Vision Based Gesture Recognition for Alphabetical Hand Gestures Using the SVM Classifier

Aseema Sultana [1]
[1]Student, M. Tech (CSE), MVJ College of Engineering
Channasandra, Near ITPL, Bangalore-67, India
aseemasultana008@gmail.com

T Rajapuspha [2]
[2]Associate Professor, Department of CSE, MVJ College of Engineering
Channasandra, Near ITPL, Bangalore-67, India
puspha.rambabu@gmail.com

Abstract— **In this paper we discuss how a Gesture Recognition system for Alphabetical Hand Gestures is built. The main motive was to develop a system that can simplify the way humans interact with computers. The system is designed using the Support Vector Machine (SVM) Classifier which is widely used for classification and regression testing. SVM training algorithm builds a model that predicts whether a new example falls into one category or other. And the SVM classifier learns from the data points in examples when they are classified belonging to their respective categories**.
Index Terms— **Computer Vision based gesture recognition, SVM classifier**.

I. INTRODUCTION

A gesture is a motion of the body that contains information. A gesture may thus be defined as a physical movement of the hands, arms, face, and body with the intent to convey information or meaning. Gesture recognition, then, consists not only of the tracking of human movement, but also the interpretation of that movement as semantically meaningful commands. Gesture recognition can be seen as a way for computers to begin to understand human body language, as in [4], thus building a richer bridge between machines and humans than primitive text user interfaces or even GUIs (graphical user interfaces), which still limit the majority of input to keyboard and mouse. Gesture recognition enables humans to interface with the machine (HMI) and interact naturally without any mechanical devices, example of a journal article in [5]-[6].

Gesture recognition can be conducted with techniques from computer vision and image processing. Vision-based hand gesture recognition has three basic processing stages including Hand Segmentation, Gesture Modeling, and finally Gesture Classification, as in [3]. The main procedure of hand segmentation is to detect hand regions in the sequence of hand gesture and separate them from backgrounds. It should be mentioned that the correctness of hand gesture recognition has a close relationship to the accuracy of hand segmentation, as in [3]. In the stage of hand gesture modeling, hand postures as well as motion patterns are calculated from the hand gesture frame sequence, and the hand gesture model is created accordingly. The final stage is hand gesture recognition in which the output of current gesture model from the second stage is compared with each model in hand gesture database where the most matched hand gesture is selected as final recognition result.

This paper summarizes about a Gesture Recognition System that can recognize alphabetical Hand Gestures. The Computer Vision Based Techniques are explained in Section II which also gives a brief overview of this project. Then as we move on we discover these techniques in detail in Section III through V. Then in Section VI of this paper, the Experimental Results for this project have been discussed.

II. COMPUTER VISION BASED TECHNIQUES FOR HAND GESTURE RECOGNITION

In general, vision-based hand gesture recognition has three basic processing stages including Hand Segmentation, Gesture Modelling, and finally Gesture Classification, as in [3]. Fig1 shows these stages in two phases namely, the Training phase and the Testing phase. The Training phase includes only the first two processing stages, whereas the Testing phase includes all three processing stages.

