2D&3D-ComFusFace: 2D and 3D Face Recognition by Scalable Fusion of Common Features

Juan Zhou,Yongping Li and Jingyan Wang

†Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai, 201800 China
†† Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai, 201800 China
†††Mathematical and Computer Sciences and Engineering, King Abdullah University of Science and Technology, Jeddah, 21534 Saudi Arabia

Summary
In traditional 2D and 3D face recognition systems, different features are extracted from 2D and 3D face images, and then are fused to improve the recognition performance. The shortage of these methods is that they neglect the intrinsic complementary features between 2D and 3D data. In this paper, we investigate the possibility of extracting and scalable fusing common features from 2D intensity and 3D depth face images, and develop a novel 2D and 3D face recognition method-- 2D&3D-ComFusFace, which represent and fuse some common global and local features of 2D and 3D data. A novel pose normalization method for 3D range data is also proposed before transiting them to be depth image. After preprocessing, two global features--2D Principle component Analysis (2DPCA), 2D Fisher Linear Discriminate Analysis (2DFLD), and a local feature--Local Binary Pattern (LBP) are extracted from both 2D intensity image and 3D depth image. Then the matching scores are computed and fused by weighted sum rule to get a further improved performance. The experiments are carried out on CASIA3D database, and significant improvements of both recognition rate and EER are achieved

Key words:
2D intensity images, 3D depth images, global feature, local feature, Fusion

1. Introduction

The recognition algorithms based on 2D face images are usually sensitive to facial variations and uncontrolled environment, Low recognition rate may occur in some conditions [1], While the 3D face contains more spatial information, which is inherent property of an object and robust to the uncontrollable environment. We believe that integrating 2D and 3D data is a promising approach to improve the face recognition performance. Published results on multimodal 2D and 3D face recognition have shown that the recognition of faces from 2D and 3D facial data results in a better performance when compared with methods using solely 2D or 3D data [2]. Published works in multi-modal 2D and 3D face recognition have rapidly increased in recent years. Surveys on approaches in 3D and multimodal 2D and 3D face recognition can be found in the contributions of [3]. Here we will list a few major related works:

(i) Reference [4] estimated the 3D shape from the 2D face using 3Dmorphable model (3DMM), and the virtual faces of different views were generated from the 3DMM to assist face recognition.
(ii) Reference [5] used feature detection and registration with the ICP algorithm in the 3D domain and LDA in the 2D domain for multi-modal face recognition.
(iii) Reference [6] presented a method by means of fusing color, local spatial and global frequency information and specifically fusing the multiple features derived from a hybrid colors pace, the Gabor image representation, the local binary patterns, and the discrete cosine transform of the input image.
(iv) Reference [1] proposed a method based on PCA to combine facial cues from the 2D and 3D images, and utilized the 2D and 3D facial information at the enrollment, image and score levels.
(v) Reference [7] used PCA on both 2D intensity images and 3D depth images, and fused 2D and 3D results to obtain the final performance.

All the above works have made great contributions to multi-modal face recognition. In this paper, we investigate the usage of common global and local features for both 2D and 3D face data. The main thrust of our work here is to find the best combination of 2D intensity image and 3D depth image when using 2DPCA, 2DFLD and LBP to get an improved recognition rate. To maximize the benefit of using both 2D and 3D facial data, we focused on the preprocessing, feature extraction and score fusion steps for both of them.

Manuscript received March 5, 2012
Manuscript revised March 20, 2012

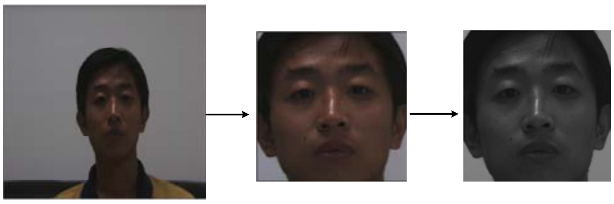
2. Preprocessing and Normalization

2.1 Preprocessing of RGB Image

For RGB image, we use Adaboost algorithm to extract the face region. This process is done assisted with the open source code provided by open CV yahoo group [8]. Then we transform RGB images of the face region to YC_bC_r color space, which is defined as follows:

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.4810 & 128.5530 & 24.9660 \\ -37.7745 & -74.1592 & 111.9337 \\ 111.9581 & -93.7509 & -18.2072 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

In which Y component will be used as the intensity images. Also we normalize all the images to be the same size of 200×200 .This process is shown in Fig.1. (a) is the original image, (b) is the face region extracted by Adaboost algorithm, and (c) is the Y component of the face region.



