Feature-Based Face Recognition

Erik Hjelmås*
Department of Informatics
University of Oslo
P. O. Box 1080 Blindern, 0316 Oslo, Norway
email: erikh@hig.no

Abstract

In this paper we propose a face recognition system based on local features. Interesting feature points in the face image are located by Gabor filters, which gives us an automatic system that is not dependent on accurate detection of facial features. The feature points are typically located at positions with high information content (such as facial features), and at each of these positions we extract a feature vector consisting of Gabor coefficients. We report some initial results on the ORL dataset.

1 Introduction

The problem of face recognition in cluttered scenes, is a typical task in which humans are still superior to machines. Several algorithms have been proposed and implemented in computers [2], but most of them rely on accurate detection of facial features (e.g. exact location of eye corners) or some kind of geometrical normalization. In this paper we take a more general approach, and propose an algorithm which relies only on general face detection.

Face recognition systems developed so far can roughly be classified as belonging to the "holistic feature" category or in the "local features" category. Most of the systems in the "holistic feature" category are based on PCA (eigenfaces) or similar approaches (fisherfaces, LDA or PCA combined with LDA) [8] [9] [1]. The "local features" category mostly consist of systems based on graph matching or dynamic link matching [4] [6] [5], or some derivative from these. Our system also belongs in this category and uses the same approach for feature description as most other "local features" system do, namely Gabor coefficients.

The main advantage with our approach is that it is automatic. This is not entirely true, since we assume general face detection, but it is automatic in the sense that we have no need for accurate detection of facial features or geometrical normalization of the input face images. Since we find all local features automatically, and make no demand on the features extracted to be pre-determined facial features (such as eyes/nose/mouth), we also claim to have a robust strategy. By this we mean that we do not base our system on locating specific facial features, we base our system on locating feature points in the face images which contain interesting information. These are usually located around facial features, but does not have to be. Our approach is also easily extendible to face recognition in videoseqences and situation where the training data is of varying quality and quantity.

In the next section we present the datasets we report results on. Section 3 provides an overview of our proposed face recognition algorithm, while section 4 and 5 goes somewhat more into details of our approach to feature detection, extraction, and recognition. Our experimental results are presented in section 6, while we conclude og discuss future directions in section 7.

2 Datasets

To initially test our system, we have used the ORL (Olivetti Research Laboratoy) dataset. The ORL dataset consists of 10 face images from 40 subjects for a total of 400 images, with some variation in pose, scale, facial expression and details. The resolution of the images are 112×92, with 256 gray-levels.

3 Overview of the algorithm

We have two separate algorithms for training and testing (algorithm 1 and 2). However, the basic procedure is equivalent in both algorithms. When processing a face image (either for training or testing), we filter the image with a set of Gabor filters as described in the next section. Then we

*Also with Faculty of Technology, Gjøvik College, P. O. Box 191, 2802 Gjøvik, Norway.

multiply the filtered image with a 2-D Gaussian to focus on the center of the face, and avoid extracting features at the face contour. This Gabor filtered and Gaussian weighted image is then searched for peaks, which we define as interesting feature points for face recognition. At each peak, we extract a feature vector consisting of Gabor coefficients and we also store the location and class label. A visualized example from the testing algorithm is shown in figure 1.

Algorithm 1 The training algorithm.

1. Gabor filter face image
2. apply Gaussian weighting
3. locate peaks in image
4. extract feature vector at located peaks
5. if this is first training image of subject, store feature vector, location and class label for all extracted peaks, else store only those who are misclassified (with respect to the current gallery)

Algorithm 2 The testing algorithm.

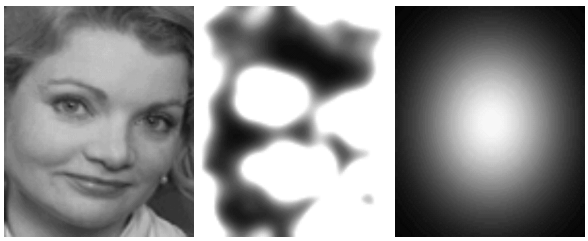
1. - 4. same as training algorithm
2. for each extracted feature vector, compute distance to all feature vectors in gallery
3. based on class label to the nearest matching feature vectors, assign points to corresponding class

4 Feature detection and localization

We wish to locate points in the face image with high information content. Typically these are located around the eyes, nose and mouth. We are not interested in the face contour or hair, since we know that the most stable and informative features in the human face are located in the center of the face. To enforce this we apply Gaussian weighting to focus our attention on the center of the face.