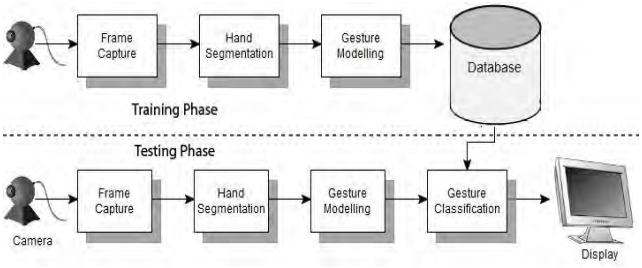


Fig. 1. Techniques for Hand Gesture Recognition

2.1 Hand Segmentation

The main procedure of hand segmentation is to detect hand regions in the sequence of hand gesture and separate them from backgrounds, as shown in Fig1. It should be mentioned that the correctness of hand gesture recognition has a close relationship to the accuracy of hand segmentation, as in [3]. Many conventional methods of hand segmentation take advantages of colour cues. However, the accuracy of hand gesture segmentation tends to be affected easily by several factors such as the skin colour differences between humans, the sensitivity of colour to illumination, and especially the situation when hand suffers in the presence of objects with similar skin colour. Therefore, many works have tried to combine the information from colour cue and other cues (i.e., motion, shape) in order to provide more accurate hand detection or segmentation. Besides, with the intensive study of object tracking, many object tracking methods such as parameterized deformable template based tracking algorithm, and particle filtering based tracking algorithm are employed for hand localization. In addition, hand detection also belongs to the research scope of object detection. Thus, many pattern recognition methods based on classifiers are always used for hand detection and recognition.

2.2 Gesture Modelling

In the stage of hand gesture analysis, hand postures as well as motion patterns are calculated from the hand gesture frame sequence, and the hand gesture model is created accordingly, as shown in the training and testing phases in Fig1.

2.3 Gesture Classification

The final stage is hand gesture recognition in which the output of current gesture model from the second stage is compared with each model in hand gesture database where the most matched hand gesture is selected as final recognition result. This has been illustrated in the testing phase of Fig1.

Different hand gesture modeling methods have diverse recognition approaches. The hand gesture is recognized by counting the number of active fingers. The hand gesture is modeled as the star skeleton, and the recognition is performed by distance signature. Other features such as hand position and direction, finger position and direction, and distance between the fingers are always used for establishing spatial model of hand gesture. As to dynamic hand gesture, we should further take the changes of spatial model in the temporal sequence into account. By integrating the temporal and spatial characteristics, many works have used a trajectory in high dimension space to represent dynamic hand gesture, and typically have used Hidden Markov Model (HMM) to recognize hand gesture. A non-trajectory gesture representation method is also used, which mainly uses statistical features (i.e. statistical moment) for hand gesture recognition.

III. SEGMENTATION OF HAND

Image Processing Techniques take images as input which gives out somewhat detail description of the scene. Most of the processing techniques perform Segmentation as a first step towards producing the description. Here input and output are images but in an abstract representation of the input. Segmentation technique basically divides the special domain, on which the image is defined, into meaningful parts or regions. This meaningful region may be a complete object or may be a part of it.

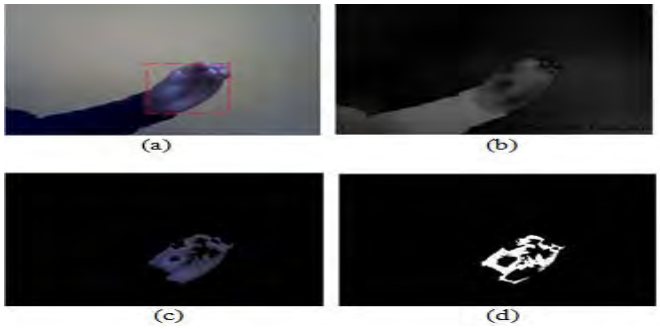


Fig. 2., (a) Input Frame, (b) Difference Image, (c) Skin Segmentation, (d) Binary Image.

Given above in Fig 2, are the snapshots of the Hand Segmentation. (a) shows image of the input video. (b) is an image which is obtained from the difference of background image and the current frame. (c) gives the skin segmentation, and (d) shows the binary image obtained by applying threshold to all the pixels in the frame.

A. *Feature Thresholding*

Consider an image that contains two types of regions, and the distinctness of the regions is reflected by the feature value at the pixels belonging to them.

Suppose there exists a threshold t such that feature values of all pixels that actually belong to regions of first type are less than or equal to t, and the gray values of all pixels that actually belong the regions of the second type are greater than t.

In this case, the segmented image is obtained as

b(r,c) = { 1 if p(r,c) <= t
{ 0 if p(r,c) > t

Where p(r,c) is the feature value at pixel (r,c), b(r,c) is a binary image where 1-pixels contribute foreground or object of interest and 0-pixels contribute the background.

