Selection of Location, Frequency and Orientation Parameters of 2D Gabor Wavelets for Face Recognition

Berk Gökberk, M.O. İrfanoğlu, Lale Akarun, and Ethem Alpaydın

Boğaziçi University, Department of Computer Engineering,
TR-34342, Istanbul, Turkey
{gokberk, irfanoglu, akarun, alpaydin}@boun.edu.tr

Abstract. In this paper, a two-level supervised feature selection algorithm for local feature-based face recognition is presented. In the first part, a genetic algorithm is used to determine the useful locations of the face region for recognition. 2D Gabor wavelet-based feature extractors are used for local image descriptors at these locations. In the second part, the most useful frequencies and orientations of Gabor kernels are determined using a floating feature selection algorithm. Our major aim in this study is to examine the relevance of the two common assumptions in the local feature based face recognition literature: first, that the contribution of a specific feature to the recognition performance is independent of others, and secondly, that feature extractors should be placed over the visually salient points. In this paper, we show that one can obtain better recognition accuracy by relaxing these two assumptions.

1 Introduction

In all of the computational face processing tasks such as face recognition, detection, and tracking, it is now widely accepted that the representation of a human face plays a very important role. In much of the recent works, researchers try to find an efficient representation method for a given task. In the face recognition literature, several approaches emerged during the last few years. These approaches can be broadly classified into two groups: local feature-based approaches including template-based methods, and global statistical approaches[1].

Feature-based approaches try to code face images using several different methodologies. In the most simplistic way, one can represent a face image using geometrical relations among various face regions. As this method, most feature-based approaches try to localize various facial feature points, such as the coordinates of eyes, mouth, nose, and eyebrows. Once these points are found or tracked, you can represent the face image by features extracted from these points.

2D Gabor wavelet-based methods are frequently used in feature-based face representation approaches as local feature extractors. Gabor kernels are similar to the receptive fields of simple cells in the primary visual cortex. In addition, multi-resolution and multi-orientation capabilities of Gabor kernels make them attractive for face representation. Since the full convolution of face images with different Gabor kernels is very costly, sparse sampling is generally used where local feature vectors are formed at each positions of Gabor kernels.

Typically, the features extracted by 2D Gabor wavelets have a very large dimensionality. It is, therefore, essential to analyze the contribution of each feature component to performance of the task at hand. In the most general case, one should examine three parameters of a Gabor kernel: location, frequency and orientation [2, 3].

There were several studies that tried to emphasize the importance of Gabor kernel parameters for face recognition. 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 frontal faces 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.

The determination of the importance of local Gabor features has also been used in face authentication systems. For example, in [6], an improvement for the elastic graph matching approach was proposed where local Gabor features extracted at the nodes of the graph are used to compute a local similarity response. In the local similarity calculation, Fisher's discriminant ratio is used to learn the coefficients which map feature vectors to a local similarity response. Global similarity is then defined as the sum of local similarities. Authors have shown that it is not necessary to take into account all local discriminants in order to obtain good performance. So, they use only a fraction of local discriminants. Their experimental results showed that eyes are more important at high frequencies whereas chin and cheek areas are important at lower frequencies for frontal face authentication.

In a recent statistical analysis of 2D Gabor wavelet-based feature detectors [7], 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 [8].

In almost all of the previous studies, either the importance of the locations of Gabor kernels or the importance of the used frequencies and orientations are examined. Only in [7], all the three parameters are examined, but they assume the independence of each feature dimension. This independence assumption is actually not valid, so one needs a more complex methodology to infer the usefulness of each local feature element. In this paper, we have formalized our approach as a subset selection problem, and removed the independence assumption. In addition, we also wanted to show the validity of the commonly used sampling technique of placing the Gabor kernels at salient facial feature positions such as the corners of eyes, mouth and the tip of the nose, etc. 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 convolved the image with Gabor kernels having 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 [9], 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) [10]. Genetic algorithms (GA) and tabu search are also proposed as solutions for a subset selection problem [10].

In order to apply feature selection algorithms to the task of finding optimal Gabor kernel locations, and finding useful frequency/orientation parameters, we have decomposed the problem into two parts by separating location finding problem and frequency/orientation selection problem. In the first part, location selection module tries to find optimal face regions in a supervised manner by using all of the 40 different Gabor kernels having full frequency and orientation range. Then, in the second part, frequency and orientation selection module tries to come up with an efficient subset of all of the different Gabor kernels at the found locations.

3.1 Kernel Location Selection

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.

3.2 Kernel Frequency and Orientation Selection

After finding useful kernel locations, similar feature selection methodology as in the previous part should be carried out in order to eliminate irrelevant feature dimensions. For this purpose, we have applied another layer of SFFS-type feature selection mechanism to the outputs of the location selection module. Frequency and orientation selection module also works in a supervised manner, and it produces a subset of a useful frequencies and orientations at each kernel location. Since, we use all the information produced by location selection module together, without dividing them according to the kernel locations, this methodology also produces an almost optimal subset by taking into account the dependence of each feature dimension.

4 Experiments and Results

In our experiments, we have used a subset of the FERET face database [11]. The used part of the database contains normalized frontal images of 146 subjects. Each subject has 4 gray scale images of resolution 150×130. 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 Kernel Location Selection

BIF, SFS, and SFFS based 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.

