

Real Time Gesture Recognition Using Gaussian Mixture Model

Aisha Meethian, B.M.Imran

Abstract— This paper investigates a real time gesture recognition system which recognizes sign language in real time manner on a laptop with webcam. Real time performance is achieved by using combination of Euclidistance based hand tracking and mixture of Gaussian for background elimination. In this paper gesture reorganization is proposed by using neural network and tracking to convert the sign language to voice/text format. The aim of research to develop a Gesture Recognition Hand Tracking (GR-HT) system for hearing impaired community. The experimental result shows that the proposed GR-HT system achieves satisfactory performance in hand gesture recognition.

Index Terms— Neural network, GMM, Centroid, Euclidean distance, Gesture recognition.

1 INTRODUCTION

The latest computer vision technologies and the advanced computer hardware capacity make real-time, accurate and robust hand tracking and gesture recognition promising. Many different approaches have been proposed such as appearance-based approaches and 3D hand model-based approaches [1]. Most of these approaches deal the hand gesture as a whole object and try to extract the corresponding mathematical description from a large number of training samples. These approaches analyze hand gestures without breaking them into their constituent atomic elements that could simplify the complexity of hand gestures. As a result, many current approaches are still limited by the lack of speed, accuracy and robustness. They are either too fragile or demand too many prerequisites such as markers, clean backgrounds or complex camera calibration steps, and thus make the gesture interaction indirect and unnatural. Currently there is no real-time vision-based hand tracking and gesture recognition system that can track and identify hand gestures in a fast, accurate, robust and easily accessible manner. The goal of this work is to build a real-time 3D hand tracking and gesture recognition system for the purpose of human-computer interaction. To achieve this goal, the characters of hand gestures need to be taken into account and the principles that can improve the system's performance in terms of speed, accuracy and robustness need to be applied.

There are several methods had been proposed in recent years to translate the sign language using gestures. In 1997, Ko and yang developed a finger mouse that enables a user to specify commands with the fingers as in [2]. Then, Raymond and Fitzgibbon develop gesture recognition using deterministic boosting[3]. Other methods include the colored glove

based method, skin color segmentation, video sequence appearance modelling and Hidden Markov Model (HMM) systems as in [4]- [6]. Neural networks, HMMs, distance-based, skin color based, and other statistical methods have successfully solved the sign language recognition when considered for Word/sentence recognition, as in [7]- [12]. It only requires video-based data collection and hence leads to a better natural interface for the user. Sign language when compared to the spoken language has different grammar. In spoken language, the speech is group of sentences where words in the sentence are linear (i.e. One word followed by another) whereas in sign language, Simultaneous structure exists with a parallel temporal and spatial configuration.

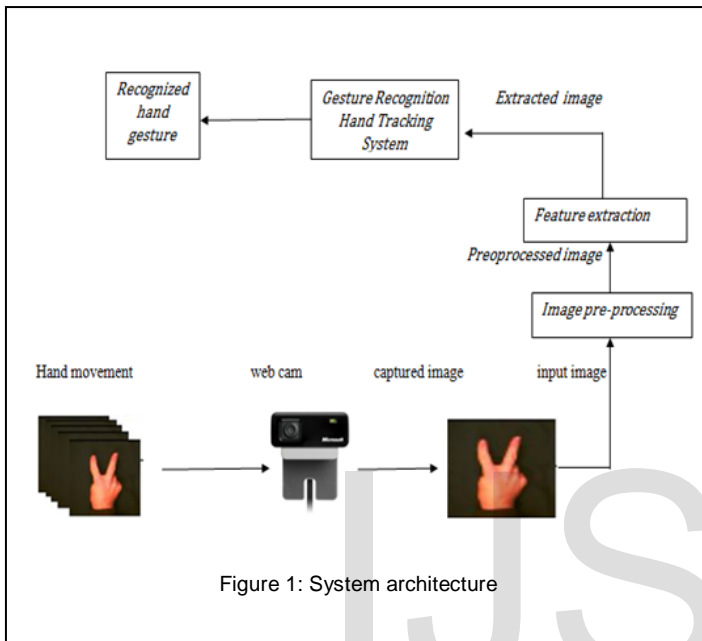
In this paper, we focus our attention to sign language recognition of hand gestures. The first part of the paper provides hand tracking of observed and recorded by typical video cameras. The second part of the paper presents a specific approach taken to gesture Recognition intended to support natural interaction Most of the complete hand tracking systems comprises of three layers: detection, tracking and recognition. The detection layer is responsible for defining and extracting visual features that can be attributed to the presence of hands in the field of view of the camera(s). The tracking layer is responsible for performing temporal data association between successive image frames, so that, at each moment in time, the system may be aware of " what happening there" Last, the recognition layer is responsible for grouping the spatiotemporal data extracted in the previous layers and assigning the resulting groups with labels associated to particular classes of gestures.

