

A Study on Different Hand Gesture Recognition Techniques

Anju S R

Department of Computer Science and Engineering
SCT College of Engineering,
Trivandrum, India.

Subu Surendran

Department of Computer Science and Engineering
SCT College of Engineering
Trivandrum, India.

Abstract -- *Gesture recognition is an area in computer science and language technology that aims in defining human gestures via mathematical algorithms. With gesture recognition it is possible for humans to interact naturally with machines without the aid of any mechanical devices. Hand gesture is one of the most expressive and most frequently used among a variety of gestures. Applications of hand gesture recognition are varied from sign language to virtual reality. This paper provides a study on four different techniques in hand gesture recognition based on hand segmentation, real time gesture recognition, neural network shape fitting and finger earth mover's distance.*

Keywords – *Gesture, Hand Segmentation, Colour Models*

I. INTRODUCTION

Gesture is a means of communication through any bodily motion or state which commonly originate from face or hand. The aim of gesture recognition technology is to interpret human gestures through mathematical algorithms. Primitive systems used Text User Interfaces or Graphical User Interfaces which limit the majority of input to keyboard and mouse. Through gesture recognition it is possible for humans to communicate naturally with machines without using any mechanical devices.

Hand gesture is the most powerful and frequently used gesture in people's daily life. In linguistics, it is an important component of body language. Applications of hand gesture recognition includes teleconferencing, interpretation and learning of sign languages, telerobotics, controlling television set remotely, enabling hand as a 3D mouse and so on. Hand gesture recognition can be considered as a complex problem since gestures vary between individuals and for the same individual based on different contexts. The challenges in this technology include uncontrolled environment and lighting conditions, skin colour detection, rapid hand motions and self occlusions.

Several approaches have been introduced for handling hand gesture recognition in which some makes use of mathematical algorithms whereas some is based on soft computing. The practical implementation of gesture

recognition requires devices or gadgets such as those for tracking hand motion, imaging etc. Glove based techniques require user to wear devices that needs loads of cables to connect it to the computer thereby reducing the ease of interaction. Vision based techniques makes use of cameras and can obtain properties such as texture and colour of hand for identifying gesture, while dealing with problems due to illumination changes, complicated background, camera movement, specific user variance and self occlusions.

This paper provides an overview of four different techniques for hand gesture recognition. In the first technique hand gesture recognition is based on a static gesture set and two different colour models. In real time hand gesture recognition hand tracking and segmentation along with scale-space feature detection is the technique implemented. Hand gesture recognition in third method is based on a neural network based shape fitting technique. In the final method that we discuss, a kinect sensor is used to capture the hand shape and makes use of Finger-earth mover's distance metric for hand dissimilarity measure.

II. GESTURE RECOGNITION SYSTEM

The gesture recognition system consists of a sensor or a tracking technology that captures the gestures of the user. This can be either a digital camera or sensors like flex sensors, kinect sensors or devices like wired gloves. Gestures can be sign language, hand gesture or facial gesture. These gestures captured by the tracking media is then fed as input to the recognition algorithm. Various processes such as feature extraction, segmentation, region detection, gesture recognition and mapping are carried out by the gesture recognition system. The system then carries out the appropriate process or gives the desired response as output to the user.

In most of the currently implemented hand gesture recognition systems, kinect sensors or wired gloves or depth aware cameras are used for tracking hand gesture. The techniques for the recognition used in algorithms include hand region detection and segmentation. This section gives an overview of such terms and processes related to hand gesture recognition.