IV. MODELLING OF THE GESTURES

In the previous module we saw that segmentation technique divides an image into a set of meaningful regions each of which is either an object or a part of an object. To generate the description of the scene or for image understanding, we usually need to recognize each of these regions. One of the prerequisites of identification and recognition is 'Feature Extraction'. By the term Feature Extraction we mean determining various attributes as well as properties associated with a region or object.

A. Boundary Based Description and Contour Generation

Boundary of an object may be an interior boundary or an exterior boundary. Interior boundary consists of pixels that belong to the object(s) itself, and pixels of exterior boundary belong to the background. That means the exterior boundary of an object is same as the interior boundary of the background.

Boundary based description is the most common way of depicting a shape and is capable of storing the local features more automatically. Boundaries encompass regions and they are often decomposed to smaller and regular regions which are easy to represent. Boundary is a collection of pixels at least one of whose neighbours belong to either background or any other region.

0000000000000000
0000011110000000
0000011111000000
0000111111100000
0000111111110000
0001111001111000
0001110000111000
0011100001110000
0011100111100000
0011111110000000
0011111000000000
0011110000000000
0001111110000000
0000111111110000
0000011111111000
0000000000000000

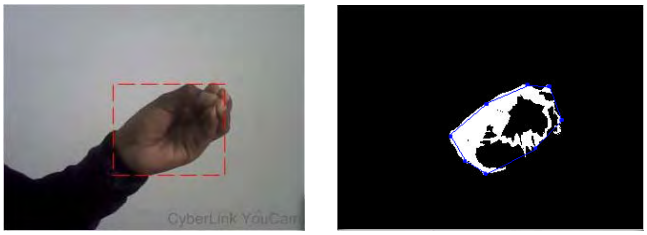
Fig. 3, Example of a Binary Image for Contour Generation

Suppose we walk along the boundary of an object in a binary image and the direction of movement change rapidly. An example of such an object is as shown in Fig 3. Hence the object in the figure can be represented by the boundary pixels namely => {(1,5), (1,6), (1,7), (1,8), (2,9), (3,10), (4,11), (5,12), (6,12), (7,11), (8,10), (8,9), (9,8), ...}.

B. Hull Points Extraction

The minimal convex polygon containing an object is known as the convex hull of that object.

The main idea is to select any two extreme points from the polygon, divide the polygon into two regions or planes. Now take one of the planes and pick a maximum distant point in that plane. Now eliminate all points that come under the imaginary line joining these 3 points. Then consider the first and the third point as extreme points and repeat the process recursively. So at any point we have an array of Hull points, an array of valid points, and an array of eliminated points. And the process is repeated until all valid points become Null. The resultant hull points may be as shown in Fig 4 (b).



(a) (b)

Fig. 4(a) Input Image, (b) Binary Image for Hull points Extraction.

C. Dynamic Time Wrapping

It is here where the Scaling, Rotation and Registration of the images come into picture. Registration is a process of mapping between temporal sequences of image frames. Here we establish a correspondence. Meaning, matching of identical shapes in the related image pair is done. Consider Fig 5 (a-f) which shows different angles in which the object may be detected and notice how each consecutive frames may have to be aligned.

