Graph-based Matching of Occluded Hand Gestures

Atid Shamaie Alistair Sutherland

School of Computer Applications, Dublin City University, Dublin 9, Ireland
{ashamaie, alistair}@compapp.dcu.ie

Abstract
Occlusion is an unavoidable subject in most machine vision areas. Recognition of partially-occluded hand gestures is an important problem. In this paper a new algorithm is proposed for the recognition of occluded and non-occluded hand gestures based on matching the graphs of gestures in an eigenspace.

1. Introduction
The use of hand gesture recognition in Human Computer Interaction has been addressed in the literature. [1] It is a more natural way compared with the use of keyboards, mice, etc.
Current approaches to the recognition of hand gesture are often based on statistical methods to extract the features of different shapes of hand in a sequence of images. However, many other different approaches have been introduced as well. Spatio-temporal hand gesture recognition using neural networks [2][3], spatial modelling of gestures [4][5], recognition of gestures using Hidden Markov Models [4][6][7], Principal Component Analysis [8][9][10][11], position-based gesture recognition [12] and many other techniques have been used to improve the process of gesture recognition.
Although there is a lot of work on graphs [13] and graph matching [14][15], no very efficient technique has been reported in the literature about the application of graph matching techniques in hand gesture recognition [16].
Also recognition of hand gestures in the case of occlusion is an open area. In this paper we will discuss the problem of gesture recognition of the human hand and approaches based on Principal Component Analysis (PCA). A new algorithm for the recognition of un-occluded and partially occluded gestures based on PCA and graph matching is presented.
In the next section the problem of recognition of gesture is addressed. In Section 3 we discuss PCA, its application to gesture recognition and we give an overview of some other work in this area. In Section 4 a very general overview of the problem of graph matching is discussed. Section 5 is dedicated to our proposed new algorithm based on graph matching and its power in recognising partially occluded hand gestures. In Section 6 we discuss the parallel nature of the proposed algorithm. In Section 7 we give the results of some experiments. And some discussion and conclusions are stated at the end of the paper.

2. Problem statement
An approach to hand gesture recognition uses Principal Component Analysis (PCA) and projects the sequence of hand shapes into an eigenspace. From the mathematical point of view PCA is "reducing the dimensionality of a data set in which there is a large number of interrelated variables, while retaining as much as possible of the variation present in the data set." [17]
Many different applications of PCA have been addressed in the literature [17][18] in machine vision and pattern recognition as well as chemistry and other areas.
Herein, it is assumed that in the sequence of input images the hand has been segmented from the background. This leads to the following definition of the problem:
"Given a sequence of images containing a hand gesture, find the best fit of the sequence to a known gesture, the direction of the gesture, the start and end point of the gesture in the case of complete and partially occluded gestures."

3. Projection of gesture into eigenspace
By projecting the sequence of images into the eigenspace, each frame of the input sequence maps to a point in the eigenspace.
In this research we kept the variation of illumination as small as possible and in experiments we tried to use the same illumination for different gestures. Based on this hypothesis we collect many examples of each gesture, calculate the covariance matrix and form a subspace. During the training phase each gesture generates its own subspace. The projection of the input sequence of images into its own subspace forms the "main manifold" of that subspace. Figure 1 shows the main manifold of a gesture used in our experiments.
This manifold belongs to the movement of the V sign from the upright position to the horizontal position (Figure 2).

0-7695-1245-3/01 $17.00 © 2001 IEEE

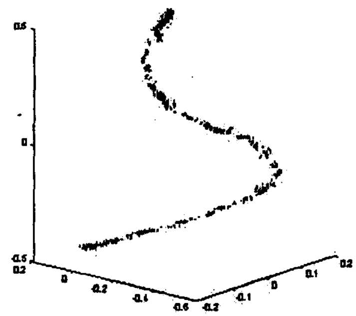

Figure 1. The main manifold of a subspace

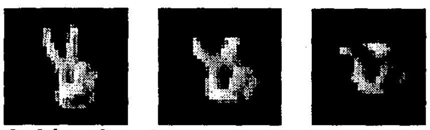

Figure 2. A hand gesture.

The resolution of the images is 32x32. Therefore the resulting covariance matrix is 1024x1024 and each eigenvector has 1024 elements.

4. Bipartite graph matching

The bipartite graph matching problem is finding a set of pairwise disjoint edges of a bipartite graph based on a special characteristic of edges.

