# Pose and expression invariant face recognition with one sample image per person

Hui-Fuang Ng (黃惠藩)
Dept. of CSIE, Asia University
pang@asia.edu.tw

Hong-Wen Chen (陳宏文)
Dept. of CSIE, Asia University
nru123kimo@yahoo.com.tw

## Abstract

Face recognition under pose and facial expression variations is a challenging problem. In this paper, we present a method for pose and expression invariant face recognition using only a single sample image per person. The method utilizes the similarities of a face image against a set of faces from a prototype set taken at the same pose and expression to establish pose and expression invariant similarity vector which can be used for comparing face images of a person taken in different poses and facial expressions. Experimental results indicate that the proposed method achieves high recognition rate even for large pose and expression variations.

**Keywords**: face recognition; pose invariant; expression invariant; principle component analysis; similarity vector.

## 1. Introduction

One of the key problems in face recognition is how to handle the variability in appearance due to changes in pose and facial expression (Zhao et al. 2003). Pose and expression invariant recognition capability is crucial to a face recognition system because in general it is difficult, if not possible, to control the imaging direction and to restrict a person's facial expression when acquiring images of human faces, such as in video surveillance.

A common solution to handling pose variations in face recognition is the view-based method (Pentland and Moghaddam 1994; Murase and Nayar 1995). In this method, face images of the individuals to be recognized are acquired from different view angles. Images of the same view are used to construct an eigenspace representation for each view, and the view-specific eigenspace representations are then used for recognizing a person in different poses. A major limitation of this approach is the need to acquire and store a large number of views for each face. This technique is impractical in situations where only one or a few views of the face to be recognized are available—for example, a passport photo.

Several approaches had been proposed to reduce the effect of facial expression on face recognition problem. For examples, Li et al. (2006) utilized the idea of separating geometry and texture information in a face image and modeled the two types of information by projecting them into separate PCA spaces which are specially designed to capture the distinctive features among different individuals. Subsequently, the texture and geometry attributes are re-combined to form a classifier which is capable of recognizing faces with different expressions. Chen and Lovell (2004) presented an illumination and expression invariant face recognition method which required only one sample training image per person. Their method first applies PCA analysis to construct a subspace for image representation and then rotates and warps the subspace according to the within-class co-variance and between-class covariance of samples to improve class separability under

variations in lighting conditions and facial expression. Both methods, however, do not handle pose variations.

To address the single training sample problem, Beymer and Poggio (1995) used a parallel deformation technique to synthesize virtual views from a real view using 2D images of rotating prototype faces. The combined set of one real view and multiple virtual views is used to provide sample views for pose invariant recognition by the view-based method. A similar approach was reported by Lando and Edelman (1995). The virtual view concept has been extended to include full 3D face model for synthesizing virtual views for invariant face recognition (Blanz and Vetter 2003; Jiang et al. 2005). A 3D model of a human face is generated from a sample view and it is used to synthesize the appearance of the face under different poses, expressions, and lighting conditions. The virtual view approaches require precise correspondence between feature points in the sample view and the stored view; however, finding such correspondence reliably is difficult in practice. Several other methods have been proposed to address the similar problem—for a detailed review of the methods, see Tan et al. (2006) and Wang et al. (2006).

The authors had proposed an efficient method for pose invariant face recognition using only one sample view per person (Ng 2006). Our method is based on the observation that the resemblance between the faces of two individuals should be rather consistent across different viewing directions. That is, if two faces look alike in the frontal view, they should also look alike in the 45 degree view. Similarly, if two faces look different in one view, they should also look different in other views. This is not always true, because the human face is three dimensional; if the 3D shapes of two faces are significantly different, the resemblance between them might not be consistent across different views. In the majority of cases, however, the above observation should be valid. In addition, facial features such as eyes and mouth can usually generalize better to new views than the face in its entirety. Therefore, a similarity indicator based on the resemblance of a face and its facial features against a large enough set of other faces (prototype faces) represents a robust pose invariant measure that can be used for comparing face images of a person taken in different poses.

Given a new face image, we compute the similarities between the image (whole face and facial features) and a set of prototype face images taken at similar or close to similar view. These similarity values form a similarity vector that is a pose invariant representation for the new face image. Given another image of the same face in a different view, we generate another similarity vector for the image against the face images in the prototype set taken in the new view. Pose invariant recognition is achieved by comparing the two similarity vectors, which should be highly correlated if they are coming from the same face. The proposed approach requires only one sample view of a person to be recognized. Multiple views are needed only for the faces in the prototype set.

