IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Volume-03, Issue-3 E-ISSN: 2347-2693

Vision Based Gesture Recognition System

Priyanka Parvathy Dali[1*] and Hema C.R[2]
[1,2] *Department of Computer Science & Engineering, Karpagam University, Coimbatore, India*

www.ijcseonline.org

Received: Feb/14/2015 Revised: Feb/22/2015 Accepted: Mar/13/2015 Published: Mar/31/2015

***Abstract*—** Lately, the research world has been aggressively exploring techniques and means to develop effective systems for Human Computer Interaction (HCI). Humans interact with computers using different interfaces like Graphical User Interfaces (GUI), Voice User Interfaces (VUI) and Gestures. Among these methods gesture interaction provides a very handy means of dealing with computers. The goal of gesture recognition is to recognize and differentiate the human gestures and make use of these gestures for applications in specific areas. The aim of this paper is to provide a survey on numerous techniques employed for gesture recognition, and to outline preliminary system architecture for implementation.

***Keywords*—**Gesture Recognition, Segmentation, Feature Extraction, Gesture Classification.

I. INTRODUCTION

Gestures are so innate in humans that we tend to use them even when it is not required, like while we are in a telephone conversation. A gesture [1] can be defined as a movement of our body that conveys some information. The human hand is considered to be a very normal mode of communication in human interaction, which can be extended further to be used effectively for communicating with machines as well. Regardless of lots of previous work in this area, the task of building an efficient hand gesture recognition system still remains as a challenging problem.

There have been diverse approaches to handle gesture recognition [2], which includes mathematical models like Hidden Markov Models [3], approaches based on soft computing [4] and use of different imaging and tracking devices or gadgets like data gloves, color gloves [5], kinetic sensors [6], markers etc. Off late vision based gesture recognition systems are being extensively researched and are gaining popularity because of its ease of communication. The performances of existing vision-based approaches [7] are greatly restricted by factors such as the input image quality, lighting conditions and background clutters.

Gesture recognition primarily involves three stages of development: segmentation, feature extraction and gesture classification "Fig 1". The robustness of a system is determined by how well the hand is detected from the whole image and how competent the system is, in rightly identifying the gesture communicated by the hand.

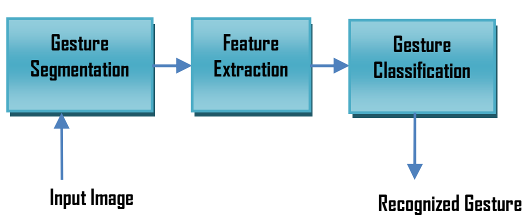

Fig 1: Block diagram for gesture recognition system

II. HAND SEGMENTATION METHODS

Image segmentation is the process of separating an image into numerous regions for the purpose of identifying objects or other significant information [8]. Two main property of the image is considered for this purpose: discontinuity (sharp change in the image intensity) and similarity (grouping together similar values to form regions). In order to recognize manual gestures from video streams, both spatial and temporal gesture segmentation techniques have to be applied. Spatial gesture segmentation technique is employed for the purpose of determining the location of the gesturing hand in each video frame. And in order to establish the start and end positions of the gesture we apply temporal gesture segmentation process [9]. Various segmentation techniques have been used to isolate the hand from the background, some of the common techniques are summarized as follows:

A. Thresholding

Image thresholding is a way of segmenting an image into a foreground and background. This technique isolates objects by converting grayscale images into binary images. It is most effective in images with high levels of contrast. Several thresholding techniques were also widely used for hand segmentation [10] such as static thresholding [11], adaptive thresholding, P-tile method, edge maximizing technique and histogram dependant techniques.

B. Color based Segmentation Techniques

Segmentation in colored images makes use of the skin color information to detect the hand and other body parts from the image [12]. Many approaches where put forward for segmentation based on skin color modeling such as Gaussian mixture model [13] (statistical method), HSV color space model [7] [13] and histogram based model (non-parametric statistical model) [14] [15]. These techniques are rotation and scaling invariant as well as easy to implement.

Other segmentation techniques include optical flow method [16], Otsu segmentation [31][17] (Kulkarni & S.D.Lokhande, 2010), Video Object Planes [18] and Blob detection [3] [19].

III. FEATURE EXTRACTION TECHNIQUES

Feature extraction is the process of reducing the amount of information required to describe a large set of data. The input data to an algorithm can contain redundant information that can be removed, so as to obtain a reduced feature set which would represent the original data set. For hand gestures, the extracted features can be the centroid of the hand, finger shapes, palm locations, finger counts, position of the hand etc.

