Comparison of Feature Extraction Techniques in Automatic Face Recognition Systems for Security Applications

S. Cruz-Llanas, J. Ortega-García, E. Martínez-Torrico, J. González-Rodríguez

Dpto. Ingeniería Audiovisual y Comunicaciones, EUIT Telecomunicación, Univ. Politécnica de Madrid, Spain

{cruzll, jortega, etorrico, jgonzalz}@atvs.diac.upm.es. http://www.atvs.diac.upm.es

ABSTRACT

This paper is aimed to analyze the performance of two different state-of-the-art automatic face recognition systems. One of the key issues regarding face recognition is the election of convenient features for representing identity in facial images. Multivariate analysis and Gabor analysis are widespread alternatives for accomplishing this feature extraction stage. Consequently, two different approaches to the face recognition problem, one based in multivariate analysis, whereas the other based in Gabor analysis, are proposed. A brief review of the theoretical foundations of both systems, together with some tests, conducted for comparing them, are addressed in this paper.

INTRODUCTION

Biometrics-based person recognition is currently one of the key issues in security applications. Many biometric signals (speech, iris, fingerprint, signature, etc.) are being used in this field. In this paper, we concentrate on automatic face recognition, which is one of the less expensive modalities (in terms of user constraints and cost of acquisition devices) and the closest one to visual human recognition of other human beings.

Biometrics-based person recognition can be generically considered as a particular case of the pattern recognition problem, and many techniques have been described and developed to cope with it. Nevertheless, prior to the pattern matching stage, a feature extraction stage is mandatory, in order to obtain the face characteristics and to accomplish the recognition task. As a result of this feature extraction stage, it would be desirable to have a simple and reliable representation of the input signal (eliminating so the redundant information), but retaining, at the same time, all the important cues for recognition.

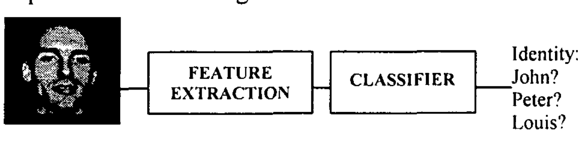

Figure 1.- Block-diagram of a face recognition systems in terms of a generic pattern recognition problem.

The input to our system, a gray-level image, is a high memory-consuming signal, so it is quite difficult to deal with it. Furthermore, in our particular problem, variability produced by factors like illumination conditions, pose, expression, etc., is important. Taking this into account, we can state that compact and invariant representations to all these factors are required for efficient face recognition.

Two types of techniques have been widely used for feature extraction in face recognition. These are techniques based on multivariate analysis and Gabor analysis. Our goal here is to compare their suitability in different situations. For doing so, we will start by describing the two systems (one based on each technique) that we have been using. We will describe afterwards some experiments that we have conducted in order to compare them. Conclusions to this work and future research lines will be finally presented.

SYSTEM BASED ON MULTIVARIATE ANALYSIS

Multivariate analysis ([1]) is an important branch of statistics whose purpose is to study random systems with more than one random variable of interest. It is so a broad area of knowledge which provides useful tools in an important number of problems. In our particular problem, regarding face recognition, several of these tools have shown their utility. Our system considers (describing each separately) the following tools, namely: a) nearest neighbor classifier, b) (linear) principal component analysis, c) (linear) discriminant analysis, and d) probabilistic principal component analysis.

Figure 2.- An extract of Yale database showing different facial expressions.

1. Nearest neighbor classifier (NN) is a well known paradigm in pattern recognition problems. The

0-7803-5965-8/00/$10.00 ©2000 IEEE

basic idea and its translation to our problem are simple. In order to recognize faces from images, the system needs to 'know' these faces, so looking in our set of examples (the training dataset) will allow us to choose the closest example as belonging to the correct identity. In this section, facial images as taken as vectors, containing each the gray-level values of all the pixels. The euclidean distance between two of these vectors is employed in order to determine if the vectors are close enough. This simple idea can perform fairly well, but it has important drawbacks; for instance, the enormous computational cost of computing distances in so high-dimensional spaces (for images of 128x128 pixels, a 16,384-dimension space is required). Furthermore, if the training dataset has many examples, we have to compute lots of distances like this. Nevertheless, the most important drawback is the fact that this kind of systems are very weak when the conditions during training and testing vary, even for slight variations. The reason is that we are not really extracting any information from our signals, but simply using them as a whole. The following methods try to overcome these drawbacks, although this basic technique remains the primary reference, as it can be implicitly found in the majority of the other techniques.

