

Vision Based Approach to Sign Language Recognition

Himanshu Gupta, Aniruddh Ramjiwal, Jasmin T. Jose

School of Computer Engineering, VIT University, Vellore-632014, India

Article Info

Article history:

Received Dec 19, 2017

Revised Mar 4, 2018

Accepted May 2, 2018

Keyword:

Detection

Gesture

Recognition,

Segment

Track

ABSTRACT

We propose an algorithm for automatically recognizing some certain amount of gestures from hand movements to help deaf and dumb and hard hearing people. Hand gesture recognition is quite a challenging problem in its form. We have considered a fixed set of manual commands and a specific environment, and develop a effective, procedure for gesture recognition. Our approach contains steps for segmenting the hand region, locating the fingers, and finally classifying the gesture which in general terms means detecting, tracking and recognising. The algorithm is non-changing to rotations, translations and scale of the hand. We will be demonstrating the effectiveness of the technique on real imagery.

Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Himanshu Gupta,

School of Computer Engineering,

VIT University,

Vellore-632014, India.

Email: guptahimanshu12@gmail.com

1. INTRODUCTION

The sign language is basically a language in which there is manual communication to convey meaning, contrary to conveyed sound patterns. This involves simultaneous combinations of orientation, hand shapes and movement of the person's hands, arms and body to express his thoughts. A number of sign languages share a lot of similarities with the spoken languages, that depends primarily on sound, and linguists considered both to be the types of natural language. Although there are many significant differences between spoken and signed languages, like as how they use space grammatically, sign languages show the same linguistic properties and use the same language faculty as do spoken languages.

Still it is not known that how many sign languages exist. One of the very common misconception is all sign languages are same worldwide or sign language is international. Aside from the pidgin International Sign, almost every country has its own, native sign language, and some countries have more than one because of different languages in the country itself based on the regions. The 2013 edition of Ethnologue lists 137 sign languages. Some sign languages have achieved some form of legal recognition, while there are many others which have no status at all.

In South Asia, Indo-Pakistani Sign Language (IPSL) is the predominant sign language used by at least several hundred thousand deaf signers. As with many sign languages, it is highly difficult to count numbers with any certainty, as the Census of India has not listed sign languages and most of the studies have focused on the north and urban areas.

Gestures are a very powerful mode of communication among humans. Amongst different types of modality of body, the hand gesture is the simplest and natural way of communication mode. The real time, vision based hand gesture recognition is more probable because of the latest advances in the field of computer vision and pattern recognition and image processing but it has yet, to be totally explored for Human

Computer Interaction (HCI). With the wide applications of HCI, now days, it becomes active focus of research.

In recent years, vision-based automatic hand gesture recognition has been an active research topic with a lot of motivating applications such as sign language interpretation, robot control and human computer interaction (HCI). But the problem is challenging because of a number of issues which includes the complicated nature of dynamic and static hand gestures, occlusions and complex backgrounds. We need very intensive computer resources to attack a problem in its generality which requires elaborate algorithms. Our motivation to do this work is a robot navigation problem, because we are highly interested in controlling a robot by hand signs given by a human being. We are interested in computationally efficient algorithms because of real-time operational requirements. In hand gesture recognition problem in earlier approaches there was a use of markers on the finger tips in a robot control context. To identify which are the fingers that are active in the gesture to detect the color and presence of the markers, an associated algorithm was used. But what makes this as an infeasible approach is placing markers on the user's hand which is very inconvenient. Recent methods use more accurate computer techniques and they don't require markers. Hand gesture recognition is performed through a curvature space method in, which involves finding the boundary contours of the hand. It is a robust approach that is translation, scale and rotation invariant on the hand pose, yet it is computationally demanding. A multi-system camera is used to pick the COG of the hand and points it with farthest distances from the center, giving us the locations of the finger tips, which are then used to obtain a skeleton image, and then finally for the gesture recognition and it is proposed in a vision-based hand pose recognition technique using skeleton images. A technique was proposed for gesture recognition for sign language recognition. Particle filters, Fourier descriptors, principal component analysis, orientation histograms, neural networks and specialized mappings architecture are included in the 3D and 2D hand gesture recognition which are the computer vision tools. Our focus is the recognition of a fixed set of manual commands by a robot, in a reasonably structured environment in real time. Therefore the important thing for us is speed, hence simplicity of the algorithm. We develop and implement such a procedure in this work. Our way involves segmenting the hand based on size constraints and skin color statistics. Based on the preprocessing steps which are finding the farthest point from the center of gravity and finding the center of gravity (COG) of the hand region, we will derive a signal which carries information on the activity of the fingers in the sign. Finally we identify the sign based on that signal. Our algorithm is non-changing to rotations, translations and scale of the hand. Also, this technique does not need any storage of a hand gesture database in the robot's memory.

