A Technique for Hand Gesture Recognition on Real Time Basis

Ayushi Shrivastav
Department of Computer Science and Engineering,
Shri Ramdeobaba College Of Engineering and Management
Maharastra, India

Radhika Agrawal
Department of Computer Science and Engineering,
Shri Ramdeobaba College Of Engineering and Management
Maharastra, India

S. G. Mundada
Department of Computer Science and Engineering,
Shri Ramdeobaba College Of Engineering and Management
Maharastra, India

ABSTRACT

Sign gesture is a non-verbal visual language, different from the spoken language in terms of medium of communication, but serves the same function for hearing & speech impaired community. Gesture Recognition, and more specifically hand gesture recognition, is one of the typical methods used in sign language for non-verbal communication. It is often very difficult for the hearing & speech impaired community to communicate their ideas and creativity to the normal humans. This paper focuses on discussing different methods to identify the gesture. Method for hand segmentation is discussed in terms of the different approaches to sub-components of the identifying the gesture. The judgement parameters are accuracy in real time performance, processing time, processor utilization, etc.

Keywords

Hand gesture recognition, Image processing, Human computer interaction (HCI), K-means clustering, Hand segmentation, hand gestures

1. INTRODUCTION

Sign language is a non-verbal language in which signs are made by moving one or both the hands, combined with facial expressions and postures of the body. The signs have a particular meaning and the people who understand the language knows what the sign stands for. It is one of widely used communication methods by hearing & speech impaired community. In sign language, gestures are considered to be any specific patterns or movements of the hands, face or body to make certain sense or meaning. In other words, gestures can be expressed with the help of facial expressions, limb movements or any meaningful bodily state.

A static hand gesture is a gesture which can be represented by a single image i.e. the meaning can be delivered by particular fixed position or posture captured in the image. The typical input for a static gesture is a single picture representing the gesture. A dynamic hand gesture can be said to be a series of hand postures connected in continuous motion over a definite period of time. These series or sequences of action are taken from a video recorder as a video, can be real time as well, and given as an input to the application deployed for the task of hand gesture recognition. The video can then be processed as frames and the gesture can be identified by applying certain recognition patterns, such as neural network, hand segmentation, etc.

In this paper, a methodology based on vision-based hand gesture recognition is proposed for a dynamic input, and hand segmentation, centroid calculation and direction tracking is performed on the frames and a string of code bits is generated, which is then used to recognize the gesture from already defined database. All this is done on video by processing it into frames, with the assumption that the continuous motion is a gesture representing a meaning. The proposed method is broken down into the following steps: Selection of frames, Background segmentation using K-means algorithm, Centroid calculation for detecting change in direction, thresh-holding for generating string of code bits, hand gesture interpretation and speech generation.

2. LITERATURE REVIEW

Hand gesture recognition methods can be broadly classified into three categories: Glove-based, Depth-based and Vision-based hand gesture recognition. [5][10] Glove-based hand gesture recognition methods predict the gesture by data extracted by the hardware called data glove. Data glove, also known as wired glove, uses various sensors, for tactile sensing and fine-motion control sensing, to capture information about positioning of the hand and joints in order to extract the gesture. S. Oniga & I. Orha[5] have implemented the hardware based approach for gesture recognition, by using a bracelet that captures the movement of the hand using accelerometers and Field Programmable Gate Array (FPGA) and then modeled, trained and simulated the desired network using Neural Network Toolbox[6].

In vision based approach, a camera or video recorder is used to capture and extract information about the hand position and analyze it to understand the gesture. Y. Fang et. al. [2] uses extended Adaboost method for hand detection and hand segmentation is done by collecting the color of the hand from neighborhood of features mean position. They further use the scale-space feature detection to detect blob and ridge structures, i.e. palm and finger structures.

Another method for hand gesture recognition is Depth-based recognition. The Senz3d camera captures a RGB video frame along with the associated depth data. Depth based thresholding is performed to remove the background. Then segmentation based on depth data is performed for the object closest to the camera. This may also include pixels that belong to the arm region. A color based filtering is performed on these pixels to check if these actually represent the hand pixels based on a predefined color model. If these are not recognized to belong to the hand region, then the algorithm waits for the next frame. [9]

Further, we explore the different methods for hand segmentation, hand tracking, feature extraction and gesture identification in dynamic gesture recognition.

