© IJEDR 2018 | Volume 6, Issue 3 | ISSN: 2321-9939

Implementation of gesture recognition system using shape based feature

[1]Bhavana R.M , [2]Manikeshwari B.S
[1]Assistant Professor, [2]P.G Student
[1]Department of Computer Science and engineering,
[1]Visvesvaraya Technological University, Center for PG Studies, Kalaburgi, India

Abstract - Pattern recognition and Gesture recognition are the growing fields of research. It is possible to recognize and classify ten hand gestures based solely on their shapes. Hand Gesture recognition system provides us an innovative, natural, user friendly way of interaction with the computer which is more familiar to the human beings. A hand gesture recognition system can provide an opportunity for a mute person to communicate with normal people without the need of an interpreter. Gesture Recognition has a wide area of application including human machine interaction, sign language, immersive game technology etc. In this paper, we discuss a real-time human computer interaction system based on hand gesture. The whole system consists of three components: hand detection, gesture recognition and human-computer interaction (HCI) based on recognition; and realizes the robust control of mouse and keyboard events with a higher accuracy of gesture recognition.

Index Terms - Gesture recognition, convolution neural network, human-computer interaction, mouse cursor control. (key words)

I. INTRODUCTION

Computerized hand gesture recognition has received much attention from academia and industry in recent years, largely due to the development of human-computer interaction (HCI) technologies and the growing popularity of smart devices such as smart phones. As the most flexible part of human body, hands play an important role in human's daily life. It is very natural for a person to use his/her hands when he/she wants to physically manipulate an object or communicate with other people.
Traditionally, HCI is accomplished with devices such as mouse and keyboard, which are limited in terms of operational distance and convenience. By contrast, hand gesture recognition provides an alternative to these cumbersome devices, and enables people to communicate with computer more easily and naturally. In nowadays, hand gesture recognition technologies have already been successfully implemented in a wide range of applications, such as virtual reality [1], consumer electronics control [2], video games, sign language recognition [3] etc. Note that there is a clear distinction between hand posture and hand gesture, although people often consider them to be identical. Hand posture refers to a single *static* pose of the hand (e.g., the victory sign) and can be represented by a single image. In contrast, hand gesture is *dynamic*, which is composed of a sequence of changing postures posed in a short period of time. For example, waving goodbye belongs to this category. In the following part of this paper, we will focus on dynamic hand gesture recognition, although both areas do share some common techniques.

To enable hand gesture recognition, numerous approaches have been proposed, which can be classified into various categories. A common taxonomy is based on whether extra devices are required for raw data collecting. In this way, they are categorized into data glove based hand gesture recognition [4], vision based hand gesture recognition [5], and color glove based hand gesture recognition [6] (Fig. 1).

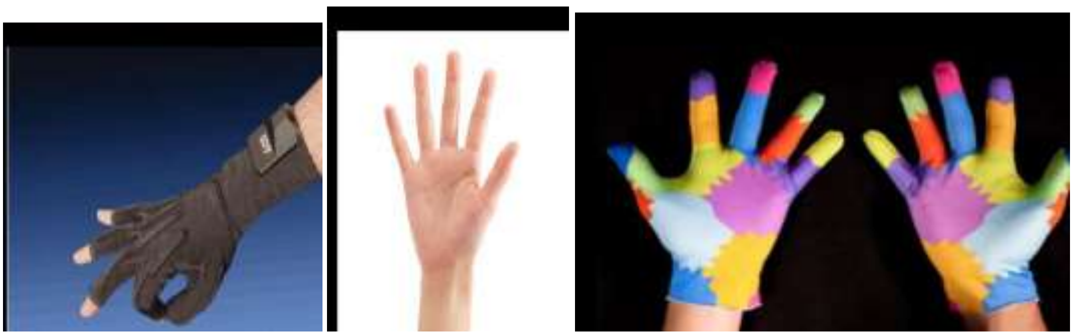


Figure 1. Data glove based, vision based and colour glove based hand gesture recognition (from left to right).

Data glove based approaches require the user to wear a cumbersome glove-like device, which is equipped with sensors that can sense the movements of hand(s) and fingers, and pass the information to the computer. The advantages of these approaches are high accuracy and fast reaction speed. However, these techniques are not very natural and flexible, and the data gloves can be quite expensive.

