

Hand & Upper Body Based Hybrid Gesture Recognition

Prerna Sharma^{#1}, Naman Sharma^{*2}

[#] *Research Scholar, G. B. P. U. A. & T. Pantnagar, India*

^{*} *Ideal Institute of Technology, Ghaziabad, India*

Abstract— Communication in today's world is performed through the vocal sounds and body language. Vocal sounds are main tool for interaction, where body language and facial expressions also have important support. Even in few cases, interacting with physical world by using expressive movements like gestures and postures is much easier. In this paper, an approach is designed for Upper Body Pose Recognition and Hand Gesture Classification. This approach focuses on vision based technique. 6 classes of poses and 6 gesture classes are recognized during the testing process of this approach and neural network is used for classification.

Keywords— Singular Value Decomposition (SVD), Principal Component Analysis (PCA), Artificial Neural Network (ANN), Pose Recognition, Hand Gesture Recognition.

I. INTRODUCTION

Body language has broad range of activities like eye expressions, light change in skin color, variation of the vibrations in vocal sounds etc. But most important body language expressions are performed by using hands. Hand gestures would be ideal for conveying information in most cases like representing a number, expressing a feeling etc. Hand gestures are primary interaction tools for sign language and gesture based computer control. So, a system is designed in this area which works in two modules – Upper Body Pose Recognition and Hand Gesture Recognition. In upper body pose recognition, it determined the arm location in various poses and in case of hand gesture recognition; some specific types of gestures are recognized.

Gesture and Posture recognition can be used in various areas like 3D animation [12], Tele-presence [13], Virtual reality [14], sign language recognition [11] and many other areas. Applicability of proposed work is mainly in Human Computer Interaction. Singular value Decomposition (SVD) and Principal Component Analysis (PCA) are used for extracting the features and feed forward neural network is used for classification purposes.

A. SVD

It's a method used for data dimension reduction and feature extraction. SVD is based on a theorem from linear algebra which says that a rectangular matrix A can be broken down into the product of three matrices - an orthogonal matrix U , a diagonal matrix S , and the transpose of an orthogonal matrix V . [10] The theorem is usually presented like this:

$$A_{mn} = U_{nm} S_{nm} V_{nm}^T$$

where $U^T U = I$, $V^T V = I$; the columns of U are orthonormal eigenvectors of AA^T , the columns of V are orthonormal eigenvectors of $A^T A$, and S is a diagonal matrix containing the square roots of eigen values from U or V in descending order.

B. PCA

PCA belongs to linear transforms based on the statistical techniques. This method provides a powerful tool for data analysis and pattern recognition which is often used in signal and image processing as a technique for data compression, data dimension reduction or their de-correlation as well. [10]

The theorem is usually presented like this:

$$Y = PX$$

where X is a correlated variable set matrix and Y is uncorrelated variable set matrix that is calculated by multiplication of X and a co-variance matrix, P .

C. Feed-Forward Neural Network

“An ANN consists of a sequence of layers; each layer consists of a set of neurons. All neurons of every layer are linked by weighted connections to all neurons on the preceding and succeeding layers”. [9]

Each neuron in the network is able to receive input signals, to process them and to send an output signal. Each neuron is connected with at least one neuron, and each connection is evaluated by a real number, called the weight coefficient, that reflects the degree of importance of the given connection in the neural network.

There are two functions defined for the behavior of a neuron in a particular layer

1. Input function
2. Output/Activation function

Output function, net of a network layer i input x_i and weight w_i is defined as

$$net_i = \sum w_i x_i + bias$$

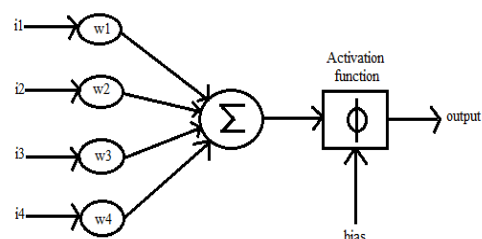


Fig. 1 Architecture of ANN

II. RELATED WORK

Gesture can be recognized using two approaches:

A. Vision Based Approach

In vision based methods the system requires camera(s) to capture the image required for the natural interaction between human and computers. Segmentation is done using skin or color segmentation. Although these approaches are simple but a lot of challenges are raised such as the complex background, lighting variation, and other skin color objects with the hand object.

B. Instrumented Glove Approach

Instrumented data glove approaches use sensor devices for capturing hand position, and motion. These approaches can easily provide exact coordinates of palm and finger's location and orientation, and hand configurations however these approaches require the user to be connected with the computer physically [1].