This problem has been widely used in various subfields of science and technology [16][14].

Finding the optimal solution of the graph matching problem has been addressed widely in the literature [13][14][16]. Due to the NP-Completeness of the problem finding the optimal solution requires exponential time. However, a suboptimal or approximative solution can be satisfactory in some cases with polynomial processing time [16].

The graph-matching algorithm we introduce here finds a suboptimal match of two graphs. It is based on finding the shortest edges of a complete bipartite graph.

Each sequence of input frames generates a manifold in its own eigenspace. Other gestures generate different manifolds in the subspace as shown in Figure 3.

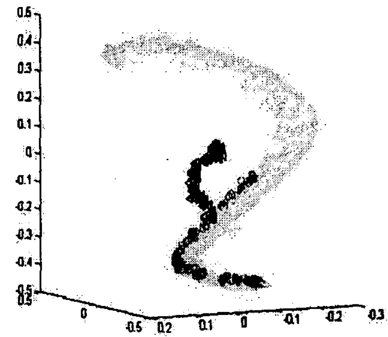

Figure 3. Projection of two different gestures into a subspace. The big manifold is the main manifold of the subspace.

Finding the best (suboptimal) match between the main manifold of a particular gesture and the manifold of an unknown gesture is the main problem in this paper.

5. New algorithm for gesture classification

5.1. Constructing subspaces

Like other recognition systems we have to have a training phase as well as recognition phase.

In the training phase we collect a certain number of known gestures and we construct the eigenspace of each individual gesture. By projecting every input sequence into its own subspace we get the main manifold of that subspace. In the recognition phase we find the best match of a given unknown gesture to the already known gestures.

By projecting the input sequence of unknown gestures into all the n subspaces we get n manifolds of the unknown gesture, Figure 4.

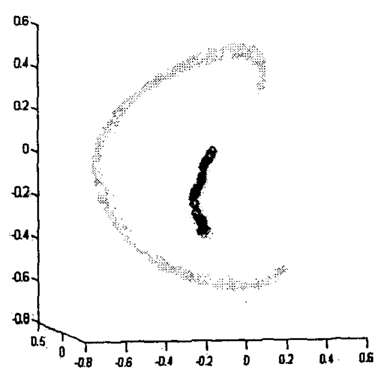

Figure 4. Projection of an unknown gesture into $n=3$ different subspaces.

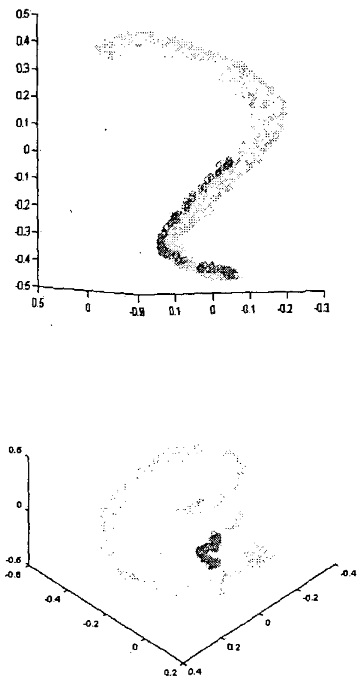

Figure 4 (continued). Projection of an unknown gesture into *n=3* different subspaces.

5.2. Constructing graphs

Each manifold can be represented as a directed graph whose vertices (nodes) are the means of groups of points in the manifold. So we can divide the manifold into v vertices connected by $e = v - 1$ edges. Labelling the vertices with numbers from 1 to v, Figure 5, identifies the direction of the graph.

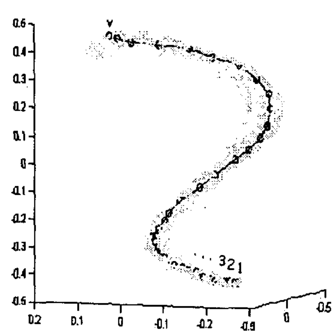

Figure 5. The graph of the main manifold and the numbered vertices.

Now let the graph G = (V, E) of a manifold where $v = |G|$ is order of graph and $e = \|G\|$ is the number of edges.

5.3. Matching graphs

Now we start to find the best (suboptimal) match of every pair of directed graphs $M_i = Match(G_i, G_i')$, i=1,...,*n*. The graph-matching algorithm is as follows:
Given two graphs we construct a complete bipartite graph of their sets of vertices as Figure 6.

