

An Efficient and Robust System for Hand Gesture Recognition and Interpretation

Apoorva M A

Dept. of Computer Science
GSSSIETW Mysuru

Harshitha M S

Dept. of Computer Science
GSSSIETW Mysuru

Chaitra S

Dept. of Computer Science
GSSSIETW Mysuru

Akshitha V

Dept. of Computer Science
GSSSIETW Mysuru

Rajath A N

Assistant Professor
Dept. of Computer Science
GSSSIETW Mysuru

Abstract:- This paper devises a vision-based hand detection system which converts sign language to text. The present work is an attempt to develop a convenient system that is helpful for the people who have hearing difficulties and in general who use very simple and effective method; sign language. This system can be used for converting sign language to text. The camshift algorithm and PCA classifier is used. To improve the performance of the system efficient features and visual properties of human hand are selected. In addition, the detection is accelerated due to several optimization methods, including the method for fast calculation of histogram features and skin color pre-detection. Experiments were performed on our self-constructed dataset. The result showed faster output.

Keywords— Hand Detection; Sign Language; Histogram Features; Camshift Algorithm; Principal Component Analysis; Recognized Text

I. INTRODUCTION

Hand gesture recognition is one of the primary topics in human-computer interaction. In the early study of this technique, gesture recognition was accomplished by using hand gloves [1] or color markers on the fingertips [2]. As those devices make the usage very inconvenient, unnatural and do not produce robust segmentation results, most of recent works focus on the vision-based gesture recognition with naked hands. However, the difficulty is how to segment and track the hands in complex backgrounds, such as light variation, face interference, movements of other objects and so on. Recent studies in the field of computer vision and pattern recognition show a great amount of interest in content extraction from images and videos. The content can be in the form of objects, color texture, shape as well as the relationships between them. A gesture is a motion of the body that conveys information. An activity of conveying meaningful information is called communication. The primary necessity in human-computer interaction involves hand gesture recognition. Human hand movements have significant functions and most of these functions are controlling functions (e.g. object grasping). There are some logically explainable functions that human hand performs like Sign language, pointing etc. Training the machine to understand these functions is a hard task.

Hand gesture recognition can be done either dynamically or statically. Most of the methods focus on dynamic method of capturing the hand gestures, since it is closer to the real time. These gestures are distinguished based on the shape of the hand and nature of motion. Some of the gesture recognition methods can work only on the constrained environments such as using gloves, uniform or fixed backgrounds. Various steps are involved in hand gesture recognition such as capturing hand movements, segmentation, feature extraction, training dataset, classification and finally gesture recognition. Each of these steps can use different techniques.

For detecting hand there are various algorithms including skin color based algorithms. YCbCr segmentation method for hand gesture recognition can be used, while the background of images should be clear, simple and uniform. YCbCr color model is used in order to improve the detection accuracy [5]. HSV color space can be used to extract the skin-like region by estimating the parameter values for skin pigment [6]. These methods get rid of the background information which can split image in reversed side to enhance the performance.

Most of recently developed tracking algorithms use the following principles: correlation methods, optical flow, background subtraction, particle filtering, methods based on probability density evaluation, etc. The correlation and optical flow methods are distinguished with their high computational complexity making them hardly suitable for real-time applications [3].

In our design, we use an approach based on the probability density estimation algorithm known as Camshift. This simple technique has low computational cost and can perform in real-time mode. Camshift has sufficient reliability, is able to track non-rigid objects when the camera is moving, shows low sensitivity to noise and occlusions.

The rest of the paper is organized as follows. Section 2 consists of introduction to sign language. A detailed description of proposed methods used for segmentation,

tracking and recognition is explained in section 3. The Section 4 contains experimental results. Finally the conclusion and further work are presented in section 5.

II. SIGN LANGUAGE

It is easy to find a wide number of sign languages all over the world and almost every spoken language has its respective sign language, so there are about more than 200 languages available. There are several sign languages available such as American, British, German, French, Italian, and Turkish Sign Language. American Sign Language (ASL) is well-known and the best studied sign language in the world. Figure 1 shows alphabets of ASL.