Figure 2: Examples of Hand Gesture

Vision based approaches do not require the user to wear anything (naked hands). Instead, video camera(s) are used to capture the images of hands, which are then processed and analyzed using computer vision techniques. This type of hand gesture recognition is simple, natural and convenient for users and at present they are the most popular approaches to gesture recognition. However, there are still several challenges to be addressed, for instance, illumination change, background clutter, partial or full occlusion etc.

Color glove based approaches represent a compromise between data glove based approaches and vision based approaches. Intrinsically, they are similar to the latter, except that, with the help of colored gloves, the image preprocessing phase (e.g., segmentation, localization and detection of hands) can be greatly simplified. The disadvantages are similar to data glove based approaches: they are unnatural and not suitable for applications with multiple users due to hygiene issues.

Vision based hand gesture recognition can be further categorized into appearance based approaches and 3D model based approaches [7]. Appearance based approaches extract the low-level features from 2D images and compare them with predefined template gestures. They are relatively simple in computation but the loss of the depth information makes it susceptible to background disturbance. In contrast, 3D model based approaches can exploit the depth information and are much more computationally expensive but can identify hand gestures more effectively.

II. RELATED WORK

HGR is the natural way of Human Machine interactions, and it can be applied to games, vision enabled robots, from virtual reality to smart home systems. There are many approaches associated with the accuracy of hand gesture recognition. Since the insufficient motion capture data is delivered by the KinectSDK, two major issues are the keys to success in hand gesture recognition: One is the hand region segmentation, and another is hand feature points positioning.

Zhou Ren [1] focused on building a robust part-based hand gesture recognition system using the Kinect sensor. In that system, in order to get the hand shapes, Finger-Earth Mover's Distance (FEMD) was used to measure the dissimilarity between hand shapes. HiaathamHasan[2] explored a multilayer neural network-based approach to recognize the hand gestures. Javier Molina[3] used static and dynamic models to get a real-time users independent hand gesture.

Radhikacentre [4] presented a computer vision method to recognize the hand gesture from the image captured by a webcam.

In the processing of hand region segmentation, Dan Xu[5] proposed a method which can locate hands simultaneously in real-time by using skin-color detection and K-means in conjunction with stereoscopic depth information, and using a YCbCr color space filter to locate hand regions. The histogram color-based image threshold can be used to detect skin on the human body, and use the GMM model to segment human hand regions[6]. The morphological filtering technique can be used to effectively remove background and object noise from binary images[7]. Zhong Yang et al introduced the HSV color space skin filter[8]. Antonis A. Argyro et al [9] used a Bayesian classifier which is boot strapped with a small set of training data to detect skin-colored objects [10, 11].

In terms of positioning feature points, traditional approaches mainly focus on looking for the center of a maximum inscribed circle [12]. Jagdish L Raheja et al located the palm center by using distance transformation. After analyzing images captured by Kinect, we extend the idea presented [11] by positioning feature points based on morphological operations.

The Graham Scan is used in [10] to find the convex hull of a hand including the fingertips, but this process is not always precise due to the capture device factors and dynamic hand movements. Wang C[14] presented a hand gesture recognition algorithm by using the depth and skeleton from Kinect. Raheja, J.L.et al assumes that the fingertip is the point with minimum depth in each finger point cluster [13]. The problem with this method is it cannot precisely conduct certain fingertips when they are blending. JongShill Lee et al. use the average distance between the point on the contour and the center point as the feature property of gesture recognition. Since the feature is relatively simple, just single float type, so they judge the gesture according to the value without using an additional classifier [15]. Manavender R. Malgireddy et al. propose a dynamic recognition framework using Sub-gesture model [16]. They describe the method of positioning gesture fragment by Sub gesture and use the classifier based on HMM to experiment for zoom and rotate gesture. Zhong Yang uses a six dimensional feature vector including Hand Position. Hand Velocity, Hand Size, Hand Shape and other feature information in the lecture [20]. He proposes a gesture fragment positioning method based on state and the data alignment method when using the classifier based on HMM.

III.ARCHITECTURE

The architecture diagram of the proposed system is shown in Figure 3.

- **Preprocessing**; localize and track the hand to generate its motion trajectory (gesture path).
- **Feature extraction**: Clustering extracted features to generate discrete vectors, which are used as input to HMMs recognizer.
- **Classification**: The gesture path is recognized using discrete vector and Left-Right Banded topology.