The paper is organized as follows. Section II, describes about architecture of proposed hand tracking system. Section III describes Real-Time hand gesture recognition, section IV describes Results and Analysis, and section V draws the Conclusion.

- Aisha meethian is currently pursuing masters degree program in Applied Electronics in Ilahia college of engineeing& tech, MG University, India. E-mail: aishamethian@gmail.com
- B.M.Imran is currently professor & Head of ECE Dept in Ilahia College of engineeing& tech, India. E-mail: professor.imran@gmail.com

2 ARCHITECTURE OF PROPOSED HAND TRACKING SYSTEM

The system proposed in this paper is designed to recognize hand gesture in real-time. The technique that is used to recognize hand gesture is based on computer vision. The overall system architecture is shown in figure 1. The whole system of hand gesture recognition is divided into four phases: Image Acquisition, Image Pre-processing, Feature Extraction and Hand Gesture Recognition.



Reading the video, Frame extraction and Pre-processing comes under video acquisition module. The initiation of the acquisition is first done manually. A camera sensor is needed in order to capture the features/web cam. Local changes such as noise and digitised errors should not change the image scenes and information's. In order to satisfy the memory requirements and the environmental scene conditions, pre-processing of the raw video content is highly important [13]. Various factors like illumination, background, camera parameters, and viewpoint of camera add complexity to the system. These conditions adversely affect images dramatically. The first most step of pre-processing block is filtering. It is used to remove the unwanted noise from the image scenes.

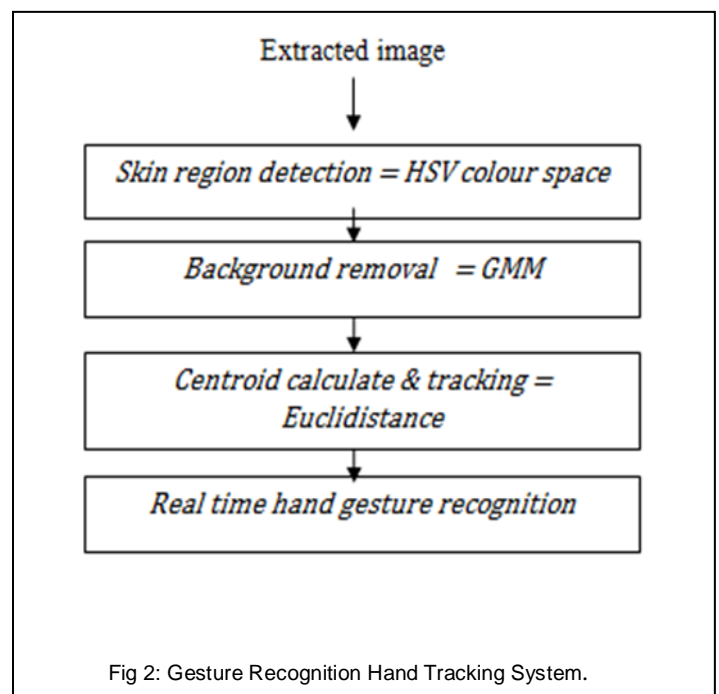
3 REAL TIME HAND GESTURE RECOGNITION SYSTEM

The communication between human and machines or between people can be done using gestures called sign language [14]. The use of sign language plays an important role in the means of communication method for the hearing impaired community [15]. American Sign Language (ASL) is the 3rd most-used language and the choice for most deaf people in the United States. Now a days about 500,000 to 2,000,000 people use sign language as their major daily communication tool.

