Proc. Vision Interface (VI'2000), Montreal, Quebec, Canada, May 2000.

Face Recognition in Fourier Space

Hagen Spies
Interdisciplinary Center for Scientific Computing, Univ. of Heidelberg
INF 368, 69120 Heidelberg, Germany
Hagen.Spies@iwr.uni-heidelberg.de
and
Ian Ricketts
Department of Applied Computing, Univ. of Dundee
Dundee DD1 4HN, UK
ricketts@computing.dundee.ac.uk

Abstract

This paper describes a simple face recognition system based on an analysis of faces via their Fourier spectra. Recognition is done by finding the closest match between feature vectors containing the Fourier coefficients at selected frequencies. The introduced method compares favourably to three other competing approaches implemented on the same database.
Keywords: Face recognition, Fourier coefficients, Neural Networks, Eigenfaces

1 Introduction

The face is one of several features which can be used to uniquely identify a person. It is the characteristic that we most commonly use to recognise others and it plays a vital role in our social interactions. No two human faces are identical which makes them well suited for use in identification schemes, in the same way as fingerprints or DNA samples are used.

Besides being a challenging problem in itself the importance of face identification systems lies in their potential applications (access control, passport and personal identification, etc.). The obvious advantage of a face recognition system compared to competing methods is its low level of intrusion. It only requires looking into a camera.

Automated face recognition systems generally evolved along two main routes, either the analysis of grey level information (often called template based) or the extraction of mainly geometrical features (such as shape, profile or hair colour). It can be expected that the most reliable systems will combine both approaches and [5] describes how the combination of shape and grey level information can achieve better results than each approach used separately.

The most successful systems are reported to correctly classify previously unseen faces to belong to the correct person in 95-100% of the cases [5, 8, 11, 12, 2, 1]. As [6] point out it is, however, not sufficient to simply compare these percentages. Mainly because each database used will be assembled under different constraints regarding lighting, head tilt, facial expressions, gender and the ethnic origin of the subjects.

The work presented here comprises a novel template based approach that achieves 98 percent correct recognition. Considering it's simple algorithm this compares very well to other methods that have been used on the same database of facial images. Those methods consist of a system using Hidden Markov Models [8], a principal components analysis approach termed *eigenfaces* [11] and a backpropagation neural network that was implemented by the authors for comparison purposes.

According to [7] humans are thought to view faces primarily in a holistic manner and experiments suggest that holistic approaches are superior to geometrical recognition systems [3]. Our technique is based on the Fourier spectra of facial images, thus it relies on a global transformation, i.e. every pixel in the image contributes to each value of its spectrum. The Fourier spectrum is a plot of the energy against spatial frequencies, where spatial frequencies relate to the spatial relations of intensities in the image. In our case this translates to distances between areas of particular brightness such as the overall size of the head, or the distance of the eyes. Higher frequencies describe finer details and we found them less useful for identification of a person, just as humans can recognise a face from a brief look without focusing on small details.

The recognition of faces is done by finding the closest match, in a Euclidean sense, between a newly presented face and all those faces known to the system. The distances are calculated between feature vectors with entries that are the Fourier transform values at specially chosen frequencies. As few as 27 frequencies yield excellent results (98%), this

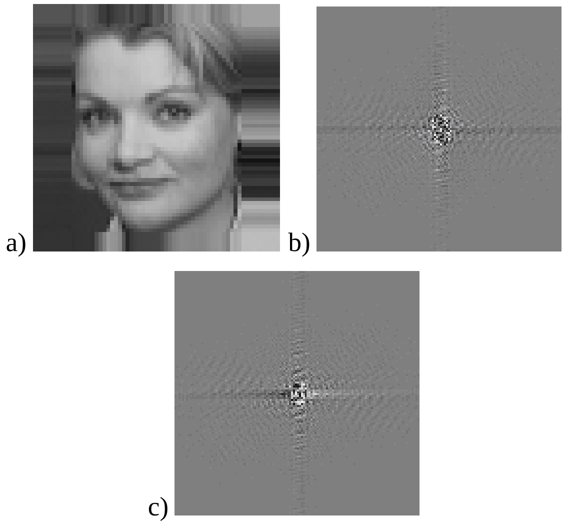


Figure 1: Padded Image (a), real (b) and imaginary (c) spectra.

small feature vector combined with the efficient Fast Fourier Transform makes our system extremely fast.

