Dynamic Hand Gesture Recognition using PCA, Pruning and ANN

S.M. Shitole
Department of E&TC, Sinhgad College Of Engineering, Pune.

S.B. Patil
Department of E&TC, Sinhgad College Of Engineering, Pune.

S.P. Narote
Department of E&TC, Sinhgad College Of Engineering, Pune.

ABSTRACT
Gesture recognition has the potential to be a natural and powerful tool supporting efficient and intuitive interaction between the human and the computer. The primary goal of hand gesture recognition research is to create a system which can identify specific hand gestures and use them to convey information for device control. This paper proposes a vision-based dynamic hand gesture recognition system using PCA, pruning and ANN. The system will find hand location in each frame. The hand segmentation is done using color marker for PCA. Skin color detection method is used for pruning and ANN. After connecting all these hand locations, the template is generated at the end of Segmentation. Template is used to extract the feature for training and testing purpose. The implementation uses the eigenspace which is determined by processing the eigenvalues and eigenvectors of the image set. The experimental results show ANN method gives more recognition rate than PCA and pruning.

Keywords
Dynamic hand gesture recognition, Eigen values and Eigen vector, human-computer interaction, PCA, Pruning, ANN etc.

1. INTRODUCTION
Hand gesture recognition has gain a lot of importance in the research due to the increasing demands for human-computer interaction(HCI) in recent years. vision based gesture recognition system has many advantages and potential to replace devices like Keyboards, mouse, or electronic gloves. As computer become important part of today's lifestyle, providing human-computer interaction (HCI) will have a positive impact on their use because hand gesture provides a natural and intuitive way of communication for human-computer interaction. In this paper, an effective hand gesture recognition method is demonstrated, which can be use in many applications like augmented reality (AR), human-computer interface, sign language recognition image/video coding, content-based image/video retrival, and video games etc.

Gesture can be static or dynamic. Static gesture means user with certain fix pose or configuration. Dynamic means gesture with prestroke, stroke, and poststroke phases. In simple language, for static gestures, image has to process and for dynamic gestures, frames of the video has to process. Some gestures also have both static and dynamic elements like sign language. Lot of work has done in static gesture recognition. Dynamic gestures are more challenging to handle than static one. To recognize the gesture, human body position, configuration (angles and rotation), and movements (velocities) should be sensed accurately. There are two ways to do this i.e. two types of gesture recognition systems, appearance based and vision based. In appearance based system, the above parameters are sensed by using magnetic field trackers, instrumented (data) gloves, and body suits. These tracking devices can detect fast and subtle movements of the fingers when user's hand is moving. Each sensing technology varies along several dimensions, including accuracy, resolution, latency, range of motion, user confront and cost. In this method, user has to wear a burdensome device and carry a load of cables connecting device to the computer. This affects the comfort and naturalness of the user's interaction with computer. Vision based techniques overcome this disadvantages. In this method, different necessary parameters are sensed using cameras and different computer vision techniques. This type of system uses image properties such as texture and color for analyzing a gesture, while tracking device cannot. Vision based techniques can also vary among themselves. So different issues need to consider are number of cameras, their speed and latency, background or user constrain, etc[1][2]. This paper discuss a vision based dynamic hand gesture recognition system, which will recognize the dynamic gestures for 0,1,2,3, and L made by user in the video. There are two systems : System using color marker and system without color marker. Color marker makes the segmentation algorithm easy but it impose restrictions on user. For the first system PCA is used and for the second pruning and ANN is used. Total 60 videos are processed and result is analyzed for pruning, ANN.

The paper is organized as follows, Section 2 summarizes literature related to different hand gesture recognition systems. Section 3 describes system architecture which includes preprocessing, segmentation, and feature extraction. In section 4, classifiers are discussed in brief. Section 5 and 6 includes results and conclusions of this paper.

2. LITERATURE SURVEY
Gesture-based interaction was firstly proposed by M.W. Krueger as a new form of human–computer interaction in the middle of the seventies [3], and there has been a growing interest in it recently. Good literature survey available of Vladimir[1] and Ali Erol[2].

