Face Recognition Using Invariance with a Single Training Sample

Qian Tian
National ASIC Engineering Research Center, Southeast University, Nanjing, China
e-mail: tianqian@seu.edu.cn

Abstract

For the limits of memories and computing performance of current intelligent terminals, it is necessary to develop some strategies which can keep the balance of the accuracy and running time for face recognition. The purpose of the work in this paper is to find the invariant features of facial images and represent each subject with only one training sample for face recognition. We propose a two-layer hierarchical model, called invariance model, and its corresponding algorithms to keep the balance of accuracy, storage and running time. Especially, we take advantages of wavelet transformations and invariant moments to obtain the key features as well as reduce dimensions of feature data based on the cognitive rules of human brains. Furthermore, we improve usual pooling methods, e.g. max pooling and average pooling, and propose the weighted pooling method to reduce dimensions with no effect on accuracy, which let storage requirement and recognition time greatly decrease. The simulation results show that the proposed method does better than some typical and nearly-proposed algorithms in balancing the accuracy and running time.

***Keywords**: invariance model, single training sample, face recognition*

1. Introduction

In recent years, face recognition technique has been applied on intelligent mobile terminals for varieties of applications. However, the terminals have some resource limits such as low power, small memories and low computing performance. In this case, the small training base and simple algorithms are preferred. So it is necessary to study strategies for face recognition with one training sample to keep a balance of the accuracy and running time. Actually, there are some fruits [1]-[5] in academics. The improved discriminative multi-manifold analysis (DMMA) method has been proposed in [1] by partitioning each face image into several non-overlapping patches to form an image set for each person. Although the recognition accuracy of DMMA [1] was higher than 80 percents for AR database, its running time on PC is as 100 times as typical methods such as PCA due to its complexity. Thus, one problem is addressed that whether an invariant factor we can acquire for face recognition using simple methods to perform less running time and an acceptable accuracy. This paper is trying to solve this problem.

The idea in this paper is inspired by the invariance in the visual recognition [6]-[8], which imitates the tuning properties of view-tuned cells in infero-temporal (IT) cortex [9],[10]. One of the corresponding famous models is Hierarchical Model and X (HMAX) model proposed by Poggio and his research group [11],[12]. HMAX was a four-layer hierarchical model, in which two types of computations: linear summation and non-linear max operation alternated between layers. HMAX model was first proposed for object recognition, not spatially for face recognition. The research group of Poggio later proposed the method for face recognition based on HMAX, which described cognitive characteristics in accuracy as a face descriptor [6]-[8]. This method is simulated on PC, not considering the operating platform, especially for the terminals without enough resources such as storages and computing abilities. The HMAX based methods mainly make best of Gabor filters for feature extraction, while Gabor filters could not get enough local details that represent the distinct features of different faces. Thus, we have to find a kind of wavelet filters instead of Gabor filters to get enough details on face features at different scales. To improve the accuracy using one training sample, the method of patch segments is also used for feature extraction. Furthermore, to reduce the quantities of data and still keep the invariance of the facial images, we take advantages of invariant moments [13],[14] and pooling techniques

Received August 17, 2014; Revised October 14, 2014; Accepted November 4, 2014

to keep the geometric invariance, such as rotation, shift and zooming, of the images as well as reducing the data storage.Based on these considerations, we propose an invariance model and its corresponding algorithms for face recognition using one training sample in conditions of low storage and low computing performance platform.

This paper is organized as five sections. Section II introduces some related work including HMAX model and moment invariants. Section III illustrates the proposed invariance model and the corresponding algorithm in details. The simulation results based on public face databases are discussed in section IV. Section V makes a conclusion and discusses the further work.

2. Related Work

2.1. HMAX Model

There are two simple units and two complex units in HMAX model. The schematic of HMAX is shown in Figure 1.

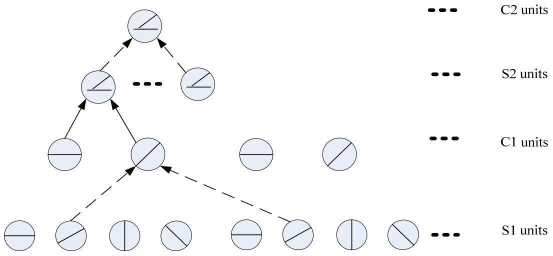


Figure 1. Schematic of HMAX model

In Figure 1, S1 and S2 are simple units, and C1 and C2 are complex units. At S1, arrays of two-dimensional filters, generally Gabor filters, at four different orientations are used for the input raw images. Gabor filters are tightly tuned for both orientation and frequency but cover a wide range of spatial frequencies. Then, the groups of cells at same orientation but at a slightly different scales and positions are fed from S1 into C1 using MAX operation [11]. At S2,within each filter band, a square of four adjacent and non-overlapping C1 cells in 2×2 arrangement is grouped as cells, which are again fed from S2 into C2, finally achieving size invariance over all filter sizes in the four filter bands and position invariance over the whole input images.

