Vision Based Analysis of Hand Gesture Recognition for Human Computer Interaction (HCI)

Nidhi Sikri[1], Navleen Singh Rekhi[2]

[1]Research Scholar, DAVIET Jalandhar, Department of ECE

[2]Assistant Professor, DAVIET, Department of ECE

Abstract: *This paper presents a static hand gesture recognition based on function classification using Multi Class Support vector Machine (SVM). Gestures have long been studied as a communication technique that can possibly deliver more instinctive, a powerful tool supporting efficient and perceptive methods for interaction with our computers. The data acquisition was achieved using high resolution Logitech c270 web camera. The different hand gestures were mathematically processed individually using Discrete Cosine Transform (DCT) and singular Value Decomposition (SVD) and then classified using SVM. The accuracy rate of function classification for DCT and SVD were 94 % and 98%.*

Keywords: Hand gesture, DCT, SVD and Multi Class SVM

1. Introduction

Human-computer interface is the process of communication between the computer and the human being [1, 9]. The flow of information between the two is known as the loop of iterations. The loop of iterations can be done in different forms and in different environments:

1) Visual based: In this, the communication is done in the form of visuals i.e. images. This is the widespread are Human-Computer Interaction research.
2) Audio based: in this the different audio signals are taken to perform the communication process.
3) Task environment: the user has to set the goals and the limits to perform the communication with the help of different software.
4) Machine environment: the environment in the computer is connected to i.e. the laptop or the digital camera.
5) Input flow: the flow of information in the task environment.
6) Output flow: the flow of information from the task environment.
7) Feedback: the loops through the interface from the human to the computer and back to the human.

The methodology to obtain Hand gesture is classified as below:

a) Data Acquisition
The data acquisition of hand gestures [6] can be achieved either through portable camera which is interfaced to the computer or the inbuilt camera to the device which transmit the data via the networking devices or wireless to the main computer or through control computer:

b) Gesture Interpretation
To interpret the gestures for HCI, two methods are used commonly:

1) *Contact-based:* In this method, the mechanical or optical sensors are attached to the gloves that convert the flexions into electrical signals to determine the gesture.
2) *Vision-based***:** This method is easier to use as no hardware is required in this method.

c) Feature Extraction of Hand Gestures
In image processing, initial set of data is measured and features are derived which are supposed to be informative and error free [4], used for continuous learning which leads to better human interpretations. Various features can be classified based on:

General features: color, texture and shape i.e. application independent features. They can be further divided into: Pixel level features, Local features and Global features: entire image or sub area of an image is used for features calculation. Also, Domain specific features: human faces, finger prints, etc are some feature dependent features can be specified as domain specific features.

d) Function Classification
Image classification is done with the help of many mathematical tools. Some of them are described below:

Hidden Markov Model [5]: It is a statistical model with hidden states, but the output is dependent on the visible states. The word 'hidden' refers to the state sequence through which the model passes, not the parameters. The temporal characteristics are provided by the tracker, which is used characterize the gestures.

Neural networks [9]**:** It is based on large collection of artificial neurons. Each neural unit is attached to others and links enhance the activation state of next neural units. Each neural unit computes. Each individual neural unit computes using summation function. These systems are self learning and trained, and excel in solution of feature classification.
Fuzzy C means Algorithm [2]: It is a clustering technique in which data set is grouped into clusters which is belonging to every cluster to a certain degree. It starts with an initial guess to mark the mean location of each cluster which is most likely incorrect. The iteration process starts with each step moving towards the right location of cluster center with data set. This iteration is based on minimizing the distance from given point to center weighted by that data point grade.

Support Vector Machine: It is a classifier which is defined by the hyper plane. The working of the SVM algorithm is based on the largest minimum distance given to the samples by the hyper plane. Given a set of samples, an SVM training algorithm makes a model that gives new samples to one categories or the other.[3,7]
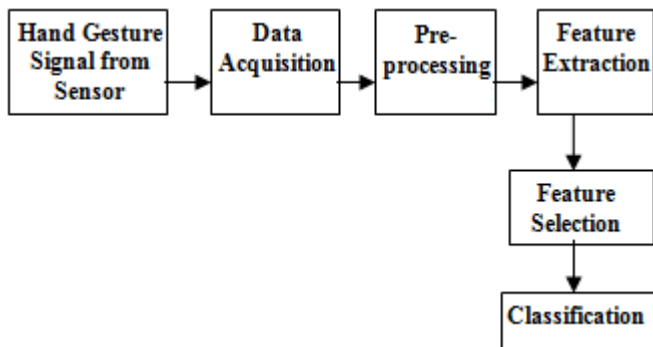