Any gesture recognition system generally consist of three important part gesture segmentation, gesture analysis(Feature extraction), and gesture recognition. Surveys on gesture segmentation is given nicely in [4]. In hand gesture segmentation stage, regions which represent the hand gestures are to be distinguished from the background. Various image processing techniques have been applied here to segment the region of interest. Basically there are three way to do gesture segmentation image information, motion based analysis, and

multiple cue. Detailed study of these three method is given in [4]. From the previous work it is clear that, skin color modeling can be done well in HIS, HSV colorspace rather RGB colorspace[5].

Once the region of interest is obtained, next step is to extract features from it which will be input to gesture recognition stage. The selection of features is depends on modeling technique. In order to interpret gesture accurately, it is important to first consider which model to use. Scope of a gestural interface for HCI is directly related to the proper modeling of hand gestures. Gesture modeling is generally depends on application A very coarse and simple model may be sufficient for simple application like tracking [6][7]. However, if the purpose is a natural-like interaction, a model has to be established that allows many if not all natural gestures to be interpreted by the computer. Good review on hand gesture modeling is available in literature [1][2]. It also gives us different features can be used and advantages and disadvantages of the same.

The last step is the hand gesture classification, whereby meaningful hand gestures are defined and recognized using the training data and estimated parameter. Gesture recognition is the phase in which extracted features are analyzed to recognized specific gesture. According to application and requirement, different approaches like Statistical modeling, computer vision and pattern recognition, image processing, connectionist systems are used in this phase. Survey on gesture recognition tools are given in [8].

Psarrou et al. [9] who used HMM for behavior recognition. He uses HMM to recognize sequences of tennis strokes based on quantized time-sequential binary images. However, these methods are too much dependent on probability factors, time constraint, iterations, etc., resulting in being computationally expensive. The proposed eigenspace method, on the other hand, does not require geometrical calculation or partial segmentation of the models. Therefore, the application of this technique to hand gessture recognition is simpler and computationally less expensive than any other methods. The eigenspace method is often used to analyze huge amount of data. M. Masudur et al. [10] has used eigenspace for human motion recognition.

Rybach [11] investigated appearance-based features for the person-independent vision-based recognition of continuous sign language. In this method segmentation of the images is not done. The image itself acts as a feature. He used combination of features including PCA (for dimensionality reduction), hand position, hand velocity and hand trajectory. He used RWTH-BOSTON-104 database is used for algorithm testing. Solomon [11] has used Principle Component Analysis (PCA) algorithm for gesture recognition system. PCA has the distinction of being the optimal linear transformation for keeping the subspace that has largest variance. The principal component feature vector y is a point in the orthogonal coordinate system defined by the eigenvectors.
Jonathan Alon[6] has introduce the concept of pruning to speed up the gesture classification process. The main idea is to take consideration of only important feature while rest of the feature can simply prune or discard. The author also introduces the subgesture reasoning algorithm to deal with subgesture problem, i.e. the fact that some gestures may be very similar to subgestures of other gestures. For example, gesture for symbol 5 is the subgesture of symbol 8. So for large dataset, to avoid the confusion of the classifier, subgesture reasoning algorithm can be used.

One of the pattern recognition technique is ANN which discriminate the functionality from extracted features. ANNs are formed of cells simulating the low-level functions of biological neurons. ANNs are particularly useful for complex pattern recognition and classification tasks. The capability of learning from examples, the ability to reproduce arbitrary non-linear functions of input, and the highly parallel and regular structure of ANNs make them especially suitable for pattern classification tasks [13].

3. SYSTEM ARCHITECTURE

Two systems are discussed in this paper for dynamic hand gesture recognition. First system uses the color marker where other does not. Overview of these systems are given in this section.

3.1. System using color marker

The aim of this paper is to present a model for dynamic gesture recognition. Proposed system classify dynamic gestures into four category :Zero, one, two, and three. The idea of user made gesture will get clear from fig. 1. To reduce the complexity of segmentation algorithm three assumptions are made. They are,

1) Only hand region is present in the video.
2) Background is static.
3) Red LED or marker is present at the tip of finger.

