A Multistage Hierarchical Algorithm for Hand Shape Recognition

Mohamed Farouk

School of Computing
Dublin City University
Ireland
mfarouk316@hotmail.com

Alistair Sutherland

School of Computing
Dublin City University
Ireland
alistair@computing.dcu.ie

Amin A. Shoukry

Computer & Systems Eng. Dept.
Alexandria Univ.
EGYPT
aminsh@alex.edu.eg

Abstract—**This paper represents a multistage hierarchical algorithm for hand shape recognition using principal component analysis (PCA) as a dimensionality reduction and a feature extraction method. The paper discusses the effect of image blurring to build data manifolds using PCA and the different ways to construct these manifolds. In_order to classify the hand shape of an incoming sign object and to be invariant to linear transformations like translation and rotation, a multistage hierarchical classifier structure is used. Computer generated images for different Irish Sign Language shapes are used in testing. Experimental results are given to show the accuracy and performance of the proposed algorithm.**

I. INTRODUCTION

Sign language is a visual language that can help deaf people to communicate during their everyday life. Automatic sign language recognition is a challenging area of research where automatic sign language systems can be used to help deaf people to communicate in an easy way with hearing persons who do not know that language.

Signs in a hand tracking system for sign language recognition can be represented as a sequence of movements with different changes in the signer hand like position and shape [1]. In order to acquire the needed information to track the signer hand, some techniques require wearing electronic gloves [2,3]. These gloves are used as a computer human interaction interface to obtain the features of a gesture. To facilitate tracking the signer hand, some other techniques start each gesture from a predefined position [4].

The proposed algorithm is an automatic sign language classifier that helps in detecting the shape of a signer hand from twenty shapes in Irish Sign Language. The only restriction it imposes is a black background and wearing black clothes to ease the hand segmentation from the video frames. The proposed algorithm is invariant to linear transformations like rotation and translation. It detects the signer arm rotation angle and its position for each individual video frame from the acquired sequence by using translation manifolds that represent different shapes at different rotations.

The rest of the paper is organized as follows. Section 2 reviews the related work. Section 3 overviews the multistage hierarchical classifier. Section 4 discuses the different variations of the algorithm. Experimental results are discussed in section 5 followed by the conclusion in section 6.

II. RELATED WORK

Principal component analysis (PCA) has been used in many applications to reduce a complex dataset to a lower dimensional space to reveal the sometimes hidden, simplified structure that often underlies it. In Vision applications it has been used for the representation and recognition of faces [5,6], recognition of 3D objects under varying pose [7] and tracking of deformable objects [8,9]. PCA assumes that the set of model images lie in a low dimensional subspace spanned by the eigenvectors of the covariance matrix for a group of images. Images can be considered as objects in a high dimensional space spanned by the values of the individual pixels. PCA is used to map these images to a low dimensional linear space that represent the images of interest. A manifold in a low dimensional space is constructed by projecting the original images into that eigenspace using a number of eigenvectors that have the highest eigenvalues. However, manifolds are often non-linear. The proposed algorithm tests the effect of image blurring on reducing the non-linearity of the manifolds by using a Gaussian filter.

III. PROPOSED ALGORITHM

The proposed algorithm is based on calculating a series of manifolds to represent the translation, rotation and shape. Applying PCA on a set of example images generates these

manifolds. The eigenvectors with the largest eigenvalues represent the directions in feature space along which there is the greatest variation in the data. We select for example three eigenvectors with the largest eigenvalues to form the basis of a low-dimensional space. We project the example images into this eigenspace and each image is represented by a point in that eigenspace. PCA can be used as part of a classification algorithm. Suppose we have two classes of objects. We can calculate an eigenspace for each class. An unknown object can be classified by calculating the "perpendicular distance" of it from each of the two eigenspaces.