Figure 1: Methodology for recognition of Vision based Hand Gesture

2. Related Work

Every new experiment is specifically dedicated to increase the intelligence level of computer and help in making complex communication between the human with the computer. The result oriented efforts done by the researchers made it possible to create the human-computer interfaces. Shital M. et al[2015] had done the hand gesture movement tracking in JAVA using the MATLAB Image processing tools. The result showed that the mouse movements were replaced with the tracking of hand and the cursor movement and click functions was done with hands. Prerna Sharma et al[2015] used the hand and the upper body with the feed forward Neural Network for training and classification of static images. SVD-PCA features of upper body and hand part features were extracted and then used for the training purpose. The results showed that the proposed work had given the accuracy rate of 95.9%. Rajat Agarwal, et al[2015] used 2D Discrete Wavelet Transformation Decomposition on the images to have 7th level as feature vector, which was used in SVM for training and testing purpose. The system was able to tolerate 0.5 intensity of salt and pepper noise. The data set was of static gestures where size, orientation, illuminations variations was taken. The accuracy of the system was calculated with 94% for about 350 samples of 7 hand gestures.

Samruddhi Kahu et al [2013] had used the SVD for image compression technique instead of DCT for the images having higher pixel quality. It was observed that the compressed images was nearby original image. As compared to other compression techniques SVD gave good compression results with less computational complexity. Le Hoang Thai et al [2012] used Artificial Neural Network and Support Vector Machine for image classification for Roman Numerals. The sub-images were separated for the images on the basis of features and the used by ANN for the classification process, further compiled the results using SVM. The precision recognition was done by drawing Roman Numerals on drawing canvas and the result was shown in classification canvas. The precision rate for the recognition was 86%.Malvika Bansal et al [2011] recognized dynamic hand

gesture in real time using HMM by using Computer vision. The database size was reduced by standardizing the axis through centroid. The model parameters were used by Forward and Backward Algorithm, by the Viterbi Algorithm, Posterior Decoding, and Baum-Welch Algorithm. The result was expected to be high.

The aim of the proposed work to recognize the five hand gestures of whap-whap, throw, get, stop and go position. The dimensionality reduction was attained and verified individually for hand gesture using SVD and DCT. The functional classification was done using Multi Class Support Vector Machine supported by LIBSVM toolbox to find the better solution for the accuracy rate. The work is carried out in three phases (Figure 1) viz..(1)Data Acquisition of Vision Based Hand Gesture. (2)Feature Extraction using Discrete Cosine Transformation (DCT) and Singular Value Decomposition and function classification using Multi Class Support Vector Machine based on one versus rest.

3. Work Done

1) Data Acquisition
The data acquisition involves collection of the samples of vision based hand gestures. To detect the hand gestures movement of the user, the sensor is required. Sensor in the form of camera is used. The Logitech HD C270 camera captures the images at a fixed rate and resolution of 3 mega pixels is used. The images were acquired by the webcam in real time without projecting the arm and were saved in .jpeg format. In order to initialize the system user have to wave his/her hand in front of webcam. The image acquisition toolbox of MATLAB acquires the output of the webcam in terms of frames. The basic block diagram for the data acquisition is shown in fig. **2.**
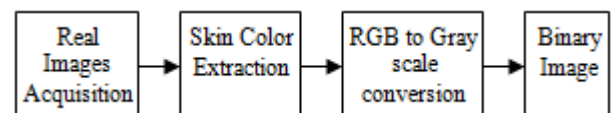

Figure 2: Basic Block Diagram of Data Acquisition

Figure 3: Different Static Hand Gesture Positions Stop, Whap-whap, Throw, Get and Go

For the purpose of image acquisition, required training for the desired motions of the hand was given to each subject individually. The hand gestures of different positions like go, get, throw, stop and whap-whap position were taken. For recording of the data, five people of different age groups between 15-20 years, 20-30 years, 30-40 years and 40-50 years were taken as subjects.