Fig. 2 Glove Based and Vision Based Approaches [1]

A research [2] is performed on hand gesture that used orientation histogram as a feature vector in neural network for recognizing static hand gesture of American Sign languages. Wah [3] have considered a vision based system for interpreting hand gestures in real time. They have used a segmentation procedure to extract binary hand blob from each frame. Fourier Descriptors were used to represent the shape of hand blob and gave this as an input to Radial Basis Function Network for pose classification. They have applied Hidden Markov Model and Recurrent Neural Network at last for recognizing the gestures.

Licsar [4] developed a vision based hand gesture recognition system which used Modified Fourier Descriptor for the classification of static hand gestures and Background Subtraction method for removing the background. A method [5] is described for pose recognition using Particle Swarm Optimization. They used a subdivision body model with an underlying skeleton layer to estimate the body poses. In the first step, they extracted the silhouettes from the video sequences and then, match them to the project model in a pose suggested by PSO.

A survey paper [6] is published on gesture recognition. They have discussed on many gestures like hand and arm gesture, head and face gesture, body gestures. They provided a survey on various techniques like Hidden Markov Model, Particle Filtering and Condensation Algorithm, Finite State Machine Approach and Artificial Neural Network on all types of gestures. Hu [7] proposed a

novel method in 2009 based on upper body pose recovery method via efficient joints detection. In this method, firstly face, skin and torso is observed, and then, recognize the proper joint locations. Finally, sample-based Markov Chain Monte Carlo is employed to determine the final pose.

Nguyen [8] in 2013 presented a method supporting hand gesture recognition in the static form, using Artificial Neural Network. They were taken the image and detected the hand region using skin color filters and background subtraction. Then, they applied median filtering for removing the noise and cropping the hand region. Finally recognize the gestures using a trained dataset.

III. SAMPLE LIBRARY

Here, some gesture training images are shown from the library:



Fig. 3 Hand Gesture Recognition

Here, some posture training images are shown from the library:



Fig. 4 Pose Recognition

IV. PROPOSED APPROACH

In this section, a novel method is given for Pose Recognition and Hand Gesture Recognition. Feed-forward neural network is used for training and classification. Proposed methodology can be explained using the block diagram for both modules: arm model and hand gesture recognition.

A. Arm Model

This block diagram explains the complete procedure for pose recognition:

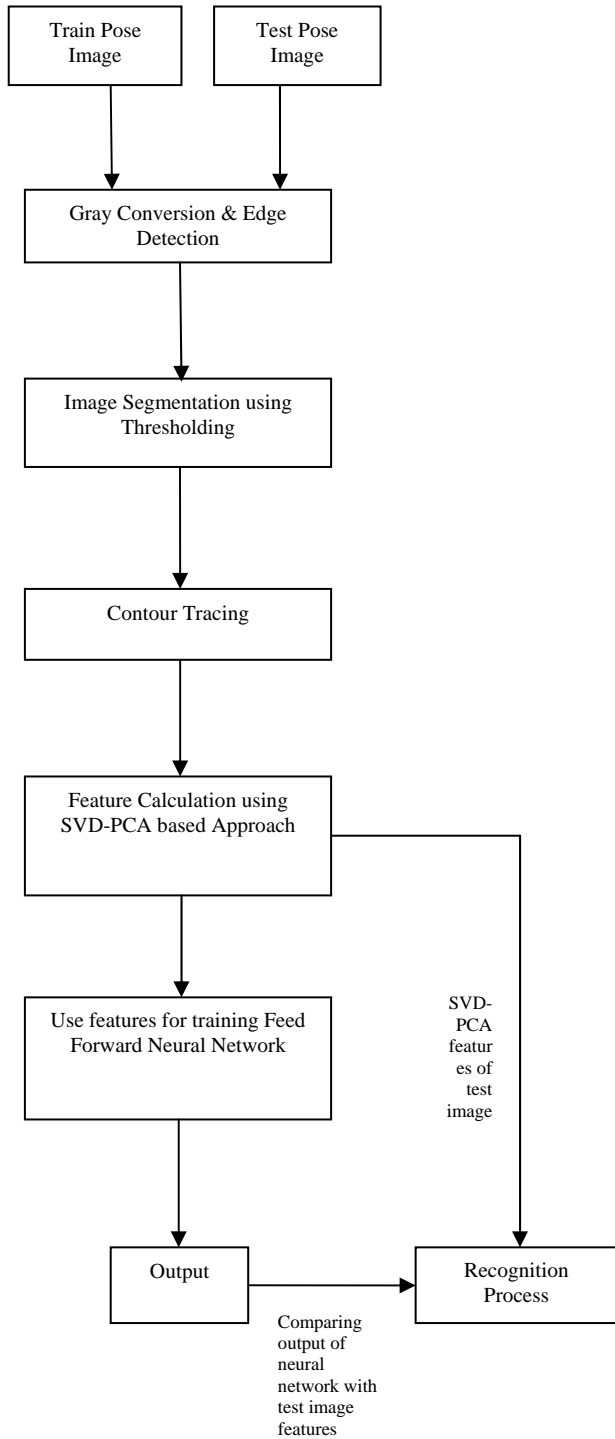


Fig. 5 Arm Model Block Diagram

B. Hand Gesture Recognition

This block diagram explains the complete algorithm for hand gesture recognition:

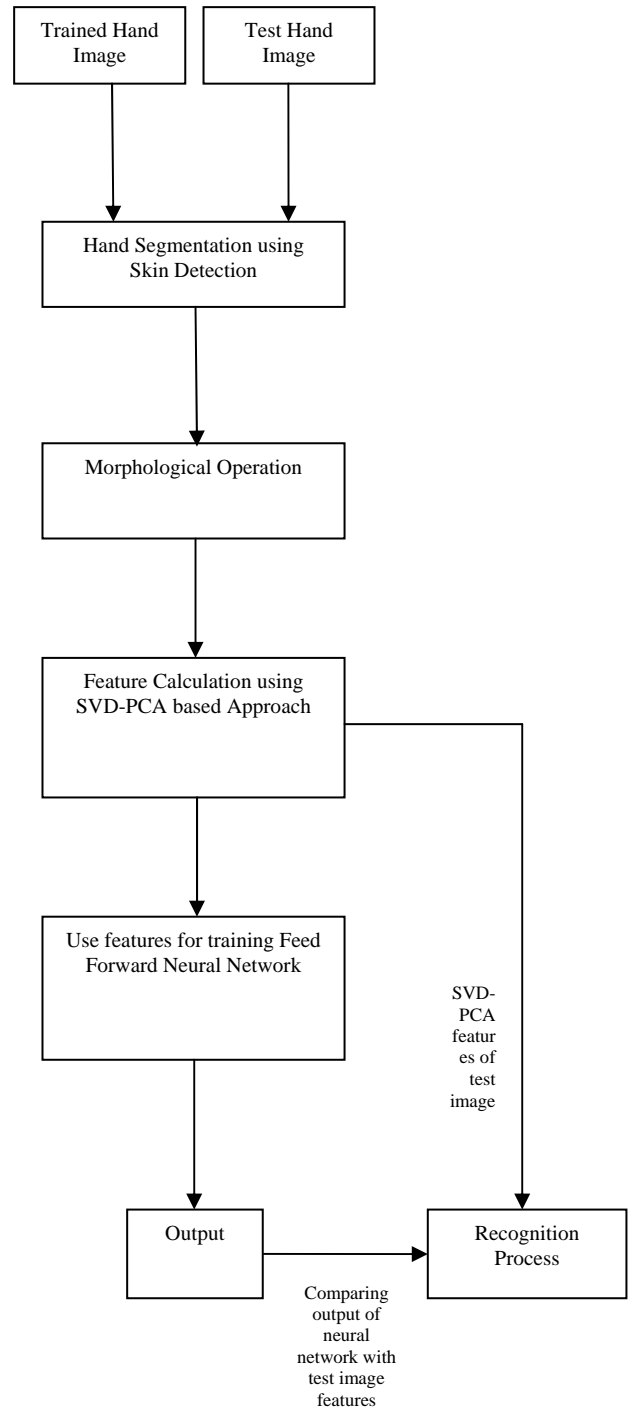


Fig. 6 Hand Gesture Recognition Block Diagram

C. Algorithm

The following steps will represent proposed method:

- 1) **Input Data & Training Data:** House dataset is created that contains an upper body image contained hand and arm portion clearly. Hand gesture and arm location are recognized in these images for recognizing the poses and gestures. And Sabastien Marcel Hand gesture dataset [15] is used for hand gesture recognition in comparison.



Fig. 7 Test case

- 2) **Pre-processing :**
Hand Gesture Recognition:
 Firstly, skin pixels are detected using RGB color model and segmented the hand portion. And perform some morphological operations for improving the results.
Pose Recognition:
 Firstly gray conversion is performed and then edges are detected using canny edge detection and contours are obtained.
- 3) **Extracting Features & Training:**
Hand Gesture Recognition:
 SVD-PCA features that contains Signal matrix, principal components matrix, diagonal matrix of eigen vectors of hand image are extracted and use them for training the network. Matrix of PC is shown in figure.

