
Android Based Portable Hand Sign Recognition System

Jagdish L. Raheja, A. Singhal, Sadab and Ankit Chaudhary

These days mobile devices like phones or tablets are very common among people of all age. They are connected with network and provide seamless communications through internet or cellular services. These devices can be a big help for the people who are not able to communicate properly and even in emergency conditions. For a disabled person who is not able to speak or a person who speaks a different language, these devices can be a boon as understanding, translating and speaking systems for these people. This chapter discusses a portable android based hand sign recognition system which can be used by disabled people. This chapter shows a part of an ongoing project. Computer Vision based techniques were used for image analysis and PCA was used after image tokenizer for recognition. This method was tested with webcam results to make the system more robust.

Jagdish L. Raheja, A. Singhal, Sadab
Machine Vision Lab, CSIR-CEERI
Pilani, India
e-mail: jagdish@ceeri.res.in, {asinghal, sadab1933}@gmail.com

Ankit Chaudhary
Dept. of Computer Science, Truman State University
Missouri, USA
e-mail: dr.ankit@ieee.org

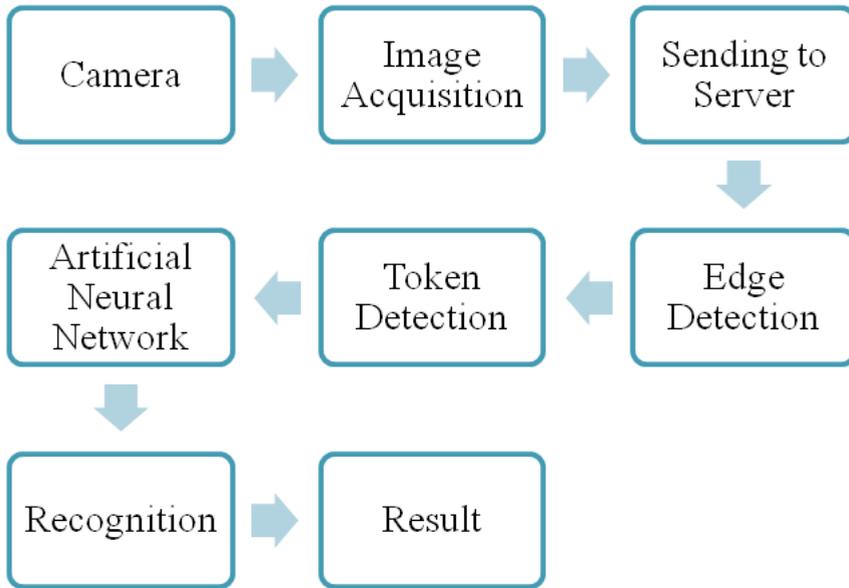


Figure 1.1: Block diagram of the system.

1.1 Introduction

Image processing is a rapidly growing area in diverse applications, such as multimedia computing, secured data communication, biomedical, biometrics, remote sensing, texture understanding, pattern recognition, content-based retrieval, compression, and many more. This is all about how a computer can sense pictorial data after processing an image. Among the set of gestures intuitively performed by humans when communicating with each other, pointing gestures are especially interesting for communication and is perhaps the most intuitive interface for selection. They open up the possibility of intuitively indicating objects and locations, e.g., to make a robot change moving direction or simply mark some object. This is particularly useful in combination with speech recognition as pointing gestures can be used to specify parameters of location in verbal statements.

This technology can be a boon for disabled people who are not able to speak hence cannot communicate. Also if the person speaks a different language than the receiver, then also, it can be used as a translator. The development of a natural interaction interface, where people interact with technology as they are used to interact with the real world had always been a challenge. A hand free interface, based only on human gestures, where no devices are attached to the user, will naturally immerse the user from the real world to the virtual environment.

In this chapter, an application for the disabled people has been discussed which has been developed on Android device. Android device brings the long-expected technology to interact with graphical interfaces to the masses. Android device captures the user's movements without the need of a controller. The basic block diagram of a hand

gesture recognition system is given in Fig.(1.1). The video sequences captured by an android camera are processed to make it suitable for extracting useful information about the user. In this work, a real time object selection and recognition method is proposed for the application for disabled people.