2 The Recognition System

2.1 Fourier Transformation

Let the dimensions of an image be N and M. There are two frequencies u and v corresponding to the two coordinates x and y. If $f_{x,y}$ is the grey value at location (x,y) then the two-dimensional discrete Fourier transform values $f'_{u,v}$ are given by:

$$f'_{u,v} = \sum_{y=0}^{M-1} \sum_{x=0}^{n-1} f_{x,y} e^{2\pi j (\frac{xu}{N} + \frac{yv}{M})} \qquad j = \sqrt{-1} \tag{1}$$

As the actual implementation was achieved using the Fast Fourier Transformation an additional requirement was that the image dimensions be a power of two. The resolutions used in the available database of images were 46x56 and 92x122. Thus the images had to be padded to sizes of 64x64 and 128x128 respectively. Padding was achieved by repeating the outermost grey values until the desired dimensions were reached. This type of padding minimised the introduction of sharp changes and their attendant higher frequency content in the spectra. Figure (1) shows an example padded image (128x128) with its real and imaginary spectra. As the images are real both spectra are symmetric around the origin which means only half of each carries valuable information.

2.2 Selecting Frequencies

From the spectra it can be seen that almost all the information is contained near the centre, i.e. within the low frequencies. Thus it seems reasonable to assume that these frequencies will also provide the best ground for a recognition system. Even though appealing on first sight this needs some further justification. It is not necessarily the case that those frequencies which contain most of the information within an image will also carry most information about the differences between the images - and all a recognition system needs to know are the differences. Therefore the variance of each frequency across the whole training set was calculated to identify which frequency values varied the most. The variance is given by:

$$Var(x_1, ..., x_N) = \frac{1}{N-1} \sum_{k=1}^{N} (x_k - \overline{x})^2 \tag{2}$$

$$\text{with} \quad \overline{x} = \frac{1}{N} \sum_{l=1}^{N} x_l \tag{3}$$

The next step is then to order the frequencies with respect to their variance and only use the earlier frequencies - how many will depend on the implementation. However, examining which frequencies do actually vary most one finds that it is in fact the low frequencies. Figure (2) (a) and (b) show the location of the first 200 thus ordered frequencies in the spectra - they are shown as white pixels while the remainder are black. Indexing the frequencies according to the above order gives the first basic set-up for our system, this arrangement will be called *variance based*.

When looking at Figure (2) it can be seen that the valuable frequencies do not lie in a circle around the origin but more in a rhombus shaped region. As mentioned above one half of the spectra contains all the information. However further reduction is possible, the highest variance is found directly on the axes, where one of the frequencies is zero. As we use only half of the spectra only one axis is entirely present. However, because the image is real there is an additional symmetry along this axis, i.e. only half of it holds useful information.

As the most important part (the axis) does not contribute to the information content anyway it then seems a promising idea to totally discard another quadrant. Hereby reducing the amount of data still further.

One might argue that a face is in fact an even picture with respect to one dimension - since both halves are almost exactly the same. This is true to some extent for perfect frontal views but most of the images considered here contain slightly tilted faces and their spectra were not found to be even.