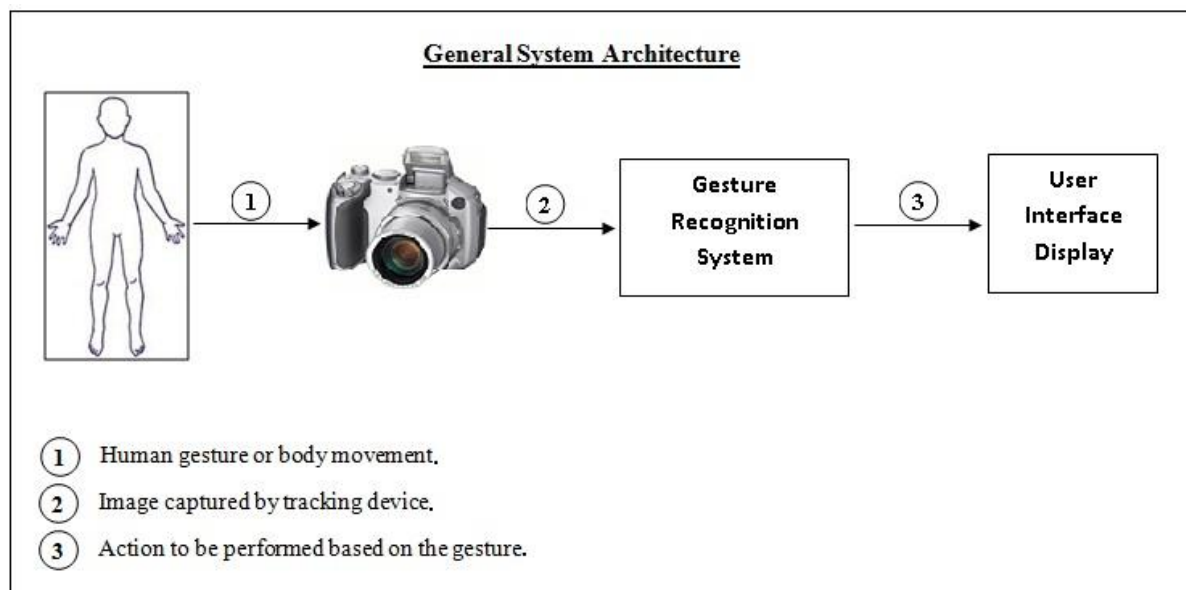


Fig 2.1 General System Architecture

A. Colour Models: Colour is a very powerful descriptor for object detection. Colour information is needed to perform the various image processing techniques on an image. Colour models are used to specify a colour in a standard way. A colour model is a specification of a coordinate system or subspace within that system where each colour is represented by a single point [1]. Three colour spaces that are commonly used are discussed below.

1. RGB (Red Green Blue): Each colour in this colour model appears in its primary spectral components of red, green and blue. Simplicity is the advantage of this colour space but it is perceptually not uniform. It does not separate the luminance and chrominance components of light and also the red, green and blue components are highly correlated.

2. CMYK (Cyan Magenta Yellow Black): Cyan, Magenta and Yellow are the secondary colours of light and primary colours of pigments. When a surface coated with cyan pigment is illuminated with white light, no red light is reflected. This is because cyan subtracts red light from reflected white light. Similarly magenta subtracts green and yellow subtracts blue light respectively. Instead of the muddy looking black obtained by combining cyan, magenta and yellow, colour black was added to CMK colour model to obtain CMYK model.

3. HSI (Hue Saturation Intensity): Humans describe a colour object based on its hue, saturation and brightness. Hue is a colour attribute that describes pure colour. Saturation gives a measure of the degree to which a pure colour is diluted by white light. Brightness is a subjective descriptor that is practically impossible to measure and is one of the key factors in describing colour sensation. This model decouples the intensity component from colour-carrying information in a colour image.

4. CIELAB: CIE $L^*a^*b^*$ is the most complete colour space specified by International Commission on Illumination (CIE). L^* represents the lightness of the colour, a^* represents the position between red/magenta and green and b^* represents position between yellow and blue. This colour space separates a luminance variable from two perceptually uniform chromaticity variables.

B. Image Segmentation: Image segmentation is the process of partitioning a digital image into multiple segments or set of pixels. It is typically used to locate objects and boundaries in images and also for image compression, image editing or database lookup. Each of the pixels in a region is similar with respect to some computed property or characteristic such as colour, intensity or texture. When the object of interest in an application is isolated, segmentation should stop. Segmentation algorithms are generally based on one of two properties of intensity values, discontinuity and similarity. Discontinuity is to partition an image based on sharp changes in intensity and similarity is to partition an image into similar regions according to a set of predefined criteria. The quality of segmentation depends on the clarity of the image captured.

C. Image Histogram: It acts as a graphical representation of tonal distribution in a digital image. It plots the number of pixels in the image with a particular brightness value. Image histograms can be useful for thresholding. The information contained in the graph is a representation of the pixel distribution as a function of tonal variation, hence image histograms can be analyzed for peaks or/and valleys which can then be used for determining a threshold value. This threshold value can then be used for edge detection or image segmentation.