2.2. Moment Invariants

Feature moment as a global invariant is often used for the feature selection to reduce the input for classifiers. The representations of seven moment invariants based on the second-order and third-order normalized center moments are given by [7]. For a two-dimensional M× N image function f(x,y), the definitions of p+q order geometry moment mpq and the center moment μ_{pq} are given by (1) and (2) .

$$m_{pq} = \sum_{i=1}^{M}\sum_{j=1}^{N} i^p j^q f(i, j) \tag{1}$$

$$\mu_{pq} = \sum_{i=1}^{M}\sum_{j=1}^{N} (i-\overline{x})^p (j-\overline{y})^q f(i, j) \tag{2}$$

Here, $\overline{y}=m_{01}/m_{00}$ $\overline{x}=m_{10}/m_{00}$ is the gravity center on the direction of x in the image, is the gravity center on the direction of y in the image. Zero-order center momentμ_{pq} is used to normalize the moments. So the normalized center moment is given by (3).

$$\eta_{pq}=\frac{\mu_{pq}}{\mu_{00}^{r}} \tag{3}$$

Then, the seven moment invariants are given by (4). The definition of the seven invariants are given in (4) for the second and third order moments. The seven invariants are useful for not only pattern identification independently of position, size and orientation but also independently of parallel projection.

$$\begin{aligned}
&\varphi_1=\eta_{20}+\eta_{02}\\
&\varphi_2=(\eta_{20}-\eta_{02})^2+4\eta_{11}^2\\
&\varphi_3=(\eta_{30}-3\eta_{12})^2+(3\eta_{21}-\eta_{03})^2\\
&\varphi_4=(\eta_{30}+\eta_{12})^2+(\eta_{21}+\eta_{03})^2\\
&\varphi_5=(\eta_{30}-3\eta_{12})(\eta_{30}+\eta_{12})[(\eta_{30}+\eta_{12})^2-3(\eta_{21}+\eta_{03})^2]\\
&+(3\eta_{21}-\eta_{03})(\eta_{21}+\eta_{03})[3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2]\\
&\varphi_6=(\eta_{20}-\eta_{02})[(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2]\\
&+4\eta_{11}(\eta_{30}+\eta_{12})(\eta_{21}+\eta_{03})\\
&\varphi_7=(3\eta_{21}-\eta_{03})(\eta_{30}+\eta_{12})[(\eta_{30}+\eta_{12})^2-3(\eta_{21}+\eta_{03})^2]\\
&-(\eta_{30}-3\eta_{12})(\eta_{21}+\eta_{03})[3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2]
\end{aligned} \tag{4}$$

3. Invariance Model and Proposed Algorithms

As the illustration in section II, Gabor filters are generally used for feature extraction at the first layer in HMAX model. Gabor transformation could simulate the human visual system by decomposing the retinal image into a set of filter images. And the each filter image can represent changes in the intensities of the frequency and direction in the local scope of the raw image. So the texture features can be obtained by a group of the multi-channel Gabor filters. Equation (5) shows the Gabor filter, the key parameters are the frequency functions and the window size of Gauss function. Actually, Gabor transformation uses Gauss function as a window function to make local Fourier transformation by choosing the frequency and Gauss parameters. Although Gabor filters have tempo-spatial characteristics, they are not orthogonal so that different feature components have redundancies which lead to the low efficiency for texture feature extraction.

$$\begin{gathered}
G(\mathrm{x},\mathrm{y},\theta,\mathrm{f})=\exp(-\frac{1}{2}((\frac{x'}{sx})^2+(\frac{\mathrm{y}'}{sy})^2))\sin(2\pi\,\mathrm{fx}')\\
\mathrm{x}'=\mathrm{x}\cos\theta-\mathrm{y}\sin\theta\\
\mathrm{y}'=\mathrm{y}\cos\theta+\mathrm{x}\sin\theta
\end{gathered} \tag{5}$$

Here, sx and sy are the variances along x and y axes respectively, and f is the frequency of the sinusoidal function, θ is the orientation of the Gabor filter.