1.2 Hand Gesture Recognition

In the past decade, the computational power of computers has doubled, while the human computer interface (HCI) has not changed too much. When we work with a computer, we are constrained by intermediary devices (keyboards and mice). However, these are inconvenient and have become a bottleneck in human-computer interaction. In our daily life, we use speech to communicate with each other and use gestures to point, emphasize and navigate. They are the most natural and preferable means to interact with computers for human beings. To make computers understand this however is not an easy task.

Gesture recognition is a topic in computer science and language technology with the goal of interpreting human gestures via mathematical algorithms. Gestures can originate from any bodily motion or state but commonly originate from the face or hand. Gesture recognition can be seen as a way for computers to begin to understand human body language, thus building a richer bridge between machines and humans. Gesture recognition enables humans to communicate with the machine and interact naturally without any mechanical devices. Gesture recognition can be conducted with techniques from computer vision and image processing.

Gestures of the Hand are read by an input sensing device such as a mobile or a computer. It reads the movements of the human body and communicates with the computer that uses these gestures as an input. These gestures are then interpreted by using algorithm either based on statistical analysis or artificial intelligence techniques. The primary goal of gesture recognition research is to create a system which can identify specific human hand gestures and use them to convey information. By recognizing the hand symbols of a man it can help in communication with deaf and dumb people. It helps in taking prompt action at that time. Many Hand Gesture recognition projects are discussed in [9, 8, 11, 10, 17]. Many researchers [12, 13, 21] have tried with different instruments and equipment to measure hand movements like gloves, sensors or wires, but in these techniques the user has to wear the device which does not make sense in practical use. So people thought about a way of a contactless gesture recognition that could be considered as a research area in Computer Vision and which would be as natural as human to human interaction [20].

Gestures are usually understood as hand and body movement that can pass information from one person to another. Since we consider only hand gestures, the movement of the hand that expresses or emphasizes an idea, sentiment or attitude belongs to a gesture. Hand gestures can be classified into four categories according to the different application scenarios: conversational gestures, controlling gestures, manipulative gestures and communicative gestures [9]. We use conversational gestures to help express ourselves more clearly in our daily life. They are very subtle and need careful psychological studies. Controlling gestures can be designed to navigate in a virtual

environment. For example, we can ask the computer to drive a car to the south by pointing in that direction. Manipulative gestures serves as a natural way to interact with virtual objects [6, 7, 18]. Sign language is an important case of communicative gesture [19]. Deaf people rely on it to talk to each other. It is objective and well defined and rarely causes ambiguity which makes it suitable as a test-bed for gesture recognition systems.

The approaches can be broadly classified into “Data-Glove based” and “Vision-based”. Many recognition systems are based on the data-glove, an expensive wired electronic device. Various sensors are placed on the glove to detect the global position and relative configurations of the hand. One limitation of this method is the price of the glove and another problem with this approach is that it needs a physical link between the user and the computer. Because of this shortcoming, more and more researchers show interest in vision-based systems, which are wireless and the only thing needed is one or multiple cameras. In this work we present a vision-based system using one camera in an android device. Digital image processing allows a much wider range of algorithms to be applied to the input data, and can avoid problems such as the build-up of noise and signal distortion during processing [16].

1.3 Proposed Method

Signs are one form of hand gestures. Sign language is used as a communication medium among deaf and dumb people to convey the message with a normal person. A person who can talk and hear properly (normal person) cannot communicate with a deaf and dumb person unless he/she is familiar with the sign language. Same case is applicable when a deaf and dumb person wants to communicate with a normal person or a blind person. In order to bridge the gap in communication among deaf and dumb community and normal community, Video Relay Service (VRS) is being used nowadays. In VRS a manual interpreter translates the hand signs to voice and vice versa to help communication at both ends. A lot of research work has been carried out to automate the process of sign language interpretation with the help of image processing and pattern recognition techniques.

A lot of researchers initially used morphological operations to detect hand from image frames. It is an important research area not only from the engineering point of view but also for its impact on the society. Sign languages are non verbal visual languages, different from spoken languages, but they serve the same function. There are different sign languages all over the world such as the American Sign Language (ASL), British Sign Language (BSL), Japanese Sign Language, French Sign Language, Australian Sign Language, Indian Sign Language (ISL) etc.