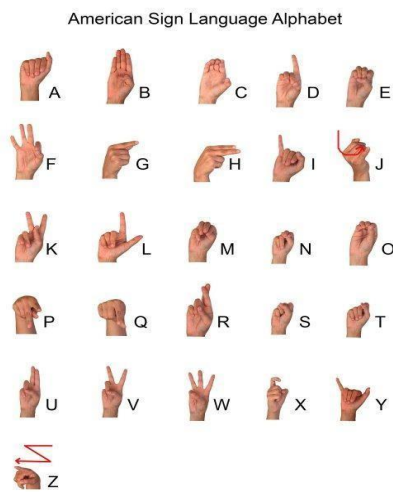


Figure 1: American Sign Language

III. PROPOSED ALGORITHM

The figure below shows the proposed method for gesture acquisition and recognition.

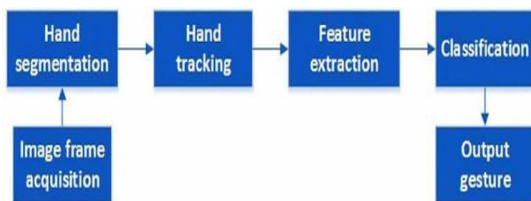


Figure 2: Proposed method for hand gesture recognition

As shown in the figure 2, the proposed method consists of following 6 steps: (1) video acquisition, (2) Hand segmentation, (3) Hand tracking, (4) feature extraction, (5) Classification, (6) Output Gesture.

Video acquisition is done by using webcam. The webcam captures the video, identifies the hand and recognizes the gestures. Depending on the digital camera's resolution there is variation in the input image's size. Normally the

resolution starts from 1 Megapixel. In the beginning, the input image is down-sampled by an integral multiple so that the dimension of the image is diminished to the close multiple of 0.35 Megapixels. This is done so that the proposed algorithm should work smoothly. Afterwards, the down-sampled image was changed to 8-bit grayscale.

Hand Segmentation is done by using skin color and chrominance components. For most images, the RGB color space is considered as the default color space. In order to convert into other color spaces, we can apply linear or non-linear transformation on the RGB components. In this algorithm, the input RGB image is converted in to YCbCr images due to the fact that RGB color space is more sensitive to different light conditions so we need to transform the RGB values in to YCbCr.

A. Camshift Algorithm

Camshift algorithm is used for *hand detection* and tracking. The original camshift algorithm uses one-dimensional histogram as a captured object model. The histogram consists of the hue (H) channel in HSV colour space.

The object search is being conducted through finding the probability distribution maximum obtained from a so-called histogram back-projection procedure. In order to reduce the amount of calculations, the colour probability distribution is scanned not over the whole image. Instead, we restrict ourselves with calculating the distribution in a smaller image region surrounding the current search window.

The camshift algorithm reduces to the following sequence of steps [5]:

- 1) Set the calculation Region of Interest (ROI) based on the probability distribution to be equal to the whole frame.
- 2) A two-dimensional mean shift search window is chosen as the initial location.
- 3) Skin tonal distribution is calculated in the 2D region on the chosen search window location in an area slightly larger than mean shift window size.
- 4) The maximum density of the skin tonal colour is searched using mean shift parameter. The number of iterations are set. Area and the size of the region to be tracked is stored.
- 5) For the next video frame, the search window is placed in the chosen position and fixed in step 4. Window size is made resizable ion of interest. Go to step 3.

As mentioned above, on step 2 the mean shift is calculated search window by the back-projection procedure. This low level operation puts the pixel values in the image into correspondence with the values of target's histogram bins.

If the target histogram contains a significant number of features that belong to a background image or some adjacent objects, target position and scale can hardly be

determined accurately. To cope with this problem, we used the ratio between target's histogram bins and the respective background histogram bins (the background corresponds to images outside the initial search window) – the so-called histogram weighing.

