International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 2 Issue 7 ǁ July. 2014 ǁ PP.25-32

Gesture Recognition Review: A Survey of Various Gesture Recognition Algorithms

Simerjeet Kaur[1], Ashish Kashyap[2]
[1]*(M.tech , E.C.E, Chandigarh Group of Colleges, Mohali, India)*
[2]*(Asst. Prof., E.C.E, Chandigarh Group of Colleges, Mohali, India)*

Abstract-*This paper presents simple as well as effective methods to realize hand gesture recognition. Gesture recognition is mainly apprehensive on analysing the functionality of human Intelligence. The main aim of gesture detection and recognition is to design an efficient system which is able to recognize particular human gestures and use these detected gestures to transfer information or for controlling devices. Hand gestures enable a vivid complementary modal to communicate with speech for expressing ones thought of idea. The information which is associated with hand gestures detection in a conversation is extent or degree, detection discourse structure, spatial and temporal design structure. Based on the above given points the paper discusses various models of gesture detection and recognition.*

Keywords – *HMM, camshaft, YUV, FSM, Gesture Recognition*

I. Introduction

Hand gestures provide a mean of communication medium for expressing ones thought and ideas. The approaches at present stage can be divided into methods Data-Glove and Vision Based. The Data-Glove methodology uses sensor devices in order to digitize hand and finger motion into multi-parameter interpretation of data. The extra sensory networks make it an easy work, when it comes to data collection, hand configuration and hand movement. However, all the devices involved are quite expensive and bring a dynamic experience to users. On the other side, the Vision Based methods has pre-requisite of only a camera module, thus establishing an interaction between humans and computers, removing the use of any extra needed devices. All these systems tend towards complementing the biological vision system by defining artificial visions that are implemented in software and hardware designs. A challenging problem arises as these systems require being background invariants, light insensitive, person or camera independent to give real time performance.

II. METHODS

2.1 PIXEL BY PIXEL COMPARISON

This method involves pixel by pixel comparison between the frame captured and every image in the database. It may be an easy method to implement but the results may not be accurate. Before implementing this method background of all the images in the database are made uniform by selecting a threshold using Otsu’s method. If ‘s’ is the selected threshold and ‘I’ is the pixel intensity value then s=0 if s is less than I and s=255 if s is greater than I.Now subtraction of each pixel of the captured image is done with corresponding pixels of the entire number of images in the database so as to calculate the Euclidean distance between the images. Lesser the value of the Euclidean distance closer will be the match.

2.2 EDGES METHOD

The objective of this method is to find out in what portion of the image the highest gradient value lays. This will be followed by applying threshold in the gradients so that the good one’s cancel out most the noise in the image background. The magnitude of gradient is given by sum of derivatives in x and y direction respectively.

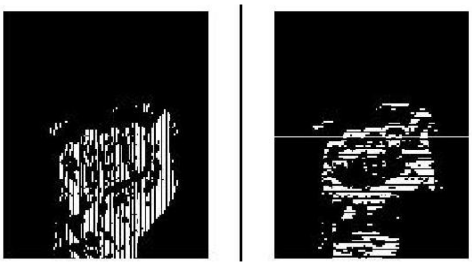

Fig.1. Edge detection using X-Y filter

In order to get good gradient, magnitude filters were used to blur the image and like in the pixel by pixel comparison method background of images are made uniform. The threshold will remove low magnitude gradient. Now we calculate Euclidean distance between vectors all 10 images in database and that of the image captured. With the help of this distance one is able to ascertain whether the gesture is recognized or not.

2.3USING ORIENTATION HISTOGRAM

This method is dependent on feature vector called orientation histogram for pattern recognition. The feature vector forms a histogram based on the edges of the image. The system is trained by giving hand postures as commands. First the image is captured using a webcam then again the image is converted into grey scale image. Next step is to find histogram of the image. In the histogram 360 degrees is grouped into 36 groups with each group of ten degree, for every pixel (x,y) in an image, the gradient of the pixel is given by

$$\mathbf{dx=I(x,y)-I(x+1,y) \text{ and } dy=I(x,y)-I(x,y+1)} \qquad \mathbf{(1)}$$

Gradient direction and gradient magnitude are given by arctan(dx, dy) and sqrt(dx*dx+dy*dy) respectively. If gradient magnitude is greater than threshold, the group of gradient direction is found and the frequency is incremented. Now the histogram is saved as a training pattern. Thereafter for recognition of image the same steps are followed i.e. capture, conversion to grey scale and histogram calculation. Then Euclidean distance is found between the new image captured and the various training patterns .The pattern with least distance is found. The advantage of this method is that it is very fast, robust and translation invariant, whereas the disadvantage being it is rotation dependent.