Gesture recognition is generally based on two different ap-

proaches. Primarily, glove-based analysis [16-18] where either mechanical or optical sensors are attached to a glove that transforms finger flexions into electrical signals to determine the hand posture [17]. Currently, the vision-based analysis [19-21] is used mostly, which deals with the way human beings perceive information about their surroundings. The database for these vision-based systems is created by selecting the gestures with predefined meaning, and multiple samples of each gesture are considered to increase the accuracy of the system [20]. In this paper, we have used the vision-based approach for our gesture recognition application. Several approaches have been proposed previously to recognize the gestures using soft computing approaches such as artificial neural networks (ANNs) [21-22], and genetic algorithms [4]. ANNs are the adaptive self-organizing technologies that solved a broad range of problems such as identification and control, game playing and decision making, pattern recognition medical diagnosis, financial applications, and data mining in an easy and convenient manner.

Hand Gesture is habitually used in everyday life style. It is so natural way to communicate. Hand gesture recognition method is used in the application area such as for Controlling mouse and/or keyboard functionality, 3D World, Human/Robot Manipulation and Instruction Communicate at a distance. This paper introduces a real time hand gesture recognition system. Fig 2 shows the gesture recognition hand tracking system. This system consists of three stages: background removal using GMM, skin feature extraction using HSV colour space, hand tracking using Euclidistance and finally gesture recognition. In the first stage input image of hand gestures are acquiesced by digital camera in approximate frame rate. In second stage a rotation, translation, scaling and orientation invariant feature extraction method has been introduced to extract the feature of the input image based on moment feature extraction method. Finally, a neural network is used to recognize the hand gestures.



3.1 Skin region detection

HSV is used to detect skin region. The range of the skin colors depends up on the lighting conditions, thus the HSV color model which is robust to lighting Change is adopted for skin color localization. One of the most important advantages of this color model in skin color segmentation is that it allows users to intuitively specify the boundary of the skin color class in terms of the hue and saturation. In this paper, the ranges of the hue and saturation are set in between 0° and 5° and 0.23 to 0.68, respectively, as specified in [24]. Fig. 3 shows one of the segmentation results.

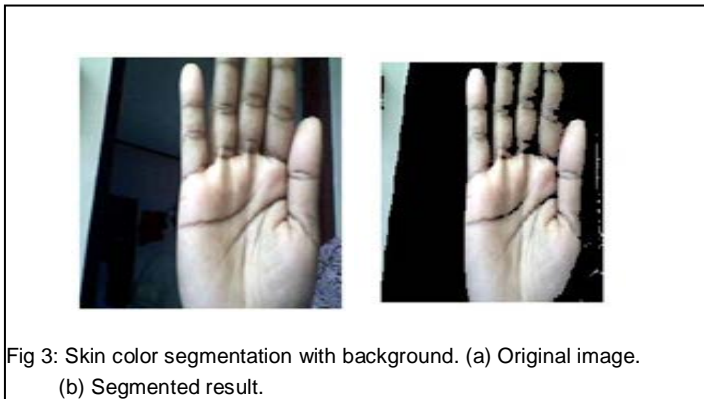


Fig 3: Skin color segmentation with background. (a) Original image. (b) Segmented result.

3.2 Background Removal

GMM is used for background removal. Modelling the values of all the pixels as one particular type of distribution, we simply model the values of a particular pixel as a mixture of Gaussians. Based on the persistence and the variance of each of the Gaussians of the mixture, we determine which Gaussians may correspond to background colors. Pixel values that do not fit the background distributions are considered foreground until there is a Gaussian that includes them with sufficient, consistent evidence supporting it. Our system adapts to deal robustly with lighting changes, repetitive motions of scene elements, tracking through cluttered regions, slow-moving objects, and introducing or removing objects from the scene. Slowly moving objects their color has a larger variance than the background. Also repetitive variations are learned, and a model for the background distribution is generally maintained even if it is temporarily replaced by another distribution which leads to faster recovery when objects are removed.