A. Finger Tracking and Contour Detection

If we utilize a model that uses finger trajectories, then we require the finger tips [11]to be extracted first. Here the whole image is used as the feature and this is a common practice in appearance based models where image sequences are used for temporal modeling [20]. Dung and Mizukawa [21] presents a new stable method to extract the hand centre, the directions and the finger tip positions. Connected component labeling (CCL), distance transform (DT) and Hough transform (HT) are used as features to represent the hand. The contour [11]of the hand image is another feature that can be used to represent the hand, where edge detection algorithms [3] can be employed. Certain studies also use the Kalman filter [22] to track the finger tip trajectories from frame to frame.

B. Principal Component Analysis

Principal Component Analysis (PCA) [23]is another popular method used for extracting the features of the hand. It is an eigen vector based method which explains the variance of the data in the image. Gopalan & Behzad [24] employs PCA to estimate the orientation direction of the hand.

Some other methods which are not so commonly employed have but been studied and implemented for feature extraction includes Discrete Fourier transform and 7 Hu moments [7], Wavelet Transform [25] and Mel frequency Ceptral Coefficients [26].

IV. GESTURE CLASSIFICATION

Gesture classification is the process in which the image features are identified and the gesture will be recognized. Several classifiers are used for this purpose.

A. Hidden Markov Model

The Hidden Markov Model (HMM) can be used to efficiently model spatio–temporal information in a normal way. The process involves ignoring the spatial component of the dynamic gesture while the temporal component is represented through an HMM classifier. Then the classifier is trained using a set of data and finally verified using test data. The HMM is popular for its high recognition rates. Researchers have also made use of HMM merged with other classifiers, as in [3] where Kalman Filtering and Graph Matching is combined with HMM for efficient recognition.

B. Finite State Machine

In the Finite State Machine (FSM) [18] approach, a gesture is represented as an ordered sequence of states. The training of the model is done off-line using many samples of each gesture as training data, and this outputs the parameters of each state. The trained FSM can be used for recognizing gestures online. When a gesture is inputted into the FSM, if it is successful in reaching the final state then it is understood that the gesture has been recognized.

C. Support Vector Machine

Positioning Support Vector Machines (SVM) are supervised learning models which are controlled by learning algorithms that analyze data and recognize patterns that can be used for gesture classification. If we have two classes and a set of points that belong to either of these classes then a linear SVM finds the hyper plane which separates the points in such a way that points of the same class fall on the same side. And at the same time the distance of the classes from the hyper plane will be maximized.SVM has been used in its basic form [24] or in other forms such as Multiclass SVM.

D. Artifical Neural Networks

Neural Network is another classifier which is widely used and gaining popularity because of its self-learning ability on-line. Neural Networks are composed of small units called neurons which exhibit properties similar to the biological neuron. These networks have to be trained with training samples so as to improve their self learning. Here the error ratio decreases with increase in training sample sets [32]. Some of these networks are K-mean based radial basis function neural network [17], back propagation neural network [27] [25] [20] [28] [16] and Multi layer Perceptron Networks (MLP).

Other methods that have been used for gesture classification are Kalman filter [22], PCA [29], Nearest Neighborhood method [8] and minimum distance classifiers.

The overall review summary is presented in the table below "Table 1"

Method [Ref No.]	Segmentation Technique	Feature Extraction Technique	Gesture Classifier	Accuracy
[19]	HSV color space,		PCA	95%
[18]	Video Object Planes	Hand Centroid	FSM	97%
[24]	Skin Color Segmentation	PCA +IDSC descriptor	SVM	85%
[15]	Histogram		Kalman Filter	87% 93%
[3]	Contour detection + Blob Analysis		Kalman Filter + HMM + Graph Matching	
[16]	Optical flow		Back propagation NN	83.3%
[28]	Contour detection	Finger tip	Back propagation NN	95%
[20]		Temporal Contour	Back propagation NN	93%
[25]	Background subtraction	Wavelet Transform	Back propagation NN	97%
[27]	Thresholding + Edge detection (Sobel operator)	Hand contour	Back propagation NN	86.38%
[17]	Ostu Segmentation	Localized contour	Radial Basis Function NN	99.6%
[8]	HSV color space	Discrete Fourier Transform + 7Hu Moments	Nearest Neighbourhood	96%
[11]	Thresholding	Contour detection Finger count		

Table 1: Comparison of HGR systems

V. CHALLENGES IN GESTURE RECOGNITION SYSTEMS

Illumination conditions - Changes in the lighting condition can cause false detection of hand skin region.
Background problem- A cluttered or complex background can result in poor segmentation of the hand.
Rotation problem – The hand region can be rotated in any direction in the image and this can prevent the gesture being identified properly.
Scale Invariant problem- Gesture images captured from different users at different distances can result in the hand being in varied sizes. This could cause difficulty in identifying the gestures.
Translation problem- This problem occurs due to change in in hand position in the image.