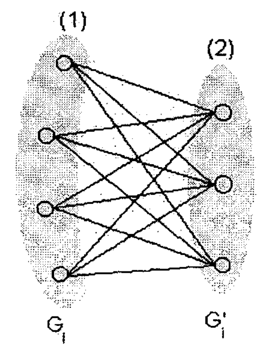

Figure 6. A complete bipartite graph.

We label each edge of the bipartite graph with its length. The length of the edge is the Euclidean distance of adjacent vertices in the eigenspace. So with $v \, x \, v'$ edges we have:

$$l_{ii'} = \text{Euclidean Distance}(\gamma, \gamma')$$

Where:
$$i = 1, \ldots, v$$
$$i' = 1, \ldots, v'$$
$$\gamma \in V \quad , \quad \gamma' \in V'$$

For the first set of vertices (say G_i) in the bipartite graph, for each node we find the shortest edge connected to that node and eliminate other edges. At the end of this stage we have two sets of vertices V and Γ' where $\Gamma' \subseteq V'$.

Γ' is a subset of V' because the number of vertices of second set (G_i') at the end of this stage must be equal or smaller than this number in the graph G_i'. This happens because at the end of this stage we eliminate the nodes in the second set with no edge connected to them, Figure 7. As can be seen from Figure 7 a vertex from G_i' can remain with more than one edge because a vertex of G_i'

can be the closest vertex to more than one vertex of G_i. The second stage of the matching process finds a graph with vertices that each has only one edge connected to it.

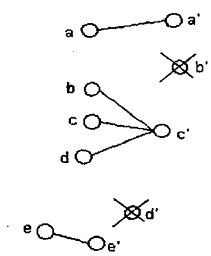

Figure 7. In this graph some of the vertices remain because of their shortest edge to the first set.

For G_i' we find the shortest edge connected to each of its vertices and eliminate other edges. At the end of this stage we remove all the vertices of G_i without any edge connected to them. Therefore we get a set of nodes Γ which is a subset of first set ($\Gamma \subseteq V$). Now we have two sets of vertices Γ, Γ' with the same number of nodes and one-to-one edge connected between two sets.

5.4. Classification and decision-making

As we have seen the graph-matching algorithm reduces the number of nodes of two sets of vertices while it's finding the shortest edges of the bipartite graph. So one can conclude that if the vertices of two graphs are distributed along the same trajectory in the eigenspace the number of vertices after the matching process will be greater than in the case of two graphs with different trajectories, Figure 8.

For example in the second case in Figure 8 some of the vertices of say first set are closest to many vertices of second point and vice versa.

We could say the unknown gesture belongs to the group of known gestures with maximum number of matched vertices. Of course for almost the same number of vertices we consider the mean value of the distances of matched vertices.

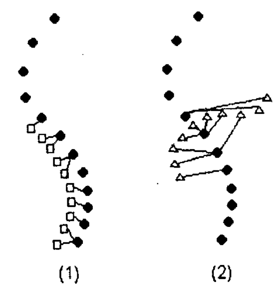

Figure 8. Different distributions of graph nodes with respect to the main manifold of the subspace.

5.5. Occlusion recognition

Figure 9 shows an occluded gesture with respect to the main manifold. The points of the partially occluded gesture are distributed along some parts of the main manifold.

After the graph matching process the remaining vertices of the graph of the main manifold show the start and end point of the partial gesture with respect to the main manifold.

It's necessary to say that we only deal with the input images that have the whole hand not the images in which a part of hand appeared. That is another problem that we have to work on in the future.

5.6. Sense detection

The "sense" of a gesture means doing gesture in a direction or opposite one. In other words, for example, moving a special shape of the hand from left to right or from right to left, Figure 10.

By looking at the sequence of labels of vertices of two sets of matched graphs we could recognize the direction of unknown gesture with respect to the direction of the main manifold (gesture).

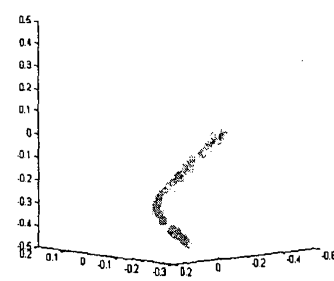

Figure 9. Partially occluded gesture's manifold and graph.

