FEATURE SELECTION FOR POSE INVARIANT FACE RECOGNITION

B. Gökberk, L. Akarun, E. Alpaydın
Computer Engineering Dept.
Boğaziçi University
{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 have designed a feature based pose estimation and face recognition system using 2D Gabor wavelets as local feature information. The difference of our system from the existing ones lies in its simplicity and its intelligent sampling of local features. Intelligent feature selection can be carried out by learning a set of parameters where the aim is the optimal performance of the overall system. In this paper, we give comparative analysis of the performance of our system with the standard modular Eigenfaces approach and show that local feature based approach improved the performance of both pose estimation and face recognition. For efficient coding, we have employed Principal Component Analysis(PCA) to the outputs of local feature vectors. Intelligent feature selection also reduced the space and time complexity of the system while retaining almost the same estimation and recognition accuracy.

1. Introduction

A computer based face recognition system designed for a real–life application must take into account the pose variation of faces because faces often undergo large rotations in depth. In–depth pose variation problem is not trivial to solve because it introduces nonlinear transformations. The situation becomes worse when lightning differences, occlusions and self shadowing of facial features are present.

A good solution would be to use information about three dimensional nature of a face [4]. However, 3D model based approaches are both expensive and difficult to develop. Alternatively, a view–based approach would be an efficient solution. View–based methods generally use stored canonical 2D images of different views, and employ a pose estimation module prior to identity recognition [1, 7]. The view–based and 3D model based approaches were combined in [9] to generate new views of a face.

PCA was found to be informative about the pose manifolds (eigensignatures) that occur when a face undergoes pose changes [6]. In [2], a variant of eigensignature based approach was presented. This idea is based upon the concept that an individuals's face exhibits a distinctive characteristic through a pose-varying eigenspace.

However, PCA is sensitive to illumination conditions, scale, translation or rotation in the image plane. In order to overcome these problems, local, preferably directional image features are used. In [3], 2D Gabor wavelet based pose manifolds are used to estimate pose and meaningful Gabor frequency/orientation parameters were also analyzed and certain directions were found to be important for pose estimation. Similarly, [5] has performed a statistical analysis of Gabor kernel parameters for frontal faces. In [8], a face pose estimation module was proposed using Gabor wavelet networks which represent an object as a linear combination of Gabor wavelets.

In this work, we propose a local feature based system that uses an intelligent/optimal sampling methodology. In the rest of the paper, we explain our Gabor wavelet based representation and optimal parameter selection scheme. We then give comparative analysis of our results with a standard modular eigenface approach, and examine the space and time complexity of our approach.

2 Feature Based Pose Invariant Face Recognition

2.1. Gabor Features

A biologically motivated representation of face images is to code them using convolutions with 2D Gabor–like filters. In order to represent face images using Gabor filters, we have placed a square grid over the face region in the image. At each grid point on the image we have convolved the image with Gabor kernels. The set of convolution coefficients for kernels of different orientations and frequencies at one image pixel is called a *jet* [5]. A *jet* contains responses of convolutions in an image, $I(\vec{x})$ around a given pixel $\vec{x}=(x,y)$. It is based on a wavelet transform, defined as a convolution with a family of Gabor kernels

$$\psi_j(\vec{x}) = \frac{k_j^2}{\sigma^2} e^{-\frac{k_j^2 x^2}{2\sigma^2}} [e^{i\vec{k_j}\vec{x}} - e^{-\frac{\sigma^2}{2}}] \quad (1)$$

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 $v=0,\ldots,4$, and 8 orientations, with $w=0,\ldots,7$,

$$\begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_v \cos\varphi_\mu \\ k_v \sin\varphi_\mu \end{pmatrix}, k_v = 2^{-\frac{v+2}{2}}\pi, \varphi_\mu = \mu\frac{\pi}{8}, \quad (2)$$

with index $j=\mu+8v$. The width σ/k of the Gaussian is controlled by the parameter $\sigma=2\pi$.

We want to learn which frequencies and which orientations are useful for both pose estimation and face recognition. So, in our system, we carried out experiments to show independent contributions of each frequency and orientation on the pose estimation and recognition performance. Thus, by using the most important subset of frequency and orientation parameters, we can speed up the feature representation phase and have a more compact form of feature vectors.

2.2. Intelligent Sampling Grid Selection

We have used a sparse sampling of of face images using a rectangular lattice, thus avoiding the problems that can arise when precise facial feature localization is used. It can be shown that each sampling lattice point(grid point) has a different importance for pose estimation and recogniton. For example, with a rectangular grid, the grid points on the borders are more efficient in face localization and outline extraction, whereas inner grid points that represent inner facial landmarks (eyes, mouth, forehead) are more useful for face recognition. Therefore, grid points mat be prioritized for each application.

We have employed a learning algortihm which tries to learn the importance of each grid point for a given task. Learning algorithm gives a weight to a grid point based on the estimation (or, recognition) performance using that points' feature vector only. This is an off–line process in which a sufficiently large training database can be used to extract the weights of each grid point. After extracting grid weights, one can take into account those weights that optimize the performance of the system. For this purpose, we have tried two schemes; a hard scheme and a soft scheme. In the hard scheme, the grid points whose weights are below a certain threshold are discarded. In the soft scheme, after discarding the least important grid points, the distance between feature vectors is weighted by the inverse of the normalized importance of grid points.