(a) (b) (c)

Fig. 1 Preprocessing of RGB image

2.2 Normalization of Range Data

Assume that the given range data represents a human face, therefore the knowledge involving the face and facial features can be exploited. For each shot, the prominence of nose can be localized easily and robustly. So we can extract the face region and normalize them to be a nearly frontal view by:

(i) Find out the nose tip and set a certain semi-diameter R , chose those points from which to the nose tip is not far than R as the region of interest (we set $R = 900mm$).

(ii)Transform points set (region of interest) S_w as below:

$$O_m = \frac{1}{N_m}\sum_{i=1}^{N_m} p_i \quad (2)$$

$$C_{cov} = \frac{1}{N_m}\sum_{i=1}^{N_m}(p_i - O_m)(p_i - O_m)^T \quad (3)$$

$$A = \begin{pmatrix} v_2^x & v_2^y & v_2^z \\ v_1^x & v_1^y & v_1^z \\ v_3^x & v_3^y & v_3^z \end{pmatrix} \quad (4)$$

$$p_i^{'} \leftarrow A \times (p_i - O_m) \quad (5)$$

In which, O_m is the center of S_m , $p_i \in S_m$, N_m is the number of S_m , C_{cov} is the covariance matrix of S_m , Assume that C_{cov} has three eigenvalue $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and corresponding eigenvector v_1, v_2, v_3 , $p_i^{'}$ is the normalized point.

(iii)For CASIA3D database, the nose tip of every face usually has a maximal value in z axis. After pose normalization of step (ii), the range data has a nearly frontal view, then we transit them to be depth image by interpolating at the integral x and y coordinates and storing the corresponding z coordinates in the depth image matrix using x as a horizontal index and y as a vertical index. The whole processing is shown in Fig.2. (a) is the original data, of which the face region specified and normalized to be nearly frontal face is show in (b), and (c) is the x and y index by interpolating at the integral x and y coordinates, (d) is the mesh graph of the face region, and finally the depth image is shown in (e).

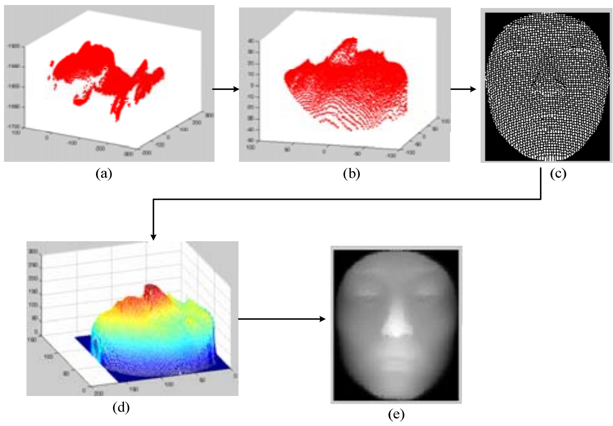


Fig. 2 Normalization of range data:

3. 2D&3D-ComFusFace Method for 2D+3D Multi-modal Face Recognition

In this section, we introduce the novel 2D and 3D face recognition method: 2D&3D-ComFusFace. We first describe the two global and one local features used by 2D&3D-ComFusFace, then the score fusion scheme is proposed.

