quad \|Ax - y\|_2 \leq \varepsilon$$

3. Find the class with least residual by solving:

$$\hat{i} = \arg\min_i \ \|y - A\,\delta_i(\hat{x})\|_2$$

Output: Identity(y) = $\hat{i}$

Applying such framework to biometric watermarking scheme, we could verify the credibility of fingerprint image by checking the validation of extracted pattern. Namely, instead of solving the conventional watermark detection problem of *whether a watermark is presented*, we extract a bit stream from the fingerprint image anyway, and determine *whether the extracted pattern is valid face feature*. Since forge images from unreliable source contain no watermark, they only provide random patterns that can not be properly represented by the dictionary matrix A. Thus, we adopt outlier rejection strategy in face recognition [4] for watermark verification, by applying a threshold to the proportion of minimum residual to sub-minimum residual. Fig. 3 demonstrates the non-valid watermark pattern can be effectively distinguished with a threshold.

5 Experimental Results

Since face images are captured from legal users as their biometric watermark, it is reasonable to assume that the data collection environment could be well controlled. Therefore the face data utilized in our experiments is of less challenging: a subset of Yale database B which consists 38 subjects. For each subject, 40 images with frontal pose and slight illumination change are selected. We randomly

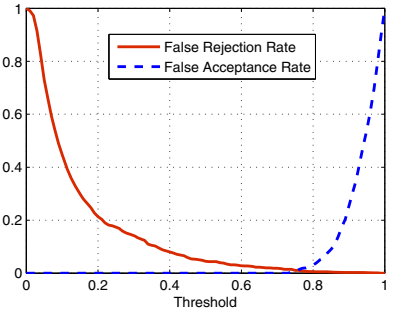


Fig. 3. Performance of watermark verification

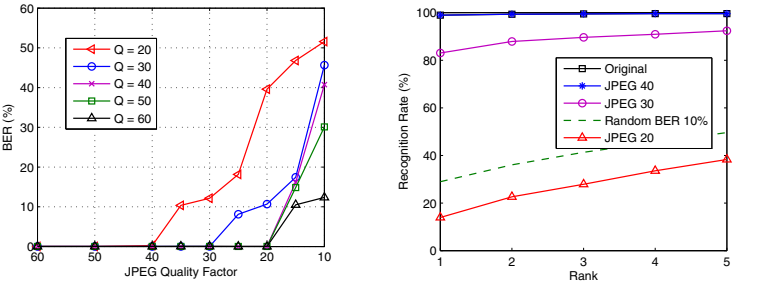
use half of them for training and the rest for testing. The fingerprint subjects are randomly picked from FVC2002 DB2, with each one consists 8 images size of 296×560. The fingerprint matching score is calculated using the minutiae based method in [9]. For watermarking scheme, the length of spread spectrum sequence K is set to 21, the Q value of 5 layers are duplicate values.

In order to evaluate the distortion introduced by watermark, we conduct embedding experiments with different quantization step Q. The quality of watermarked fingerprint sets are demonstrated in Table 1. It can be observed that, with the increase of embedding strength, conventional image quality metric PSNR (dB) and SSIM show an obvious decreasing trend. However, despite the slight fluctuation due to the naive NN classifier, the recognition performance of fingerprint set is basically unaffected with original recognition rate (RR) equals 86.97%. Therefore, we may conclude that in biometric watermarking application, the practical utility of host image is discriminability rather than visual quality. It is reasonable to increase the embedding strength to guarantee watermark robustness while preserving the discriminant biometric feature.

Table 1. Effects of watermarking to fingerprint images

	$Q = 20$	$Q = 30$	$Q = 40$	$Q = 50$	$Q = 60$
PSNR	39.18	35.82	33.48	31.72	30.32
SSIM	0.973	0.949	0.923	0.899	0.878
RR	**85.9%**	**88.0%**	**87.8%**	**86.6%**	**86.5%**

The next experiment is conducted to evaluate the robustness of face watermark under JPEG compression. Fig. 4(a) shows the bit error rate (BER) of embedding strength Q versus JPEG quality factor. From the result we can observe that, the proposed method has good robustness to JPEG compression. Even with a small embedding strength ($Q = 20$), all watermark bits can be perfectly decoded under a low JPEG quality of 40. Fig. 4(b) illustrates the recognition curves using face feature (embedd with $Q = 20$) extracted from fingerprint sets compressed by different JPEG quality factors. Note that, when JPEG quality



(a) Bit error rate (b) Face recognition performance

Fig. 4. Evaluation of face watermark robustness

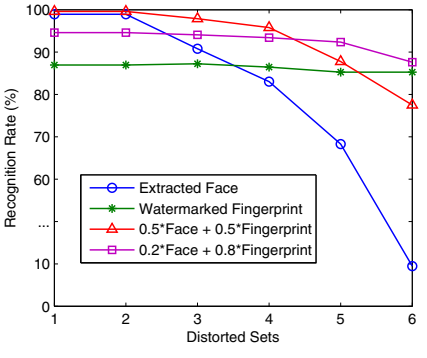


Fig. 5. Multimodal recognition performance after various distortions ($Q = 30$): 1. Original; 2. JPEG ($q = 20$); 3. Gaussian noise $(\mu, \sigma^2) = (0, 0.045)$; 4. Gaussian blur (window size, σ^2) = ([9 × 9], 0.5); 5. Salt & pepper noise ($d = 0.02$); 6. Salt & pepper noise ($d = 0.04$).

decreases to 30, although the watermark has a BER of around 10% (See Fig. 4(a)), it can still act as reliable identity information. However, when we randomly destroy 10% bits of the original face watermark, a great degeneration is shown in recognition performance (the dashed line in Fig. 4(b)). The significant gap between the two results just verify the effectiveness of differentiating bit priority. In fact, as for the former JPEG compression case, the error decoded 10% bits are mostly in less significant bitplane. The high priority bits embedded in low frequency with larger robustness, can still be correctly extracted, and thus greatly preserve the energy of feature vector. Further, slight noise in the extracted watermark can also be resisted during sparse reconstruction.

We apply a simple fusion strategy of min-max normalization followed by the weighted sum of scores to test the multimodal recognition performance of fusing face watermark (embedding strength $Q = 30$) with host fingerprint image. As Fig. 5 illustrates, the face watermark can provide an additional performance under moderate distortions and retains certain identity information even after severe damage. However, fusion recognition rate may decrease in some extreme

cases (distorted set 6 in Fig. 5). Thus, the predefined fusion strategy should give a larger weight to the feature with higher robustness. Furthermore, the distortion which is large enough to severely damage the watermark, also affects the creditability of the host biometric image. Under such condition, the data can be rejected as unreliable content in previous watermark verification stage.

6 Conclusions

In this paper, we have proposed a robust watermarking method to embed a compact face feature in fingerprint for multimodal biometric security. By employing the framework of sparse representation based classification, the watermark can serve as a reliable source to verify the credibility of biometric data and increase multimodal recognition reliability. Besides, the proposed bit priority based watermarking scheme can increase the robustness of numeric watermark.

Acknowledgements. This work is funded by the National Basic Research Program of China (No. 2010CB327902), the National Natural Science Foundation of China (No. 60873158, No. 61005016, No. 61061130560) and the Fundamental Research Funds for the Central Universities.

References

1. Jain, A.K., Uludag, U.: Hiding biometric data. IEEE Trans. Pattern Analysis and Machine Intelligence 25(11), 1494–1498 (2003)
2. Jain, A.K., Nandakumar, K., Nagar, A.: Biometric template security. EURASIP Journal on Advances in Signal Processing (2008)
3. Kim, W., Lee, H.: Multimodal biometric image watermarking using two-stage integrity verification. Signal Processing 89(12), 2385–2399 (2009)
4. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Trans. Pattern Analysis and Machine Intelligence 31(2), 210–227 (2009)
5. Sheikh, M., Baraniuk, R.: Blind error-free detection of transform-domain watermarks. In: Proc. ICIP, pp. 453–456 (2007)
6. Miller, M.L., Doerr, G.J., Cox, I.J.: Applying informed coding and embedding to design a robust high-capacity watermark. IEEE Trans. Image Processing 13(6), 792–807 (2004)
7. Candes, E.J., Tao, T.: Decoding by linear programing. IEEE Trans. Information Theory 51(12), 5406–5425 (2005)
8. Candes, E., Romberg, J.: ℓ_1-Magic: Recoverry of sparse signals (2005), http://www.acm.caltech.edu/l1magic/
9. Jea, T.Y., Govindaraju, V.: A minutia-based partial fingerprint recognition system. Pattern Recognition 38, 1672–1684 (2005)

A Review of Recent Advances in Ear Recognition

Li Yuan*, Zhi-Chun Mu[1], and Fan Yang[2]

[1] School of Automation and Electrical Engineering, University of Science and Technology Beijing, 100083, China
[2] LE2I Laboratory, University of Burgundy, 21078 Dijon Cedex, France
yuanli64@hotmail.com

Abstract. Ear recognition has become a rapidly growing research area in recent years. This paper gives an up-to-date review of research works in ear recognition based on 3D data and 2D ear images. For ear recognition in 3D, recent works on 3D feature extraction and model matching are presented and discussed. For 3D ear recognition based on 2D images, 3D ear model reconstruction is mainly discussed. For ear recognition in 2D, main research works on ear detection, ear feature extraction and classification are presented and discussed. Some possible future research interests are proposed in the end.

Keywords: ear recognition in 2D, ear recognition in 3D, ear detection, ear image dataset.

1 Introduction

Increasingly stringent security requirements call for more robust, convenient, user-friendly personal identification systems. Ear recognition has received much more attention in recent years. The human ears have rich and stable structure that is preserved from childhood into old age [1], and they do not suffer from changes in facial expression, aging, psychological factors or cosmetics. The long history of using ear shapes or ear prints in forensic science has shown their use for automatic human identification [2, 3]. Similar ways acquiring the biometric data at distance to the face biometric make the ear biometric an appealing candidate for video surveillance and non- contact biometric recognition.

Similar to face recognition, the research interests of ear recognition include automatic real-time ear tracking and detection, feature extraction and ear recognition /verification. This paper provides an up-to-date review of recent trends and major research works in ear recognition methods based on the two major imaging modalities: 3D range images and 2D images. The rest of the paper is organized as follows: Section 2 covers a general introduction to ear databases most commonly used and approaches to automated ear detection. Feature extraction and classification methods are overviewed for 3D ear recognition using range images, 3D ear recognition based on 2D images, 2D ear recognition and multimodal recognition

* This paper is supported by the National Natural Science Foundation of China under the Grant No. 60973064; Beijing Municipal Natural Science Foundation under the Grant No. 4102039.

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 252–259, 2011.

based on face and ear. Section 3 concludes the paper with some suggestions for future research directions in this area.

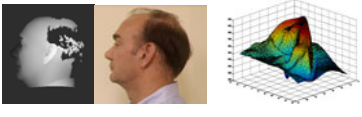
2 Principal Methods for Ear Recognition

In this section, we will start with presenting some commonly used ear image datasets. Then we will discuss automatic ear detection methods, and some principal methods for ear recognition based on 3D data and 2D images.

2.1 Ear Image Datasets

At present, the ear images used in most works mainly come from four datasets: UND dataset[4], UCR dataset[5], USTB dataset[6] and XM2VTS dataset[7]. More details about these datasets are shown as following:

(1) The UND ear dataset is taken by University of Notre Dame. The acquisition equipment is Minolta Vivid 910 camera. UND dataset includes three subsets as follows: Collection E (464 visible-light ear images from 114 subjects), Collection F (942 3D + corresponding 2D ear images from 302 subjects, now a subset of Collection J2), Collection J2 (1800 3D + corresponding 2D ear images from 415 subjects). In collection J2, 24 subjects with images taken at four different viewpoints. Figure 1 shows some example images of this dataset.



