ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

Improved Hand Gesture Recognition with SVM, HOG and HMM

Gopinadh Sasubilli[1], Uday Shankar Sekhar[2]

Research Scholars

[1]gopinadh_sasubilli@bcbstx.com, [2]uday.sekhar@ey.com

ABSTRACT: This paper aims to improve Human Computer Interaction (HCI) using "Hand Gesture Recognition". Human Computer Interaction (HCI) is the boon to new advancements in the ease of using computer and various devices. Use of body parts like hand, face expressions, etc gives a new and an easy way to interact with the computer. Making a gesture using hand is one of the simplest ways to give a command to a device to perform a specific task. Hand gesture recognition has many applications such as Sign Language Recognition, Communication in Video Conference, using a finger as a pointer for selecting options from menu and Interacting with a computer by easy way for children or also in gesture based games.

KEYWORDS: Hand Gesture Recognition, Histogram of Oriented Gradients (HOG), Support Vector Machine (SVM), Hidden Markov Model.

I. INTRODUCTION

Gesture recognition is based on three main stages: preprocessing, feature extraction and classification. In preprocessing stage, skin detection and segmentation is carried out. Color information is used to detect both hands and face in connection with morphological operation. After the detection of the skin parts, hand will be classified among all parts. We create a Hand Detector using Histogram of Oriented Gradients (HOG) features. Then tracking will take place in further step in order to determine the motion trajectory so-called gesture path. The second stage, feature extraction, enhances the gesture path which gives us a pure path and also determines the orientation between the center of gravity and each point in a pure path. Thereby, the orientation is quantized to give a discrete vector (DV) that used as input to HMM. In the final stage, the gesture is recognized by using Left-Right Banded model (LRB) in conjunction with Baum-Welch algorithm (BW) for training the parameters of HMM.

In our experiment, 45 trained gestures are used for training and also 15 tested gestures for testing. Our method recognizes the various complex gestures and achieves an average recognition rate of 90.3%.

II. RELATED WORK

The use of hand images for the gesture recognition has been gaining popularity. Various works have been done by different researchers. In this section we will review the work done by various researchers on hand gesture recognition

Chang-Yi Kao. et. al.[3] presented a hand gesture recognition technique based on path of hand motion by using HMM as a classifier. In their approach eight different kinds of gestures have been developed using either single hand or both hands. In the proposed methodology first of all face localization is done and then maximum circle plate mapping is used to locate palm of the hand from skin region and in feature extraction procedure orientation is used as a main feature and the experiment shows 96% accuracy with HMM recognition model.

Amit Gupta, et. al. [4] introduced hand gesture recognition technique which utilize an FPGA based smart camera for gesture analysis. The experiment is performed using shape based features as the system is modeled on FGPA board and shape based features consume less FGPA area and requires less effort for computation. The features for used for

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

experiment are - area of hand, perimeter of hand, thumb detection and radial profile and angular position. For image processing and hand segmentation, illumination compensation technique and skin color segmentation model is utilized keeping in view the changing background. For performing the experiment images of 25 different persons is captured and is found that system is able to recognize 10 different gestures with accuracy 94.40%.

Jing Lin, et. al. [5] proposed a technique based on histrograms of oriented gradients (HOG) in order to remove the hindrance caused by cluttered background during hand localization and modeling. Histrograms of oriented gradients and SVM are applied to localize the hand and then motion trajectory features of temporal gestures are extracted and standard database is created. For recognition Mahalanobis distance is used. The technique is tested on 6 standard gestures and average accuracy of 91.7% is obtained but this system is able to work for complicated gestures.

Zhou Ren, et. al. [6] developed a robust part based technique using kinect sensor keeping in view the limitations of glove based and vision based techniques. In their approach, first of all kinect sensor is used to capture the both color images and depth maps corresponding to that image. Using depth maps, hand can be easily detected even in cluttered background also by using depth thresholding. After detecting the hand it is represented by its finger parts using time series curve. Then for gesture recognition a dissimilarity measure called Finger-Earth Mover's Distance (FEMD) is proposed which can recognize noisy hand contours as compared to other recognition methods and is robust to change in scale, orientation, local distortions and background conditions. For experiments dataset consisting of 10 person, 10 gestures and 10 cases/gestures id used and achieves accuracy of 93.2% and system is tested on two real time applications.