To illustrate our algorithm we apply it to a dataset of computer generated images of a human arm and hand see Figure 2. The dataset contains 20 different hand shapes from the Irish Sign Language alphabet. Each hand-shape is rotated from +90 degrees to +180 degrees in intervals of 2 degrees. "Translation manifolds" are generated by translating each hand-shape image from -5 to +5 pixels in the horizontal and vertical directions to build a manifold of 121 objects as shown in Figure 3. At each rotation angle we create a set of 20 translation manifolds representing 20 different hand-shapes. As the range of angles to be represented increases, the number of manifolds increases and consequently the space of manifolds to be searched becomes larger.

Our objective is to classify the hand-shape and measure the rotation angle and the translation position in an efficient and accurate way.

A. The different stages of the algorithm

A multistage hierarchical structure is used to reduce the search space from one stage to another to find the right manifold to search in and hence decide the shape, rotation, and translation position of an incoming sign object. Figure 1 shows the different stages of the proposed algorithm. At stage 1 it estimates the rotation angle of the incoming object. At stage 2 it searches through the 20 manifolds for the different shapes at the estimated angle. At stage 3 it searches within the translation manifold of the estimated shape and hence detects the translation position.
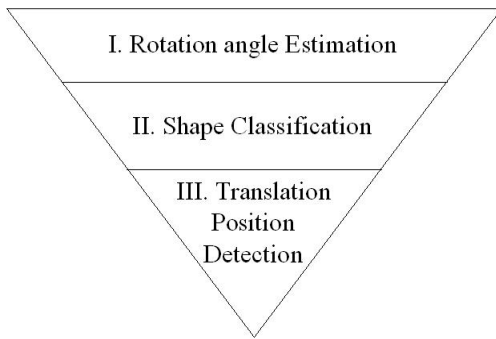
Figure 1. The multistage hierarchy

The proposed algorithm explores the effect of image blurring using Gaussian Kernels. Image blurring has the effect of removing small changes between objects. At this level of blurring the dominant effect will be the rotation angle and it will be easy to estimate it. Figure 2 shows the effect of blurring on a computer generated image for sign 'A' in the Irish sign Language.
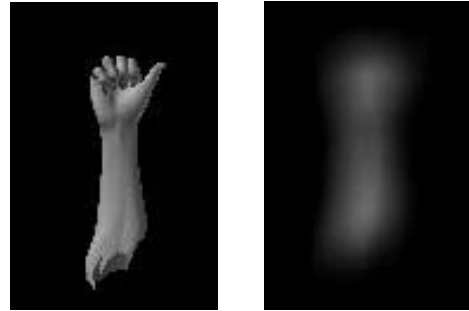
Figure 2. The effect of blurring on the images

Image blurring not only has the effect of grouping shapes together but also it has an effect on the structure of the translation manifolds. Image blurring reduces the nonlinearity of the translation manifolds and makes the manifolds more flat. This makes it easier to calculate the perpendicular distance to a manifold and to differentiate between neighbouring manifolds. Figure 3 shows two translation manifolds for sign 'A' (projected onto the 2nd and 3rd eigenvectors) before and after blurring.
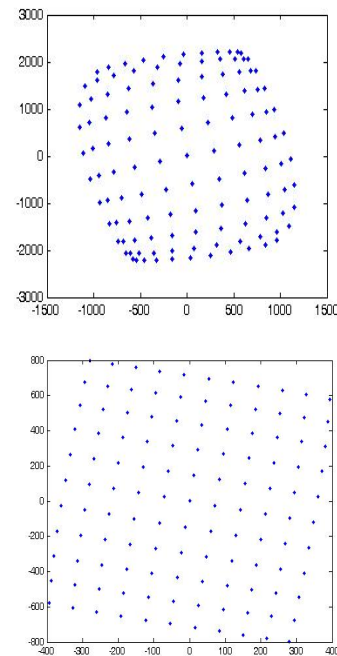
Figure 3. The effect of image blurring on a translation manifold