Image segmentation is typically performed to locate the hand object in image. Proper hand segmentation background is very crucial for the overall efficacy and the effectiveness of a hand gesture recognition algorithm. The background is usually so chosen that maximum variation in pixel intensities between the hand image and background is observed, and any occlusion of hand with other body parts is avoided. Hand segmentation can indirectly be referred to as detecting Region of Interest. S. Bhowmick, S. Kumar and A. Kumar[10] have used a skin colour based hand segmentation technique that exploits a hybrid HSV+YCbCr colour model. These colour spaces have the advantage over RGB colour space in the sense that colour intensity has to be varied individually for each colour in RGB frame while on the other hand, H component (hue) and Y component control the colour intensity in their respective frames.

Another method that can be used for hand detection is background subtraction. Rather than detecting the ROI, the background is subtracted by applying clustering algorithms [1]. The proposed method uses K-means algorithm for the same.

Tracking in computer vision refers to the technique which constantly monitors the consecutive positions/locations of the region of interest (ROI). [10] The Region of Interest in this case is hand we're tracking. In order to find the gesture trajectory, the centroid / center of gravity (CoG) of the segmented hand is found out first. The centroid can be found out by moment calculation. A moment is a gross characteristic of a contour computed by integrating or summing over all of the pixels of the contour.

3. PROPOSED METHOD

The problem statement is divided into two modules: Gesture recognition, when the input is video, and Speech conversion of the identified gesture. For Speech conversion, a text-to-speech API is used. A combination of direction and codebit is used to identify the gesture. The gesture recognition method involves selecting frames, calculating direction & codebit of the gesture, detecting change in gesture (in case of multiple gesture) and passing the identified gesture to speech-to-text API. The preprocessing steps involve selecting frames from the video, detecting codebit and direction of the gesture, calculating least square error between frames and then identifying the gesture/string of gestures.

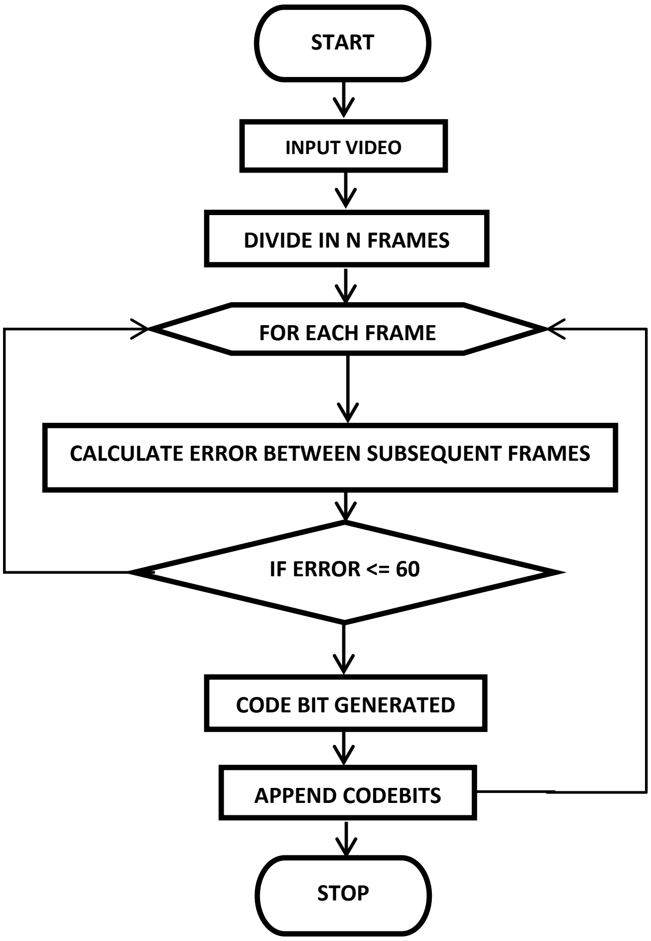

Fig1. Flowchart of the proposed algorithm

3.1. Splitting of video into frames

This video is preprocessed by splitting into frames and these frames are used to calculate gestures. Each frame has a unique frame_id, and using the frame_rate found out by inbuilt openCV method, total number of frames in the video is calculated. First frame per minute is selected and the proposed algorithm is applied on the selected frames. The number of frames is calculated by using the formula :

