GABOR WAVELET BASED POSE ESTIMATION FOR FACE RECOGNITION

B. Gökberk, L. Akarun, E. Alpaydın
Dept. of Computer Engineering, Boğaziçi University, 80815 Bebek, İstanbul, Turkey
{gokberk,akarun,alpaydin}@boun.edu.tr

Abstract. One of the major difficulties in face recognition systems is the in-depth pose variation problem. Most face recognition approaches assume that the pose of the face is known. In this work, we use a variation of Gabor wavelet transform for the representation of human face images to efficiently solve the pose estimation problem. Parameters of the Gabor wavelets, namely frequency and orientation, are adjusted to gain better performance. Principal Component Analysis is performed to reduce the dimensionality without a significant loss in the performance. Our results show that Gabor wavelet based filtering of images improves the performance of the pose estimation module.

1. Introduction

There has been a significant improvement in automatic recognition of human face images over the last five years. However, the task of robust face recognition is still difficult as recent tests revealed problems related to pose variation and illumination problem. Pose variation is a serious problem especially when there exist in-depth rotation variations in input images.

Various methods have been proposed to handle the rotation problem. Basically they can be grouped into three approaches: 1) multiple-image based methods when multiple images per person are available, 2) hybrid methods when multiple training images are available during training but only one database image per person is available during recognition, and 3) single image based methods when no training is carried out.

Among the first class of approaches, Beymer (Beymer 1993) proposed a template-based correlation matching scheme where 2D affine transformation and optical flow algorithm is used to align images before obtaining correlation scores. The main restrictions of this method are 1) many images of different views per person are needed, 2) no lighting variations are allowed, and 3) the computational cost is high.

Second type of approach includes linear class based methods (Vetter and Poggio 1997), graph matching based methods (Wiskott et al. 1997), and the view-based eigenface (Pentland et al. 1994) methods. In (Vetter and Poggio 1997), an image synthesis method based on the assumption of linear 3D object classes is implemented to calculate the correspondence between the reference image and the other example images using optical flow. In (Wiskott et al. 1997), a robust face recognition scheme based on Elastic Bunch Graph Matching is proposed using the learned transformations of 'jets' under face rotation. The drawback of this approach is the requirement of accurate landmark localization which is not an easy task. In

(Pentland et al. 1994), the popular eigenface approach has been extended to view-based eigenface method in order to achieve pose-invariance. This method codes the pose information by constructing an individual eigenface for each pose. Third class of approaches consist of 3D model based methods and invariant feature based methods (Gordon 1991).

In this paper we aimed to design a pose estimation module which can be used in a view-based system prior to the recognition step. For the sake of simplicity and to avoid the drawback of accurate localization of facial landmarks, we used a uniform sampling of feature points through a square mesh structure. At each grid point, we have extracted feature vectors using the outputs of multiscale and multiorientation 2D Gabor wavelets. For classification, we have used nearest neighbour classifier. Our results show that Gabor based preprocesing of images significantly improves the pose estimation performance. We also observed that dimensionality reduction of feature vectors using PCA can be used without a significant loss of estimation performance in the system.

2. Gabor Wavelets

The processing of facial images by Gabor filter is chosen for its biological relevance and technical properties. The Gabor filter kernels have similar shapes as the receptive fields of simple cells in the primary visual cortex (Wiskott et al. 1997). They are multi-scale and multi-orientation kernels. The response describes a small patch of gray values in an image $I(\vec{x})$ around a given pixel $\vec{x} = (x, y)$. It is defined as a convolution

$$J_j(\vec{x}) = \int I(\vec{x}')\psi_j(\vec{x}-\vec{x}')d^2\vec{x}' \qquad (1)$$

with a family of Gabor kernels

$$\psi_j(\vec{x}) = \frac{k_j^2}{\sigma^2}\exp\left(-\frac{k_j^2 x^2}{2\sigma^2}\right)\left[\exp(i\vec{k}_j\vec{x}) - \exp\left(-\frac{\sigma^2}{2}\right)\right] \qquad (2)$$

in the shape of plane waves with wave vector $\vec{k}_j$, restricted by a Gaussian envelope function. We employ a discrete set of 5 different frequencies, with $v = 0,...,4$, and 8 orientations, with $w = 0,...,7$,

$$\vec{k}_j = \left(k_{jx}, k_{jy}\right) = \left(k_v \cos\varphi_w, k_v \sin\varphi_w\right),\ k_v = 2^{-\frac{v+2}{2}}\pi,\ \varphi_w = w\frac{\pi}{8} \qquad (3)$$

where $j = w + 8v$. The width $\frac{\sigma}{k}$ of the gaussian is controlled by the parameter $\sigma = 2\pi$. Graphical representations of the real and imaginary parts of a Gabor wavelet kernel and the results of the full convolution of Gabor filters for different frequencies and orientations are shown in Figure 1 and Figure 2, respectively.