1) Estimating the rotation angle: The incoming object is compared with data manifolds for a selected shape from the 20 shapes at discrete angles through the range of angles to be classified. Perpendicular distance is used to compute the distance between the incoming object and the centroid of each manifold to find the first and the second nearest manifolds. These two manifolds represent the start and the end of an interval for a smaller range of angles to search in. As rotating the signer arm in 2D space dramatically affects the classification process and could mislead in classifying the shape, a further narrow search is done through the half of the estimated interval to get the closest angle and hence get the right class in the next stage. According to how much the incoming object is similar to the start or the end manifolds of that interval, the decision is taken to search in the manifolds of the first half or the second interval half. As blurring removes only small changes between objects, a single selected shape is not enough to compare with in this stage. However, six objects from the 20 shapes are chosen to compare with. To pick them, a k-medoid clustering algorithm can be used where these objects should be the most centrally located objects between the 20 shapes. Figure 4 shows the flowchart for stage1.
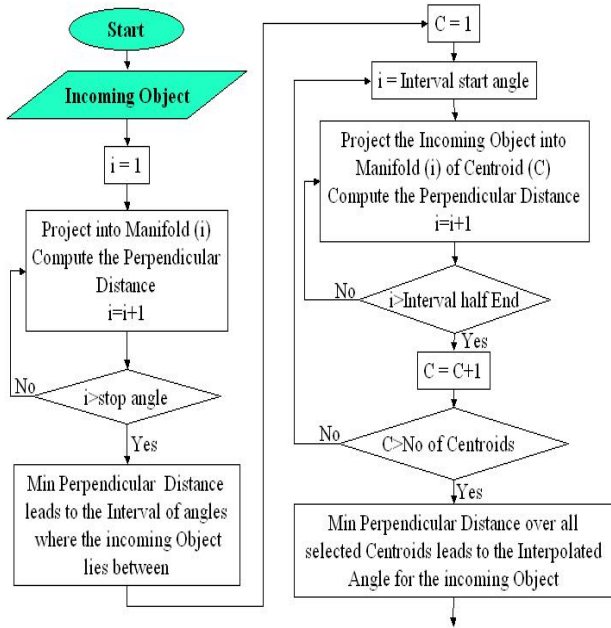

Figure 4.   Stage 1 flow chart

2) Shape Classification: As the angle of rotation has been estimated in stage 1, a search through the 20 translation manifolds at the estimated angle is needed. The perpendicular distances between the incoming object and the different manifolds for the 20 shapes are computed. The shortest distance leads to the nearest neighbour shape. Figure 5 shows the flow chart for stage 2 and 3.
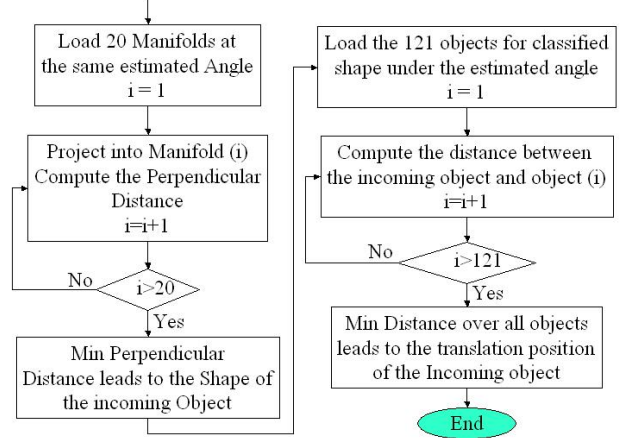

Figure 5.   stage 2 and 3 flow chart

3) Detecting the translation position: From stage 1 and 2, the rotation angle and the shape for the incoming object are classified. At stage 3, a search through the 121 objects that represent the translation manifold for the classified shape at the estimated angle is needed. The incoming object should be projected into that eigenspace where the Manhattan distance is used to compute the distances between it and the translation objects within that manifold. The nearest neighbour between the 121 objects leads to the translation position of the signer arm. Figure 7 summaries the multistage hierarchical structure.