Step 1: Image Preprocessing and Segmentation:

Image preprocessing is a necessary step for image enhancement and for getting good results. The RGB images are captured using a 7 MP webcam. The input sequence of RGB images has to first be converted into YCbCr images as RGB color space is more sensitive to different light conditions. YCbCr is a family of color space where Y is a luma component that deals with luminance information of image, Cb and Cr are the blue difference and red difference chromo components that deal with the color information of the image. As luminance creates many problems, it is desirable to process only the chrominance component. Image segmentation is typically performed to separate the hand object in the image from its background. Segmentation based on YCbCr requires a plain and uniform background. Hence, we have to convert the YCbCr images to binary images.

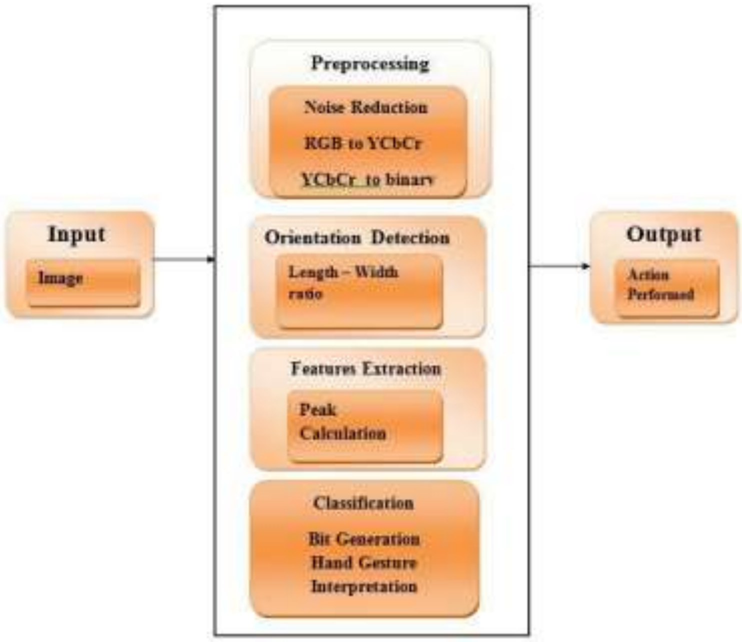


Figure 3: Architecture of the proposed system

Step 2: Orientation Detection.

For orientation detection, we have used a dual approach in order to increase the accuracy. To find out whether the orientation of the hand is horizontal or vertical, we compute the length to width ratio of the bounding box. It is assumed that if the length of the bounding box is greater than the width i.e. if the length to width ratio is greater than 1, the orientation is vertical. Likewise, if the length to width ratio is less than 1, the orientation is horizontal. Secondly, we implement the edge detection algorithm for finding the edges of the hand. Edge detection is used to identify the points at which the brightness in the image changes very sharply. To detect the edges of the hand, we scan the image and extract those pixels in the image where the pixel value changes rapidly from 0 to 1 i.e. from black to white. After edge detection, to identify the orientation of the hand, we scan the boundary matrices or edges of the hand in the binary image. Whenever we get the x-boundary equal to 1 along with increasing value of y-boundary for some time span, we classify it as a horizontal hand and if we get the y-boundary equal to maximum size of the image along with increasing value of x-boundary, it is classified as a vertical hand.

Step 3: Features Extraction

1) **Centroid**: Centroid of the hand is calculated to partition the hand into two halves, one part having fingers and the other having no fingers. Centroid is also called as centre of mass if the image is uniformly distributed and it divides the hand into two halves at its geometric centre.

2) **Thumb Detection**: Thumb detection is performed to detect the presence or absence of the thumb in the hand gesture. The thumb is considered as an important shape feature for classification of hand gestures. To detect the presence of the thumb we consider 30 pixels width from each side of the previously calculated bounding box and crop the bounding box into two boxes i.e. left box and right box. We then count the number of white pixels present in each box of the binary image.
3) **Finger Region Detection**: The finger region is obtained by computing the centroid. In this step, we determine the number of fingers raised and the number of fingers folded in the hand gesture. Peaks are used to represent tips of fingers. Vertical and horizontal images are processed in a different manner for finger detection. For the vertical hand, only y-coordinates of the boundary matrices are considered. When the y-coordinate values start increasing after sharp decrement in y-coordinate values, it is considered a peak value or fingertip. For the horizontal hand, only x-coordinates of the boundary matrices are traced. When the x-coordinate values start decreasing after continuous increment in x-coordinate values, it is considered a peak value.

Step 4: Classification and Bits Generation.