Figure (2) (c) shows the numbering of the frequencies for only one quadrant (lower right corner). This number-

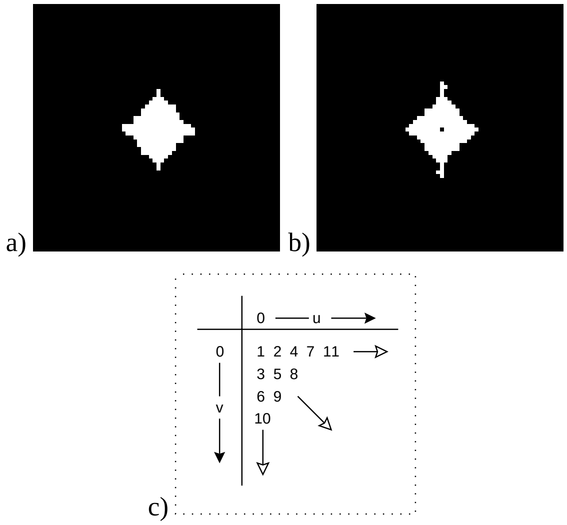


Figure 2: Most variant Frequencies: (a) real, (b) imaginary and (c) selected numbering.

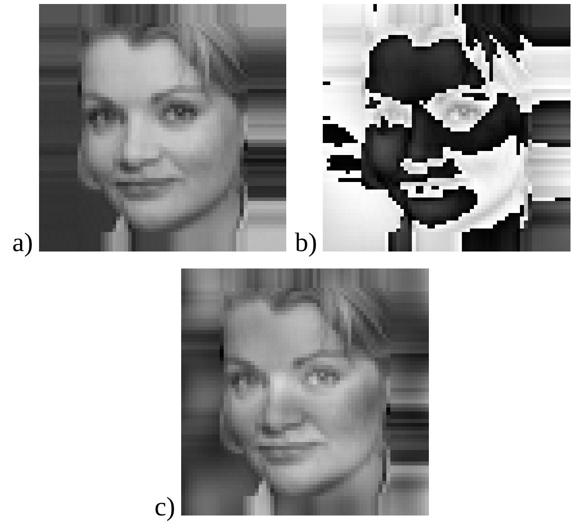


Figure 3: Effect of removing frequencies: (a) Original, (b) 5 most variant frequencies (real and imaginary part) removed and (c) after setting the imaginary part of the 40 most variant frequencies to zero.

ing is chosen for its simplicity and does not correspond precisely to the amount of variance the frequencies carried but it makes the system independent of the variance calculation, i.e. no large data set is needed for initialisation. For the imaginary part of the spectrum the (0,0) frequency is omitted as it is always zero. This set-up is called *lower quadrant*.

How much information is contained within these frequencies? To illustrate the answer to this question some of the frequencies with most variance were removed (set to zero) from the spectra followed by an inverse transformation so the effect could be observed. Figure (3) contains three such images: The first is the original; the second has the 5 most variant frequencies (real and imaginary part) removed; the last shows the removal of the imaginary part of the 40 most variant frequencies.

Inspection of these images suggests that removing relatively few frequencies from the spectra produces images that are quite difficult to identify. Thus it can be concluded that these frequencies contain valuable information for face identification. Another observation is that the real part seems to be more useful than the imaginary one.

2.3 Classifying Faces

Having established which frequencies to use the next task is to classify the faces. Each frequency (real or imaginary part) gives a value that is used as an entry in a feature vector. Choosing frequencies equates to the selection of axes onto which the faces are projected to give the coordinates i.e. the entries in the feature vector.

As the functions (combinations of cosines and sines) onto which the images are projected are orthonormal they form an Euclidean space. Hence classification can be done by means of the Euclidean distance. The distances between the feature vector of a presented test face and all the feature vectors in the training set are calculated. The smallest distance then gives the closest match. If that match belongs to the same person the classification was successful. This allows the use of a threshold in order to gain some confidence in the decision, if the distance of the closest match is greater than the selected threshold the system rejects the test image. Use of a threshold makes it possible to reject faces that do not belong to the trained subjects - a very important property for a face recognition system.

There is no apparent reason why the feature vectors of different faces should be linearly separable. Therefore a non-linear classifier was also constructed using a backpropagation neural network and its performance compared to that of the linear Classifier.

3 Experiments

3.1 The database

The experiments used the Olivetti Research Ltd (ORL) face database. The database contains 400 images compiled from 40 subjects with 10 images each. The images were taken against a dark homogeneous background with varying lighting conditions. The faces are in a frontal upright position and show a range of expressions. Side movement and head tilt were tolerated to a limited extent only. The subjects include 4 females and 36 males within the age range of 18 to

Figure 4: Three example face images.