2. Principal component analysis (PCA) ([2],[3]). Its use is so widespread regarding our particular problem that a specific name (eigenfaces) is usually applied to it. The basic idea is to take advantage of the redundancy existing in the training set (as all images come from faces and have common parts) for representing it in a more compact and meaningful way. Instead of keeping all the vectors from the training set, in this technique we only keep a smaller amount of vectors that linearly represent and permit the reconstruction of the training data. These vectors are the mean of the training set and a set of eigenfaces. Expansion of images in the training set in terms of eigenfaces is optimal in the sense that, between all linear transformations, it is the one that guarantees minimum euclidean distance between these images and their corresponding reconstruction.

It can be shown that eigenfaces are obtained as eigenvectors of the empirical covariance matrix of the training set. These eigenvectors have a statistical interpretation, as they define orthogonal axis (principal components) which explain the main causes of variability in the training set. Their corresponding eigenvalues take into account the relative importance of the source of variability for each eigenvector. Eigenvectors pointing to directions where the variations of the data in the training set are important have higher eigenvalues than eigenvectors pointing to less relevant directions.

Two reasons may explain the suitability of eigenfaces to the automatic face recognition problem:

- They produce a compact representation, and expansion of images in the set of eigenfaces requires only the ensemble of weights obtained when projecting the images onto the eigenset. Instead of having to classify now 16,384-dimension vectors, the number of eigenfaces is the new dimension of our space.
- More meaningful representation, as we express our data now in terms of the main directions of variability (see Figure 3). For recognizing with the NN classifier, if we keep all the eigenfaces, the resulting nearest neighbor will be the same in the low and high-dimensional spaces. Dropping eigenfaces (usually some of the last are discarded) does not imply losing performance in recognition. If the eigenfaces that we drop do not code properly the identity (but other factors as illumination or noise) the system can even gain performance. Other metrics (as Mahalanobis metrics weighted by the eigenvalues) can lead also to good results.

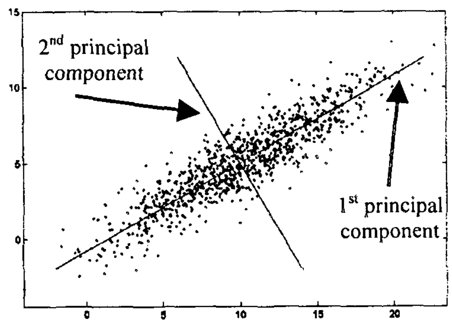

Figure 3.- Example of principal components analysis in a 2-D distrubution of data.

Figure 4.- Example of mean image (left) and 5 first eigenfaces derived from the database.

3. Linear discriminant analysis (LDA). ([4],[5]) Based on the same principles than the previous one (more compact and meaningful representations), the idea here is to change the criterion with which these goals are attained. Instead of optimal reconstruction here the purpose is that, when projecting vectors to a lower dimensional space, images from a same subject remain as close as possible, and images from different subjects as separated as possible. The mathematical formulation (eigenvectors, eigenvalues, etc.) is similar than before. The differences here are due to the fact that now, instead of considering one single covariance matrix, we consider two: the within-class scatter matrix and the between-class scatter matrix. The eigenvectors which expand the new subspace are called fisherfaces. Previous PCA is mandatory to avoid singularity problems.

Figure 5.- Example of mean image (left) and 5 first fisherfaces. Note the differences between the subjects are emphasized.

4. Probabilistic principal component analysis (PPCA). ([6],[7]) It is an extension of ordinary PCA with the additional hypothesis that the underlying distribution for the high dimensional vectors representing face images for a given subject is a multivariate gaussian. The mean of each gaussian is estimated as the mean image for the corresponding subject in the training set. However, an accurate estimation of the covariance matrix would require a lot of images per subject. This problem can be avoided if a common covariance matrix is estimated for all the gaussians, so that we have more data for estimating only one covariance matrix. As for the eigenfaces case, the size of the covariance matrix is so large and the data are so scarce that a full estimation is unfeasible, and only some principal components (intra-eigenfaces), conserving the main variations of the data, will be used for the covariance matrix estimation. For non-principal components (until a dimension of 16,384), the covariance matrix takes a model of 'white noise', so only one variance (or degenerate eigenvalue) has to be specified. For recognition, the decision is made here in terms of likelihood of the gaussians instead of distances. For an incoming image the identity of the subject producing maximum likelihood is attributed.