We are developing a system using an android device and will also be tested with a webcam. This system will be useful for a deaf and dumb person carrying an android device or a system with a webcam. The Android device and the iBall camera used in this project are shown in Fig.(1.2). The algorithms for detection using a webcam and an android are shown in Fig.(1.3) and Fig.(1.4).

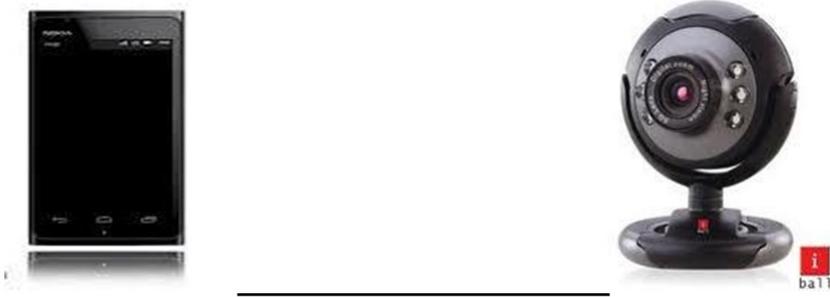


Figure 1.2: Android device and iBall camera.

1.3.1 Principal Component Analysis (PCA)

Like face recognition, which is inherently a classification problem in a high dimensional feature space, we also treat the recognition of hand gestures as a problem in the field of pattern recognition, and indeed, many techniques have been proposed in this area. PCA [2] is a standard tool in modern data analysis - in diverse fields from neuroscience to computer graphics - because it is a simple, non-parametric method for extracting relevant information from confusing data sets. It is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since it is hard to find patterns in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analyzing data [2]. The other main advantage of PCA is that once you have found these patterns in the data, you can compress the data, i.e. by reducing the number of dimensions, without much loss of information. This technique is used in image compression. PCA is a rather general statistical technique that can be used to reduce the dimensionality of the feature space [3].

Given a set of training objects represented by their feature vectors, x_i, y_i ($1 < i < n$), where n is the number of samples, the training set can be written as

$$\begin{bmatrix} x_1, y_1 \\ x_2, y_2 \\ x_3, y_3 \\ \dots\dots\dots \\ x_n, y_n \end{bmatrix}$$

whose mean vector would be $x_i - \text{mean}(x)$ and $y_i - \text{mean}(y)$ and covariance matrix is \mathbf{R}_x , defined as:

$$C^{n \times n} = (c_{i,j}, c_{i,j} = \text{cov}(Dim_i, Dim_j)),$$

where $C^{n \times n}$ is a matrix with n rows and n columns, and Dim_x is the x -th dimension.

The covariance matrix is a real symmetric square matrix of size N by N , where N is the length of the feature vector. The training set x, y corresponds to a cluster of