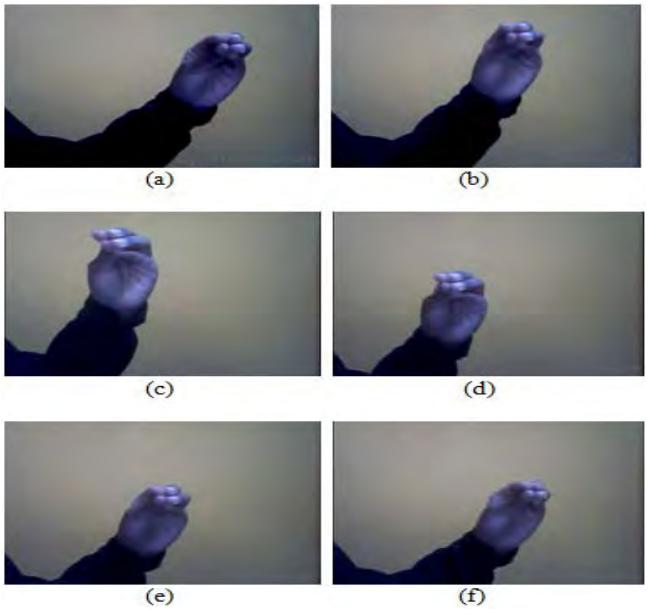


Fig. 5, Examples of different angles of Rotation and Scaling

The basic problem that DTW attempts to solve is how to align two sequences in order to generate the most representative distance measure of their overall difference. If you have two signals encoded as a sequence of evenly spaced values (representing, for example, the peak frequency of the signals), then an obvious way to compare the signals is to sum the differences in frequency at each point along the signals. However, a problem arises if there is any discrepancy in the alignment of the signals – if for example one of the signals is stretched or compressed compared to the other: how do you decide which points to compare with each other? For this bilinear interpolation is done before calculating the displacement between every two frames.

D. Haar Wavelet Transformation

Wavelet analysis is similar to Fourier analysis in that it allows a target function over an interval to be represented in terms of an orthonormal function basis. The Haar wavelet is also the simplest possible wavelet. The technical disadvantage of the Haar wavelet is that it is not continuous, and therefore not differentiable. This property can, however, be an advantage for the analysis of signals with sudden transitions, such as monitoring of tool failure in machines.

V. CLASSIFICATION OF GESTURES

The final stage is hand gesture recognition in which the output of current gesture model from the second stage is compared with each model in hand gesture database where the most matched hand gesture is selected as final recognition result. Here we use the Support Vector Machine (SVM) Classifier which is widely used for statistical classification and regression analysis. The SVM classifier learns from the data points in examples when they are classified belonging to their respective categories. SVM training algorithm builds a model that predicts whether a new example falls into one category or other as mentioned in [15]. Consider the example below, in Fig 3.

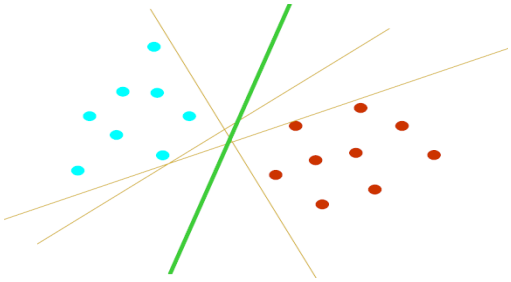


Fig. 3. Optimal separating hyperplane

Here there are many possible linear classifiers that can separate the data, but there is only one that maximises the margin (maximises the distance between it and the nearest data point of each class). This line

ar classifier is termed the optimal separating hyperplane. Intuitively, we would expect this boundary to generalise well as opposed to the other possible boundaries.

Suppose that the input video to be interpreted is that of alphabet ‘a’ then the resultant would look as shown in Fig 6.

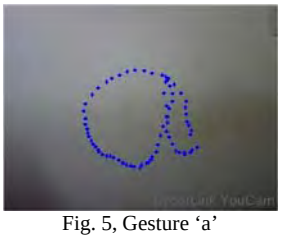


Fig. 5, Gesture ‘a’

VI. EXPERIMENTAL RESULTS

The proposed system is implemented on a Computer System running Windows Vista. The dynamic hand gesture sequences with 24-bit true colour are captured by an ordinary webcam attached to the system with a resolution of 320x240 pixels. The system is coded by VC++, and the OpenCV SDK is used. In the system, the hand gesture capture thread runs in parallel with the hand gesture analysis thread. The system satisfies the demand of real-time interaction. In this paper, 5 alphabetical dynamic hand gestures shown in Fig.6 are used. Each gesture is trained and analysed separately one at a time. Since one gesture could be finished in 20-30 seconds, the expected gesture sequence length is 30x5 frames. Therefore, the overall time taken for recognition of one gesture equals the expected sequence length in seconds. When the gesture features are extracted and the gesture is displayed, up to 80% accurate results can be obtained, provided the gesture input is correctly given. In other words, the output would be as accurate as the input would be. Note that if there is a great change of lighting during the hand gesture sequence then there are chances of the hand features not getting extracted accurately. The recognition success rate of each hand gesture depends on individual gesture itself. For example gestures ‘a’ and ‘b’ can have a high success rate since they look less similar to each other. Whereas, if we consider gestures ‘i’ and ‘j’ then it is so possible that the success rate would be low since the two gestures are very much similar to each other.

