

OPTIMAL GABOR KERNEL LOCATION SELECTION FOR FACE RECOGNITION

B. Gökberk, M. O. İrfanoğlu, L. Akarun, and E. Alpaydın

Boğaziçi University, Computer Engineering Dept.
Istanbul, TURKEY
{gokberk,irfanogl,akarun, alpaydin}@boun.edu.tr

ABSTRACT

In local feature-based face recognition systems, the topographical locations of feature extractors directly affect the discriminative power of a recognizer. Better recognition accuracy can be achieved by the determination of the positions of salient image locations. Most of the facial feature selection algorithms in the literature work with two assumptions: one, that the importance of each feature is independent of the other features, and two, that the kernels should be located at fiducial points. Under these assumption, one can only get a sub-optimal solution. In this paper, we present a methodology that tries to overcome this problem by relaxing the two assumptions using a formalism of subset selection problem. We use a number of feature selection algorithms and a genetic algorithm. Comparative results on the FERET dataset confirm the viability of our approach.

1. INTRODUCTION

In recent years, automatic human face identification problem has become one of the challenging problems of pattern recognition. As in most pattern recognition problems, one has to find a suitable representation of input patterns for a given task. Thus, a significant amount of research in the face recognition area has focused on finding an efficient representation of face images for a recognition task [1].

There are mainly two approaches for face representation, holistic, and feature based. In the holistic approaches, statistical methods are used to represent face images as a whole. On the other hand, feature based methods try to extract local features from face images. Often, feature based approaches require the localization of fiducial points. Early attempts use typical geometrical features such as the location and relative positions of facial features (i.e. eyes, nose, and mouth) [1]. Recently, 2D Gabor wavelet-based methods are used for local feature-based human face representation [2].

In 2D Gabor wavelet-based methods, sparse sampling at the fiducial points of face images is carried out where local features are extracted using multi-frequency and multi-orientation Gabor kernels. Typically, the features extracted

using this method have large dimensionality. It is, therefore, required to statistically analyze the contribution of each feature vector to the recognition task so that one can measure the relative importance of topological relevance and the effect of the frequencies and orientations used [3].

In [4], the discriminative power of the nodes of a graph that is placed over face features is examined. The aim is to learn the weights of nodes for face discrimination. The problem is formulated as an optimization problem and simplex algorithm is used. According to their results, the eyes are more important for discrimination of half profiles and frontals compared to the mouth and chin. A similar approach was employed in [5] where the aim of the learning algorithm is to find out a suitable subgraph which only contains the nodes important for head finding and pose identification.

In a recent statistical analysis of 2D Gabor wavelet-based feature detectors [6], univariate analysis of variance of 2D Gabor kernel activations has been used to weight the contribution of each parameter (kernel location, frequency, and orientation) in the representation according to its power of predicting similarity of faces. The results show that the hairline area with the forehead and eye regions provide useful information while the mouth, nose, cheek and lower part of the outline region are the least useful part of a face for face recognition. In a similar work, results confirm that the eyes and mouth are more stable for recognition, whereas hair and nose region have large variations [7].

In this paper, we formalize the salient feature selection problem as a special case of a subset selection problem. In all of the previous approaches, a common assumption for location of Gabor kernels is that these should be located at points. Our aim is to test the validity of this approach by formulating these as a feature selection problem. Another restrictive assumption is that each feature's discriminative power is considered to be independent. However, a more complete approach would be to consider the relationships between features. In this formalism, one should use a more complicated feature selection methodology where the independence assumption is relaxed. In the rest of the paper, we explain our local image descriptors in section 2, and feature

selection methodology in section 3. In section 4, we give experimental results for FERET face database.

2. IMAGE REPRESENTATION USING 2D GABOR WAVELETS

Local features are represented using the convolution results of the face image with 2D Gabor wavelets at the convolution points. At each image point, we have used five different frequencies and eight different orientations. The Gabor kernel resolution is selected as 15×15 pixels in order to reduce the overlapping of kernels. The magnitudes of complex outputs of Gabor convolutions are used as feature descriptors, giving a feature vector of size 40 at each image point.

