

Automatic Face Representation and Classification

Daniel B Graham and Nigel M Allinson
Image Engineering and Neural Computing
Department of Electrical Engineering and Electronics
UMIST, Manchester, M60 1QD, United Kingdom.
danny@sound.ee.umist.ac.uk|allinson@umist.ac.uk

Abstract

A working face recognition system requires the ability to represent facial images in such a way that permits efficient and accurate processing. The human visual system effectively stores, recognises and classifies familiar facial images under a wide variety of viewing conditions, albeit with various degrees of accuracy. We describe a system which automatically determines a representation for pose-varying facial images - a representation with inherent classification properties, an ability to generalise from one viewing condition to another, and which uses fast computational procedures.

1 Introduction

Representing a human face under a large range of viewing conditions and being able to recognise images of familiar people in these conditions is one of the more remarkable abilities of the human cognitive system. Whilst the relatively simple procedure of matching frontal face images under *consistent* viewing conditions has been implemented and commercially exploited [17, 23], we have yet to see a system capable of recognising faces over significant variation¹ in viewing conditions and generalising to novel conditions. The technique described in this paper demonstrates a system which attempts both of these difficult tasks for an arbitrary set of viewing conditions. We provide an automatic method for constructing efficient representations of faces - for example over 90° head rotation is described by a one hundred-dimensional vector. Within this representation we utilise our prior experience of known behaviours to facilitate the recognition of faces in novel viewing conditions. Our system demonstrates the use of this approach for pose-varying faces and we show that, having previously only seen one image of an individual, we can then generalise to poses as far as 90° away from the training pose with 50% accuracy. The general purpose method described here is capable of learning representations of any object set in any previously experienced viewing conditions and of relating object appearance in one condition to another.

Machine vision techniques are often inspired by human visual processes and one of the key issues for recognition algorithms is that of *representation* - how do we represent objects (and faces) in order to be able to recognise them under a large variety of viewing

¹For instance [9] in which the pose range is limited to $\pm 20^\circ$ causing no occlusion of features.

conditions? Two competing theories concerning the nature of object and face representation in the brain can be simply described as the *three dimensional (3D) approach* and the *two dimensional (2D) approach*. Essentially, does the human visual system (HVS) attempt to recover the 3D information from a single (or multiple) image (e.g. [22]) and use this information to generalise to novel viewing conditions or does the HVS (in some manner) store a number of 2D images (for instance [20]), under various conditions, and interpolate between and extrapolate beyond these conditions in order to generalise to novel views? Recently there have been several experiments [14, 8, 1, 18] which have supported the (class-based) 2D approach rather than the more traditional 3D approach. This paper provides an automatic, view based, 2D approach for face representation and uses the representation itself to determine characteristic appearance changes for new individuals.

Frontal face recognition systems based upon the principal component analysis techniques [19, 12] or the Gabor wavelet approach (dynamic link architecture) [7, 23] have been shown to exhibit good performance characteristics when the viewing conditions between the training and testing phases remain relatively constant. However their performance is usually severely degraded when illumination or pose changes become significant. We are primarily concerned here with learning an efficient, yet characteristic, representation of human faces under different viewing conditions. Using this representation we can then attempt recognition and classification. We are also interested in the condition-dependent task performance within the representation. Furthermore, what we consider to be recognition is *always* in a viewing condition considerably different from the training view (by as much as a 90° pose change in the experiments reported here) and our representation facilitates this generalisation ability.

The remainder of this paper is organised as follows: section 2 describes our method of generalisation and recognition within an appropriate subspace and section 3 describes the manner in which we automate the construction of view-based representations. Our experimental results are described in section 4 where we show that this automatically constructed representation performs equivalently to a (painstakingly formed) manual one. Finally, in section 5 we describe how such representations may be used for classifying faces through various attributes.

2 Subspace Generalisation

In our experiments, we employ the pose varying eigenspace representation used by Murase and Nayar [15] for object recognition and McKenna et al [11] for face recognition. We do not attempt to construct an optimal eigenspace for recognition, rather we use a small but representative sample of our facial image set to form an eigenspace. Typically we sample less than one hundred images from an image set of over 500. The eigenspace is constructed, as usual, from the covariance matrix of the sampled images - resulting in a number of image-sized eigenvectors. We eliminate all but the largest ten eigenvectors which are used for the subsequent analysis. We have shown previously that the performance of systems based on pose-varying eigenspaces with dimensionality greater than ten is effectively constant [3]. After the image processing stage, our database of faces consists of a number of individuals, in a number of poses, each image being represented as a single point in a 10D subspace.