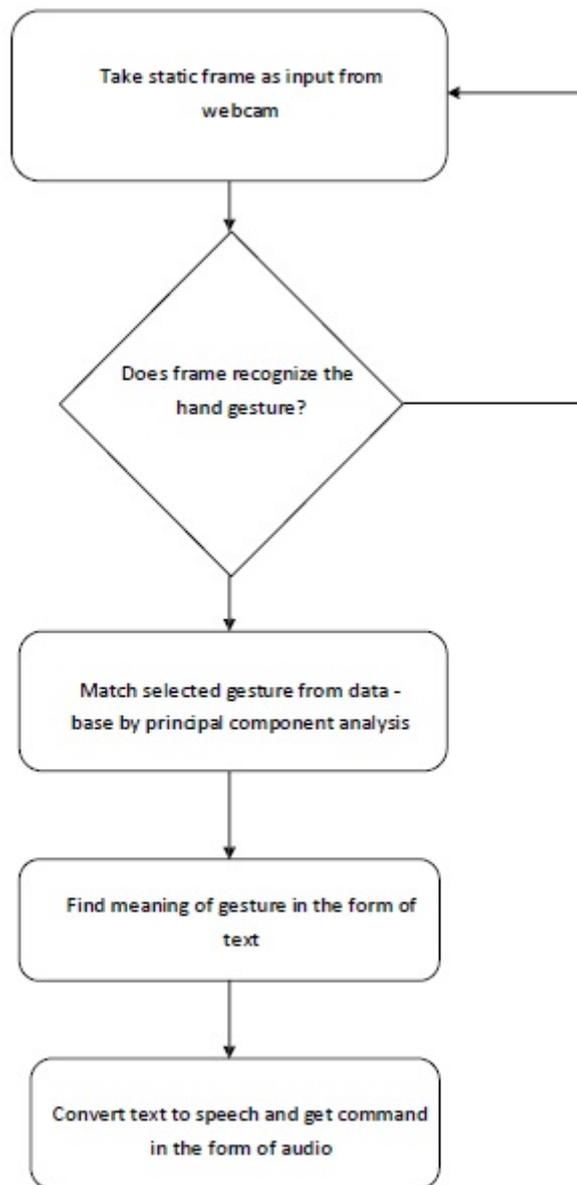


Figure 1.3: Block diagram of Sign Language Recognition model using webcam.

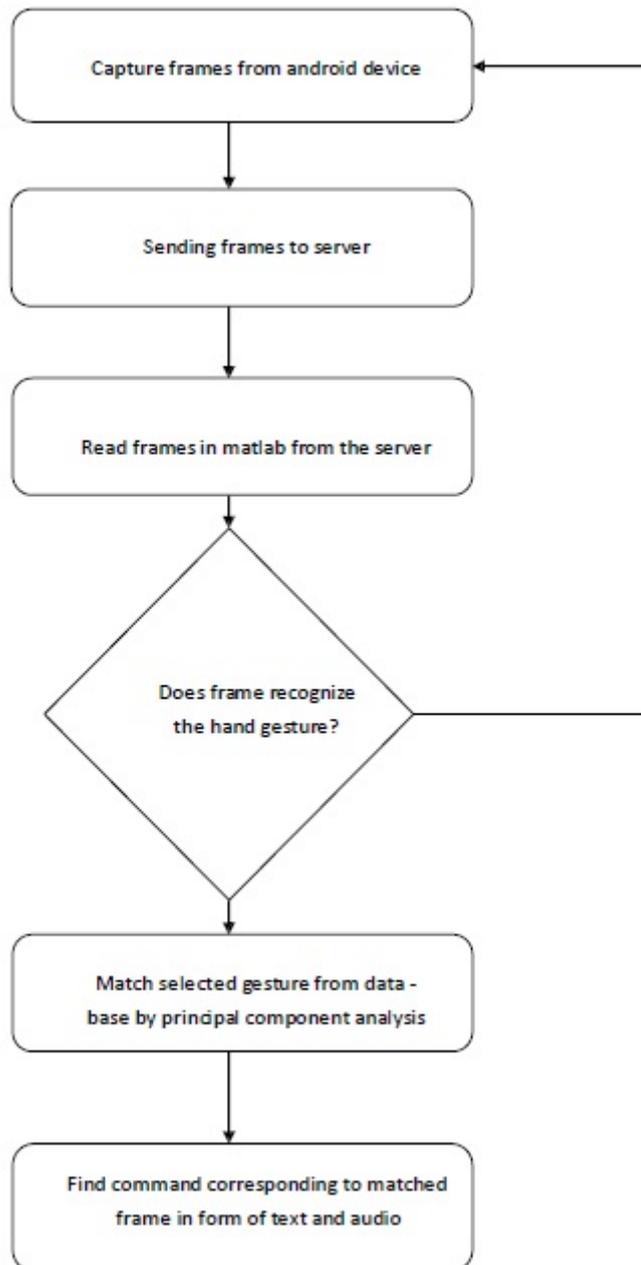


Figure 1.4: Block diagram of Sign Language Recognition using Android device.

data points in an N dimensional feature space. There exists redundancy in the feature space since the features, i.e the dimensions, are not independent to each other. By PCA we can eliminate the redundancy by transforming the original feature space into a so-called PC space in terms of Principal Components (PCs). Given the mean vector \mathbf{x}_m and the covariance matrix \mathbf{R}_x of a set of training samples, the eigenvector and eigenvalue would be calculated [2].

1.3.2 Extracting Frames from Live Video

This work was done for live realtime constraints, so we have to segment video into frames in such a memory efficient and fast way that complex processing on that frame can be done in real time for hand gesture detection. Here the system captures a frame per 1/3 second and three continuous frames are analyzed to determine motion in frame. It tries to capture the frame which is static. The difference between three continuous frames is less than a specified value. Here we can say that hand gesture is shown in this frame.