There are several methods to represent whole image using local jets. At one extreme, images can be represented by the full convolution with Gabor kernels at each pixel. Another approach would be to place a face graph where the nodes of the graph lie on facial features. This approach requires a fine localization of facial feature points. In between these two approaches, one can use a rectangular sampling grid that is placed over the face region.

3. FEATURE SELECTION METHODOLOGY

In feature selection, the goal is to find a subset maximizing a selected criterion. This criterion can be inter-class distance measure or the classification rate of a classifier. The optimal solution could be found by using exhaustive search. However, for higher dimensional problems, this solution is unusable. Branch and bound type of algorithms can also give optimal solutions [8], but their application is only limited to monotonic criterion functions, which does not hold in our case. Alternative to optimal algorithms, several fast sub-optimal algorithms can be used. Among them, the most frequently used ones are: sequential forward selection (SFS), sequential backward selection (SBS), plus-L-minus-R, and floating search methods (SFFS, SFBS) [9]. Genetic algorithms (GA) and tabu search are also proposed as solutions for a subset selection problem [9].

In order to find the most discriminative image locations of faces for recognition, we have designed several feature selection scenarios. These scenarios are, namely: best individual features (BIF), forward selection(SFS), floating forward search (SFFS), and genetic algorithm. In the first three approaches, we represented the face images using both rectangular grids (lattice) and manually positioned face graphs. Lattice-based sampling is done via placing a 7×7 grid centered on the face region. As a face graph, we have identified 30 facial feature points, as seen in Figure 2.a , and used them as the nodes of our graph. In the GA approach, we have used full convolutions of Gabor wavelets at each pixel in the image.

4. EXPERIMENTS AND RESULTS

In our experiments, we have used a subset of the FERET face database [10]. The used part of the database contains normalized frontal images of 146 subjects. Each subject has 4 gray scale images of resolution 150x130. In all of the experiments, we have put 2 images of a subject into training set, and the rest of the images into test set. Faces in the dataset contain facial expression and illumination variations. In the recognition part, we have used 1-nearest neighbor classifier.

4.1. BIF, SFS, and SFFS based feature selection

In a recognition problem, each individual local feature has a certain degree of recognition power. Therefore, it is useful to learn the importance of each local feature in order to obtain a better discriminator. The best heuristic to measure the importance of each local feature is to look at its individual recognition performance. In the BIF approach, one can simply combine the most important N features into a final feature vector. This simple idea can perform well only if each local descriptor contributes independently to the discrimination performance, irrespective of the existence of other local features. However, in many cases, it would be proper to design a feature selector which additionally takes into account the relative information gain when used with an existing feature set. Thus, we have used SFS algorithm in order to consider this relative gain. More formally, we add the most informative local feature at each step to an existing previously selected subset S .

SFFS algorithm takes this idea one step further by backtracking to remove the least useful features from an existing feature subset to overcome the nesting effect. Specifically, SFFS adds the most useful feature and then searches for a feature in the existing subset S to discard if the removal of that feature improves the discriminative power.

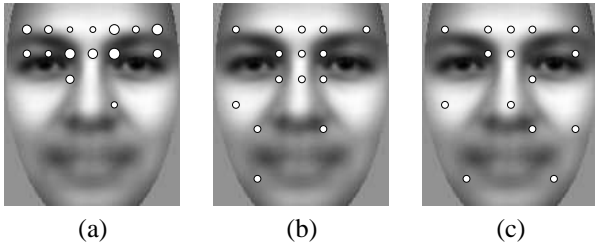
In our experiments, we have placed a rectangular lattice of size 7×7 over the face region, and look for a useful subset of grid points for efficient face representation for recognition task. In the second column of Table 1, the recognition performances of each method are presented. The recognition accuracies of BIF, SFS, and SFFS are 84.54, 90.38, and 91.07 percent respectively. The recognition performance of using all of the grid points is 86.94 percent. In Figures 1.a, 1.b and 1.c, the most important 15 local feature positions are shown for BIF, SFS, and SFFS algorithms. In Figure 1.a, circle sizes are proportional to each points recognition performance. The best performance is obtained using the SFFS approach, where the selected subset performs even better than using all of the grid points, and as expected, BIF approach performed worst among all of these methods since it considers each feature independently.