In this paper, we extend the similarity vector concept to handle the variations due to both pose and facial expression. The similarities of a face to be recognized against a set of prototype faces of similar view and facial expression are used to form the pose and expression invariant similarity vector for the new face, which is later used for recognizing the same face taken at different poses with different facial expressions. The rest of the paper is organized as follows. Section 2 describes the algorithms for constructing pose and expression invariant similarity vector for a face image and how the similarity vectors are used for face recognition. Section 3 presents the experiments of applying the proposed method to the CMU PIE face database (Sim et al. 2003) and provides discussion of the results. Section 4 contains concluding remarks and discussions about the future directions.

## 2. Similarity Vector

The algorithm for constructing and using pose and expression invariant similarity vector consists of the following steps. First, the face images in the prototype set are grouped according to their views. Each group consists of images with several different facial expressions for every prototype face taken at the same pose. The images are used to build the view-specific eigenspace representation for each view using principle component analysis (PCA). Second, the similarity vector for an input face image is constructed by projecting the image into the eigenspace of the appropriate view and by computing the minimum distances of the input image to each of the prototype faces in the eigenspace. Finally, the similarity vector of the input image is compared to the pre-stored similarity vectors of the sample images using normalized correlation.

### 2.1 View-specific eigenspace representation

Eigenspace representation has been extensively applied for the task of face recognition (Turk and Pentland 1991; Pentland et al. 1994; Ruiz-del-Solar and Navarrete 2005). Eigenspace-based methods project input faces into a dimensional reduced subspace and the distance in the subspace is used as similarity function for recognition. Eigenspace representation is created using PCA.

The face images in the prototype set are first grouped according to their views. Each group consists of images with several different facial expressions for every prototype face taken at the same pose. Assuming there are $p$ prototype faces and each face has $m$ expressions, the total number of images in a group will be $p \times m$. The images of the same view are used to build a view-specific eigenspace representation. For $n$ views, there are $n$ sets of view-specific eigenspace representations for the prototype images, each capturing the variation of the faces in a common view. This is similar to the view-based eigenspace approach (Pentland et al. 1994).

Assuming the size of the prototype face images to be $M \times N$, these images can be represented as a vector of dimension $MN$, by scanning the images left to right and top to bottom. Let $\mathbf{X}_1$, $\mathbf{X}_2$,…, $\mathbf{X}_{p \times m}$ be the $p \times m$ prototype images in a group, the covariance matrix $\mathbf{C}$ of the prototype images is obtained as

$$\mathbf{C} = \frac{1}{p \times m} \sum_{i=1}^{p \times m} (\mathbf{X}_i - \mathbf{A})(\mathbf{X}_i - \mathbf{A})^T \qquad (1)$$

where $\mathbf{A}$ is the average of the prototype images. Let $\mathbf{E}_j$ and $\lambda_j$ represent the eigenvectors and eigenvalues computed from the covariance matrix. We can obtain the optimal approximation of an image by selecting the $k$ ($k \leqq p \times m$) most significant eigenvectors with the largest corresponding eigenvalues and representing each image by a linear combination of the major $k$ eigenvectors as

$$\hat{\mathbf{X}}_i = \mathbf{A} + \sum_{j=1}^{k} w_{ij} \mathbf{E}_j \qquad (2)$$

where

$$\mathbf{W}_i = [w_{ij}] = [\mathbf{E}_1, \mathbf{E}_2, \cdots, \mathbf{E}_k]^T (\mathbf{X}_i - \mathbf{A}) \qquad (3)$$

The vector subspace formed by the $k$ eigenvectors is referred as the eigenspace. $\mathbf{W}_i$ is the projection of image $i$ in the subspace, and it represents the coordinates of the image in the eigenspace. The cumulative proportion $\mu^{(k)}$ is useful for determining the number of eigenvectors. A value of 0.95 was used throughout this paper.

$$\mu^{(k)} = \sum_{i=1}^{k} \lambda_i \Big/ \sum_{i=1}^{p \times m} \lambda_i \qquad (4)$$

As mentioned before, to reduce the effect of the 3D nature of the human face in the similarity measure between two faces, we use both the global (whole face) and local (eyes, nose and mouth) features for similarity measurement. Therefore, for each

view, there are three eigenspaces generated: one for the face, one for the eyes, and one for the nose and mouth. See Section 3.1 for the extraction of face and facial features.