We randomly select a facial image from ORL database shown in Figure 2. Figure3 shows the filter images of Figure 2 using (5). The images in Figure3 are normalized for easy observation. The upper row images in Figure 3 are the four filter images at one scale and four orientations. The lower row images are the four filter images at another scale and four orientations. Figure 3 shows that there are very slightly variances of the eight images and the more local details are not enough. On the other hand, for the same raw image, we make wavelet transformation at four scales using the wavelet 'db5'. Figure 4 shows the result, in which the upper left image, the upper right image, the lower left image and the lower right image are the detail images at scale four, scale three, scale two and scale one respectively. Comparing Figure 3 and Figure 4, wavelet transformation can obtain more details and local features than Gabor transformations. Actually, we let the most left image of the upper row in Figure 3 subtract the other seven images of Figure 3 respectively to get seven difference matrices, and then calculate the corresponding standard variance of the seven matrices. Similarly, we conduct the same operation on Figure 4. The computation results are shown in Figure 5. Obviously, the standard deviation values of the filtered images using wavelets are much bigger than that using Gabor. The standard deviation averages using wavelets and Gabor are 5.2420 and 0.7182 respectively.

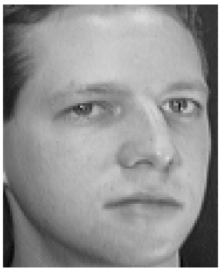


Figure 2. The raw image in ORL database

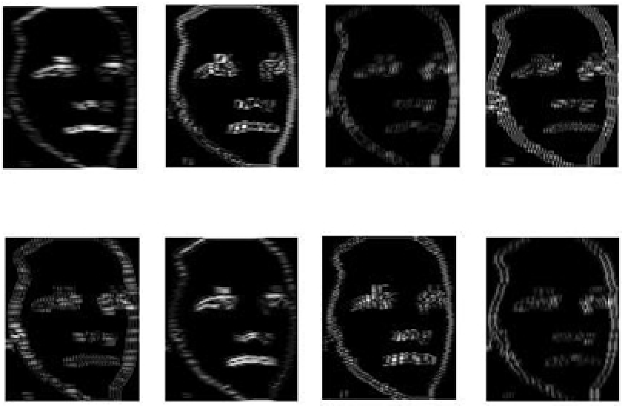


Figure 3. The upper row images are the results using Gabor filters with the window size of (2, 1), and the orientation of $2\pi/12, 5\pi/12, 8\pi/12\ and\ 11\pi/12$ respectively. The lower row images are the results using Gabor filters with the window size of (4, 2), and the orientation of $2\pi/12, 5\pi/12, 8\pi/12, 11\pi/12$respectively

Therefore, we chose wavelet transformation instead of Gabor filters to get more local features. The typical orthogonal wavelet 'db5' is used as the kernel to make wavelet transformations. For each image, the wavelet transformation is conducted at four scales to get the corresponding approximation coefficients whose matrix is named A0 at the fourth scale and the detail coefficients whose matrices are named D1, D2, D3, and D4 at four different scales respectively. Then, A0, D1, D2, D3 and D4 are respectively singly used to make reconstruction so that to generate the reconstructed five images, e.g. Figure 4.

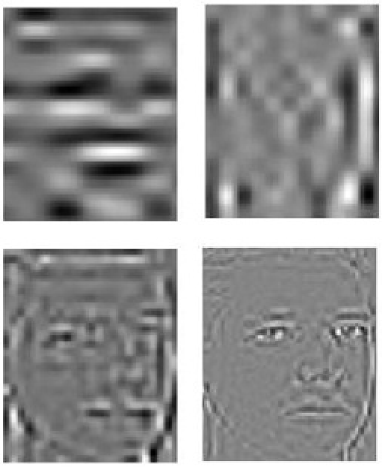


Figure 4. The reconstruction images using the detail coefficients of the wavelet transformation at four scales respectively

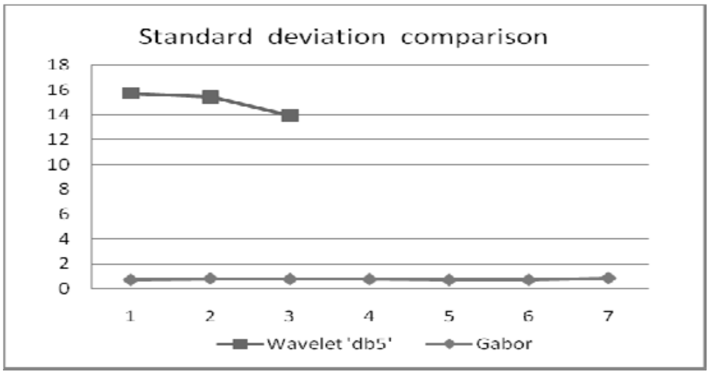


Figure 5. The standard deviation comparison using wavelets transformation and Gabor transformation

After wavelet transformations and reconstructions, one image is decomposed to generate five images at different scales so that the quantities of data increased by five times. To solve this problem, the technique of dimension reduction has to be considered. Although PCA is a typical method of dimension reduction, it operates on the global features without considering the local details. While in case of only one training sample for each face, the local details play a very important role.

