2.5D Face Recognition Using Gabor Discrete Cosine Transform

Ali Cheraghian, Farshid Hajati, Soheila Gheisari, Yongsheng Gao

Abstract—In this paper, we present a novel 2.5D face recognition method based on Gabor Discrete Cosine Transform (GDCT). In the proposed method, the Gabor filter is applied to extract feature vectors from the texture and the depth information. Then, Discrete Cosine Transform (DCT) is used for dimensionality and redundancy reduction to improve computational efficiency. The system is combined texture and depth information in the decision level, which presents higher performance compared to methods, which use texture and depth information, separately. The proposed algorithm is examined on publically available Bosphorus database including models with pose variation. The experimental results show that the proposed method has a higher performance compared to the benchmark.

Keywords—Gabor filter, discrete cosine transform, 2.5D face recognition, pose.

I. INTRODUCTION

FACE recognition has been an active area of research for many decades. It still remains a challenging problem due to the complexity and variability of human face images. One of the challenges in the face recognition is pose variation. Algorithms presented for pose variation are divided in two main categories depends on the type of gallery images [1]. The first category consists of multi-view face recognition systems, which require several poses for each subject in the gallery. The second category identifies probe faces, which have different poses compared to the gallery. In real applications, a frontal face is in the gallery while probe faces have unpredicted pose variations. Therefore, the system should be robust to these pose variations.

Local Binary Pattern (LBP) [2] is applied for feature extraction. Compared to holistic approaches, LBP is more robust to pose changes because it does not require exact locations of patterns but relies just on the histogram of the pattern in a region. Recently, a high-order local pattern descriptor, local derivative pattern (LDP) [3], is proposed for robust face recognition. LDP is a micro-pattern representation, which can also be modeled by histogram. Liu and Chen [4] proposed a Probabilistic Geometry Assisted (PGA) face recognition algorithm to handle pose variations. In their

F. Hajati and Y. Gao are with School of Engineering, Griffith University, Nathan Campus, 170 Kessels Road, Brisbane, QLD 4111, Australia (e-mail: f.hajati@griffith.edu.au, yongsheng.gao@griffith.edu.au).
S. Gheisari is with the School of Software, University of Technology Sydney, Sydney, NSW 2007, Australia (e-mail: soheila.gheisari@student.uts.edu.au).
F. Hajati and A. Cheraghian are with Electrical Engineering Department, Tafresh University, Tafresh 3951879611, Iran (e-mail: hajati@tafreshu.ac.ir, a.cheraghian@tafreshu.ac.ir).
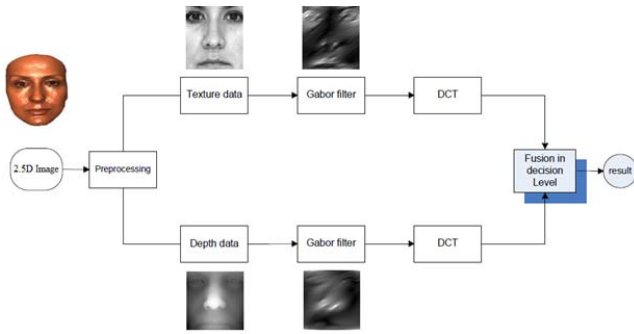
algorithm, human heads are approximated as an ellipsoid whose radiuses, locations, and orientations are estimated based on a universal mosaic model.

Most face recognition researches are focused on the 2D intensity images. As shown in Face Recognition Vendor Test 2002 [5], one of the problems of these algorithms is the pose variation. Zhang et al. [6] proposed a reconstruction of the personalized 3D face shape by using multi-level quadratic variation minimization. From 3D feature points specified manually on the frontal and side views of an input face, the 3D face model was reconstructed by minimizing a quadratic cost function of surface variations, which ensures a second order smoothness. Recently, face 3D. Xua et al. [8] used Gabor wavelet for feature recognition algorithms, which combine 2D and 3D data, are proposed. Chang et al. [7] showed that combining 2D and 3D results by using a simple weighting scheme outperforms either 2D or extraction, proposes a novel system by combining the depth and intensity information. A recognition rate of 91% for face images with pose variation was achieved.