The hand gestures are classified based on the features calculated in the previous section. A 5 bit binary sequence is generated in which each bit position corresponds to a finger. Significant peaks or raised fingers are encoded as 1 and insignificant peaks or folded fingers are encoded as 0 in the binary sequence. The leftmost bit in the 5 bit binary sequence is reserved for status of the thumb in hand image. If the thumb is present, leftmost bit will be set to 1, otherwise set to 0.

IV.METHODOLOGY

The work is mainly concentrated on the following methodologies which will perform the task required. Which are explained in the following,

1. Non-linear Image De-noising
2. Pixel Orient analysis
3. Skin mapping
4. Euclidian distance

1. Non-linear Image De-noising:

The noise is an undesirable signal present in the image. It might have been caused during the storage transmission or processing. Some of the noises like Gaussian noise, impulse noise, coherent noise, granular noise, poison noise etc.

It is a simple process of removing the possible noise which may be present in the image input. The removing of noise from the image is called Image De-noising.

Images always get corrupted due to the transmission errors, storage time or addition of unwanted noise to the image. Image de-noising has been a major issue in digital image processing for many years.

Many of the techniques have been proposed like image de-noising with Wavelet transformation, Fuzzy logic transformation, and threshold based de-noising, Spread spectrum, Multilevel clustering of the pixels etc.

The Discrete Wavelet Transform is accordance to a sub band coding, it was set up for the speedy computation of Wavelet Transform. The major advantage of Discrete Wavelet Transform is that it is effortless to implement and reduce the instant management for the resources.

2. Pixel Orient analysis

The image captured is initially processed by the user based on the orientation of the angle of deviation. The hand gesture has to be processed based on the system threshold. The initial movement captured in the image feature of considered in the view angle of the aperture. The image orientation is analyzed by the following, process

1. input image
2. skin mapping for the area localization
3. Skin pixel masking and labeling
4. Skin area extraction
5. Estimate the alignment and orientation
6. Calculate the centric , area as features
7. Result passed to the later stages

Image skin pixel mapping is important aspect the input image may consider the undesired objects in the path . hence to solve and remove the unwanted parts the skin pixel mapping by GLCM is important. The Gray level co-occurrence matrix is used for the skin pixel extraction

The gray-level co-occurrence matrix P[i, j] is consist of displacement vector(d) in which the pixel values are grouped in[3*3] array vector. The pixel value of each in the original image are copied and replaced by the equitant image pixel in gay level scale

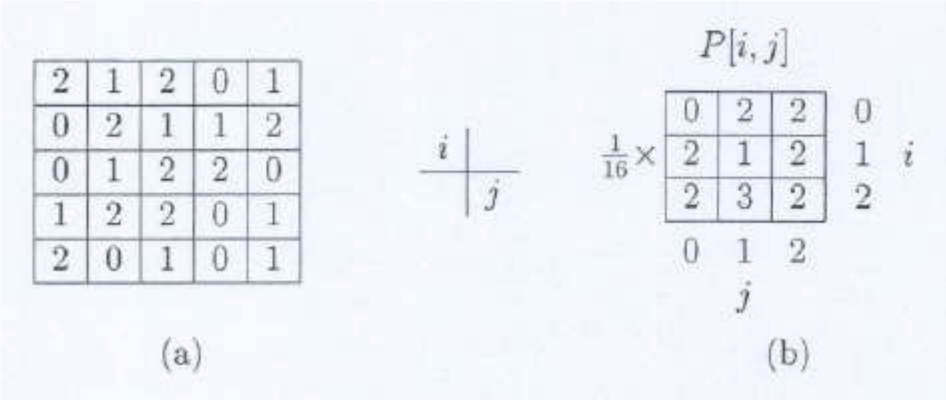


Figure 3: (a) A 5×5 image with three gray levels 0, 1, and 2. (b) the gray level co-occurrence matrix for d= (1, 1).