A variable separation technique can be applied, instead of searching linearly through the 121 objects to reduce the time of search as shown in Figure 6. The object with the cross sign represents an incoming object with translation 2 pixels to the right and 3 pixels to the up. The variable separation is done first by searching through the 11 objects, which represent the horizontal translations with zero vertical translation. The nearest neighbour leads to the horizontal translation of the incoming object, where the vertical translation is detected by searching through the 11 objects that represent the different vertical translations at the classified horizontal translation as shown in the figure. The variable separation reduces the search from 121 objects to only 22 objects.
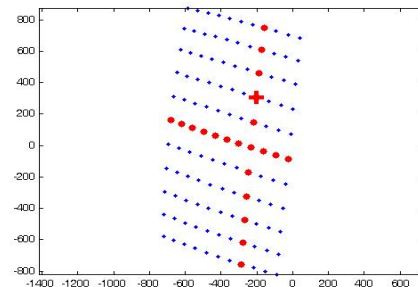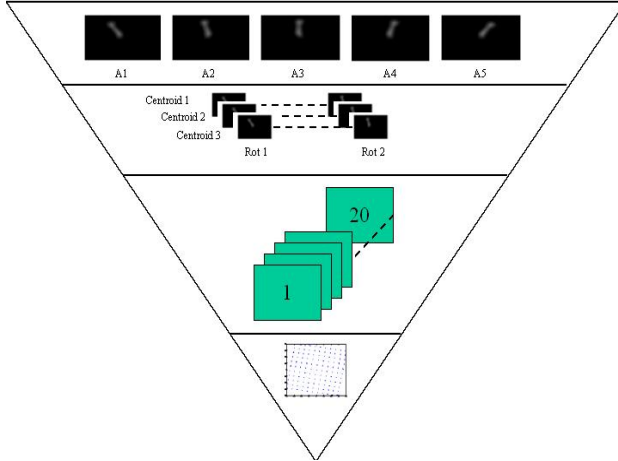

Figure 6.   Variable Separation

Figure 7.   Overview of the algorithm

B.   *Fine Search*

When moving in a top down direction or when moving from a higher to a lower level through the multistage hierarchical algorithm, the scope of the search narrows. Therefore, if the higher level decision is wrong there is no way to backtrack. Therefore, a fine search strategy is needed to increase the accuracy of the proposed algorithm. Instead of taking a hard decision about the angle of rotation after stage 1, stage 2 is applied both to the first and the second nearest estimated angle for the incoming object. The perpendicular distances between the incoming object and the 40 manifolds for the 40 shapes of the two estimated angles are sorted. A decision about the shape and the translation position is taken at stage 3 where the incoming object is compared with the data objects within the manifolds instead of depending only on the manifolds centroids (as in stage 2). To fine search the results stage 3 is applied to the first N nearest neighbours from stage 2 as will be discussed in the experimental results.

IV.   VARIATIONS ON THE ALGORITHM

In this section we present some variations on the proposed algorithm, which may offer improvements in speed and accuracy. In the algorithm, which we have just presented, a separate manifold for each hand-shape at each angle of rotation is generated offline. The number of eigenvectors used for the classification process is important as it can affect both efficiency and accuracy. The size of each eigenvector is also important. In each separate manifold, the size of each eigenvector is actually the size of the original image. In-order to reduce the size of the eigenvectors, we introduce the concept of "shape manifolds" and "shape sub-manifolds".

As shown in Figure 8, a "shape manifold" is a manifold that holds different shapes within the same eigenspace. It is constructed by applying PCA directly to the data objects of the 20 shapes at the same angle of rotation. The 20 shapes