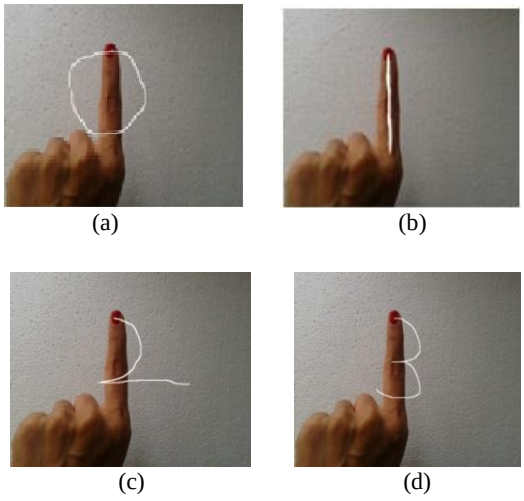


Fig 1: Example digit gestures for classes "0", "1", "2", and "3", as performed by a user.

The block diagram of our proposed system is shown in fig. 2. Simple pattern recognition technique to the problem of hand gesture recognition. This method has training phase and testing phase. In the training phase, the user shows hand gestures which were captured using USB based Fronttech e-cam camera. The algorithm is tested for dynamic gesture of zero, one, two, and three. All videos are captured at 30 FPS in AVI format and saved to disk. For implementation of this method, an ordinary workstation with no special hardware beyond a web camera is used.

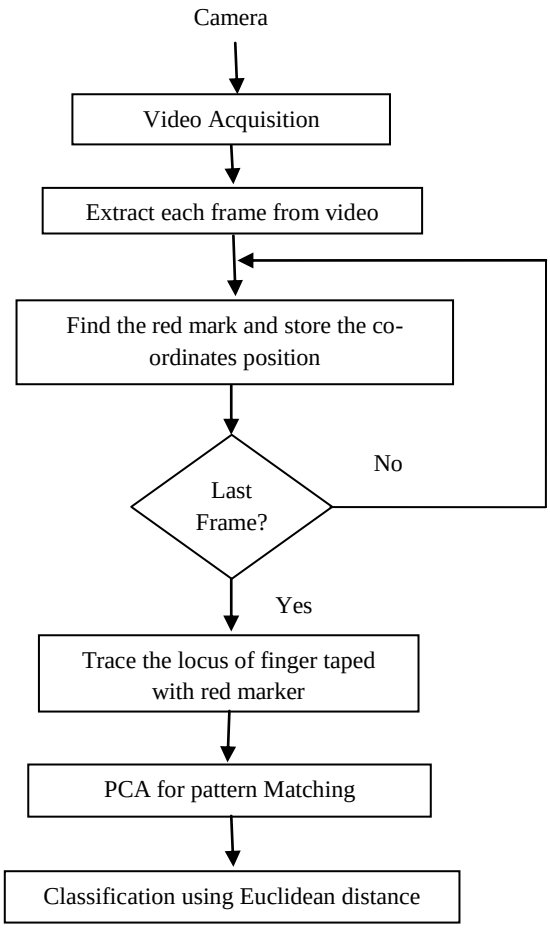


Fig 2: Algorithm for system using color marker.

Once gesture is detected in video, the fingertip has to trace. In each frame, red spot will be detected and its position is stored. By using the positions of the red spot, at the end, gesture matrix set is form. This matrix will be used for pattern matching to detect the gesture. So following are the general steps.

Hand gestures are acquired using a web camera that monitors the hand gestures. Then the frames are separated from video. Segmentation is achieved using a simple low-cost red color filtering in the RGB color space in each frame. The homogeneous background is chosen as it provides contrast with the users' red color marker which allows for a robust finger tip detection in real-time. The complexity of segmentation reduces due to homogeneous background which is helpful for gesture segmentation. The position of red spot in each frame is stored. After finishing with all frames, next step is to trace the locus of finger taped with red marker. This is done by forming binary image using stored location of red spot. Now these spot may or may not connected so morphological operation has to perform to get proper image. Now this image will be input to PCA. The network will be trained using same procedure.