2.2 Similarity vector

The similarity vector for an input face represents the similarities between the input face and the faces in the prototype set with similar pose and facial expression. To obtain the similarity vector for an input face, we project the input face image into the appropriate view-specific eigenspace and compute the Euclidean distance between the input image and the prototype images in the eigenspace. Note that each prototype face has *m* images of different facial expressions, thus the similarity between an input face and a prototype face of similar facial expression is the minimum distance among the *m* images to the input face image.

Since there are three eigenspaces (face, eyes, nose and mouth) associated with each view, the similarity values are computed separately for each eigenspace and the results are cascaded to form the similarity vector. The length of the similarity vector is equal to three times the number of faces in the prototype set.

$$\mathbf{S} = [d_1, d_2, \cdots, d_m]^T \quad , \quad m = 3p \qquad (5)$$

where **S** is the similarity vector. $d_i$ is the minimum Euclidean distance between input face image and the images of a prototype face *i* in the eigenspaces:

$$d_i = \sqrt{\sum_{j=1}^{k} (v_j - w_{ij})^2} \qquad (6)$$

where $[v_j]$ denotes the projection of input image in the eigenspace. The distance values indicate how similar the input face is to each of the faces of the same view and expression in the prototype set.

For an input image, we determine the view for the image by selecting the eigenspace which best describes the input image. This is accomplished by finding the eigenspace that produces minimum reconstruction error (referred to as "distance-from-face-space" by Turk and Pentland (1991)). The reconstruction error (*err*) is the difference between the input image **Y** and its approximation $\hat{\mathbf{Y}}$ defined as.

$$err = diff(\mathbf{Y}, \hat{\mathbf{Y}}) = \sqrt{\sum_{i=1}^{MN} (y_i - \hat{y}_i)^2} \qquad (7)$$

where $\hat{\mathbf{Y}}$ is obtained from the eigenspace's eigenvectors as defined in Eq. (2).

2.3 Face recognition using similarity vector

During the learning phase, the similarity vectors for each sample face of the persons to be recognized are constructed and stored using the procedure described in Section 2.2. The pose and facial expression of the sample faces can be arbitrary, as long as they are in the range covered by the prototype set.

During the recognition phase, the similarity vector of an input image is computed using the same procedure and compared to the pre-stored similarity vectors of the sample images using normalized correlation. Normalized correlation between two similarity vectors $\mathbf{S}_a$ and $\mathbf{S}_b$ is computed as

$$NC(\mathbf{S}_a, \mathbf{S}_b) = \left( \frac{\sum_{i=1}^{3p} (d_{ai} - \bar{d}_a)(d_{bi} - \bar{d}_b)}{\sqrt{\sum_{i=1}^{3p} (d_{ai} - \bar{d}_a)^2 \times \sum_{i=1}^{3p} (d_{bi} - \bar{d}_b)^2}} + 1 \right) \times \frac{1}{2} \qquad (8)$$

where $\bar{d}_a$ and $\bar{d}_b$ denotes the averages of the elements in $\mathbf{S}_a$ and $\mathbf{S}_b$ respectively. The correlation values range from 0 ~ 1, a high value indicates that the two vectors are correlated.

A correct match is declared if the highest score is larger than a predefined threshold value and comes from the same person as in the input image.

## 3. Experiments

The performance of the proposed method was evaluated using the CMU PIE face database (Sim et al. 2003). The CMU PIE face database contains large number of face images with different sources of pose, illumination, and expression variations from 68 individuals. For our purposes, we selected images with three different facial expressions (neutral, smile, blink) from five different pose variations under environment light. Fig. 1 shows sample images of an individual taken at the five camera positions with three facial expressions. There are 15 images per individual, and a total of 1020 images are used in the experiments.

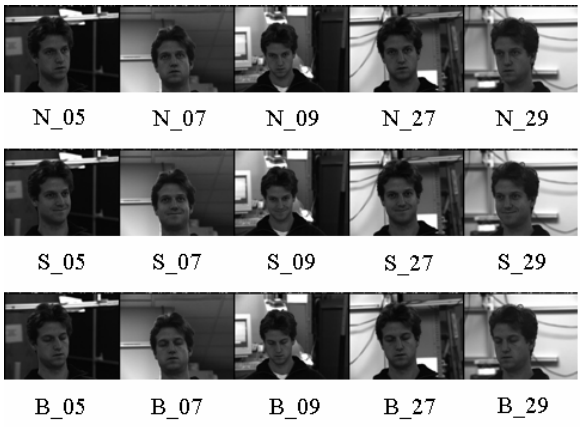


Fig. 1 Sample images from the CMU PIE face database (Sim et al. 2003) taken at five different camera positions and three facial expressions.