In this paper, we present a novel 2.5D face recognition system using Gabor Discrete Cosine Transform (GDCT) algorithm. In this algorithm, Gabor wavelets are used for feature extraction. Since the extracted data has a high dimensionality, Discrete Cosine Transform (DCT) is used. In the next step, results of the texture and depth information are combined in the decision level to get a higher accuracy. The proposed method is evaluated using a wide range of experiments conducted with publically available Bosphorus database. The encouraging experimental results show the proposed method has a higher performance compared to the benchmark.

This paper is organized as follows. Section II presents the proposed Gabor Discrete Cosine Transform (GDCT). Section III discusses the experimental results and Section IV concludes the paper.

II. PROPOSED GABOR DISCRETE COSINE TRANSFORM (GDCT)

The overall scheme of the proposed method is shown in Fig. 1. The proposed method consists of four steps: Preprocessing, Gabor filter, Discrete Cosine Transform, and Classifier. The detail of each step will be described in the following:

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:10, No:2, 2016

Fig. 1 Overall scheme of the proposed method

A. Preprocessing

Since acquired 3D faces are noisy, they are not in correct coordinates and choosing the Region of Interest (ROI) in the face is essential. Therefore, a preprocessing is needed before the feature extraction. The preprocessing includes three steps. In the first step, faces are aligned to fixed coordinates. In order to align faces to the fixed coordinates the ICP algorithm is applied [9]. In this algorithm, a frontal face with neutral expression is selected as a fixed model and all 3D faces are rotated and translated to have a minimum distance with the model. In the next step, a ROI is selected for each face. Hairs, shoulders, and neck are not useful for face recognition. Therefore, to select a suitable ROI a 120×120 rectangular region is used. In the third step, in order to remove spike noise, a median filter is used. A 3×3 window was utilized to remove spike noise. Using larger windows may cause loss of information. These three steps provide clean data to extract a depth and texture data. The flow diagram of these steps is presented in Fig. 2. Depth and texture images are obtained from registered 3D data. Each 3D point has both 3D spatial coordinates and RGB color information.
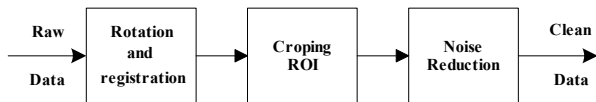

Fig. 2 A flow diagram of the preprocessing procedure

B. Gabor Filter

2D Gabor wavelet is used for feature extraction from depth and texture data. Gabor wavelets were defined by [10] and were extended to two-dimensional space by [11]. A Gabor wavelet is complex sinusoids, which are modulated by a 2D Gaussian function and represents the properties of spatial localization, orientation selectivity, spatial frequency selectivity, and quadrature phase relationships [12].

The Gabor representation of an image, called the Gabor image, is the convolution of the image with the Gabor kernels defined as

$$\psi(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} \exp\left(-\frac{\|k_{u,v}\|^2 \|z\|^2}{2\sigma^2}\right)\left[\exp(jk_{u,v}z) - \exp\left(-\frac{\sigma^2}{2}\right)\right] \quad (1)$$

where $z = \begin{pmatrix} x \\ y \end{pmatrix}$, $k_{u,v} = \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_v \cos\phi_u \\ k_v \sin\phi_u \end{pmatrix}$, $k_v = \frac{\pi}{2^{(v+2)/2}}$, $\phi_u = u(\pi/8)$, $v = 0, \ldots, v_{max} - 1$, $u = 0, \ldots, u_{max} - 1$, v is the frequency, u is the orientation, $v_{max} = 5$, $u_{max} = 8$ and $\sigma = 2\pi$.

In this paper, five different scales and eight different orientations are used. Each image has two Gabor parts: amplitude and phase. Here, we just use the amplitude part for face recognition. For each pixel of the image, we have totally 40 Gabor magnitude coefficients, so we can obtain 40 Gabor magnitude and 40 Gabor face images from a single input face image.

C. Discrete Cosine Transform (DCT)

In this study, Discrete Cosine Transform (DCT) is applied to reduce the dimensionality and redundancy of the Gabor images for face recognition. DCT transforms images from the spatial domain to the frequency domain, where an image is decomposed into a combination of various and uncorrelated frequency components. DCT is able to extract the features in the frequency domain to encode different facial details that are not directly accessible in the spatial domain. Due to the specific properties, DCT has been successfully applied in face recognition systems [13].