To keep local features as well as reduce data dimensions, we first divide the reconstructed images into patches. Combining the cognitive law and the facial image characteristics, we divide the facial image into 9 patches shown in Figure 6. Each patch just represents one physiological region of the face. Patch1 is a left forehead, patch2 is a middle forehead, patch3 is a right forehead, patch4 is a left cheek as well as a left eye, patch5 is a nose, patch6 is a right cheek as well as a right eye, patch7 is a left chin, patch8 is a lip, and patch9 is a right chin.

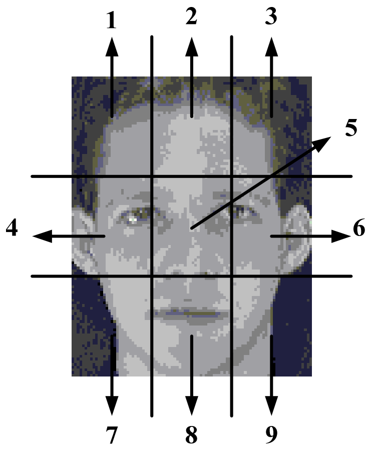


Figure 6. Patch segments of the facial image

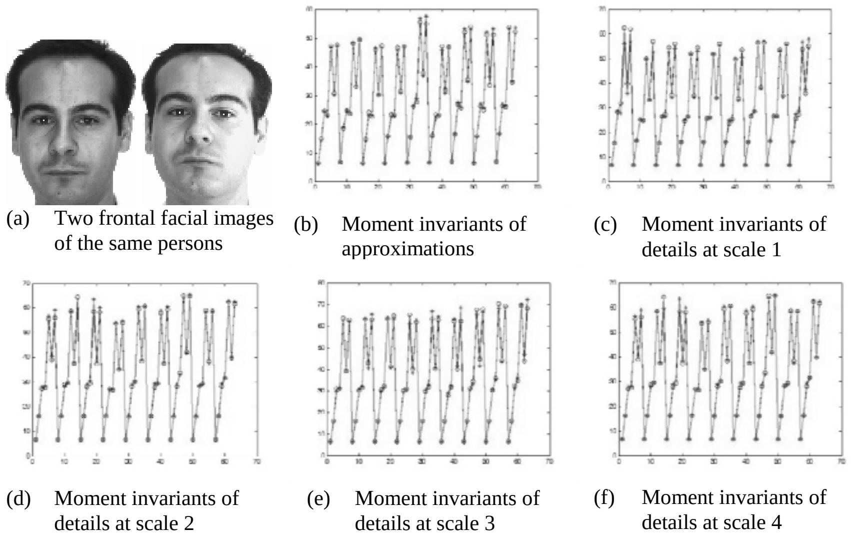
Among so much feature data, the key features have to be preserved to represent the characteristics of each facial image. Because even if the images of the same face, they still have slightly variance in different conditions such as illumination, the invariance of features become more and more important. Thus, the technique of moment invariants is considered to preserve the invariance of feature data due to its invariance in scale, zoom and rotation. The details of moment invariants are introduced in section II.

For each reconstructed image, we first divide it into 9 patches, and then compute the seven moment invariants of each patch, and finally one column of 63 data is obtained. The corresponding algorithm is summarized in Table I.

Table 1. Algorithm 1(A.1):Feature Extraction

A.1 Feature extraction
Step 1. Wavelet decomposition For each training facial image, wavelet transformation is made at four scales to get approximation coefficients named A0, and detail coefficients named D1, D2, D3, D4 at four scales.
Step 2. Image reconstruction. We make wavelet reconstruction respectively using A0, D1, D2, D3 and D4 to generate five reconstructed images (e.g. Figure5).
Step 3. Patch segments For each reconstructed image, we divide it into 9 patches (e.g.Figure6).
Step 4. Moment invariants We calculate seven moments for each patch so that totally 9 groups of seven moments are obtained for each reconstructed image. Let the 9 groups of data be one column vector with the size of 63 data. Thus, one vector of 63 data represents the features of one reconstructed image (e.g.Figure7 and Figure8).
Step 5. Feature extraction According to the results of step 4, there are 63 feature data for one reconstructed image. Because one facial image is corresponding to five reconstructed images, there are totally 315 feature data for one facial image.

