

A Novel Mathematical Based Method for Generating Virtual Samples from a Frontal 2D Face Image for Single Training Sample Face Recognition

Reza Ebrahimpour

Assistant Professor, Department of
Electrical Engineering
Shahid Rajaee University
Tehran, P. O. Box 16785-136, Iran

ebrahimpour@ipm.ir

Masoom Nazari

Department of Electrical Engineering
Shahid Rajaee University
Tehran, P. O. Box 16785-136, Iran

innocent1364@gmail.com

Mehdi Azizi

Department of Electrical Engineering
Shahid Rajaee University
Tehran, P. O. Box 16785-136, Iran

azizi_php@gmail.com

Mahdiah Rezvan

Islamic Azad University south Tehran Branch
Tehran, P. O. Box, 515794453, Iran

mhrezvan@gmail.com

Abstract

This paper deals with one sample face recognition which is a new challenging problem in pattern recognition. In the proposed method, the frontal 2D face image of each person is divided to some sub-regions. After computing the 3D shape of each sub-region, a fusion scheme is applied on them to create the total 3D shape of whole face image. Then, 2D face image is draped over the corresponding 3D shape to construct 3D face image. Finally by rotating the 3D face image, virtual samples with different views are generated. Experimental results on ORL dataset using nearest neighbor as classifier reveal an improvement about 5% in recognition rate for one sample per person by enlarging training set using generated virtual samples. Compared with other related works, the proposed method has the following advantages: 1) only one single frontal face is required for face recognition and the outputs are virtual images with variant views for each individual 2) it requires only 3 key points of face (eyes and nose) 3) 3D shape estimation for generating virtual samples is fully automatic and faster than other 3D reconstruction approaches 4) it is fully mathematical with no training phase and the estimated 3D model is unique for each individual.

Keywords: Face Recognition, Nearest Neighbor, Virtual images, 3D face model, 3D shape.

1. INTRODUCTION

Face Recognition is an effective pathway between human and computer, which has a lot of applications in information security, human identification, security validation, law enforcement, smart cards, access control etc. For this reasons, industrial and academic computer vision and pattern recognition researchers have a significant attention to this task.

Almost the face recognition systems are related to the set of the stored images of a person, which called training data. Efficiency of these types of systems considerably falls when the size of training data sample is small (Small Sample Size Problem). For example in ID card verification and mug-shot we have only one sample per person. Several methods have done with the mentioned problem which we will introduce some of them that our idea is given from.

From the primary and most famous appearance based methods we can mention to PCA [1]. Then for one training sample per person, J. Wu et al. introduced $(PC)^2A$ [2] method. In this method, at first a pre-process

on image is done to compute a projection matrix of face image and combine it with the original image, then PCA have being applied on projection combined image. Then, S.C. Chen et al. offered $E(PC)^2A$ [3] method which was the enhanced version of $(PC)^2A$. To increase the efficiency of system they could increase the set of training samples by calculating the projection matrix in different orders and combining it with the original image. In [4] J. Yang offered 2DPCA method for feature extraction. 2DPCA is a 2D extension of PCA and has less computational load compared to PCA with higher efficiency compared to PCA for few training samples.

From another point of view, one can generate virtual samples to enlarge the training set and improve its representative ability, variant analysis-by-synthesis methods are put forward, i.e., the labeled training samples are warped to cover different poses or re-lighted to simulate different illuminations [5-8]. Photometric stereo technologies such as illumination cones and quotation image are used to recover the illumination or relight the sample face images. From this point of view, Shape from shading algorithms [9-11] has been explored to extract 3D geometry information of a face and to generate virtual samples by rotating the result 3D face models.

In our proposed method, we divided the frontal face to some sub-regions. After estimating the 3D shape of each sub-region, we combined them to create 3D shape of whole face. Then, we add the 2D face image with its 3D shape to construct 3D face models. Finally, different virtual samples with different views can be obtained by rotating the 3D face model in different angels.

Compared to previous works [8], this framework has following advantages: 1) only one single frontal face is required for face recognition and the outputs are virtual images with variant views for the individual of the input image, which avoids the burdensome enrollment work; 2) this framework needs only 3 key point of face (eyes and nose) 3) the proposed 3D shape estimation for generating virtual samples is fully automatic and faster than other 3D reconstruction approaches 4) this method has no training phase and is fully mathematical and also the estimated 3D model is unique for each individual .

