RESEARCH ARTICLE OPEN ACCESS

Image Based Sign Language Recognition on Android

Prutha Gandhi[1], Dhanashri Dalvi[2], Pallavi Gaikwad[3], Shubham Khode[4]
[1,2,3,4](Computer Department, Modern Education Society's College of Engineering, and Pune)

Abstract:

A mediator person is required for communication between deaf person and a second person. But a mediator should know the sign language used by deaf person. But this is also not possible always since there are multiple sign languages for multiple languages. It is difficult for a deaf person to understand what a second person speaks. And therefore deaf person should keep track of lip movements of second person in order to know what he is speaking. But the lip movements do not give proper efficiency and accuracy since the facial expressions and speech might not match. To overcome the above problems we have proposed a system, an Android Application for recognizing sign language using hand gesture with the facility for user to define and upload their own sign language into the system. The features of this system are the real time conversion of gesture to text and speech. For two-way communication between deaf person and second person, the speech of second person is converted into text. The processing steps include: gesture extraction, gesture matching and conversion of text to speech and vice-versa. The system is not only useful for deaf community but can also be used by common people who migrate to different regions and do not know local language.

Keywords **— Gesture extraction and detection, pattern matching, text generation, text to speech and speech to text conversion.**

I. INTRODUCTION

A Deaf person is very much dependent on the sign language to communicate with the other person. So the person who is interacting with the deaf person needs to know the sign language in order to understand and communicate effectively. Since many people are not familiar with sign language, it very difficult for a deaf person to have interaction with the society.

The previously implemented system had a predefined database with a limited scope. Thus, we are facilitating an application that will allow a user to define their own database i.e. to define and upload his own sign language in the system. This feature will help deaf people to communicate with other people varying from different countries or regions. Our application includes these phases:

- Digitization and image capture
- Compression (coding)
- Segmentation
  - Edge and feature detection
- Scene Analysis
  - Color
  - Motion
  - Object recognition
- Pattern matching
- Text generation
- Text to speech conversion
- Speech to text conversion

II.RELATED WORK

In [1] M. Mohandes, M. Deriche, and J. Liu, designed an ArSLR (Arabic Sign Language Recognition System) for alphabet recognition, isolated word recognition, and Continuous signer recognition using image based and sensor based approach. The image based approach mostly depends on the coloured gloves and knuckles. This approach works efficiently for determining geometric features and body/facial expressions. The sensor based approach utilizes the glove specs

(power, cyber, and data gloves), mostly statistical features, 3D position information.

In [2] Ravikiran J, Kavi Mahesh, Suhas Mahishi, Dheeraj R, Sudheender S, Nitin V Pujari, designed a highly accurate image processing algorithm for recognizing American Sign Language. Their algorithm implementation does not require use of any gloves or markers. The implemented system detects the number of open fingers using the concept of boundary tracing which also combines finger-tip detection.

In [3] Son Lam Phung, Abdesselam Bouzerdoum and Douglas Chai have analysed three important issues related to pixel wise skin segmentation: colour representation, colour quantization, and classification algorithm. They found that the Bayesian classifier with the histogram technique and the multilayer perceptron classifier have higher classification rates compared to other tested classifiers.

In [4] Ashish Sethi, Hemanth S, Kuldeep Kumar,Bhaskara Rao N,Krishnan R, designed an application for dumb and deaf person. The application was the integration of already existing methods. Their application processing includes: gesture extraction, gesture matching and conversion to speech. They used histogram matching, bounding box computation, skin colour segmentation and region growing methods for gesture extraction. For gesture matching they used feature point matching and correlation based matching techniques. Integrating all these methods their paper provides following four approaches:
Approach A: Skin colour segmentation with Feature point matching using SIFT
Approach B: Region Growing with Feature point matching using SIFT
Approach C: Skin colour segmentation with Correlation matching
Approach D: Region Growing with Correlation matching
The application also includes gesture to text conversion.

In [5] William T. Freeman, Michal Roth, provided orientation histogram as a feature vector for gesture classification and interpolation. This method is simple and fast to compute, and provides robustness to scene illumination changes. They provide two categories of gestures: Static gestures and Dynamic gestures. Static gesture includes a particular hand configuration and pose represented by a single image.Moving gesture, represented by a sequence of images comes under dynamic gesture. They used a pattern recognition system, converts the a image or sequence of images into a feature vector, which then compared with the feature vectors of a training set of gestures. They also used a Euclidean distance metric and video digitizer. In the run phase, the computer compares the feature vector for the present image with those in the training set, and picks the category of the nearest vector, or interpolates between vectors. The methods are image-based, simple, and fast.