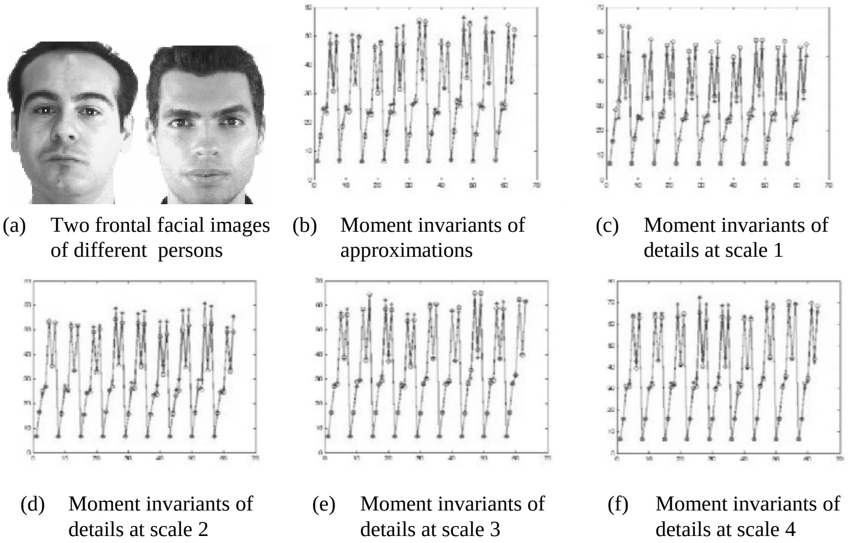
We randomly chose some images shown in Figure 7 and Figure 8 to illustrate feature extraction. Figure 7 shows that the moment invariants of the two faces of the same person have only slight differences, especially, among the seven moment invariants referring to equation (4), φ_1 and φ_2 are almost the same. While Figure 8 shows the moment invariants of the two different facial images. Obviously, there are differences between the circle lines and the cross lines, especially in the reconstructed images at scale 1, 2 and 3(see Figure 8(c),(d),(e)).



(a) Two frontal facial images of the same persons (b) Moment invariants of approximations (c) Moment invariants of details at scale 1

(d) Moment invariants of details at scale 2 (e) Moment invariants of details at scale 3 (f) Moment invariants of details at scale 4

Figure 7. Comparison of wavelet reconstructed images between two frontal facial images of the same person. In from (b) to (f), the circle lines represent moment invariants of reconstructed images of the left face, and the cross lines represents moment invariants of the reconstructed images of the right face



(a) Two frontal facial images of different persons (b) Moment invariants of approximations (c) Moment invariants of details at scale 1

(d) Moment invariants of details at scale 2 (e) Moment invariants of details at scale 3 (f) Moment invariants of details at scale 4

Figure 8. Comparison of wavelet reconstructed images between two frontal facial images of the same person. In from (b) to (f), the circle lines represent moment invariants of reconstructed images of the left face, and the cross lines represent moment invariants of the reconstructed images of the right face

To reduce the dimensions of feature data furtherly, we also proposed a weighted average pooling method. Observing the patches in Figure 6 and combing the cognitive rules of human brains, we suppose patch 4, 5, 6 and 8 have more contributions than other patches for representing one face, so we distribute comparatively bigger weights named 'w4','w5','w6','w8' to patch 4, 5, 6 and 8 than those named 'w1','w2','w3','w7','w9' to patch 1,2,3,7,9. All of the nine weights are experiencing values from many experiments, and the sum of all the weights is equal to 1. The details are shown in Table II. Using the weighted average pooling method, each reconstructed image has seven feature data so that each train sample has 35 feature data. The dimensions of training data are greatly reduced. That is, the storage requirement decreases which is a very important advantage for embedded device, e.g. sensor nodes of small size and low cost.

Table 2 Algorithm 2(A.2) Dimension reduction

```
A.2 Dimension reduction
Step 1: Weights distribution
Distributing '$w_1$','$w_2$''$w_3$','$w_4$','$w_5$','$w_6$','$w_7$','$w_8$','$w_9$' to patch 1,2,3,4,5,6,7,8,9 respectiv
sum of the nine weights is 1.
Step2 :Weighted average pooling
Supposing the moment invariants vectors of the nine patches are respectively
$v_1$,$v_2$,$v_3$,$v_4$,$v_5$,$v_6$,$v_7$,$v_8$, and $v_9$, in sequence, then the feature vector of each reconst
is
$^{\cdot} v = \sum_{i=1}^{9} v_i * w_i$
```

In a summary, an invariance model of two layers is constructed. Figure9 shows the invariance model structure composed of invariance attributes (IA) layer and invariance cluster (IC) layer. In IA, wavelet transformations are conducted at four scales resulting into four-scale detail coefficients and approximation coefficients, which are respectively used to generate new five reconstructed images. To keep the invariance and reduce quantities of data, we take advantages of the technique of patch segments and invariant moments to keep the global invariance of each reconstructed images. The corresponding algorithm of AI is summarized in A.1. On the base of AI, feature cluster has to be conducted to reduce dimensions as well as keep feature invariance. So in IC, the improved weighted pooling technique is used to select the key features. The corresponding algorithm of IC is summarized in A.2. Finally, a vector of 35 feature data represents the raw image.