For detecting frame of interest, following processing is applied: Suppose three frames are A , B , and C . Motion parameter = $(A \text{ XOR } B) \text{ OR } (A \text{ XOR } C)$ Suppose size of frame is $M \times N$ pixels. If motion parameter is less than $(M \times N) / 100$ then frame A is considered as static and this frame is captured. This system is developed in MATLAB and in Java. It is tested and used successfully as a sub function in many works [1].

1.3.3 Image Acquisition

Image acquisition is the first step in any vision system. In this application it is done by using IPWebCam android application. The application uses the camera present in the phone for continuous image capturing and a simultaneous display on the screen. The image captured by the application is streamed over its Wi-Fi connection (or WLAN without internet as used here) for remote viewing. The program access the image by logging to the device's IP, which is then showed in the GUI. The result of image capture is shown is Fig.(1.5).

1.3.4 Edge Detection

In this program the Sobel edge detector is used as an edge detection technique being described in [14, 15]. The image captured is then passed through the Sobel filter and the results are shown in Fig.(1.6).

1.3.5 Hand Token

Here we will use a Neural Network which would be based on cosines and sinus angles of the shape. These angles would represent the criteria for a recognition system. Figure 1.7 shows an image with tokens. Each square represents a point on the shape of the hand image from which a line to the next square is drawn.

On zooming a part of Fig.(1.7) it shows a right-angled triangle between the 2 consecutive squares, as shown in Fig.(1.8a). This and the summary of all triangles of



Figure 1.5: Original image captured from Android device.

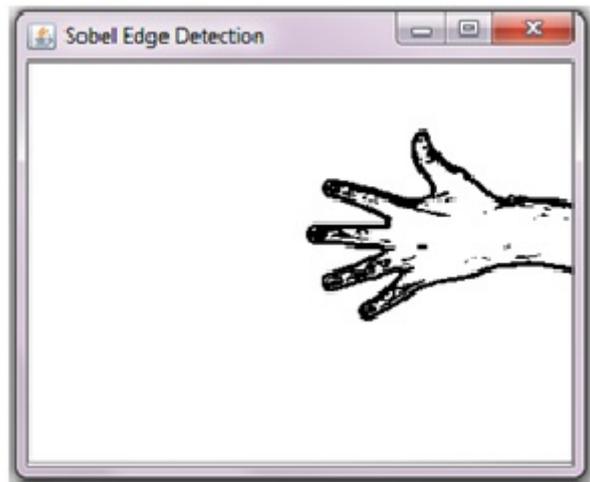


Figure 1.6: Sobel edge filtered image.

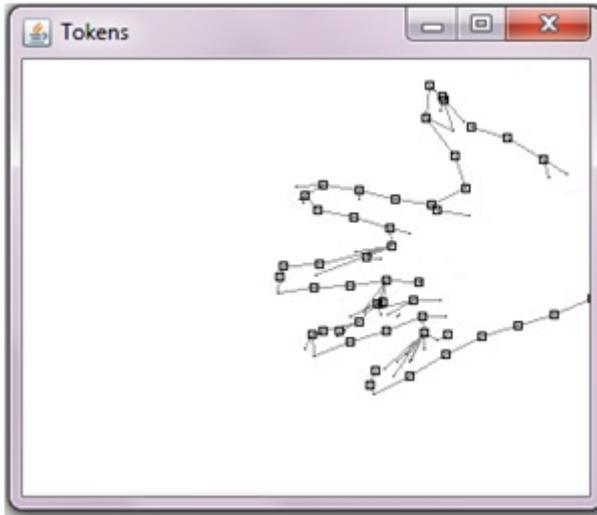


Figure 1.7: Generated token of the original image.

a hand image are the representation of the tokens of a hand from which we can start the neural network calculations.

The right-angled triangle in Fig.(1.8b) represents a token of a single hand image. The angles A and B are the two necessary parts which will be fit into the neural network layers. With the two angles we can exactly represent the direction of the hypotenuse from point P_1 to P_2 which represents the direction of a hand image.