3.1 Global Feature Extraction

Global feature of 2DPCA. Principal Component Analysis (PCA) is the subspace method in human face recognition and used to get the eigenvalue of facial images. 2DPCA was developed for image feature extraction based on 2D matrices as opposed to the standard PCA, which is based on 1D vector. It was proposed by [9] to cut the computational cost of the standard PCA. Unlike PCA that treats images as vectors, 2DPCA views an image as a matrix and has been proved having a higher recognition rate than that of PCA and it can be described as follows:

Let $A_j(j=1,2,\cdots,M)$ be the j_{th} face in the train database, and each face is of size $m\times n$, M is the total number of train images, and $\bar{A}$ is the average image of the database, thus the covariance matrix C can be written as

$$C=\frac{1}{M}\sum_{j=1}^{M}(A_j-\bar{A})^T(A_j-\bar{A}) \tag{6}$$

The criterion is to maximize the equation of

$$J(x)=x^TCx \tag{7}$$

If x_{opt} can maximize Eq.(6), it can be seem as the optimal projection vector. Of course, one can compute k optimal projection vectors, which are the k leading eigenvetors of C, that is

$$\begin{cases}\{x_1,x_2,\cdots,x_k\}=\arg\max(J(x))\\ x_i^Tx_j=0, i\neq j, j=1,2,\cdots,d\end{cases} \tag{8}$$

$$\begin{cases}B_{ji}=A_j^{'}x_i(i=1,2,\cdots,k)\\ B_j=[B_{j1},B_{j2},\cdots,B_{jk}]\end{cases} \tag{9}$$

In Eq.(8) $A_j^{'}=(A_j-\bar{A})$, and B_j is the projection image of A_j in the feature space. Let P be an image to be identified, and Q is the projection image of P in the feature space, then the distance from the model B_j to Q will be calculated:

$$d(Q,B_j)=\sum_{i=1}^{k}\left\|Q_j-B_{ji}\right\|_2 \tag{10}$$

If $d(Q,B_j)=\min_j[d(Q,B_j)]$, image P belongs to A_j.

Global feature of 2DFLD. 2DFLD is somewhat similar to 2DPCA, the difference is the definition of criterion function $J(x)$, which make use of Fisher Linear Criterion to maximize the ratio of between scatter matrix and within scatter matrix:

$$J(x)=\frac{x^TS_Bx}{x^TS_wx} \tag{11}$$

In which $$S_B=\frac{1}{L}\sum_{i=1}^{L}N_i(A_i-\bar{A})^T(A_i-\bar{A}) \tag{12}$$

$$S_w=\frac{1}{P_i}\sum_{k=1}^{P_i}(A_k-\bar{A}_i)^T(A_k-\bar{A}_i) \tag{13}$$

Since 2DFLD feature extraction method is also based on projection image comparison in the feature space, and the processing is similar to 2DPCA from Eq.(8) to Eq.(10), here we will no longer repeat describing it.

3.2 Local Feature Extraction

LBP was initially used to depict the texture image. Reference [10] introduced it to face recognition and used it to describe face feature, and the basic LBP operator can be described as in Fig.3(a), in which $P_i(i=1,2,\cdots,8)$ will be compared with the central pixel value P_o, and be valued 0 or 1 by the threshold:

$$b_i=\begin{cases}1, P_i\geq P_o\\ 0, P_i<P_o\end{cases} i=1,2,\cdots,8 \tag{14}$$

Then $b_i(i=1,2,\cdots,8)$ will be arranged clockwise, and a binary number of 8-bit is calculated and translated into a decimal number which will be the tag value of the central pixel. In this paper we use a round neighborhood as is shown in Fig.3 (b) to calculate LBP image.

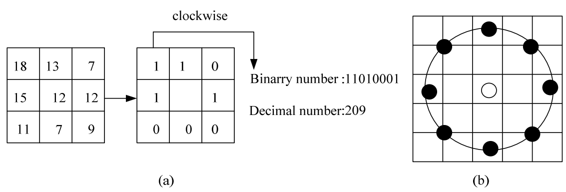


Fig. 3 LBP operator

After preprocessing and normalization step, we get Y intensity image and 3D depth images, then we will use LBP to encode both of them and subdivide them to three parts according to face characteristics. After calculating histograms of each part, they will be cascaded to be one. The sketch map is shown in Fig.4. (a) is intensity image and depth image tagged by LBP operator, after subdivided into three parts are shown in (b), their histograms of three parts being cascaded are shown in (c).

The similarity of histogram can not be measured by Euro distance or cosine similarity. We use histogram intersection (HI) as the matching criterion for two different histogram sequences. It can be defined as:

$$S_{HI} = \sum_{b=1}^{N} \min(H_1(b), H_2(b)) \tag{15}$$

In which, H_1 and H_2 are two different histograms, b is frequency bar, and N is the total number of b, in this paper $N = 256 \times 3 = 768$.