III. PROPOSED WORK

Our method is designed to classify the gesture path that is generated from a single hand motion by using HMM. In particular, the alphabet gesture method consists of three main stages:-

(1) Preprocessing, localize and track the hand to generate its motion trajectory (gesture path).

(2) Feature extraction, enhance the gesture path and then determine the discrete vector by quantization of the orientation.

(3) Classification, the gesture hand graphical is recognized by using discrete vector and Left-Right Banded (LRB) model with 6 states.

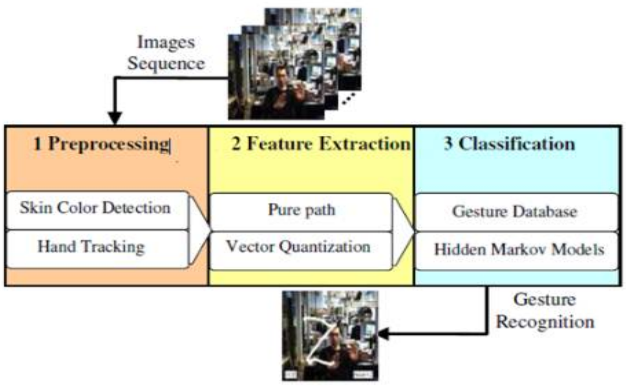

Fig.1: Gesture Recognition Stages

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

IV. PREPROCESSING STAGE

Preprocessing is the first stage in our method which contains two steps. The first step is "skin color detection"and the second step is "hand localization and tracking".

Skin Color Detection **-** In images and videos, skin color is an indication of the existence of humans in such media. Skin detection means detecting image pixels and regions that contain skin-tone color.The appearance of skin in an image depends on the illumination conditions where the image was captured. Therefore, an important challenge in skin detection is to represent the color in a way that is invariant or at least insensitive to changes in illumination. So we tried using various color spaces for skin detection and found the result to be good in HSV and YCbCr color space.

V. FEATURE EXTRACTION USING HISTOGRAM OF ORIENTED GRADIENTS (HOG)

Histogram of oriented gradients (HOG)

The histogram of oriented gradients (HOG) is a feature descriptor used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image. This method is similar to that of edge orientation histograms, scale-invariant feature transform descriptors, and shape contexts, but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy.
HOG descriptor has a few key advantages over other descriptors. Since it operates on local cells, it is invariant to geometric and photometric transformations, except for object orientation. Such changes would only appear in larger spatial regions. The HOG descriptor is thus particularly suited for human detection in images. But we will implement HOG to detect the hand features.

VI. SVM CLASSIFIER

The final step in object recognition using histogram of oriented Gradient descriptors is to feed the descriptors into some recognition system based on supervised learning. The support vector machine (SVM) classifier is a binary classifier which looks for an optimal hyperplane as a decision function. Once trained on images containing some particular object, the SVM classifier can make decisions regarding the presence of an object, such as a human, in additional test images. After the skin segmentation of a frame (image) is complete, a bounding box is created onto the skin patches. Only those skin patches having area greater than the threshold value are sent to the SVM for classification. This threshold value of area depends on the type of camera used and its pixel density. Each of the skin patch is resized to a particular value and then there HOG features are extracted. This is done to make the Feature vector of the each patch equal in size so that it can be compared to the standard database of the SVM to classify each patch as hand or non-hand. Our project is based on the assumption that only single hand is used for making gestures. So as soon as any skin patch is classified as hand than recognition part is complete and tracking of the hand starts.

Data Set- We used a data set where 870 positive hand images were used and 800 negative hand images.

The data set was partially created by us and some images were taken from various sources like face dataset used by Voila Jones and from the help of Google images.

VII. FEATURE EXTRACTION STAGE

The feature extraction is a very important part in our method to recognize the alphabet gesture path. There are three basic features as location, orientation and velocity. The previous researches showed that the using of orientation feature is the best in terms of results. So, we will rely upon it in our method. A gesture path is spatio-temporal pattern which consists of centroid points (Xhand; Yhand). The gesture path may be containing unmovable points notably at the first and end of gesture path, so we enhance the gesture path to obtain a pure path as follows