Figure 6.- Mean image for some of the subjects (upper row) in the primary reference experiment; note that subjects can still be recognized. 5 first intra-eigenfaces (lower row) used for the estimation of the common covariance matrix.

SYSTEM BASED ON GABOR FEATURES

In this second feature extraction approach statistics do not play a crucial role, as a more deterministic signal processing approach is applied: the problem is considered as a particular application of wavelets or multiresolution pyramids theory.

In the current implementation, the feature extraction stage consists of representing the images as multiresolution Gabor pyramids. These pyramids are composed by the set of signals obtained when applying a bank of filters to the original images and subsampling afterwards by suitable factors. The family of Gabor filters employed is designed to analyze the textures of the images in four specific orientations and three normalized frequencies in as exponential scale, as shown in Figure 7. Gabor character leads to gaussian shape of the filters in the frequency domain (isotropic in the current implementation) or, alternatively, a product between a complex exponential and a gaussian in the spatial domain.

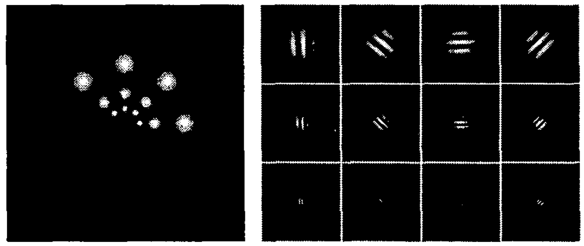

Figure 7.- Gabor filters in the Fourier domain (left). Each fiter is an isotropic gaussian. Real part of the filters in the spatial domain (right). Each row represents a resolution and each column an orientation.

For recognition, all the images (training and testing sets) are expressed as pyramids of that kind. From every image of size 128x128, we obtain 12 complex signals (Figure 8).

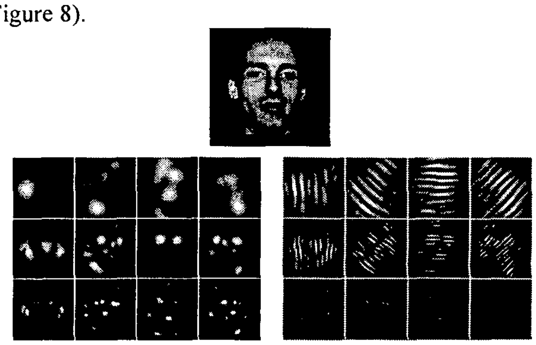

Figure 8.- Example of the application of the filter-bank to the image in the top, producing 12 complex signals. The left block of images shows modulus and the right block real part. Note that all these images can be subsampled without loss of information.

The question that arises now is: why should recognition be performed better on these new features rather than on the raw images? The operation we have performed (Gabor transform) has some interesting properties that justify this choice. It is quite invariant to some adverse factors for recognition, as variable illumination conditions. Compared to multivariate techniques in the previous section, it produces a localized representation, which means that the localization of the facial attributes has not been lost, being the relations between the positions of the different facial attributes very important cues for recognition. Mathematically, it can be shown that Gabor transform does not sacrifice the spatial localization in spite of frequency resolution, which means that we can find accurate positions of objects and analyze regularities (textures) at the same time, performing both operations on Gabor features. Moreover, Gabor features are plausible models for some stages of the human brain processing (simple cells in primary visual cortex, [8]).

Nevertheless, Gabor pyramids are more difficult to manage than the previous features, as before they were simply low-dimensional vectors and now we have to deal with a whole pyramid structure whose amount of data is, in our case, larger than the original raw image. Furthermore, there are various methods for using Gabor features for pattern or face recognition, as can be seen in [9] and [10].

