UDC 57.087.1:004.932
Original scientific paper
Received: 16.10.2017.

Sparse representation for pose invariant face recognition

Zhi Chen(1), **Wei Shen**(1), **Yumin Zeng**(2)

(1) *Department of Electronic and Information Engineering, Taizhou Polytechnic College, Taizhou, Jiangsu Province, CHINA*

(2) *School of physical science and technology, Nanjing Normal University, Nanjing, Jiangsu Province, CHINA*
e-mail: 17784068@qq.com

SUMMARY

Face recognition is easily affected by pose angle. In order to improve the robustness to pose angle, we need to solve the pose estimation, face synthesis and recognition problem. Sparse representation can represent a face image with linear combination of atom faces. In this paper, we construct different pose dictionaries using face images captured under the same pose angle to estimate pose angle and synthesize front face images for recognition. Experimental results show that sparse representation can estimate pose angle accurately, synthesize near frontal faces very well and significantly improve the recognition rate for large pose angles.

KEY WORDS: *face pose estimation, sparse representation, face synthesis, face recognition.*

1. INTRODUCTION

As a research hotspot of pattern recognition and computer vision, face recognition has attracted more and more attention from researchers [1, 2]. Compared with other biometric identification techniques, face recognition has the advantages of easy access and no needs of manual authentication, so it has extensive application prospects and higher research value [1, 2]. Although the automatic face recognition under certain conditions (such as frontal face images under controlled lighting) has obtained comparable accuracy to human beings, the performance of automatic recognition has dropped significantly due to the face images easily affected by lighting, pose, expression and age. How to achieve robustness of face recognition under various conditions has become a very important research direction.

Under different pose angles and illumination conditions, difference in the face images of the same person is usually larger than that between the face images acquired under the same condition from different persons [2]. One solution is to synthesize normal illuminated frontal face images for recognition [3]. This approach generally needs to solve the following three problems: pose estimation, face synthesis and face recognition to deal with pose invariant face recognition.

Pose estimation methods can be divided into two categories based on 2D images and 3D depth images [4, 5]. 2D image based approaches can be further categorized into the following two groups. One approach first localizes key facial fiducial points, such as eyes, nose and mouth etc. Then use the well trained model to estimate the pose angle based on the local facial features around fiducial points [6]. The other approach attempts to learn the relationship between facial appearance and the pose angle then utilize statistical method to estimate pose angle [7, 8]. Methods based on 3D depth images need special sensors to obtain the depth information, but usually no 2D appearance information is acquired. Meanwhile these methods also need to locate specific facial features (such as eyes and nose) and estimate pose angles through models [9]. In this paper, we only deal with the texture information, so we only discuss the 2D image methods.

Face synthesis can also be grouped into 2D and 3D methods [10]. 2D methods generally models relationship between facial fiducial points and pose angles and use 2D affine transformation to map the local appearance of facial points of certain pose angle to the corresponding facial points at other pose angles [11, 12]. 3D methods utilize the deformable 3D face model to integrate the shape information and texture information [13], which can get a good estimation of pose to synthesize texture information.

Face recognition across different poses can be classified into four categories [10]: 1. face image normalization, in which both gallery and probe images are normalized to frontal view based on the model and classification is based on the normalized images [14]; 2. face image synthesis, in which multiple virtual face images at various poses are synthesized for each gallery and the probe image is compared with the synthesized gallery images of the same pose [15, 16]; 3. pose robust features, which builds a Pose Adaptive Filter (PAF) to select representative feature points to extract facial features that are less sensitive to the pose for identification [10]; 4. parameter matching, which applies existing model to represent facial shape and texture information as different parameters and identification is performed by parameter matching [17].

Generally several different methods are used for face pose estimation, synthesis and recognition. However, if a system uses several different methods to deal with different problems, it will become more complex, which is not conducive to system development, maintenance and wider application. Sparse representation technology proposed in this paper is applied to solve the problem of pose estimation, image synthesis and face recognition. Different from most existing 2D methods that need to landmark face fiducial points for pose estimation and face synthesis, the propose approach use reconstruction error and linear combinations of face images for pose estimation and face synthesis without complex fiducial points detection.