2) Dimensionality Reduction
i. Discrete Cosine Transform (DCT)
DCT is a technique for converting a signal into elementary frequency components. The still images may have the regions which may contain same value of pixels in their

neighborhood. To minimize such redundancy, JPEG based DCT compression was implied [15]. The results for each hand gesture are as shown in Figure 4. Each image is broken in to 8x8 blocks. Then each block is compressed through quantization.

The DCT is defined as follows:

Table 1: PSNR for different Hand Gesture

Hand Gesture	PSNR
Go	233.9625
Whap	235.7926
Stop	232.6110
Throw	233.1556
Get	233.7904

$$P_{u,v} = \frac{C_u \times C_v}{4} \sum_{x=0}^{n} \sum_{y=0}^{n} p_{x,y} D_{x,u} D_{x,v} \qquad ....(1)$$

Where $P_{u,v}$ is the coefficients, $p_{x,y}$ is the intensity of picture elements and $C_{u,v}$ are the constants. The reconstructed image is calculated as follows:

$$p_{x,y} = \frac{C_u \times C_v}{4} \sum_{u=0}^{n} \sum_{v=0}^{n} P_{u,v} D_{x,u} D_{x,v} \qquad ....(2)$$

To measure the quality of reconstructed image, PSNR(Peak Signal to Noise Ratio) was calculated for each hand gesture. Higher the value of PSNR, more accurate is the reconstructed image. PSNR is easily defined via the mean squared error (*MSE*). For a given $m \times n$ image and its noisy approximation *MSE* is defined as:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2 \qquad .....(3)$$

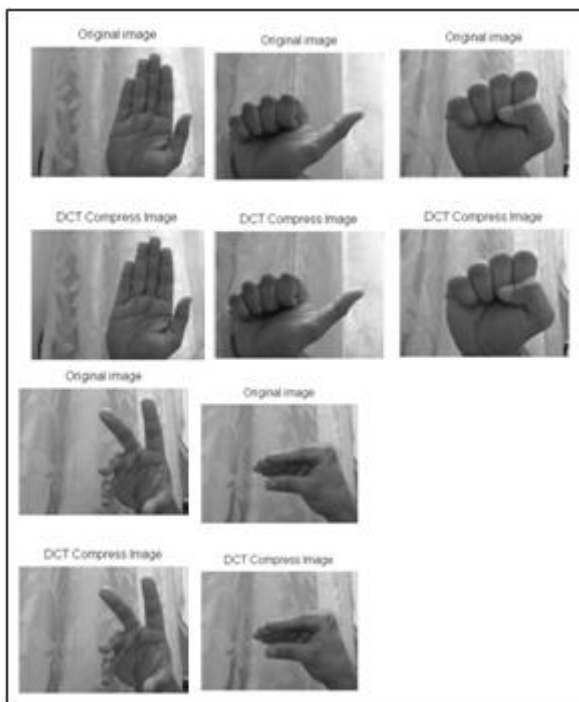$$PSNR = 20 \log_{10} \left(\frac{MAX^2}{MSE} \right) \qquad ......(4)$$

Figure4: DCT Compression for the respective Hand gesture

3) Singular Value Decomposition

For each image, the singular value decomposition was implemented. Singular value decomposition is a linear algebra technique that decomposes matrices into constituent parts, a left-hand and right-hand matrix separated by a descriptive diagonal matrix, the singular values, as their weighting[11]. It takes a rectangular matrix of data in which the *n* rows represents the text of SEMG data, and the *p* columns represents the wavelet coefficients.

The SVD theorem states:

$$X_{nxp} = U_{nxn} S_{nxp} V^T_{pxp} \qquad .... (5)$$

Calculating the SVD consists of finding the Eigen values and eigenvectors of XX^T and X^TX. The eigenvectors of X^TX make up the columns of *V*, the eigenvectors of XX^T make up the columns of *U*. Also, the singular values in **S** are square roots of Eigen values from XX^T or X^TX. The singular values are the diagonal entries of the *S* matrix and are arranged in descending order. The singular values are always real numbers. If the matrix *X* is a real matrix, then *U* and *V* are also real.
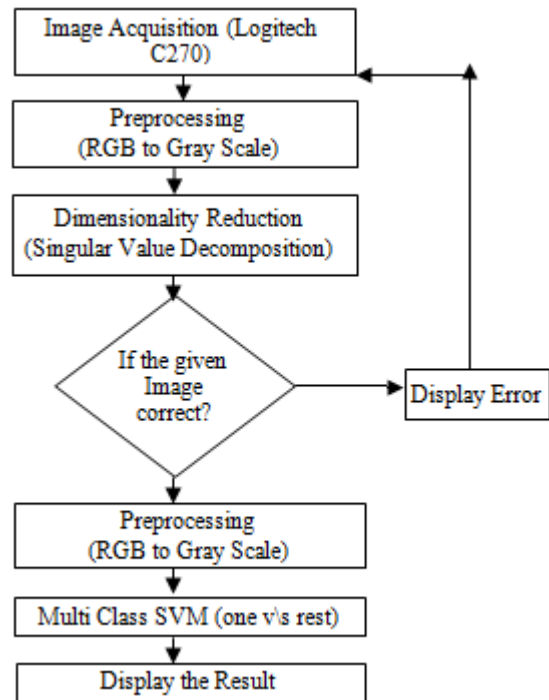

Figure 5: Flowchart for SVD Technique

4) Function Classification

The proposed method has used the LIBSVM for the classifying of the feature vector generated from the static hand gesture into the respective classes. SVM have proved to be a remarkably robust classification method across a wide variety of applications[14]. Here, a multi-class one versus rest classification problem is considered. Essentially, the SVM attempts to find a hyper-plane of maximum "thickness" or *margin* that separates the data points of the two classes. This hyper-plane then forms the decision boundary for classifying new data points. Let w be the normal to the chosen hyper-plane. Then, the classifier will label a data point x as +1 or −1, based on whether w · x + b is greater than 1, or less than -1. Here, b is chosen to maximize the margin of the decision boundary while still

classifying the data points correctly. This leads to the following learning algorithm for linear SVM. For the classifier to correctly classify the training data points $x_1, x_2, ..., x_n$ with labels $y_1, y_2, ..., y_n$ drawn from drawn from ±1, the following constraints must be satisfied:

$$w.x_i + b \geq 1, \quad \text{if} \quad y_i = 1 \ ......(6)$$

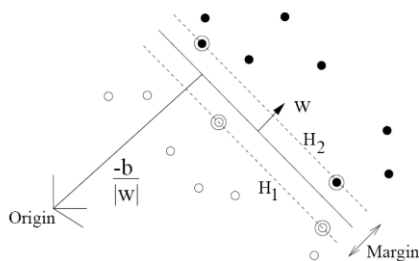$$w.x_i + b \leq -1 \quad \text{if} \quad y_i = -1 \ ......(7)$$

Figure 4.1: The optimal separating hyper-plane is the solid line. Support vectors (highlighted with an extra circle) are the points which lie on hyper-planes (the dashed lines) that have unit (functional) distance to the optimal separating hyper-plane.

The following tabulated results shows the classification of hand gestures for DCT and SVD methods. The recognition is tested with real time static gesture and is verified from the dataset of 10 images for each hand gesture.

Table 2: Results From The Multi-Class SVM (one vs rest) Radial Basis Function Kernel, Method:- SVD

S. No.	Gesture	Output For tested Data For Each Gesture					Accuracy (%age)
1	Stop	1	1	1	1	1	100
2	Get	2	2	2	2	2	100
3	Throw	3	3	3	4	3	92
4	Whap	4	4	4	4	4	100
5	Go	5	5	5	5	5	100
	Overall						98

Table 3: Results From The Multi-Class SVM (one vs rest) Radial Basis Function Kernel, Method:- DCT

S. No.	Hand Gesture	Output For tested Data For Each Gesture					Accuracy (%age)
1	Get	1	1	1	1	1	100
2	Stop	2	2	2	2	2	100
3	Throw	3	3	4	3	3	87
4	Whap	4	4	4	4	3	90
5	Go	5	5	5	5	5	100
	Overall						94

4. Conclusion

The work presented is the real time recognition of hand gesture for developing the Human Computer Interaction. The technique used for function classification is one versus rest Multi-class SVM from LIBSVM toolbox. The results from the techniques are further compared which shows that the classification success rate of almost 98% percent is achieved from SVD and approximately 94% from DCT method.

It can be concluded that the SVM is definitely going to have a dramatic effect in the research field of hand gesture recognition. The recognition was implemented to five groups of classes viz. go, get, whap, throw and stop. In future, more number of symbolic images can be inculcated to further enhance the usage for vast HCI. Furthermore, different function classification tools can be investigated to increase the efficiency of the recognition method.