The two-dimensional DCT of an image A with size M-by-N is defined as,

$$B_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos\frac{\pi(2m+1)p}{2M} \cos\frac{\pi(2n+1)q}{2N}$$

$$\alpha_p = \begin{cases} \frac{1}{\sqrt{M}}, p = 0 \\ \sqrt{\frac{2}{M}}, 1 \le p \le M-1 \end{cases} \quad \alpha_q = \begin{cases} \frac{1}{\sqrt{N}}, q = 0 \\ \sqrt{\frac{2}{N}}, 1 \le q \le N-1 \end{cases} \quad (2)$$

To obtain the feature vector representing a face, its DCT is computed, and only a subset of the obtained coefficients is retained. Different DCT coefficients are corresponding to different spatial information. The upper-left subset in the DCT domain encodes most of the energy in a face image. Based on this assumption, a sub-image with size 20×20 is selected starting from the upper-left region.

D. Classifier

In this paper, k-Nearest Neighbor method [14] is used for classification. In this method, we classify x by assigning it to the class label ĉ (most frequently represented among the k-nearest neighbors) which has the minimum distance among all the classes,

$$d(x, x_{ci}) = \min_{1 \le c \le C, 1 \le i \le n_c} d(x, x_{ci}) \quad (3)$$

where, $d(x, x_{ci}) = \|x - x_{ci}\|$ is the Euclidean norm of the vector.

Texture and depth data are combined in the decision level. Combination of the 2D and 3D systems has been performed by the weight factor w as in (4).

$$D = \sqrt[2]{d_T^2 + w \times d_D^2} \quad (4)$$

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:10, No:2, 2016

where d_T is the distance from the gallery set of the texture image and d_D is the distance from the gallery set of the depth image. The weight factor w is determined in a training process.

III. EXPERIMENTAL RESULTS

In this section, we evaluate the robustness of the proposed method under different face rotations. In this experiment, we used the publically available Bosphorus database [15], which contains 2.5D face images in different rotations. This database has 105 subjects. Facial data are acquired using a structured-light based 3D system. Acquisitions are single view, and subjects were made to sit at a distance of about 1.5 meters away from the 3D digitizer. The sensor resolution in x, y, and z (depth) dimensions are 0.3mm, 0.3mm, and 0.4mm respectively, and color texture images are high resolution (1600x1200 pixels). We use faces with 10º, 20º, and 30º pose rotation in this study. Fig. 3 illustrates an arbitrary subject from the database in different pose rotations.

At first, we should find the value of the weight factor in the proposed algorithm. 105 subjects with 10, 20 and 30 degrees pose rotations were selected randomly, and the rank-1 recognition rate of the system was obtained for various values of weight factor and considering the w corresponding to the maximum recognition rate. For those values of w which are close to zero and for very large amounts of w, the performance of the proposed method approaches to the performance of the system which is based just on the texture data and depth data, respectively. For some specific values of w, the performance of the proposed method becomes the maximum.

Fig. 3 Some examples of Bosphorus database: (a) Frontal, (b) Right-turn 10º (respect to Y axis), (c) Right-turn 20º (respect to Y axis), (d) Right-turn 30º (respect to Y axis)
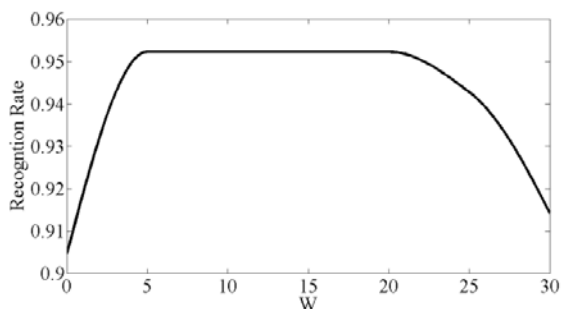