2.4 THINING METHOD

In order to find the histogram of image the centre of image is taken as reference. Consider 6*6 windows in the middle of the image. This window which is a RGB image is converted into YCbCr. A map of chrominance was created using a training set of 10 images. Cb range was found to be 75-100 and that for Cr was 125-160. Now if Cb and Cr values of pixel belong to the range specified, then pixel are converted to white, or else the pixel is converted into black. The resultant image will be gray scale image. Next step is to convert the image to binary image which is done by automatic selection of threshold using Otsu's method. This Threshold varies as the lighting changes, Binary image is then thinned. While undergoing the procedure of thinning the image, some noise or undesirable segments may crop up this is then required to be removed from image capture to grey scale then to binary finally to thinning and cleaning of the image

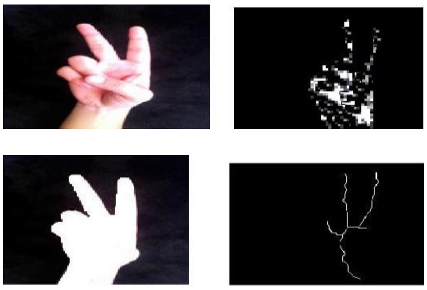

Fig2: Step by step transformation of the image

From the binary image as shown in figure above end points and joint points are found out. End points and joint points are distinguishable by the fact that end points contain only one connectivity neighbour whereas the later contains more than one. If the separation between end points and joint points is less than 15% of the length of the hand, then the segment is considered as noise and is subsequently removed from the hand. Now the feature for gesture recognition will be an angle between lines joining each end points the centre point. Another feature which is used is the distance from end point to the vertical extended line. With the help of these two features raised finger is identified and hence gestures are recognized.

III. HAND DETECTION AND RECOGNITION

3.1 Hidden Markov Models

This method (Hidden Markov Model [1]) computes the dynamic properties of gestures. All gestures are extracted from a video by converting it to frame images and tracking the skin-colour blobs corresponding to the hand into a body– face space on the face of the user information. The aim is to detect two streams of gestures: deictic and symbolic. The image is filtered using a fast look–up indexing table of skin colour pixels in YUV colour space. After filtering, skin colour pixels are gathered into blobs. Blobs are statistical objects based on the location (x,y) and the colour-metry (Y,U,V) of the skin colour pixels in order to determine homogeneous areas. A skin colour pixel belongs to the blob which has the same location and colour-metry component. Deictic gestures are pointing movements towards the left (right) of the body–face space and symbolic gestures are intended to execute commands (grasp, click, rotate) on the left (right) of shoulder.

3.2 YUV Colour Space and CAMSHIFT Algorithm

This method deals with recognition of hand gestures. It is done in the following five steps.

1. First, a digital camera records a video stream of hand gestures.
2. All the frames are taken into consideration and then using YUV colour space skin colour based segmentation is performed. The YUV colour system is employed for separating chrominance and intensity. The symbol Y indicates intensity while UV specifies chrominance components.
3. Now the hand is separated using CAMSHIFT [2] algorithm .Since the hand is the largest connected region, we can segment the hand from the body.
4. After this is done, the position of the hand centroid is calculated in each frame. This is done by first calculating the zeroth and first moments and then using this information the centroid is calculated.
5. Now the different centroid points are joint to form a trajectory .This trajectory shows the path of the hand movement and thus the hand tracking procedure is determined.

3.3 Naïve Bayes' Classifier

This method is an effective and fast method for static hand gesture recognition. This method is based on classifying the different gestures according to geometric-based invariants which are obtained from image data after segmentation; thus, unlike many other recognition methods, this method is not dependent on skin colour. Gestures are extracted from each frame of the video, with a static background. The segmentation is done by dynamic extraction of background pixels according to the histogram of each image. Gestures are classified using a weighted K-Nearest Neighbours Algorithm which is combined with a Naïve Bayes [5] approach to estimate the probability of each gesture type. When this method was tested in the domain of the JAST Human Robot dialog system, it classified more than 93% of the gestures correctly.