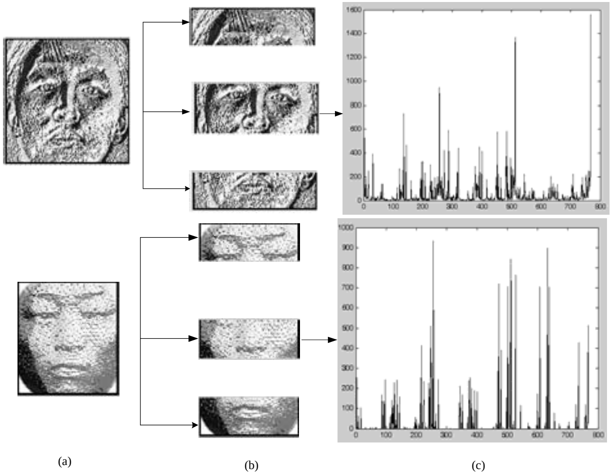


Fig. 4 Histograms of intensity image and depth image

3.3 Score Fusion

Similarity scores from different classifiers may differ in value and dimensional unit. When fusing them together, they need to be normalized into a same region [11]. In this paper, we normalized both recognition scores of intensity images and depth images to be in region $[0,1]$ by $\min-\max$ rule:

$$S_{normal} = \frac{S - \min(S)}{\max(S) - \min(S)} \tag{16}$$

After score normalization by Eq.(16), two scores from different classifiers will be fused together by weighted sum rule:

$$S_{fusion} = \omega * S_{normal1} + (1-\omega) S_{normal2} \tag{17}$$

In which, $\omega = [0, 0.1, 0.2, \cdots, 1]$, $S_{normal1}$ and $S_{normal2}$ are the normalized similarity scores of 3D depth image and Y intensity image, respectively, and S_{fusion} is the fusion score of the two by weighted sum rule.

4. Experimental Consideration

4.1 Database of CASIA3D

This paper uses the released Face Database provided by Center for Biometrics and Security Research (CBSR (2004)) called CASIA3D database [12], which is captured by non-contact 3D digitizer, Minolta Vivid 910, Between August 2004 and September 2004, consisting of 4624 scans of 123 persons, not only with variations of poses, expressions and illuminations, but also with combined variations of expressions under illumination and poses under expressions. Each person contains 37 or 38 scans, and from each scan, one 2D color image and one 3D facial triangulated surface are also generated. In our experiments we only use the 3D range data and the 2D color image of the first 30 people (DB1) and chose 16 nearly frontal view as our experimental database. And for each people, we use the first 3 scans with neural expression and 5 scans with different expressions (smile, laugh, anger, surprise, eye closed) as the train database, others as the test model.

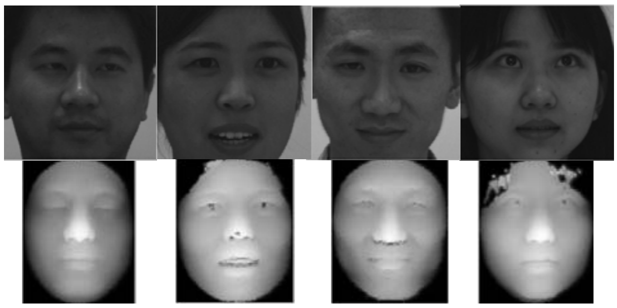


Fig. 5 several samples of intensity images and depth images

After preprocessing and normalization step, we get 2D intensity images of size 200×200 and 3D depth images of size 150×130 , and several samples are shown in Fig. 5, of which the first row are the intensity images, and the second row are the corresponding depth images.

4.2 Experiment Setup

We listed our recognition rate in 6 experiments:

(i) we use 2DPCA to project both 2D intensity image (Y image) and 3D depth image to the feature space, and then chose 60 leading eigenvectors (60 leading eigenvectors can achieve a good performance for 2DPCA and 2DFLD) to denote 2D and 3D images, respectively. Then the similarity scores of the two are calculated and normalized. Finally, we fuse the two scores by weighted sum rule.