Fig. 4 Recognition rate versus w

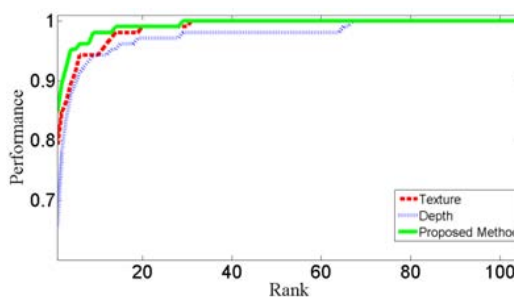In this study, for those values of w between 5 and 20, the

optimum state of the system efficiency is achieved. For these values of w, the system performance reaches 95%. The results of this experiment are shown in Fig. 4. As can be seen, the recognition rate of the system for w=5,…, 20 is the maximum. Therefore, we set w=12.5 in the following experiments.
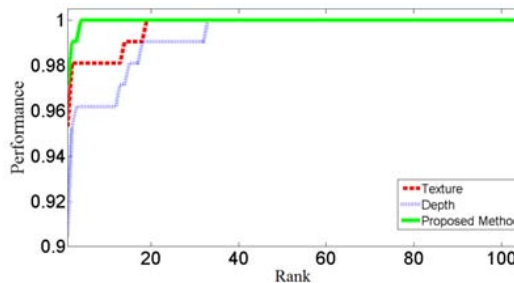
(a)
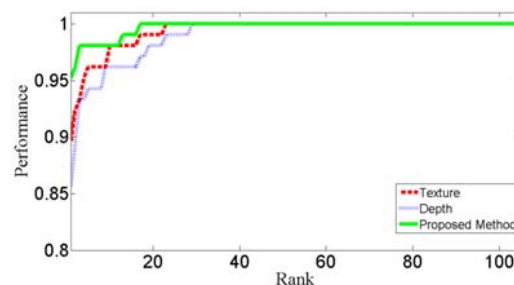
(b)

(c)

Fig. 5 Cumulative Match Characteristics for (a) the right-rotated in 10º, (b) the right-rotated in 20º, (c) The right-rotated in 30º

Since our goal is to recognize faces with pose variation from one sample per person, one frontal face with neutral expression of each person is utilized for the gallery and other faces are used as the probe. In this experiment, system is examined by faces, which have pose rotation in right directions (10º, 20º and 30º). Comparison of the recognition rates for the proposed method with 2D and 3D systems in term of Commutative Match Characteristics (CMC) is depicted in Fig. 5. As can be seen, the recognition accuracy of the proposed method is considerably improved when the depth and the texture information are combined. For 10º pose rotation, the performance of the proposed system (97%) is improved 2% and 8% compared to the texture and the depth data, respectively. For 20º pose rotation, the performance

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:10, No:2, 2016

(95%) improved 5% and 10% compared to the texture and the depth data, respectively. Finally, for the 30º pose rotation the system accuracy (85%) gained 5% and 15% improvement compared to the texture and the depth data, respectively. As a result, the texture data has better performance compared to the depth data and is more robust to pose variations. After the fusion of the depth and the texture data in the decision level, the proposed method has a better performance.

The proposed method is also compared with multimodal PCA as a benchmark. In the multimodal PCA algorithm [7], the PCA is used for feature extraction and combination of 2D and 3D information. The results of the comparison are tabulated in Table I. It is clear that the proposed method has a better performance. For the faces with 10º pose rotation, the performance of the proposed method improved 5.6% compared to the benchmark. In addition, for faces with 20º and 30º pose rotation, around 5.5% and 16.5% improvement is observed. It is obvious that for the faces with large pose rotation, the efficiency of the proposed method is much higher than the performance of the b.

TABLE I
COMPARISON OF THE PROPOSED METHOD WITH MULTIMODAL PCA

Considered Rotation	Multimodal PCA	Proposed Method
10º right	91.4%	97%
20º right	89.5%	95%
30º right	68.5%	85%
Average	**83.1%**	**92.3%**

IV. CONCLUSION

In this paper, a 2.5D face recognition method based on Gabor Discrete Cosine Transform (GDCT) feature was presented. This method combined 2D and 3D information in the decision level, which was shown a higher performance than systems use separate information. A sufficient investigation was considered on the pose variation problem in the face recognition. The results of the experiments demonstrated significant accuracy improvement compared to the benchmark. According to the obtained experimental results, the proposed method achieved the overall recognition rate of 92.3% on faces with pose variation compared to 83.1% for the benchmark.

REFERENCES

[1] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face Recognition: A Survey," Pattern Recognition Letters, Vol. 28, no. 14, pp. 1885–1906, 2007.

[2] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971–987, Aug. 2002.

[3] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local Derivative Pattern Versus Local Binary Pattern: Face Recognition with High-Order Local Pattern Descriptor," IEEE Trans. Image Processing, vol. 19, no. 2, pp. 533-544, Feb. 2010.