Suppose that each pixel can be represented as a mixture with more than one particular distribution, the GMM method maintains the persistence of K- Gaussians that correspond to the time series of background pixel values.

For each pixel, the probability can be calculated by the following function:

$$P(X_t) = \sum_{i=1}^K w_{i,t} \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (1)$$

Where X_t is each pixel value of the t th image frame, and $w_{i,t}$ is a weight vector for the corresponding Gaussian distribution that can be expressed as follows

$$\eta(X_t, \mu_{i,t}, \Sigma_{i,t}) = \frac{\exp\left[-\frac{(X_t - \mu_t)^T}{2} \Sigma^{-1} (X_t - \mu_t)\right]}{2\pi^{n/2} |\Sigma|^{1/2}} \quad (2)$$

Where $\mu_{i,t}$ is the mean value of the i th Gaussian, and $\Sigma_{i,t}$ is the covariance matrix of the i th Gaussian.

3.3 Centroid calculates & hand tracking

Euclidean distance is used to calculate the Centroid and to track the hand movement. Hand tracking is an application in field of tracking system. In our paper, the Euclidean distance is used to track the hand because of its simplicity and effectiveness. The Euclidean distance between two points $A(x_1, y_1)$, $B(x_2, y_2)$ as shown in Fig. 4(a) is Calculated as follows [24]:

$$\text{Euclidist}(A, B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (3)$$

Assume that in current frame, two hands are detected. The center of each hand is denoted as $ci(x_i, y_i)$, and the hand area is limited within a square with the parameter $ri = \text{Euclidist}(Chi, Bot_i)$, where the variable Bot_i is denoted on the bottom-right corner of Fig. 4(b). The green line in Fig. 4(b) indicates the trace of the hand in previous frames. The tracker in frame t has two parameters: center $t(x, y)$ and the distance dt . Now, the task is to determine center $t(x, y)$ belonging to which hand. Because the detection was implemented in real time, the distance between the center of hand I in frame t and frame $(t+1)$ is mostly less than r_i . Thus, the distance $dist_i(x, y)$ between $ci(x_i, y_i)$ and center $t-1(x, y)$ is computed. If the minimum value of $dist_i(x, y)$ is less than r_1 , then $ci(x_i, y_i)$ is the new point for tracking. According to Fig. 4(b), the satisfactory condition for tracking is $dist_1(x, y) < r_1$. Then, the center $t(x, y)$ is assigned to $C1(x_1, y_1)$. As a result, even a traced hand suddenly disappears; the system is still able to track the correct object after it appears again.

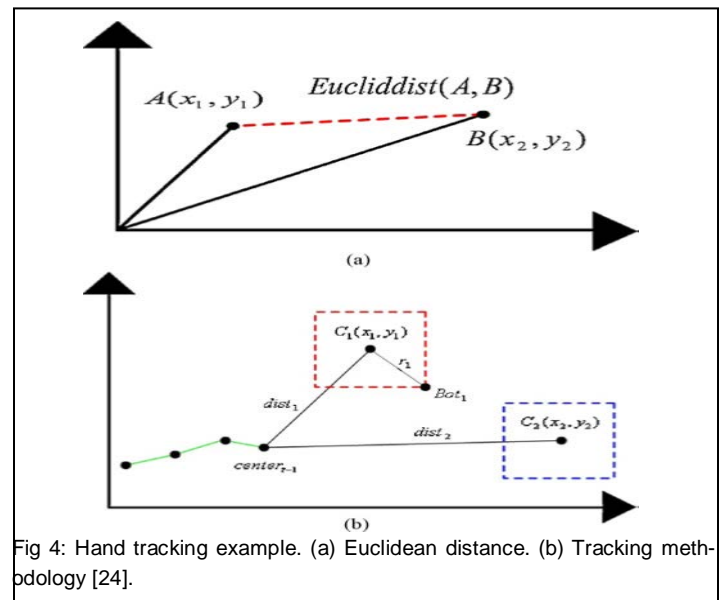


Fig 4: Hand tracking example. (a) Euclidean distance. (b) Tracking methodology [24].