In [6] V. Nayakwadi, N. B. Pokale,In this paper a survey on various recent gesture recognition approaches is provided with particular emphasis on hand gestures. A review of static hand posture methods are explained with different tools and algorithms
appliedongesturerecognitionsystem,includingconnectionistmodels,hidden Markov model, and fuzzy clustering.
Vision Based approaches: In vision based methods the system requires only cameras to capture the image required for the natural interaction between human and these approaches are simple but a lot of gesture challenges are raised such as the complex background, lighting variation, and other skin colour objects with the hand object.
InstrumentedGloveapproaches:
Markedglovesorcolouredmarkersaregloves thatwornbythehumanhandwithsomecolourstodirecttheprocessoftrackingthehandandlocatingthepalmand fingers,whichprovidetheabilitytoextract geometric features necessary to form hand shape. Gesture Recognition Techniques: Most of the researches use ANN as a classifier in gesture recognition process.
Histogram Based Feature: A method for recognizing gestures based on pattern recognition using orientation histogram.
Fuzzy Clustering Algorithm: In fuzzy clustering, the partitioning of sample dataintogroupsinafuzzywayarethemaindifferencebet

weenfuzzyclustering and other clustering algorithm, where the single data pattern might belong to different data groups.

Hidden Markov Model (HMM): HMM is a stochastic process, with a unite number of states of Markov chain, and a number of random functions so that each stat ehasarandom function. HMM system topology is represented by one state for the initial state, a set of output symbols, and a set of transitions state. HMM contained a lot of mathematical structures and has proved its efficiency for modelling spatiotemporal lin formation data. Sign language recognition, are one of the most applications of HMM.

In [7] Massimo Piccardi, reviews the Background subtraction method, which is a widely used approach for detecting moving objects from static cameras. This paper provides a review of the main methods and an original categorisation based on speed, memory requirements and accuracy. Methods reviewed include parameter I can dnon-parametric background density estimates and spatial correlation approaches. Several methods forperformingbackgroundsubtractionhavebeenprop osed, allofthesemethods try to effectively estimate the background model from the temporal sequence of the frames. The methods reviewed in the following are: Running Gaussian average, Temporal median filter, Mixture of Gaussians, Kernel density estimation (KDE), Sequential KD approximation, Concurrence of image variations, Eigen-backgrounds.

III. RESEARCH WORK

For extracting and processing of an image, process of acquisition is performed. Generally the image acquisition process involves pre-processing such as scaling etc. A scale space representation can be used. A scale space is representation of an image at multiple resolution levels. Then Difference of Gaussians can be applied on the image. Difference of Gaussians is a feature enhancement algorithm that involves the subtraction of one blurred version of an original image from another, less blurred version of the original. In the simple case of grayscale images, the blurred images are obtained by convolving the original grayscale images with Gaussian kernels having differing standard deviations. Blurring an image using a Gaussian kernel suppresses only high-frequency spatial information. Subtracting one image from the other preserves spatial information that lies between the range of frequencies that are preserved in the two blurred images. Thus, the difference of Gaussians is a band-pass filter that discards all but a handful of spatial frequencies that are present in the original grayscale image. There are many ways to handle image translation, one way to handle translation problems on images is template matching, it is used to compare the intensities of the pixels, using the SAD (Sum of absolute differences) measure. The other way include a pixel in the search image with coordinates (x_s, y_s) has intensity $Is(x_s, y_s)$ and a pixel in the template with coordinates (x_t, y_t) has intensity $It(x_t, y_t)$. Thus the absolute difference in the pixel intensities is defined as

$Diff(x_s, y_s x_t, y_t) = | Is(x_s, y_s) – It(x_t, y_t) |.$

Image processing also includes image segmentation. Segmentation procedures partition an image into its constituent parts or objects. In general, autonomous segmentation is one of the most difficult tasks in digital image processing. A rugged segmentation procedure bringstheprocess towards successful solution of imaging problems that require objects to be identified individually.

Then the representation and description of image is provided. Knowledge base is used to store the information about an image that can be later utilize for object recognition.

The algorithms for image extraction and detectioninclude background subtraction and blob detection algorithm.

Blob detection is an algorithm used to determine if a group of connecting pixels are related to each other. This is useful for identifying separate objects in a scene, or counting the number of objects in a scene.