This algorithm proceeds in three main steps. The first step is to segment and label the objects of interest and to extract geometric invariants from them. Next, the gestures are classified using a K-nearest neighbour algorithm with distance weighting algorithm (KNNDW) to provide suitable data for a locally weighted Naïve Bayes' classifier. The input vector for this classifier consists of invariants of each region of interest, while the output is the type of the gesture. After the gesture has been classified, the final step is to locate the specific properties of the gesture that are needed for processing in the system—for example, the fingertip for a pointing gesture or the centre of the hand for a holding-out gesture.

IV. Vision Based Hand Gesture Recognition

4.1 3D Hand Model Based Approach

Three dimensional hand model based approaches rely on the 3D kinematic hand model with considerable DOF's, and try to estimate the hand parameters by comparison between the input images and the possible 2D appearance projected by the 3D hand model. Such an approach is ideal for realistic interactions in virtual environments. This approach has several disadvantages that have kept it from real-world use. First, at each frame the initial parameters have to be close to the solution, otherwise the approach is liable to find a suboptimal solution (i.e. local minima). Secondly, the fitting process is also sensitive to noise (e.g. lens aberrations, sensor noise) in the imaging process. Finally, the approach cannot handle the inevitable self-occlusion of the hand.

4.2 Appearance Based Approach

This method use image features to model the visual appearance of the hand and compare these parameters with the extracted image features from the video input. Generally speaking, appearance based approaches have the advantage of real time performance due to the easier 2D image features that are employed. There have been a number of research efforts on appearance based methods in recent years. A straightforward and simple approach that is often utilized is to look for skin coloured regions in the image. Although very popular, this has some drawbacks like skin colour detection is very sensitive to lighting conditions. While practicable and efficient methods exist for skin colour detection under controlled (and known) illumination, the problem of learning a flexible skin model and adapting it over time is challenging. This only works if we assume that no other skin like objects is present in the scene. Another approach is to use the Eigen space for providing an efficient representation of a large set of high-dimensional points using a small set of basis vectors. Based on the observations regarding Hand Detection and Tracking, we can conclude that using YUV Skin Colour Segmentation followed by CAMSHIFT algorithm will help in the effective detection and tracking as the centroid values can easily be obtained by calculating the moments at each point, later we could combine Hidden Markov Training for further applications. It is better when compared to Time-Flight Camera where one has to find the bounding box and then use Iterative Seed Fill algorithm.

4.3 HMM

A time-domain process which demonstrates a Markovian property if the CDF of the current event, given all present and past events, depends singularly on the *j*th of recent event. In case the current event depends on recent past event, then the process is termed a first order Markov pro-cess. This is a useful assumption to make, when considering all the positions and orientations of hands gesture through time space. The HMM [7], [8] are double stochastic processes controlled by: 1) underlying Markov chain having a finite number of states and 2) pair of random functions with each associated to one state. In every discrete time instance, this process has any one of the states and generates observatory points according to the random function to the current state of the process. The probabilities are as follows:

1) transitional, provides the probability for transition;
2) output, a conditional probability of emittinance of an output from finite alphabet of a state.

The HMM is highly optimized and efficient in mathematical structures and has been proved to efficiently model spatio–temporal information in a natural way. The model is termed “hidden” because all that can be seen is only a sequence of observations.

4.4 Condensation Algorithm

This algorithm was developed on the principle of particle filtering. It was originally applied in tracking motion of objects in clutter [5]. Here, one of the gesture models involves an augmented office white-board with which a user can make simple hand gestures to grab regions of the board, print them, save them, etc. In this approach, the authors allow compound models that are very like HMMs, with each state in the HMM being one of the defined trajectory models. The other part deals with human facial expressions, using the estimated parameters of a learned model of mouth motion [9].

4.5 FSMs for Hand Gesture Recognition

As discussed in Section II-C, a gesture can be modelled as an ordered sequence of states in a spatio–temporal configuration space in the FSM approach. This has been used to recognize hand gestures [10], [11], [13].A method to recognize human-hand gestures using a FSM-model-based approach has been used in [11]. The state machine is used to model four qualitatively distinct phases of a generic gesture—static start position (static at least for three frames), smooth motion of the hand and fingers until the end of the gesture, static end position for at least three frames, and smooth motion of the hand back to the start position. The hand gestures are represented as a list of gesture vectors and are matched with the stored gesture vector models based on vector displacements. Another state-based approach to gesture learning and recognition has been presented in [9]. Here, each gesture is defined to be an ordered sequence of states, using spatial clustering and temporal alignment. The spatial information is first learned from a number of training images of the gestures. This information is used to build FSMs corresponding to each gesture is used to recognize gestures from an unknown input image sequence.