The generalisation from single views (and hence single points in the subspace) to

novel views requires the ability to capture the manner in which a subject moves through the eigenspace as they undergo the viewing condition change. Essentially we require some prior experience of similar transformations and our ability to generalise to novel individuals is dependent upon the extent of that experience. For this we utilise a Radial Basis Function (RBF) Network [13] - a type of neural network containing one hidden layer of Gaussian activation nodes fully connected by a set of linear weights to the output layer. Radial Basis functions are trained in three stages; first we select the coordinates of the hidden unit centers, next we select the width of the Gaussian activation functions - centered at the positions chosen in stage one, finally we calculate the weights connecting the hidden and output nodes - as this is a linear summation it can be performed by a matrix inversion.

For our face database we train a RBF network on a set of single images - represented by their 10D vectors in the eigenspace - to predict the remaining N poses of the same individuals in this space. Thus a 10D input vector produces $N \times 10D$ output vectors corresponding to that individuals' representation at all N poses in the eigenspace - referred to as a *virtual eigensignature* [4].

Formally, the recognition of faces from previously unseen views requires a function Γ which maps a real point p to a virtual eigensignature Υ , i.e $\Gamma(p) = \Upsilon$, where Υ consist of N points $\{q_1, \dots, q_N\}$ where $\{p, q_{1 \dots N}\} \in \mathbb{R}^{10}$ - points in the 10D eigenspace. Γ here is the RBF network described earlier. Recognition of a novel view consists of calculating the eigenvalues of a test image to give a point p_t in the eigenspace. This is then used to find a minimum distance between p_t and q_n^i - where q_n^i is the n -th member of eigensignature i (Υ_i). Several distance functions are possible, we have investigated the Euclidean distance, the Mahalanobis distance [2] (where each dimension in the eigenspace is weighted according to the magnitude of the corresponding eigenvector) and more sophisticated measures. All have been found to perform similarly and in this paper we report only on the use of the Euclidean distance.

For a large database we have many Υ_i and each $\Upsilon_i \in \mathbb{R}^{10N}$ such that an exhaustive search in $10N$ dimensional space can require extensive computation. Thus we can restrict the search space of the recognition by attempting to classify according to which pose we think the target belongs to (q_n) or by using a data structuring technique such as that of Nene and Nayar [16].

3 Representation Building

As previously mentioned our chosen representation for faces consists of a series of ten dimensional points in an eigenspace $q_{1, \dots, N}$. These points cover the range of viewing conditions from which we wish to attempt recognition. Given a small number of people and conditions it is possible, but time-consuming, to manually select the training data (q_n) for each individual in the training set. This approach restricts the applicability of the method. A more useful and insightful approach would be to automatically construct a condition set, given our experience of facial images and then to re-sample the database according to this condition set.

Our method of automatic representation building employs the K-means clustering algorithm of MacQueen [10]. Here we use our experience of facial images to determine local clusters of data in the eigenspace. These clusters then form the basis centers from