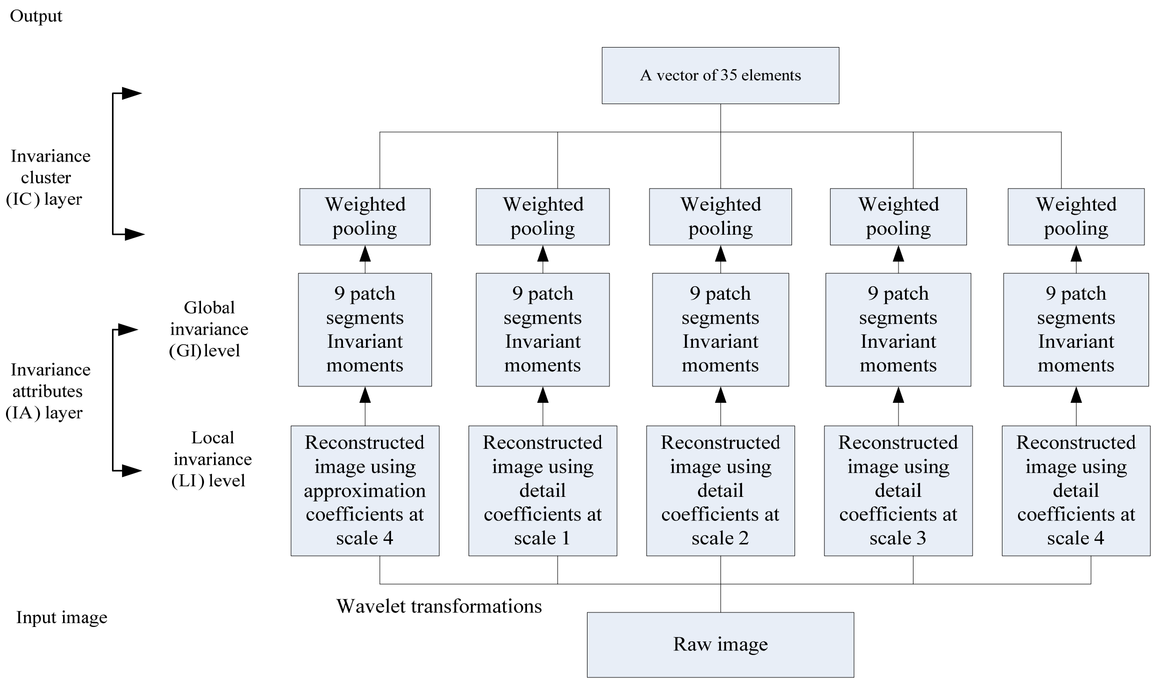

Output
A vector of 35 elements
Invariance cluster (IC) layer
Weighted pooling
Weighted pooling
Weighted pooling
Weighted pooling
Weighted pooling
Global invariance (GI) level
9 patch segments Invariant moments
9 patch segments Invariant moments
9 patch segments Invariant moments
9 patch segments Invariant moments
9 patch segments Invariant moments
Invariance attributes (IA) layer
Local invariance (LI) level
Reconstructed image using approximation coefficients at scale 4
Reconstructed image using detail coefficients at scale 1
Reconstructed image using detail coefficients at scale 2
Reconstructed image using detail coefficients at scale 3
Reconstructed image using detail coefficients at scale 4
Wavelet transformations
Input image
Raw image

Figure 9. The invariance model structure

4. Simulations

Three public face databases: ORL database, AR database and FERET database are used for simulations [15]. The details of selected facial images are shown in Table II. The simulation is running on PC with 3.3GHz CPU and 4G RAM.

Table 3. Characteristics of selected images in ORL, AR and FERET

Name	No. of people	No. of pictures per person	conditions	size per image
ORL	40	10	frontal and slight tilt;light expressions; variable light	112×92
AR	120	26	Frontal view with different expressions; variable illuminations and occlusions	80×100
FERET	200	7	Multiple pose, different time face for per person	80×80

We first chose the proposed method, called IM short for invariance model, including A.1 and A.2, to make feature extraction and dimension reduction, and then chose KNN and LDA, the simple classifiers to make classification. To evaluate the proposed method, after step 2 in A.1, we use PCA, the typical dimension reduction technique to make dimension reduction instead of moment invariants, and then chose the same classifiers to make classification. Figure10, Figure11, Figure12 and Figure13 show the simulation results of different methods using one training sample respectively. Because the proposed method is not considering the occlusion, the facial images with occlusion in AR database are not used for the simulation. Table IV shows the recognition time for one test image supposing the training data have been ready.