III. GESTURE RECOGNITION TECHNIQUES

In this section we discuss about the four gesture recognition techniques used in four different scenarios. An overview of the methods implemented in each of these techniques is specified and at the end of this section a table listing the various characteristics of each of the techniques is provided for quick reference.

A. Hand Gesture Recognition based on Static Gesture set and HSI and CIELAB colour space: This technique makes use of a system model $S=\{I,G,M,F,O\}$ where, I is the set of input hand gestures, G represents a set of single handed anticipated static gestures, M represents mouse operations, F represents feature vector for G and O represents output with application interface. Static gesture is a specific posture assigned with a meaning. If the feature vector of the input gesture matches with the feature vector of any of the gesture in static gesture set then the gesture has been recognised and corresponding output is given to application interface [1].

To find out the feature vector from input gesture image two methods were used which is based on two different colour spaces HSI and CIELAB [2]. In hand segmentation method using HSI colour space, the values of hue, intensity and saturation were calculated and H-S histogram was created. Gestures were classified based on the histogram values. Skin colour samples needed to be passed to the algorithm for skin colour detection. The drawback of this method was that it was sensitive to even little variation in colour brightness. In the segmentation method using CIELAB colour space, the captured image was first converted into a^* and b^* planes. Then thresholding was done on the converted image. To obtain superior hand shape morphological processing was performed. This method works for skin colour detection but is sensitive of complex backgrounds.

B. A Technique for Real-Time Hand Gesture Recognition: The performance of vision based gesture interaction is prone to be influenced by illumination changes, complicated backgrounds, camera movement and specific user variance. In this technique, to trigger hand detection a specific gesture was designed. Skin colour based hand detection is unreliable for the difficulty to be distinguished from other skin-coloured objects and sensitivity to lightning conditions. Hence extended adaboost method was made use of for classification [3].

Hand tracking was implemented using a multi-modal technique that combines optical flow and colour cue to obtain stable hand tracking. For hand segmentation, a single Gaussian model is used to describe hand colour in HSV colour space. Histogram based method is on the assumption that no other exposed skin colour part of user is in certain area around the hand. Scale-space feature detection method is used for gesture recognition [4]. The method reduces computation expense by detecting multi-scale feature across binary image.

C. Hand Gesture Recognition Using Neural Network Shape Fitting Technique: In this technique hand gesture recognition is based on hand gesture fitting procedure via Self-Growing and Self-organised Neural Gas (SGONG) network [5]. SGONG is an unsupervised neural classifier. The first stage of this technique is hand region detection which is achieved through colour segmentation i.e. classification of the pixels of input image into skin colour and non-skin colour clusters [6]. Next step is to approximate the hand's morphology which is accomplished with the help of SGONG network.

The third stage in this technique is to identify finger. In finger recognition, the number of raised fingers is estimated along with the extraction of hand shape characteristics and finger features. In final phase of recognition process the features distribution is calculated over the training set of images. Then this values are subjected to likelihood based classification and a final classification which is performed by choosing the most probable finger combination[7] [8].

D. Finger Earth Mover's Distance with Commodity Depth Camera Technique for Hand Gesture Recognition: This method uses kinect sensors as input device to capture the colour image and it has a dept map of 640x480 resolutions. To handle the noisy hand shape obtained from the kinect sensor, a novel distance metric for dissimilarity measure called Finger Earth Mover's Distance (FEMD) is used [4]. It matches the fingers and not the whole hand shape. Time series curve is used to record the relative distance between each contour vertex to a centre point. Shape representation is carried out by analyzing the time series curve.

For gesture recognition, template matching is carried out. The input hand is recognised as the class with which it has the minimum dissimilarity distance which is found out using FEMD. Thresholding decomposition method is used for finger recognition [4]. The finger is defined as a segment in time series curve whose height is greater than a particular threshold.