References

[1] Andrew Wilson and Aaron Bobick, "Learning visual behavior for gesture analysis," In Proceedings of the IEEE Symposium on Computer Vision, Coral Gables, Florida, pp. 19-21, November 1995.

[2] E. Holden, R. Owens and G. Roy, "Hand Movement Classification Using Adaptive Fuzzy Expert System", Journal of Expert Systems, Vol. 9(4), pp. 465-480, 1996.

[3] Lalit Gupta and Suwei Ma "Gesture-Based Interaction and Communication: Automated Classification of Hand Gesture Contours" , IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, vol. 31, no. 1, February 2001

[4] Ryszard S. Chora´s, "Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems" ,International Journal Of Biology and Biomedical Science, Issue 1, Vol 1, 2007

[5] Ramage D (2007) Hidden Markov models fundamentals, Lecture Notes. http://cs229.stanford.edu/section/cs229-hmm.pdf

[6] Zhengmao Zou, Prashan Premaratne, Ravi Monaragala, Nalin Bandara, Malin Premaratne, "Dynamic Hand Gesture Recognition System using Moment Invariants" Information and Automation for Sustainability(ICIAFs), 5th International Conference on Dec 2010

[7] Liu Yucheng and Liu Yubin, "Incremental Learning Method of Least Squares Support Vector Machine", International Conference on Intelligent Computation Technology and Automation" VCL-94-104, 2010

[8] Malvika Bansal, Shivin Saxena, Devendra Desale ,Dnyaneshwar Jadhav," Dynamic Gesture Recognition Using Hidden Markov Model in Static Background" IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 6, No 1, November 2011 , ISSN (Online): 1694-0814

[9] S. S. Rautaray, A. Agrawal, "Vision Based Hand Gesture Recognition for Human Computer Interaction: A Survey", *Springer*, 06 November 2012.

[10] Le Hoang Thai, Tran Son Hai, Nguyen Thanh Thuy," Image Classification using Support Vector Machine and Artificial Neural Network", I.J. Information Technology and Computer Science, 2012, 5, 32-38

[11] Samruddhi Kahu, Ms. Reena Rahate, Image Compression using Singular Value Decomposition" International Journal of Advancements in Research & Technology, Volume 2, Issue 8, August-2013

[12] Shital M. Chavan, Smitha Raveendran," Hand Gesture Movement Tracking System for Human Computer Interaction" International Research Journal of Engineering and Technology (IRJET) Volume: 02 Issue: 08 Nov-2015.

[13] Prerna Sharma, Naman Sharma," Hand & Upper Body Based Hybrid Gesture Recognition "International

Journal of Computer Science and Information Technologies, Vol. 6 (2) , 2015, 1258-1262 (IJCSIT).

[14] Rajat Agarwal, Balasubramanian Raman, Ankush Mittal," Hand Gesture Recognition using Discrete WaveletTransform and Support Vector Machine", 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)

[15] Kavitha Jaganathan Dr. Lili Nur Liyana, Dr. Razali Yakob Dr. Lili Nur Liyana, Dr. Razali Yakob Dr. M. Jaganathan," Single Hand Gesture Recognition Based On Dwt And Dct Feature Extraction And Neuro-Fuzzy Classifier" E-Proceeding of the International Conference on Social Science Research, ICSSR 2015 ,8 & 9 June 2015, Meliá Hotel Kuala Lumpur, Malaysia.

Author Profile

Nidhi Sikri, research scholar is perusing M TECH degree from DAV Institute of Engineering and Technology in Electronics and Communication with specialization in Signal Processing. She received the B TECH degree in Electronics and Communication Engineering from Rayat Institute of Engineering and Information Technology in 2008.

Navleen Singh Rekhi was born in Jalandhar, Punjab India. He had received his B.Tech from Dr. Baba Saheb Ambedkar Marathwada University, Maharashtra. He had pursued hi M.Tech from SLIET Longowal, Punjab and currently pursuing PhD from IKG PTU Jalandhar..He had a teaching experience for more than 10 years. He is currently posted as Assistant Professor in Department of ECE, DAVIET, Jalandhar, Punjab. He has more than 20 publications in reputed International and National journals in his expertise field of Digital Signal & Image Processing for Human Computer Interaction.