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

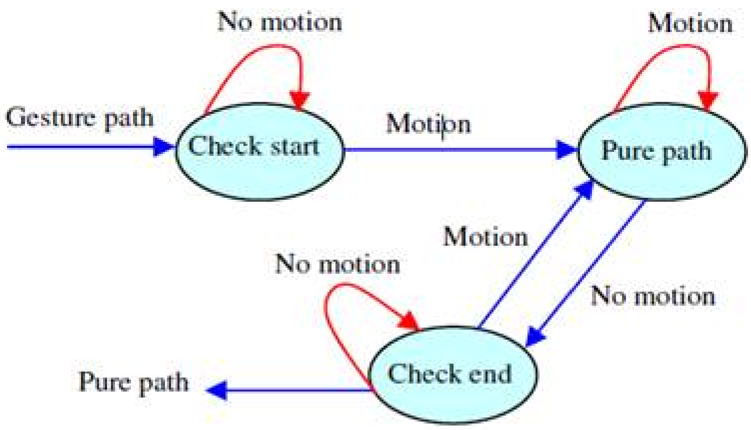

Fig.2: Gesture Path Enhancement

- Firstly, the gesture path is entered to a check start state to see, if there is no motion then it takes the next point, else the pure path is being generated. So a minimum velocity Vstart is used such that if velocity of hand detected is less than the Vstart than the gesture is not considered as started. The velocity from one frame to next frame is also dependent on the frame rate.
- Secondly, the pure path state continues to generate the pure path while the points of gesture path input moveat a velocity greater than Vstop. When the velocity is less than Vstop, we call to the check end state and the gesture is ended. Both Vstart and Vstop were calculated by performing repeated experiments for various gestures to obtain the optimum results.
- Finally, the pure path is generated from the check end state when the input gesture path ended (i.e. there is no point). Also, while there are points in gesture path, this state perform a check to see if the point not moves then delete it and takes the next point, else return to the pure path state.

Once the hand is detected in the video, to increase the speed we localize the search area from the next frame to the bounding box region of the hand detected in the previous frame. If sometimes hand is not detected in one of the frames so from the next frame we increase the search area using an error term ‘e’. In contrast to the enhancement gesture path to obtain a pure path, our method is based on the angle (orientation) as a basic feature. Therefore, the orientation is based on (Xc; Yc) where Xc and Yc are the center of gravity for pure path and are determined as given below. Since the location of pure path for the same gesture according to the start point is different, we calculate the orientation between any point in pure path and the center of gravity as shown

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

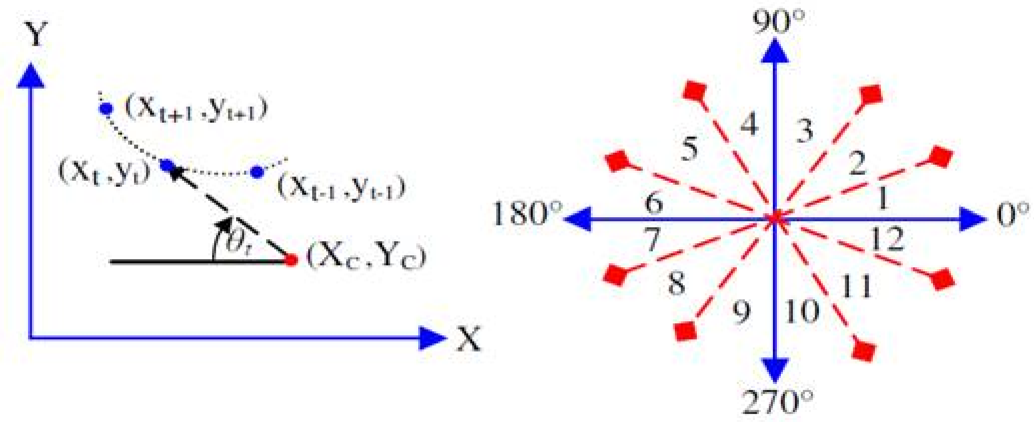

Fig.3 Orientation and Vector Quantization Range

VIII. CLASSIFICATION

Classification is the final stage in our method. Throughout this step, the pure path of hand graphical is recognized by using Left-Right Banded model with 6 states and building gesture database. The gesture database includes 15 paths for each gesture where these paths are determined from the training video that is captured for each gesture. According to this stage, Baum-Welch algorithm (BW) is used for training the initialized parameters of HMM to provide the trained parameters. Moreover, the trained parameters and discrete vector are used as input to Viterbi algorithm in order to obtain the best path. By this best path and gesture database, the pure path is recognized by selecting the gesture which best matches the path in the gesture database videos. The following subsections describe this stage in details.