(ii) we use 2DFLD to project both 2D intensity image and 3D depth image to the feature space and chose 60 leading eigenvectors to denote 2D and 3D images, respectively. Then the similarity scores of the two are calculated and normalized. Finally, both scores are fused by weighted sum rule.

(iii) We use LBP to encode both 2D and 3D images, and subdivide them into three parts according to face characteristics. Then histogram of each part is calculated and cascaded, later we calculate histogram intersection of 2D and 3D data as there similarity scores, respectively. Finally, we fuse both scores by weighted sum rule.

(iv) We use LBP to encode both 2D and 3D images and then follow method of (1) to calculate recognition rate.

(v) We use LBP to encode both 2D and 3D images and then follow method of (2) to calculate recognition rate.

(vi) Based on the results of the above five experiments, we combine the best schemes of intensity images which achieved highest recognition rate and the best schemes of depth images. Then we fuse their scores by weighted sum rule to calculate recognition rate. This scheme will be denoted as “Scalable fusion” in this paper.

4.3 Results and Discussions

All of the recognition results are listed in Table 1, and the corresponding ROC curves of intensity images, Depth images and Fusion methods are shown in Fig.6, Fig.7 and Fig.8, respectively.

From Table 1 and the corresponding ROC curves, we can make the following observations:

(i) Low recognition rate occurs when use homogenous face information of intensity images or of Depth images. In our experiments, the intensity image get a lowest recognition rate when use 2DFLD to extract face features (50.13%), because intensity images are sensitive to changes in illumination, 2DFLD can not overcome this problem. The depth image get a lowest recognition rate when use LBP and 2DPCA to extract face features (66.38%)), because depth image is the depth information of a face, if two people has a similar change of depth information, their LBP representation will be similar, and 2DPCA can not discriminate it, which will downgrade the recognition rate.

(ii) After LBP encoded the intensity images, they obtain a better performance when 2DPCA, 2DFLD implemented, and achieved a highest recognition rate of 93.82% when use LBP and 2DFLD to extract face features. Since LBP is a nonparametric and computationally simple descriptor of local texture patterns, it is invariant to monotonic gray scale transformation. Hence the LBP representation may be less sensitive to changes in illumination. So after LBP encoded intensity images, the recognition rate raised greatly.

(iii) Depth image has a higher recognition rate than that of intensity image when use 2DPCA, 2DFLD or LBP to extract face features, and when use LBP it achieved a highest recognition rate of 91.88%. The reason is that depth image is a representation of inherent property of a people. It can never be affected by illumination changes. In contrast, the intensity images can be affected greatly by that.

Table 1: Recognition rate of 6 schemes in CASIA3D (30 people)

Experimental schemes	Intensity images	Depth images	Scores fusion
2DPCA	79.97%	83.85%	87.87%
2DFLD	50.13%	88.76%	88.76%
LBP	87.52%	91.88%	92.66%
LBP+2DPCA	85.78%	66.38%	87.08%
LBP+2DFLD	93.82%	88.29%	93.82%
Scalable fusion	93.82%	91.88%	94.68%