(a) (b) (c)

Fig. 1. Example images of UND ear database. (a) range image; (b) corresponding color image; (c) 3D image of the ear.

(2) The UCR dataset is taken by University of California at Riverside. The acquisition equipment is Minolta Vivid 300 camera. It includes 902 shots of 3D and corresponding 2D ear images from 155 subjects.

(3) The USTB ear dataset is generated by University of Science and Technology Beijing. The acquisition equipment is CCD camera. The USTB dataset includes the following four subsets:

Subset1: 60 subjects, 3 ear images per subject.

Subset2: 77 subjects, 4 ear images per subject. The images are taken under illumination variation and pose variation within ±35°.

Subset3: 79 subjects, 40 profile images per subject. The images are taken under different poses (frontal ear, left(+), and right(-) rotation of ±5°, ±10°, ±15°, ±20°, ±25°, ±30°, ±35°, ±40°, ±45°). For the 24 subjects in this dataset, partially occluded images are collected to form the subset of partial occlusion.

Subset4: 500 subjects, 85 profile images per subject. The images are taken under pose variation (the five poses are frontal, tilt ±30°, yaw ±30°). For each pose, 17 images at

10° interval are acquired. This subset is designed for ear recognition under pose variation and multimodal recognition using face and ear. Figure 2 shows some example images of this dataset.

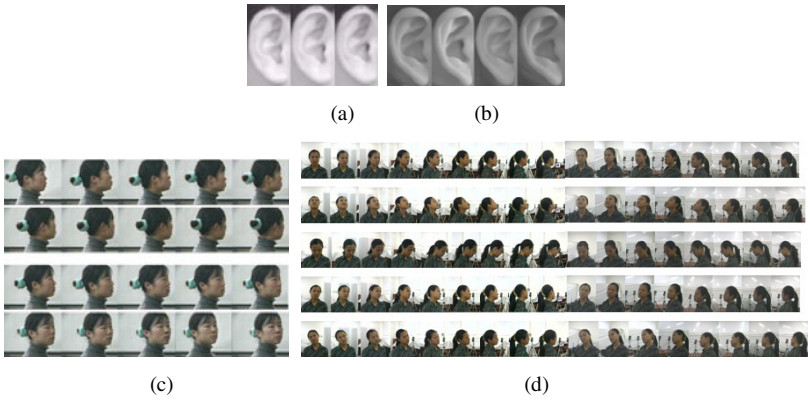


Fig. 2. Example images of USTB ear dataset. (a) ear image from subset1; (b) ear image from subset2; (c) ear image from subset3; (c) ear image from subset4.

(4) The Head Rotation Shot belongs to XM2VTSDB dataset. The acquisition equipment is Sony VX1000E digital cam-corder. There are a total of 2,360 images, one left profile and one right profile image per subject. Figure 3 shows some example images of this dataset.

Fig. 3. Example images of XM2VTS face profile dataset

2.2 Automated Ear Detection

This step mainly focuses on detecting and tracking human ear from the input video and returning the location and extent of each ear in the frame if one or more ears are present. In [8] and [9], automated ear detection methods under complex background using cascaded Adaboost were proposed. These methods had two stages: off-line cascaded classifier training and on-line ear detection. In the off-line training stage, both methods added some extended haar-like features to build the weak classifiers. Ref. [8] used traditional Gentle AdaBoost algorithm to train the strong classifiers to form the cascaded multi-layer ear detector. The training process in Ref. [8] is time consuming (about 20 days). So Ref. [9] applied a modified cascaded AdaBoost algorithm reducing the training time to about 8 hours. In the on-line detection stage, the detector is scanned over the test images in different sizes and locations. In Ref. [8], the full cascade of 18 stages achieved a false reject rate of 1.61% and a false accept rate of 2.53% on a total of 434 images from three image datasets. In Ref. [9],

the full cascade of 19 stages achieved a false reject rate of 5.31% and a false accept rate of 3.50% on a total of 2113 images from four image datasets. Future work for these automated ear detection may focus on improving the system performance in case of partial occlusion and rotation variation.

Ref. [10] proposed a four-step ear extraction method using 2D images and 3D range images: profile extraction by skin detection, ear pit detection using curvature estimation, ear extraction using active contour algorithm and ear cropping with 3D ear image. The disadvantage of this method is the dependence on the specific location of the nose tip. And the active contour algorithm may fail if there are no gradient changes in either colour or depth image. An improvement might focus on using shape and texture constraints to help the segmentation.

2.3 Ear Recognition Using 3D Data

Ear recognition using 3D data addressed such problems as illumination variation or orientation variation very well. There are two kinds of recognition methods using 3D ear data. The first kind is based on model matching. In [10], an automatic ear biometric system using 2D and 3D information was presented. An ICP-based approach was applied for 3D shape matching. On UND collection J2, this method achieved a rank-one recognition rate of 97.8% for identification and an EER of 1.2% for verification. However, the ICP algorithm involves in intensive computation and it is easy to fall into local minimum. Future research work may focus on developing a fast method for model matching and defining appropriate distance measurements regarding to the measurement of similarity among models.

Another kind of ear recognition method using 3D data is to extract distinctive 3D features first and then design some matching criterion. In [5], the author used local surface patch (LSP) for 3D feature extraction and a modified Iterative Closest Point (ICP) algorithm for ear matching. The root mean square (RMS) registration error was used as matching criterion. Later in their work [11], they firstly reduced the high dimensionality of LSP feature space by FastMap algorithm. Then the similarity between a model-test pair was computed using the LSP features. The similarities for all model-test pairs were ranked using SVM to generate a short list of candidate models for verification. The verification was performed by aligning a model with the test object using ICP algorithm. On the UND collection F, they got 96.7% for rank-1 recognition and EER of 1.8% for verification. The LSP descriptor can be regarded as 3D structure feature. In the future work, we may consider extracting more 3D ear recognition features including both the 3D structure information and the 3D texture information.

2.4 3D Ear Recognition Based on 2D Images

Most of current 3D recognition research works utilize range scanner for data acquisition. And to produce accurate scan results, many restrictions are imposed on the subject, e.g. the subject has to remain complete still during scanning for at least several seconds. Therefore, a more convenient and ideal way for non-intrusive recognition is to use 2D data sources for 3D recognition, which could be obtained with properly designed imaging system using regular video camera.

The key issues of 3D ear recognition based on 2D images are 3D ear reconstruction and 3D ear recognition. For ear reconstruction, there are mainly three kinds of methods: SFS (Structure from Shading) [12], SFM (Structure from Motion) [13] and binocular stereo vision [14, 15]. In [12], 3D ear reconstruction method based on SFS was proposed. The 3D ear model was constructed from single image. But the proposed method was sensitive to the illumination condition. In [13], 3D ear reconstruction based on SFM was proposed. This method was semi-automatic because the feature points were selected in a way of human-computer interaction. The reconstruction accuracy of this method depends on the location accuracy of image feature points as well as the accuracy of camera inner and outer parameter calibration. In [14], an automatic 3D ear reconstruction method based on binocular stereo vision was proposed. Firstly, SIFT feature based matching approach was used to compute the seed matches. Then match propagation algorithm with epipolar geometry constraint was used to obtain quasi-dense correspondence points. Finally, the 3D model was reconstructed by triangulation. In [15], 3D ear reconstruction based on epipolar geometry was proposed. The author used DAISY descriptor to describe the feature points, and epipolar constrain for dense matching.

The amount of vertices created by different methods are compared in Table 1. For 3D ear recognition based on 2D images, there still are many unresolved problems including 3D feature representation (such as 3D structure descriptor construction and 3D texture representation), fast 3D model matching.

Table 1. Comparison of different reconstruction methods

Methods	Automatic/ interactive	Vertices
range data [11]	automatic	7000-9000
multiview [13]	interactive	300-600
3D reconstruction [14]	automatic	2000-3000
3D reconstruction [15]	automatic	5000-8000

2.5 Ear Recognition in 2D

At present, 3D ear recognition based on 2D images needs expensive computation and more than one facility, so most of the reported publications on ear recognition are focused on 2D images. Ear recognition in 2D can be categorized into three kinds: ear recognition under constrained environment, ear recognition with pose variation, ear recognition under partial occlusion.

(1) Ear recognition under constrained environment

Here, ear recognition under constrained environment means the training image and test image are taken under standard environment without illumination and pose variations.

In [16], the author extracted geometric feature of the ear image including shape and structural features, and applied nearest neighborhood classifier for recognition. In [17], the author combined SIFT descriptors and an auricle geometric feature to form the feature vector. In [18], the author proposed a multi-matcher system based on image division. Each matcher was trained using features extracted from the convolution of each sub-window with a bank of Gabor Filters. The most discriminative matchers were

combined for the final recognition. In [19], the author proposed force field transformation based ear recognition. The force field of the ear images were taken and then converted to convergence fields. Then Fourier based cross-correlation techniques were used to perform multiplicative template matching on ternary thresholded convergence maps. In [20], the author compared face and ear biometrics using PCA, and reached the conclusion that they had similar recognition performance on the Human ID database.

(2) Ear recognition with pose variation
Most research papers proposed improved kernel based method or manifold learning algorithm for ear recognition with pose variation. In [21], the author proposed an improved locally linear embedding algorithm for multi-pose ear recognition. In [22], the author applied improved independent component analysis to extract the local and global features, then the features were combined by series weighed strategy and the dimension of new feature was descended by kernel principal component analysis.

(3) Ear recognition under partial occlusion
In applications, ear occlusion by hair or earrings is always a problem in uncontrolled environments. In [23], the author built an ear model using a stochastic clustering on the SIFT key points detected on the training images, and used the log-Gabor filter to exploit the variations between the boundary curves of the ear. In [24], the ear was treated as a planar surface, and ear recognition was done by creating a homography transform using SIFT feature matches. In [25], the author proposed a sparse representation based ear recognition approach. A test ear image to be identified can be represented as the sparse linear combination of the training images plus the sparse err produced by noise or partial occlusion on the test ear image. These methods showed good recognition performance when the occlusion percentage is within 20%.

2.6 Multimodal Recognition Using Face and Ear

For a robust non-intrusive biometrics recognition, eliminating the influence of pose variation, occlusion, expression etc. turns out to be a more demanding work. Similar ways acquiring the biometric data at distance and obviously complementary physical position of face and ear make the fusion of the two an appealing, natural and convenient multi-modal identification scheme. In open literature, the research works of multimodal fusion of ear and face that have been reported so far are the fusion recognitions based on: 2D images of ear and frontal face [20,26,27], laser scanned 3D data of ear and frontal face[28,29], as well as reconstructed 3D ear and 2D frontal face [30]. However, research on the fusion of 3D reconstructed ear, frontal face and profile face are still open issues.

The multimodal recognition of ear and face are performed on the following aspects: a). feature level fusion based recognition method [26-28]. The feature level fusion aims to enhance the discriminating power of the feature vector and improve classifier performance. To start with forming fusion feature, ear feature and face feature will be separately combined first to form new high-dimensional feature vectors. And then, in the corresponding new feature spaces, fusion features are selected. The main issue on this aspect is to design the learning based robust feature selection method, capable of improving the discriminating ability while reducing the dimension of the feature

vectors. b). Score level based recognition strategy[29,30]. Separate classifiers are designed for ear and face respectively, and the matching scores of each classifier are combined according to fusion rules to realize more robust multimodal recognition. The key issue on this aspect is to evaluate the performance of different fusion rules and propose new fusion strategies subjected to pose variation or occlusion.

3 Conclusions and Future Interests

Ear recognition has become an active research field for its potential use in commercial and law enforcement applications such as access control, security monitoring and video surveillance. Like face recognition, ear recognition is a user-friendly non-intrusive system for personal identification. This paper reviews recent advances in ear recognition methods focusing on 3D recognition and 2D recognition. Some representative works in these fields are discussed in the paper. The review shows that research on ear recognition is still a growing field. There still exists a gap between the research work and real application.