4.1 Filtering for information

There are several possible ways to locate points with high information content. We intend to pursue and explore several approaches in the extension of this work, but for now we have selected the Gabor filter appraoch. The Gabor filters $\psi_{\boldsymbol{k}}$ are generated from a wavelet expansion of a Gabor kernel, parameterized (determining the wavelength and orientation) by the vector $\boldsymbol{k}$ [4]:



(a) original image (b) Gabor filtered image (c) 2-D Gaussian

(d) Gabor*Gauss (e) located feature points

Figure 1. Example of the face recognition procedure (training algorithm). In (e), the black cross on a white background indicates an extracted and stored feature vector at this location while white cross on black bacground indicates an ignored feature vector.

$$\psi_{\boldsymbol{k}}(\boldsymbol{x}) = \frac{\boldsymbol{k}^2}{\sigma^2} e^{-\frac{\boldsymbol{k}^2 \boldsymbol{x}^2}{2\sigma^2}} \left(e^{i\boldsymbol{k}\boldsymbol{x}} - e^{-\frac{\sigma^2}{2}} \right), \qquad (1)$$

where

$$\boldsymbol{k} = \begin{pmatrix} k_\nu \cos\phi_\mu \\ k_\nu \sin\phi_\mu \end{pmatrix}, \quad k_\nu = 2^{-\frac{\nu+2}{2}}\pi, \quad \phi_\mu = \mu\frac{\pi}{8}. \qquad (2)$$

For filtering we use kernels of three sizes ($\nu \in \{1, 2, 3\}$) and eight orientations ($\mu \in \{0, ..., 7\}$), manually selected based on image resolution. The resulting filtered image consists of the sum of the magnitudes of the Gabor filter coefficents at each location in the image. The magnitude from the filter is good to apply in this case due to its smoothly changing behavior [4]. A typical Gabor filtered image is shown in figure 1b. After Gaussian filtering this image, we end up with the image in figure 1d, which exhibits the kind of behavior we are looking for (namely high values around interesting locations such as the eyes nose and mouth).

4.2 Selecting maximas

To select the maximas (peaks) from the image in figure 1d we scan the image with a $\omega \times \omega$ window and select locations satisfying the following constraints:

1. center pixel value is larger than all other pixel values in window
2. all pixel values are above average pixel value in the entire face image

At each such location we apply the feature extraction procedure described in the next section. Initially we have manually selected $\omega = 7$.

5 Representation and recognition

5.1 Feature extraction and representation

At each located feature point (peak) $\boldsymbol{x}$ we would like to extract a feature vector which describes the local neighbourhood surrounding $\boldsymbol{x}$. We choose to do this in the same manner as the system of Lades et al. [4], namely with a set of Gabor filters (similar to the filtering procedure described in the previous section, but with different kernel sizes). For the Gabor filtering we used kernels of three sizes. For the feature extraction and representation we use an additional smaller and an additional larger kernel which gives us a 40-dimensional feature vector consisting of the magnitudes of the complex Gabor coefficients.

During training, the extracted feature vectors are stored in the gallery accoriding to the rule "if this is the first training image of the subject, store all extracted feature vectors, if not: for each extracted feature vector find the closest feature vector already stored in the gallery and compare the class labels, if the class labels match ignore the feature vector, otherwise add to the gallery". The reason for applying this rule is that we do not have any intention of establishing classes in feature space (the gallery). We want a gallery consisting of single local feature points with high information content, thus we want to add only feature vectors which add new information to the gallery (and not "overpopulate feature space"). The rule above is a rough approximation to this, but some initial experiments have shown that performance does not decrease due to this rule.

5.2 Classification

Initially we compute Euclidian distance to all feature vectors in the gallery and rank them accordingly. Furthermore, we apply two strategies for classification. One is to see how well we can do with a single feature point. In this case we select the feature vector (of all the extracted feature vectors in the current test image) which is closest to (in Euclidian distance) some feature vector in the gallery and assign this test image the class label of that feature vector. The second strategy involves all the extracted feature vectors in the test image. In this case (for each feature vector) we select a number n of the closest feature vectors in the gallery, and assign (award) points to the possible classes ranging from n (best match) to 1. For each class we sum the number of points assigned by each extracted feature vector. The test image is assigned the class label of the class with the most points.

6 Experimental results and discussion

In [3], we examined how well we could do in face recognition if the only available information were the eyes. On the ORL dataset, we reported 85% correct classification with feature vectors consisting of Gabor coefficients, which was the best result compared to eigenfeatures (PCA) and gray-level features. From this we know that we can do reasonably well with only a few feature points, but of course we want to use as much information as possible.