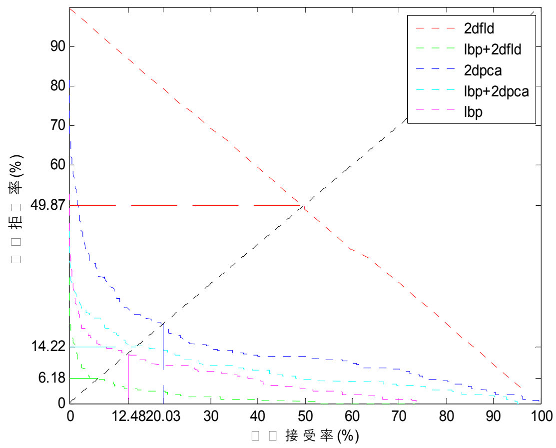


Fig. 6 ROC curves of intensity images

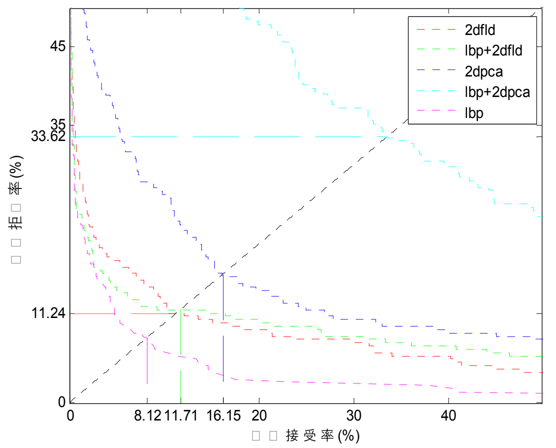


Fig. 7 ROC curves of depth images

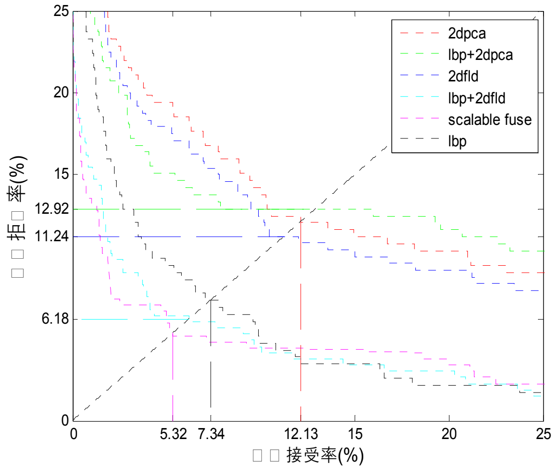


Fig. 8 ROC curves by fusing scores of intensity and depth images

Table 2: Comparison of recognition rate in CASIA3D (30 people)

Reference	Fusion schemes	Recognition rate
Paper[13]	3D points and G-H	82.4%
Paper[14]	2D+3D using CCA	87.04%
2D&3D-ComFusFace	Scalable score fusion	94.68%

(iv) In most cases, fusion scores of both intensity images and depth images improved the final recognition rate and can overcome low recognition rate which may exist in method using mono-information. When combine the best feature extraction method of intensity image and that of Depth image, we get a further improved recognition rate of 94.68%, and this is the best combination of 2D intensity image and 3D depth image for face recognition we sought out. It be named as 2D&3D-ComFusFace.

Base on the experiments mentioned above, we construct our final 2D&3D-ComFusFace method by scalable fusion of 2D intensity images and 3D depth image scores. To compare the performance of 2D&3DComFusFace with that of state-of-the-art, we also refer to paper [13] and paper [14], who use the same database. The former use 3D point clouds and then proposed a novel shape variation representation based on Gaussian-Hermite moments to characterize an individual, and the recognition rate of it is shown in the second row of Table 2. The later use the CCA to learn the mapping between 2D face image and 3D face data, and the recognition rate of it is shown in the third row of Table 2.

We can see that our approach outperforms and has a higher recognition rate. The experiments on the CASIA3D database show that the proposed approach can work well with 3D depth images and 2D intensity image, and also can deal with variation of pose and some changes of expression to a certain extent.

5. Conclusion

This paper combines 2D intensity information and 3D depth information for face recognition. In the preprocessing step we use a novel pose normalization method for 3D range data and then transit them to be depth image. After that, we use 2DPCA, 2DFLD, LBP and their combination to extract features of both 2D intensity image and 3D depth image and calculate the similarity scores. Finally we fused the scores by weighted sum rule to get a further improved performance.

The proposed feature extraction schemes show that the combination of LBP and 2DFLD can do a good performance for 2D intensity image. Meanwhile, LBP and HI can reach highest recognition rate for 3D depth image, Fusing scores of the two achieved the highest recognition rate. We think it is the best combination of 2D intensity and 3D depth images in CASIA3D database when use 2DPCA, 2DFLD and LBP to represent a face. The proposed method of 2D&3D-ComFusFace may do a good job in a real face recognition system, if it captures 2D intensity image and 3D range data.

Acknowledgment

The work was supported by the Open Research Funding of HPCSIP Key Laboratory, Ministry of Education, China, under Grant No. HS201107, the National Grand Fundamental Research (973) Program of China under Grant No. 2011CB911102 and 2010CB834303, the National Natural Science Foundation of China under Grant No. 60973154.

The authors would like to express their cordial thanks to CBSR for providing database of CASIA3D.