For 3D recognition, faster feature extraction and matching techniques should be developed to be applicable to large scale recognition applications. For 2D recognition, some future research interests may be focused on ear registration and alignment, developing approaches to deal with illumination, pose problems and occlusion. New feature descriptors representing the unique feature of ears are yet to be discovered for ear recognition. Last but not least aspect is the realization of real time ear recognition system for real applications.

References

1. Iannarelli, A.: Ear Identification. Forensic Identification Series. Paramount Publishing Company, Fremont (1989)
2. Choraś, M.: Intelligent Computing for Automated Biometrics, Criminal and Forensic Applications. In: Huang, D.-S., Heutte, L., Loog, M. (eds.) ICIC 2007. LNCS, vol. 4681, pp. 1170–1181. Springer, Heidelberg (2007)
3. Alberink, I., Ruifrok, A.: Performance of the FearID earprint identification system. Forensic Science International 166(2), 145–154 (2007)
4. http://www.nd.edu/~cvrl/CVRL/CVRL_Home_Page.html
5. Bhanu, B., Chen, H.: Human Ear Recognition by Computer. Springer, Heidelberg (2008) ISBN: 978-1-84800-128-2
6. http://www.ustb.edu.cn/resb/
7. http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/
8. Yuan, L., Zhang, F.: Ear Detection based on Improved Adaboost Algorithm. In: Proceedings of the 2009 International Conference on Machine Learning and Cybernetics, pp. 2414–2417 (2009)
9. Abaza, A., Hebert, C., Harrison, M.A.F.: Fast Learning Ear Detection for Real-time Surveillance. In: Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS), pp. 1–6 (2010)
10. Yan, P., Bowyer, K.W.: Biometric Recognition Using 3D Ear Shape. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(8), 1297–1308 (2007)

11. Chen, H., Bhanu, B.: Efficient Recognition of Highly Similar 3D Objects in Range Images. IEEE Transactions on Pattern Analysis And Machine Intelligence 31(1), 172–179 (2009)
12. Cadavid, S., Abdel-Mottaleb, M.: 3D Ear Modeling and Recognition From Video Sequences Using Shape From Shading. IEEE Transactions on Information Forensics And Security 3(4), 709–718 (2008)
13. Liu, H., Yan, J.Q., Zhang, D.: 3D Ear Reconstruction Attempts: Using Multi-view. In: International Conference on Intelligent Computing, pp. 578–583 (2006)
14. Zeng, H., Mu, Z.C., et al.: Automatic 3D Ear Reconstruction Based on Binocular Stereo Vision. In: 2009 IEEE International Conference on Systems, Man, and Cybernetics, Texas, USA, pp. 1–5 (2009)
15. Sun, C., Mu, Z.C., Zeng, H.: Automatic 3D Ear Reconstruction Based on Epipolar Geometry. In: ICIG 2009, pp. 496–500 (2009)
16. Yuan, L., Mu, Z., Xu, Z.: Using Ear Biometrics for Personal Recognition. In: Li, S.Z., Sun, Z., Tan, T., Pankanti, S., Chollet, G., Zhang, D. (eds.) IWBRS 2005. LNCS, vol. 3781, pp. 221–228. Springer, Heidelberg (2005)
17. Tian, Y., Yuan, W.Q.: Ear Recognition based on Fusion of Scale Invariant Feature Transform and Geometric Feature. Acta Optica Sinica 28(8), 1486–1491 (2008)
18. Nanni, L., Lumini, A.: A multi-matcher for Ear Authentication. Pattern Recognition Letters 28(16), 2219–2226 (2007)
19. Hurley, D., Nixon, M., Carter, J.: Force field feature extraction for ear biometrics. Computer Vision and Image Understanding 98(3), 491–512 (2005)
20. Chang, K., Bowyer, W.K., et al.: Comparison and Combination of Ear and Face Images in Appearance-Based Biometrics. IEEE Trans. on Pattern Analysis and Machine Intelligence 25(9), 1160–1165 (2003)
21. Xie, Z.X., Mu, Z.C.: Ear Recognition Using LLE and IDLLE Algorithm. In: The 19th International Conference on Pattern Recognition (2008)
22. Dun, W.J., Mu, Z.C.: An ICA Baased Ear Recognition Method through Nonlinear Adaptive Feature Fusion. Journal of Computer Aided Design & Computer Graphics 21(3), 383–388 (2009)
23. Arbab-Zavar, B., Nixon, M.: Robust Log-Gabor Filter for Ear Biometrics. In: Proc. 20th International Conference on Pattern Recognition (2008)
24. Bustard, J.D., Nixon, M.: Robust 2D Ear Registration and Recognition Based on SIFT Point Matching. In: Proc. 2nd IEEE Conference on Biometrics: Theory, Applications and Systems (2008)
25. Yuan, L., Fu, W., Mu, Z.C.: An Automatic Ear Recognition Approach. In: Proceedings of the 30th Chinese Control Conference, China, pp. 3310–3314 (2011)
26. Yuan, L., Mu, Z.C.: Multimodal Recognition using Ear and Face. In: 5th International Conference on Wavelet Analysis and Pattern Recognition, Beijing (2007)
27. Dun, W.J., Mu, Z.C.: Multi-modal Recognition of Face and Ear Images Based on Two Types of Independent Component Analysis. Journal of Computational Information Systems 4(5), 1977–1983 (2008)
28. Theoharis, T., Passalis, G., et al.: Unified 3D Face and Ear Recognition using Wavelets on Geometry Images. Pattern Recognition 41(3), 796–804 (2008)
29. Islam, S.M.S., Bennamoun, M., Mian, A.S., Davies, R.: Score Level Fusion of Ear and Face Local 3D Features for Fast and Expression-Invariant Human Recognition. In: Kamel, M., Campilho, A. (eds.) ICIAR 2009. LNCS, vol. 5627, pp. 387–396. Springer, Heidelberg (2009)
30. Mahoor, M.H., Cadavid, S., Abdel-Mottaleb, M.: Multi-modal Ear and Face Modeling and Recognition. In: Proceeding of International Conference on Image Processing (2009)

SDUMLA-HMT: A Multimodal Biometric Database

Yilong Yin, Lili Liu, and Xiwei Sun

School of Computer Science and Technology, Shandong University,
Jinan, 250101, China
ylyin@sdu.edu.cn, ll_liu@yahoo.com.cn, sun_xiwei@hotmail.com

Abstract. In this paper, the acquisition and content of a new homologous multimodal biometric database are presented. The SDUMLA-HMT database consists of face images from 7 view angles, finger vein images of 6 fingers, gait videos from 6 view angles, iris images from an iris sensor, and fingerprint images acquired with 5 different sensors. The database includes real multimodal data from 106 individuals. In addition to database description, we also present possible use of the database. The database is available to research community through http://mla.sdu.edu.cn/sdumla-hmt.html.

Keywords: Multi-modal, Homologous, Biometrics, Face, Finger vein, Gait, Iris, Fingerprint.

1 Introduction

The last decades has witnessed great advances in biometric recognition techniques. And lots of biometric traits have been used for identification authentication, such as fingerprint, face, iris, etc. Currently, biometric recognition system can reach very high accuracy when tested on the open biometric databases. However, due to inherent properties of biometric traits and the constraints of sensing technologies, the performance of individual biometric system is limited. Therefore, many researchers recently put their efforts to multimodal biometric fusion [1][2][3].

Multimodal biometric fusion, which combines two or more biometric traits, is an effective way to alleviate some of the limitations of single biometric system. It can improve the overall matching accuracy and make the biometric systems invulnerable to security threats. There are several approaches to study biometrics fusion. One approach is to use heterogeneous database [3], i.e., combine biometric trait (e.g. fingerprint) from a database with biometric trait (e.g. face) from another database. From the experiment point of view, these combined biometric traits belong to the same person. And the resultant person is called as *chimeric user*. Although this approach has been widely used in multimodal literature, it was questioned that whether this approach was reasonable during the 2003 Workshop on Multimodal User Authentication [4]. Poh et al. [5] studied this problem and showed that the performance measured with experiments carried out on *chimeric users* does not necessarily reflect the performance with true multimodal users. Obviously, the best way to study biometrics fusion is to use homologous multimodal biometric databases, which means the different biometric traits are truly come from the real same person.

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 260–268, 2011.

However, there are only a few multimodal biometric databases publicly available. And most of the existing multimodal databases are composed two modalities. BANCA [6] and XM2VTS [7] include face and voice; MYCT [8] includes fingerprint and signature. Besides, there are also several databases including more than two modalities, such as BIOMET [9] which includes face, voice, fingerprint, hand and signature, and BioSec [10] including fingerprint, face, iris and voice. These existing databases have several limitations, e.g., lack of import traits or lack of diversity of sensors/traits. Therefore, we build the SDUMLA-HMT database which is composed of homologous multimodal biometric traits, including face images from 7 view angles, finger vein images of 6 fingers, gait videos from 6 view angles, iris images from an iris sensor, and fingerprint images acquired with 5 different sensors. To the best of our knowledge, gait is the only biometric trait which can be recognized from a far distance, and finger vein is a kind of biometric traits which use information hidden inside of skin. Therefore, gait is non-invasive and it is difficult to conceal or disguise while finger vein cannot be damaged or forged easily. So, it is very profound to add gait and finger vein to the family of multimodal biometric database. The detailed characteristics of the SDUMLA-HMT database are described in Section 2; Section 3 gives conclusions.

2 SDUMLA-HTM Database

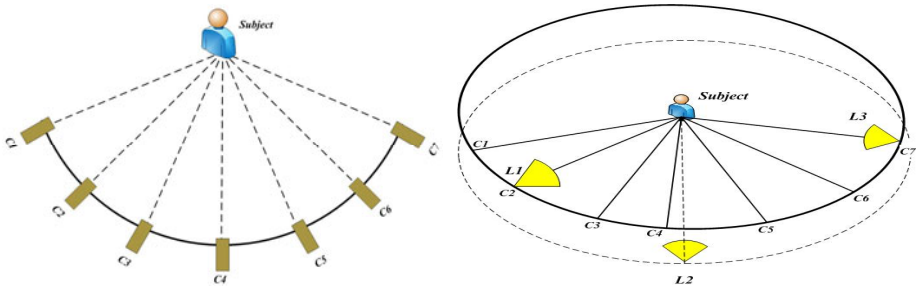
SDUMLA-HMT was collected during the summer of 2010 at Shandong University, Jinan, China. 106 subjects, including 61 males and 45 females with age between 17 and 31, participated in the data collecting process, in which all the 5 biometric traits – face, finger vein, gait, iris and fingerprint are collected for each subject. Consequently, there are 5 sub-databases included in SDUMLA-HMT, i.e., a face database, a finger vein database, a gait database, an iris database and a multi-sensor fingerprint database. It is to be noted that in the 5 sub-databases, all the biometric traits with the same person id are captured from the same subject. We will detail the 5 sub-databases in the next 5 subsections.

2.1 Face Database

Face recognition is a relatively mature technology in biometrics. The face database included in the SDUMLA-HMT is aiming at real-world face recognition. In order to simulate real-world settings, we capture faces with different poses, facial expressions, accessories and illuminations. Section 2.1.1 introduces the carefully designed simulate environment, followed by section 2.1.2 describing the database.

2.1.1 Environmental Setting

Camera Setting: We use 7 ordinary digital cameras to symmetrically capture both sides of the face. As shown in Fig. 1(a), all the cameras are fixed in a circle centered at the subject with radius 50 *cm* and angle interval 22.5°. And the subject was asked to sit towards the center camera *C4*. The height of the cameras is set to be 110 *cm* by tripods. All the 7 cameras work simultaneously.