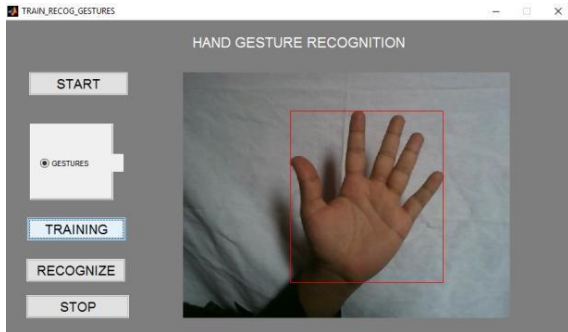


Figure 3: Hand detection and tracking using Camshift Algorithm

B. PCA Classifier

We used Principal Component Analysis to find a match in the database. This technique is also used in image compression. We need a set of data images we store in our program. After we have used PCA on the database, the original data will be in the forms of the eigenvectors.

After the image is input to the system, we will measure the difference between the eigenvectors in input images with the original images in the dataset, and then we need to determine which picture has the least difference to identify the input image.

Principal component analysis (PCA) is a technique used to emphasize variation and bring out strong patterns in a dataset. It's often used to make data easy to explore and visualize. PCA reduces the dimensionality of the dataset, which takes the dimensions that encode the most important information and removes the dimensions that encode the least important information.

By reducing the number of dimensions, the data utilizes less space, thus allowing classification on larger datasets in less time. Further, by taking only the salient dimensions, PCA projects the dataset onto dimensions that hold the most meaning, thus drawing out patterns in the dataset.

A dataset usually has two dimensions, like (height, weight). This dataset can be plotted as points in a plane. But if we want to tease out variation, PCA finds a new coordinate system in which every point has a new (x,y) value. The axes don't actually mean anything physical; they're combinations of height and weight called "principal components" that are chosen to give one axes lots of variation.

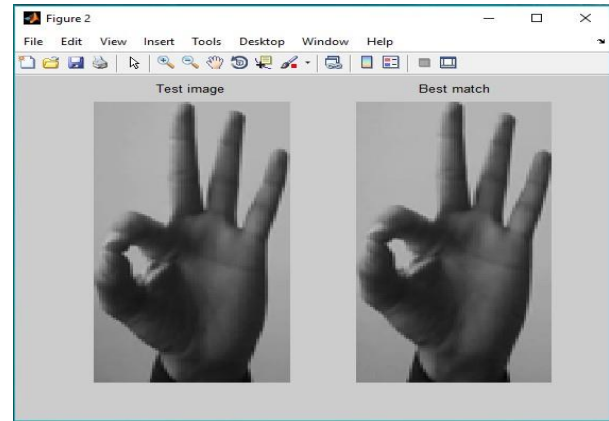


Figure 4 : Working of PCA classifier

C. Dataset

In recent years, more and more datasets have been created by researchers in order to develop, train, optimize and evaluate algorithms; several of them have been made publicly available to developers and researchers. For our experiment we have used our own data set. This dataset contains gestures performed by 4 different people, each performing 26 different gestures for all English alphabets repeated 200 times each, for a total of 20,800 gestures. Figure 5 shows the dataset being used for training.

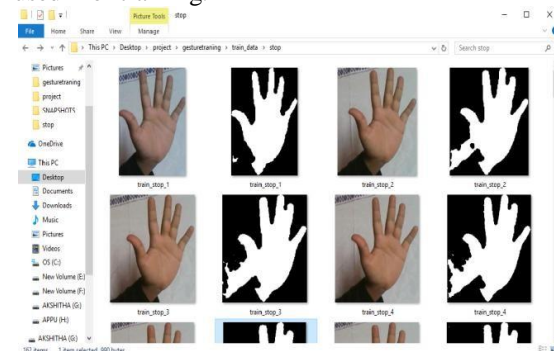


Figure 5: Dataset used for training

IV. EXPERIMENTAL RESULTS

We were able to recognize gestures for all the 26 letters of the English alphabet with different circumstances and variations. We have also designed a graphic user interface which allows the user to recognize the pre-defined gestures. Once the gesture is recognized, the corresponding alphabet appears on the top of the recognized gesture.