1.3.6 Training Data

Another main part of this work is the integration of a feed-forward back-propagation neural network. As described earlier the inputs for this neural network are the individual

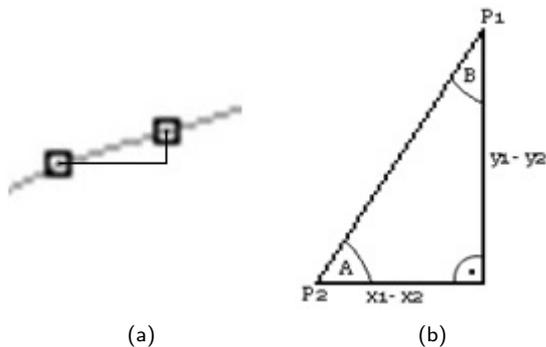


Figure 1.8: (a) zoomed image of image tokens and b) effective right angled triangle.



Figure 1.9: Sign Database.

S R T H X A G C I E

Figure 1.10: Sign Database means.

tokens of a hand image, and as a token normally consists of a cosines and sinus angle, the amount of input layers for this network are the amount of tokens multiplied by two. The implemented network just has one input, hidden and output layer to simplify and speed-up the calculations on that java implementation. For training purpose the database of images located on the disk is used. It contains different types of pre-defined gestures. These gestures are shown in Fig.(1.9) and also in Fig.(1.11). These gestures are first processed and then the tokens generated are passed to the network for training purpose. This process of training network from set images is done automatically when the application is initialized.

1.3.7 Webcam Implementation

Database has been developed for 10 sign gestures using webcam as shown in Fig.(1.9). These gestures have been taken from Indian Sign Languages that uses both hands except few gestures. Database of corresponding Alphabets has also been developed as shown in Fig.(1.10). Gestures have been captured in black background with single light source. One can capture and store more than one image of gestures in database to improve the matching rate. Images in database should be resized into same dimension (60×80).

1.3.8 Android Implementation

A different database has been developed for 10 sign gestures using android device as shown in Fig.(1.11). These gestures have been taken from Indian Sign Languages that uses both hands except few gestures. Database of corresponding Alphabets has also been developed are shown in Fig.(1.12). Gestures have been captured in black background with single light source. One can capture and store more than one image of gestures in database to improve the matching rate. Images in database should be resized into same dimension (100×100).



Figure 1.11: Sign Database.

G S C H I A
R T L X

Figure 1.12: Sign Database means.

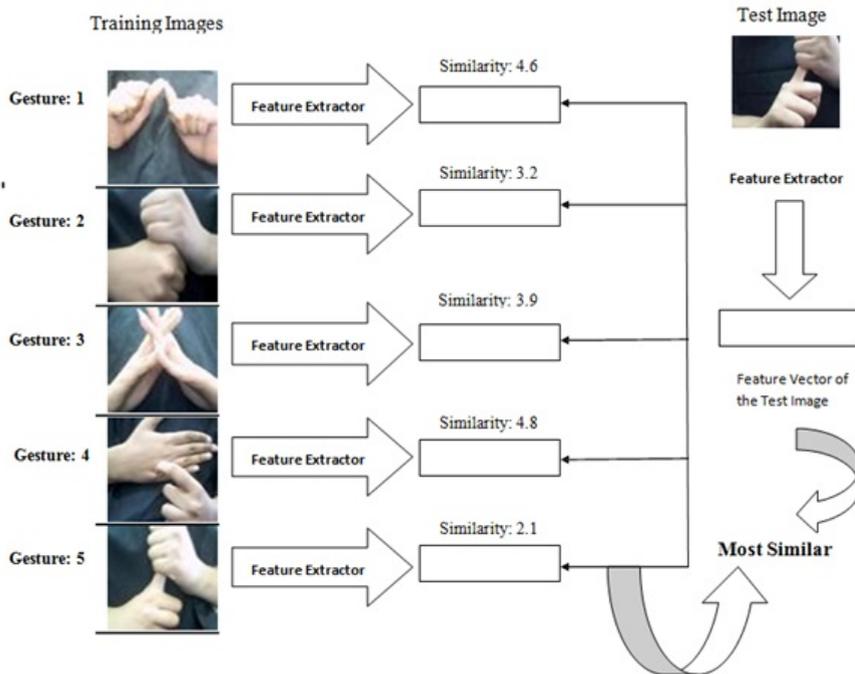


Figure 1.13: Recognition process.

1.3.9 Recognition

Recognition is the final step of the application. PCA was used for this. Principal component was calculated and stored for database images. As any new gesture would be shown to screen, its feature vector would be calculated and compared to database vectors. It gives percentage of recognition to each gesture with highest percentage closely matching and lowest to the farthest matching and the closest match is considered as the result. The match gesture would lead to actuation of the system. The recognition system is shown in Fig.(1.13).