(a) Camera Setting (b) Illumination Setting

Fig. 1. Environmental Setting

Illumination Setting: To simulate varying illumination condition, we design different illuminations using 3 lamps labeled as *L1*, *L2* and *L3*. As shown in Fig. 1(b), *L1* is nearby *C2, L3* is nearby *C7* and *L2* is under *C4*. It is to be noted that the direction of *L2* is from down to up to the subject. Each time only one lamp is on.

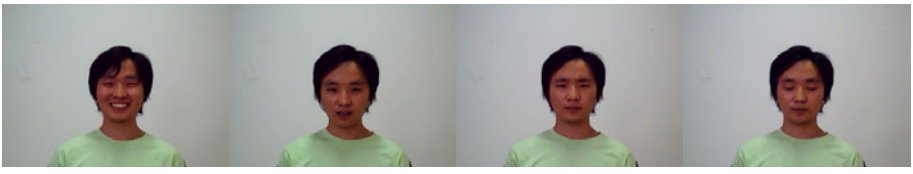
Accessories: To simulate the influences of different accessories, we prepared two common accessories -- a pair of glasses and a hat.

2.1.2 Database Description

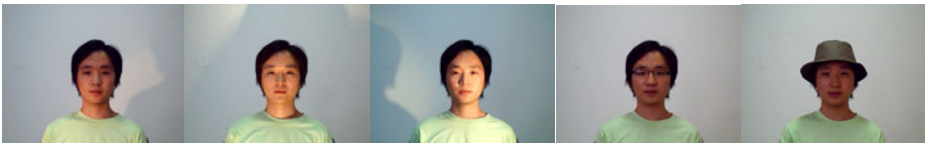
Four variations, including poses, facial expressions, illuminations, and accessories, are considered in the face data acquisition process. Using normal illumination, i.e., no lamp is on, face images with 3 type of poses (look upward, forward, and downward), 4 type of expressions (smile, frown, surprise, and close eyes) and 2 type of accessories (glasses and hat) were captured in our face database. And each time with one lamp on, we captured face images with 3 types of illuminations. Sample images of our face database are shown in Fig. 2.

(a) Pose Variation

Fig. 2. Sample images in the face database



(b) Expression Variation



(c) Illumination Variation (d) Accessory Variation

Fig. 2. (*continued*)

The face database contains 7× (3+4+2+3) ×106=8,904 images. All images are 24 bit *"bmp"* files and the size of image is 640×480 pixels. The total size of the face database is 8.8G.

2.2 Finger Vein Database

Finger vein recognition is a recently developed research hotspot. We include in SDUMLA-HMT a finger vein database which, to the best of our knowledge, is the first open finger vein database.

The device used to capture finger vein images is designed by Joint Lab for Intelligent Computing and Intelligent Systems of Wuhan University. In the capturing process, each subject was asked to provide images of his/her index finger, middle finger and ring finger of both hands, and the collection for each of the 6 fingers is repeated for 6 times to obtain 6 finger vein images. Some sample images are shown in Fig. 3.

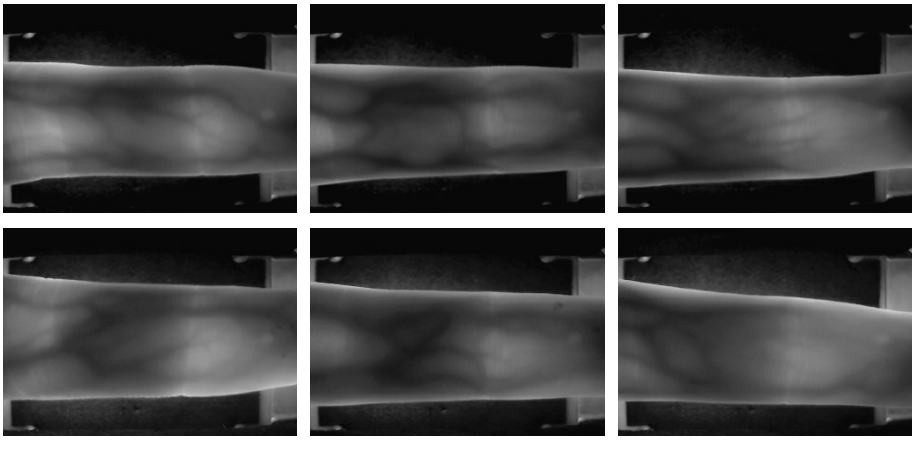


Fig. 3. Sample images in the finger vein database

The finger vein database is composed of 6×6×106=3,816 images. Every image is stored in *"bmp"* format with 320×240 pixels in size. The total size of our finger vein database is around 0.85G Bytes.

2.3 Gait Database

Gait recognition is a newly arisen research area in biometrics. So, there are few open gait databases. To the best of our knowledge, the gait database included in SDUMLA-HMT is one of the largest databases among the existing gait databases, and SDUMLA-HMT is the first multimodal biometric database which contains gait. As the gait capture process needs complicated environment, we first introduce environmental setting in Section 2.3.1, and then describe the details of the gait database in Section 2.3.2.

2.3.1 Environmental Setting

We capture gait data in a specially designed room. As shown in Fig. 4, we assigned 6 ordinary digital cameras labeled *C1* to *C6* along the right hand side of the subject, and all the cameras were fixed in a circle centered at the middle of walking road with radius 6 *m*. Here *C3* was set in the vertical direction of the walking path, and *C2* and *C1* (*C5* and *C4*) symmetrically spread around *C3* with angle interval 22.5°. *C6* was fixed along walking direction to capture the frontal view. To avoid reflection on the floor, we paved an $8m \times 2m$ sized carpet along the walking path. Besides, a calibration tap with $10cm \times 20cm$ white-yellow alternative blocks was placed to facilitate the reconstruction of geometry information, and the height of the tap is 2 *m*.

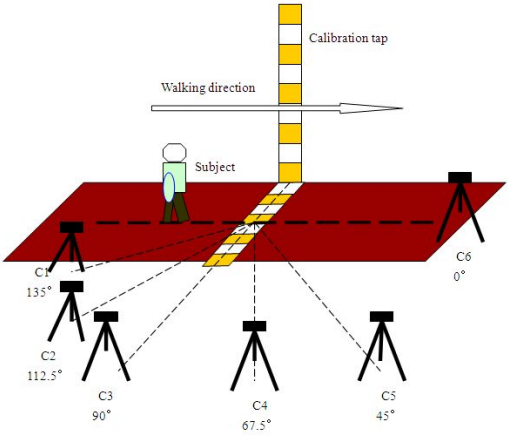


Fig. 4. Environmental setting for gait collection

2.3.2 Gait Database

Three variations, including view angle, accessory and motion type, are considered in the gait data acquisition process. For each subject, we first captured 6 background videos using the 6 cameras before his/her walking. Then the subject was asked to

walk naturally along the walking direction for 6 times. After that, the subject was asked to carry a bag and walked twice again. The bag could be a knapsack, a satchel, or a handbag chosen according to the subject's preference. Furthermore, we also recorded twice of the subject's running videos. As a result, there are totally 66 videos recorded for the subject. Fig. 5 shows some sample snapshots of the gait videos.

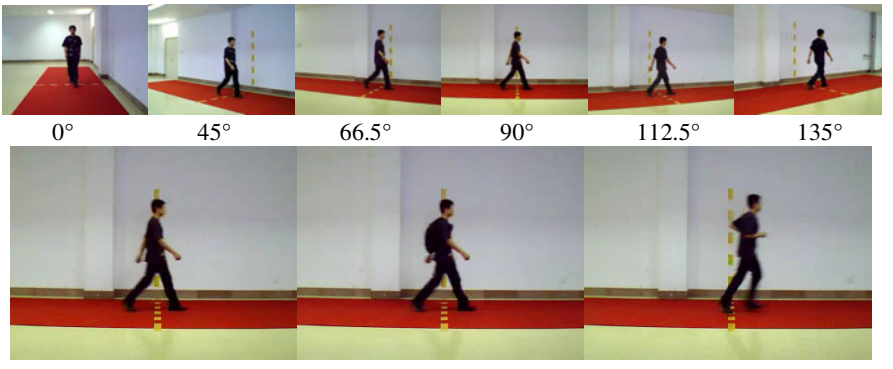


Fig. 5. Sample snapshots in the gait database

The gait database is composed of 6×11×106=6,996 videos in total and each video records about 2 to 3 gait cycles. All the videos are stored in *"avi"* format encoded with XviD codec. The frame size is 320×240 *pixels*, and the frame rate is 25 *frames per second*. The total size of the gait database is about 1.6G Bytes.

2.4 Iris Database

Iris recognition is a top research focus in recent years. Statistical tests made in [11] show that in all the biological characteristics, iris has the most reliable and most stable features. Therefore, we include in SDUMLA-HMT an iris database.

We collected the iris data with an intelligent iris capture device developed by University of Science and Technology of China under near infrared illumination. To avoid reflection, the subjects were asked to take off their glasses and to keep the distance between the eye and the device within 6*cm* to 32*cm*. Every subject provided 10 iris images, i.e., 5 images for each of the eyes. Fig. 6 provides 4 sample images in our iris database.

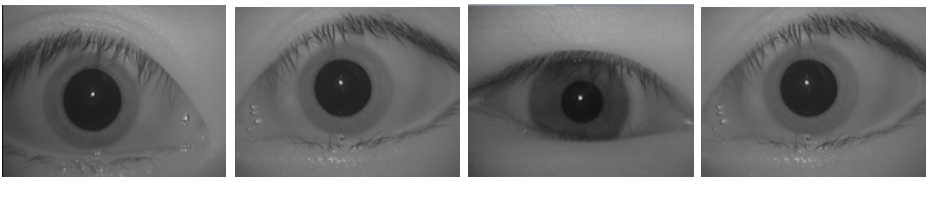


Fig. 6. Sample images in the iris database

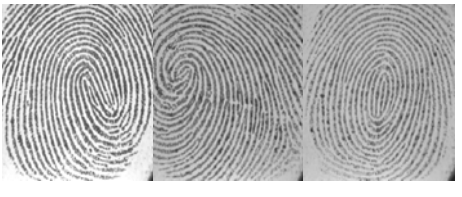
The iris database is composed of 2×5×106=1,060 images. Every iris image is saved in 256 gray-level *"bmp"* format with 768 × 576 *pixels* in size. The total size of our iris database is about 0.5G Bytes.

2.5 Multi-sensor Fingerprint Database

Fingerprint is one of the most commonly used biometric traits in biometric authentication. Therefore, we include in a fingerprint database in SDUMLA-HMT. In addition, the fingerprint database acquired with 5 sensors will also enrich research on the sensor interoperability of fingerprint recognition which is a hotspot in fingerprint recognition recently.

Our fingerprint database includes fingerprint images captured from thumb finger, index finger and middle finger of both hands. In order to explore the sensor interoperability, we captured each of the 6 fingers with 5 different type of sensors, i.e., AES2501 swipe fingerprint scanner developed by Authentec Inc, FPR620 optical fingerprint scanner and FT-2BU Capacitive fingerprint scanner both developed by Zhongzheng Inc, URU4000 optical fingerprint scanner developed by Zhongkong Inc and ZY202-B optical fingerprint scanner developed by Changchun Institute of Optics, Fine Mechanics and Physics, China Academy of Sciences. It is to be noted that 8 impressions were captured for each of the 6 fingers using each of the 5 sensors. Some sample images of the fingerprint database are shown in Fig. 7.

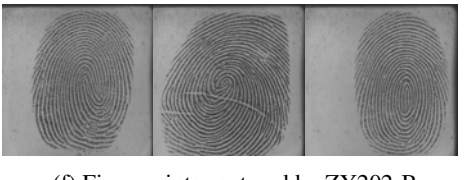
(a) Fingerprints captured by AES2501



(b) Fingerprints captured by FPR62

(c) Fingerprints captured by FT-2BU

(d) Fingerprints captured by URU4000



(f) Fingerprints captured by ZY202-B

Fig. 7. Sample images in the fingerprint database

Our fingerprint database contains 6×5×8×106=25,440 fingerprint images in total. Every fingerprint image is saved in 256 gray-level *"bmp"* format but the size varies according to the capturing sensors. Table 1 lists the size of images captured by the 5 sensors. It is worth noting that AES2501 is a swipe fingerprint scanner and the image size varies in different swiping processes, which has been shown in Fig. 7(a). The total size of the interoperability-based fingerprint database is about 2.2G Bytes.