2. RELATED WORK

Even after much of the efforts there are many scholarly articles in which different authors have used different techniques. In [1] the author intends to create a two device on wearable and another one that is desk based. The device will record video of person talking in asl. Wearable device will record the video of person wearing it while desk based will record video of other people. While there are two authors yuntao cui y and john wengz [2] researched discriminant analysis which are characterized as follows: one has two types of multivariate observations. The first, called training samples, are those whose class identity are known. The second type, referred to as test samples, consists of observations for which class identity are unknown and which have to be assigned to one of the class. One of the author said that a series of hypotheses and tests, where a hypothesis of model parameters at each step is generated in the direction of the parameter space (from the previous hypothesis) achieving the greatest decrease in mis-correspondence [3]. While there was a category in which a approach called "ruled based approach" is used. Rule-based approaches consist of a set of manually encoded rules between feature inputs. Given an input gesture a set of features as in [4] are extracted and compared to the encoded rules, the rule that matches the input is outputted as the gesture.

There has been many reviews written on hidden markov model. Hidden-markov-models (HMM) are a very strong and powerful technique for detailed examination of nondeterministic time signals. They are widely used in gesture recognition, speech recognition and sign language recognition, since they have the ability to handle spatio-temporal variations. In 2002 suat akyol [5] found a special class of HMMs, the so called linear models, only have forward transitions, i.e. A transition never leads to another state, which connects directly or indirectly to the current state. This topology is supposed to model the causality condition of time signals.

There has been a significant amount of research based on skin color. In [6] extracts skin, clothes and elbow region by matching template of initial person's region. Then it tracks the face, hands and elbow at each frame using skin colour and template of shape. While in 90s [7] there have been efforts for color of human skin which is described by spherical influence field of color prototypes. Its representative color features are extracted by using the method of color prototype density estimation in rce training procedure. In a paper

published in 2009 authors liwicki, stephan, and mark everingham [8] classifies pixels as hand or non-hand by a combination of three components: a signer-specific skin color model, a spatially-varying 'non-skin' color model, and a spatial coherence prior. Emil m. Petriu et al.[9] haar like features to tell the ratio between dark and bright areas, can operate much faster than pixel based system. It also robusts noise and lighting changes.

However, it is difficult to separate out hand from the image because of mixture of skin color in that still some authors researched about that. In [10] performing region-based segmentation of the hand, eliminating small false-alarm regions that were declared as "hand-like," based on their color statistics. Finding the center of gravity (cog) of the hand region as well as the farthest distance in the hand region from the cog. In one of the papers with the help of kinetic sensors we can acquire colour image input and depth map input as said in [11]. With the help of both, hand can be separated out. Colour image input will help to locate the hand while depth image input will distinguish hand from other objects in screen removing problem from cluttered backgrounds or other body parts mixing with hands. Some authors discussed in their papers and tried to solve illumination problem by adding another level to hand detection but [12] does not so we can say that this will be the least effective method. It uses hts algorithm for hand tracking and segmentation. It uses edge detection technique to it.