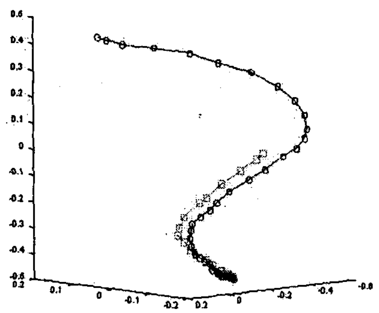

Figure 9 (continued). Partially occluded gesture's manifold and graph.

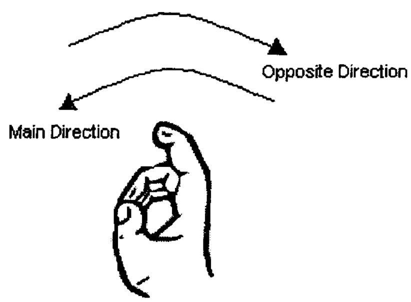

Figure 10. Different directions of a gesture.

5.7. The algorithm

In the following algorithm we refer to $v_{(i)}$ and $e_{(i)}$ as a vertex and edge of a graph respectively. The algorithm is summarized as follow:

1. ***Training Phase***

   1.a. Capture several instances of a gesture:

   $$a_{(i)} = ConcatenateFrames(a_{(i)}, Capture\ gesture(i))$$

   Where:

   $a_{(i)}$ is a sequence of images of gesture *i*

   1.b. Repeat the step 1.a for all classes of the gestures.

   1.c. By using PCA make subspace of each class of gestures and project each gesture into its own subspace to form the main manifold of the subspace:

   $$z_{(i)} = A_i' a_{(i)}$$

   Where:

   A_i' is the orthogonal matrix whose kth column is the kth eigenvector of covariance matrix of a sequence of images represented as *a(i)*, and *z(i)* is the projection of image sequence *a(i)* into its subspace.

   1.d. Obtain the main graph of each subspace:

   $$G_{(i)} = (V_{(i)}, E_{(i)})$$

   Where:

   $$V_{(i)} = \{v_{i1}, v_{i2}, \ldots, v_{im}\}$$

   v_{ik} = MeanPoint (kthslice of the Principal Components)

   $$E_{(i)} = \{e_{i1}, e_{i2}, \ldots, e_{im-1}\},\ e_{ik} = v_{ik}v_{ik+1}$$

2. ***Recognition Phase***

   2.a. Capture a gesture *b*

   2.b. Project the gesture *b* into all the subspaces:

   $$t_{(i)} = A_i' b$$

   Where:

   $t_{(i)}$ is the projection of the image sequence b into the ith subspace

   2.c. Construct the graphs of the gesture *b* in every subspace:

   $$G'_{(i)} = (V'_{(i)}, E'_{(i)})$$

   2.d. Construct the complete bipartite graph with the two sets of vertices:

   $$K_{(i)}(m_1, m_2) = (V''_{(i)}, E''_{(i)})$$

   Where:

   $$V''_{(i)} = V_{(i)} \cup V'_{(i)}$$

   and every edge in $E''_{(i)}$ labled with the Euclidean Distance of its two vertices

   2.e. Find the match subgraph of the complete bipartite graph:

   2.e.1. Start from a set of vertices $V'_{(i)}$ and find the edge with the smallest label incident with every vertex of the set $V'_{(i)}$

   2.e.2. Keep the edge with the smallest label and remove the other incident edges for each vertex.

   2.e.3. Repeat Step 2.e.2 for all the vertices in the set $V'_{(i)}$

   2.e.4. Remove the vertices of the set $V'_{(i)}$ with no incident edge and obtain $\Gamma'_{(i)} \subseteq V'_{(i)}$

2.e.5. Repeat Steps 2.e.1 to 2.e.4 with the second set of vertices $V_{(i)}$ of the main manifold, obtain $\Gamma_{(i)} \subseteq V_{(i)}$ and the matched bipartite subgraph $H_{(i)}$, with no adjacent edge, for each subspace.

2.f. The index of the matched subgraph $H_{(i)}$ with the largest number of vertices, between the matched subgraphs of all the subspaces, represents the most similar gesture in the training set to the given gesture.

2.f.1. For almost the same number of matched vertices in different subspaces the minimum mean value of distances of matched vertices represents the most similar gesture.