Table 1. Image size captured by the 5 sensors

Sensor	Image Size
AES2501 swipe fingerprint scanner	not fixed
FPR620 optical fingerprint scanner	256×304
FT-2BU capacitance fingerprint scanner	152×200
URU4000 optical fingerprint scanner	294×356
ZY202-B optical fingerprint scanner	400×400

3 Conclusions

In this paper, we describe a homologous multimodal biometric database — SDUMLA-HMT which contains 5 different biometric traits of 106 subjects. This database will be very useful to study different kind of biometric fusions, which is a profound research area of biometric recognition. And the database is available to research community through http://mla.sdu.edu.cn/sdumla-hmt.html.

Acknowledgments. Thanks to all of the members of MLA Group who participate in this work. This work is partly supported by National Natural Science Foundation of China under Grant No. 61070097, the Research Found for the Doctoral Program of Higher Education under Grant No. 20100131110021 and Natural Science Foundation of Shandong Province under Grant No. Z2008G05.

References

1. Kittler, J., Matas, G., Jonsson, K., Sanchez, M.: Combining evidence in personal identity verification systems. Pattern Recognition Letters 18(9), 845–852 (1997)
2. Snelick, R., Indovina, M., Yen, J., Mink, A.: Multimodal Biometrics: Issues in Design and Testing. In: Proc. of The 5th International Conference on Multimodal Interfaces (IMCI 2003), Vancouver, British Columbia, Canada (November 2003)
3. Ross, A., Jain, A.: Information fusion in biometrics. Pattern Recognition Letters 24, 2115–2125 (2003)
4. Dugelay, J.-L., Junqua, J.-C., Rose, K., Turk, M. (Organizers): In: Workshop on Multimodal User Authentication (MMUA 2003), Santa Barbara, CA (December 2003)

5. Poh, N., Bengio, S.: Can Chimeric Persons Be Used in Multimodal Biometric Authentication Experiments? In: Renals, S., Bengio, S. (eds.) MLMI 2005. LNCS, vol. 3869, pp. 87–100. Springer, Heidelberg (2006)
6. Bailly-Baillière, E., Bengio, S., Bimbot, F., Hamouz, M., Kittler, J., Mariéthoz, J., Matas, J., Messer, K., Popovici, V., Porée, F., Ruiz, B., Thiran, J.-P.: The BANCA Database and Evaluation Protocol. In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003. LNCS, vol. 2688, Springer, Heidelberg (2003)
7. Poh, N., Bengio, S.: Database, protocols and tools for evaluating score-level fusion algorithms in biometric authentication. Pattern Recognition 39, 223–233 (2006)
8. Ortega-Garcia, J., Fierrez-Aguilar, J., et al.: MCYT baseline corpus: a bimodal biometric database. IEEE Proc. Vis. Image Signal Process. 150, 395–401 (2003)
9. Garcia-Salicetti, S., Beumier, C., Chollet, G., et al.: BIOMET: a Multimodal Person Authentication Database Including Face, Voice, Fingerprint, Hand and Signature Modalities. In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003. LNCS, vol. 2688, pp. 845–853. Springer, Heidelberg (2003)
10. Fierrez, J., Ortega-Garcia, J., Toledano, D.T., Gonzalez-Rodriguez, J.: Biosec baseline corpus: A multimodal biometric database. Pattern Recognition 40, 1389–1392 (2007)
11. Mansfield, T., Kelly, G., Chandler, D., Kane, J.: Biometric product testing final report. Nat. Physical Lab., Middlesex, U.K (2001)

Study of Human Identification by Electrocardiogram Waveform Morph

Gang Zheng[1,2], Zhong-Yi Li[1,2], Tong-Tong Liu[1,2], and Min Dai[1,2]

[1] Laboratory of bio signal and intelligent processing, Tianjin University of Technology, 300384, Tianjin, China
[2] School of Computer and Communication Engineering, Tianjin University of Technology, 300384, Tianjin, China
kenneth_zheng@vip.163.com, sailangwudi@163.com, 765351075@qq.com, daimin@tjut.edu.cn

Abstract. A new biometric recognition material electrocardiogram (ECG) waveform was developed rapidly in recent ten years. Except the common feature of biometric recognition, its unique "aliveness" and individual difference of heart geometric structure, This make ECG waveform has becoming a kind of high security level biometric. The paper proposed a similarity measurement strategy to do recognition work by ECG waveform. It uses the ECG waveforms collected from one person as his or her ECG waveform sample set. These ECG waveforms were partitioned into single ECG waveform (which is generate by one heart beat) firstly. And the discrete points that formed single ECG waveforms were overlaid into a two dimension coordinate. The points which have same coordinate will be accumulated, and the color of this coordinate will be changed into a 8 bit color system according to the overlay number of point. After the procedures, it presents a hierarchy color changing image, which can be used as a tunnel like ECG morph. Moreover, the inclusive degree can be computed by color clustering status. In the end, the morph will become the morph model of the person to judge the belonging of new ECG data. The ECG data used in the paper is from MIT/BIH ECG standard data set. From the results, the recognition accurate of ECG recognition can reach to 95.1% averagely.

Keywords: Biometric recognition, ECG identification, Similarity measurement, ECG waveform, ECG morph, Inclusive value.

1 Introduction

In the study of information security, the appearance of new biometric will be broadly concerned by many researching fields. The fusion with other developed biometrics (that is the generalization and application of multi bioinformatics identification model) will provide ability to meet the requirement of higher level security. And ECG (Electrocardiogram) is one of this biometric.

Since the correspondence between ECG and heart disease, the reliability of heart disease diagnosis by ECG had been proved. In the near decades, the study of ECG has

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 269–278, 2011.

been expanded to bioinformatics, and especially in biometric which is used for human identification. One of the most important characteristic in ECG is its "aliveness". The ECG waveform of a person is changing all the time, especially in different postures, active status, time period and environment. This advantage cannot be acquired from other biometrics, like fingerprint, iris, palm print, face, etc… And depended on this, ECG waveforms have become the material which is use to build up password system.

Previous study of this had gotten some achievements, but all of them are used to describe the first circumstance. Hoekema[1] proposed and proved that the difference of different person's ECG signal is mostly originated from the electrophysiology information that reflected by heart geometric structure. Simon[2] presented the same idea in 1997 and describe the feasibility of ECG as biometric. Bie[3] provided realization strategy of ECG identification, and describe the "aliveness" of ECG in the same year. In his paper, the features of ECG were presented and these features (such as time period of key point, amplitude and slope) were used for human identification. Other researchers [4-7] also provide some supporting information on the personality of ECG waveform. These individual characteristics are produced because of the geometrics of heart.

2 Related Works

Since 1997, lots of researchers paid their attention on the study of ECG identification. From then to now, two studies directions are formed. One is focusing on the reorganization of ECG's key points. The other one adopts frequency analysis on ECG waveforms, especially by wavelet transformation.

In the first direction, lots of progress had been made. In 1997, M.Ogawa[8] use wavelet and neural network to distinguish two individuals by ECG. And in 1999, Biel[3] extracted 30 features from ECG waveform as candidate features. The analysis of a correlation matrix was employed to reduce the dimensionality of features to 12. With 20 participants, the experimental results yielded identification rate between 90-100% using empirically selected features.

In 2001, M. Kyoso [9] distinguished 9 individuals by Mahalanobis distance. In 2002, Shen[10] extracted 7 temporal and amplitude features from the QRST wave. 20 subjects were distinguished by combining template matching and a decision-based neural network(DBNN), and the experiment result showed that the rate of correct identity verification was 100%. In 2005,Kim[11] extracted the time period of ECG as features, and tested on 10 individuals with 30 seconds ECG data. In 2006 Wei[12] extracted 14 features based on the fiducial points of ECG waveform from 50 individuals, and Bayes' theorem based classification method was adopted. In the paper, they proved that VI and VII lead of ECG data is better than other leads in recognition. SAIC（Science Applications International Corporation）moved their eyes on ECG identification. In 2005 and 2008, Israel and Irvine[4][13] extracted 15 features that were time duration between fiducial points. The Wilkes' lambda method was employed to select features and Linear Discriminant Analysis for classification. In paper [13] the eigenPluse of ECG was used to improve identification rate. They merged

ECG and face recognition to identify human, and the result is better than face recognition alone [14].

Another research group is Edward S. Rogers from electronic & computer engineering department of Toronto University. They started the study from 2003. In 2006, To improve the identification accuracy, Wang[15] proposed a hierarchical pattern recognition and suggested an integration of 21 analytic features（temporal and amplitude features）. The hierarchical scheme achieves subject recognition rate of 100% for both data sets, and a heartbeat recognition rate of 98.90% for PTB and 99.43% for MIT-BIH. In 2008, Gahi[16] extracted 24 features based on fiducials of ECG on 16 individuals, the accuracy was 100%.From the achievements of this study field, the recognition accuracy rate is high, and sometime reached 100%. But the dataset was relatively small, usually below 30. Since ECG waveform is a very weak signal, and lots of interfere noise is still along with it, and the morph of same person's ECG waveform will be different, especially when some heart diseases happened which will make ECG waveform changing. And in other aspect, the locations of fiducials of ECG are a problematic issue, such as P wave, T wave, etc…… Therefore, the identification work cannot really count on it. Another direction of study is doing identification work without fiducial.

In 2006 Chan[17] collected ECG data from 50 subjects during three data recording sessions on different days. From 2006, Hatzinas group of Toronto University turned their study direction to non-fiducial ECG identification study. They[18] used AC (Autocorrelations) to blend in all samples in the ECG window to a sequence of sum of products, and the actual location of fiducials would not be depended exactly. Waveform P, QRS, and T were proved to have larger contribution on identification. After that, DCT (Discrete Cosine Transform) was used to reduce parameter numbers. In the end Euclidean distance and Gaussian log likelihood were used as distinguish method. And the FP (False Positive) and FN(False Negative) are less than 1%.

Same group[19] studied 12 leads ECG data in 2008, and combining AC and LDA (linear discriminant analysis) were used to extract features. Tests were done on 14 individuals from PTB, and 100% identification rate was reached. In 2009, wavelet transform was used as feature selection method [20], and one of the novelties of this paper is the design of personalized heartbeat template so that the dataset consists of only one heartbeat per individual. The correlation coefficient was used of classification, and the identification accuracy rate was more than 99%. In 2010, on the base of AC/LDA pattern, periodicity transform (Heart Rate Various) was added as parameter [21], and the distinguish rate reached 92.3%. In same year, phonocardiogram and electrocardiogram were combination as fusion information, and wavelet transform is adopted as feature selection methods yet, after testing on 21 individuals, 97% recognition rate was reached.

Wavelet transform was also used by Ye [22], which selected 118 features and 18 ICA (Independent Component Analysis) components. These parameters were reduced to 26 by PCA (Principle Component Analysis). The recognition rate that depended on the 26 parameters reached 99.6%. In 2008,Chiu [23] employed wavelet transform to select the features of ECG to do the identification work, and the Euclidean distance was adopted for classification. According to the experiment results, identification rate

of 100% was reached on 45 normal individuals and 81% was reached on 10 arrhythmia patients.

On other study area, Irvine[13] proposed a strategy on the idea of face recognition in 2008. The ECG waveform data were transferred to a matrix, and the identification work was done by eigenPulse of the matrix. The recognition rate was reached 95%. And Shi[24] transferred the ECG waveform to a cipher key, which was used to identify individual in BAN(Body Area Network). It was used in IOT(Internet of Thing) and Intelligent Family.