We first introduce sparse representation in section 2. Then we discuss in details of applying sparse representation for pose estimation, face image synthesis and face recognition in section 3, followed by empirical evaluation of the propose method on the above three problems in section 4. Finally, we draw a conclusion and discuss future work in section 5.

2. SPARSE REPRESENTATION

Sparse representation is based on the long-term observation that people found if the basis of the signal subspace is appropriately selected, the natural signal can be accurately represented. It was originally used in signal compression at lower sampling rate [18 - 20]. Suppose a signal needs n numbers to be represented in the spatial domain $x \in \mathbb{R}^n$, if the basis of the subspace is properly chosen, then x can be expressed by a linear combination of only $d (d<n)$ atomic signals. The collection of all the atomic signals constitutes a dictionary D of the signal subspace.

In the signal subspace, a dictionary D can be used to represent all possible signals, and each signal can be accurately approximated by a few atoms. That is to say that any signal x can be sparsely represented with the signal dictionary D. The sparse representation can be solved by the following l_0 norm optimization:

$$min_w \|w\|_0 \text{, subject to } x = D_w \tag{1}$$

where $\|w\|_0$ means l_0 norm, that is, the number of nonzero coefficients. However, the above problem is non-convex and finding the unique sparsest solution is non-deteministic polynomial-time hardness (NP-hard) and difficult to approximate [21]. In general, finding the sparsest solution cannot be more efficient than exhausting search of all subsets of the combination of atomic faces. In a greedy search method based on orthogonal matching pursuit is proposed to approximate the l_0 optimization [22]. Recent research discover that under sufficiently sparsity constraint of x, the above equation can be transformed into the following l_1 norm optimization:

$$min_w \|x - Dw\|_2^2 + \lambda \|w\|_1 \tag{2}$$

This problem can be solved in polynomial time by standard linear programming methods [23]. From equations (1) and (2) we can assure that signal x can be recovered (synthesized) by the signal dictionary D as Dw, which would be used in our proposed approach for pose estimation and face image synthesis.

3. SPARSE REPRESENTATION FOR POSE INVARIANT FACE RECOGNITION

Sparse representation has been applied successfully on various computer vision tasks, including detection, segmentation and classification [24 - 28]. In this section, we will discuss in detail about how to apply sparse representation to simultaneously handle pose estimation, face image synthesis and face recognition.

3.1 POSE ESTIMATION BASED ON SPARSE REPRESENTATION

We use the reconstruction error between the original face image and the synthesized image using various pose dictionaries to estimate the pose angle. Due to the large face image difference between various poses, if we apply the atomic signal under the same pose α to compose the corresponding pose dictionary D_α, based on sparse representation hypothesis, the face image under pose α can be accurately reconstructed by D_α, while images of other poses can not be accurately recovered by D_α. Therefore, according to the residual error of image restoration, the pose of face image can be determined. If n face pose dictionaries are constructed, for the facial image x of a certain person, according to Eq. (2) using each pose dictionary D_i to recover x, the pose angle the dictionary corresponding to the smallest residuals can be considered as the pose of image x. The above description can be expressed as the following equation:

$$min_\alpha \|x - D_\alpha w_\alpha\|_2 \tag{3}$$

$$w_\alpha = argmin_w \|x - D_\alpha w\|_2^2 + \lambda \|w\|_1 \text{, } \alpha \in (1, \ldots, n) \tag{4}$$

Equation (3) indicates that the choice of the pose dictionary with smallest residuals to reconstruct image x, and Eq. (4) is the sparse representation solution for each dictionary using l_1 norm optimization. The pose angle of the dictionary with the minimal residual error can be regarded as the pose angle estimation of image x, complying with the assumption that face image can be reconstructed precisely using the pose dictionary of the same pose angle.