2.g. By looking at the index of the vertices of the best-matched graph $H_{(i)}$ recognize direction, the start and the end points of occluded gesture

6. Parallelising the algorithm

Although as the number of vertices of a graph is incremented the processing time of the graph matching algorithm increases as well, the whole classification algorithm is naturally parallel at different levels.

While the projection process of an unknown gesture into many different subspaces and finding the matched subgraph of each subspace can be distributed to many processing units, the graph matching process algorithm is inherently parallel and can be distribute to many processing elements as well, Figure 11.

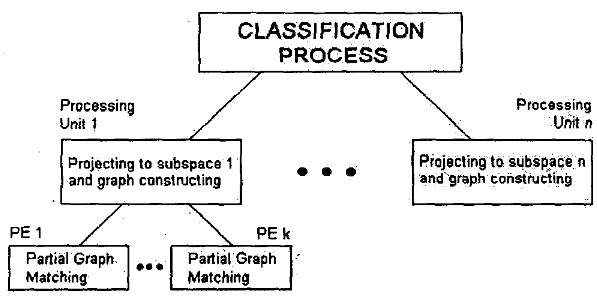

Figure 11. Parallelising at different levels.

7. Experimental results

We did hundred experiments to recognise the input occluded and non-occluded gestures and classify them into 10 different known gestures. Figure 14 shows the gestures we used in our experiments.

With this number of experiments we got 99% recognition rate by using 7 principal components (PC) and 40 vertices in each graph.

In Figure 12 the recognition rate versus number of principal components is shown. With 7 PCs we have the highest recognition rate (equal to 99%). The effect of noise makes the results worse using more than 7 PCs.

Also the number of vertices in each graph has significant effect on the recognition rate. Figure 13 shows the recognition rate versus the number of vertices in the graph of test set gestures.

Different number of vertices changes the position of vertices in the space because different sets of points participate in positioning the vertices. And different positions of points in the space are because of noise.

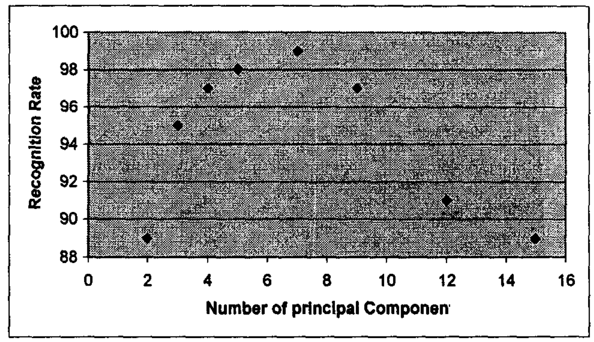

Figure 12. The plot of the recognition rate versus the number of principal components.

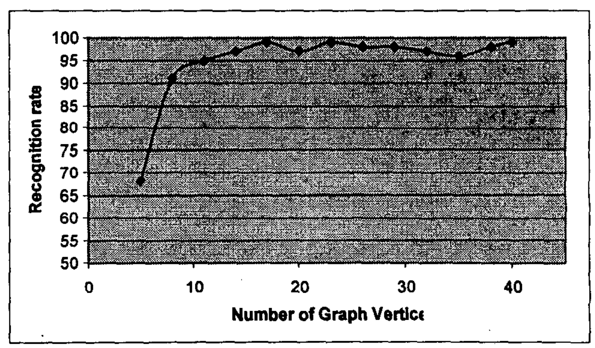

Figure 13. The plot of recognition rate versus the number of graph vertices.

8. Discussion

Although the proposed algorithm is not the perfect solution to the problem of hand gesture recognition it has good advantages because of its features in recognizing partially occluded gestures, detection of start and end point of gesture, direction of gesture and its parallel nature.

Since at the moment we take out the image frames in which the shape of the hand is occluded, one of the future tasks is the automatic recognition of the occluded frames. Also recognition of "co-articulation" , where one gesture influences the next, is another point to be worked on.

9. Conclusion

In this paper we have introduced a new algorithm based on a graph matching technique for classification of different gestures. The main advantages of this algorithm were discussed in detail and a view of expected future improvements to this algorithm described. At the end some experimental results were shown.