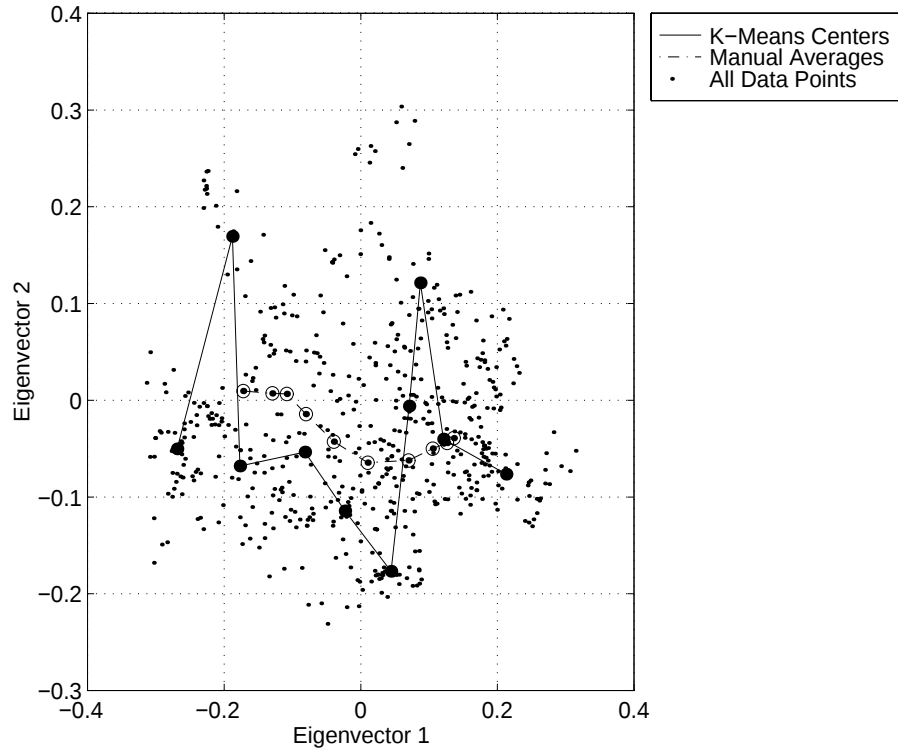


Figure 1: K-means center locations (with manual averages)

which we choose q_n . We take a sequence of images for each individual and only retain those images which are nearest to each basis center. Thus we form a real eigensignature for each individual in the training set.

The K-means clustering was applied algorithm on our database of 564 images of 20 individuals. In order to compare the performance of this automatic method with our previous manually-formed one [4] ten basis centers were chosen. As this is a partially stochastic approach we applied the K-means algorithm several times in order to form a more accurate picture of the system performance. Effectively we are forming a different representation for each run of the K-means algorithm and so establishing the recognition characteristics of the technique for a large number of cluster locations. Figure 1 illustrates the calculated K-means centers for one run. For illustration purposes we show only the first two eigenvalues of each image and K-means center. For comparison the average points of each person, at each pose, in the manually-formed representation used in [4] is shown. It can be seen clearly that the clusters represent more of the data in the sample than the manual averages - but that the characteristic nature of the K-means centers may be more difficult to obtain with the RBFN due to the increased distance between adjacent centers.

It has often proved useful to visually examine the eigenvectors in such experiments (e.g. [21]) indeed the *face-like* nature of the eigenvectors has led to them being called *eigen-faces* [19]. Figure 2 shows the image reconstruction of these points in the first ten

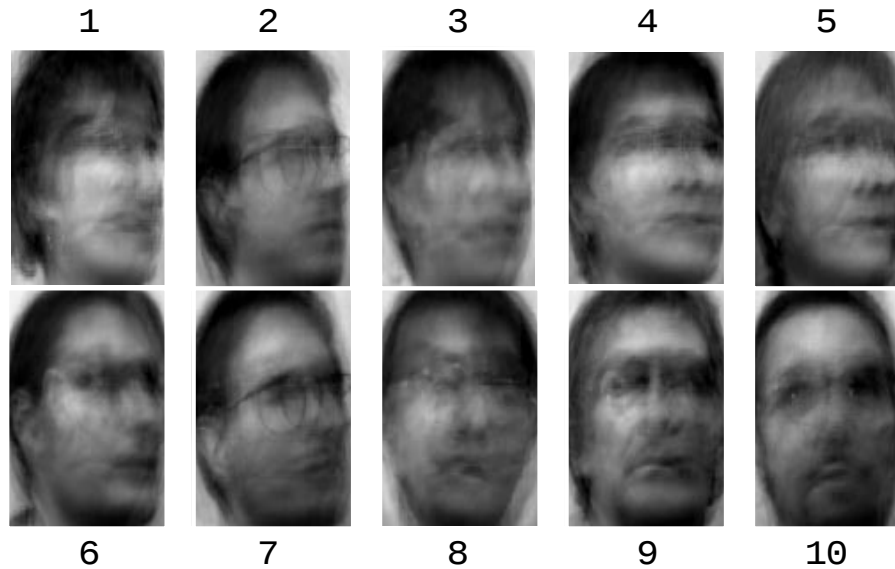


Figure 2: K-Means reconstructed centers

eigenvectors. Clearly these images cover the range of poses represented in our database. Interestingly, there are apparent features such as the presence of glasses (eigenvectors 2 and 7) and the presence of a beard (eigenvector 10). It is these apparent features that have led us to use the automatically formed representation for classification of faces by features. These experiments are described in section 5.