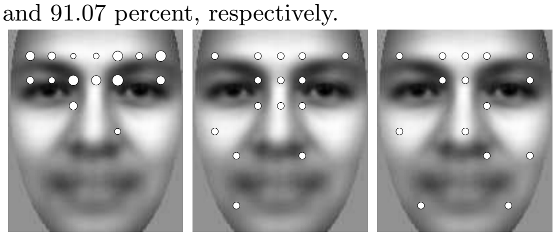
In all of the three approaches, most of the selected grid points are at the upper part of the face region. This result is largely due to the expression variations present in the dataset especially, in the mouth region. In SFFS, the combination of features extracted from eyebrows, the lower-center part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

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

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

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.



(a) (b) (c)

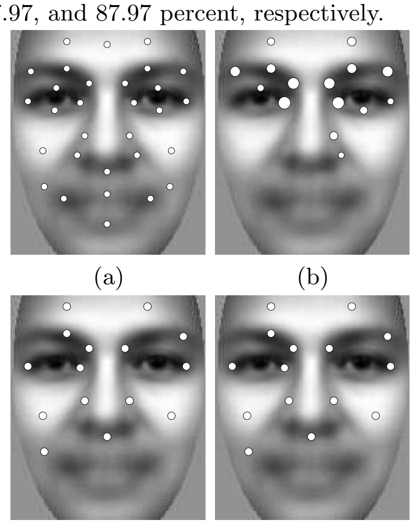
2.c, the locations of 15 useful fiducial points are shown. The points selected for SFS and SFFS methods are the same, and they are generally at the upper face region. Eyebrows, the corners of eyes, forehead, cheeks, and the outline of nose seem to carry the most discriminative information.

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 periphery of fiducial points.

Genetic algorithm-based 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 [10].

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 possible. 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

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.



(a) (b) (c) (d)

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.

4.2 Kernel Frequency and Orientation Selection

In the second part of our two-level feature selection approach, we determine the most useful orientations and frequencies of the selected kernel locations using SFFS-based methodology. The output of the kernel location module is a feature vector of size 15×40, where each kernel contributes only one jet having a dimension of 15. In frequency and orientation selection part, we further study an efficient subset of the output of the kernel location module. Therefore, in applying SFFS, we start with an empty set of selected features, and gradually add additional features. Note that, each added dimension corresponds to a specific

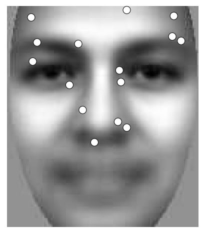


Fig. 3. The locations of important facial feature combinations for genetic algorithm.

frequency and orientation pair of the outputs of a previously selected Gabor kernel at some specific location. Again, the feature selection criteria in SFFS is the supervised classification accuracy of the selected subset.

Using this policy, we can extract a better feature subset of the original set, because of the large dimensionality of the original set. In order to find the near optimal subset, we have forced the SFFS algorithm to find a subset of dimension 600. Then, we select the minimal subset having a peak performance. In this way, we improved the performance of the face recognition system on the test set from 96.50 percent to 99.32 percent, by using a subset of dimensionality 230 out of 600.

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. As a second phase, we have introduced a selection of frequency and orientation parameters using a sequential floating subset search. By extracting useful frequencies and orientations at specific face locations, we have eliminated the irrelevant parts of the original feature vector, and also improved the recognition performance significantly.

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, we will extend this methodology for pose invariant face recognition. Another important research direction would be to analyze the aging effects where facial regions and their characteristic features which could be invariant under age variations may be found using the proposed architecture.

References

1. Chellappa, R., Wilson, C.L., Sirohey, S.: Human and machine recognition of faces: A survey. Proceedings of IEEE **83** (1995) 705–740
2. Gokberk, B., Akarun, L., Alpaydin, E.: Feature selection for pose invariant face recognition. In: Proceedings of the 16th Iinternational Conference on Pattern Recognition. (2002)
3. Gokberk, B., Irfanoglu, M.O., Akarun, L., Alpaydin, E.: Optimal gabor kernel location selection for face recognition. In: Proceedings of the IEEE International Conference on Image Processing. (2003)
4. Kruger, N.: An algorithm for the learning of weights in discrimination functions using a priori constraint. IEEE Transactions on Pattern Analysis and Machine Intelligence **19** (1997) 764–768
5. Kruger, N., Potzsch, M., Malsburg, C.: Determination of face position and pose with a learned representation based on labelled graphs. Image and Vision Computing **15** (1997) 665–673
6. Duc, B., Fischer, S., Bigun, J.: Face authentication with Gabor information on deformable graphs. IEEE Transactions on Image Processing **8** (1999) 504–516
7. Kalocsai, P., Malsburg, C., Horn, J.: Face recognition by statistical analysis of feature detectors. Image and Vision Computing **18** (2000) 273–278
8. Grudin, M.: On internal representations in face recognition. Pattern Recognition **33** (2000) 1161–1177
9. Jain, A., Duin, R.P.W., Mao, J.: Statistical pattern recognition: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 4–37
10. Kudo, M., Sklansky, J.: Comparison of algorithms that select features for pattern classifiers. Pattern Recognition **33** (2000) 25–41
11. Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.: The feret database and evaluation procedure for face recognition algorithms. Image and Vision Computing **16** (1998) 295–306