3.2 IMAGE SYNTHESIS BASED ON SPARSE REPRESENTATION

If we know the sparse representation of a signal, the signal can be reconstructed by Eq. (1). Therefore, if we know the sparse representation of a frontal face, we can recover the frontal face directly according to Eq. (1). But the sparse representation is obtained under the condition that signal is known. If the face images are given in other poses, how can we synthesize their corresponding frontal faces? The face of the same person in different views can form a smooth manifold, and the coefficient of synthesis in different perspectives remains the same [29]. That is to say, if the atomic faces of different pose dictionaries are from the same set of person and are in the same order, then face images for the same person at different pose angles can be represented with a consistent sparse representation using the corresponding pose dictionary. For example, if a frontal face image of person x can be reconstructed using atomic faces *1*, *3*, *10* from the frontal dictionary with weight *0.1*, *0.5*, *1.0* respectively. Then the face image at pose angle α of person x can also be reconstructed using atomic faces *1*, *3*, *10* from the corresponding pose dictionary D_α with the same weight. A requirement is that atomic faces *1*, *3*, *10* of the frontal face dictionary are the same person as the atomic faces *1*, *3*, *10* of pose dictionary D_α respectively. Therefore, when we know the face image x_α of a certain pose, we can get the sparse representation w_α based on the pose dictionary D_α. If we want to get the image of the same person's face in other poses, we can use the corresponding pose dictionary D_β to replace D_α, which is indicated by the following equation:

$$x_\beta = D_\beta w_\alpha \,,\ \ subject\ to\ \ w_\alpha = arg\,min_w \left\| x - D_\alpha w \right\|_2^2 + \lambda \left\| w \right\|_1 \tag{5}$$

For the convenience of cross pose face recognition, we usually synthesize the frontal face.

3.3 FACE RECOGNITION BASED ON SPARSE REPRESENTATION

When sparse representation is used for face recognition, theoretically the coefficient w of the same person's face is unique and sparse. However, due to the noise and the error of sparse optimization, w is not consistent and contains many small non *0* items, which will reduce the robustness of the recognition. In this paper, the minimum residual method is used to make the identification. That is, find the face from all frontal faces with a minimal error to the synthesized frontal face image x_β.

$$identy(x) = arg\,min_i \left\| x_\beta - x_0^i \right\|_2 \tag{6}$$

By far we have used sparse representation to solve the problems in face pose estimation, face synthesis and face recognition. Given a face image of a certain person, firstly we use the Eq. (3) and (4) to estimate its pose angle α, and obtain its sparse representation w_α in pose dictionary D_α. Then we use Eq. (5) to synthesize the corresponding frontal face. Finally we compare the synthetic face with all front faces in the database, and determine the identity according to Eq. (6).

4. EXPERIMENTAL RESULTS AND ANALYSIS

We have selected the FERET (the face recognition technology) database [30] and choose the pose subset with different face poses to verify the performance of the propose approach on face classification across different poses for three aspects including pose estimation, face synthesis and face recognition. The pose subset contains a total of *1800* face images from *200* people with *9* different poses. Face images are aligned, cropped and normalized to *64* by *64* pixels according to the position of the eyes provided by FERET dataset, see Figure 1. We have selected *900* images of *100* individuals to build the pose dictionary on *9* poses respectively, with the rest of the *900* images for testing.

Fig. 1 *Face image samples of FERET pose subset*

4.1 POSE ESTIMATION

In this subsection, we will evaluate the accuracy of sparse representation on pose angle estimation. According to the experimental results, we set the parameter $\lambda=7$. Table 1 lists the pose estimation results for *9* different poses. The first row of Table 1 shows the actual face pose angle, and the second row demonstrates sparse representation pose estimation accuracy. The third row is the estimation accuracy based on the relaxed pose constraint. That is if the estimated angle of sparse representation is the nearest angle to the ground truth, it is also considered as the correct estimation. For example, when estimating *+40*° face pose image, if the estimation of the sparse representation is *+25*°, it is also regarded as the correct estimation. Table 1 indicates that the estimation accuracy on sparse representation for larger view angle is higher than that of the smaller view angle. This is because the larger the view angle, the more significant the difference between the face images across neighborhood views. Thus the image reconstruction residual of larger view angle by the neighborhood pose dictionary is relatively greater. Images of smaller view angles (*+15*°~*-15*°) are relatively similar, thus the corresponding pose dictionary can reconstruct the images of neighborhood view angles with less error. From Table 1 we can see that, sparse representation can estimate the pose angle of face image quite well, with an average accuracy of *81.2%*, and for the relaxed constraint pose estimation, the estimation accuracy is close to *100%*.