4 Experimental Observations

Our previous face recognition experiments [4], which employed the manually formed representation described, demonstrated the use of the *virtual eigensignatures* formed by the RBF-based framework described in section 2 for unfamiliar view face recognition. In [4], we examined both the training and testing pose dependent nature of the system. Views intermediate to the frontal and profile views were superior views for both training and testing. In [5], we proposed that this property is a direct consequence of the distance between individuals at these intermediate poses and the relatively smaller distance between them at profile and frontal views. Section 4.1 will establish the performance characteristics of the K-means algorithm for determining representations whilst section 4.2 will analyse the distance relationships of such representations to see if they support the proposed relative distance theory.

4.1 Recognition

In [4] we determined the performance characteristics for unfamiliar view face recognition of the RBF-based approach. In summary we found that our manually formed representation could generalise to novel views from a single training image with a mean correct

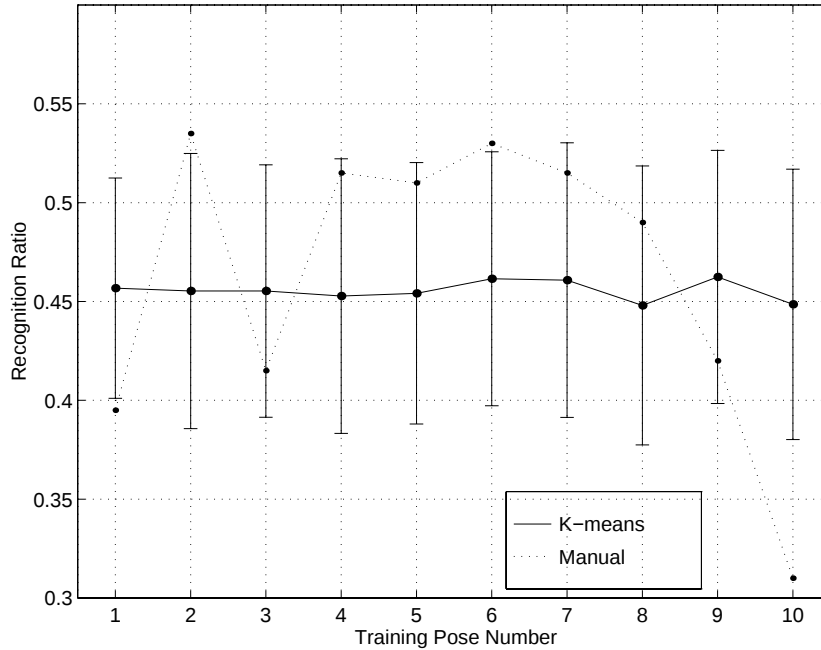


Figure 3: K-Means vs Manual recognition ratios

recognition ratio of 0.4635 ($\sigma = 0.0747$). This result is surprising when one considers the actual task and how we achieve these results: namely that all we are using is a RBF network trained to capture the characteristic behaviour of a set of individuals as they rotate through an eigenspace - and use this network to predict how a previously unseen person would undergo the same change. We are capturing some of the information available in such a characteristic process - how much more is extractable within this and similar frameworks is the subject of ongoing research.

According to the scheme described in section 3 the K-means algorithm was applied 100 times on our database of 564 images this produces 100 sets of cluster locations in the eigenspace. Each set was sorted according to its first eigenvalue and these locations used to determine the poses to store from each sequence of the 20 individuals in the database. The RBF network was trained on every combination of 19 from 20 of the individuals to achieve the characteristic learning and the network was then used to characterise the remaining individual from a single pose. The procedure was repeated for each pose. Figure 3 displays the mean recognition ratios and standard error for each of the training poses. For comparison we show the performance characteristic of the manual representation for each training pose. The K-means algorithm achieves a mean correct recognition rate of 0.4556 ($\sigma = 0.0050$) which is equivalent to the manual representation. As the K-means algorithm is designed to cover the data evenly, we see little pose dependency in the K-means representations (0.5%) compared with the pose dependent deviation of the manual one (7.5%). The *inter-set* standard deviation (i.e. the mean pose-varying standard deviation for each set of cluster locations) was 6.6% - which is similar to the manual

Measure	Correlation
Neighbour distance	-0.11 (± 0.31)
Mean cluster separation	0.02 (± 0.36)
Cluster width	-0.16 (± 0.34)
Cluster membership	-0.21 (± 0.32)

Table 1: Cluster property correlations

representation. From this we can conclude the each individual choice of representations may exhibit pose-dependent behaviour but, on average, there is no intrinsic dependency in such representations.

