Review paper of Character recognition using image processing

Pooja Sharma

Assistant Professor, D. A. V. College, Abohar, Punjab, India.
poojasharma8aug@gmail.com

Abstract— The paper proposes a system for gesture or hand images recognition to provide a easy method for communication between deaf-dump and vocal people it allow the user to deal with the system in a natural way. The hand can be classified into one of the four gesture classes or one of the four different movement directions. The aim of this paper to develop a system which converts the 32 combination of binary number 2^5 into decimals and then the decimals number converted into a sign language. In this paper we develop a system by using feature extraction method, HMM (Hidden Markov Model), canny edge detector and image processing tool.

Keywords- Hand gestures, human computer interaction, Graphical user interface.

I. INTRODUCTION

This paper introduce a hand recognizer with a flexible model that utilizes result from a character detector. Sign language is a form of manual communication which has developed as an alternative to speech amongst the deaf and vocally impaired. Although many deaf people can speak clearly (particularly those whose hearing impairment was acquired after early childhood) and can use skills such as lip-reading when communicating with hearing people, such methods of communication are generally inappropriate for communication within the Deaf community. Therefore the hands have become the primary means of communication within these communities. The hands are also widely utilised during communication between the vocal community, with gestures often used to augment speech. However such gestures bear very little similarity to the signs that make up sign language. Such data is accurate and easier to work with, however the gloves are expensive and cumbersome devices thus affecting the user friendliness of the system. Vision-based systems on the other hand, provide a more natural environment for capturing the gesture data. The flipside of this method is that working with images requires intelligent feature extraction and image processing techniques which might not be too accurate and may add to the computational complexity of the system[2]. The aim of this work is to provide signed contents for multimedia applications and to create a Spanish Sign Language (LSE) synthesizer that can be used by application designers and programmers who have no prior any knowledge of sign language (SL) synthesis or sign definition[4]. Some researchers mainly concentrated on the capture, recognition and classification of the gestures of sign language. Arabic[2], American[3], Span[4] and Korean sign languages contributed more in this area. Many more techniques and algorithms have been proposed in using a variety of methods based on the sensor fusion signal processing, image processing, and pattern recognition methods till now[5]. Hand gesture recognition research has gained much attention because of its applications for interactive human–machine interface and virtual environments. The common technologies are glove-based method and vision-based method: the Special glove-based devices have been developed to extract hand posture, some of which also allow recovering hand position and orientation. However, vision-based recognition of hand gestures is an extremely challenging interdisciplinary project due to the following three reasons: hand gestures are rich in diversities, multi meanings, and space–time varieties; human hand is a complex non rigid object; computer vision itself is an ill-posed problem[7]. A sign language is a small subset of possible forms of gesture communication. Sign languages are highly structured and most of them have symbolic natures i.e. the meaning is not transparent from observing the corresponding gesture. Sign language communication involves manual and non-manual signals; however, the main concern of this paper is manual signing. Some recent related systems are pointed out in the following [8]. The authors divided the task into three different problems:

• Automatic speech to text conversion

• Automatic translation of arbitrary English text into suitable representation of Americana Sign language

• Display of this representation as a sequence of signs using computer graphics techniques.

The authors used a virtual human (avatar) that they have already developed for other application for performing signs[1].

II. PROPOSED METHODOLOGY

In the proposed method, 32 combinations of binary images each representing _UP' and _DOWN' position of fingers shown in the Fig 1:-

