A Gesture Tool for Hand Gesture Detection and Recognition

Bhagyalakshmi Joshi, Dr. Manali Godse

M.E.(EXTC), D. J. Sanghvi College of Engineering, Vile Parle, Mumbai-68
HOD, Biomedical Department, D. J. Sanghvi College of Engineering, Vile Parle, Mumbai-68
bhagyalakshmi.joshi@gmail.com, manali.godse@djsce.ac.in

Abstract: A real time Hand Gesture Detection and Recognition system has been developed using Human Computer Interaction (HCI). This system will be useful for developing a communication channel between the deaf and dumb and the rest of the world. The system operates in two stages; the first stage is the “Training” of the system, and the second stage is the “Recognition”. The Training stage includes training 100 samples of each gesture and storing those samples on to the system, for using them in real time testing stage. The Training stage uses Viola Jones algorithm and the Recognition stage uses Principal Component Analysis (PCA). Hand and Gesture Detection is carried out using Haar Cascade Classifier algorithm, developed by Viola and Jones, which was originally developed for the detection of the Human Face. For the training stage, a set of positive images of desired hand gesture and a set of negative images are used to obtain the required Haar-like features to detect the hand and gesture. During the Recognition stage, PCA uses Eigen Vector from the Co-Variance Matrix of the image. This Eigen Vector termed as Weighted Vector is used to obtain the Euclidean Distance. This distance is compared with the Threshold Distance, to identify the correct gesture. Mahalanobis Distance is used to calculate the accuracy of each gesture. This entire process is designed in C++ with OpenCV library. RobotBASIC is used as GUI for the representation of the system.

Keywords: Human Computer Interaction (HCI), Viola Jones algorithm, Principal Component Analysis (PCA), Haar Cascade Classifier algorithm, Co-Variance Matrix, Threshold Distance, Mahalanobis Distance, RobotBASIC.

1. INTRODUCTION

Communication is the medium through which one can convey information and develop understanding among each other, which helps build transfer of information, analysis and development. With hearing impairment and muteness, deaf and dumb people require non-verbal communication mode. Non-verbal communication plays a vital role in such people to convey and emphasis the information, which is done by symbolizing with gestures, which acts as a replacement to speech communication.

To convey information using body language, hand gestures plays a vital role in conveying the information. The human hand has 27 Degrees Of Freedom (DOF): 4 in each finger, 3 for extension and flexion and one for abduction and adduction; the thumb is more complicated and has 5 DOF, leaving 6 DOF for the rotation and translation of the wrist[1]. Hand gestures fall into two categories, namely static and dynamic.

2. LITERATURE SURVEY

The use of hardware is eliminated by using the vision based Human Computer Interaction through Real time Hand Detection and Gesture Recognition. For this purpose, a webcam is used to track user’s hand and recognize the gestures to initialize specific interactions [10]. The technology is based on the assumption that all hand movements are properly coordinated. Many hand detection and gesture recognition systems have been proposed. Earlier systems used markers or colored gloves for recognition. Later, methods for skin detection, shape etc came up for detecting the hand region [2,3]. But these systems are not robust enough to deal with dynamic environments, since they require uniform background, illumination. This can be overcome using boosted classifiers, which allow robust and fast detection of hands [4,5,6].

Boosted classifiers has been used in both hand and gesture detection. This system is based on Viola & Jones’ cascade of boosted classifiers, which is applied on static images [7].

An extension to this system [8] on real time, where boosted classifiers are employed for hand detection. However, this system [9] is very time consuming (since it is using a tree classifier rather than a single cascade), and is not applicable for interactive applications.
3. ALGORITHMS

3.1 HAAR CASCADE CLASSIFIER

The Haar transform is one of the simplest and basic transformations from the space/time domain to a local frequency domain, which reveals the space/time-variant spectrum. The attracting features of the Haar transform, including fast for implementation and able to analyze the local feature, make it a potential candidate in modern electrical and computer engineering applications, such as signal and image compression.