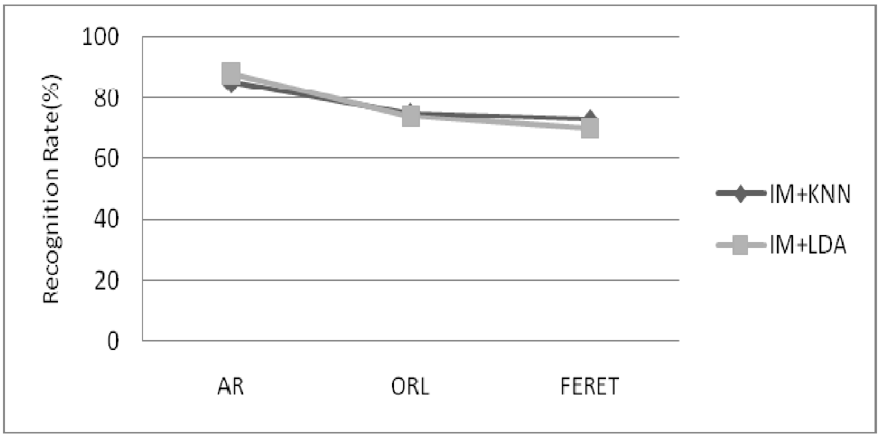


Figure 10. Recognition rates using the proposed method

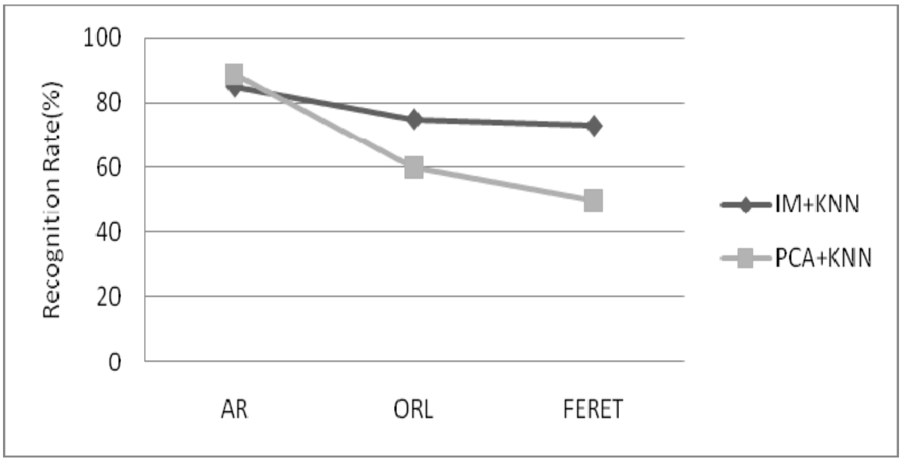


Figure 11. Recognition rates using the both methods

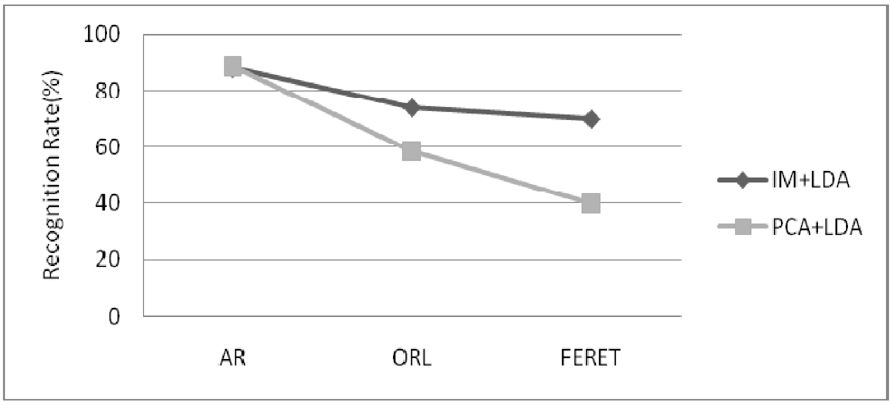


Figure 12. Recognition rates using the both methods

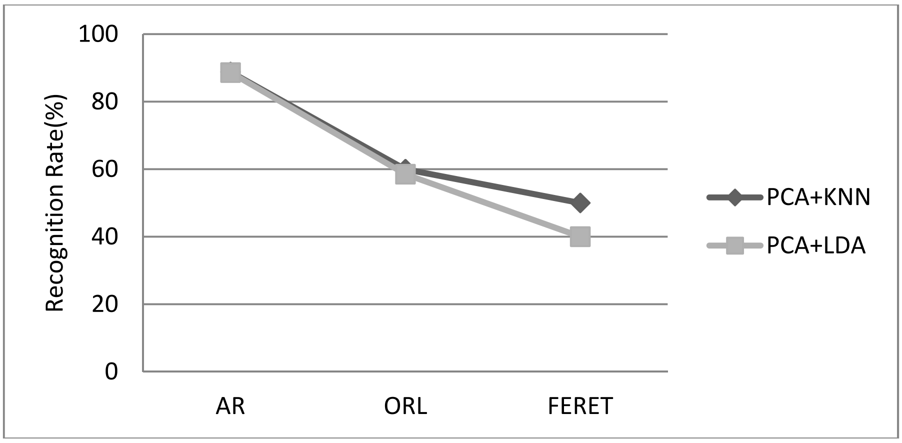


Figure 13. Recognition rates using PCA

Table 4. Average recognition time of one test image

	AR	ORL	FERET
IM+KNN	0.00013s	0.00005s	0.00022s
IM+LDA	0.00014s	0.00006s	0.00022s
PCA+KNN	0.00014s	0.00064s	0.00031s
PCA+LDA	0.00014s	0.00055s	0.00035s