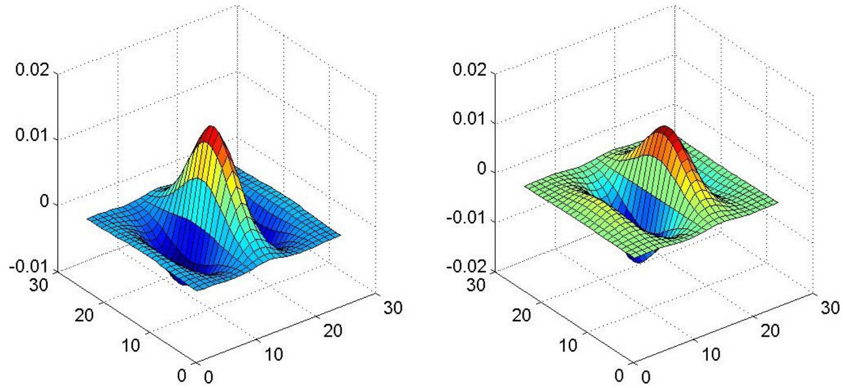

Figure 1. Real and Imaginary parts of Gabor wavelets

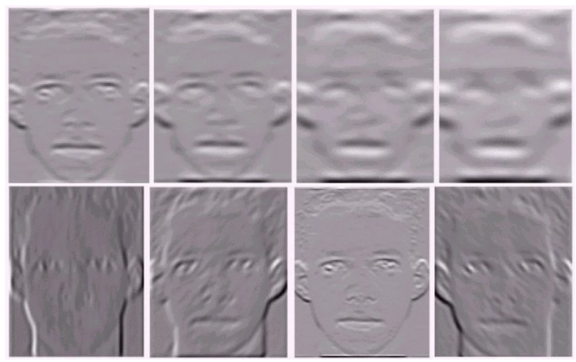

Figure 2. Top row shows Gabor-filtered images for frequencies *v*=1,2,3,4.
Bottom row shows Gabor filtered images for orientations *w*=0,2,4,6.

3. Pose Estimation

We have used a database contaning images of 19 subjects (Demir et al. 2000). Each subject has 6 different images for each of the 3 different poses. For each pose, there are slight variations in facial expression and both in-depth and in-plane head rotation angle (see Figure 3).

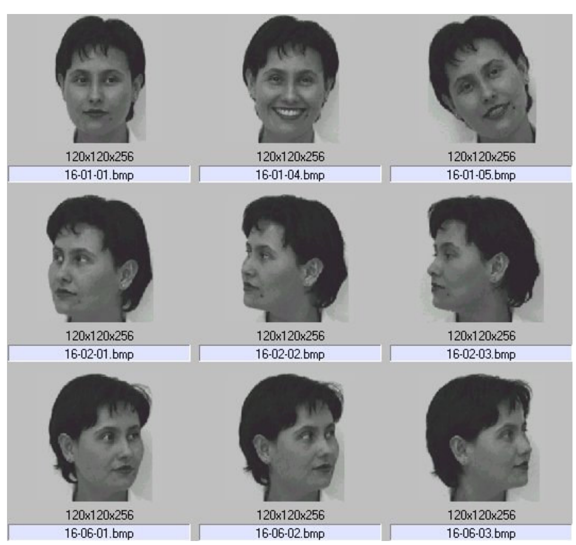

Figure 3. Sample images from a face database

We have implemented 2 different methods for pose estimation. In the first method, face images are transformed into the eigenface domain using PCA mapping (Demir et al. 2000). Then, using a classifier in the eigenface domain, we estimate the pose of a given sample. In this method, no preprocessing of images is performed. As a classifier, we investigate nearest neigbour classifier. We have used city-block distance and euclidean distance metrics. Performance results for this method with various image sizes are given in Table.1. For each pose, one image is used for training data and the remaining five images are used as test data. Performance results are the averages of 6 runs for each image in a specific pose class.

TABLE 1. Raw image vs PCA-of-Raw image pose estimation performance

	Image Size					
	16x16 (dim=256)		32x32 (dim=1024)		64x64 (dim=4096)	
	City	Euc	City	Euc	City	Euc
Raw Image	91,75	90,76	91,28	90,35	91,46	90,81
PCA-of-Raw Image (dim=20)	86,96	84,39	87,83	85,49	87,78	85,85
PCA-of-Raw Image (dim=40)	87,13	84,50	87,72	85,56	87,95	85,90

In the second method, we have placed a square grid of size nxn to the centre of the image. For each grid point, we have convolved the image with Gabor wavelets with *v* different frequencies and *w* different orientations. Convolution operation gives complex valued feature values. We took the magnitudes of these complex numbers as our features. Each grid point is then represented by a feature vector of size *v* x *w* called a 'jet'. So for a grid of size n x n, we represent an image with a feature vector of size (n x n) x (*v* x *w*).

The effect of changing the grid size that is placed on an image can be seen in Table 2. Although increasing the grid size improved the performance slightly, when time and space complexity is more important than accuracy, it would be more efficient to use smaller grids.