A complete face recognition solution involves segmentation of faces (face detection) from the background, facial feature extraction from the face region, and recognition. The purpose of the experiments performed here was to evaluate the performance of our recognition method. Faces and facial features were extracted manually as a preprocessing step (section 3.1) in all experiments. Refer to Yang et al. (2002) for a review of methods for localizing faces and facial features.

### 3.1 Preprocessing and normalization

Before constructing the eigenspaces for the prototype images, a series of preprocessing steps were applied to the face images. First we localized the face in an image by manually locating the centers of both eyes and the mouth. Second, face alignment was performed by centering the eyes and the mouth in the same relative positions. Face alignment is critical for good recognition performance (Martinez 2002). Third, the face region was extracted and resized to 60×70 pixels. Finally, histogram equalization was performed on the extracted image and the image was normalized to unit energy.

After the face region is extracted and normalized, the region of the eyes (50×16 pixels) and the nose and mouth (30×40 pixels) are extracted from the face. Fig. 2 shows examples of preprocessing and facial feature extraction results, before energy normalization.

(a) Image N_27

(b) Image N_05

Fig. 2 Examples of preprocessing and facial feature extraction. The left column contains the original images; the right column contains the processed face and facial feature regions.

3.2 Experiments and results

From the CMU PIE face database, we randomly selected *p* individuals to form the prototype set. The prototype images were used to generate view-specific eigenspaces, one for each view. The optimal value for *p* was determined experimentally, as discussed below.

In the learning phase, images from the frontal view with neutral expression (image N_27, see Fig. 1) were selected as sample images. The similarity vectors for each of the sample images were computed and stored. In the recognition phase, the similarity vector of a test image (image with pose and facial expression different from the sample image) was computed and correlated to the similarity vectors from the sample image. If the highest score came from the same person as in the test image, a correct recognition was declared.

We performed tests for the five poses and three facial expressions other than the frontal view and neutral expression as shown in Fig. 1. All 68 individuals in the database were used for testing. In each test, we varied the number of prototype images for generating the eigenspaces to determine how the number of prototype images affects the performance of the proposed method. Table 1 to 3 shows the recognition rates of the proposed method under different poses, facial expression, and different numbers of prototype images. The recognition rates using virtual faces method from Jiang et al. (2005) are also shown in the bottom row of Table 1 for comparison.

As shown in Table 1, for pose only variations, the recognition rates of the proposed method are over 97% when the number of prototype images is 30 or more. The recognition rate increases as the number of prototype images increases. The recognition accuracy of the proposed method is significantly better than the virtual face method.

For both pose and facial expression variations, the recognition rates of the proposed method are over 90% for all three facial expressions (Table 1 to 3) when the number of prototype images is 50 or more. The recognition rates drop about 10% when the number of prototype images is reduced to 30. One explanation for this is that when the variations are large, a low number of prototype images (low dimensionality of similarity vector) might not have enough discriminative power to capture the variation in appearance due to large pose and expression changes. The experimental results show that the similarity vector approach performs well on handling variations due to changes in pose and facial expression.

Table 1 Recognition rates for the proposed method and for virtual faces method under different poses and neutral expression.

| | Test Image Label | | | |
|---|---|---|---|---|
| Number of Prototype Images | N_05 | N_07 | N_09 | N_29 |
| 30 | 100 | 99 | 97 | 99 |
| 40 | 100 | 100 | 100 | 100 |
| 50 | 100 | 100 | 100 | 100 |
| 60 | 100 | 100 | 100 | 100 |
| Virtual Face* | 85 | 92 | 93 | 68 |

* Jiang et al. (2005)

Table 2 Recognition rates for the proposed method under different poses and smile expression.

| | Test Image Label | | | | |
|---|---|---|---|---|---|
| Number of Prototype Images | S_05 | S_07 | S_09 | S_27 | S_29 |
| 30 | 88 | 81 | 88 | 91 | 88 |
| 40 | 88 | 88 | 87 | 91 | 88 |
| 50 | 94 | 91 | 94 | 99 | 93 |
| 60 | 94 | 90 | 94 | 99 | 96 |

Table 3 Recognition rates for the proposed method under different poses and blink expression.

| | Test Image Label | | | | |
|---|---|---|---|---|---|
| Number of Prototype Images | B_05 | B_07 | B_09 | B_27 | B_29 |
| 30 | 85 | 79 | 91 | 85 | 82 |
| 40 | 90 | 82 | 93 | 90 | 91 |
| 50 | 91 | 90 | 96 | 93 | 93 |
| 60 | 96 | 94 | 96 | 96 | 97 |

## 4. Conclusions

We have presented an efficient method for pose and expression invariant face recognition which requires only one sample person. Our method uses the similarities between a face and a set of prototype faces of similar view and facial expression to establish a pose and expression invariant similarity vector which can be used for comparing faces taken in different poses and facial expressions. Experiments using the CMU PIE face database have shown that the proposed method can achieve high recognition rate for significant pose and expression variations. Compared to the virtual faces method, the proposed method is relatively simple and fast. Adding a new face to the face database is simply by creating a similarity vector for the new face, no retraining is required.