The approach followed here is explained in detail in [11]. Our basic idea is the following: as an incoming image is presented, one score is computed for each of the subjects in the training dataset (supposing that in the training dataset there is only one image per subject). This score can be computed for lowest, medium or highest resolution and, between all the subjects, the identity of the one giving maximum score is taken as the best match. A key point in this implementation is that every time one score has to be computed, one correspondence problem is solved. The correspondence problem is a difficult one in image processing -given two images containing the same object, it consists in finding in the images-coordinates frames points having the same origin in the 3-D real world. In the present approach, we look for candidates in the incoming image having one corresponding point in the training images. The training images are always taken as the reference. The grid of pixels of the incoming image at the present resolution can be distorted in order to find good correspondences. The final score is a weighted sum of the quality of the correspondences and the distortion we have been forced to apply for finding them. The correspondences are established in a hierarchical way (it means that for finding the correspondences in high levels of resolution, we need to have found them before in lower levels). Of course, all the computations (correspondences, scores, etc.) are carried over the pyramids, and the original faces are not needed any longer.

Figure 9 shows the process of establishing correspondences.

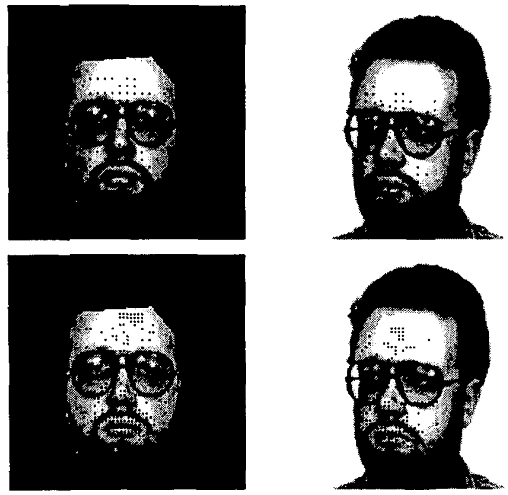

Figure 9.- Example of the correspondence problem at 3 given resolutions (16x16, 32x32 and 64x64, from coarse to fine resolutions), showing black points where correspondence between model (left) and test image (right) is found.

EXPERIMENTS AND RESULTS

We have designed and conducted some experiments in order to compare the two different approaches described before. The aim was to directly compare the performance of these approaches when mismatch occurs between training and testing conditions. Three important causes of variation are monitored during testing: illumination conditions, facial expression and non-frontal faces. As we have previously seen, the features in both systems were so different that they impose also differences in the recognition strategy. The analysis of the results cannot be explained only in terms of the feature extraction (as it would be desirable), but the complete recognition system has to be taken into account.

Datasets and preparation of the data. For the case of changes in expression and illumination we have employed de Yale database ([12]). The experiments have been carried over 15 of the subjects of this database. Regarding illumination and facial expression, the database is originally labeled. For each subject, one image is available with frontal illumination, one with left illumination and one with right illumination. One image is also available for each of the five following expression for all the subjects: happy, sad, sleepy, surprised and wink. All the image examples shown in the analysis multivariate section have been obtained from different facial expressions in this database. One 'normal' image has been employed for training and as a reference.

Regarding non-frontal views, we have employed the PICS database, acquired at the University of Stirling ([13]). We have worked with 2 images per subject (one frontal, and one where the face is rotated 45° between frontal and profile) and with 34 subjects.

The systems work with images in pgm format. We have converted all the images to this format and cropped and resized conveniently to have always the same size (128x128). For the Gabor system, we needed some previous segmentation of the training set, in order to establish correspondences only for points inside the face and not in the full image. A suitable rectangle was selected for each database, so that most of the points inside fell in the face for the training set.

Changes in facial expression. First of all, we have used for training only the 15 'normal' available images (one per subject) and we have tested over the 75 images (15 subjects x 5 expressions) showing facial expressions. For the multivariate analysis, the results are given in Table 1.

	NN	PCA-5	PCA-10	PCA-14
Euclid.	88%	88%	88%	88%
Mahalan	-	85%	83%	88%

Table 1.- Correct identification results considering changes in facial expression. PCA-n indicates n eigenfaces are considered (the maximum is 14). Euclidean and Mahalanobis metrics are employed.

LDA and probabilistic-PCA method could not be applied, as they require more than one training image per subject in the training set. With 15 training images, the mean and 14 eigenfaces expand the whole training set, so it does not make sense to consider more eigenfaces. Moreover, performance does not vary dramatically when reducing the number of eigenfaces.