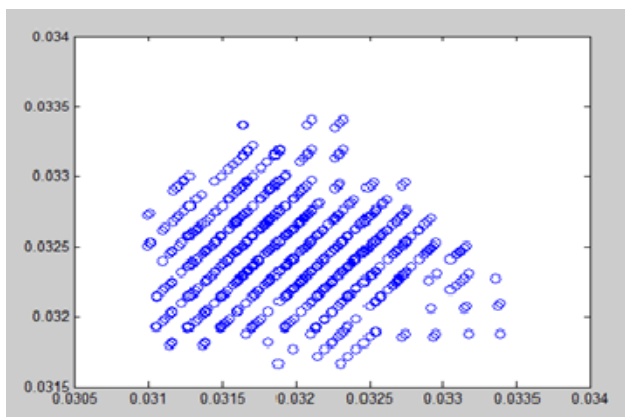


Fig. 8 PC matrix graph of given test case

Pose Recognition:
 SVD-PCA features of upper body image are extracted; these are also the same features discussed in hand part and use them for training the network.

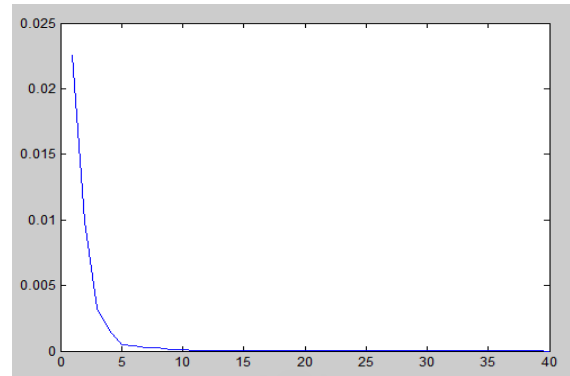


Fig. 9 Eigen Vector graph of given test case

This graph shows the variation level corresponding to number of eigen vectors. Eigen vectors are plotted on X-axis and variation level is plotted on Y-axis.

V. RESULTS AND COMPARISON

Confusion matrix and performance plot of this approach are obtained and shown. Confusion matrix shows the accuracy of system and corresponding error where other middle values in the matrix shows the intermediate iterations and their values. In case of performance plot, after number of epochs, system will get the optimum result i.e. trained performance meets the best performance. After that value of epoch, error becomes constant to minimum value.

Hand Gesture Recognition:

Output Class	1	2	3	4	5	6	
1	9 18.4%	1 2.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	90.0% 10.0%
2	0 0.0%	6 12.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	1 2.0%	11 22.4%	0 0.0%	0 0.0%	0 0.0%	91.7% 8.3%
4	0 0.0%	0 0.0%	0 0.0%	10 20.4%	0 0.0%	0 0.0%	100% 0.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	9 18.4%	0 0.0%	100% 0.0%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 4.1%	100% 0.0%
	100% 0.0%	75.0% 25.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	95.9% 4.1%
	1	2	3	4	5	6	