$$no.\ of\ frames\ =\ frameID \mod frameRate$$

3.2. Hand detection & Background elimination

For this step, K-means cluster can be used to form two clusters, i.e. a background cluster and a hand cluster. The K-means clustering algorithm is an iterative technique which is used to segment the image into K clusters. An initial set of centroid seeds are picked randomly and rest are assigned to its closest seed. After each assignment, the assigned centroid is updated by adding in the coordinates of the new point. Assigning all points to a set of successively updated centroids constitutes one iteration of k-means algorithm. Each iteration consists of a re-assignment of all points, until no point can be moved to a centroid closer than the one for the cluster it is already a member of. Every time a point is re-assigned, its old centroid must be down dated and its new centroid must be updated.

The RGB image is converted to a black and white image by clustering the pixels and changing the pixel's color depending on its cluster. If the pixel belongs to the hand cluster, the color of this pixel is changed to white, and if it belongs to the background cluster, the color is changed to black. So, applying K-means clustering, with k=2, on the image will result in a black and white image with two clusters, one is the hand region and the other cluster would contain the background region. The background cluster is discarded and hand cluster is used to perform further pre-processing steps.

Centroid Calculation & Finger Detection
Using the first frame as the input, the centroid of the hand is calculated using image moment, which is the weighted average of pixel's intensities of the image[1]. The centroid is calculated by first calculating the image moment using this formula.

$$M_{ij} = \sum \sum x^i y^j I(x, y) \qquad (1)$$

where Mij is image moment, I(x, y) is the intensity at coordinate (x, y).

Equation (1) is used to calculate moments used for the calculation of coordinates of the centroid.

$$\{\, \bar{x}\, \bar{y} \,\} = \{M_{10} / M_{00} \,, M_{01} / M_{00}\} \qquad (2)$$

By using equation (2), coordinates of centroid are computed. ($\bar{x}$, $\bar{y}$) are the coordinate of centroid and M00 is the area for binary image.

For peak detection, the concept of convex hull is used. The palm is separated from the hand by creating a palm mask, i.e. eroding the fingers away from the hand mask, by successive erosion and dilation morphological operations. Then by using convexity defects, the center of the hand and the number of fingers are detected.

Direction observation
Centroid calculated for each frame is used to observe any change in direction of the gesture. Consecutive frames and change in centroid of the frame are observed. If change in direction is observed, and if the change is in x co-ordinate, and the change is positive, then the direction is taken to be right and if it is negative, the change is taken to be left. Similarly, the change in y coordinate is also observed and according to the sign of the change, the direction is assigned as up or down. This change in direction is stored and used later to detect the gesture.

Thresh-holding and conversion into code
Now, in order to classify the raised and folded fingers i.e. significant and insignificant peaks among the detected peaks, the distances of the peaks from the centroid are used. For classifying the peaks, distance of each peak from the centroid is calculated using city block distance. Now, a threshold, T_h, is set. If the distance > T_h, the code bit value would be set to 1 and if the distance < T_h the value would be set to zero. These values would indicate whether the finger is open or close. If the value is 1, it indicates that the finger is open; and if the bit generated is 0, the finger is considered to be closed.

A string of code bits are obtained and this indicates the code of the input image which will be used for mapping the gesture to the meaning of the gesture.

3.3. Finding error between frames

For each frame, error is calculated using the least squared error formula. This calculated error is compared with its previous frame. If the error is less than 60% it is ignored. If it is more than 60%, it indicates that there is a change in gesture and the next frame is treated as a different gesture and is sent for processing of code bits.

3.4. Mapping of code with words/characters and speech as output

The code obtained in the above step is used to find a matching code in the stored code-word pair, where each code corresponds to characters/words i.e. the meaning of the gesture. The word is extracted from the code-word pair and passed as text to speech conversion API, thus giving the speech as output.