VI. PRELIMINARY SYSTEM ARCHITECTURE

Based on the study data, we present a preliminary architecture "Fig 2" for a hand gesture recognition system that will be used for giving instructions to a robot. MATLAB R2012a is the tool selected for processing.

Acquisition: Video stream is captured at a rate of 24 frames per second. For uniformity, few assumptions are made: the background is kept static (light background), images are captured under natural lighting and the hand gesture has to be held for few seconds in front of the camera. Our intention is to spot the frame that contains the gesture which is to be recognized. A similar method to [19] is employed for this purpose, 4 frames are captured every second and the total pixels of mismatch between these frames are computed. If the mismatch greater than zero and less than a threshold value then the system accepts it as a gesture to be processed.

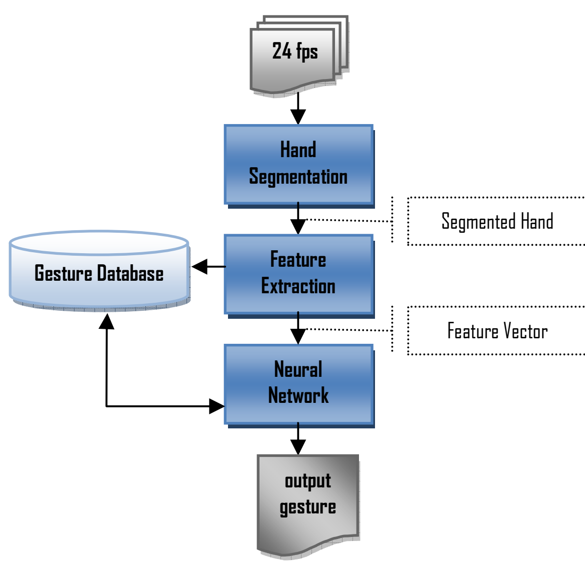

Fig 2: Preliminary system architecture

Segmentation: This phase aims at segmenting all unwanted data from the image and obtaining only the hand. Firstly the captured images are subjected to background subtraction so as to isolate only the hand. But this could also result in isolating any other new object other than the hand that appears in the frame. To improve recognition of hand gestures, the captured frames are converted from RGB to HSV color space subjected to skin color detection. This process introduces some noise and it is removed by application of a Gaussian filter. It is then further converted to binary by application of global thresholding technique.

Feature Extraction: The feature extraction phase is responsible for converting the hand gesture into a numerical representation of the original image. The extracted feature matrix should be in such a way that it should optimize the performance of the Neural Network classifier in the next phase.

Classifier: Radial Basis Function Neural Network has been selected as the classifier because of its self learning ability online. The output of the network will have 10 output nodes for the 10 selected gestures. The input nodes depend on the feature vector obtained in the feature extraction stage.

VII. DISCUSSION AND CONCLUSION

Effectiveness of a gesture recognition system depends on the right selection of methods that will be used in the three different stages. This choice of appropriate methods is highly dependent on the environmental conditions and system requirements. On studying and comparing various techniques we have picked out methods that would bring out an efficient system for recognizing hand gestures. The tentative system architecture has been outlined. Back propagation feed forward Neural Network was the choice for classifier because of its online self learning ability and high recognition accuracy.

REFERENCES

[1] Wikipedia, "Gesture recognition," [Online]. Available: http://en.m.wikipedia.org/wiki/Gesture_recognition.

[2] 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, pp. 677 - 695, 1997.

[3] M. C. A. Burande, P. R. M. Tugnayat and Prof.Dr. Nitin K. Choudhary, "Advanced Recognition Techniques for Human Computer Interaction," *IEEE,* vol. 2, pp. 480-483, 2010.

[4] S.Mitra and T.Acharya, "Data Mining:Multimedia, Soft Computing and Bioinformatics".

[5] R. Y.Wang and J. Popovic, "Real-time Hand Tracking with a Color Glove," *ACM Transactions on Graphics,* 2009.

[6] Z. Ren, J. Meng, J. Yuan and Z. Zhang, "Robust Hand Gesture Recognition with Kinect Sensor," *ACM,* 2011.

[7] A. S. Ghotkar and G. K. Kharate, "Vision based Real Time Hand Gesture Recognition Techniques for Human Computer Interaction," *International Journal of Computer Applications,* vol. 70, no. 16, pp. 1-6, 2013.

[8] A. S. Ghotkar and G. K. Kharate, "Hand Segmentation Techniques to Hand Gesture Recognition for Natural Human Computer Interaction," *International Journal of Human Computer Interaction,* vol. 3, no. 1, pp. 15-25, 2012.