3.2. System without color marker

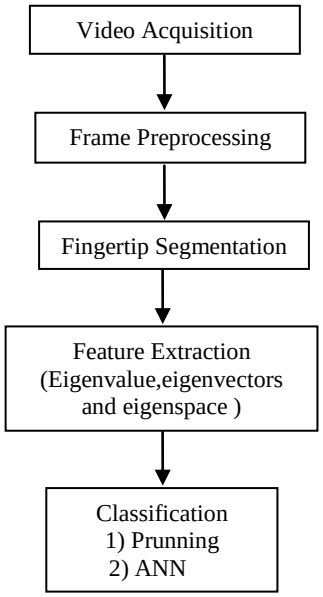


Fig 3: Block Diagram of system without color marker.

Block diagram of a vision–based dynamic hand gesture recognition system is as shown in Fig. 3; which involves following steps: video acquisition, video- preprocessing, fingertip segmentation, feature extraction, and classification by using either pruning or ANN.

3.2.1. Pre-processing

The video was captured by web camera with the frame resolution of the video is 320×240. The video frames are first normalized by subtracting the mean of the image from each pixel Which can also be performed using histogram equalization but it takes relatively more number of computations. The normalization removes most of the frame differences due to different lighting condition and other noises added due to the use of a low image quality camera. After preprocessing image is send to segmentation block.

3.2.2. Fingertip Segmentation

The gesture points are aquired using the skin tone of the candidate. The empirically set value of H,S and V are used because HSV colorspace can separate skin color efficiently than RGB colorspace. The hue plane is not taken into consideration in this system as it is taken care in the normalization process. The gesture points i.e. skin color pixel is obtained and the co-ordinates of those are recorded. Training videos are analyze to decide the values H, S and V for skin color detection. In each frame, scanning will start from rightmost corner. Whenever skin color is detected, the respective co-ordinate position is stored. At the end, each set of gesture points are connected together to get the final template which is used for further processing in the operation of the classification.

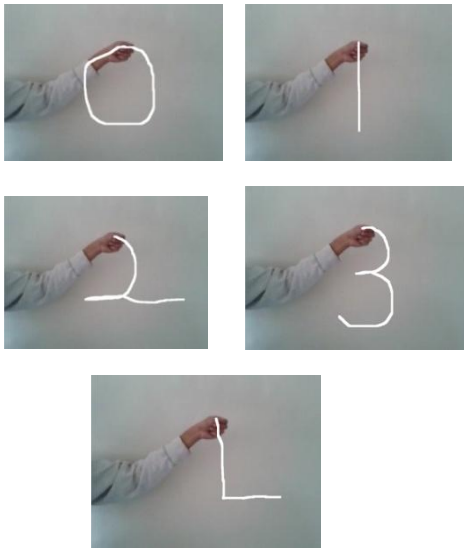


Fig 4: Example of digit gestures for classes "0", "1","2", "3", "L".

Fig 5: Final template at the gesture segmentation step.

3.2.3. Feature Extraction

Selection of features is the important steps as the accuracy and robustness of the classification is depends on that. At the of segmentation process, there will be formation of the binary image , in which gesture symbol can be seen as shown in fig 5.

An image I(x,y) is represented by a two dimensional array of intensity values. This image can also be considered as a vector of dimension equal to number of pixels in the image. This is obtained by reading pixel brightness values in a raster scan manner. After representing image in vector space, calculation for eigen vlues, eigen vector , eigen space is done which will act as a features for classification purpose.

Let us consider, five different gestures has to classify using this algorithm and have p number of training video. Where p is multiple of 5 that means for every gesture there are equal number of gesture videos. During training, at the end of segmentation block, p gesture templates or images Ip (p = 1,2,3,……,p) will be available. The sampled image (or posture) Ip having 320×240 pixel size is converted into a column vector of the form

$$X = \left[\sum_{i=1}^{320}\left(\sum_{j=1}^{240} X\text{ij}\right)\right]^{\text{T}}$$

By arranging pixels in a raster scan manner, the size of this column vector is 76800×1. This procedure is repeated for p images which and placing these vectors side by side there will be the final matrix of size 76800×P.
Then we will normalized this vector to 1 i.e,

$$\|X\| = 1.$$

The next step involves computation of covariance matrix :

$$C = X^{T}X$$

This matrix is of size p×p.
Eigenvalues λ_j with its corresponding eigenvectors e_j of matrix C is calculated using an eigenequation $Ce_j = \lambda_j e_j$. The N dimensional space defined by all the eigenvectors of C is reduced via the principal component analysis [11]. As a result, the chosen K ($1 \leq K << N$) eigenvalues λ_k (k = 1,2 ,. . . ,K) and corresponding eigenvectors e_k are obtained. The K-dimensional space defined by the eigenvectors e_k is called an eigenspace.