References

[1] G. P. Kusuma, C.S. Chua, "PCA-based image recombination for multimodal 2D+3D face recognition," Image and Vision Computing, vol.29, pp.306-316, 2011.

[2] A. F. Abate, M. Nappi, D. Riccio, "2D and 3D face recognition: A survey", Pattern Recognition Letters, vol.28, pp.1885-1906, 2007.

[3] T. Joshi, S. Dey, D. Samanta, "Multimodal biometrics: state of the art in fusion techniques," Int. J. of Biometrics, vol.1, pp.393-417, 2009.

[4] L. Wang, L. Ding, X. Ding, "Improved 3d assisted pose invariant face recognition," In: Proc. of ICASSP, pp. 889-892, 2009.

[5] X. Lu, A. K. Jain, D. Colbry, "Matching 2.5d scans to 3d models," IEEE Transactions on Pattern Analysis and Machine Intelligence vol.28(1),pp.31-43, 2006.

[6] Z. Liu, C. Liu, "Fusion of color, local spatial and global frequency information for face recognition," Pattern Recognition, vol.43, pp.2882-2890,2010.

[7] K. I. Chang, K. W. Bowyer, P. J. Flynn, "An evaluation of multimodal 2d+3d face biometrics," IEEE Transactions on

Pattern Analysis and Machine Intelligence, vol. 27(4), pp.619-624, 2005.
[8] http://tech.groups.yahoo.com/group/OpenCV
[9] J. Yang, D. Zhang, “Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition,” IEEE Trans. Pattern Anal. Machine Intell. (PAMI), vol.26(1), pp.131-137, 2004.
[10] T. Ahonen, M. Pietikainen, “Face Recognition with Local Binary Pattern,” In: Proc. Of European Conference on Computer Vision, pp.469-481, 2004.
[11] A. S. Mian, M. Bennamoun, R. A. Owens, “An efficient multimodal 2d-3d hybrid approach to automatic face recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.29(11), pp.1927-1943,2007.
[12] http://biometrics.idealtest.org/dbDetailForUser.do?id=8
[13] X. Chenhua, W. Yunhong, T. Tieniu, “3D Face Recognition Based on G-H Shape Variation,” SINOBIOMETRICS 2004, pp.233-243, 2004.
[14] W. Yang, D. Yi, Z. Lei, “2D-3D face matching using CCA,” In: Proc. of IEEE International Conference on Automatic Face and Gesture Recognition, pp.1-6, 2008.

Juan Zhou received her MS Degree in biomedical engineering from Northeast University, China in 2008. Currently she is a Ph.D. candidate in Shanghai Institute of Applied Physics, Chinese Academy of Sciences. Her research interests are 3D face recognition, multimodal face recognition, medical images processing and 3D reconstruction.

Yongping Li received MS degree in the Electrical Engineering in 1989 from the Graduate School of Chinese Academy of Sciences, and received Ph.D. degree in Pattern Recognition, with emphasis on Biometrics, in 2000 under the supervision of Professor Josef Kittler. Currently, he is a research professor and Ph.D. student supervisor at Shanghai Institute of Applied Physics. He is also a part-time Ph.D. student supervisor at the Center for Biometrics and Security Research (CBSR), Institute of Automation, Chinese Academy of Sciences. His current research interests include digital signal processing for MCU and DSP based instrumentations, biometric algorithms for face, voiceprint and fingerprint recognition, multiple biometrics, biometric solution integration, biometrics system performance evaluation, testing, and standardization. Dr. Li is a member of the IEEE, IEEE computer society and IEEE instrumentation and measurement.

Jingyan Wang received his B.S. degree from the School of Information Science and Engineering, Central South University, China in 2003, and his Ph.D. degree from Graduate University of Chinese Academy of Science, China in 2012. Currently, he is a postdoctoral research fellow in the Mathematical and Computer Sciences and Engineering Division, King Abdullah University of Science and Technology, Saudi Arabia. His research interests are Machine Learning, Pattern Recognition, Information Retrieval and their applications to Bioinformatics, Biometrics, Medical Imaging and Computer Vision. Jingyan Wang is a Member of IEEE, IEEE Computer Society, and IEEE Engineering in Medicine and Biology Society.