3.4 Real time hand gesture recognition

The primary step in gesture recognition systems is the detection of hands and the segmentation of the corresponding image regions. A large number of methods have been proposed in the literature that utilize a several types of visual features and, in many cases, their combination. In this paper Neural Network is used to recognize hand gesture. Neural Network creates new networks. Basic of Neural Network is Euclidean distance. There are three layers in neural network. They are three types of layers Input layer, Inner layer and Output layer. The number of inner layers gives the perfection of the system. Properties of neural network are Training and Testing. In the gesture classification stage, a simple neural network model is developed for the recognition of gestures signs using the features computed from the video captured. Assumptions are made that there is no overlap between the two hands. Sign language recognition using neural networks is based on the learning of the gestures using a database set of signs.

4 .EXPERIMENTAL RESULTS

The system is implemented using MATLAB tool box and made the work easy because of the simplicity in design and easy use of toolbox. General software that can perform gesture recognition is MATLAB, Microsoft Visual C#, Microsoft Visual C++, and Microsoft Visual Basic etc. The most common software used are MATLAB and Microsoft Visual C# both of which are very powerful tools. MATLAB is chosen over others because it is a high-performance language for technical computing, perfect for speeding up development process as it allows the user to work faster with the features of toolboxes and provides ease and flexibility to the design of the programming. It is a tool of choice for high-productivity research, analysis, and development. It is preferred based on the previous research developed by Maung [25], Maraqa and Abu-Zaiter, [4] and on the review in [27, 28]. It integrates programming, computation, and visualization in user-friendly environment where problems and solutions are presented in common mathematical notation [29]. Also in industry, MATLAB tool is widely used for high-productivity research, development, and analysis [29].

4.1 GR-HT SYSTEM

The main objective of this paper is to track a hand and system is designed to recognize simple gestures or signs act as a media of communication for hearing impaired people. The design is very simple and does not require any kind of gloves also the system is applicable to different backgrounds. In this implementation, Microsoft LifeCam VX -700 is used for testing in real time, and the captured images are of size 160 × 120 under 30 f/s with natural lighting conditions. The Toshiba laptop with CORE i3 2.4GB and 4GB RAM is employed as the platform.

For training date the user has to keep the hand in front of the webcam to specify one or more regions on the hand to generate the color distribution at the initial stage. Before testing, samples are processed with the histogram equalization for resist against the lighting effect. In our paper, the processing time highly de-

pends on the complexity of the background. The Figure 5 represents the output of the Gesture Recognition Hand Tracking System Performed using the MATLAB.