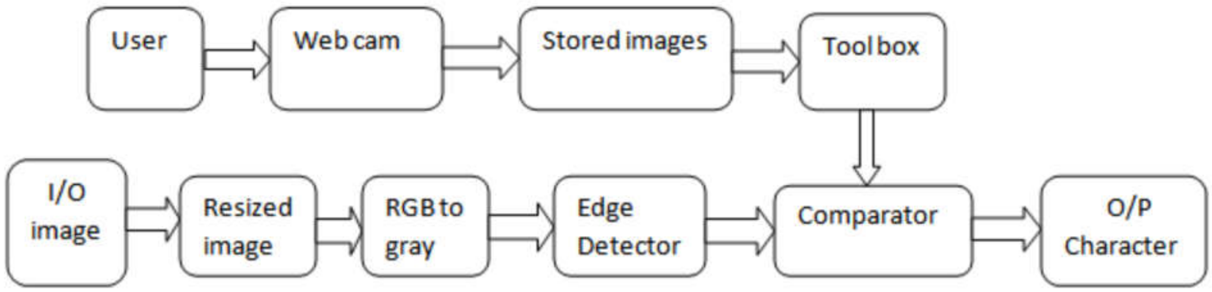

Fig 1:- shows the proposed method for character recognition

Each image is taken through web camera in different positions for 10 times. Thus we obtained 320 images. All these images are stored in a separate directory[5]. Then it uses the tool box, in tools box the parameters of each image are stored in it. Now firstly we take the input image then resized the image 160 x 120 then convert the resized image into RGB to gray. Now apply the canny edge detector on to the gray image then gray image converted into the canny image . now the canny image compare with tool box parameters images and give the result in form of character.

III. TECHNIQUES:-

A. Canny edge detector

To find edges, you can use the edge function. This function looks for places in the image where the intensity changes rapidly, using one of these two criteria:-

- Places where the first derivative of the intensity is larger in magnitude than some threshold.
- Places where the second derivative of the intensity has a zero crossing

edge provides a number of derivative estimators, each of which implements one of the definitions above. For some of these estimators, you can specify whether the operation should be sensitive to horizontal edges, vertical edges, or both. edge returns a binary image containing 1's where edges are found and 0's elsewhere. The most powerful edge-detection method that edge provides is the Canny method. The Canny method differs from the other edge-detection methods in that it uses two different thresholds (to detect strong and weak edges), and includes the weak edges in the output only if they are connected to strong edges. This method is therefore less likely than the others to be fooled by noise, and more likely to detect true weak edges. The Canny edge detector is an edge detection operator that uses a multi-stage algorithm to detect a wide range of edges in images. The Canny Edge Detector is one of the most commonly used image processing tools, detecting edges in a very robust manner. It is a multi-step process, which can be implemented on the GPU as a sequence of filters. An edge in an image may point in a variety of direction so The Canny algorithm uses four filters to detect horizontal, vertical and diagonal edges in the blurred image.From this the edge gradient and direction can be determined:

$$G = \sqrt{{G_x}^2 + {G_y}^2}$$
$$\Theta = \arctan\left(\frac{G_y}{G_x}\right).$$

The goal of edge detection is to mark the points in an image (sign image) at which the intensity changes sharply. Sharp changes in image properties usually reflect important events and changes in world properties[3]. We used Canny edge detection technique because it provides the optimal edge detection solution[3]. Here in the proposed system an ‘optimal’ edge detector means:

- good detection—the algorithm should mark as many real edges in the image as possible.
- good localisation—edges marked should be as close as possible to the edge in the real image.
- minimal response—a given edge in the image should only be marked once, and where possible, image noise should not create false edges.

B. Feature point extraction

One of the most important requirements for a feature point is that it can be differentiated from its neighboring image points. If this were not the case, it wouldn't be possible to match it uniquely with a corresponding point in another image. Therefore, the neighborhood of a feature should be sufficiently different from the neighborhoods obtained after a small displacement. First, to represent the motion, the forward prediction error of consecutive sentences is computed. This can be achieved by computing pixel-based differences of successive images. It can be justified that the difference between two images of similar background results in an image that only preserves the motion between the two images. These image differences are then converted into binary images by applying an appropriate threshold. Subsequently, zig-zag scanning is used to retain only a required number of frequency coefficients. This process is known as Zonal coding. The number of coefficients to retain or the DCT cutoff is elaborated upon in the experimental results section. These coefficients obtained in a zig-zag manner make up the feature vector that is used in training the classifier[2].

Feature point extraction technique can be classified into six sub-categories[3]:

1. Template matching.
2. Feature extraction and analysis.
3. Active shape models ‘‘smart snakes’’.
4. Principal component analysis.
5. Linear fingertip models.
6. Causal analysis.

The difference of the height between this point and fixed reference point (x_0, y_0) colour is set in red colour at the bottom-up approach. This is calculated by using *Euclidean Distance* formula[5].