3 ECG Waveform Morph

3.1 Related Formal Description

ECG waveform can be expressed by a periodical time sequence. In the paper, it was described by equation (1), and (2). In the equation, B is a period of ECG waveform (The length can be several seconds to several hours or more), Bi is one of individual ECG waveform, which is produce in a heartbeat.

$$B = (B^1, B^2, ..., B^m) \tag{1}$$

m is the number of individual ECG waveforms in B.

B^i is an individual ECG waveform.

$$B^i = (b_1, b_2, ..., b_n) \tag{1}$$

n is the number of digitals that form the single bio-signal waveform.

A point in bio-signal is signed as B^i_j, *i* is the i^{th} bio-signal waveform, *j* is the j^{th} point in bio-signal B^i.

The target of bio-signal analysis is to classify the waveforms. They can be variety of abnormal waveforms. Therefore, an assumption is presented, that is, the morph of bio-signal waveform changes mildly. They can be affected by abrupt big changes, but within a rather low probability. By introducing appropriate similarity measurement strategy, the partition rate is improved. And so does the inclusive of bio-signal waveform in classification. This is different from many previous studies on bio-signal waveform analysis.

3.2 ECG Waveform Comparison

According to the above analysis of feature selection on ECG, the comparison of two individual ECG waveforms is crucial to identification. But, the problem of ECG comparison is a linearly non separable case, traditional comparison and measurement, such as Euclidean distance, are unacceptable. We need to design a non-linear comparison method or measurement. It can distinguish the ECG in fuzzy and tolerant way. That is means; it highlights the common features, and minimized the different of ECG

on same person. And at the time, on the contractor, it highlights the difference, and minimized the common features of ECG on different person.

From the previous works on ECG human identification, most of them turned their study from key points' orientation to methods by filter or transformation analysis in frequency domain. And some achievements had been acquired. Since ECG waveform has morph characteristics, an idea of tunnel was proposed in the paper.

3.3 Idea of ECG Waveform Morph

A morph was shown in figure 1, ECG waveform *A* was drawn by a tunnel, the other waveforms which are similar to waveform *A* are contained in this tunnel. If the tunnel morph was defined, a class of ECG waveform was select and this can be used to identify individuals. The rest study will focus on the definition of tunnel width, tunnel description and tunnel measurement.

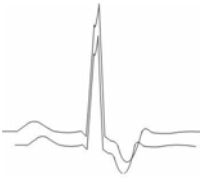


Fig. 1. Schematic diagram of ECG waveform morph

3.4 ECG Data Map to Morph

Map an individual ECG waveform to a dataset *A** based on two-dimensional coordination. Horizontal direction represents time, sign as T（$T = j \ \ and \ \ 1 \le j \le t$）, vertical direction is the discrete signal value a_j^i at time T. It is shown as Equation (3):

$$A^* = \begin{pmatrix} a_1^1 & a_2^1 & \cdots & a_t^1 \\ a_1^2 & a_2^2 & \cdots & a_t^2 \\ \vdots & \vdots & \vdots & \vdots \\ a_1^p & a_2^p & \cdots & a_t^p \end{pmatrix} \tag{2}$$

The mapping dataset *A** can form a m×n counter matrix *E*, the maximum and minimum value in dataset *A** was shown from Equation (4) to (7).

$$a_{min} = \min A^* \tag{3}$$

$$a_{max} = \max A^* \tag{4}$$

$$m = a_{max} - a_{min} + 1 \tag{5}$$

$$n = t \tag{6}$$

And after accumulating the emerging time (frequency, sign as *f*) of a_j^i in j^{th}, and established counter matrix *E*.

$$f^i = \sum_{j=1}^{t} E\left(a_{\max} - a_j^i, j\right) \tag{7}$$

3.5 Tunnel Morph Description Based on RGB Color

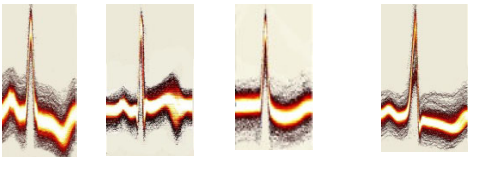
A strategy was proposed to create a link between the overlapping frequency of ECG signal and RGB color. The color is lightening along with the increasing of *f*. A simple example was given to describe the tunnel by a 8 bit RGB color system, expressed by Equation (9):

$$(0,0,0) \to (0,0,255) \to \cdots \to (255,255,255) \tag{8}$$

In Equation (9)， color changes in 8bit RGB color system. The map function between RGB and *f* is Equation (10):

$$\frac{f \times 2^8}{\max(E)} \to (r,g,b) \tag{9}$$

An ECG waveform dataset was selected to show the morphology of tunnel morph strategy. The Fig.2(a),(b),(c) and (d) gave several tunnel morphs established by ECG waveforms.



(a) (b) (c) (d)

Fig. 2. Tunnel liked morph of ECG waveform

3.6 Edge Extraction

In Figure 2, the red area was supposed the edge of tunnel w, it was expressed by two numbers L and H, L was low edge, and H is high edge.

$$w = [L, H] \tag{10}$$

$E[i,j]$ was the point in morph, Suppose that a_j^i is the point on edge, and its *f* exists in $E[i,j]$, its position can be judged by Equation (12) .

$$a_j^i \to \begin{cases} h_i \in H, H = (h_1, h_2, \cdots h_n) & when\ E[i-1,j] < E[i,j] < E[i+1,j] \\ l_i \in L, L = (l_1, l_2, \cdots l_n) & when\ E[i-1,j] > E[i,j] > E[i+1,j] \end{cases} \tag{11}$$

3.7 Individual Judgment

The attribution judgment of an ECG waveform can be done by Equation (12). Suppose an ECG sample data $S=(s_1,s_2,\ldots,s_t)$, it can be judged by Equation (12), by which whether signal *S* belongs to the tunnel can also be judged.

$$S \to T \begin{cases} S \in M^p & when \bigcap_{i=1}^{t} l_i^p \le S_i \le h_i^p = True \\ S \notin M^p & when \bigcap_{i=1}^{t} l_i^p \le S_i \le h_i^p = False \end{cases} \tag{12}$$

4 Experiment

4.1 Data Description

The paper use MIT/BIH ECG database as experiment data. Three cases (210, 223, 233) from the MIT-BIH Arrhythmia Database. And three cases (e0104, e0208, e0613) from European ST-T Database.

4.2 `Judging

After the establishment of tunnel, a random selected ECG waveform that belongs to the individuals can be judged. Table 1 gave the identification rate. In Figure 3 , 210-b, 223-b, 233-b, e0104-b, e0208-b and e613-b are the examples of a dataset belonging to their own tunnels. Contrary to that, 210-a, 223-a, 233-a, e0104-a, e0208-a and e0613-a are the negative examples.

Table 1. Identification rate on MIT/BIH ECG dataset

Item	Description						
MIT/BIH data No.	116	119	223	233	210	105	213
Accurate(%)	97	100	97	95	97	84	96

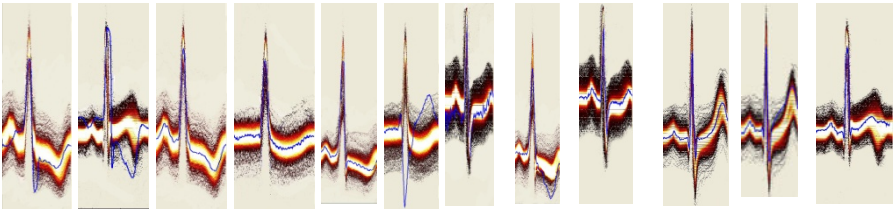


Fig. 3. Tunnel morph examples

5 Discussion

Our target is to establish an ECG based judging model to distinguish the difference of different person and the identity of same person. It is will satisfy the characteristic of

inclusion. Because of over precision on digitalized ECG waveform, one person's ECG gotten at different time may be distinguished as two people. To avoid such misjudge, the similarity measurement of ECG should have some inclusive ability like human.

For comparison of the measurement strategies need some fuzzy decision. The inclusive value I was introduced and calculated by equation (14).

$$I = (S_m - S_t) / S_t \tag{13}$$

S_m is the margin waveform value; S_t is that of waveform on point T. I is a number of positive or negative, when it is negative and smaller, the inclusive effect is better. When I is positive and larger, the inclusive effect of the measurement strategy is worse.

From the observation on the experiment result, high accurate was reached in some ECG data like, 116, 119, 223, 233, 210, 213. Since there are lots of arrhythmia waveform in Data No. 105, that will lead an another problem. That is, single ECG morph model may not satisfied the distinguishing requirement.

6 Further Study

In the future, ECG data of one person will be acquired from his or her 24 hours ECG continuing waveform. It contains the ECG waveform in different status, like standing, sitting, running, climbing ladder, etc… More different ECG morph will be used to identity an individual. Therefore, the clustering of those ECG waveforms, and ECG pattern waveform will be taken into consideration.

Acknowledgment. The paper was supported by Tianjin Natural Science Foundation (10JCYBJC00700) and Tianjin Key Foundation on Science Supporting Plan (10ZCKFSF00800).

References

1. Hoekema, R., Uijen, G.J.H., van Oosterom, A.: Geometrical aspect of the interindividual variability of multilead ECG recordings. IEEE Trans. Biomedical. Engineering 48, 551–559 (2001)
2. Simon, B.P., Eswaran, C.: An EGG classifier designed using modified decision based neural network. Computer and Biomedical Research 30, 257–272 (1997)
3. Biel, L., Patterson, O., Philipson, L., Wide, P.: ECG analysis: a new approach in human identification. In: Proceedings of the 16th IEEE Instrumentation and Measurement Technology Conference, Venice, Italy, pp. 557–561 (1999); and IEEE Transactions on Instrumentation and Measurement, pp.808-812 (2001)
4. Israel, S.A., Irvine, J.M., Cheng, A., Wiederhold, M.D., Wiederhold, B.K.: ECG to Identify Individuals. Pattern Recognition 38(1), 138–142 (2005)