The family of $N$ Haar functions $h_k(t)$ are defined on the interval $0 \leq t \leq 10[26]$. The shape of the Haar function, of an index $k$, is determined by two parameters: $p$ and $q$, where

$$k = 2^p + q - 1$$

and $k$ is in a range of $k = 0, 1, 2, \ldots, N - 1$.

When $k = 0$, the Haar function is defined as a constant $h_0(t) = 1/\sqrt{N}$; when $k > 0$, the Haar function is defined as

$$h_k(t) = \begin{cases} 
\frac{1}{\sqrt{N}} & (q - 1)/2^p \leq t < (q - 0.5)/2^p \\
-\frac{1}{\sqrt{N}} & (q - 0.5)/2^p \leq t < q/2^p \\
0 & \text{otherwise} 
\end{cases}$$

From the above equation, one can see that $p$ determines the amplitude and width of the non-zero part of the function, while $q$ determines the position of the non-zero part of the Haar function $0$.

The Haar transform $HT^n(f)$ of an $N$-input function $X^n(f)$ is the $2^n$ element vector

$$HT^n(f) = H^nX^n(f)$$

The Haar transform cross multiplies a function with Haar matrix that contains Haar functions with different width at different location.

Viola Jones algorithm introduced for the first time face detection system, where three ingredients working in concert to enable a fast and accurate detection:

- The integral image for feature computation,
- Adaboost for feature selection and
- An additional cascade for efficient computational resource allocation.

3.2 PRINCIPAL COMPONENT ANALYSIS

The main purposes of a principal component analysis are the analysis of data to identify patterns and finding patterns to reduce the dimensions of the dataset with minimal loss of information. Here, our desired outcome of the principal component analysis is to project a feature space (our dataset consisting of $n \times d$-dimensional samples) onto a smaller subspace that represents our data "well". A possible application would be a pattern classification task, where we want to reduce the computational costs and the error of parameter estimation by reducing the number of dimensions of our feature space by extracting a subspace that describes our data "best".
4. FRAMEWORK

Our system enables hand gesture detection and recognition of user’s hand fingers. Fig. 1 shows the algorithm used in our approach to detect and recognize the gestures created on screen using the Haar Cascade Classifier and PCA.

![Algorithm for Hand gesture Detection and Recognition](image)

**Figure 1.** Algorithm for Hand gesture Detection and Recognition

![Haar Cascade Classifier Algorithm for Hand gesture Detection](image)

**Figure 2.** Haar Cascade Classifier Algorithm for Hand gesture Detection

A webcam captures the real time image and sends the captured information for processing. The captured real time RBG image (aspect ratio is 4:3) is converted into Grayscale image. This image is resized to a generalized dimension 120*90, irrespective of change in the aspect ratio during processing. This creates a new image of the detected region.

Next, a reference region is created i.e. region of interest. The grayscale image being processed is a 32-bit floating image which is converted, using DFT (Discrete Fourier Transform) to obtain an 8-bit pixel data. This grayscale floating 8-bit image is then stored in BMP format.

The Haar Cascade Classifier marks the hand region of the image captured. Haar-like features focus on the information within a certain area of the image rather than each single pixel. In order to have better classification accuracy and achieve the real-time performance, the AdaBoost (Adaptive Boost) help select the best features in each step and combine them into a stronger classifier. AdaBoost learning algorithm takes a set of “positive” samples, which contains the object of interest (in our case hand postures) and “negative” samples i.e. image which does not contain object of interest. At the training stage, distinctive Haar-like features are selected to classify the images containing the object of interest at each stage. This hand detection region returns a rectangle for the detected region in the given image. With this algorithm, the system is trained and an array of each gesture is created, which is stored in a database.
During testing, the image recognition is carried out using PCA. Here, the test image is projected upon the PCA subspace. PCA reduces the dimension of the image set i.e. it helps obtain only the basic components, stripping away all the unnecessary parts. It obtains only the hand region and deletes the background. To perform PCA on an image, we require the eigen vector and eigen value of the image. The eigen vector and eigen values exists in pairs i.e. every eigen vector has a corresponding eigen value. Eigen value is a number which helps obtain the variance, determining the hand gesture data. Hence, the eigen vector of the image obtained is stored in the form of image. This processed image is compared with the stored images in the database to obtain the correct match of gesture. For obtaining better accuracy and most likely gesture, Mahalanobis Distance is used to obtain the nearest similarity image.