3. Skin mapping

Skin mapping of\r the skin pixel labeling place an important role as it's the base for the later stages where the image is proceeds based on the extracted pixel values and the features form the image are extracted by these labeled pixels. Hence the probability of the image pixel in the form of $c = [Cb\ Cr]^T$, is

$$p(c/skin) = \sum_{j=1}^{M} p(c/j).P(j) \qquad (1)$$

,*M*- number of Gaussian components model

P(j), the probability pixel at jth location which is ,

$$p(c/j) = \frac{\exp\left[-\frac{1}{2}(c-\mu_j)^T \Sigma_j^{-1}(c-\mu_j)\right]}{2\pi\sqrt{|\Sigma_j|}} \qquad (2)$$

$\boldsymbol{\mu_j}$- mean

$\boldsymbol{\Sigma_j}$ -covariance matrix of the j^{th} pixel

4. Euclidian distance

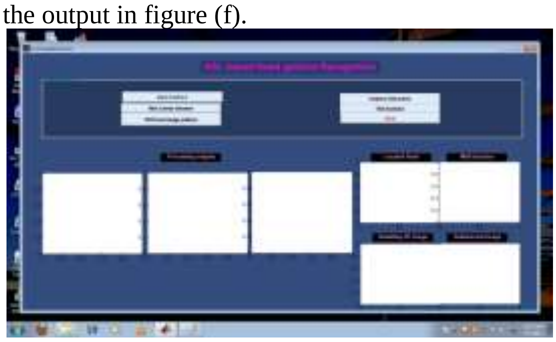
The Euclidian distance is the simple but yet effective way to classify the images based on the stats available in the database. The two image values are dumped into the system aiming which the image values which is nearest in distance is choose as the result to the user n general, for an *n*-dimensional space, the distance is

$$d = \sqrt{\sum_{i=1}^{v} (p_{1i} - p_{2i})^2}$$

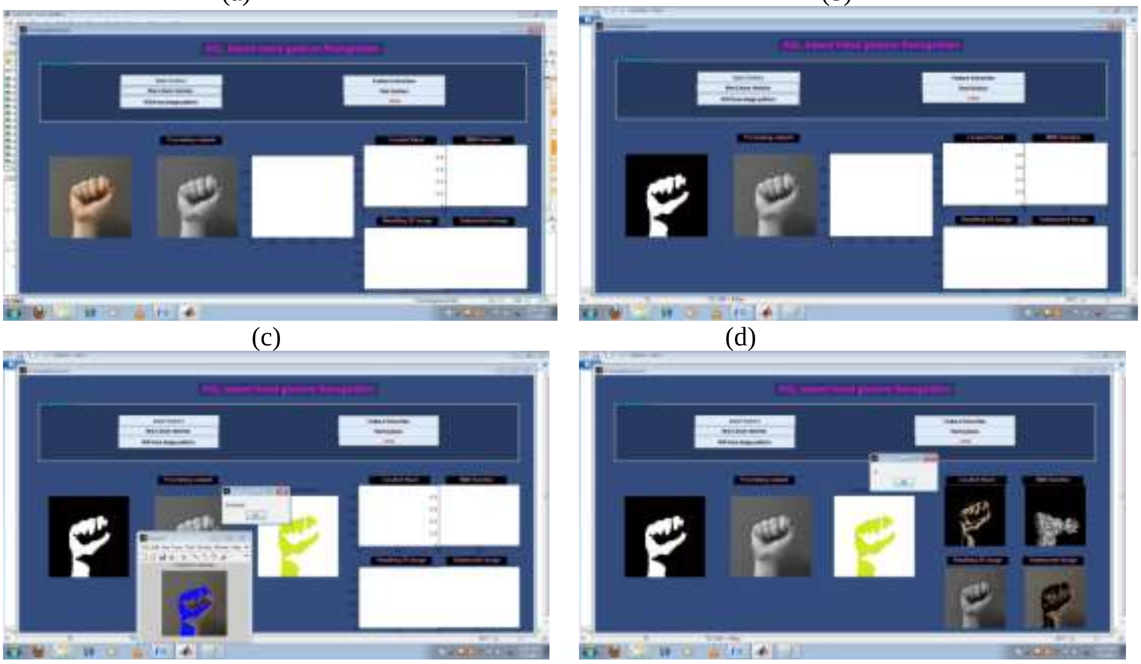
Normalized Euclidian distance, where v u is stats vector of pixel list 1 and pixel list 2 which are images in our scenario.

V. Result And Discussion

We have applied the above discussed algorithm and with this algorithm we have tested some images with few different patterns. By using these effective shapes based features and encoded bit sequence, we can recognize and classify few different hand gesture patterns. On the basis of generated binary bit sequences we can assign different-different task to support human computer interaction are show in figure 3. Figure (a) indicates initial HGR for the project execution where it consists of different options like input image, non linear de-noise etc. which will help in the processing. Figure (b) it indicates a hand image which is inserted by the button select image .The image will be loaded from the available data base and display it for the user. Figure (c) extracts the skin part of the given input image. Figure (d) indicates that pattern components options where it extracts the position of the region of interest. Figure (e) indicates the options that is Extracted. Finally once we click the button resulting emotion it displays the output in figure (f).