In all of the three approaches, most of the selected grid

Table 1. Comparative analysis of BIF, SFS, and SFFS for lattice- and face graph-based sampling methods. The numbers in parentheses show the number of feature points for each representation

	Lattice (49 pts)	Face Graph (30 pts)
All pts	86.94	83.85
BIF (15 pts)	84.54	82.13
SFS (15 pts)	90.38	87.97
SFFS (15 pts)	91.07	87.97

Fig. 1. The locations of important facial feature combinations for a) BIF, b) SFS, and c) SFFS approaches. The grid size in all figures is 7×7 . The recognition performances are 84.54, 90.38, and 91.07 percent, respectively.

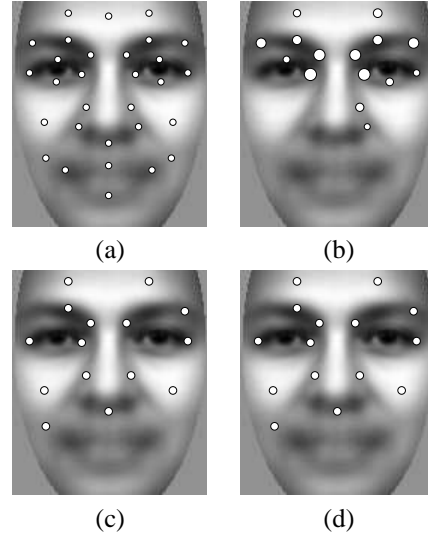


points are at the upper part of the face region. This result is largely due to the expression variations present in the dataset especially, in the mouth region. In SFFS, the combination of features extracted from eyebrows, the lower-center part of the forehead, the nose region, and the lower part of the mouth seems important.

Similarly, we have performed the same feature selection analysis to the manually positioned face graph, in order to see the importance of the generally used fiducial points. In the third column of Table 1, the recognition accuracies of BIF, SFS, and SFFS is shown. Using all of the 30 points in the face graph gives 83.85 percent classification performance, whereas an SFS-, or SFFS-based subset selection can improve the performance to 87.97 percent. In Figures 2.b, 2.c, and 2.c, the locations of 15 useful fiducial points are shown. The points selected for SFS and SFFS methods are the same, and they are generally at the upper face region. Eyebrows, the corners of eyes, forehead, cheeks, and the outline of nose seem to carry the most discriminative information.

When lattice and face graph based sampling is considered, lattice-based approach performs better. Our results show that although fiducial points are important, feature selection from a set of fiducial points greatly improves performance. Furthermore, our experiments with the lattice approach show that superior results can be achieved at the

Fig. 2. The locations of important facial feature combinations for manually positioned face graph. a) the locations of 30 fiducial points used, b) subset of 15 points for BIF c) subset of 15 points for SFS, d) subset of 15 points for SFFS. The recognition performances are 82.13, 87.97, and 87.97 percent, respectively.



periphery of fiducial points.

4.2. Genetic algorithm-based feature selection

One of the key motivations of our research was to try to understand whether it is better to choose the locations of facial features for local image descriptors. Therefore, we aimed to search for useful combination of face locations from data, without using any a priori information, such as fiducial point coordinates. In contrast to the sparse sampling methods (lattice, face graph), we have a much larger search space. The complexity of the search space is determined by the exhaustive search of a combination of N feature points selected from all of the pixels in the face region. However, in higher dimensions, such as in our problem, exhaustive search is unusable. So, we have used a genetic algorithm which is sub-optimal but faster. It was shown that genetic algorithms can reach near-optimal solutions quickly in feature selection [9].