[4] X. Liu and T. Chen, "Pose-Robust Face Recognition Using Geometry Assisted Probabilistic Modeling," in Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, 2005, pp. 502–509.

[5] P. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, and M. Bone, "Face Recognition Vendor Test 2002" , in Proc. AMFG, 2003.

[6] X. Zhang, Y. Gao, and M. Leung, "Recognizing Rotated Faces from Frontal and Side Views: An Approach towards Effective Use of Mugshot Databases," IEEE Trans. Information Forensics and Security, vol. 3, no. 4, pp. 684–697, 2008.

[7] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An Evaluation of Multimodal 2D+3D Face Biometrics," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 4, pp. 619-624, Apr. 2005.

[8] C. Xua, S. Lia, T. Tana, and L. Quan., "Automatic 3D Face Recognition from Depth and Intensity Gabor features," Pattern Recognition, vol. 42, no. 9, pp. 1895-1905, 2009.

[9] P. J. Besl and N. D. McKay, "A Method for Registration of 3D Shapes," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pp. 239–256, 1992.

[10] D. Gabor, "Theory of Communication," J. Inst. Elect. Eng., vol. 93, pp. 429–457, 1946.

[11] J. Daugman, "Uncertainty relation for Resolution in Space, Spatial Frequency and Orientation Optimized by Two-Dimensional Visual Cortical Filters," Journal of the Optical Society of America, vol. 2, no. 7, pp. 1160-1169, 1985.

[12] J. Jones and L.Palmer, "An Evaluation of the Two-Dimensional Gabor Wavelet Model of Simple Receptive Fields in Cat Striate Cortex," Journal of Neurophysiology, vol. 58, no. 6, pp. 1233–1258, Dec. 1987.

[13] Z. M. Hafed and M. D. Levine, "Face Recognition Using the Discrete Cosine Transform," Int. J. Comput. Vis., vol. 43, no. 3, pp. 167–188, 2001.

[14] T. M. Cover and P. E. Hart,"Nearest Neighbor Pattern Classification," IEEE Trans. Information Theory, IT-13, vol. 1, pp. 21–27, 1967.

[15] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus Database for 3D Face Analysis," The First COST 2101 Workshop on Biometrics and Identity Management (BIOID 2008), Roskilde University, Denmark, 7-9 May 2008.

A. Cheraghian received the B.Sc. degree in electronic engineering from Shiraz University of Technology, Iran, in 2008 and M.Sc. degree in electronic engineering from Amirkabir University of Technology, Iran, in 2011. He is currently with the Machine Vision and Image Processing (MVIP) lab at Tafresh University, Iran. His research interests include face recognition, biometrics, pattern recognition, computer vision, and multispectral image analysis.

F. Hajati received the B.Sc. degree in electronic engineering from K. N. Toosi University of Technology, Iran, in 2003 and M.Sc. and Ph.D. degrees in electronic engineering from Amirkabir University of Technology, Iran, in 2006 and 2011, respectively. From 2009 to 2010 he was a visiting scholar with the School of Engineering, Griffith University, Australia. From 2011, he has been the head of the electronics group and the Machine Vision and Image Processing (MVIP) lab at Tafresh University, Iran. In 2012, he was a visiting fellow with the School of Engineering, Griffith University, Australia. Also, he has been an adjunct researcher at the Institute for Integrated and Intelligent Systems (IIIS), Griffith University, Australia. His research interests include digital image processing, pattern recognition, computer vision, neural networks, and face detection and recognition.

S. Gheisari received the B.Sc. degree in control engineering from K. N. Toosi University of Technology, Iran, in 2004 and M.Sc. degree in electronic engineering from Central Tehran Branch, Azad University, Iran, in 2013. From 2013, she has been with the Machine Vision and Image Processing (MVIP) lab at Tafresh University, Iran. Her research interests include 3D image processing, pattern recognition, and computer vision.

Y. Gao received the B.Sc. and M.Sc. degrees in electronic engineering from Zhejiang University, Hangzhou, China, in 1985 and 1988, respectively, and the Ph.D. degree in computer engineering from Nanyang Technological University, Singapore. Currently, he is a Professor with the School of Engineering, Griffith University, Australia. His current research interests include face recognition, biometrics, biosecurity, image retrieval, computer vision, pattern recognition, environ-mental informatics, and medical imaging.