A number of authors conducted their research based on the frames. In [13] a tracking system is used to obtain head and hand positions, these co-ordinates are then clustered to obtain a quantized description for each frame which are then temporally concatenated to create biframe features. In 2013 one of the author in her technique said [14] that it uses video input from microsoft kinetic 3gear technology which sense the motion of the hand in a skeletal frames and give the value of coordinates in each frame. The proposed method takes into account both the local and global features associated with a gesture. In [15] for recognition of dynamic gestures implies the analysis of temporal sequences of frames, which contain the various poses assumed by the hand during its movement. Step by Step Process of Proposed Work as shown in Figure 1.

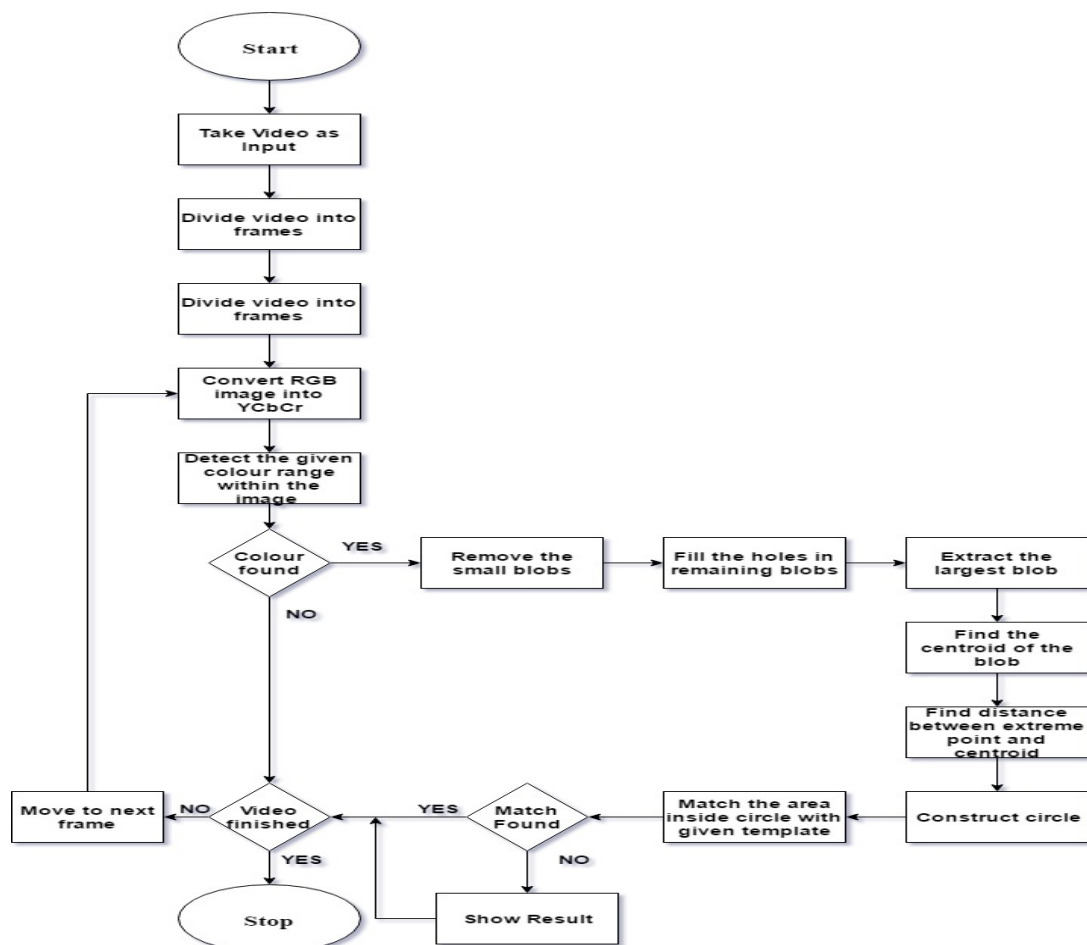


Figure 1. Step by step process of proposed work

3. PROPOSED WORK

We ask user to upload the video of his hand where he is using sign language to communicate. Video is collection of frames that are played together. If we divide the video into frames and consider each frame one by one. We process each frame as single image and move on to next frame till the end of the video we will be able to process each frame in video.