5. Irvine, J.M., Wiederhold, B.K., Gavshon, L.W., Israel, S.A., McGehee, S.B., Meyer, R., Wiederhold, M.D.: Heart Rate Variability: A New Biometric for Human Identification. In: International Conference on Artificial Intelligence (ICAI 2001), Las Vegas, Nevada, vol. III, pp. 1106–1111 (2001)
6. Irvine, J.M., Israel, S.A., Wiederhold, M.D., Wiederhold, B.K.: A New Biometric: Human Identification from Circulatory Function. Joint Statistical Meetings of the American Statistical Association, SanFrancisco, 7 pages (2003)
7. Israel, S.A., Irvine, J.M., Wiederhold, B.K., Wiederhold, M.D.: The Heartbeat: The Living Biometric. In: Boulgouris, N.V., Micheli-Tzanakou, E., Plataniotis, K.N. (eds.) Biometrics: Theory, Methods, and Applications. John Wiley and Sons/IEEE (March 2010)
8. Ogawa, M., et al.: Fully Automated Biosignal Acquisition System for Home Health Monitoring. In: Proc. of the 19th IEEE EMBS Intl. Conf., Chicago, USA (1997)
9. Kyoso, M., Uchiyama, A.: Developrnent of an ECG Identification System. In: Proc. of the 23th IEEE EMBS Intl. Conf., Istanbul, Turekey (2001)
10. Shen, T.W., Tompkinsl, W.J., Hu, Y.H.: One-Lead ECG For Identity Verification. In: Proceedings 2001 2nd Joint Conference of the IEEE Engineering in Medicine and Biology Society and the Biomedical Engineering Society, EMBSBMES, Houston, TX, USA, October 23-26 (2002)
11. Kim, K.-S., Yoon, T.-H., Lee, J.-W., Kim, D.-J., Koo, H.-S.: A Robust Human Identification by Normalized Time-Domain Features of Electrocardiogram. In: Proceedings of the 27th Annual Conference on 2005 IEEE Engineering in Medicine and Biology, Shanghai, China, September 1-4 (2005)
12. Zhang, Z., Wei, D.: A new ECG identificatiom mothed using bayes' theorem
13. Irvine, J.M., Israel, S.A., Scruggs, W.T., Worek, W.J.: eigenPulse: Robust Human Identification from Cardiovascular Function. Pattern Recognition 41, 3427–3435 (2008)
14. Israel, S.A., Scruggs, W.T., Worek, W.J., Irvine, J.M.: Fusing face and ECG for personal identification. In: Proceedings of 32nd Applied Imagery Pattern Recognition Workshop, vol. 1, pp. 226–231 (2003)
15. Wang, Y., Plataniotis, K.N., Hatzinakos, D.: Integrating Analytic And Appearance Attributes For Human Identification From Ecg Signals. In: Proceedings of Biometrics Symposiums (BSYM), Baltimore (September 2006)
16. Gahi, Y., Amrani, M., Zoglat, A., Guennoun, M., Kapralos, B., El-Khatib, K.: Biometric Identification System Based on Electrocardiogram Data. In: 2nd IEEE International Conference on New Technologies, Mobility and Security (NMTS 2008), Tangier, Morocco, November 5-7 (2008)
17. Chan, A. D. C., Hamdy, M. M., Badre, A., Badee, V.: Wavelet Distance Measure for Person Identification Using Electrocardiograms. IEEE Transactions on Instrumentation and Measurement 57(2) (February 2008)
18. Plataniotis, K.N., Hatzinakos, D., Lee, J.K.M.: Ecg Biometric Recognition Without Fiducial Detection. In: Biometrics Symposium (2006)
19. Agrafioti, F., Hatzinakos, D.: Fusion of ECG sources for human identification. In: ISCCSP (2008)
20. Zahra Fatemian, S., Hatzinakos, D.: A New Ecg Feature Extractor For Biometric Recognition. IEEE (2009)
21. Agrafioti, F., Hatzinakos, D.: Signal validation for cardiac biometrics. In: Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing, Dallas, TX, vol. 1, pp. 1734–1737 (2010)

22. Ye, C., Coimbra, M.T., Vijaya Kumar, B.V.K.: Investigation of Human Identification using. Two-Lead Electrocardiogram (ECG) Signals
23. Chiu, C.-C., Chuang, C.-M., Hsu, C.-Y.: A Novel Personal Identity Verification Approach Using a Discrete Wavelet Transform of the ECG Signal. In: Proceedings of International Conference on Multimedia and Ubiquitous Engineering, Busan,Korea, vol. 1, pp. 201–206 (2008)
24. Shi, J., Lam, K.-Y.: VitaCode: Electrocardiogram Representation for Biometric Cryptography in Body Area Networks
25. Li, Z., Yuan, J., Yang, H.: Analysis distances for similarity estimation by Fuzzy C-Mean algorithm. In: The Eighth International Conference on Machine Learning and Cybernetics, Baoding, China, vol. (1), pp. 582–587 (2009)

Non-user-Specific Multivariate Biometric Discretization with Medoid-Based Segmentation

Meng-Hui Lim and Andrew Beng Jin Teoh

School of Electrical and Electronic Engineering,
College of Engineering, Yonsei University,
Seoul, South Korea
menghui.lim@gmail.com, bjteoh@yonsei.ac.kr

Abstract. Univariate discretization approach that transforms continuous attributes into discrete elements/binary string based on discrete/binary feature extraction on a single dimensional basis have been attracting much attention in the biometric community mainly to derive biometric-based cryptographic key derivation for security purpose. However, since components of biometric feature are interdependent, univariate approach may destroy important interactions with such attributes and thus very likely to cause features being discretized suboptimally. In this paper, we introduce a multivariate discretization approach encompassing a medoid-based segmentation with effective segmentation encoding technique. Promising empirical results on two benchmark face datasets significantly justify the superiority of our approach with reference to several non-user-specific univariate biometric discretization schemes.

Keywords: Multivariate Biometric Discretization, Medoid-based Segmentation, Quantization, Encoding.

1 Introduction

Biometric discretization is a process that maps continuous features into a discrete space mainly for biometric key derivation for subsequent cryptographic purpose [4][7][8][9][10]. Apart from this, this transformation process can not only produce a concise summarization of the continuous features for facilitating better understanding towards the data, but also make learning or data analysis faster and more accurate. Furthermore, biometric discretization is able to improve predictive classification accuracy in most cases if important distinctions of features can be preserved. The general block diagram of binary biometric discretization is illustrated in Fig. 1.

Fig. 1. General block diagram of binary biometric discretization

Z. Sun et al. (Eds.): CCBR 2011, LNCS 7098, pp. 279–287, 2011.

Biometric discretization can be decomposed into two essential components:

- **Quantization / Segmentation:** This first component is a domain partitioning process where the feature space is partitioned into a number of non-overlapping segments.
- **Encoding:** The second component is a labeling process of the quantization segments where every segment is labeled with a unique discrete index/binary codeword or a string of indices/codewords. An unknown test object that falls within any segment will be mapped to the label associated to such a segment.

Most of the works in biometric discretization have been focusing on a univariate approach. Equal-width [10] and equal-probable [1][4][9] discretizations are among the two common univariate non-user-specific discretization schemes. Prior to encoding, the former segments every single-dimensional feature space into a number of equal-width quantization intervals; while the latter partitions every single-dimensional feature space into a couple of quantization intervals containing equal mass of the background probability distribution.

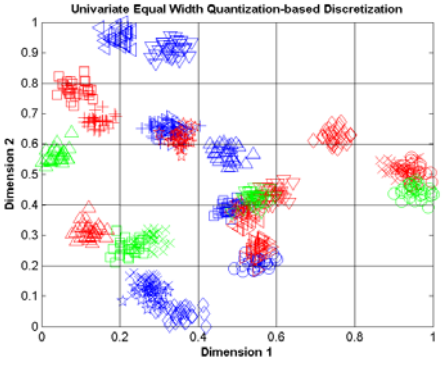
However, when each single-dimensional feature space containing locally sorted feature instances is quantized in isolation of the other dimensions, univariate discretization may destroy important interactions with such attributes, thus potentially causing these discretization techniques to perform sub-optimally when the training features are not distributed uniformly.

On the contrary, multivariate discretization takes into account the interdependence among the attributes in defining high dimensional segments, making it a potential approach with better predictive accuracy. However, it would certainly be more time-consuming and complicated to search for the optimal boundaries when taking the interdependence with other attributes into consideration in the discretization process.

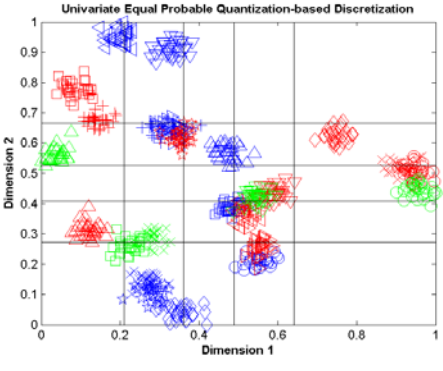
Our work in this paper centers on the proposal of a multivariate discretization technique based upon a representative-based clustering with a distinct discrete encoding technique. In specific, our algorithm basically employs a medoid-based clustering technique to induce voronoi segments in the first phase and adopts an effective discrete segmentation encoding approach to preserve the separation among the segments to a certain extent. Comparing to univariate techniques that derive hyper-rectangular/hyper-cubical segments from a high-dimensional point of view, our algorithm offers great flexibility in the shape of segments formed, inducing convex irregular hyper-polygonal segments that could better adapt to the object distribution within each segment, as illustrated in Fig. 2.

Note that this work is intended to serve as a preliminary step in developing a multivariate binary biometric discretization algorithm. At the present stage of our work, our algorithm is only able to produce a discrete user representation rather than a practical binary representation from each user's biometric features. Therefore, for the time being, we will deal with discrete encoding in this paper.

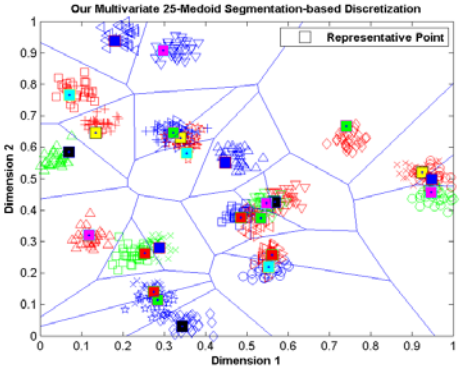
The structure of this paper is organized as follows. In the next section, we detail the representative-based segmentation algorithm and segment encoding algorithm of our proposed approach. In Section 3, we evaluate the equal error rate performance of our scheme and subsequently vindicate the superiority of our approach over the univariate discretization schemes. Finally, concluding remarks are drawn and potential future works are discussed in Section 4.



(a) Square segments by univariate equal-width quantization



(b) Rectangular segments by univariate equal-probable quantization



(c) Irregular polygonal voronoi segments by our medoid-based segmentation

Fig. 2. A comparison of 2-dimensional segments induced by our multivariate discretization scheme with respect to classical univariate techniques

2 Our Proposed Multivariate Discretization Approach

2.1 Representative-Based Segmentation

Our discretization approach adopts a representative-based segmentation algorithm to perform partitioning over the set of D-dimensional data set in producing k segments from a data set of n training samples. This algorithm partitions the data set into groups through making an attempt to optimize a specific objective function. In representative-based segmentation, each segment contains a representative point known as *medoid*, whose average dissimilarity to all the objects in the segment is minimal.

In algorithm 1, we describe the medoid segmentation-based quantization in pseudocode. This algorithm begins by initiating a random segmentation solution SL_{init} (i.e. a set of medoids) and recursively converges to a good local solution SL_{final} through optimizing the following objective function $f(SL)$:

$$f(SL) = \sum_{j=1}^{k} \sum_{i=1}^{|Clus_j|} \sum_{d=1}^{D} \left| x_{ji}^{d} - m_j^{d} \right| \tag{1}$$

where $|Clus_j|$ denotes the cardinality of the j-th segment. Through minimizing (1), the SL_{final} intuitively offers minimal within-segment variation potentially that minimizes the sum of pairwise dissimilarities within each segment defined by the corresponding medoid.

The same process is repeated for R times with a different SL_{init} each time and the best converged solution is returned in order to guarantee a "sufficiently good" solution being achieved.

Algorithm 1: Medoid-based Segmentation

REPEAT *R* TIMES
  Create an initial solution *SL* by randomly selecting a non-overlapping set of *k* medoids (representatives), such that *SL*:= {$\boldsymbol{m}_1$, $\boldsymbol{m}_2$, …, $\boldsymbol{m}_k$}.
  WHILE SL remains unchanged for not more than *L* runs DO
    1. Identify the non-representative points associated with every medoid and compute *f*(*SL*) for the initial objective values.
    2. Pick a random set of *p* medoids from the *m* current medoids.
    3. Determine *n* (non-representative) neighbouring points of each of the *p* medoids.
    4. Replace each of the *p* medoids with a random neighbour candidate. If no remaining candidate of a medoid is found, the replacement of such a medoid is aborted.
    5. Check if any overlapping selection occurs
      i. the selection of a common neighbour or more than one different neighbours with identical components for multiple medoids;
      ii. the selection of a neighbour that has identical feature components as any of the *m*-*p* unselected medoids in Step (2);

If it occurs, discard the overlapping point from the neighbour candidates of the evaluated medoid and repeat Step (4) for the overlapping medoid. Otherwise, with the new solution *SL'*, proceed to Step (6).

6. Identify non-representative points associated with every medoid and compute f(*SL'*).
7. IF (f(*SL'*) < f(*SL*)) THEN *SL*:= SL';
   ELSE terminate and return SL as the solution for this run.