Experimental results on ORL dataset also prove the efficiency of our proposed method than traditional methods in which only the original sample of each individual uses as training sample.

2. OVERVIEW OF THE PROPOSED SCHEME

Aiming to solve the problem of recognizing a face image with single training sample, an integrated scheme is designed which is composed of two parts: database image synthesis part and face image recognition part. Before recognition, synthesis work would be done on frontal pose of face image. Through the synthesis part, the training database will be enlarged by adding virtual images with other different views. In the recognition stage, One Nearest Neighbor is used to classify the test images. Therefore, the most important part in the proposed scheme is the synthesis one which has crucial affect on the recognition accuracy.

2.1 Face image synthesis

This section gives a summary of the synthesis proposed in our scheme and introduces briefly the key techniques utilized for generating virtual views.

As we know the general shape of human face is almost uniform. It means that the main regions of face such as eyes, nose and mouth nearly have the uniform shape for all human. For example if we consider a typical 3D face image of human in frontal, the region around the eyes has some notch and also the region around the nose has some nub which begin from the center of the brow and its nub increases approximately linearly till tip of the nose. In the proposed method we divided the frontal face image to some sub-regions. After estimating the 3D shape of each sub-region, we combine them to create 3D shape of whole face.

To obtain the 3D shape of the face, we require a distance matrix which can be easily computed from the distance between two lenses in 3D cameras. But in 2D images we need to estimate the distance matrix. Consider the 2D image of face in 3D space as shown in Figure 1.

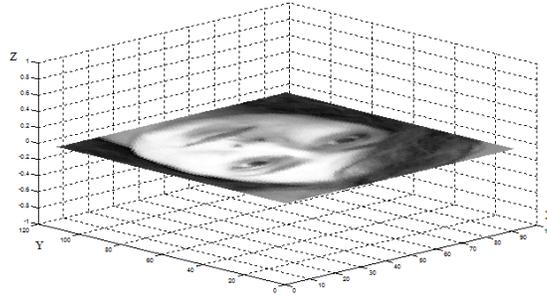


FIGURE 1: A 2D image of face in 3D dimension

Each pixel of this image represents one point in Cartesian X-Y coordinate system and Z can be regarded as distance axis of the image. In our proposed method we aim to estimate the Z matrix of face image to create virtual face images with different views that illustrated in detail as following:

It is worth noting that all of the equations used in our proposed method are obtained heuristically by some manipulation of different values and functions.

1- Consider an $m \times n$ face image. We locate three key points on face image (eyes and nose) automatically using the following method.

Note that to find the location of eyes and nose we need to crop the region of face. To accomplish our method for generating virtual sample, because it is the first step of the proposed scheme and affects the next steps dramatically, the region of face should be cropped with 100% accuracy,. There is no automatic algorithm with the accuracy of 100% until now (although some algorithms [12] with high accuracy need a manually located point of face such as nose location). Thus we crop all of ORL dataset manually as you see in Figure 2.



FIGURE 2: a sample of manually cropped face image

This method finds the region of eyes and nose as following:

- a) Illumination Adjusting the face image and converting into a binary face image (see Figure 3)



FIGURE 3: converting illumination adjusted face image into binary image

- b) Dividing the binary image into three regions to locate the position of left eye, right eye, nose and lips (Figure 4).

To accomplish this task, we have used an eye detector, based on histogram analysis. In order to eye, nose and lip localization, the following steps are performed: 1. Compute the vertical and horizontal projections on the face pixels 2. Locating the top, down, right and left region boundaries where the projecting value exceeds a certain threshold.

We assume that eyes should be located in the upper half of face skin region. Once the face area is found, it may be assumed that the possible eye region is the upper portion of the face region. By analyzing the curve, we find the maximum and minimal point of the projection curve. Figure 4 shows the corresponding relation between these points and the position of facial organ: eyes, nostril, and mouth. Only the positions of eyes and lips are calculated in this case.