share the same eigenspace where one eigenspace is needed to project them, unlike separate manifolds, where 20 eigenspaces were used for separate manifolds in-order to get the right class or shape for the incoming object.
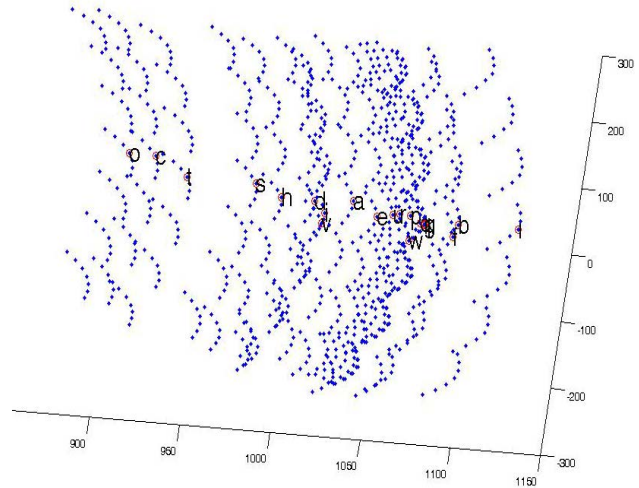

Figure 8.   Shape Manifold

"Shape sub-manifolds" are separate manifolds within a shape manifold eigenspace. They are generated by applying PCA a second time to a shape manifold. The data is first projected into the eigenspace of its shape manifold. Then PCA is applied to each set of 121 projected objects for each shape to get its sub-manifold. The size of the eigenvectors for the shape sub-manifolds is lower because it is equal to the number of eigenvectors used from the parent eigenspace to build them. Stage 2 in the original algorithm loads 20 eigenspaces for the 20 shapes at the estimated angle of rotation to classify the shape of an incoming object. The Shape sub-manifolds could be used to enhance the performance in stage 2 because the size of the eigenvectors is lower than the size of the eigenvectors when using separate manifolds.

To extend the idea of using shape sub-manifolds, rotation manifolds can be generated by building one eigenspace for each shape at different rotations. Figure 9 shows a rotation manifold for sign 'A'.

A "rotation manifold" is constructed by applying PCA directly to the data objects for a certain shape at different rotation angles. These objects share the same eigenspace in a rotation manifold. "Rotation sub-manifolds" are separate manifolds within a rotation parent eigenspace. They are generated by applying PCA a second time to each set of 121 objects at different rotations in a rotation manifold eigenspace.

Rotation sub-manifolds can be used in stage 1 instead of separate manifolds. Stage 1 loads different manifolds at different rotations to estimate the angle of rotations for an incoming sign object. Rotation sub-manifolds are efficient

to be used in stage 1, as the size of their eigenvectors equals to the number of eigenvectors in the parent rotation manifold eigenspace used to generate them.
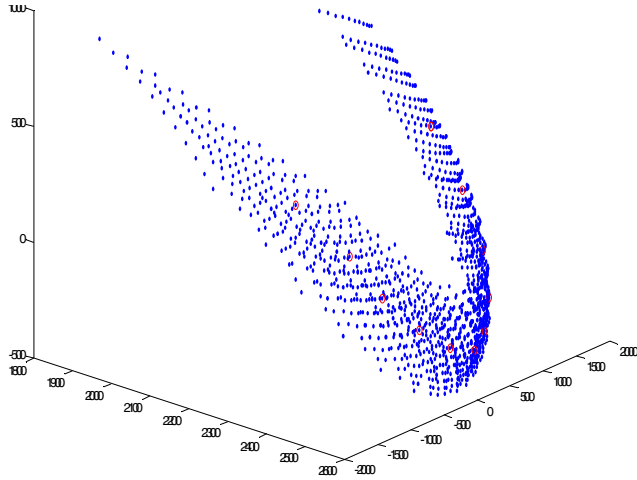

Figure 9.   Rotation Manifold

V.   EXPERIMENTAL RESULTS

The proposed multistage hierarchical algorithm has been implemented in Matlab under Windows XP operating system. All the experiments are done on Intel® Pentium® Dual CPU T2370 @ 1.73GHz, 2.00 GB of RAM. The resolution of the images used in the testing is 250x330. All the images are converted into grayscale before blurring. Blurred images are created using a two-dimensional Gaussian lowpass filter of size [6,6] and with standard deviation equal to 10. In-order to examine the performance and the accuracy, the proposed algorithm has been tested using different data manifolds structures at the same blurring level based on selecting signs 'A', 'C', 'H', 'I', 'P', 'R' for stage 1 as the most centrally located objects between the different shapes in the all experiments. The results are based on a data sample of 14520 (6x20x121) objects. The sample represents the 20 shapes at the first 6 angles from the range of angles that represents the two images in Figure 10.