2.3. Classifier

When there is pose variance in the face database, there are two ways to deal with the pose problem. In the parametric approach, a single eigenspace will encode both pose and identity. The second approach, named view–based approach, builds a view–based set of P separate eigenspaces, each capturing the variation of the individuals in a common view. Once the pose is determined, the image is classified by the eigenvectors of that view space. Previous studies [1] showed that view–based coding has a performance advantage when compared to the parametric approach. In our work, we have used view–based approach. Nearest neighbor classifier is used to determine the pose of the input image where L_1 norm is used as a distance measure.

3. Results

3.1. Face Databases and Methodology

We have used two face databases. In the first database, named the ESRA face database, there are 20 subjects, each having 18 images for each of the three pose classes, namely left (full profile), right (full profile), and centered (frontal). Recognition results are the averages of two runs where six images for each pose is put into training set and the rest is put into test set. In the CVL database, there are 113 subjects and three pose classes: right (full profile), right half profile, frontal. For each pose class, there are only two images. We have put one image for each pose class into the training set, and put the remaining image into the test set. Recognition results for the CVL face database are the averages of two runs (by swapping training and test sets for the second run).

3.2. Modular Eigenface Approach

We have used view–based eigenface technique to transform 32×32 images into a lower dimensional space. Table 1 shows the results for the ESRA face database. First two columns show the feature vector dimension (in the reduced PCA space), where *nview* denotes the dimensionality for the view–based approach, and the next three columns displays the recognition performance of view–based approach for each pose class. The last column shows the pose estimation performance that is carried out prior to view–based coding. we see that increasing the feature vector dimensionality slightly improved the performance. We have chosen *nview* dimensionality that performed best for later comparisons.

Table 1. Eigenface performance

ESRA				
nview	view1	view2	view3	pose
30	94.07	94.70	91.50	94.86
60	93.39	94.90	91.54	95.00
360	93.39	95.30	91.97	95.14
CVL				
nview	view1	view2	view3	pose
28	6.58	9.06	19.03	60.47
57	9.97	10.91	19.27	60.47
113	10.21	10.70	20.10	60.17

Second part of Table 1 displays the same results for the CVL database. As expected, both recognition and pose estimation performance is very low since in the CVL database, there is only one image for the training set, and one image for the test set. Also, it is much more difficult for a pose estimator to differentiate between half profile view between frontal and full profile views. These two reasons make the CVL face database harder for a typical face recognition task.

3.3. Frequency and Orientation Selection

In order to examine the importance of frequency and orientation of a 2D Gabor wavelet, we have placed a square grid of size 3×3 on 32×32 face images, and convolved the image with specific frequencies and orientations. In Table 2, recognition performances of different frequency parameters are shown for the ESRA face database. First column shows the selected frequency, and the second column shows the orientation range. View–based results for each three pose classes are averaged and displayed as the fifth column. Rows 2–6 show the effect of different frequencies from high to low. It is clear that lower frequencies are more useful for both recognition and pose estimation. Rows 7–14 show the effect of different orientation parameters from 0 degree to 157.5 degree. It is found that orientations which stress the horizontal facial features carry most of the information for the classification of face images. Therefore, we have used lower–frequencies [2 3 4] and horizontal feature sensitive orientations [2 3 4 5 6] in parallel, reducing the system complexity by 37.5 percent.

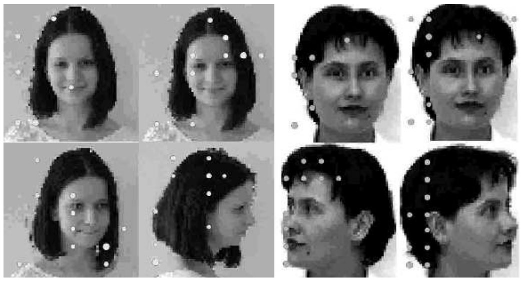
3.4. Grid Point Selection

We have used a more complex lattice that has dimensions of 7×7 to examine the importance of grid points on the lattice which provides finer resolution. Using selected frequencies and orientations, convolution is performed. To select the most informative grid points for the pose estimation task, we have used each grid points jet response as an isolated feature vector for a given test image. Using only this feature vector, we have calculated the pose estimation performance of the system. Higher estimation performance means that the selected point contains useful information, thus deserves a higher weight proportional to its estimation performance. Applying this scheme to each grid point on the lattice, we can get weights for each point. In the view–based approach, we must repeat this scheme for each pose class for a face recognition task. Figure 1(a) and 1(b) show important grid points that are found by our approach for ESRA and CVL database. Image in the upper–left corner shows important grid points for the pose estimation task, upper right, lower left and lower right images show important grid points for the view class that the background image indicates. It is observed that for pose estimation, grid points that are close to face outline become more important as expected. Our results confirm the findings of [5] that grid points on the forehead region are important for face recognition of frontal images. For full profile images, the vertical ear axis carries the most weight. In general we can conclude that in frontal images, horizontally located grid points in the upper half of the image carry more weight whereas in half and full profile images, a vertical axis dominates.