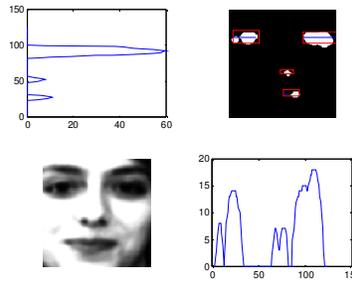


FIGURE 4: Eye, nose and lip localization using vertical and horizontal projections on the face pixels. The red rectangles indicate the boundaries of eyes, nose and lips and blue lines indicate the central lines of them.

Let the center of left and right eye and the tip of nose e_l , e_r and p_n respectively.

2- By using the position of eyes, we can compute the distance between eyes and also the middle point of the distance as shown in equation (1).

$$c_x = \frac{(e_l(x) + e_r(x))}{2}, \quad x \in [1, n] \quad (1)$$

$$c_y = \frac{e_l(y) + e_r(y)}{2}, \quad y \in [1, m]$$

$$\sigma = \sqrt{(e_r(x) - e_l(x))^2 + (e_r(y) - e_l(y))^2}$$

Where σ is the distance and c is its middle point between left eye and right eye.

3-Through the equation (2), we make the face border (including the ears and head border) more sunken.

$$z_{gf}(x) |_{y \in [1, m]} = \frac{2000}{(50\sigma + (x - c_x) / \sigma)} \quad (2)$$

4-We know that the brow region is nearly slick and plane from the side-view and after the eyebrows there is the pone of eyes. By using eye situation, we find the part of the matrix Z which represents brow region and call it Z_b . Thus we can have a good estimate of brow region according to equation (3).

$$z_{fh}(y) |_{x \in [1, n]} = \frac{1}{1 + e^{(y - c_y - (\sigma / (.2\sigma))) / 5}} \quad (3)$$

5- In the previous stage, the points under brow sunk whereas the cheek must be salient. To fix this notch and signalize the cheek region, we can use equations (4).

$$z_{c1}(y) |_{x \in [1, n]} = \frac{1}{1 + e^{(y - c_y - (\sigma / (.08\sigma))) / 5}}$$

$$z_{c2}(y) |_{x \in [1, n]} = \frac{1}{1 + e^{(y - c_y - (\sigma / (.12\sigma))) / 5}} \quad (4)$$

$$Z_{cheek} = z_{c1}(x, y) \cdot z_{c2}(x, y)$$

6- If we pay attention to the downward regions of face, we would find out that in most faces the notch of borders increases nearly exponentially with respect to the center of face. According to equation (5), we estimate the matrix that does this task.

$$z_{df}(x) |_{y \in [1, m]} = e^{-(x - c_x)^2 / \sigma^2} \quad (5)$$

7- In this stage we obtain an estimate for the distance matrix of the nose. By little attention to the general form of human face we find out that the bridge begins from the middle point between of eyes and its nub almost increase linearly with the slight slope till nose tip and then decrease nearly with sharp slope while both side of nose sink exponentially as shown in equation (6).

$$Z_{nose}(x,y)=\begin{cases} \frac{-(x-cx)^2}{(y-cy).(e^{0.05\sigma^2})} & , \quad cy < y < pn \\ 0 & , \quad others \end{cases} \quad (6)$$

8-In the preceding stages, we estimated some different sub-matrixes for the distance matrix (Z) that each of them can estimate one part of the face excellently and in the other regions cause increase in error. Thus, the only important point is how to combine these matrixes. Since in each sub-region of the face corresponding matrix must be used, we used equation (7) to combine estimated local matrixes in order to obtain the total estimate matrix for 3D shape of face image.

$$Z(x,y)=Z_{gf}(x,y).((1-Z_{lf}(x,y)).(1-Z_{fl}(x,y)).(1-Z_{cheek}(x,y))) + Z_{nose}(x,y) \quad (7)$$

Figure 5 schematically represents our proposed method for estimating 3D shape of face.