Figure 10.   Range of rotation angles

The proposed algorithm is tested using different numbers of eigenvectors. As the number of eigenvectors increases, the experimental results show that the accuracy increases monotonically until saturation. It was interesting to find the best number of eigenvectors to use under the examined test data. Figure 11 shows the accuracy of using separate manifolds under different numbers of eigenvectors from 2 to 10.
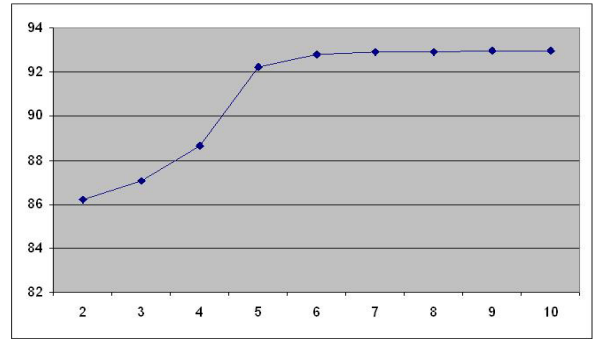

Figure 11.   Separate Manifolds Accuracy

There are different data manifolds structures as discussed in section 4. To build the shape sub-manifolds and rotation sub-manifolds, the first highest 7 eigenvectors from the parent eigenspace are used. Experimental results have been done on these different aspects to examine the performance and the accuracy of them.

TABLE I.   ACCURACY AT THE BEST NUMBER OF EIGENVECTORS

Method	Number of Eigen- Vectors	Accuracy %
Separate Manifolds	7	92.9
Shape Manifolds	3	48.5
Shape Sub-Manifolds	5	90.9
Shape Sub-Manifolds + Rotation Sub-Manifolds	5 , 3	87
Shape Sub-Manifolds + Rotation Sub-Manifolds + Fine Search	5 , 3	93.2

Table I lists the best number of eigenvectors for each of the different aspects used to implement the proposed algorithm, to get the highest accuracy. The best accuracy is for separate manifolds. To enhance the performance shape manifolds are used, where the accuracy starts to get down. Using shape sub-manifolds could enhance the accuracy and get more speed up over the previous two aspects as shown in Table II. Using rotation sub-manifolds beside shape sub-manifolds enhance the performance and to get a comparable accuracy fine search is used.

TABLE II.   THE PROPOSED ALGORITHM EXECUTION TIME

Method	Speed (sec)
Separate Manifolds	3.5
Shape Manifolds	1.3
Shape Sub-Manifolds	1.21
Shape Sub-Manifolds + Rotation Sub-Manifolds	0.25
Shape Sub-Manifolds + Rotation Sub-Manifolds + Fine Search	0.35

Figure 12 show a bar chart for the performance of the multistage hierarchical algorithm.
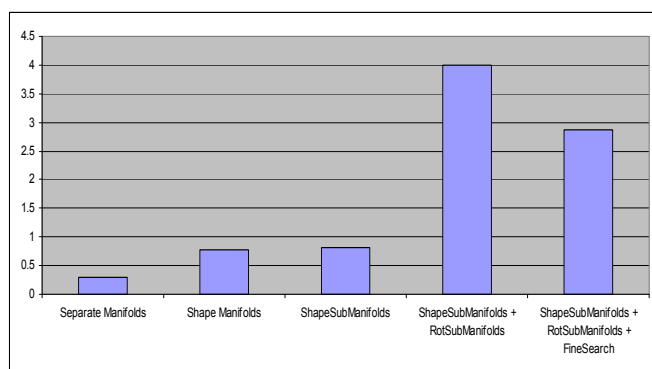