Table 1 *Pose angle estimation accuracy of sparse representation*

angle	*+60°*	*+40°*	*+25°*	*+15°*	*0°*	*-15°*	*-25°*	*-40°*	*-60°*	*average*
Accuracy 1	*0.9*	*0.81*	*0.79*	*0.79*	*0.78*	*0.8*	*0.75*	*0.77*	*0.92*	*0.812*
Accuracy 2 (relaxed constraint)	*1.0*	*0.99*	*1.0*	*1.0*	*1.0*	*1.0*	*1.0*	*1.0*	*0.99*	*0.998*

Figure 2 illustrates the average reconstruction error of different pose images using different pose dictionaries. The x axis and y axis represent the pose angle and the brightness illustrates reconstruction error where the brighter the color is the larger the error will be. As can be seen from this figure, synthesis error of the face images using corresponding pose dictionary of the same view angle is the smallest, which is shown in the diagonal. Synthesis error by using neighborhood pose dictionary is relatively smaller. Whilst, synthesis error by other pose dictionaries is generally larger as shown in the top right and bottom left corner.

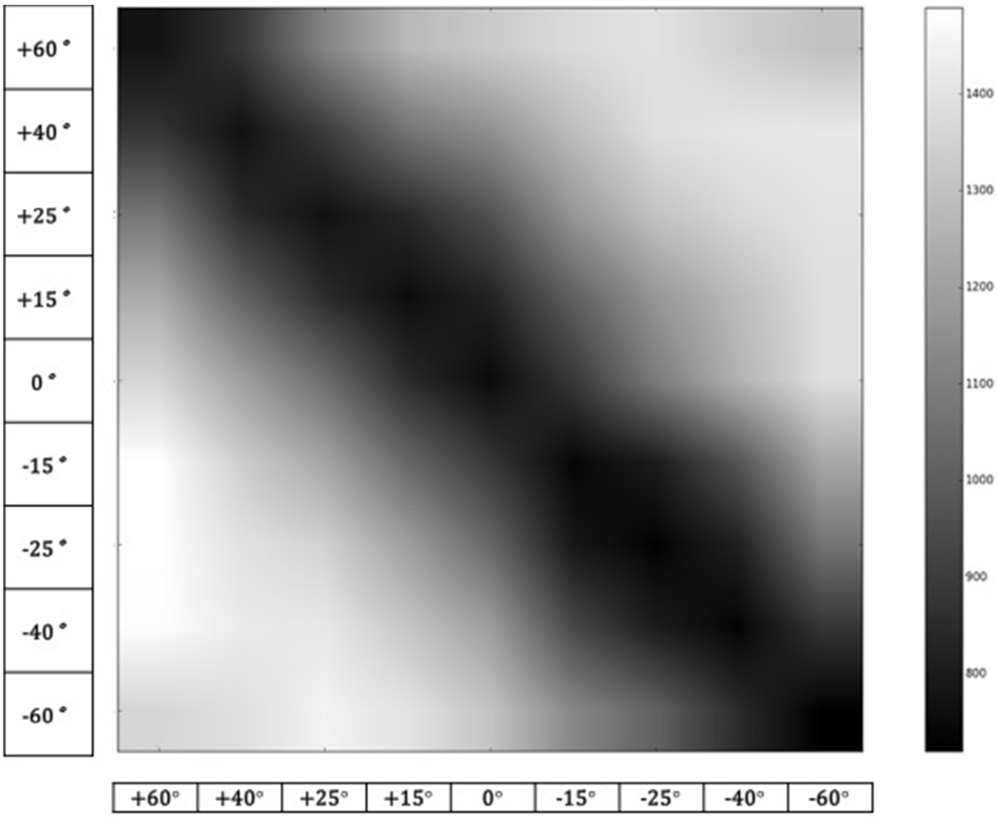

Fig 2. *Average synthesis error of different pose images using different pose dictionaries*
The brighter the grayscale is, the larger the error is