Table 2. Frequency/orientation importance

fre	ori	jet.dim	view	pose
0	0:7	8	80.69	92.36
1	0:7	8	85.74	91.14
2	0:7	8	88.43	94.72
3	0:7	8	93.09	**96.11**
4	0:7	8	**94.36**	95.97
0:4	0	5	77.30	86.97
0:4	1	5	83.97	92.50
0:4	2	5	82.99	91.42
0:4	3	5	89.11	**93.72**
0:4	4	5	89.53	92.14
0:4	5	5	**89.56**	92.92
0:4	6	5	87.44	93.08
0:4	7	5	83.32	91.03

Table 3 shows the recognition results for ESRA face database. Columns one and two show the number of grid points selected for pose estimation and face identity recognition respectively. In the first experiment, full 5×5 grids were used, yielding 25 grid points. In the second experiment, we have selected most important 6 grid points for pose estimation and used 5×5 grid for face recognition in the related pose class. When we compare (25,25) and (6,10), we see that we can come up with a more efficient



(a) (b)

Figure 1. Upper left corner images show the chosen grid points for pose estimation, and the other images show the chosen grid points for recognition of different pose classes for (a) CVL, (b) ESRA database.

Table 3. Gabor performances

Grid Sizes		ESRA			
Pose	ID	View1	View2	View3	Pose Est.
25	25	96.76	96.76	97.30	96.45
6	25	95.11	97.72	96.49	96.11
25	10	95.68	95.54	96.46	96.45
6	10	94.27	96.27	95.47	96.11
Grid Sizes		CVL			
Pose	ID	View1	View2	View3	Pose Est.
25	25	28.42	15.96	49.18	78.61
6	25	23.03	16.24	51.56	79.05
25	10	22.33	16.48	51.60	78.61
6	10	19.13	16.75	54.12	79.05

representation with a minor performance degradation. In these experiments, we have used the hard scheme by discarding the least important grid points. Second part of Table 3 shows the grid point selection results for CVL database. As expected, intelligent selection of grid points can perform almost equally when compared to more complex representation. When the soft scheme is employed, recognition performances for the ESRA database improved slightly. In some cases, the performance increase was 0.6% while in others, the performance did not change.

When the time complexity is considered, the major drawback of the Gabor sampling was the convolution process. By intelligently selecting Gabor parameters and grid points, we can reduce the time complexity by factor of $(N^2 \times F \times O)/(n \times f \times o)$ where N^2, F, and O represent the number of grid points, frequencies and orientations used, respectively. The parameters, n, f, and o, represent the number of grid points, frequencies and orientations that are used in the improved system.

4 Conclusion

In this work, we have designed a feature based pose estimation and face recognition system using 2D Gabor wavelets as local feature information. As opposed to previous approaches that uses Gabor wavelet based coding of face images, we determined optimum frequency and orientation parameters which are useful for pose estimation and face recognition tasks. We also located facial regions from which local features are extracted, in an intelligent way to both maximize the system performance and at the same time decrease system complexity. For efficient coding, we have employed PCA to the outputs of local feature vectors. It is found that by giving higher priorities to local feature extractors that uses lower frequencies and horizontaly–tuned directed wavelet kernels, face representation becomes more suited to pose estimation and face recognition. We also showed that intelligent learning of topographical information (*where* information) for efficient face coding decreased system complexity while retaining good recognition accuracy.

References

[1] M. T. A. Pentland, B. Moghadam T. Starner. View based and modular eigenspaces for face recognition. In *Proceedings on IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 84–91, 1994.

[2] N. M. A. D. B. Graham. $norm^2$–based face recognition. In *Proceedings of the Computer Vision and Pattern Recognition*, pages 586–592, 1999.

[3] E. O. J. Sherrah, S. Gong. Understanding pose discrimination in similarity space. In *Proceedings of the 10th British Machine Vision Conference*, 1999.

[4] C. v. d. M. K. Okada, S. Akamatsu. Analysis and synthesis of pose variations of human faces by a linear pcmap model and its application for pose–invariant face recognition system. In *Proceedings of Fourth International Conference on Automatic Face and Gesture Recognition*, pages 142–149, 2000.

[5] J. H. P. Kalocsai, C. von der Malsburg. Face recognition by statistical analysis of feature detectors. *Image and Vision Computing*, 14(4):273–278, 2000.

[6] S. J. M. J. J. C. S. Gong. An investigation into face pose distributions. In *Proceedings of the Second International Conference on Automated Face and Gesture Recognition*, 1996.

[7] C. J. T. T. F. Cootes, G. J. Edwards. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–684, 2001.

[8] G. S. V. Kruger, S. Bruns. Efficient head pose estimation with gabor wavelet networks. In *Proceedings of the 11th British Machine Vision Conference*, 2000.

[9] T. Vetter. Synthesis of novel views from a single face image. *International journal of Computer Vision*, 28(2):103–116, 1998.