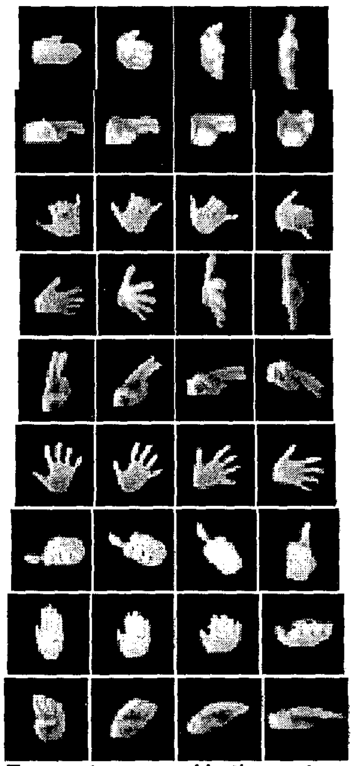

Figure 14. Ten gestures used in the system.

References

[1] R. Cipolla and A. Pentland, *Computer Vision for Human-Machine Interaction,* Cambridge University Press, 1998.

[2] M. Su, H. Huang, C. Lin, C. Huang, and C. Lin "Application of Neural Networks in Spatio-Temporal Hand Gesture Recognition," *Proc. IEEE World Congress on Computational Intelligence,* 1998.

[3] D. Lin, "Spatio-Temporal Hand Gesture Recognition Using Neural Networks," *Proc. IEEE World Congress on Computational Intelligence,* 1998.

[4] V. Pavlovic, R. Sharma, and T. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 19, No. 7, July 1997.

[5] J. Davis and M. Shah, "Toward 3-D Gesture Recognition," *Int'l Journal of Pattern Recognition and Artificial Intelligence,* Vol. 13 No. 3, May 1999.

[6] T. Starner and A. Pentland, "Visual Recognition of American Sign Language Using Hidden Markov Models," *Proc. Int'l Workshop on Automatic Face and Gesture Recognition,* Zurich, Switzerland, June 1995.

[7] J. Schlenzig, E. Hunter, and R. Jain, "Recursive Identification of Gesture Inputs Using Hidden Markov Models," *Proc. Second IEEE Workshop on Applications of Computer Vision,* Sarasota, Dec. 5-7, 1994.

[8] M. V. Lamar, M. S. Bhuiyan, and A. Iwata, "Hand Gesture Recognition Analysis and An Improved CombNET-II," *Proc. IEEE Int'l Conf. Systems, Man and Cybernetics,* Vol. 4, 1999.

[9] B. Moghaddam, W. Wahid, and A. Pentland, "Beyond Eigenfaces: Probabilistic Matching for Face Recognition," *Proc. 3rd IEEE Int'l Conf. Automatic Face & Gesture Recognition,* Nara, Japan, April 1998.

[10] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Detection," *Proc. 5th Int'l Conf. Computer Vision,* Cambridge, MA, June 1995.

[11] B. Moghaddam, "Principal Manifolds and Bayesian Subspaces for Visual Recognition," *Proc. 7th IEEE International Conf. Computer Vision,* ICCV'99, September, 1999.

[12] C. W. Ng and S. Ranganath, "Gesture Recognition via Pose Classification," *Proc. Int'l Conf. Pattern Recognition ICPR'00,* Barcelona, Spain, Sept. 2000.

[13] R. Diestel, *Graph Theory,* Springer-Verlag New York Inc., 1997.

[14] M. Karpinski and W. Rytter, *Fast Parallel Algorithms for Graph Matching Problems,* Oxford University Press, Oxford, 1998.

[15] H. Bunke and K. Shearer, "A Graph distance metric based on the Maximal Common Subgraph," *Pattern Recognition Letters,* Vol. 19, 1998.

[16] H. Bunke, "Recent Developments in Graph Matching," *Proc. Int'l Conf. Pattern Recognition ICPR'00,* Barcelona, Spain, Sept. 2000.

[17] I. T. Jolliffe, *Principal Component Analysis,* Springer-Verlag New York Inc., 1986.

[18] K. I. Diamantaras and S.Y. Kung, *Principal Component Neural Networks Theory and Applications,* John Wiley & Sons Inc., 1996.

[19] A. Pentland, T. Starner, N. Etcoff, N. Masoiu, O. Oliyide, and M. Turk, "Experiments with Eigenfaces," *Proc. Looking at People Workshop Int'l Joint Conf. Artificial Intelligence,* Aug. 1993.

[20] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 23, No. 2, Feb. 2001.