4.2 IMAGE SYNTHESIS

Figure 3 shows an example of a synthesized frontal face image based on Eq. (5) from face images captured at different pose angles. The figure shows that, the synthesized frontal face images from different view angles are very similar to each other, which are also similar to the real frontal face (the fifth picture of row 1). For large angle (±*60*°), the synthesized image of the frontal face is blurry around the mouth. This is because when we normalize the images to align and scale the face according to the eyes location, mouth area is partly missing for large view angles, which leads to the lack of information for face synthesis. If using less atoms for reconstruction, the blurry region may be reduced. However, because all the test images are unseen and using less atoms the overall reconstruction error will be much higher. Our synthesis is not attempt to reconstruct a very clear face with fine details but to reconstruct a frontal face that preserve similar facial features. In this way, the final face recognition performance is higher.

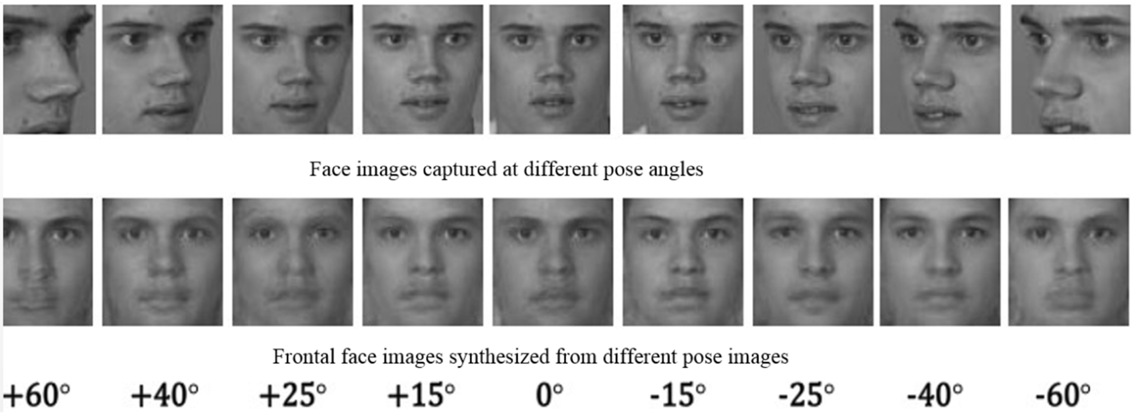

Fig. 3 *Face images with different pose*

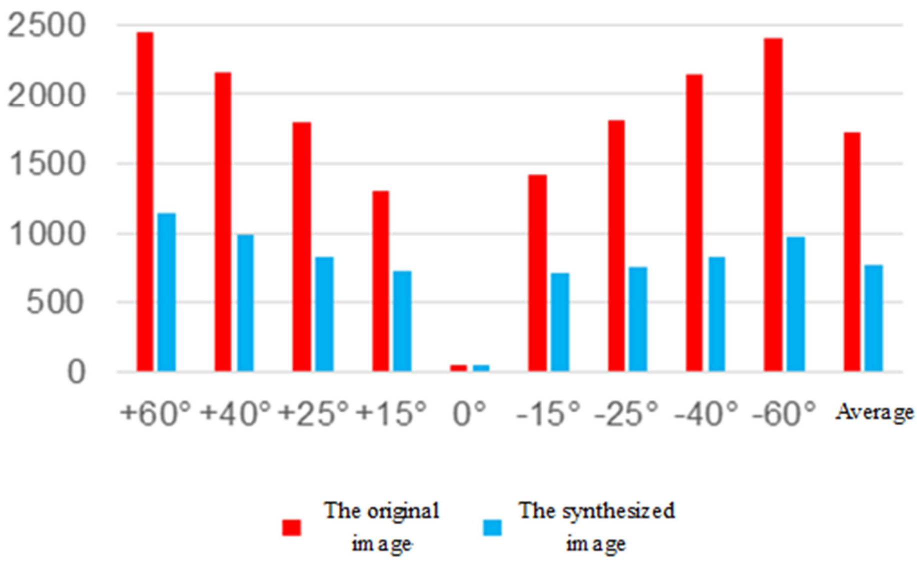

Fig. 4 *Difference between the original image and the synthesized image to the real frontal face*