Return the best out of the *R* solutions found.

2.2 Segmentation Encoding

Once the best segmentation solution is obtained, it is essential to decide the proper way to encode this set of medoid-based segments. It is important to understand that medoids are representative points that define voronoi segments and the decision of assigning an object to be a medoid in a segment may be influenced by the position of other medoids. Particularly, in order to construct an optimal segmentation boundary with respect to an objective function, the position of a medoid could simply be adjusted with respect to the adjacent medoids and may not best characterize object distribution within a segment. Therefore, a better point for use of encoding would be the mean of the objects in each segment so that a more precise discrete index can be allocated to each component of the segment with respect to all other segments for a better estimation of the segment position.

In our approach, the object mean of every segment is computed and sorted among the mean of all other individual segments accordingly in each dimension after the best segmentation solution is ascertained from the quantization phase. Every segment eventually ends up with D discrete indices corresponding to the D dimensional space segmentation. Algorithm 2 provides the pseudocode of our segmentation encoding technique.

Algorithm 2: Segmentation Encoding

1. With the best solution $SL_B = \{\boldsymbol{m_1}, \boldsymbol{m_2}, \dots, \boldsymbol{m_k}\}$, compute the mean $\grave{\imath}_j$ of every j-th segment $(j \in \{1,2,\dots,k\})$ by averaging the enclosed training objects:

$$\grave{\imath}_j = \frac{\sum_{i=1}^{|C_j|} \boldsymbol{x_{ji}}}{|C_j|}, \quad \text{for } j = 1,2,\dots,k.$$

2. Sort the mean of the segments descending according to their d-th component value to obtain the sorted indices $v_1^d, v_2^d, \dots, v_k^d$:

$$[v_1^d, v_2^d, \dots, v_k^d] = ascending\_sort(\grave{\imath}_1^d, \grave{\imath}_2^d, \dots, \grave{\imath}_k^d), \qquad \text{for } d = 1,2,\dots,D.$$

3. Assign a discrete index to each mean component, such that

$$\grave{\imath}_{v_j^d}^d \leftarrow \varsigma^d, \qquad \text{for } \varsigma^d \in \{1,2,\dots,k\}; j = 1,2,\dots,k\,; d = 1,2,\dots,D.$$

4. As a result, a D-dimensional object that lies within the j-th segment can be represented by a set of discrete indices:
$$\boldsymbol{x_{ji}} = \{x_{ji}^1, x_{ji}^2, \dots, x_{ji}^D\} \leftarrow \{\varsigma^1, \varsigma^2, \dots, \varsigma^D\}$$

To facilitate the discretization of unknown test objects, the helper data relative to discretization that needs to be stored during the training phase include the medoids and their D-dimensional discrete labels. To decide the segment association during the testing phase, an unknown test object $\boldsymbol{y} = \{y^d | d = 1, \dots, D\}$ is associated to the medoid $\boldsymbol{m_{j^\star}} = \{m_{j^\star}^d | d = 1, \dots, D\}$ of the $j^\star$-th segment that is the most similar, or geometrically closest to the object:

$$j^\star = \min_{j=\{1,2,\dots,k\}} \sum_{d=1}^{D} y^d - m_j^d \tag{2}$$

and $\boldsymbol{y}$ is then tagged with the D-dimensional label of the $j^\star$ segment.

3 Experiments and Discussions

3.1 Experiment Setup

Two popular face data sets are used to evaluate the performance of our discretization scheme with respect to that of a few univariate non-user-specific biometric discretization schemes in this section:

Yale B: The adopted data set is a random subset of the Yale B face data set [2], containing a total of 304 images with 32 images per person for 38 identities. The selected images encompass faces with moderate illumination differences.

FERET: The adopted data set is a random subset of the FERET face dataset [6], containing a total of 2400 images with 12 images per person for 200 identities. The images were collected under a semi-controlled environment with varying illumination conditions and facial expressions.

For all datasets, proper alignment is applied to the images based on standard face landmarks. To avoid possible strong variation in hair style, the face region is extracted for recognition by cropping the images to the size of 42×48 for Yale B; 61×73 for FERET data sets. Finally, histogram equalization is applied to the cropped images.

In these data sets, half of each user's images are used for training while the remaining half is used for testing. To measure the system FAR, each image of each user is matched against each image of every other user accordingly, while for evaluating the system FRR, each image is matched against every other image of the same user for every user. The total number of genuine and imposter matches are shown in Table 1.

Table 1. Experiment Settings

Face Dataset	No. users	Training samples per user	Testing samples per user	Total Genuine Matches	Total Imposter Matches
Yale B	38	16	16	4,560	179,968
FERET	200	6	6	3,000	716,400

In the experiment, discretization performance of our approach is compared to that of several existing approaches. In our algorithm, we set the number of neighbor candidates N to 20, the number of epochs R to 50 and the count of stable solutions L to 50. The dimensions of the raw image were first reduced to 6 by using the Eigenfeature Regularization and Extraction (ERE) [3] prior to discretization.

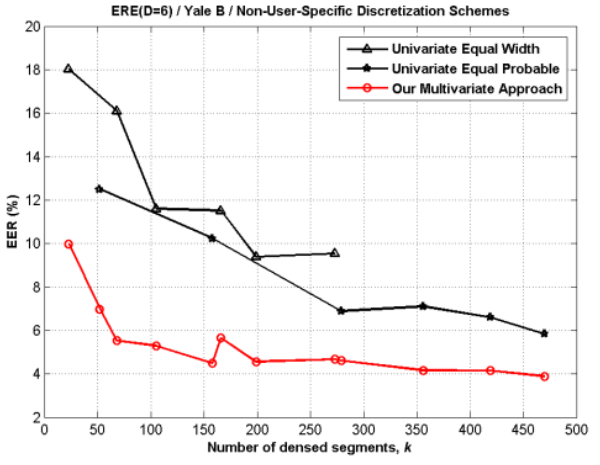
We collect multiple amounts of D-dimensional densed segments based on various equal partitioning settings on every single dimension and generate the same quantities of D-dimensional segments accordingly for performance comparison. The discretization schemes adopted for comparison include:

- Univariate equal-width quantization-based discretization,
- Univariate equal-probable quantization-based discretization,
- Multivariate k-medoid-based discretization (Our approach).

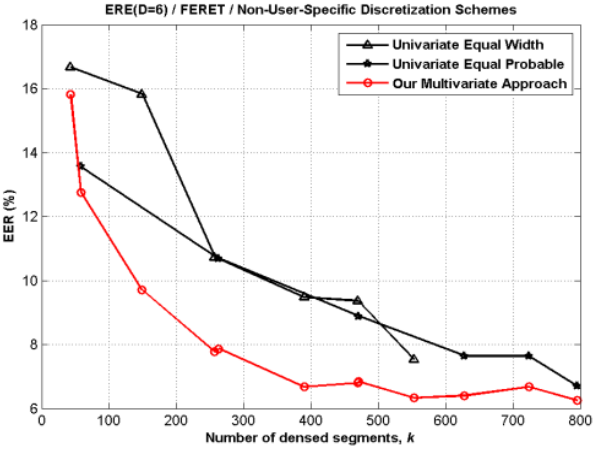
3.2 Performance Evaluation

Fig. 3 illustrates the performance comparison among the non-user-specific discretization schemes based on the ERE-extracted features from (a) Yale B, (b) FERET face data sets. It is observed that our multivariate discretization approach outperforms the univariate discretization schemes significantly. It is worth to note that for all cases where $k < u$, $k = u$ and $k > u$ with k denoting the number of densed segments and u denoting the number of users, the improvement can consistently be seen. It is worth to note that one limitation of adopting representative-based segmentation technique is that only convex segments can be induced and thus might not be able to accommodate training samples with non-convex intra-class distribution. However, by setting $k > u$, segments with non-convex shape can be approximated by a union of several neighboring segments with identical class majority to overcome the problem of non-convex intra-class samples distribution.

Consistent improvement in both evaluations basically justifies that, regardless of whether every segment is induced to represent objects of a class, our approach remains effective in extracting meaningful segmentation boundaries through effectively exploiting the interdependency among the feature attributes with respect to the univariate approaches.



(a)



(b)

Fig. 3. EER performance of our multivariate discretization scheme with reference to the common univariate discretization schemes based on (a) Yale B and (b) FERET face datasets

4 Conclusion and Future Work

In this paper, we have proposed a multivariate discretization scheme that bases upon a medoid-based segmentation with effective segmentation encoding technique. In particular, the segmentation algorithm derives segments to which unknown test objects may be associated while the encoding algorithm derives appropriate discrete indices to which the objects within a segment has to be mapped based on the mean of the training objects in each segment. Experimental results on both Yale B and FERET face datasets yield promising improvements with reference to the univariate discretization schemes.

To extract practical binary biometric representation using a multivariate approach, adopting Linearly Separable SubCode Encoding [5] is necessary for effective binary classification. A few possible future works in this direction would be to address the following two challenges: (1) efficiently representing huge amount of discrete segmentation indices in binary without incurring significant degradation in the discretization performance of the actual settings; and (2) creating segments beyond the number of training samples in order to augment the uncertainty of the binary output for security reason.

Acknowledgments. This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MEST) (No. 2011-8-1095).

References

1. Chen, C., Veldhuis, R., Kevenaar, T., Akkermans, A.: Multi-bits Biometric String Generation based on the Likelihood Ratio. In: IEEE International Conference on Multimedia and Expo (ICME 2004), vol. 3, pp. 2203–2206 (2004)
2. Georghiades, P.N., Belhumeur, A.S., Kriegman, D.J.: From Few to Many Illumination Cone Models for Face Recognition Under Variable Lighting and Pose. IEEE Trans. Pattern Anal. Mach. Intelligence 23(6), 643–660 (2001)
3. Jiang, X.D., Mandal, B., Kot, A.: Eigenfeature Regularization and Extraction in Face Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(3), 383–394 (2008)
4. Kevenaar, T.A.M., Schrijen, G.J., van der Veen, M., Akkermans, A.H.M., Zuo, F.: Face recognition with renewable and privacy preserving binary templates. In: IEEE Workshop on Automatic Identification Advanced Technologies (AutoID 2005), NY, USA, pp. 21–26 (2005)
5. Lim, M.-H., Teoh, A.B.J.: Linearly separable subcode: A Novel Output Label with High Separability for Biometric Discretization. In: 5th IEEE Conference on Industrial Electronics and Applications (ICIEA 2010), pp. 290–294 (2010)
6. Philips, P.J., Moon, H., Rauss, P.J., Rizvi, S.: The FERET Evaluation Methodology for Face Recognition Algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(10), 1090–1104 (2000)
7. Sheng, W.G., Howells, W.G.J., Fairhurst, M.C., Deravi, F.: Template-free Biometric-key Generation by means of Fuzzy Genetic Clustering. IEEE Transactions on Information Forensics and Security 3(2), 183–191 (2008)
8. Teoh, A.B.J., Ngo, D.C.L., Goh, A.: Personalised Cryptographic Key Generation based on FaceHashing. Computers and Security 23(7), 606–614 (2004)
9. Tuyls, P., Akkermans, A.H.M., Kevenaar, T.A.M., Schrijen, G.-J., Bazen, A.M., Veldhuis, R.N.J.: Practical Biometric Authentication with Template Protection. In: Kanade, T., Jain, A., Ratha, N.K. (eds.) AVBPA 2005. LNCS, vol. 3546, pp. 436–446. Springer, Heidelberg (2005)
10. Yip, W.K., Goh, A., Ngo, D.C.L., Teoh, A.B.J.: Generation of Replaceable Cryptographic Keys from Dynamic Handwritten Signatures. In: Zhang, D., Jain, A.K. (eds.) ICB 2005. LNCS, vol. 3832, pp. 509–515. Springer, Heidelberg (2005)

Author Index