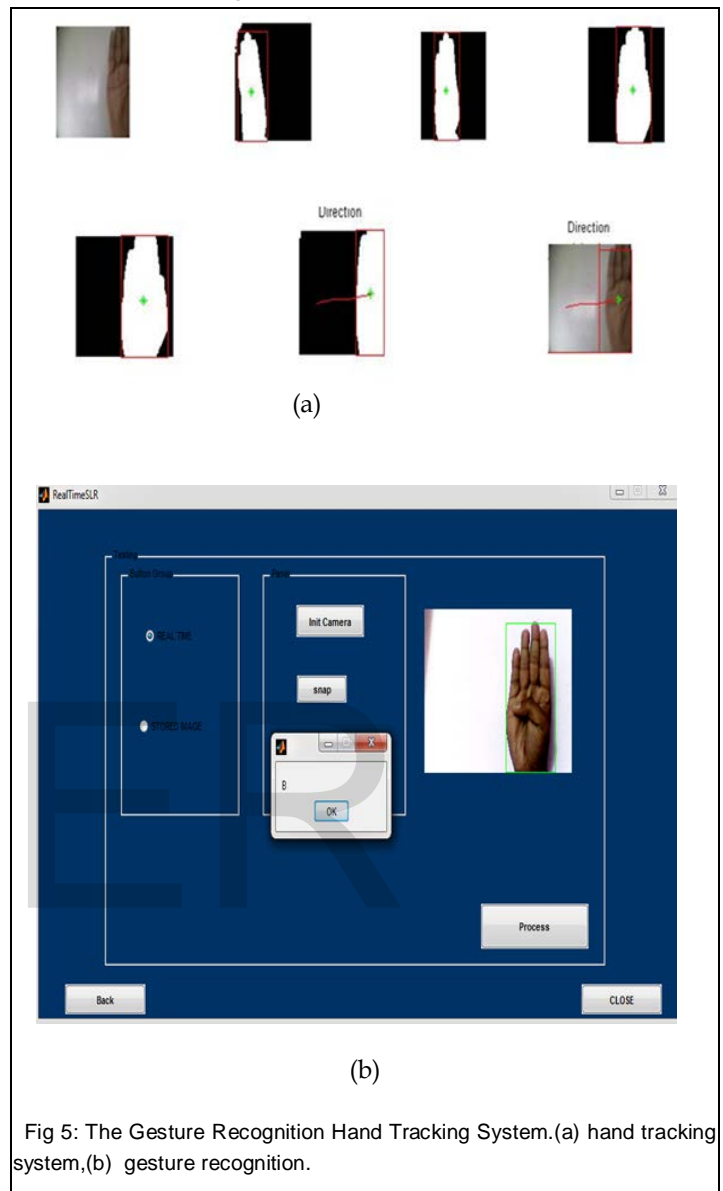


Fig 5: The Gesture Recognition Hand Tracking System.(a) hand tracking system,(b) gesture recognition.

5. CONCLUSIONS

This paper presents to recognize hand gestures in real-time with one camera as the input device. This architecture considers the hierarchical composite properties of hand gestures and combines the advantages to achieve real-time vision-based hand tracking and gesture recognition.

The goal of this paper was to reduce the required tracking time and further reducing the complexity in computation at tracking phase. According to the experimental results, the above tasks were achieved; meanwhile the tracking accuracy was maintained high. The signs for all the alphabets from A to Z are being recognized using the neural networks architecture. The advantage of using the neural networks is high processing speed so that the results in real-time manner. It is also advantageous as even noise corrupted, the signs can still be retrieved.

The system can detect and extract human hand from complex image that is an image where a human body is appeared, modify the system so that it can work in any lighting condition and expand the system to recognize the hand tracking.