TABLE 2. Effect of grid size on the classification performance

Grid Size	ImageSize=32x32		ImageSize=64x64	
	City	Euc	City	Euc
3x3	95,84	95,84	96,78	95,73
5x5	96,26	95,96	96,78	96,02
7x7	96,02	96,32	96,55	96,14

We have tried different combinations of frequencies and orientations for Gabor filtering of images. These schemes and their classification performances are in Table 3. These results were obtained from a grid of size 3x3 placed on images of size 32x32. Schemes from 1 to 13 are shown to see the effects of individual frequency and orientation on the pose estimation performance. In scheme 14, all 5 frequencies and 8 orientations were used which resulted in a feature vector of size 40 for each grid point on the image. Optimum pose estimation performance was achieved when using scheme 15 which used orientations 2 and 6, and all of the frequency range.

TABLE 3. Gabor filtering schemes

Scheme	Freq.	Ori.	Perf.
1	0	All	92,8
2	1	All	90,4
3	2	All	90,8
4	3	All	90,5
5	4	All	90,8
6	All	0	79,5
7	All	1	89,8
8	All	2	91,6
9	All	3	87,7
10	All	4	87,8
11	All	5	90,7
12	All	6	93,7
13	All	7	88,8
14	All	All	94,3
15	All	2,6	96,7

We have applied the standard Principal Component Analysis technique to the outputs of Gabor filtered images and thus reduced the dimensionality of feature vectors to relatively low values. As can be seen from Table 4, we can compress our feature vectors at a rate 12:1 with a minor degradation in performance of the system.

TABLE 4: Effect of dimensionality reduction

Feature Vector Dimension	ImageSize=64x64	
	City	Euc
Original Vector (250)	96,78	96,02
80	95,91	96,02
40	95,49	95,85
20	95,09	95,79
10	93,91	94,27

The effect of dimensionality to the pose estimation performance for the Gabor wavelet representation is shown in Figure 4. It is found that when compared to the results of raw image representation, dimensionality reduction is more efficient in Gabor wavelet representation. Best estimation performance for the raw image based representation was 91,46 and 87,95 when dimensionality is reduced to 40, whereas the best pose estimation performance for the Gabor wavelet representation was 96,78 and 95,85 when dimensionality is reduced to 40.

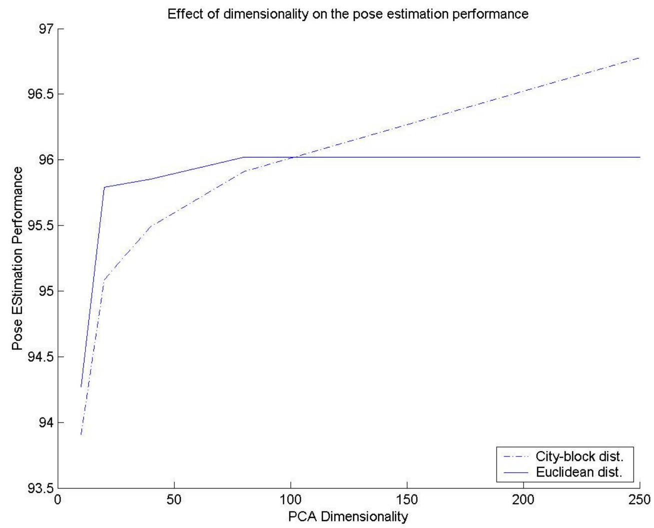

Figure 4. Dimensionality vs pose estimation performance

4. Conclusion and Future Work

In this paper, we have presented a pose estimation module for a pose-invariant face recognition system which used Gabor wavelet based representation of face images. Although we have used a basic square grid sampling of Gabor wavelets, an improvement in the pose estimation performance was observed. In reduced dimensionality, a major improvement from 87,95% to 95,85% was observed in addition to improvements in time and space complexity. We have also showed that some combinations of filter frequencies and orientations are more robust for the estimation of pose. As a future work, we will design a view-based face recognition module using our pose estimation method and will try to use other subspaces other than PCA subspace for better discrimination performance. Improvement of the Gabor representation of face images will be another research topic in future work.

REFERENCES

Beymer, D.J., “Face Recognition Under Varying Pose,” Technical Report 1461, MIT AI Lab, 1993.

Vetter, T., and T. Poggio, “Linear Object Classes and Image Synthesis from a Single Example Image,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, pp. 733-742, 1997

Wiskott, L., J.-M. Fellous, and C. Von der Malsburg, “Face Recognition by Elastic Bunch Graph Matching,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, pp. 775-779, 1997

Pentland, A. , B. Moghaddam, and T. Starner, “View-based and Modular Eigenspaces for Face Recognition,” *in Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*, 1994.

Gordon, G., “Face Recognition Based on Depth Maps and Surface Curvature,” *in SPIE Proceedings, Vol. 1570: Geometric Methods in Computer Vision*, pp. 234-247, 1991.

Demir, E., L. Akarun and E. Alpaydin, “Two-Stage Approach for Pose Invariant Face Recognition,” *Proceedings ICASSP’2000*, Istanbul, June 2000.