71 years but most subjects are between 20 and 35 years. The original resolution of the images was 92x112, 8-bit grey levels. In order to speed up the classification and to reduce the amount of data needed (in particular with respect to neural networks) this resolution was reduced to 46x56 by simply averaging four neighbouring pixels. Figure (4) shows three examples of the database. Of the 10 images per subject 5 were selected at random for training and the remaining 5 were used for testing, i.e. 200 training and 200 test images.

3.2 Classification

Experiments using pure classification (i.e. without thresholding) showed that the best results are found for the variance based system if 22 real and 8 imaginary frequency values were used. The resulting correct classification was 98% (196 out of 200). Several different combinations of frequencies from the lower quadrant system yielded the same result. The selection requiring the smallest number of variables used 15 real and 12 imaginary frequencies. In the latter arrangement using as few as 5 real plus 5 imaginary frequencies lead to 95% correct identification, i.e. 10 variables are sufficient for classification.

If instead of the Euclidean classifier a neural net was used (on the same feature vectors) the performance was found to drop by 2 to 3.5 percent for the optimum results. Those networks were fully connected with a single hidden layer and used error backpropagation.

To see if the above mentioned reduction in the resolution influenced the performance the system was tested on the original resolution (92x112) as well. No change in performance was observed.

It seems reasonable to assume that the choice of training and test images could influence the performance. To see if that was the case the system was tested using two other randomly chosen divisions of training and test sets. In both experiments the correct identification remained at 98%.

For some applications it may be desirable for the system to be adjusted to make 0% misclassifications. That is rather than classify wrongly to reject some of the presented faces. It was found that it is indeed possible to obtain 0% misclassifications, however the correct classification percentage was reduced to between 65.5% to 82% depending on the set-up.

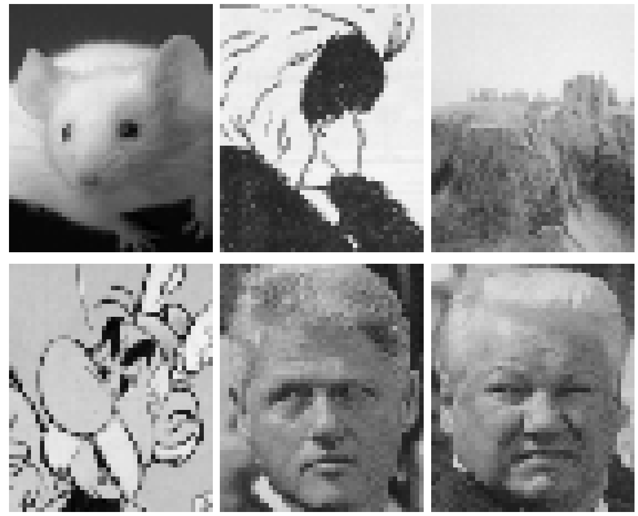


Figure 5: Images used to detect the presence of a face.

3.3 Detection of Faces

To see if the system is able to detect the presence of a face in a given image the images shown in Figure (5) were presented to the system. The distances between the closest match in the training set and the additional images were found to be significantly larger for the first three images. Interestingly the mouse was found to be closer to the human faces than the other two images. The nearest match was found to be closer for the last two additional images, which contain previously unseen faces. The distances were comparable to the largest ones encountered using the database.

However, those distances are far larger than the average and the threshold can easily be set such that these faces were not identified as belonging to any of the trained subjects. It is interesting to note how the system reacts to the cartoon. While the smallest distance normally lies in the same range as the other non- facial images it comes close to the two faces if the (0,0) frequency is not taken into account. This frequency encodes the mean grey value. Clearly the cartoon is much brighter than the other images. That can be taken as another hint that it is indeed on a rather coarse scale recognition of faces takes place [2].