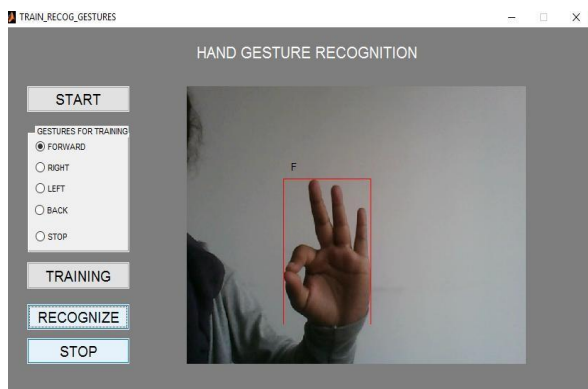


Figure 6 : Output image where the algorithm gave the perfect performance

V. CONCLUSION

This paper is about a system that can support the communication between deaf and ordinary people. The aim of the study is to provide a complete dialog without knowing sign language. In this paper, a segmentation method using YCbCr color space and RGB value of the skin are used. A suitable thresholding and morphology operations is proposed for hand detection phase in sign language recognition systems. In this method, a model for human skin color distribution is built using a database of labeled skin pixels. After thorough trials of the algorithm, new ways to its improvement became clear. Particularly, scale adaptation would give much benefit. Further it can also be converted to speech in different languages. All these methods forms a basis for future investigation and enhancements.

REFERENCES

- [1] S. S. Fels and G. E. Hinton, "Glove-Talk: a neural network interface between a data-glove and a speech synthesizer," *Neural Networks, IEEE Transactions on*, vol. 4,(1993), pp. 2-8.
- [2] J. Davis and M. Shah, "Recognizing hand gestures," in: J.-O. Eklundh (Ed.) *Computer Vision — ECCV '94*, Springer Berlin Heidelberg.
- [3] E. Roichman, Y. Solomon, Y. Moshe, "Real-Time Pedestrian Detection and Tracking" in *Proc. of the 3rd European DSP Education and Research Symposium (EDERS 2008)*, Tel-Aviv, pp. 281-288, June 2008.
- [4] H. Fu, Z. Cao and X. Cao, "Embedded omni-vision navigator based on multi-object tracking," in *Machine Vision and Applications*.

- [5] C. Chuqing and L. Ruifeng, "Real-Time Hand Posture Recognition Using Haar-like and Topological Feature," *International Conference on Machine Vision and Human-machine Interface*, 683 – 687, 2010atics(IC3I).
- [6] M.M. Hasan and P.K. Mishra, "HSV Brightness Factor Matching for Gesture Recognition System," *International Journal of Image Processing*, Vol.4, No.5, 456-467, 2011.
- [7] William C. Stokoe, Dorothy C Casterline, and Carl G Croneberg. "A Dictionary of American Sign Language on Linguistic Principles" Linstok Press, [Silver Spring, Md.], New Edition, 1976.
- [8] S. M. Canny, J. F. (1986). "A computation approach to edge detectors." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8, 34-43.
- [9] Karthikeyan, B., Vaithiyathan, V., Venkataraman, B., Menaka, M. — "Analysis of image segmentation for radiographic images" in *Indian Journal of Science and Technology* 5 (11), pp. 3660-3664.
- [10] Felzenszwalb, P. F., Huttenlocher, D.P., —"Efficient graph-based image segmentation," *International Journal of Computer Vision.*, 2004, 59, (2), pp. 167-181.
- [11] Lehmann, F. — "turbo segmentation of textured images," in *Pattern analysis and Machine Intelligence*, vol:33, pp: 16-29, 2011.
- [12] Gaurav Kumar, —"A detailed review of Feature Extraction in Image Processing System," in *2014 Fourth International Conference on Advanced Computing & Communication Technologies*.
- [13] Rohith Kumar Gupta — "A Comparative Analysis of Segmentation Algorithm for Hand Gesture Recognition" *2011 Third International Computational Intelligence, Communication System and Networks*.
- [14] M.Ali.Quresi, Abdul Aziz, Muhammad AmmarSaeed, Muhammad Hayat —"Implementation of an Efficient Algorithm for Human Hand Gesture Identification" 978-1-4577-0069-9/11, 2011, IEEE.
- [15] R. Lionnie, I. K. Timotius and I. Setyawan, — "An analysis of edge detection as a feature extractor in a hand gesture recognition system using skin color segmentation" in *Electrical Engineering and Informatics (ICEE), 2011 International Conference on* (2011), pp.1-4.