The difference between the original image, the synthesized image to the corresponding frontal face is shown in Figure 4. The smaller the view angle is, the smaller the difference between the corresponding image and the frontal face image is. Compared to the original image, the difference between the synthesized face image and the frontal face is lower than *50%*. The average difference dropped from *1718* for the original image to *772* for the synthesized image. In addition, the change of difference for synthesized images with different pose angles are much smaller than that for original images, which is consistent with samples shown in Figure 3. Therefore, frontal face image synthesis using sparse representation can significantly reduce the appearance change of face images across pose angle, leading to improved robustness for face recognition with pose variations.

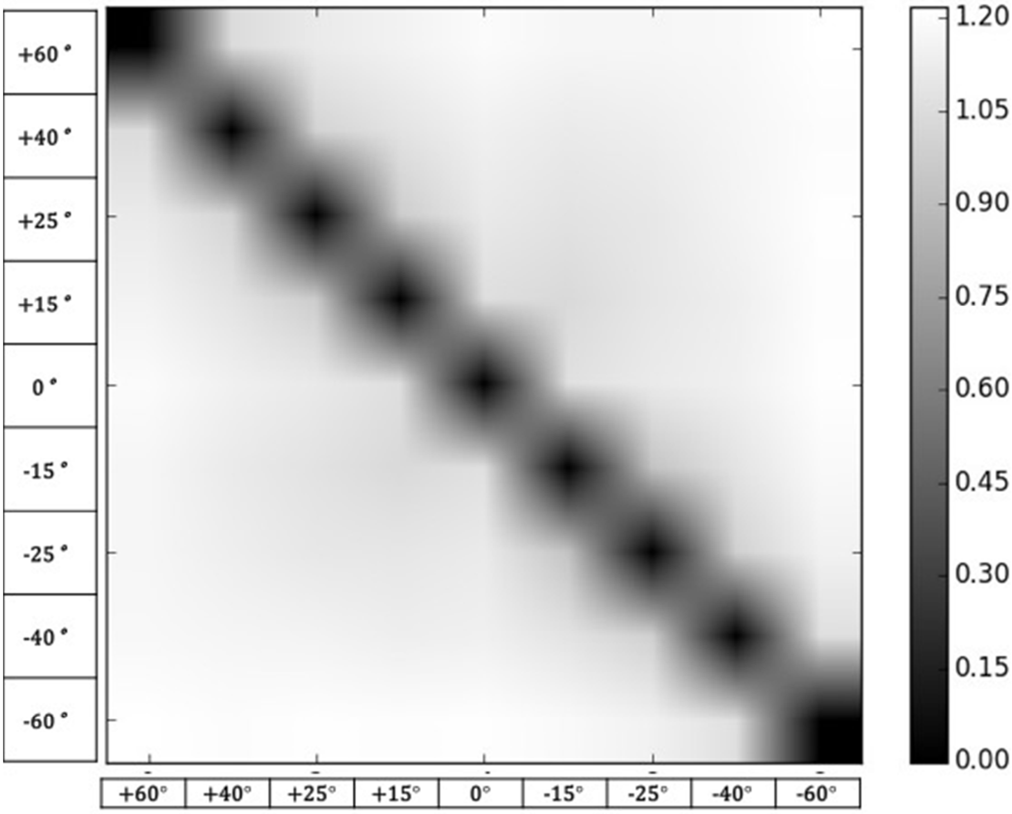

Fig. 5 *Difference between the coefficients of the face sparse representation using different pose dictionaries. The brighter the color the larger the difference is.*

In order to compare the sparse representation coefficients of human faces from different view angles with corresponding pose dictionaries, we use Eq. (7) to normalize the distance between the coefficients of the face images generated with different pose dictionaries.

$$d_{coef} = \frac{\left\| w_\alpha - w_\beta \right\|_2}{\frac{1}{2}\left(\left\| w_\alpha \right\|_2 + \left\| w_\beta \right\|_2 \right)} \quad \textbf{(7)}$$

ω_α and ω_β are the sparse representation coefficients for the same face images obtained from different pose dictionaries D_α and D_β respectively.