References

- [1] H. Zhou and T. S. Huang, "Tracking articulated hand motion with Eigen dynamics analysis," in Proc. of International Conference on Computer Vision, vol. 2, 2003, pp. 1102–1109.
- [2] K. K. Byong and H. S. Yang, "Finger mouse and gesture recognition system as a new human computer interface," in Computer & Graphics, Vol. 21, No. 5, PP. 555 - 561, 1997.
- [3] Raymond Lockton and Andrew W. Fitzgibbon "Real-time gesture recognition using deterministic boosting" Department Of Engineering Science, University of Oxford.
- [4] Manar Maraqa and Dr. Raed Abu-Zaiter, "Recognition of Arabic sign language (ArSL) using recurrent neural networks", Applications of Digital Information and Web Technologies, pp: 478 – 481, 2008.
- [5] Yang quan, "Chinese Sign Language Recognition Based On Video Sequence Appearance Modelling", Industrial Electronics and Applications (ICIEA), 2010 the 5th IEEE Conference on, pp: 1537 –15
- [6] K.Kawahigasi, Y.Shirai, J.Miura and N.Shimada "Automatic Synthesis of training Data for Sign Language Recognition Using HMM", PROC.ICCHP, pp.623-626, 2006.
- [7] R. Liang and M. Ouhyoung (1998), "Real-time continuous gesture recognition system for sign language", Proc Third. IEEE International Conf: on Automatic Face and Gesture Recognition, pp. 558-567.
- [8] Bauer and Heinz "Relevant features for video-based continuous sign language recognition", Automatic Face and Gesture Recognition, 2000. Proceedings, pp: 440 – 445, 2000.
- [9] L. Rabiner and B. Juang. An introduction to hidden markov models. IEEE ASSP Magazine, pages 4–16, January 1986.
- [10] K. Grobel. Videobasierte Gebärdenspracherkennung mit Hidden Markov-Modellen. Ph. D. Thesis, Aachen University of Technology, VDI-Verlag, Dusseldorf, 1999.
- [11] Maebatake, Suzuki, Nishida, Horiuchi and Kuroiwa S. "Sign Language Recognition Based on Position and Movement Using Multi-Stream HMM", Universal.
- [12] E.-J. Holden, G. Lee and R. Owens "Australian Sign Language Recognition, Machine Vision and Applications", Vol. 16, pp. 312- 320, 2005.42.008.
- [13] Brian L. Pulito, Raju Damarla and Sunil Nariani, "2-D Shift Invariant image Classification Neural Network, which overcomes Stability, Plasticity Dilemma", Vol 2, International Joint Conference on Neural Network, San Deigo, June 17-21, 1990.
- [14] M. M. Hasan and P. K. Misra, "Brightness factor matching for gesture recognition system using scaled normalization," International Journal of Computer Science & Information Technology, vol. 3, no. 2, pp. 35–46, 2011.
- [15] M. P. Paulraj, S. Yaacob, M. S. bin Zanar Azalan, and R. Palaniappan, "A phoneme based sign language recognition system using skin color segmentation," in Proceedings of the 6th International Colloquium on Signal Processing and Its Applications (CSPA '10), pp. 1–5, May 2010.
- [16] S. Mitra and T. Acharya, "Gesture recognition: a survey," IEEE Transactions on Systems, Man and Cybernetics, Part C, vol. 37, no. 3, pp. 311–324, 2007.
- [17] T. B. Moeslund and E. Granum, "A survey of computer vision based human motion capture," Computer Vision and Image Understanding, vol. 81, no. 3, pp. 231–268, 2001.
- [18] J. J. LaViola Jr., A survey of hand posture and gesture recognition techniques and technology [M.S. thesis], NSF Science and Technology Center for Computer Graphics and Scientific Visualization, Providence, RI, USA, 1999.
- [19] S. Meena, A study on hand gesture recognition technique [M.S. thesis], Department of Electronics and Communication Engineering, National Institute of Technology, Rourkela, India, 2011.
- [20] M. M. Hasan and P. K. Mishra, "brightness factor matching for gesture recognition system," International Journal of Image Processing, vol. 4, no. 5, pp. 456–467, 2010.
- [21] P. Garg, N. Aggarwal, and S. Sofat, "Vision based hand gesture recognition," World Academy of Science, Engineering and Technology, vol. 49, pp. 972–977, 2009.
- [22] T. H. H. Maung, "Real-time hand tracking and gesture recognition system using neural networks," World Academy of Science, Engineering and Technology, vol. 50, pp. 466–470, 2009.
- [23] Jing-Ming Guo, Yun-Fu Liu, Che-Hao Chang and Hoang-Son Nguyen, "Improved Hand Tracking System" IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 5, may 2012
- [24] T. H. H. Maung, "Real-time hand tracking and gesture recognition system using neural networks," World Academy of Science, Engineering and Technology, vol. 50, pp. 466–470, 2009.
- [25] E. Stergiopoulou and N. Papamarkos, "Hand gesture recognition using a neural network shape fitting technique," Engineering Applications of Artificial Intelligence, vol. 22, no. 8, pp. 1141– 1158, 2009.
- [26] A. Chaudhary, J. L. Raheja, K. Das, and S. Raheja, "Intelligent approaches to interact with machines using hand gesture recognition in natural way: a survey," International Journal of Computer Science & Engineering Survey, vol. 2, no. 1, 2011.
- [27] W. T. Freeman and M. Roth, "Orientation histograms for hand gesture recognition," in Proceedings of the International Workshop on Automatic Face and Gesture-Recognition, pp. 296– 301, IEEE Computer Society, Zurich, Switzerland, June 1995.
- [28] M. Donoser and H. Bischof "Real time appearance based hand tracking," in Proc. Int. Conf. Patt. Recog, pp. 1–4, 2008.
- [29] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," IEEE Trans. Patt. Anal. Mach. Intell., vol. 27, no. 10, pp. 1615–1630, 2005.