Figure 12. The performance of the proposed algorithm

To fine search the results the first N nearest neighbours are used. Figure 13 shows the accuracy versus N where the highest accuracy is obtained when N equals to 3.
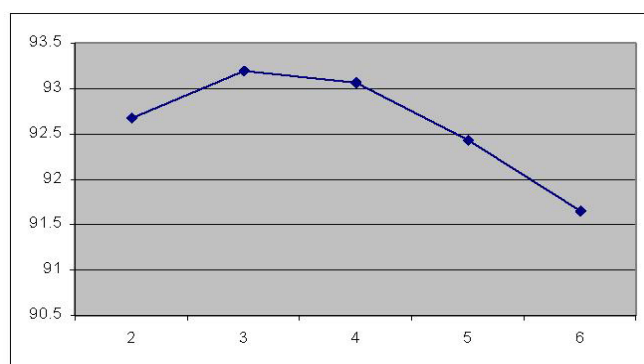

Figure 13. The Fine Search Accuracy

Referring to the error analysis of the objects based on using shape sub-manifolds and rotation sub-manifolds with fine search, Figure 14 shows the error analysis of each object according to the translation position within the translation manifolds. From the figure, it is clear that the error of the outer translated objects is greater. This is because of the non-linearity in the surface of the manifolds. Although image blurring is used to reduce the non-linearity, the surface of the generated manifolds is not completely flat.

13	14	13	15	16	16	17	16	16	19	25
9	7	6	6	9	13	14	15	15	14	18
8	6	4	4	4	5	5	7	10	14	15
4	2	1	3	4	4	4	4	3	6	12
2	3	1	1	1	2	2	2	2	5	6
7	5	4	5	5	5	4	5	5	5	6
6	4	5	4	5	6	6	7	5	5	8
6	5	5	5	5	4	6	6	7	5	8
13	8	6	5	5	4	4	4	5	6	8
17	15	13	11	10	6	5	5	4	8	8
23	22	21	21	20	17	13	10	8	6	16

Figure 14. Error Analysis

VI. CONCLUSIONS

In this paper we have proposed a multistage hierarchical algorithm to classify different hand shapes from the Irish Sign Language by building data manifolds. Blurring the images has the effect of flattening the manifolds and reducing the non-linearity. Training examples were obtained by using computer-generated images. The experimental results show that using both rotation sub-manifolds and shape sub-manifolds with fine search gives the best accuracy-speed performance measure.

VII. REFERENCES

[1] Zahedi, M., P. Dreuw, D. Rybach, T. Deselaers, and H. Ney, "Using Geometric Features to Improve Continuous Appearance-based Sign Language Recognition", BMVC, vol. 3, Edinburgh, UK, pp. 1019-1028, 2006.

[2] R. Bowden, D. Windridge, T. Kadir, A. Zisserman, and M. Brady. A linguistic feature vector for the visual interpretationof sign language. In ECCV, vol 1, pp. 390–401,2004.

[3] Vogler and D. Metaxas. A framework for recognizing the simultaneous aspects of american sign language. CVIU, 81(3):358–384, Mar. 2001.

[4] A. S. Junwei Han, George M. Awad and H. Wu. Automatic skin segmentation for gesture recognition combining region and support vector machine active learning. In FG, Southampton, UK, Apr. 2006.

[5] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. In Journal of the Optical Society of America 4, pages 519-524, 1987.

[6] M. Turk and A. Pentland. Eigenfaces for recognition. In Journal of Cognitive Neuroscience 3(1), 1991.

[7] H. Murase and S.K. Nayar. Visual learning and recognition of 3D objects from appearance. In International Journal of Computer Vision, 14(5-24), 1995.

[8] Michael J. Black and Allan D. Jepson. EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation. In Proceedings of the European Conference on Computer Vision, pages 329Œ342, Cambridge, England, 1996.

[9] Shashua, A., Levin, A., Avidan, S., Manifold pursuit: a new approach to appearance based recognition, ICPR02(III: 590-594). 2002.