[9] J. Alon, V. Athitsos, Q. Yuan and S. Sclaroff, "A Unified Framework for Gesture Recognition and Spatiotemporal Gesture Segmentation," *IEEE Transactions of Pattern Analysis and Machine Intelligence,* 2009.

[10] S. S. Al-amri, N. Kalyankar and K. S.D, "Image Segmentation by Using Thershold Techniques," *JOURNAL OF COMPUTING,* vol. 2, no. 5, pp. 83-86, 2010.

[11] A. Dhawan and V. Honrao, "Implementation of Hand Detection based Techniques for Human Computer Interaction Human Computer Interaction," *International Journal of Computer Applications,* vol. 72, no. 17, pp. 6-13, 2013.

[12] N. A. Ibraheem, R. Z. Khan and M. M. Hasan, "Comparative Study of Skin Color based Segmentation Techniques," *International Journal of Applied Information Systems,* vol. 5, no. 10, pp. 25-38, 2013.

[13] C. M. Sharma and S. Saxena, "A Context-aware Approach for Detecting Skin Colored Pixels in Images," *International Journal of Computer Applications,* vol. 71, no. 17, pp. 8-13, 2013.

[14] P. Campadelli, F. Cusmai and R. Lanzarotti, "A Color-Based Method For Face Detection".

[15] Y. Sriboonruang, P. Kumhom and K. Chamnongthai, "Visual Hand Gesture Interface for Computer Board Game Control," *IEEE,* 2006.

[16] J. Zhang, F. Zhang and M. Ito, "Image processing based remote control with robot arm simulator," *SICE,* pp. 2344 - 2348, 2009.

[17] D. K. Ghosh and S. Ari, "A Static Hand Gesture Recognition Algorithm Using K-Mean Radial Basis Function Neural Network," *IEEE,* 2011.

[18] M. Bhuyan, D. Ghosh and P. Bora, "Designing of Human Computer Interactive Platform for Robotic Applications".

[19] J. L. Raheja, R. Shyam and P. B. P. Umesh Kumar, "Real-time Robotic Hand Control Using Hand Gestures," *IEEE,* pp. 12-16, 2010.

[20] P.V.V.Kishore and P. Kumar, "Segment, Track, Extract, Recognize and Convert Sign Language Videos to Voice/Text," *International Journal of Advanced Computer Science and Applications,* vol. 3, no. 6, pp. 35-47, 2012.

[21] L. Dung and M. Mizukawa, "Fast Hand Feature Extraction Based on Connected Component Labeling, Distance Transform and Hough Transform," *Journal of Robotics and Mechatronics,* vol. 21, no. 6, 2009.

[22] Y. Sriboonruang, P. Kumhom and K. Chamnongthai, "Visual Hand Gesture Interface for Computer Board," *IEEE,* 2006.

[23] T.-N. Nguyen, H.-H. Huynh and J. Meunier, "Static Hand Gesture Recognition Using PCA Combined with ANN," *Journal of Automation and Control Engineering,* vol. 3, no. 1, 2015.

[24] R. Gopalan and B. Dariush, "Toward a vision based hand

gesture interface for robotic grasping," *IEEE,* pp. 1452-1459, 2009.

[25] H. A. Jalab, "Static Hand Gesture Recognition for Human Computer Interaction," *Information Technology Journal,* pp. 1265 - 1271, 2012.

[26] S. Gupta, J. Jaafar, W. F. w. Ahmad and A. Bansal, "Feature Extraction Using MFCC," *Signal & Image Processing : An International Journal,* vol. 4, no. 4, pp. 101-108, 2013.

[27] H. Hasan and S.Abdul-Kareem, "Static Hand Gesture Recognition using Neural Networks," *Springer,* 2012.

[28] H. Q. J. G. Jinwen Wei and Y. C. Jinwen Wei, "The Hand Shape Recognition of Human Computer Interaction with Artificial Neural Network," *IEEE,* p. 3809, 2009.

[29] J. L. Raheja, R. Shyam, U. Kumar and P. B. Prasad, "Real-Time Robotic Hand Control using Hand Gestures," *Second International Conference on Machine Learning and Computing- IEEE,* pp. 13-16, 2010.

[30] P. P. Dali and H. C.R, "Interpretation of Hand Gestures Using Neural Networks:A Review," *International Journal of Computer Science and Information Technologies,* pp. 1111-1114, 2015.

[31] V. S. Kulkarni and S.D.Lokhande, "Appearance Based Recognition of American Sign Language Using Gesture Segmentation," *International Journal on Computer Science and Engineering,* pp. 560-565, 2010.

[32] P. P. D and H. C.R, "Interpretation of Hand Gestures Using Neural Networks:A Review," *International Journal of Computer Science and Information Technologies,,* vol. 6, no. 2, pp. 1111-1114, 2015.