4. COMPARATIVE STUDY

Vision-based, glove-based and depth-based are widely used in hand gesture recognition. But the recognition method based on vision is hard to work well in bad conditions. And the recognition method based on glove also has an embarrassing situation. Although this method owns the advantages of less input data, high speed, and it can get 3D information about hands or fingers movement directly. It could also recognize a lot of hand gestures on time [1]. Being a newly developed distance measuring hardware, the depth camera gives a depth image that could reflect the 3d feature directly, which is not affected by the factors such as illumination, shadow and color. Even if there is a covered part between two objects, by using different distance information which we've got from the depth image, different parts of the covered object can be separated. But at this time, depth camera is too expensive to apply. However, the recognition method based on glove is not able to leave the support of equipment. It's impossible for users to wear bloated gloves all the time in nature condition. This obvious disadvantage destines its useless, so we need to develop a new technology to solve it. Depth-based recognition method has a high robust. And because it has the characteristics of real-time identification and high precision, it is a promising research direction. But depth camera based on the technologies such as time of flight (TOF), structure light, 3d laser scanning is so expensive that its utility has been limited.

TABLE 1: COMPARISON OF TECHNIQUES DISCUSSED

Technique	Advantages	Drawbacks
Thresholding	• Simple • Easy • Fast	• Sensitive to intensity of light
K-means Clustering	• Independent of Image intensity. • Forms clusters at run time.	• Takes background as Region Of Interest(ROI).
Convex Hull	• Easy to implement • Every point in the contour need not be accessed	• Does not detect fingers that are half folded.
Peak Detection using slope	• Detects more number of fingertips as compared to	• Need to access each point in the contour of

	convex hull. • Detects half folded fingers	ROI.

5. CONCLUSION

In this study of human hand gesture recognition, hand tracking using centroid and observing direction change were applied on the hand to detect the gesture. From this outline information, the co-ordinates of the centroid and the fingertips of the hand were obtained and the intervening differences were calculated. The time taken for detection was minimal and almost real-time. The limited number of gesture sets that we were able to detect, proved to be the only stumbling block, but we hope to circumvent this problem by further refining our algorithm in the future. Overall, this proposed method proved to be a considerable success when compared with standard methods in terms of accuracy. A further work can be carried out to show the efficiency of the system in terms of broad range of implementations.

6. REFERENCES

[1] M. Panwar (Centre for Development of Advanced Computing, Noida), 'Hand Gesture Recognition based on Shape Parameters'

[2] Y. Fang et. al. 2007, 'A REAL-TIME HAND GESTURE RECOGNITION METHOD'

[3] T. Nguyen & H. Huynh, 'Static Hand Gesture Recognition Using Artificial Neural Network', Journal of Image and Graphics, Volume 1, No.1, March, 2013

[4] M. Quraishi et. al., 'A Novel Human Hand Finger Gesture Recognition U sing Machine Learning', 2012 2nd IEEE International Conference on Parallel, Distributed and Grid Computing

[5] S. Oniga & I. Orha, 'Intelligent Human-Machine Interface Using Hand Gestures Recognition'

[6] L. Chen et. al., 'A Survey on Hand Gesture Recognition', 2013 International Conference on Computer Sciences and Applications

[7] M.Murugeswari (PG Scholar, Communication Systems, Anna University,Tamil Nadu) ,S.Veluchamy (Assistant Professor, Communication Systems, Anna University,Tamil Nadu), 'Hand Gesture Recognition system for Real-Time Application', 2014 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)

[8] M. Tao & L. Ma, 'A Hand Gesture Recognition Model Based on Semi-supervised Learning', 2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics

[9] R. Agrawal & N. Gupta, 'Real Time Hand Gesture Recognition for Human Computer Interaction', 2016 IEEE 6th International Conference on Advanced Computing

[10] Sourav Bhowmick et. al, 'Hand Gesture Recognition of English Alphabets using Artificial Neural Network', 2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)

[11] S. Gawande & Prof. N. Chopde, 'Neural Network based Hand Gesture Recognition', International Journal of Emerging Research in Management &Technology ISSN:2278-9359 (Volume-2, Issue-3)