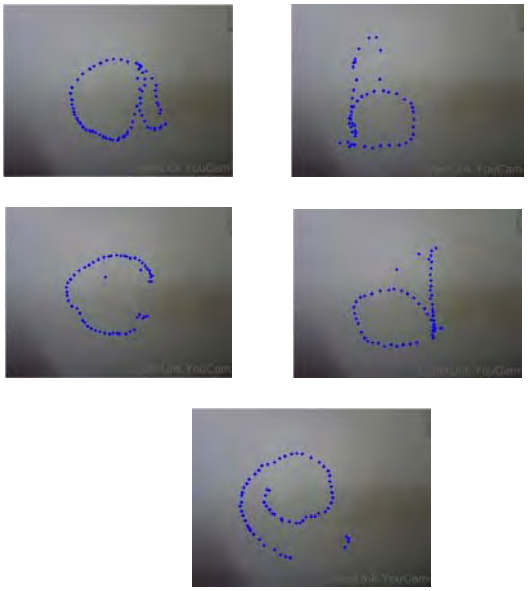


Fig. 5. First 5 alphabetical dynamic Hand Gestures

VII. CONCLUSION AND FUTURE WORK

In this paper we saw that the system could successfully interpret and recognise alphabetical hand gestures from the given input video. But in cases if the system is not able to recognise a particular gesture, then there are possibilities for the system to display wrong gesture result. So the future work can be carried on this problem.

REFERENCES

[1] http://www.google.com.
[2] http://www.wikipedia.com.
[3] Jiatong Bao, Aiguo Song, and Yan Guo "Dynamic Hand Gesture Recognition Based on SURF Tracking", International conference on Electric Information and Control Engineering, 2011.
[4] Bill Buxton, "Gesture Based Interaction", in continuum(2011), unfinished book, http://www.billbuxton.com/input14.Gesture.pdf.
[5] "A Real Time Hand Gesture Recognition Technique by using embedded device", International journal of advanced engineering sciences and technologies, 2011.
[6] S. Mohamed Mansoor Roomi, et al.., "Hand gesture recognition for human-computer interaction". Journal of Computer Science, 2010, 6(9):994-999.
[7] G. R. S. Murthy & R. S. Jadon, "A review of vision based hand gestures recognition", International Journal of Information Technology and Knowledge Management, July-December 2009
[8] Herbert Bay, "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding, 2008, 110(3):346-359.
[9] Malima, "A fast algorithm for vision-based hand gesture recognition for robot control", Proceeding of the IEEE International Conference on Signal Processing and Communications Applications. Antalya, Turkey: 2006. 1-4.
[10] D Love, "Distinctive Image Features from Scale-Invariant Key points". January 5, 2004. International Journal of Computer Vision, 2004.
[11] Konstantinos G. Derpanis, "A Review of Vision-Based Hand Gestures", Department of Computer Science. York University. February 12, 2004.
[12] E. Sanchez-Nielsen,"Hand gesture recognition for human machine interaction", Journal of WSCG, 2003, 12(1-3).
[13] Yu Zhong, Anil K.Jain, "Object tracking using deformable templates", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(5):544-549.
[14] Luo Juan & Oubong Gwun, "A Comparison of SIFT, PCA-SIFT and SURF". International Journal of Image Processing (IJIP).
[15] Zhengmao Zao & Prashan Premaratne, "Dynamic Hand Gesture Recognition using Moment Invariants". 978-1-4244-8551-2/10©2010 IEEE.