The Background subtraction is a technique in the fields of image processing and computer vision wherein an image's foreground is extracted for further processing (object recognition etc.)

The background subtraction method has some disadvantages such as:

Background subtraction can be a powerfulallied when it comes to segmenting objects in a scene. The method, however, has some build-in limitations that are exposed especially when processing video of outdoor scenes. First of all, the method requires the background to be empty when learning the background model for each pixel. This can be a challenge in a natural scene where moving objects may always be present. One solution is to median filter all training samples for each pixel. This will eliminated pixels where an object is moving through the scene and the resulting model of the pixel will be a true background pixel. An extension is to first order all training pixels (as done in the median filter) and then calculate the average of the pixels closest to the median. This will provide both a mean and variance per pixel. Such approaches assume that each pixel is covered by objects less than half the time in the training period.

Another problem is that when processing outdoor video a pixel may cover more than one background. This will result in poor segmentation of pixel during background subtraction. Another problem in outdoor video is shadows due to strong sunlight. Such shadow pixels can easily appear different from the learnt background model and hence be incorrectly classified as object pixels.

IV. ALGORITHMS

A. *Background Subtraction*

It is also known as Foreground Detection [7], is a technique in the fields of image processing and computer vision wherein an image's foreground is extracted for further processing (object recognition etc.). Generally an image's regions of interest are objects (humans, cars, text etc.) in its foreground. After the stage of image pre-processing (which may include image delousing, post processing like morphology etc.) object localization is required which may make use of this technique. Background subtraction is a widely used approach for detecting moving objects in videos from static cameras. The rationale in the approach is that of detecting the moving objects from the difference between the current frame and a reference frame, often called “background image”, or “background model”. Background subtraction is mostly done if the image in question is a part of a video stream. Background subtraction provides important cues for numerous applications in computer vision. It includes the following steps:

step1:Motion detection, it is done by using segmentation where the moving projects are segmented from the background, i.e. take an image as background and take the frames obtained at the time t, denoted by I(t) to compare with the background image denoted by B.

step 2:We can segment out the objects simply by using image subtraction technique of computer vision meaning for each pixels in I(t), take the pixel value denoted by P[I(t)] and subtract it with the corresponding pixels at the same position on the background image denoted as P[B].In mathematical equation, it is written as:

$$P\,[F(t)]=P[I(t)]-P[B]$$

step3:The background is assumed to be the frame at time t. This difference image would show some intensity for the pixel locations which have changed in the two frames.This approach will only work for cases where all foreground pixels are moving and all background pixels are static.

step4:A threshold "Threshold" is put on this difference image to improve the subtraction.

$$|P\,[F(t)]-PF(t+1)]\ |>\{\text{Threshold}\}$$

This means that the difference image's pixel's intensities are 'thresholded' or filtered on the basis of value of Threshold. The accuracy of this approach is dependent on speed of movement in the scene. Faster movements may require higher thresholds. Threshold: The simplest thresholding methods replace each pixel in an image with a black pixel if the image intensity $I_{i,j}$ is less than some fixed constant T (that is, $I_{i,j}<T$), or a white pixel if the image intensity is greater than that constant. In the example image on the right, this results in the dark tree becoming completely black, and the white snow becoming complete white.

Multiband thresholding: Colour images can also be thresholded. One approach is to designate a separate threshold for each of the RGB components of the image and then combine them with an AND operation. This reflects the way the camera works

and how the data is stored in the computer, but it does not correspond to the way that people recognize color. Therefore, the HSL and HSV color models are more often used; since hue is a circular quantity it requires circular thresholding.

B. Blob Detection Algorithm

Blob detection is an algorithm used to determine if a group of connecting pixels are related to each other[8]. This is useful for identifying separate objects in a scene, or counting the number of objects in a scene. In computer vision, blob detection methods are aimed at detecting regions in a digital image that differ in properties, such as brightness or color, compared to surrounding regions. A blob is a region of an image in which some properties are constant or approximately constant; all the points in a blob can be considered to be similar to each other. To find colored blobs, you should convert your color image from RGB to HSV format so that the colors are easier to separate. Check given images taken by camera at different times and correspondences displacements or changes. Apply Filter with Gaussian at different scales: This is done by just repeatedly filtering with the same Gaussian. Blob detectors is based on the Laplacian of the Gaussian (LoG). Given an input image f(x, y), this image is convolved by a Gaussian kernel

$$g(x,y,t) = \frac{1}{2\pi t^2} e^{-\frac{x^2+y^2}{2t^2}}$$

at a certain scale‘t’ to give a scale space representation L(x, y; t) = g(x, y, t) * f(x, y). Then, the result of applying the Laplacian operator