1.4 Results

All input images were captured by webcam or an android device. The captured image is then read in matlab and converted in binary form of size defined so that it takes less time and memory space during pattern recognition. Now by using PCA it calculates the Eigen vectors and shows the equivalent image to the input gesture with minimum equivalent distance. At last it gives out the matched hand gesture and alphabet corresponding to the given input image from the database. It also produces an audio sound indicating the output sign gesture.

The results of webcam based recognition and android device systems are shown in Fig.(1.14) and Fig.(1.15) respectively. The accuracy for both methods are shown in

Table 1.1: Test results with Webcam.

Gesture	Recognition Rate for 10 trials
A	90%
C	80%
E	70%
I	80%
G	100%
H	80%
R	90%
S	90%
T	100%
X	100%

Table 1.2: Test results with Android device.

Gesture	Recognition Rate for 10 trials
A	80%
C	70%
E	100%
I	80%
G	70%
H	90%
R	100%
S	90%
T	100%
X	90%

Table 1.1 and Table 1.2.

1.5 Conclusions

This chapter discusses a hand sign recognition system which would be deployed on an Android device. The system is developed and tested successfully with webcam and an android device. This system is useful for a deaf and dumb person carrying an android device or a system connected with webcam. All gestures have recognition rate in between 70-100% which is an acceptable range. Overall accuracy of this system is 90% (approx.) while 77% with Android. The performance of the algorithm used for detecting sign gestures can be severely decreased due to varying lighting conditions and noises in the background [5, 4]. In future, a custom camera instead of the IpWebCam app which will further enhance the success rate of the system. Other different type of gestures can also be made part of the database.



Figure 1.14: Test results with Webcam.

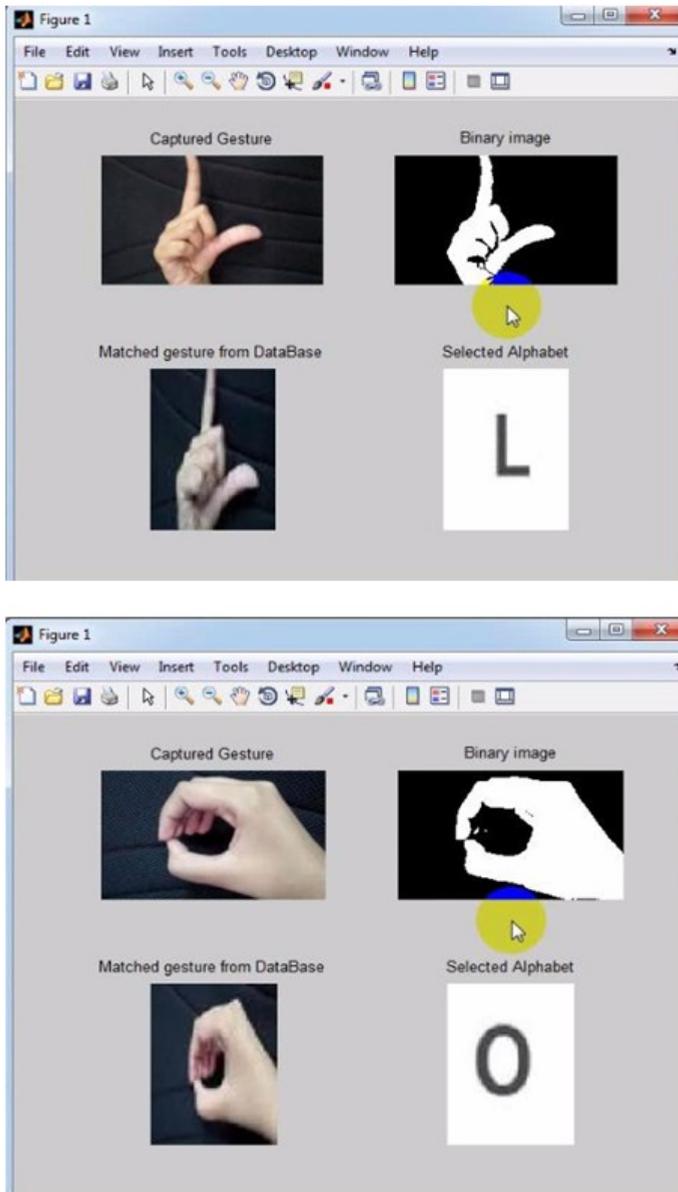


Figure 1.15: Test results with Android device.