3.4 Unknown Faces

For many uses of a face recognition system, including all access applications, it is essential that no strangers are mistaken for one of the persons in the training set. To assess how well our system meets this requirement one person's facial images were excluded from the training data and all 10 of the person's images were used for testing. This was done with three different subjects chosen at random. The limited number of test cases used in this experiment means that the results are only indicative. However, the results indicate whether such a rejection is possible and what performance penalties have to be expected.

The definition of acceptable performance depends on the final application and dictates what an appropriate test of the system would be. In our situation we have 200 examples of known and 30 examples of unknown faces, thus the incorrect acceptance of one single unknown face will have a particular statistical significance. This might be appropriate if access to a building is to be controlled, however for systems like verifying access to bank cash machine, a rejected known face is nearly as undesirable as an accepted unknown face.

We decided to simulate an application where access to a building is controlled by a face recognition system. The problem is then reduced to a simple yes or no - either known or unknown. If a known face is correctly or wrongly classified is not taken into account but treated separately as described above.

4 Other Methods

4.1 Eigenfaces

The eigenface method has been used for comparison purposes as it was one of the most successful algorithms available. This method was developed at M.I.T. by [11]. In what is called Principal Component Analysis a few parameters used for representation are extracted from a face. Those parameters are obtained by a projection of the face onto a coordinate system given by the eigenvectors of the covariance matrix of the training set. These eigenvectors, themselves images, are termed eigenfaces and span a vector space called face space. Each face is then encoded by means of its coordinates in this face space. Matching of two faces corresponds to a calculation of the Euclidean distance between their face space representations.

It was found that using 65 eigenfaces gives the best results (with as few variables used as possible) on the images used here. The classification was correct in 94% of the cases.

4.2 Neural Networks

For comparison neural networks were investigated to identify faces. The nets used were trained using the backpropagation algorithm. Each face image comprised 2576 pixels (46x56) and each pixel was assigned to one input node. As there are 40 subjects in the database the nets have 40 output nodes. The performance of the networks was considerably improved if instead of the original grey values the

Figure 6: Average face and example difference faces.

differences to a previously calculated (from the training set) average face were used as input [9]. The average face plus two examples of original and difference images are shown in Figure (6). The most successful networks were fully connected and contained a single hidden layer. Correct classification was 94% with 30 neurons in the hidden layer, 96.5% using 60 hidden units and 94% when using 70 hidden neurons. To reduce the classification error and exclude the mistaken acceptance of unknown faces a threshold for the winning output node can be used.

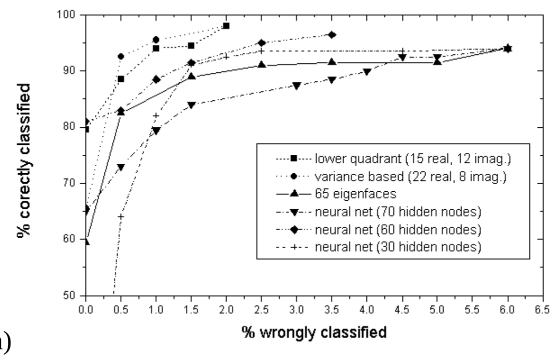
5 Comparison of the Methods

If a recognition system is to be used in a fully automated environment it is of paramount importance that the number of errors made are minimised. We therefore analysed how the performance of the various systems varied as the error rate is reduced through the application of an acceptance threshold.

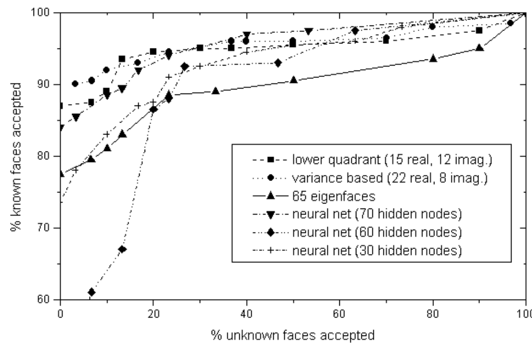
Figure (7) a) shows a graph of the percentage of known faces correctly accepted against the percentage of unknown faces wrongly accepted for all three methods provided with images from the ORL database.