4. Conclusion and Discussion

Four robust techniques for hand gesture recognition were discussed through this paper. The first technique to hand gesture recognition was tested on 70 samples of green colour glove and skin colour [1]. Green colour glove segmentation process is faster than using skin colour due to varying lighting condition. The second method was tested against 2596 frames to assess the performance of the system. Among them 2436 frames were recognized correctly. The average accuracy of the recognition experiment is 0.938[3]. Speed of this method satisfies real time requirements in human computer interface. The hand gesture recognition using a neural network shape fitting technique was tested with 180 test hand images 1800 times. The recognition rate of the system is 0.9045. Mistakes occur due to false feature extraction and false estimation of hand slope Time required for recognition on a 3Ghz CPU is 1.5s [7]. The mean accuracy of the hand gesture recognition based on finger earth mover's distance with a

commodity depth camera system is 0.906. The mean running time is 0.5s [11]. System is robust to orientations changes, because the initial point and the centre point are relatively fixed in each shape. Thus the time-series curves of the hands with different orientations are similar, and

their distances are very small. System is robust to scale changes, because the time-series curve and the FEMD distance are normalized, the hand shapes with scale changes can be correctly recognized as the same gesture. The details are summarized in a tabular form in Table 4.1.

TABLE 4.1
ANALYSIS OF DIFFERENT HAND GESTURE RECOGNITION TECHNIQUES

Method	Colour space	Segmentation technique	Gesture Recognition method	Accuracy	Remarks
Hand Segmentation Technique to Hand Gesture Recognition for Natural Human Computer Interaction	RGB, HSV, CIE-Lab	Anticipated static gesture set, HSL algorithm, HTS algorithm	Edge traversal algorithm		Best results under complex background
A Real-Time Hand Gesture Recognition Method	HSV	Skin collector method	Scale space feature detection	0.938	Speed of the system satisfy real time requirements
Hand Gesture Recognition Using a Neural Network Shape Fitting Technique	YCbCr	SGONG network	Likelihood based classification	0.904	SGONG network converges faster and captures feature space effectively
Hand Gesture Recognition Based On FEMD With A Commodity Depth Camera	RGB, HSV	Time series curve	Template matching using FEMD	0.906	System is robust to scale changes

V. REFERENCE

1. Archana S. Ghotkar & Gajanan K. Kharate, "Hand Segmentation Techniques to Hand Gesture Recognition for Natural Human Computer Interaction", International Journal of Human Computer Interaction (IJHCI), Volume (3) : Issue (1) : 2012
2. E. Stergiopoulou, N. Papamarkos, "A New Technique for Hand Gesture Recognition", IEEE-ICIP, pp., 2006.
3. Yikai Fang, Kongqiao Wang, Jian Cheng And Hanqing Lu, A Real-Time Hand Gesture Recognition Method, ©2007 Ieee
4. Y. Cui and J. Weng, "View-based hand segmentation and hand sequence recognition with complex backgrounds," in Proceedings of 13th ICPR, Vienna, Austria, Aug. 1996, vol. 3, pp.
5. Atsalakis, A., Papamarkos, N., 2005a. Color reduction by using a new self-growing and self-organized neural network. In: VVG05: Second International Conference on Vision, Video and Graphics, Edinburgh, UK, pp. 53–60.
6. Kjeldsen, R., Kender, J., 1996. Finding skin in colour images. In: IEEE Second International Conference on Automated Face and Gesture Recognition, Killington, VT, USA, pp. 184–188.
7. Kohonen, T., 1990. The self-organizing map. Proceedings of IEEE 78 (9), 1464–1480.
8. Kohonen, T., 1997. Self-Organizing Maps, second ed. Springer, Berlin. Licsar, A., Sziranyi, T., 2005. User-adaptive hand gesture recognition system with interactive training. Image and Vision Computing 23 (12), 1102–1114.
9. Huang, C.H., Huang, W.Y., 1998. Sign language recognition using model-based tracking and a 3D Hopfield neural network. Machine Vision and Applications (10), 292–307.
10. Albiol, A., Torres, L., Delp, E., 2001. Optimum color spaces for skin detection. In: IEEE International Conference on Image Processing, Thessaloniki, Greece, pp. 122–124.S.
11. Zhou Ren Junsong Yuan Zhengyou Zhang Robust Hand Gesture Recognition Based on Finger Earth Mover's Distance with a Commodity Depth Camera, IEEE Trans. on PAMI, 30:1 {11, 2008..