The Mahalanobis Distance is used to measure dissimilarity between two vectors. These two vectors would be the values obtained from the co-variance value obtained from the array of images and from the test image.

Consider vector y, be the test image vector and vector x₁ be the vector value of trained array of images. The distance between the two can be obtained by,

\[ d (x, y) = \] 

where, \( S \) is the co-variance matrix.

The Mahalanobis distance accounts for the variance of each variable and the covariance between variables. Geometrically, it is done by transforming the data into standardized uncorrelated data and computing the ordinary Euclidean distance for the transformed data. In this way, the Mahalanobis distance provides a way to measure distances that takes into account the scale of the data.

**Figure 3. Principal Component Analysis (PCA) Algorithm for Hand gesture Recognition**
5. OUR SYSTEM COMPONENTS

The components of Hand Gesture Recognition used are Web Camera, C++ working on Eclipse, and Robot Basic for GUI.

° Webcam:
The webcam captures real time videos or images. It tracks the movements of the thumbs and the index fingers of the user’s hands and is processed using algorithm developed for hand and gesture detection and its recognition.

° Open CV:
Open CV is an Image Processing library, which is created by Intel and maintained by Willow Garage. This library is an open source which is easy to use and install. We are using 2.4.3 version.

° C++:
A high-level programming language developed by Bjarne Stroustrup at Bell Labs, is one of the most popular programming language for graphical applications, which includes applications that run in Windows and Macintosh environments.

° Robot Basic:
The Robot Simulator extension of RobotBasic has incorporated a set of functions and commands into a powerful interpreted programming language which is versatile and easy to learn. This is used for GUI.

6. RESULTS

The hand detection and recognition has been simulated by using the C++ programming language to process the image from live capturing from camera to the entire processing of image. It uses additional OpenCV library to process the same.

Fig No. 6 represents GUI window in RobotBASIC. There are two options for the user to select between “Gesture to Text” and “Text to Gesture”.

![Hand Gesture Recognition GUI](image)

Figure 4. GUI Window to select “Gesture to Text” and “Text to Gesture”
On selecting the “Gesture to Text”, user gets option to choose “Alphabets”, it starts the process of detecting and recognizing the hand gestures in real time by Human Computer Interaction (HCI).

The hand region is first detected using the Haar Cascade xml file stored. If a new set of gestures, the user will have to enter the name of the new gesture which we need to store. On completing the storing of the images of the new gesture in real time, we can either begin again to store new gestures or test the stored gestures.

The detected hand region is first converted to Grayscale to find the Region Of Interest (ROI) (Fig3). Haar Cascade Classifier helps detect only the hand region, and no other objects surrounded by it. (Fig4).

The hand region detected is further resized, so that it gives the image a standard brightness and contrast, in case if it is too dark or low in contrast (Fig5). The output image obtained from the resized image is processed using histogram equalization technique, which equalizes the image (Fig6). The detection of hand along with the formation of rectangular box defines the correct hand region. Each gesture detected and recognized in the process also calculates the accuracy of the gestures, which is displayed in an output window along with the name of the gesture stored with it (Fig7).

With the above explained process, a new set gestures expressing all 26 alphabets. These alphabets are prepared using reference from the British Sign Language (BSL) and American Sign Language (ASL).
If the user chooses the “Text to Gesture”, the stored gesture will be displayed, when the user enters the alphabet it wants to view. Following is the GUI Window showing the Text to Gesture, currently displaying the Alphabet “A”.

The accuracy of the Gestures created is calculated using Mahalanobis Distance. The accuracy for alphabet “A” is given in both in tabular as well as in graphical form, shown below.
REFERENCES