(a) (b)



(c) (d)

(e) (f)

Figure 4: Experimental output of the proposed system

VI.CONCLUSION AND FUTURE WORK

The proposed system hand gesture has been made very simple to me an understand. The ASL has provided the platform for hand gesture based on multiple features like texture, pattern and shapes. This recognition system is very useful for visually impaired people as they can make use of hand gestures for writing text in an electronic document like MS office ,Notepad etc.,Each interpreted gesture is assigned with corresponding English alphabet, number and some very useful key press events for generating text.

The future works include determining different complicated gestures of hand in less time with more accuracy.

REFERENCES

[1] D. Y. Xu, “A Neural Network Approach for Hand Gesture Recognition in Virtual Reality Driving Training System of SPG,” In Proceedings of the 18th International Conference on Pattern Recognition, vol. 3, pp. 519-522, 2006.

[2] W. T. Freeman and C. D. Weissman, “Television Control by Hand Gestures,” In Proceedings of International Workshop on Automatic Face and Gesture Recognition, pp. 179-183, 1995.

[3] T. Starner, J. Weaver and A. Pentland, “Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20(12), pp. 1371-1375, 1998.

[4] J. Weissmann and R. Salomon, “Gesture Recognition for Virtual Reality Applications Using Data Gloves and Neural Networks,” In Proceedings of the International Joint Conference on Neural Networks, vol.3, pp. 2043-2046, 1999.

[5] Y. Wu and T. S. Huang, “Vision-Based Gesture Recognition: A Review,” In Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction, pp. 103-115, 1999.

[6] R. Y. Wang and J. Popovi_, “Real-Time Hand-Tracking with a Color Glove”, ACM Transactions on Graphics, vol. 28(3), Jul 2009.

[7] V. I. Pavlovic, R. Sharma and T. S. Huang, “Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19(7), pp. 677-695, 1997.

[8] Zh. Y., Y. Li, W.D. Chen, Y. Zheng, “Dynamic hand gesture recognition using hidden Markov models,” Proceedings of 2012 7th International Conference on Computer Science & Education, Melbourne, VIC, pp. 360-365, July 2012

[9] A.A. Argyros, and M.I. Lourakis, “Real-time tracking of multiple skin colored objects with a possibly moving camera,” Proceedings of 8th European Conference on Computer Vision, Prague, Czech Republic, pp. 368-379, May 2004

[10] Y. Wen, Ch. Y. Hu, G.H. Yu, Ch. B. Wang. “A robust method of detecting hand gestures using depth sensors,” 2012 IEEE International Workshop on Haptic Audio Visual Environments and Games, Munich, pp. 72-77, Oct. 2012

[11] M.K. Bhuyan, D.R. Neog, and M.K. Kar, “Hand pose recognition using geometric features,” Proceedings of 2011 National Conference on Communications, Bangalore, pp. 1-5, Jan. 2011

[12] H. Liang, J.S. Yuan, and D. Thalmann, “3D fingertip and palm tracking in depth image sequences,” Proceedings of the 20th ACM international conference on Multimedia, Nara, Japan, pp. 785-788, Oct. 2012

[13] C. Wang, and S. Chan, “A new hand gesture recognition algorithm based on joint color-depth Superpixel Earth Mover's Distance,” 2014 4th International Workshop on Cognitive Information Processing, Copenhagen, Copenhagen, pp. 1-6, May 2014

[14] J.L. Raheja, A. Chaudhary, and K. Singal, “Tracking of fingertips and centers of palm using kinect,” Proceedings of 2011 Third International Conference on Computational Intelligence, Modelling and Simulation, Langkawi, pp. 248-252, Sept. 2011

[15] J.S. Lee, Y.J. Lee, E.H. Lee, and S.H. Hong, "Hand region extraction and Gesture recognition from video stream with complex background through entropy analysis," Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Francisco, CA, USA. pp. 1513-1516, Sept. 2004
[16] M.R.Malgireddy, J.J. Corso, S. Setlur, V. Govindaraju and D. Mandalapu, "A Framework for Hand Gesture Recognition and Spotting Using Sub-gesture Modeling," Proceedings of the 20th International Conference on Pattern Recognition, Istanbul, pp. 3780-3783, Aug. 2010