References

- [1] Matlab programming tips, math works. www.mathworks.com.
- [2] Principal Component Analysis. http://en.wikipedia.org/wiki/Principal_component_analysis.
- [3] H. Birk, T.B. Moeslund, and C.B. Madsen. Real-time recognition of hand alphabet gestures using principal component analysis. In *10th Scandinavian Conference on Image Analysis*, pages 261–268, 1997.
- [4] A. Chaudhary. Finger-stylus for non touch-enable systems. *Journal of King Saud University - Engineering Sciences*, 2015. in press.
- [5] A. Chaudhary and A. Gupta. Automated switching system for skin pixel segmentation in varied lighting. In *19th IEEE International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*, pages 26–31, 2012.
- [6] A. Chaudhary, M.B.L. Manasa, and J.L. Raheja. Light invariant neuro-vision system for elder/sick people to express their needs into lingual description. In *Microsoft Research India's Annual Research Symposium (TechVista)*, 2012.
- [7] A. Chaudhary and J.L. Raheja. ABHIVYAKTI: a vision based intelligent system for elder and sick persons. In *3rd IEEE International Conference on Machine Vision*, pages 361–364, 2010.
- [8] A. Chaudhary, J.L. Raheja, and K. Das. A vision based real time system to control remote robotic hand fingers. In *IEEE International Conference on Computer Control and Automation*, pages 118–122, 2011.
- [9] A. Chaudhary, J.L. Raheja, K. Das, and S. Raheja. A survey on hand gesture recognition in context of soft computing. In *1st International Conference on Computer Science and Information Technology (CCSIT)*, volume 133 of *Advanced Computing Communications in Computer and Information Science*, pages 46–55, 2011.
- [10] A. Chaudhary, J.L. Raheja, K. Das, and S. Raheja. *Computer Vision and Image Processing in Intelligent Systems and Multimedia Technologies*, chapter Fingers' angle calculation using level-set method, pages 191–202. IGI Global, 2014.
- [11] A. Chaudhary, K. Vatwani, T. Agrawal, and J.L. Raheja. A vision-based method to find fingertips in a closed hand. *Journal of Information Processing Systems*, 8(3):399–408, 2012.
- [12] O.Y. Cho and et al. A hand gesture recognition system for interactive virtual environment. *IEEK*, 36(4):70–82, 1999.
- [13] T.S. Huang and V.I. Pavlovic. Hand gesture modeling, analysis and synthesis. In *International Workshop on Automatic Face and Gesture Recognition*, pages 73–79, 1995.
- [14] L. Kumari, S. Debbarma, N. Debbarma, and S. Deb. Image pattern matching using principal component analysis method. *International Journal of Advanced Engineering & Applications*, 1:6–9, 2011.
- [15] T. Lindeberg. *Encyclopedia of Mathematics*, chapter Edge detection. Springer, 2001.
- [16] R. Lockton and A.W. Fitzgibbon. Hand gesture recognition using computer vision. BSc. Graduation Project, Oxford University.
- [17] J.L. Raheja, A. Chaudhary, and S. Maheshwari. Hand gesture pointing loca-

- tion detection. *Optik - International Journal for Light and Electron Optics*, 125(3):993–996, 2014.
- [18] J.L. Raheja, Dhiraj, D. Gopinath, and A. Chaudhary. GUI system for elder/patients in intensive care. In *4th IEEE International Conference on International Technology Management Conference (ITMC)*, pages 1–5, 2014.
- [19] J. Rekha, J. Bhattacharya, and S. Majumder. Hand gesture recognition for sign language: a new hybrid approach. In *15th International Conference on Image Processing, Computer Vision, & Pattern Recognition (IPCV)*, 2011.
- [20] M. Sonka, V. Hlavac, and R. Boyle. *Image processing, analysis, and machine vision*. Cengage Learning, 2007.
- [21] D.J. Sturman and D. Zeltzer. A survey of glove-based input. *IEEE Computer Graphics and Applications*, 14(1):30–39, 1994.