Figure 5 shows the difference between the coefficients of the sparse representation across different poses. The *x* axis and *y* axis represent pose angles and the brightness indicates the difference. The brighter the color is, the larger the difference between two coefficients. We found that the sparse representation coefficients between neighborhood pose dictionaries are generally smaller. The greater the difference between view angles for pose dictionaries is, the bigger the difference between their coefficients, which is consistent with the results shown in Figure 2.

Non blurry background will have little impact on the face synthesis and recognition. First, the background region in each face image only occupy a very small section and mainly turn up in very high pose angles. Second, the proposed sparse representation method consider the whole face image as one atom, thus the weights extracted by the algorithm is mainly related to the facial features and is less likely affected by the non-face background. Third, after the reconstruction, the synthesized frontal view face image will not contain non-blurry background because in the dictionary all the face atoms only has blurry background. Fourth, if the frontal view image in the database has strong background, we can use the frontal pose dictionary to reconstruct the frontal view image first to remove the background.

4.3 FACE RECOGNITION

We use the K-nearest neighbors (KNN) algorithm to test the original image and the synthesized image in the FERET database. Only *100* frontal faces are used as training data, and the remaining 800 faces of different view angles are used as test data. Table 2 shows the recognition rate of the original image and the synthesized image of different view angles. From which we can see that the recognition rate of smaller view angle is higher than that of the larger view angle. This is because we use the frontal face image as the training data and the appearance of smaller view angle face images are more similar to the training data. The recognition rate of the synthesized image is *31* percentage points higher than that for the original image on average. This improvement is more significant for very large view angles (±*60°*), where recognition accuracy increases by more than *400%* from less than 10 percentage points to nearly *40* percentage points.

Table 2 *Face recognition results of synthesized images and original images on different view angles.*

viewing angle	*+60°*	*+40°*	*+25°*	*+15°*	*0°*	*-15°*	*-25°*	*-40°*	*-60°*	*average*
synthetic	*37%*	*53%*	*75%*	*83%*	*100%*	*85%*	*78%*	*68%*	*42%*	*69%*
original	*7%*	*14%*	*36%*	*69%*	*100%*	*62%*	*29%*	*14%*	*10%*	*38%*

5. CONCLUSION

In this paper, we discuss the application of sparse representation to pose estimation, face synthesis and face recognition. From the analysis of experimental results, we discover that sparse representation technique can well estimate the pose angle of face images, and it can be used to synthesize the frontal face images from different view angles with relatively high accuracy, which can greatly improve the recognition rate for large view angles. However, the proposed method is still sensitive to occlusions to certain extent. If part of the face image is invisible, for example covered by face mask or sunglasses, then the reconstruction error and face synthesis quality may be affected. This problem needs to be investigated in the future to further improve the robustness for occlusion.

6. REFERENCE

[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] W. Zhao, R. Chellappa, P.J. Phillips and A. Rosenfeld, Face recognition: A literature survey, *ACM Computing Surveys*, Vol. 35, No. 4, pp. 399–458, 2003.

[3] X. Lu, R.-L. Hsu, A.K. Jain, B. Kamgar-Parsi and B. Kamgar-Parsi, Face Recognition with 3D Model-Based Synthesis, International Conference on Biometric Authentication (ICBA), 2004.

[4] P. Chavdar, M. Tim K. and J. Michael, Real-time 3D Head Pose and Facial Landmark Estimation from Depth Images Using Triangular Surface Patch Features, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[5] F. Gabriele, G. Juergen and G. Luc Van, Real Time Head Pose Estimation with Random Regression Forests, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[6] T. Vatahska, M. Bennewitz and S. Behnke, Feature-based head pose estimation from images, IEEE-RAS International Conference on Humanoid Robots, 2007.

[7] V.N. Balasubramanian, J. Ye and S. Panchanathan, Biased manifold embedding: A framework for person-independent head pose estimation, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007.