3.1. Localising Hand Region

We assume that the user has uploaded the video in acceptable format and decent quality. Then our first task is to segment out the hand in the video from the background. We achieve this result in two steps. First, we find the pixels that belong to the hand region. Refer Figure 1.

It is been observed that YCbCr colour space gives better clustering result and computational efficiency [16]. So we convert our image into YCbCr colour space (refer to Eq. 1). In YCbCr colour space, Y is luminance, Cb and Cr are chromaticity of blue and red colour. Cb and cr are two dimensional independent. Demonstrated in Figure 2 the converted RGB colour space to YCbCr colour space :

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1440 \\ -0.1687 & -0.3313 & 0.5000 \\ 0.5000 & -0.4187 & -0.0813 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \quad (1)$$

After converting the image we extract the skin region in the image. For our system to accept people with different skin tones we apply skin threshold to separate out the skin [17].

$$80 \leq Cb \leq 120 \quad \text{and} \quad 133 \leq Cr \leq 173$$

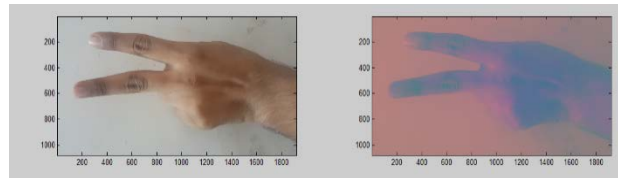


Figure 2. Converted image from RGB to YCbCr color space

3.2. Finding Centroid

After we found out the skin coloured regions in our image we calculate centroid, centre of gravity (COG) by using following formula

$$\bar{x} = \frac{\sum_{j=0}^n x_j}{n} \quad \text{and} \quad \bar{y} = \frac{\sum_{j=0}^n y_j}{n} \quad (2)$$

Where x_j and y_j are x and y coordinates of the j^{th} pixel in the hand area, and n denotes the number of pixels in the area.

After we obtain centroid (refer to Eq. 2) we find the distance from the most extreme point in the hand to the centre, normally this farthest distance is the distance from the centroid to tip of the longest active finger in the particular gesture [4].

We then draw a circle of whose radius is farthest distance from centroid. This circle contain the whole gesture.

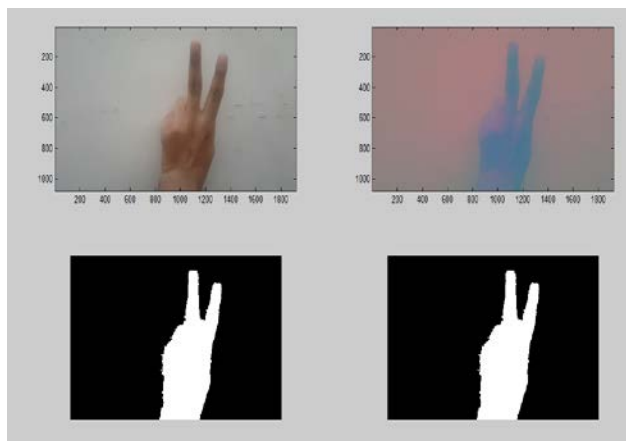


Figure.3.Expected images after conversion

3.3. Gesture Classification

Now we have to identify what the gesture means. For this we have used template matching due to its simplicity and predictable response with a limited training set. Template matching is basically the two dimensional cross-correlation of a grayscale image with a grayscale template, hence estimating the degree of similarity between the two. We compare the unknown gesture with the preset template models of individual gestures.

4. CONCLUSION

Deaf and Dumb people rely on sign language interpreters for communication. However, they cannot depend on interpreters every day in life mainly due to the high costs and the difficulty in finding and scheduling qualified interpreters. This system will help disabled persons in improving their quality of life significantly.