Image x_p is then projected into a point g_p in the eigenspace by the following equation;

$$g_p = (e_1, e_2, \ldots\ldots e_k)^T X_p.$$

4. CLASSIICATION

PCA approach is used for the system with color marker and by calculating Euclidean distance, gesture is recognized. For the system without color marker, pruning and ANN approach is used.

4.1. System using color marker

Principle Component Analysis (PCA) algorithm is used for the first system. Though PCA is sensitive to lightening changes, this algorithm has been used for gesture recognition because the environment that will be used to obtain the individual gestures is controlled and hence lighting variation can be minimized. Also, it allows us to quickly add new gesture to the gesture database, making it better suited for real time applications. This method reduces data dimensionality by performing a covariance analysis between factors.

The way pattern recognition is done is computer vision is to measure the difference between the new image and the original images, but not along the original axes, along the new axes derived from the PCA analysis. It turns out that these axes works much better for recognizing gestures, because the PCA analysis has given us the original images in terms of the differences and similarities between them. The PCA analysis identifies the statistical patterns in the data.

The gesture recognition using PCA algorithm that involves two phases :

- Training phase
- Recognition Phase

During the training phase, each gesture is represented as a column vector, with each entry corresponding to gesture pixel. These gesture vectors are then normalized with respect to average gesture. Next, the algorithm finds the eigenvectors of the covariance matrix of normalized gestures by using a speed up technique that reduces the number of multiplications to be performed. Lastly, this eigenvector matrix then multiplied by each of gesture vectors to obtain their corresponding gesture space projections.

In the recognition phase, a subject gesture is normalized with respect to the average gesture and then projected onto gesture space using the eigenvector matrix. Next, the Euclidean distance is computed between the projection and all known projections. The minimum value of these compares ions is selected and compared with the threshold calculating during the training phase.

4.2. System using color marker

This system is tested using two classifier i.e pruning and ANN.

4.2.1. Prunning

As discuss above, the eigenspace is reduced for less complexity. Only principal components are taken into consideration. The algorithm has been tested for five gestures so first five prime eigen values and corresponding vector are taken into consideration, rest of eigen values and eigen vectors are simply get prunned.

Let us consider, an image containing unknown gesture p', denoted by $I_{p'}$, it produces an edge image $E_{p'}$ and then a vector $x_{p'}$. It is projected onto a discrete point $g_{p'}$ in the eigenspace. The minimum distance d_{p*} defined by,

$$d_{p*} = \min_p \| g_{p'} - g_p \|$$

is calculated to find the nearest learned point in the eigenspace. For a certain threshold ε (>0), if $d_{p*} <$ ε holds, it concludes that gesture is similar to the one who is giving the minimum distance. In this method, minimum description length principal is employ for gesture recognition. Here threshold ε is required to avoid detection of false gesture i.e. gestures which is not present in our database.

4.2.2. ANN

An Artificial Neural Network (ANN), usually called neural network (NN), is a mathematical model or computational model that is inspired by the functional aspects of biological neural networks. A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase.

In this paper, the Cascade-feedforward network is used. Levenberg-Marquardt (LM) algorithm is used which is most widely used optimization algorithm. The algorithm determines how to adjust the weights to minimize performance by using the gradient of the performance function. Levenberg-Marquardt (trainlm) algorithm was for training because It is the fastest method for training of moderate-sized feedforward neural networks and based on numerical optimization techniques. dividing the training input data is divided: 67% for training, and 33% for testing. Furthermore, the number of data points in training set was more than sufficient to estimate the total number of parameters in the network. Early stopping method also applied to improve the generalization of the network. Two early
stopping conditions were used: either total mean squared error <=0.001 or training stopped after 20 epochs. The weights and bias of input layer and hidden layer were saved after each training session. When the simulation results are not satisfactory, the network trained again with the last saved weight and bias values. This was done to improve the network performance and to reduce the number of time for training.