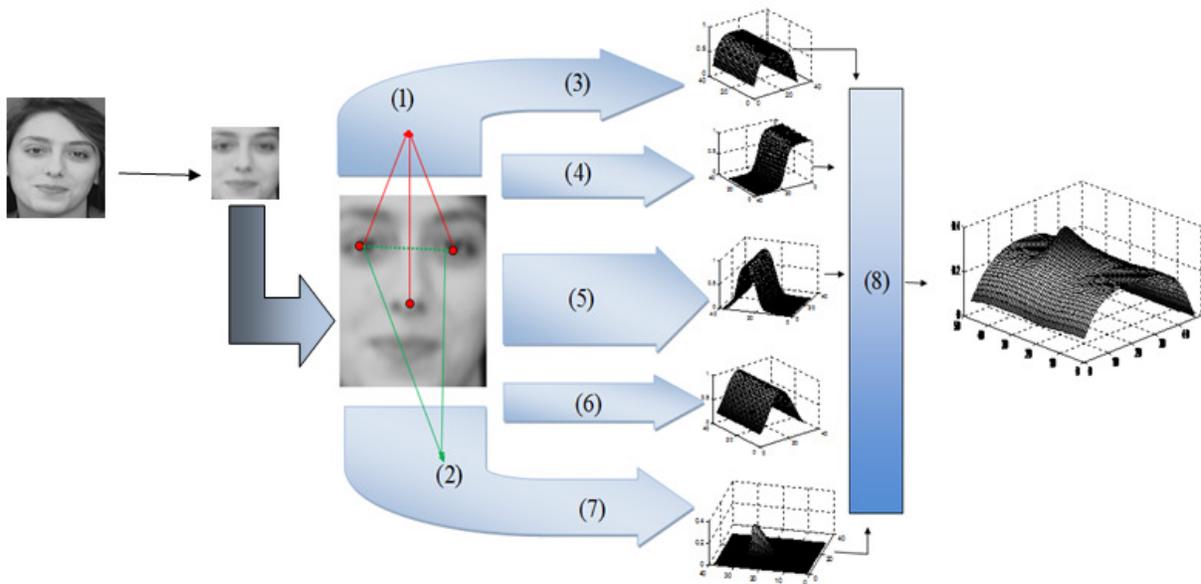


FIGURE 5: Our proposed method for estimating 3D shape of face

9-Now, since we have the distance matrix Z, we can drape the 2D face image over its 3D shape and create 3D face model as shown in Figure 6.

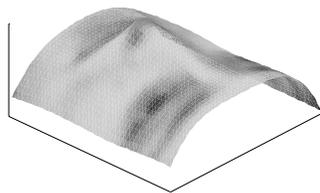


FIGURE 6: 3D face model after draping 2D face image over its 3D shape

10- Finally, we rotate the 3D face model in different views and produce virtual images in some different angles to obtain virtual images with different poses. Figure 7 shows some of virtually generated 3D faces with the proposed method on ORL dataset.



FIGURE 7: Virtual images with different views generated from only a frontal 2D face image **(a):** tilt up (6°) and angle ($-25^\circ : +25^\circ$) **b:** normal and angle ($-25^\circ : +25^\circ$) **c:** tilt down (6°) and angle ($-25^\circ : +25^\circ$) **d:** original image)

3. EXPERIMENTAL RESULTS

In the proposed method we only used frontal face image and generated 18 virtual images with different views vary from -20° to $+20^\circ$. We systematically evaluated the performance of our algorithm compared with the conventional algorithm that do not uses the virtual faces synthesized from the personalized 3D face models.

To test the performance of our proposed method, some experiments are performed on ORL face database which contains images from 40 individuals, each providing 10 different images. For some subjects, the images were taken at different times. The facial expressions and facial details (glasses or no glasses) also vary. The images were taken with a tolerance for some tilting and rotation of the face of up to 20 degrees (-20° to $+20^\circ$) and also some variation in the scale of up to about 10 percent. All images are grayscale and cropped to a resolution of 48×48 pixels. Figure 8 shows some example of ORL dataset.



FIGURE 8: Some samples of ORL database

In all the experiments, the conventional methods used only the frontal faces of each person for training and the other faces are all used for testing. The comparison experiments have been conducted to evaluate the effectiveness of the virtual faces created from the 3D face model for face recognition. We used PCA and 2DPCA for dimension reduction as well as extracting useful features and nearest neighbor for classifying the test images.

Table 1 and 2 compare the results of our proposed method and conventional method.