4.2 Distance Relationships

As discussed earlier, we believe that recognition ratios in the eigenspace are dependent upon the distance between faces at the test point. The K-means algorithm returns a statistical description of the images in the database in the form of the locations and widths of clusters (in the eigenspace) of data within the set. Using such measures we can examine the relationship between the cluster properties and the performance of the system when trained at that cluster. As before, the K-means algorithm was ran 100 times and, for each run, we determined four measures for each cluster:

- The mean distance to each neighbouring cluster,
- The mean distance to all other clusters,
- The width of each cluster,
- The number of training points which are closest to each cluster.

Table 4.2 shows the mean correlation of the 100 iterations for these four measures against the recognition ratios. The figures in brackets are the standard deviations for each measure. The lack of a strong correlation between any of these factors leads us to conclude that there is no support for the theory that the generalisation ability of the RBF networks is dependent upon the distance between each of the nodes, nor upon the relative importance of each cluster in the representation. The large standard deviations indicate that there are occasionally spurious large correlations for each of the measures.

5 Classification Experiments

As noted in section 3 there are apparent features associated with the eigenvectors of each cluster location determined with the K-means algorithm. In order to determine the relationship between the features of the faces and the locations of the cluster centers we classified all of the faces in our database manually by five attributes: *Hair (dark/light)*, *Glasses*

(*present/absent*), *Beard (present/absent)*, *Skin (Dark/Light)* and *Sex (Male/Female)*. Each of the former cases were assigned a value of $+1$ and the latter -1 .²

We again trained a RBF network on every combination of 19 individuals from 20 to predict the classification of each of the 5 attributes. For the input vector we used the vector between the pose for the individual and the cluster location. We then tested the network on the remaining individual. This was performed for each pose in the eigensignature. Figure 4 shows the mean performance per pose number of ten applications of the K-means algorithm. Note that there is effectively no pose dependency. Overall we see that that our correct classification rate is 0.6950 indicating that there is classification information available in such representations as there is for identity.

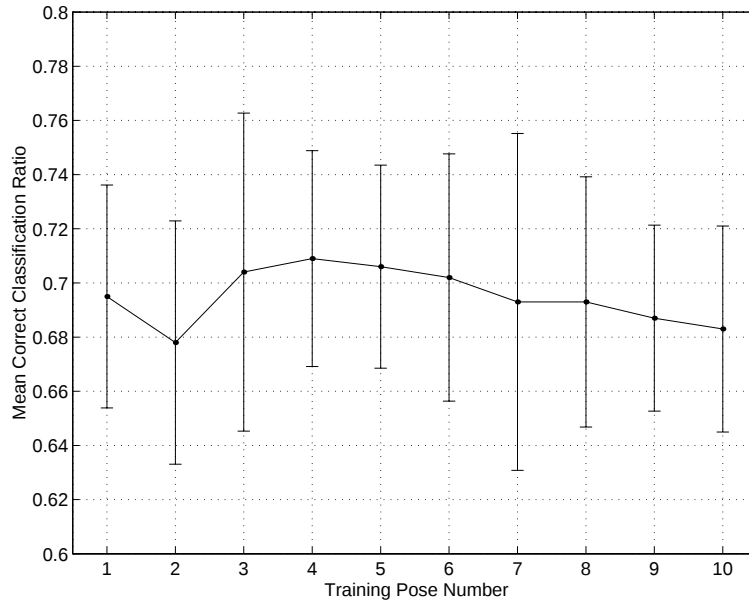


Figure 4: Classification Performance vs Training Pose

6 Discussion and Conclusion

We have introduced a means of automatically determining a view-based 2D representation of 3D objects and illustrated its use for pose-varying face recognition. We have shown that it is possible to extract identity-specific information from the characteristic nature of the representation and to extract classifiable information from the properties of the representation. The use of the K-means algorithm is perhaps a poor choice due to the inconsistent nature of its results. Future work will investigate more sophisticated approaches such as self organising maps [6]. The ability to generalise to novel views within these representations is also dependent upon the characteristics of such representations - the chosen discrete point model possibly represents the simplest approach and the

²There were 17 with dark hair, 8 wearing glasses, 4 with beards, 1 with dark skin, and 16 males.

recognition performance will be improved by a more comprehensive model.