Figure 10 shows the recognition rate of the proposed method using classifiers of KNN and LDA respectively. Obviously, KNN and LDA have almost the same recognition rate for the three databases. Especially, the proposed method achieves the best accuracy of about 85% for AR database. This is because there are all frontal faces in AR database while in ORL and FERET databases there are some slight tilt faces. Figure11 and Figure12 show that using the same classifier, the recognition rate of the proposed method is much better than that of PCA in ORL and FERET databases. While in AR databases, the accuracies of both methods are almost the same. This also estimates that the proposed method is more robust than PCA in case of non-frontal faces. Similarly in Figure 13, KNN and LDA have slight effect on the accuracy for the both methods.

To evaluate the recognition time of each facial image, we also calculate the recognition time for each test image supposing the training database is ready. Table IV shows the average recognition time of each test image in the three databases. Obviously, the proposed method running time is generally less than PCA.

5. Conclusion

In this paper, we discuss the proposed invariance model. In this model, wavelet transformations are used instead of Gabor filters by analyzing the reconstructed images from decomposition coefficients to get more accurate local details. To avoid the effects caused by shift, rotation and other inference, the moments and weighted pooling technique are used to keep the invariance and reduce dimensions greatly. The simulation results show the proposed method keeps the balance of the accuracy, storage and running time.

Acknowledgment

This work was supported in part by the project with granted No. 2011BK018 from Nature Science Foundation in Jiangsu province, China.

References

[1] J. Lu, YP. Tan, G. Wang. Discriminative multimanifold analysis for face recognition from a single traning sample per person. *IEEE Trans. PAM*. 2013; 35(1).

[2] H. Mohammadzade, D. Hatzinakos. Projection into expression subspaces for face recognition from single sample per person. *IEEE Trans. Affective Computing.* 2013; 4(1): 69-82.

[3] W. Deng, J. Hu, J. Guo. Extended SRC: Undersampled face recognition via intraclass variant dictionary. *IEEE Trans. PAMI.* 2012; 34(9): 1864-1870.

[4] A. Muntasa. New modelling of modified two dimensional fisherface based feature extraction. *Telkomnika Telecomun. Compt. Electr. Control.* 2014; 12(1): 115-122.

[5] Y. Zhang, L. Chen, Z. Zhao et. al. A novel multi-focus image fusion method based on non-negative matrix factorization. *Telkomnika Telecomun. Compt. Electr. Control.* 2014; 12(2): 379-388.

[6] L. Lsil, J.Z. Leibo, T. Poggio. Learning and disrupting invariance in visual recognition with a temporal association rule. *Frontiers in Computational Neuroscience*. 2012; 6(37).

[7] JZ. Leibo, J. Mutch, T. Poggio. Why the brain sperates face recognition from object recognition. *Advances in Neural Information Processing Systems.* 2011: 711-719.

[8] JZ. Leibo, J. Mutch, T. Poggio. Learning generic invariance in object recognition: translation and scale. *MIT-CSAIL-TR-2010-2061*.

[9] T. Serre, M. Riesenhuber. Realistic modeling of simple and complex cell tuning in the HMAX model, and implications for invariant object recognition in cortex. *AI Memo 2004-017,* MIT. 2004.

[10] C. Cadieu, M. Kouh, T. Poggio. Investigating position-specific tuning for boundary conformation in v4 with the standard model of object recognition. *Technical Report, MIT*. 2004.

[11] M. Riesenhuber, T. Poggio. Models of object recognition. *Nature Neuroscienc*e. 2000; 3(Supp): 1199-1204.

[12] R. Schneider, M. Riesenhuber. A detailed look at scale and translation invariance in a hierarchical neural model of visual object recognition. *AI Memo 2002-011, MIT*. 2002.

[13] MK. Hu. Visual pattern recognition by moment invariants. *IRE Trans. Information Theory*. 1962; 8(2): 179-187.

[14] T. JanFlusser, Z. Brbara. *Moments and Moment Invariants in Pattern Recognition.* John Wiley&Sons. 2009.

[15] Q. Tian, J. Wu. A review on face recognition based on compressed sensing. *IETE Technique Review*. 2013; E96B(4): 948-955.