Figure 2 shows results when only the single best matching feature vector is used (as described in the previous section). The results are reported in terms of cumulative match score [7], and all results are the average of 10 trials where the dataset is randomly divided into 5 images for training and 5 images for testing (for each subject). We observe that this does not give satisfactory performance (only 76.5% for rank = 1), which is expected considering we only use a very small amount of information from the face image. In this situation the classification is based only on the match of a single automatically extracted feature point in the image to

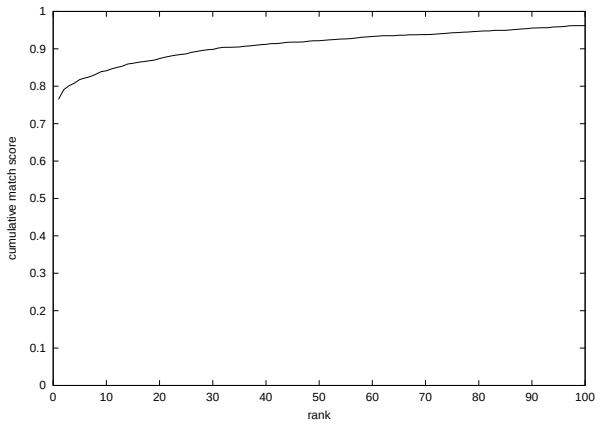


Figure 2. Performance when only the best matching single feature is used for classification.

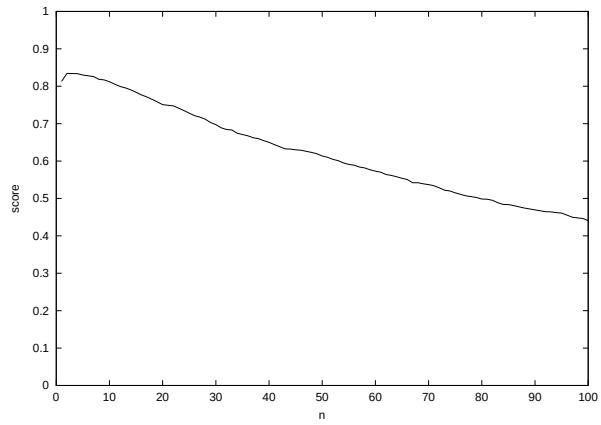


Figure 3. Performance when we assign points from all extracted feature vectors.

a stored one in the gallery (e. g. only a small neighbourhood surrounding the nose tip).

In our second strategy, where we make use of all located feature points, we would expect better performance. Figure 3 shows the results in this case, where "score" is percentage correct classification and n indicates the number of classes we assign (award) points to. We observe that performance is better (83.4% for $n \in \{2, 3, 4\}$) than for the single feature strategy, but not as good as expected. These results are not yet competitive with the best performing systems, since many systems perform 95% or better on the ORL dataset. Our task at hand is to find a better trade off between extracting information and avoiding noise. As we see from figure 3, performance increases when we increase n from 1 to 2, but decreases whenever $n > 4$ increases, which means that utilizing the lower ranked feature vectors only adds noise to the system at this level.

7 Conclusions and future work

We have presented a feature-based face recognition system together with some initial experimental results. Future work will involve detailed exploration of each component in the system, specifically parameter tuning and classification strategies.

Acknowledgements

We would like to thank the Olivetti Research Laboratory for compiling and maintaining the ORL face-database, and Rama Chellappa, Fritz Albregtsen and Jørn Wroldsen for useful discussions during the preparation of this work. This work was supported by the Research Council of Norway and was partially done while E. Hjelmås was visiting Center for Automation Research, University of Maryland.

References

[1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 1997.

[2] R. Chellappa, C. L. Wilson, and S. Sirohey. Human and Machine Recognition of Faces: A Survey. *Proceedings of the IEEE*, 83(5), 1995.

[3] E. Hjelmås and J. Wroldsen. Recognizing Faces from the Eyes Only. In *Proceedings of the 11th Scandinavian Conference on Image Analysis*, 1999.

[4] M. Lades, J. C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Würtz, and W. Konen. Distortion Invariant Object Recognition in the Dynamic Link Architecture. *IEEE Transactions on Computers*, 42(3), 1993.

[5] B. S. Manjunath, R. Chellappa, and C. von der Malsburg. A Feature Based Approach to Face Recogntion. In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1992.

[6] K. Okada, J. Steffens, T. Maurer, H. Hong, E. Elagin, H. Neven, and C. von der Malsburg. The Bochum/USC Face Recognition System and How it Fared in the FERET Phase III Test. In *Face Recognition: From Theory to Applications*. Springer-Verlag, 1998.

[7] P. J. Phillips, H. Moon, P. Rauss, and S. A. Rizvi. The FERET September 1996 Database and Evaluation Procedure. In *Proceedings of the First International Conference on Audio and Video-based Biometric Person Authentication*, 1997.

[8] M. Turk and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3:71–86, 1991.

[9] W. Zhao, A. Krishnaswamy, R. Chellappa, D. L. Swets, and J. Weng. Discriminant Analysis of Principal components for Face Recognition. In *Face Recognition: From Theory to Applications*. Springer-Verlag, 1998.