References

- [1] S. Duvdevani-Bar and S. Edelman. Similarity-based method for the generalization of face recognition over pose and expression. *Proc. of the 3rd International Conference on Automatic Face and Gesture Recognition*, pp 118-123, NARA, Japan, April 1998.
- [2] N.P. Costen, I. Craw, G. Robertson, and S. Akamatsu. Automatic face recognition: What representation? *Computer Vision*, Bernard Buxton and Roberto Cipolla (eds), ECCV'96, vol 1064 Lecture Notes in Computing Science, pages 504-513. Springer-Verlag, 1996.
- [3] D.B. Graham and N.M. Allinson. Face Recognition using Virtual Parametric Eigenspace Signatures. *IEEE Image Processing and its Applications '97*, pp 106-111, Dublin, Ireland, 1997.
- [4] D.B. Graham and N.M. Allinson. Face Recognition from Unfamiliar Views: Subspace Methods and Pose Dependency. *3rd International Conference on Automatic Face and Gesture Recognition*, pp 348-353, NARA, Japan, April 1998.
- [5] D.B. Graham and N.M. Allinson. Characterising Virtual Eigensignatures for Face Recognition. in *Face Recognition: From Theory to Applications*, Wechsler, H. and Phillips, P.J. and Bruce, V. and Fogelman-Soulie, F. and Huang, T. (eds.) NATO ASI Series F, Springer-Verlag, 1998.
- [6] T. Kohonen. *Self-Organisation and Associative Memory*. Springer-Verlag, Heigelberg, 1984.
- [7] M. Lades, J. Vorbuggen, J. Buhmann, J. Lange, C. von der Marlsburg, R. Wurtz, W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, vol 42(3), pp 300-311, 1993.
- [8] M. Lando and S. Edleman Generalization from a single view in face recognition. *Network*, vol. 6, pp 551-576, 1995.
- [9] A. Lanitis, C. Taylor and T. Cootes. A Unified Approach to Coding and Interpreting Face Images. *Proc. 5th International Conference on Computer Vision* pp 368-373, Cambridge, USA, 1995.
- [10] J. MacQueen. Some methods for classification and analysis of multivariate observations. *Proc. 5th Berkley Symposium*, 1, pp 281-297, 1967.
- [11] S. McKenna, S. Gong and J.J. Collins. Face Tracking and Pose Representation. *British Machine Vision Conference*, Edinburgh, Scotland 1996.
- [12] B. Moghaddam and A. Pentland. Face Recognition using view-based and modular eigenspaces. *SPIE* vol 2277, pp 12-21, 1994.
- [13] J. Moody and C.J. Darken. Fast Learning in Networks of Locally-Tuned Processing Units. *Neural Computation*, vol 1, pp 281-294, 1989.
- [14] Y. Moses, S. Ullman and S. Edelman. Generalization to Novel Images in Upright and Inverted Faces. *Perception*, vol. 25, No. 4, pp 443-461, 1996.
- [15] H. Murase and S.K. Nayar. Learning Object Models from Appearance. *Proc. of the AAAI*, pp 836-843, Washinton DC, July 1993.
- [16] S.A. Nene, S.K. Nayar. Closest Point Search in High Dimensions. *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, (CVPR 96), San Francisco, June 1996.
- [17] P.S. Penev and J.J. Atick. Local Feature Analysis: A general statistical theory for object representation. *Network: Computation in Neural Systems*, vol.7, No.3, pp.477-500, May 1996.

- [18] P. Sinha and T. Poggio Role of learning in three-dimensional form perception. *Nature*, vol 384, no. 6608, pp 460-463, 1996.
- [19] M. Turk and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, vol 3, No. 1, pp 71-86, 1991.
- [20] S. Ullman and R. Basri. Recognition by Linear Combination of Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 13, No. 10, October 1991.
- [21] D. Valentin and H. Abdi. Can a Linear Autoassociator Recognize Faces From New Orientations. *Journal of the Optical Society of America A - Optics, Image Science and Vision*, vol 13, No. 4, pp 717-724, 1996.
- [22] T. Vetter Learning novel views to a single face image. *Proc. 2nd International Conference on Automatic Face and Gesture Recognition*, pp 22-27, Killington, Vermont, 1996.
- [23] L. Wiskott, J. Fellous, N. Kruger and C. von der Malsburg. Face Recognition by Elastic Bunch Graph Matching. *Internal Report 96-08*, Institut für Neuroinformatik, Ruhr-Universität Bochum, April 1996.