The Levenberg–Marquardt (L–M) algorithm outperforms simple gradient descent and many other conjugate gradient methods in a wide variety of problems. For a multiple output feed forward network, the mean square error E(t) objective function is expressed as,

$$E(t) = \frac{1}{Np} \sum_{i=1}^{N} \sum_{k=1}^{k=p} \left[\left(d_k^i\right)^2 - \left(y_k^i\right)^2 \right]$$

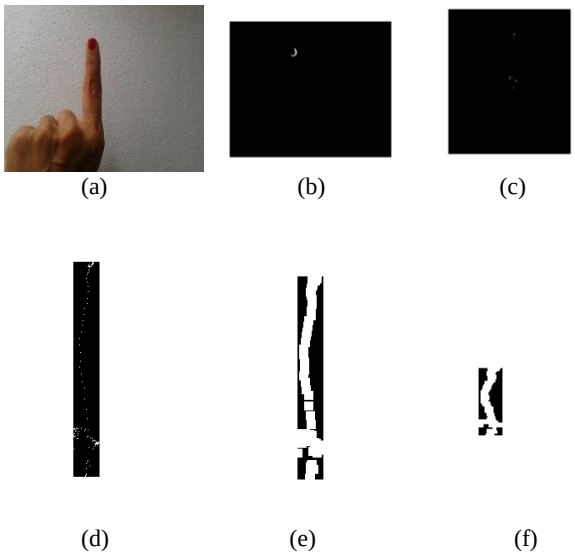
where y_k^i is the k[th] desired or target output activation of i[th] training sample. d_k^i is the k[th] actual output from the i[th] training pattern.

5. EXPERIMENTATION AND RESULTS

The initial stage of our experiment is the creation of a database. All the hand gesture video used in the experiments is collected manually, due to lack of general hand gesture sample database. Without the loss of generality, five hand gestures are selected for the evaluation of proposed system.

5.1. System using color marker

The resulted output obtained from the proposed PCA algorithm for a particular gesture no. 1 is shown in fig. 6.



(a) (b) (c)

(d) (e) (f)

Fig 6: Results of template creation for system using red marker. (a)first frame of the video. (b)Red spot detection(spatial segmentation). (c)Getting the locus of red spot in each frame. (d)Result after cropping. (e)Image after morphological operator. (f)Final template after resizing image to 42×24.

Table 1. gives the ED calculations for different symbol.

Table 1: ED calculations for different gestures.

Gesture Index	Distance to gesture no.			
	0	**1**	**2**	**3**
0	0.0008	3.1962	0.1999	0.1469
1	1.0201	0.6773	0.7654	0.9599
2	0.1106	2.5532	0.0555	0.1072
3	0.1491	3.1204	0.2102	0.0005

5.2. System without color marker

The performance of proposed algorithm was measured by collecting total 60 samples from 20 people by using the hand acquisition system. Out of these database, 40 samples serve as training set, and other serve as testing set. The algorithm was tested on a PC with AMD Turion dual core processor, 1Gb RAM. The program coding is done in Matlab 10. The videos are taken by Compaq web camera. The frame resolution of the video is 320×240.

Table 2: Result of classifier.

Method	Class	Traing No.	Test No.	Correct No.	Correct Rate(%)	Recognition Rate(%)
Prunning	0	40	20	18	90.0	84
	1	40	20	16	80.0	
	2	40	20	17	85.0	
	3	40	20	18	90.0	
	L	40	20	15	75.0	
ANN	0	40	20	18	90.0	86
	1	40	20	16	80.0	
	2	40	20	18	90.0	
	3	40	20	18	90.0	
	L	40	20	16	80	

6. CONCLUSIONS

This paper gives dynamic hand gesture recognition system which can be used for sign language recognition, interacting with virtual environment, or interacting with different applications like image browser, games, etc. Interaction becomes more natural and intuitive manner by using gesture based interfaces. Two systems and three types of classifier has been used for gesture recognition.