This paper proposes a hand gesture recognition method based on YCbCr colour space, COG and template matching. The proposed method can detect hand gesture from background. It accepts different skin tones and different illumination conditions. This method uses chromaticity in YCbCr colour space, avoiding the effect on hand gesture and uses COG to segment the hand. It then recognises the gesture with the help of template matching. This method is robust to different backgrounds and provides better result. This method will be a good addition to the ongoing research in the field of Human Computer Interface (HCI).

ACKNOWLEDGEMENTS

We would like to thank all the researchers working on this field who in one way or another guided us on achieving our goals. We would also like to express our appreciation to our guide at VIT University who was kind enough to share her views with us and offered some suggestions in making this project a success

REFERENCES

- [1] Starner, Thad, Joshua Weaver, and Alex Pentland. "Real-time american sign language recognition using desk and wearable computer based video." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.12 (1998): 1371-1375.
- [2] Tanibata, Nobuhiko, Nobutaka Shimada, and Yoshiaki Shirai. "Extraction of hand features for recognition of sign language words." *International conference on vision interface*. 2002.
- [3] Cooper, Helen, and Richard Bowden. "Learning signs from subtitles: A weakly supervised approach to sign language recognition." *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009.
- [4] Malima, Asanterabi Kighoma, Erol Özgür, and Müjdat Çetin. "A fast algorithm for vision-based hand gesture recognition for robot control." (2006): 1-4.
- [5] Chen, Qing, Nicolas D. Georganas, and Emil M. Petriu. "Real-time vision-based hand gesture recognition using haar-like features." *Instrumentation and Measurement Technology Conference Proceedings, 2007. IMTC 2007. IEEE*. IEEE, 2007.

- [6] Dong, Guo, Yonghua Yan, and M. Xie. "Vision-based hand gesture recognition for human-vehicle interaction." *Proc. of the International conference on Control, Automation and Computer Vision*. Vol. 1. 1998.
- [7] Ren, Zhou, Jingjing Meng, and Junsong Yuan. "Depth camera based hand gesture recognition and its applications in human-computer-interaction." *Information, Communications and Signal Processing (ICICS) 2011 8th International Conference on*. IEEE, 2011.
- [8] Garg, Pragati, Naveen Aggarwal, and Sanjeev Sofat. "Vision based hand gesture recognition." *World Academy of Science, Engineering and Technology* 49.1 (2009): 972-977.
- [9] Porta, Marco. "Vision-based user interfaces: methods and applications." *International Journal of Human-Computer Studies* 57.1 (2002): 27-73.
- [10] Geetha, M., et al. "A vision based dynamic gesture recognition of indian sign language on kinect based depth images." *Emerging Trends in Communication, Control, Signal Processing & Computing Applications (C2SPCA), 2013 International Conference on*. IEEE, 2013.
- [11] Murthy, G. R. S., and R. S. Jadon. "A review of vision based hand gestures recognition." *International Journal of Information Technology and Knowledge Management* 2.2 (2009): 405-410.
- [12] Akyol, Suat, and Ulrich Canzler. "An information terminal using vision based sign language recognition." *ITEA Workshop on Virtual Home Environments, VHE Middleware Consortium*. Vol. 12. 2002.
- [13] Ghotkar, Archana S., and Gajanan K. Kharate. "Hand segmentation techniques to hand gesture recognition for natural human computer interaction." *International Journal of Human Computer Interaction (IJHCI)* 3.1 (2012): 15.
- [14] Cui, Yuntao, and Juyang Weng. "Appearance-based hand sign recognition from intensity image sequences." *Computer Vision and Image Understanding* 78.2 (2000): 157-176.
- [15] Liwicki, Stephan, and Mark Everingham. "Automatic recognition of fingerspelled words in british sign language." *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*. IEEE, 2009.
- [16] Qiu-yu, Zhang, et al. "Hand gesture segmentation method based on YCbCr color space and k-means clustering." *International Journal of Signal Processing, Image Processing and Pattern Recognition* 8.5 (2015): 105-116.
- [17] Basilio, Jorge Alberto Marcial, et al. "Explicit image detection using YCbCr space color model as skin detection." *Applications of Mathematics and Computer Engineering* (2011): 123-128.