V. CONCLUSION

The work was done using static gestures, but with advances in recent technology hand gesture will have to be realized in real time i.e. dynamic gestures. While the prior work done before ours only focused on any one method this system uses three different methods for gesture recognition with which the best method can be found. One significant area of improvement has been the use of Otsu’s based on the observations regarding Hand Detection and Tracking, we can conclude that using YUV Skin Colour Segmentation followed by CAMSHIFT algorithm will help in the effective detection and tracking as the centroid values can easily be obtained by calculating the moments at each point, later we could combine Hidden Markov Training for further applications. It is better when compared to Time-Flight Camera where one has to find the bounding box and then use Iterative Seed Fill algorithm.

REFERENCES

[1] Chih-Ming Fu et.al, “*Hand gesture recognition using a real-time tracking method and hidden Markov models*”, Science Direct – Image and Vision Computing, Volume 21, Issue 8, 1 August 2003, pp.745-758

[2] Vadakkepat, P et.al, “*Multimodal Approach to Human-Face Detection and Tracking*”, Industrial Electronics, IEEE Transactions on Issue Date: March 2008, Volume: 55 Issue:3, pp.1385 - 1393

[3] E. Kollorz, J. Hornegger and A. Barke, “*Gesture recognition with a time-of-flight camera, Dynamic 3D imaging*”, International Journal of Intelligent Systems Technologies and Applications Issue: Volume 5, Number 3-4 2008, pp.334 – 343.

[4] B¨ohme, M., Haker, M., Martinetz, T., and Barth, E. (2007) “*A facial feature tracker for human-computer interaction based on 3D TOF cameras*”, Dynamic 3D Imaging (Workshop in conjunction with DAGM 2007).

[5] L. R. Rabiner, “*A tutorial on hidden Markov models and selected appli-cations in speech recognition*,” *Proc. IEEE,* vol. 77, no. 2, pp. 257–285, Feb. 1989.

[6] J. Yamato, J. Ohya, and K. Ishii, “*Recognizing human action in time-sequential images using hidden Markov model*,” in *Proc. IEEE Int. Conf.Comput. Vis. Pattern Recogn.*, Champaign, IL, 1992, pp. 379–385.

[7] Pujan Ziaie et.al, “*Using a Naïve Bayes Classifier based on K-Nearest Neighbors with Distance Weighting for Static Hand-Gesture Recognition in a Human-Robot Dialog System*” Advances in Computer Science and Engineering Communications in Computer and Information Science, 2009, Volume 6, Part 1, Part 8, pp.308-315.

[8] Pragati Garg et.al, “*Vision Based Hand Gesture Recognition*”, World Academy of Science, Engineering and Technology, pp.1-6 (2009).

[9] M. J. Black and A. D. Jepson, “*A probabilistic framework for matching temporal trajectories: Condensation-based recognition of gestures and ex-pressions*,” in *Proc. 5th Eur. Conf. Comput. Vis.*, vol. 1, 1998, pp. 909–924.

[10] “*CONDENSATION—Conditional density propagation for visual tracking*,” *Int. J. Comput. Vis.*, vol. 1, pp. 5–28, 1998.

[11] J. Davis and M. Shah, “*Visual gesture recognition*,” *Vis., Image SignalProcess.*, vol. 141, pp. 101–106, 1994.

[12] M. Yeasin and S. Chaudhuri, “*Visual understanding of dynamic hand gestures*,” *Pattern Recogn.*, vol. 33, pp. 1805–1817, 2000.

[13] P. Hong, M. Turk, and T. S. Huang, “*Gesture modeling and recognition using finite state machines*,” in *Proc. 4th IEEE Int. Conf. Autom. FaceGesture Recogn.*, Grenoble, France, Mar. 2000, pp. 410–415.

[14] Harshith.C, Karthik.R.Shastry, Manoj Ravindran, M.V.V.N.S Srikanth, Naveen Lakshmikhanth, “*SURVEY ON VARIOUS GESTURE RECOGNITIONTECHNIQUES FOR INTERFACING MACHINES BASED ON AMBIENT INTELLIGENCE*, Department of Information Technology, Amrita Vishwa Vidyapeetham, Coimbatore.