It can be seen that the Fourier Transform based system developed by the authors shows superior performance over the whole range. The neural networks show rather inconsistent results and with some of them it is not possible to eliminate all errors. However, one of them performs better than the eigenface method.

Figure (7) b) contains a graph of the performances of the three methods when they are confronted with unknown faces as described in Sect. 3.4. Again the Fourier Transform based system produces the best results. However, the neural net with 70 units in its hidden layer does achieve comparable results. The other networks and the eigenface method



a)



b)

Figure 7: (a) Classificaton results and (b) Rejection results.

perform less well.

It is remarkable that all the methods are capable of distinguishing between known and unknown faces. If no error is tolerated then the best systems are still able to grant access to known persons in more than 85 percent of the cases. As it was said before these figures should not be seen as absolute values as the number of test cases was limited, however, they show how the implemented methods compare to each other.

Recently other more involved algorithms produced comparable and even better results on the ORL database. As they have not been implemented by us the only available figure is the pure classification rate: 95% using Pseudo 2D Hidden Markov Models on the grey values [8]; 98.3% based on fractal transformations [10]; 99.5% with Hidden Markov Models used on locally computed coefficients of the discrete cosine transformation [4]; 100% using Pseudo 2D-HMM also on discrete cosine transformation coefficients [1].

6 Conclusion

A new approach to face recognition was presented which does achieve excellent recognition results both as a classifier and when confronted with faces not belonging to any of the persons used for training. The method is considerably less complicated than its competitors when one considers how well the Fourier Transformation is understood and the availability of many routines for its fast implementation. However, given the size of the database used the results mainly serve the purpose of introducing a new technique and to provide some clues as to its performance.

The number of rejections that have to be tolerated if an incorrect acceptance is not to occur is rather high. It has to be said that at least for security applications there will still be a human required to take the final decision. Yet a combination of the holistic approach presented here with a method based on the extraction of geometrical features might lead to further increases in reliability.

Acknowledgements

The authors wish to thank the Olivetti Research Ltd. for the collection and distribution of the face database used for this work. We also wish to thank the M.I.T. Media Lab for providing access to the code for the eigenface implementation.

References

[1] S. Eickeler, S. Müller, and G. Rigoll. Gesichtserkennung mit hidden markov modellen. In *Proc. DAGM*, Bonn, 1999.

[2] F. Goudail et al. Face recognition system using local autocorrelations and multiscale integration. *PAMI*, 18(6):1024–1028, 1996.

[3] S. Gutta et al. Benchmark studies on face recognition. In *Proc. Int. Workshop on Automatic Face- and Gesture-Recognition*, Zürich, 1995.

[4] V.V. Kohir and U.B. Desai. Face recognition using dct-hmm approach. In *Workshop on Advances in Facial Image Analysis and Recognition Technology*, Freiburg, 1998.

[5] C. J. Taylor Lanitis, A. and T. F. Cootes. An automatic face identification system using flexible appearance models. In *British Machine Vision Conference 1994*, pages 65–74, 1994.

[6] G. Robertson and I. Craw. Testing face recognition systems. In *British Machine Vision Conference 1993*, volume 1, pages 25–34, 1993.

[7] A. Samal and P. A. Jyengar. Automated recognition and analysis of human faces and facial expressions: A survey. *Pattern Recognition*, 25:65–77, 1992.

[8] F. Samaria. *Face Recognition Using Hidden Markov Models*. PhD thesis, University of Cambridge, 1994.

[9] H. Spies. Face recognition - a novel technique. Master's thesis, University of Dundee, 1995.

[10] T. Tan and H. Yan. Face recognition by fractal transformations. In *Int. Conf. on Acoustic, Speech and Signal Processing*, Phoenix, 1999.

[11] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[12] L. Wiskott et al. Face recognition and gender determination. In *Proc. Intern. Workshop on Automatic Face- and Gesture-Recognition*, Zürich, 1995.