In first system, red marker is used with constant background which makes gesture segmentation easy. After RGB color based segmentation, appearance based modeling has been done. Appearance based modeling is used to extract the features. Simple parameter i.e. fingertip position is used which is easy to detect and reduces the complexity of algorithm. The ease of getting the templates after segmentation, the classification of these templates can be done using the pattern recognition algorithm i.e. PCA which uses the statistical properties. This method imposes few restrictions on the user and the performance of this classifier is not satisfactory. To overcome these disadvantages, second system is developed which is without color marker. HSV color space is used for skin color detection. The segmentation algorithm is tested for different videos to decide the threshold values for skin color in HSV color space. Pruning consists of the basic classifier model which is less robust than the ANN. The classifier often gets confused about the symbol '1' and 'L' due to which the accuracy levels for the detection of two have lowered. From the results and discussion, it is seen that ANN gives better classification for multiple classes with high accuracy.

The algorithms have been tested for only five symbols and static background. These algorithms can be improved and tested for other symbols including all digits and characters, provided that there should be the consideration of Subgesture reasoning algorithm. By improving segmentation process, the algorithms can be tested for more complex and dynamic background.

7. REFERENCES

[1] Vladimir I. Pavlovic, Rajeev Sharma, Thomas S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review", IEEE Trans. Pattern analysis and machine intelligence, vol. 19, pp. 522-536, July 1997.

[2] Ali Erol, George Bebis, Mircea Nicolescu, Richard D. Boyle, Xander Twombly, "Vision-based hand pose estimation: A review", Computer Vision and Image Understanding 108 (2007) 52–73.

[3] M.W. Krueger, Artificial Reality II, Addison-Wesley, 1991.

[4] Zhi-gang Xu, Hong-lei Zhu, "Vision-based Detection of Dynamic Gesture", IEEE International Conference on Test and Measurement, 2009.

[5] Hwei-Jen Lin, Shu-Yi Wang, Shwu-Huey Yen, and Yang-Ta Kao," Face Detection Based on Skin Color Segmentation and Neural Network", International conference on Neural networks and brain, 2005, vol. 2, pp. 1144-1149.

[6] J. Alon, V. Athitsos, Q. Yuan, S. Sclaroff, "A Unified Framework for gesture recognition and spatiotemporal gesture segmentation" Ieee Transactions On Pattern Analysis And Machine Intelligence, vol. 31, no. 9, september 2009, pp. 1685-1699.

[7] Mahmoud M. Zaki and Samir I. Shaheen, "Sign language recognition using a combination of new vision based features", Pattern Recognition Letters 32 (2011), pp. 572–577.

[8] Sushmita Mitra and Tinku Acharya, "Gesture Recognation : A Survey", IEEE trans. Systems, man, and cybernetics—part c: applications and reviews, vol. 37, may 2007.

[9] Psarrou, A., Gong, S., Walter, M.,. "Recognition of human gestures and behavior based on motion trajectories". Image Vision Comput. 2002, vol.20, 349–358.

[10] M. Masudur Rahman, Seiji Ishikawa, "Human motion recognition using an eigenspace", Pattern Recognition Letters 26 (2005) 687–697

[11] Rybach, D., 2006. Appearance-Based Features for Automatic Continuous Sign Language Recognition. Diploma Thesis, RWTH Aachen University, Aachen, Germany.

[12] Solomon Raju Kota, J.L.Raheja, Ashutosh Gupta, Archana Rathi, Shashikant Sharma, "Principal Component Analysis for Gesture Recognition using SystemC ", International Conference on Advances in Recent Technologies in Communication and Computing,2009,pp. 732-737.

[13] Md. R. Ahsan, Muhammad I. Ibrahimy, and Othman O. Khalifa "Hand motion detection from EMG signals by using ANN based classifier for Human Computer Interaction", Modeling, Simulation and Applied Optimization (ICMSAO), 2011 4th International Conference on 2011, pp. 1 – 6.