$$\nabla^2 L = L_{xx} + L_{yy}$$

is computed, which usually results in strong positive responses for dark blobs of extent $\sqrt{2}t$ and strong negative responses for bright blobs of similar size. To automatically capture blobs of different (unknown) size in the image domain, a multi-scale approach is therefore necessary. Now Subtract image filtered at one scale with image filtered at previous scale. Then do the Template matching,a basic method of template matching uses a convolution mask (template), tailored to a specific feature of the search image, which we want to detect. The convolution output will be highest at places where the imagestructure matches the mask structure, where large image values get multiplied by large mask values. Implementation:1. Pick a part of the search image to use as a template: Let the search image be S(x, y), where (x, y) represent the coordinates of each pixel in the search image. Let the template be T(x_t, y_t), where (x_t, y_t) represent the coordinates of each pixel in template.2. Then simply move the center (or the origin) of the template T(x_t, y_t) over each (x, y) point in the search image and calculate the sum of products between the coefficients in S(x, y) and T(x_t, y_t) over the whole area spanned by the template. As all possible positions of the template with respect to the search image are considered, the position with the highest score is the best position. This method is also referred to as 'Linear Spatial Filtering' and the template is called a filter mask.

IV. CONCLUSION

Sign languages are one of the main communication methods used by deaf people, but opposed to common thought, there is no universal sign language: every country or even regional group uses its own set of signs. The use of sign language in digitalsystemscanenhancecommunicationinbothdirections: animatedavatarscan synthesizesignalsbasedonvoiceortextrecognition;andsignlanguagecanbetranslated into text or sound based on images, videos and sensors input. The latest is the ultimate goal of this research, but it is not a simple spelling of spoken language, so that recognizing isolated signs or letters of the alphabet (which has been a common approach)isnotsufficientforitstranscriptionandautomaticinterpretation. Thesystem will provide the output in the form of text which is equivalent to the recognized sign language hand configuration. This system will ease and encourage the interaction of common people with the handicapped people since the common people would no longer be required to learn the various sign languages in order to

communicate with them. This system could be applied at various tasks, be it commercial or non-commercial, where there is involvement of handicapped people. Handicapped people can benefit from this system in their day to day life whenever they need to easily convey their message through their sign language to common people.

ACKNOWLEDGMENT

It gives us great pleasure in presenting the preliminary project report on 'IMAGE BASED SIGN LANGUAGE RECOGNITION ON ANDROID'. We would like to take this opportunity to thank our internal guide Prof. S. S. Raskar for giving us all the help and guidance we needed. We are really grateful to them for their kind support. Their valuable suggestions were very helpful. We are also grateful to Prof. N.F.Shaikh, Head of Computer Engineering Department,Modern Education Society's College of Engineering for her indispensable support, suggestions.

REFERENCES

1. M. Mohandes, M. Deriche, and J. Liu , "Image-Based andSensor-Based Approaches to Arabic Sign Language Recognition", IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS, VOL. 44, NO. 4, AUGUST 2014

2. Ravikiran J, Kavi Mahesh, Suhas Mahishi, Dheeraj R, Sudheender S, Nitin V Pujari, "Finger Detection for Sign Language Recognition", Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol I IMECS 2009, March 18 - 20, 2009, Hong Kong

3. Son Lam Phung, Member, IEEE,Abdesselam Bouzerdoum, Sr. Member, IEEE, and Douglas Chai, Sr. Member, IEEE "Skin Segmentation Using Color Pixel Classification: Analysis and Comparison", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 27, NO. 1, JANUARY 2005

4. Ashish Sethi, Hemanth S,Kuldeep Kumar,Bhaskara Rao N,Krishnan R,"SignPro-An Application Suite for Deaf and Dumb",IJCSET May 2012 Vol 2, Issue 5,1203-1206

5. William T. Freeman, Michal Roth, "Orientation Histograms for Hand Gesture Recognition",IEEE Intl. Wkshp. on Automatic Face and Gesture Recognition, Zurich, June, 1995

6. V.Nayakwadi, N. B. Pokale, "Natural Hand Gestures Recognition System for Intelligent HCI," International Journal of Computer Applications Technology and Research, 2013

7. Massimo Piccardi,"Backgroundsubtractiontechniques: areview",IEEEInternationalConferenceonSystems,ManandCybernetics,2004

8. Anne Kaspers, "BlobDetection".