Table 1 Recognition rate comparison between face recognition with/without virtual face using PCA

<i>Dimension</i>	<i>5</i>	<i>20</i>	<i>40</i>	<i>70</i>	<i>100</i>
<i>Method</i>					
Without virtual views (%)	53.5	70.61	72.58	72.58	72.58
With virtual views (%)	70.12	73.22	78.50	78.40	78.30

Table 2 Recognition rate comparison between face recognition with/without virtual face using 2DPCA

<i>Dimension</i>	<i>(48×1)</i>	<i>(48×2)</i>	<i>(48×5)</i>	<i>(48×8)</i>	<i>(48×10)</i>
<i>Method</i>					
Without virtual views (%)	64.89	73.50	77.44	75.22	74.44
With virtual views (%)	71.37	79.10	82.10	81.20	80.83

By enlarging training data using our proposed method achieve higher top recognition rate (about 5%) than traditional methods in which only one frontal face image is used as training sample.

4. CONCLUSION AND FUTURE WORK

In this paper, we proposed a simple but effective model to make applicable face recognition task in situations where only one training sample per person is available.

In the proposed method, we select a frontal 2D face image of each person and divide it to some sub-regions. After computing the 3D shape of each sub-region, we combine the 3D shape of each sub-regions to create the total 3D shape for whole 2D face image. Then, 2D face image is draped over the corresponding 3D shape to construct 3D face model. Finally by rotating the 3D face image in different angles, different virtual views are generated and added to training sample. Experimental results on ORL face dataset using nearest neighbor as classifier reveal an improvement of 5% in correct recognition rate using virtual samples compared to the time we use only frontal face image of each person.

Compared with other related works, the propose method has the following advantages: 1) only one single frontal face is required for face recognition and the outputs are virtual images with variant views for the individual of the input image, which avoids the burdensome enrollment work; 2) this framework needs only 3 key points of face (eyes and nose) 3) the proposed 3D shape estimation for generating virtual samples is fully automatic and faster than other 3D reconstruction approaches 4) this method has no training phase and is fully mathematical and also the estimated 3D model is unique for each individual .

Our experiments also show the top recognition rate of 82.50% which still is far from satisfactory compared to average recognition accuracy that may be realized by human beings. It is expected that other techniques are needed to further improve the performance of face recognition. A possible way to achieve the mentioned goal is generating more virtual views with different

expression and illumination using more complex techniques, another possible way could be explored on classifiers with more complexity and higher accuracy.

5. REFERENCES

- [1] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.
- [2] J. Wu, Z.H. Zhou, "Face recognition with one training image per person," *Pattern Recognition Letters*, vol. 23, no. 14, pp. 1711-1719, 2002.
- [3] S.C. Chen, D.Q. Zhang, Z.H. Zhou, "Enhanced (PC)2A for face recognition with one training image per person," *Pattern Recognition Letters*, vol. 25, no. 10, pp. 1173-1181, 2004.
- [4] J. Yang, D. Zhang, "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 1173-1181, 2004.
- [5] T. Riklin-Raviv, A. ShaShua, "The quotient image: class based re-rendering and recognition with varying illuminations," *Pattern Anal. Mach. Intell.*, vol. 23, no. 23, pp. 129-139, 2001.
- [6] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 643-660, 2001
- [7] Talukder, D. Casasent, "Pose-invariant recognition of faces at unknown aspect views," *IJCNN* Washington, DC, 1999.
- [8] T. Vetter, T. Poggio, "Linear object classes and image synthesis from a single example image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 733-741, 1997.
- [9] R. Zhang, P. Tai, J. Cryer, M. Sha, "Shape from shading: a survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 8, pp. 690-706, 1999.
- [10] J. Atick, P. Griffin, N. Redlich, "Statistical approach to shape from shading: reconstruction of three dimensional face surfaces from single two dimensional image," *Neural Comput.*, vol. 8, pp. 1321-1340, 1996.
- [11] T. Sim, T. Kanade, "Combining models and exemplars for face recognition: an illuminating example," *Proceedings of the CVPR 2001 Workshop on Models versus Exemplars in Computer Vision*, 2001.
- [12] T. Jilin, F. Yun, and S. Huang, "Locating Nose-Tips and Estimating Head Poses in Images by Tensorposes," *IEEE Trans. Circuit and Systems for Video Technology*, vol. 19, no. 1, 2009