1. Hidden Markov Models: Markov model is a mathematical model of stochastic process where these processes generate a random sequence of outcomes according to certain probabilities. An HMM is a triple represented as:

- The set of states $S = s_1; s_2; \ldots.; s_N$ where N is the number of states.
- An initial probability for each state π_i, i=1, 2, ...,N such that $\pi_i = P(s_i)$ at the initial step.
- An N-by-N transition matrix $A = \{a_{ij}\}$ where a_{ij} is the probability of a transition from state S_i to S_j; 1<=i; j<=Nand the sum of the entries in each row of matrix A must be 1 because this is the sum of the probabilities.
- The set of possible emission (an observation) $O = \{o_1; o_2; :::; o_T\}$ where T is the length of pure path.
- An N-by-M observation matrix $B = \{b_{im}\}$ where b_{im} gives the probability of emitting symbol v_m from state s_i and the sum of the entries in each row of matrix B must be 1 for the same pervious reason. M represents the number of discrete symbols.

There are three main problems for HMM: Evaluation, Decoding and Training that can be solved by using Forward-Backward algorithm, Viterbi algorithm and Baum-Welch algorithm respectively. Left-Right Banded (LRB) model, where each state can go back to itself or the following state only.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

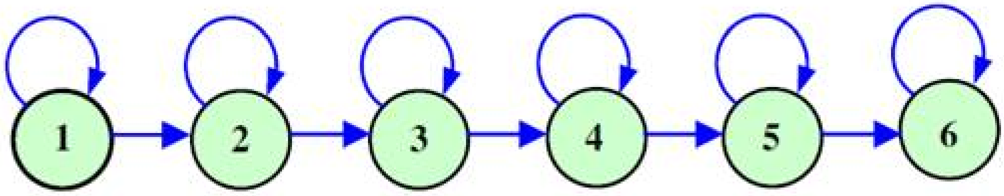

Fig. 4 Lift-Right Banded Model with 6 states

2) Initializing parameters for LRB model**:** LRB model is restricted and simple for training data that will be able to match the data to the model. Moreover, In addition, the discrete vector contains a single sequence of codebook from 1 to 12 in our method. The number of states is found by considering the number of segmented part that is contained in graphical pattern when we represent it. For example, ”L” graphical pattern contains two segmented part, thus we need only 2 states for it while ”G” pattern needs 5 states and 6 states for ”E” pattern. Thereby, the number of states is 6 states nearly for all alphabets. Also when we want to implement normal gestures like Move up, Move Down, Move Right and Move Left, then they require only 2 states.

3) Baum-Welch and Viterbi Algorithm: After the HMM parameters are initialized, we use Baum-Welch algorithm to perform the training for the initialized parameters where the inputs of this algorithm are discrete vector (DV) (obtained from feature extraction stage) and the initialized parameters (A, B). This algorithm gives us the new parameters estimation of vector pi, matrix A (A new) and matrix B (B new).

IX. RESULTS

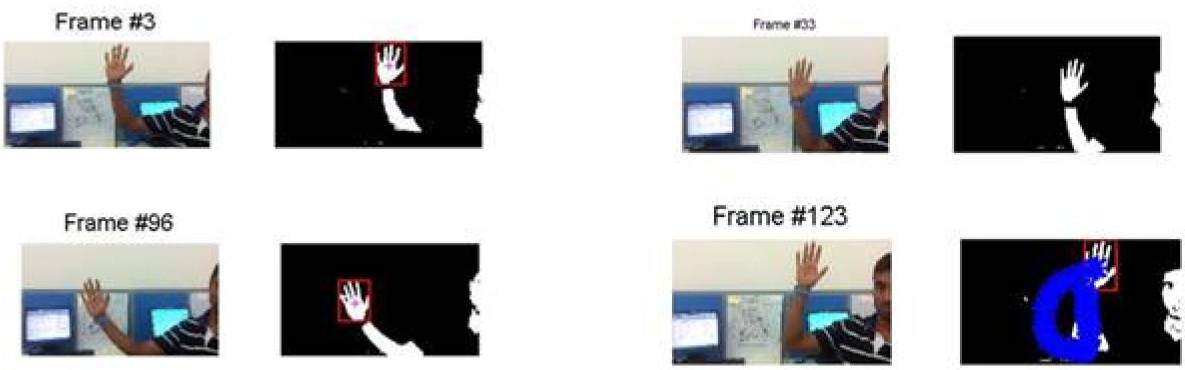

Fig.5: Outputs

The fig.5 depicts the output images of different frames used to recognize the character “D”. It clearly shows the detection of gesture in every frame and the last frame shows the completed output as “D”.