[8] G. Xin and X. Yu, Head Pose Estimation Based on Multivariate Label Distribution, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[9] V. Blanz and T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 25, No. 9, pp. 1063–1074, 2003.

[10] Y. Dong, L. Zhen and L. Stan Z., Towards Pose Robust Face Recognition, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.

[11] L. Yingchun and S. Guangda, Face Pose Estimation and Synthesis by 2D Morphable Model, International Conference of Computational Intelligence and Security, 2006.

[12] D. Gonzalez-Jimenez and J.L. Alba-Castro, Towardpose-invariant 2-d face recognition through point distribution models and facial symmetry, IEEE Transactions on Information Forensics and Security, Vol. 2, No. 3-1, pp. 413–429, 2007.

[13] B. Volker, G. Patrick, P.P. Jonathon and V. Thomas, Face recognition based on front views generated from non-front images, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[14] A. Asthana, T.K. Marks, M.J. Jones, K.H. Tieu and M.V. Rohith, Fully automatic pose-invariant face recognition via 3d pose normalization, Proceedings of IEEE International Conference on Computer Vision (ICCV), pp. 937–944, 2011.

[15] X. Zhang and Y. Gao, Face recognition across pose: A review, Pattern Recognition, Vol. 42, No. 11, pp. 2876–2896, 2009.

[16] X. Zhang, Y. Gao, and M.K.H. Leung. Recognizing rotated faces from frontal and side views: An approach toward effective use of mugshot databases, IEEE Transactions on Information Forensics and Security, Vol. 3, No. 4, pp. 684–697, 2008.

[17] B. Michael D., K. Daniel, W. Thibaut and G. Luc Van, Real-Time Face Pose Estimation from Single Range Images, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[18] C. Emamnuel J., Compressive sampling, Proceedings of the International Congress of Mathematicians (ICM), 2006.

[19] D.L. Donoho, Compressed sensing, IEEE Transactions on Information Theory, Vol. 52, No. 4, pp. 1289–1306, 2006.

[20] J. Che and X. Huo, Theoretical results on sparse representations of multiple-measurement vectors, IEEE Transactions on Signal Processing, Vol. 54, No. 12, pp. 4634–4643, 2006.

[21] E. Amaldi and V. Kann, On the Approximability of Minimizing Nonzero Variables or Unsatisfied Relations in Linear Systems, Theoretical Computer Science, Vol. 209, pp. 237-260, 1998.

[22] J.A. Tropp and A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Transactions on Information Theory, Vol. 53, No. 12, pp. 4655–4666, 2007.

[23] S. Chen, C. Sanderson, M.T. Harandi and B.C., Lovell, Improved Image Set Classification via Joint Sparse Approximated Nearest Subspaces, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.

[24] J. Wright, A. Yang, A. Ganesh, S. Sastry and Y. Ma, Robust Face Recognition via Sparse Representation, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 31, No. 2, February 2009.

[25] G. Peyre, Sparse modeling of textures, Preprint Ceremade 2007-15, 2007.

[26] R. Raina, A. Battle, H. Lee, B. Packer and A.Y. Ng, Self-taught learning: transfer learning from unlabeled data, International Conference on Machine Learning (ICML), pp. 759–766, 2007.

[27] S. Rao, R. Tron, R. Vidal and Y. Ma, Motion segmentation via robust subspace separation in the presence of outlying, incomplete, and corrupted trajectories, Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[28] S. Chen, D. Donoho and M. Saunders, Atomic Decomposition by Basis Pursuit, Society for Industrial and Applied Mathematics (SIAM) Review, Vol. 43, No. 1, pp. 129-159, 2001.

[29] H. Xinyu, G. Jizhou, C. Sen-Ching S. and Y. Ruigang, Manifold Estimation in View-based Feature Space for Face Synthesis across Poses, Asian Conference on Computer Vision (ACCV), 2009.

[30] A.P.J. Phillips, H. Moon, P.J. Rauss and S. Rizvi, The FERET evaluation methodology for face recognition algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 22, No. 10, October 2000.