Fig. 10 Confusion Matrix

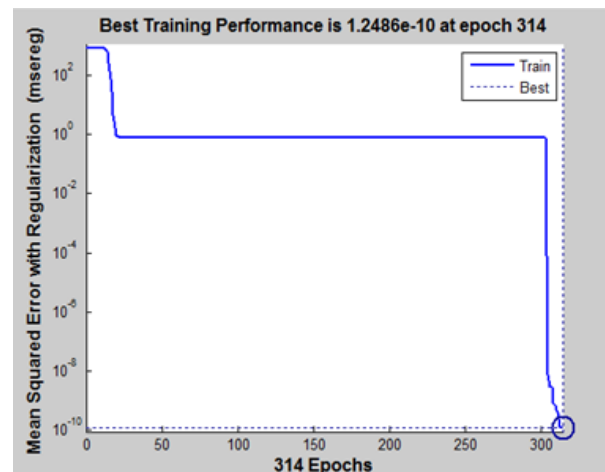


Fig. 11 Performance plot

Arm Model



Fig. 12 Confusion Matrix

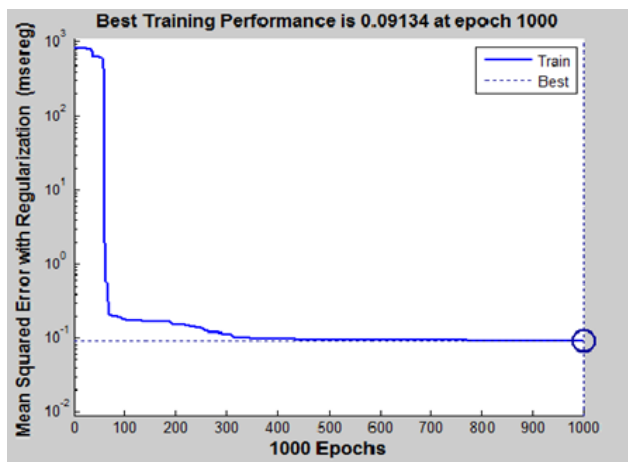


Fig. 13 Performance Plot

The accuracy of the approach is tabulated as

Table I
Accuracy Table

Dataset	No. of Classes	Accuracy
Gesture	6	95.9 %
Posture	6	90.3 %

Gesture and postures are detected after training the feed forward neural network and compared with the results of Marcel Hand Dataset.

Table II
Comparison Table

Dataset	Accuracy
Existing work on Marcel Dataset[15]	93.7 %
Proposed Work on Marcel Dataset	95.9 %

VI. CONCLUSION AND FUTURE SCOPE

Proposed work recognized specific poses and gestures in an upper body image. And accuracy is much better than existing approach. So, SVD-PCA features are better to use for training purpose of Feed-Forward neural network. It can be observed in previous work that hand gestures and poses are detected on different platforms but in proposed work, they are detected on single platform, so it can be a new endeavor to the complete body gesture recognition.

For future scope, proposed approach is work on uniform background, it can be extended for complex background and number of gestures and posture can be extended. This work tracks arm movements only in pose recognition. It can also be extended to complete upper body pose recognition which can be used in traffic control systems and other areas like 3D games, sign language recognition. Proposed work is performed on static images, for future aspects, it can be performed on dynamic images or videos which are more often useful in real time systems.

REFERENCES

- [1] N.A. Ibraheem, R.Z. Khan, 2012. Vision based gesture recognition using neural networks approaches: a review. *International Journal of human Computer Interaction (IJHCI)*. 3(1).
- [2] K. Symeonidis, 2000, Hand Gesture Recognition Using Neural Networks, MS Thesis. *University of Surrey, UK.*
- [3] C.W. Ng, S. Ranganath, 2002, Real-time gesture recognition system and application, *Image and Vision Computing*, Vol. 20, Issues 13-14, pp. 993-1007.
- [4] A. Licsar, T. Sziranyi, 2004, Dynamic training of hand gesture recognition system, *IEEE*, 0-7695-2128-2/04.
- [5] S. Ivekovic, and E. Trucco, 2006. Human body pose estimation with PSO. *IEEE Congress on Evolutionary Computation Sheraton Vancouver Wall Centre Hotel. Vancouver, BC, Canada.*
- [6] S. Mitra, T. Acharya, 2007, Gesture Recognition – A survey, *IEEE Transactions on systems, man, and cybernetics—Part C: Application and Reviews*, Vol. 37, No. 3.
- [7] Z. Hu, G. Wang, X. Lin, H. Yan, 2009. Recovery of upper body poses in static images based on joints detection. *Pattern Recognition Letters* (30). pp. 503–512.
- [8] T. N. Nguyen, H. H. Huynh, 2013, Static hand gesture recognition using artificial neural network, *Journal of Image and Graphics*, Volume 1, and No.1.
- [9] J. A. Anderson, 1997. An Introduction to Neural Network. 3rd Ed. Library of Congress Cataloging in publication Data. 651p.
- [10] K. Baker, 2005. Singular Value Decomposition Tutorial. 24p.
- [11] T. Brun, 1974. Teckensprks Lexikon. Bokforlaget Spektra AB, Halmstad.
- [12] T. Starmer, and A. Pentland, 1995. Real-time American sign language recognition from video using hidden markov models. *Technical Report No. 375*, M.I.T Media Laboratory Perceptual Computing Section.
- [13] J. Schlenzig, E. Hunter, and R. Jain, 1995. Recursive spatio-temporal analysis: Understanding Gestures. *Technical report, Visual Computing Laboratory*, University of San Diego, California.
- [14] H. Grant, C. K. Lai, 1998. Simulation modeling with artificial reality technology (smart): an integration of virtual reality simulation modeling. *Proceedings of the Winter Simulation Conference*.
- [15] S. Marcel, 1999, “Hand Posture Recognition in a Body Centered Space”, *IEEE*.