For future work on the method, we plan to test the performance of the method on larger face databases and to test how well the method handles interpolation and extrapolation between views. We will also attempt to determine the optimal number of prototype images, and to investigate the potential of using the method to handle other sources of variation such as illumination

## 5. References


[1] D. Beymer and T. Poggio, "Face recognition from a single example view", *Proceedings of the 5th International Conference on Computer Vision*, pp. 500-507, 1995.

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

[3] S. Chen and B. Lovell, "Illumination and expression invariant face recognition with one sample image", Proceedings of the 17th International Conference on Pattern Recognition, Vol. 1, pp. 300-303, 2004.

[4] D. Jiang, Y. Hu, S. Yan, L. Zhang, H. Zhang, and W. Gao, "Efficient 3D reconstruction for face recognition", *Pattern Recognition*, Vol. 38, No. 6, pp. 787-798, 2005.

[5] M. Lando and S. Edelman, "Generalization from a single view in face recognition", *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, pp. 80-85, 1995.

[6] X. Li, G. Mori, and Z. Hao, "Expression-invariant face recognition with expression classification", *The 3rd Canadian Conference on Computer and Robot Vision*, pp. 77, 2006.

[7] A. Martinez, "Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 6, pp. 748-763, 2002.

[8] H. Murase and S. Nayar, "Visual learning and recognition of 3-D objects from appearance", *International Journal of Computer Vision*, Vol. 14, pp. 5-24, 1995.

[9] H. Ng, "View-invariant face recognition from a single sample", *Proceedings of the 10th World Multiconference on*

*Systemics, Cybernetics and Informatics*, Vol. 5, pp. 212-216, 2006.

[10] A. Pentland, B. Moghaddam, and T. Starner, “View-based and moduler eigenspaces for face recognition”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 84-91, 1994.

[11] J. Ruiz-del-Solar and P. Navarrete, “Eigenspace-based face recognition: a comparative study of different approaches”, *IEEE Transactions on Systems, Man and Cybernetics*, Part C, Vol. 35, No. 3, pp. 315-325, 2005.

[12] T. Sim, S. Baker, and M. Bsat, “The CMU pose, illumination, and expression database”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 12, pp. 1615-1618, 2003.

[13] X. Tan, S. Chen, Z. Zhou, and F. Zhang, “Face recognition from a single image per person: A survey”, *Pattern Recognition*, Vol. 39, No. 9, pp. 1725-1745, 2006.

[14] M. Turk and A. Pentland, “Eigenfaces for recognition”, *Journal of Cognitive Neurosicence*, Vol. 3, No. 1, pp. 71-86, 1991.

[15] J. Wang, K. Plataniotis, J. Lu, and A. Venetsanopoulos, “On solving the face recognition problem with one training sample per subject”, *Pattern Recognition*, Vol. 39, No. 9, pp. 1746-1762, 2006.

[16] M. Yang, D. Kriegman, and N. Ahuja, “Detecting faces in images: A survey”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 1, pp. 34-58, 2002.

[17] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, “Face recognition: a literature survey”, *ACM Computing Surveys*, pp. 399-458, 2003.

## 由單一樣本影像做不受姿勢與表情影響的人臉辨識

### 摘要

在不同姿勢與表情變化的情況下做人臉辨識是一個挑戰性的問題。在這篇文章中，我們提出一個由單一樣本影像做不受姿勢與表情影響的人臉辨識方法。此方法利用一個人臉與一組在相同拍攝姿勢與表情的樣本人臉之相似度來建立一個不受拍攝姿勢與表情影響的相似度向量，並利用此相似度向量來比對從不同拍攝姿勢與表情得來的人臉影像。實驗結果指出，即使在較大的姿勢與表情變化的情況下，本方法依然可達到相當高的辨識率。

關鍵字：人臉辨識、不受姿勢影響、不受表情影響、主成份分析、相似度向量