X. SUMMARY AND CONCLUSION

This paper describes whole method of how we were able to recognize gestures using the motion trajectory of a single hand using HMM. We used our own gesture database and was successful in creating database for 3 gestures namely L, D, N. The method consists of three main stages. The first stage is the preprocessing where the hand is localized and tracked to produce the gesture path. In the second stage so-called feature extraction, the discrete vector is obtained by

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

quantization of the orientation where this vector is used as input to HMM. The final stage is the classification which can be able to recognize the hand gesture by using BW algorithm, Viterbi algorithm and LRB model. Our results show that, an average recognition rate of 85% for testing videos. Where 45 trained and 20 tested gestures are used.

XI. FUTURE WORK

The method in this report made use of various assumptions and the future work can be done to eliminate that assumption. The assumptions made and the work than can be done on that is as follows:-

1. The skin color values are only for Asian/Caucasian skin. Solution: **-** We can first make use of a face detection algorithm like Voila Jones to detect the face area followed by noting the pixel values of the skin of face and thereby implementing a dynamic Skin segmentation. (See references)
2. The background in the image should not contain any skin colored objects.
3. There is no occlusion of hand and face while making the gestures. Solution: **-** Using the stereo vision cameras or other techniques, find out the depth map of the image. Depth map is an image that contains information relating to the distance of the surfaces of scene objects from a viewpoint. With the help of this we can easily solve the hand-face occlusion. (See references)
4. The hand movement speed while making the gestures is dependent on the Frame Rate of the camera used to capture gesture.
5. Only one hand is used for gesture and the other hand should not appear in the gesture.
6. The gesture maker should wear full sleeves shirt so that only palm area is visible Solution:**-** Arm removal techniques can be used .One such technique is implemented using extended HOG i.e. SCHOG(Skin Color HOG)(See references).Also multi stage HOG can be implemented to remove arm in addition with face , fist ,complex background and simple background(See references).

In future our research focuses on enhancement of color detection using Gaussian Mixture Model (GMM) and the motion trajectory will be determined by a fingertip instead of the centroid point for the hand region. Both hands will be used for gesture purpose. Palm area will be separated from hand and tracking will be done.

REFERENCES

1. A Survey on Pixel-Based Skin Color Detection Techniques (Vladimir Vezhnevets,Vassili Sazonov, Alla Andreeva)
2. Skin segmentation using color pixel classification: analysis and comparison (Son Lam Phung, A. Bouzerdoum, D. Chai)
3. Skin Detection - a Short Tutorial(Ahmed Elgammal, Crystal Muang and Dunxu Hu)
4. Histograms of Oriented Gradients for Human Detection(Navneet Dalal and Bill Triggs)
5. Real-Time Hand Detection Based on Multi-Stage HOG-SVM Classifier(Jiang Guo ,Jun Cheng,Jianxin Pang,Yu Guo)
6. An extended HOG model: SCHOG for human hand detection(Xingbao Meng, Jing Lin, Yingchun Ding)
7. Computer vision for pedestrian detection using Histograms of Oriented Gradients(Thesis- Jose Marcos Rodriguez Fernandez)
8. MIT Courseware:-https://www.youtube.com/watch?v=_PwhiWxHK8o
9. A temporal hand gesture recognition system based on hog and motion trajectory(Jing Lin, Yingchun Ding) 2013
10. Hand gesture recognition using Histogram of Oriented Gradients and Partial Least Squares Regression(Arindam Misra, Abe Takashi,) 2011
11. Gesture Recognition for Alphabets from Hand Motion Trajectory Using HiddenMarkov Models (Mahmoud Elmezain, Ayoub Al-Hamadi, Gerald Krell, Sherif El-Etriby, Bernd Michaelis) ISSPIT 2007
12. Low Cost Hand Gesture Learning and Recognition System Based on Hidden Markov Model(Jinjun Rao, Tongyue Gao, Zhenbang Gong and Zhen Jiang)
13. A Novel Approach for Depth Extraction from Single 2D Image(Ranran Feng Jiahuang Ji)
14. www.images.google.com
15. www.google.co.in
16. www.wikipedia.org
17. INRIA Dataset
18. Voila Jones Dataset for face recognition