The Gabor system reached a performance of 95% of correct identification when using scores at the higher resolution level. The same performance is attained at the medium resolution level and 92% (still better than using multivariate methods) at the lowest resolution.

The results are not surprising considering the previous description of the systems. The Gabor system performs well with facial expressions because the differences are mainly due to changes in the shape of few facial attributes, which can be compensated with the correspondences as explained before. Nevertheless, for the multivariate case the situation is different, as performance of the system is based on statistics, and statistics in the training set were here very poor (only one image per subject, without learning of facial expression). In order to verify this hypothesis, we modified the experiment randomly redistributing the 90 images used (15 for training + 75 for testing) in two new sets of 45 images (each of one with 3 images per subject). One was used now for training and the other for testing. Now the statistics of the training set are richer and contain the variability we have in our test set. Moreover, now we can employ all the multivariate methods. As it can be seen (Table 2), results improve significantly, outperforming in all cases Gabor approach:

NN	PCA-44	PCA-44 Mahala.	LDA-14	PPCA-30
98%	98%	96%	96%	100%

Table 2.- Correct identification rates for the different techniques used.

Changes in illumination conditions. The set of experiments carried out under these conditions follow the same scheme as the previously described, as 15 'normal' images are used as training set and the other 45 (with a labeling of their illumination conditions) are used for testing, in both Gabor and multivariate approaches. Figure 10 shows severe variations in the illumination conditions, even for the same subject in the database.

Figure 10.- Different illumination conditions considered in the experiments.

For the system based on multivariate analysis, the results are given separately (Table 3) for the ensemble of all testing images and, specifically, for those corresponding to central illumination, which represents a smaller variation if compared with the whole training set.

	NN	PCA-14	PCA-14 Mahala
All images	47%	47%	47%
Central il.	80%	80%	80%

Table 3.- Correct identification rates for the different techniques used, separating testing results for all illumination conditions and only for central illumination conditions.

As it can be derived from the table, in this case the system performs worse than for facial expression variations. These results coincide with our expectation, as these methods are 'global' in the sense that they consider the image as a whole (a high dimensional vector). Changes due to facial expression affect few components in this vector, as they are very localized in small regions of the image as the eyes or the mouth. Major illumination changes, however, affect many components of the vectors (gray values) showing a more sensible behavior to these changes. Nevertheless, Gabor approach shows its inherent robustness, as 93% of the images corresponding to the test set are correctly classified when using scores from the highest resolution analysis. It also shows the advantages of a multiresolution approach, taking into account that only 33% of the images are correctly classified when using the scores from the intermediate resolution analysis. Lower resolutions, being more 'global' in the sense previously defined, discriminate the boundaries between more and less illuminated zones, while higher resolutions are subtler.

As previously, we have increased the training set, reducing at the same time the test set (45 images in the training set, 15 in the test set, one of each subject) in order to appreciate the improvement with a better learning situation. The results are shown in Table 4.

NN	PCA-44	PCA-44 Maha.	PCA-(4-44)	LDA-14	PPCA-30
47%	47%	73%	67%	93%	93%

Table 4.- Correct identification rates. (4-44) means keeping eigenfaces between 4 and 44.

In this case, the top performance is not higher than that of the Gabor system. It is important to emphasize here the improvements achieved with the Mahalanobis distance and LDA, and with PCA without considering the 3 first principal components. Some possible explanations to this can be stated:

- Illumination variations are coded in a certain way in the lowest principal components. That could be the reason for the improvement when removing the three first principal components.
- Mahalanobis distance tends to equalize the effect of all principal components. As those having more relevance in coding the illumination lose relevance with this metric, the performance increases.
- Even if the theoretical development looks more consistent, LDA performs usually worse than PCA. One reason could be that this technique is perhaps more sensible to non-linear causes of variation than PCA. Nevertheless, it can be shown that under reasonable assumptions, illumination changes produce linear variations in the data.

The first principal eigenfeature is shown for PCA, PPCA and LDA. Illumination variations are strongly present in the first two, as can be seen in Figure 11.

Figure 11.- Some eigenfeatures showing the relevance of illumination conditions.