In our setting, genetic chromosomes contain the coordinates of the selected number of face locations. We decided to use 15 points for face representation. As fitness function, we have used the recognition performance of local image descriptors of each gene in the chromosomes. The crossover and mutation parameters are 0.5 and 0.1, respectively. In both operators, we require that the coordinates of face points in a single chromosome do not overlap too much in order to extract independent local information as much as possi-

ble. This minimum overlap distance between facial point is selected to be 9 pixels. Mutation of a gene is handled by adding a random number within a specified range. This range is dependent on the image resolution. As the populations evolve, we iteratively narrow this range for better convergence. The selection of new population is based on the probability distribution of fitness values of each chromosome. For quick convergence, elitism is employed, where the elitism ratio is 0.05. As an initial population size, 200 is used.

In Figure 3, the 15 feature points found by the GA is shown. The recognition performance of this feature subset is 96.50 percent which is even better than the best sequential feature selection algorithm, namely SFFS. Again, all of the feature points are gathered over the upper face region. Similar to results of the SFFS algorithm, the outer corners of eyebrows, forehead region, and the outline of nose provide the most useful information.

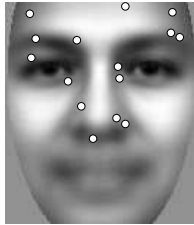


Fig. 3. The locations of important facial feature combinations for genetic algorithm.

5. CONCLUSION AND FUTURE WORK

In this work, a methodology to represent human faces in a local feature-based approach is presented. Previous research on feature selection for face recognition mainly focuses on the individual, independent contribution of each face point to the recognition performance. We have shown that, it is better to formulate the problem as a feature subset selection, where the addition or subtraction of a feature point is evaluated with respect to an existing feature subset.

Another common assumption in previous approaches was to extract local features from fiducial points. To test the validity of this assumption, we have used different sampling methods coupled with different feature selection algorithms. Our results show that although fiducial points are important, feature selection from a set of fiducial points greatly improves performance, and superior recognition accuracy can be achieved at the periphery of fiducial points.

In our experiments, sequential forward selection algorithm and genetic algorithm gave the best performances, while the latter is superior to the SFFS. In order to compare their performance with methods that selects features

based solely on their individual importance, we have implemented Best Individual Feature (BIF) algorithm. As expected, both SFFS and GA outperformed BIF-based feature selection. In general, eyebrows and face points at the outline of nose seem to provide the most discriminatory information for face recognition. As future work, frequencies and orientations of different Gabor kernels at specific locations will be examined in a similar fashion. We are also working on feature selection for pose invariant face recognition.

6. REFERENCES

- [1] S. Sirohey R. Chellappa, C. L. Wilson, "Human and machine recognition of faces: A survey," *Proceedings of IEEE*, vol. 83, pp. 705–740, 1995.
- [2] N. Kruger C. Malsburg L. Wiskott, J. M. Fellous, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, 1997.
- [3] E. Alpaydin B. Gokberk, L. Akarun, "Feature selection for pose invariant face recognition," in *Proceedings of the 16th International Conference on Pattern Recognition*, 2002.
- [4] N. Kruger, "An algorithm for the learning of weights in discrimination functions using a priori constraint," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 764–768, 1997.
- [5] C. Malsburg N. Kruger, M Potzsch, "Determination of face position and pose with a learned representation based on labelled graphs," *Image and Vision Computing*, vol. 15, pp. 665–673, 1997.
- [6] J. Horn P. Kalocsai, C. Malsburg, "Face recognition by statistical analysis of feature detectors," *Image and Vision Computing*, vol. 18, no. 4, pp. 273–278, 2000.
- [7] M. Grudin, "On internal representations in face recognition," *Pattern Recognition*, vol. 33, no. 7, pp. 1161–1177, 2000.
- [8] J. Mao A. Jain, R. P. W. Duin, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.
- [9] J. Sklansky M. Kudo, "Comparison of algorithms that select features for pattern classifiers," *Pattern Recognition*, vol. 33, no. 1, pp. 25–41, 2000.
- [10] J. Huang P. Rauss P. J. Phillips, H. Wechsler, "The feret database and evaluation procedure for face recognition algorithms," *Image and Vision Computing*, vol. 16, no. 5, pp. 295–306, 1998.