Rotating views. The last case that we have considered consist of training with frontal poses and testing with a rotating view of 45° (between frontal and profile) of the same persons (Figure 12). It is a harder problem as the 3-D perspective of the pattern (the face) cannot be easily derived. This additional difficulty explains why researchers usually design systems to cope only with a single view and use previous stages for aligning images with this view.

Figure 12.- Rotating views of a face

The performance in this case drops dramatically. As we used only one image per subject for training, for the multivariate system we have tested only with NN and PCA with 33 and 15 eigenfaces (euclidean and Mahalanobis metrics). The highest correct identification rate was only 18%. The Gabor system reached (with the scores for the highest resolution) a correct identification rate of 38%. This improvement must be due to the establishment of correspondences, as these perform better with the local deformations present in this case. However, 45° seems to be too much for establishing good correspondences. Previous similar experiments with the Gabor system on the Bochum database with 15° of rotation gave rates over 90% of correct identification.

CONCLUSIONS AND FUTURE LINES

We have described two different approaches for face recognition. Each of them is based on a different technique for feature extraction: multivariate analysis and Gabor analysis. In the experiments we have conducted, we have shown that none of them is better suited than the other, as their relative quality depends on specific conditions. If training material contains all the variations that can be found in real applications, the first methods (based on statistics) can perform well. Gabor analysis, not being a statistical approach, is not so constrained as multivariate analysis, and still reaches good performances. However, statistics is an important source of information and it seems to be reasonable to use it for recognition.

We conclude that both kind of features are somehow complementary and a system using wisely both (as perhaps the human brain does!) should outperform actual systems based in a single technique. Our future work should be oriented for joining in one single system the advantages of both feature extraction methods; for instance, Gabor representations can be used, instead of the original images, as an input to multivariate analysis systems.

Other methods, like non-linear statistical multivariate methods (as Independent Component Analysis or Kernel Principal Component Analysis) have also to be explored in the future.

ACKNOWLEDGMENTS

This work could not have being carried out without the invaluable collaboration of Dr. Rolf P. Würtz, Ramón López-Viana and Isabel López-Aranda. The Gabor system was developed by first author at Institut für Neuroinformatik (Ruhr-Universität Bochum, Germany), which also provided the mentioned Bochum database. Some tests have also been carried by first author at Institut Dalle Molle d'Intelligence Artificielle Perceptive (IDIAP, Martigny, Switzerland). We would like also to thank the institutions that have made publicly available their databases (Yale and PICS) in the Web.

REFERENCES

[1] R.A. Johnson, D.W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice Hall International, London, 1992.

[2] M. Kirby, L. Sirovich, "Application of the Karhunen-Loeve Procedure for de Characterization of Human Faces", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 12, No. 1, pp. 103-108, 1990.

[3] M. Turk, A. Pentland, "Eigenfaces for recognition", Journal of Cognitive Neuro Science, Vol. 3, pp. 71-86, 1991.

[4] P. N. Belhumeur, J.P. Hespanha, D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Using Class Specific Linear Projection", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, July 1997.

[5] D. L. Swets, J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 18, No. 8, 1996.

[6] B. Moghadam, A. Pentland, "Probabilistic Visual Learning for Object Representation", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 696-710, July 1997.

[7] B. Moghadam, C. Nastar, A. Pentland "Beyond Eigenfaces: Probabilistic Matching for Face Recognition", MIT Media Laboratory, Perceptual Computing Section, Technical Report No. 443, 1997.

[8] J. G. Daugman, "Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression", IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. 38, No. 7, pp. 1169-1179, 1998.

[9] M. Lades, J.C. Vorbrüggen, J. Buhman, J. Lange, C.v.d. Malsburg, R.P. Würtz, W. Konen, "Distortion invariant object recognition in the dynamic link architecture", IEEE Trans. on Computers, Vol. 42, No. 3, pp. 300-311, 1993.

[10] L. Wiskott, J.M. Fellous, N. Krüger, C.v..d Marlburg, "Face Recognition by Elastic Bunch Graph Matching", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol 19, No. 7, pp. 775-779, 1997.

[11] Rolf P. Würtz, "Object Recognition Robust Under Translations, Deformations, and Changes in Background" IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 769-775, 1997.

[12] http://cvc.yale.edu/projects/yalefaces/yalefaces.html

[13] http://pics.psych.stir.ac.uk/