C. HMM (Hidden Markov Model)

A **hidden Markov model** (**HMM**) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (*hidden*) states. An HMM can be considered as the simplest dynamic Bayesian network. In simpler Markov models (like a Markov chain), the state is directly visible to the observer, and therefore the state transition probabilities are the only parameters. In a *hidden* Markov model, the state is not directly visible, but output, dependent on the state, is visible. Each state has a probability distribution over the possible output tokens. Therefore the sequence of tokens generated by an HMM gives some information about the sequence of states. Note that the adjective 'hidden' refers to the state sequence through which the model passes, not to the parameters of the model; even if the model parameters are known exactly, the model is still 'hidden'. Hidden Markov models are especially known for their application in temporal pattern recognition such as speech[1], handwriting, gesture recognition[6][7], part-of-speech tagging, musical score following, partial discharges and bioinformatics. Hidden Markov Model (HMM) is used for recognition of speech signal from the user and translated to cue symbols for vocally disabled people. The proposed task is a complementary work to the ongoing research work for recognizing the finger movement of a vocally disabled person, to speech signal called “Boltay Haath”[1]. Many approaches have been proposed that combine Dynamic time wrapping or Hidden Markov Models (HMMs) with discrimination classifier for recognizing speech, handwriting or Sign Language[5]. Signs are represented by using sets of view models, and then are matched to stored gesture patterns using dynamic time warping developed a non-HMM-based system, which can recognize 28 different gestures in front of complex backgrounds[7]. HMM is used to find the most probable path through the network and log likelihood of the sign is calculated. To identify a sign, a decision procedure is carried out as follows: if one feature is considered, the sign with maximum log likelihood is chosen. If two features are considered, each feature is treated independently as above and the final sign is obtained from the two chosen signs using maximum log likelihood. If the three features are considered, each feature is treated also independently and a voting system is used to select the final sign depending on the number of features choosing it (occurrences of the sign). However if the three features chose three different signs their maximum log likelihood is again employed[8].

D. Comparison with other works

The average sentence recognition rate of 75% and word recognition rate of 94% are obtained using a natural vision-based system with no restrictions on signing such as the use of gloves[2]. The proposed system proved to be robust against changes in gestures, position, size and direction. This is because the extracted features method used proved to be translation, scale, and rotation invariant. The proposed system was able to reach a recognition rate of about 98.5% for training data and 80% for testing data[3]. The recognition rate of the eSign system is 81% for isolated signs and 61% for complete sentences. determination of the recognition rate for complete sen- tences in the eSign system involved evaluation of both the synthesis and the machine translation system, thus a direct comparison cannot be presented. Both systems show a decrease in recognition rate when complete sentences are compared to isolated signs. It may be that the larger difference found in the eSign system is due to only the machine translation[4]. When it is tested through static image recognition method, except 10 images out of 320 were given 96.87% sign recognition accuracy[5]. the new method is invariant to the scale and plane movement for the normalization to the hand region. But the rotation is sensitive to the recognition result. The total recognition rate is 92.2%[7]. the PCA has been used for dimensionality reduction (not as a descriptor as it has been used here) also his features combinations were PCA only, PCA and hand position, PCA and hand velocity, PCA and hand trajectory and two model combination (PCA and hand velocity with PCA and hand trajectory), at which the error rate is 17.9%. Both works indicate that the vision based approach with sensible features can be relied upon to mitigate the challenges involved in ASL recognition[8].

IV. CONCLUSION

This paper presents an aproch towards image recognition for vocally disabled people. the system proposed could efficiently recognize the image using HMM and generate an equivalent character. This proposed work is very important for deaf and dump people to communicate with vocal people.

REFERENCES

[1] Ibrahim Patel, Dr. Y. Srinivasa Rao , “ Technologies Automated speech recognition approach to finger spelling ”, IEEE 2010 second international conference on computing, communication and networking.

[2] Khaled Assaleh1, Tamer Shanableh2, Mustafa Fanaswala1, Harish Bajaj1, and Farnaz Amin1,” Vision-based system for Continuous Arabic Sign Language Recognition in user dependent mode”, Proceeding of the 5th International Symposium on Mechatronics and its Applications (ISMA08), Amman, Jordan, May 27-29, 2008.

[3] Qutaishat Munib *, Moussa Habeeb, Bayan Takruri, Hiba Abed Al-Malik, “American sign language (ASL) recognition based on Hough transform and neural networks”ww w.elsevier.com/locate/eswa.

[4] Fernando Lo´ pez-Colino n, Jose´ Cola´ s 1, “Spanish Sign Language synthesis system”, journal homepage: www.elsevier.com/locate/jvlc.

[5] P. Subha Rajam [a], G. Balakrishnan [b], “Recognition of Tamil Sign Language Alphabet using Image Processing to aid Deaf-Dumb People”, International Conference on Communication Technology and System Design 2011.

[6] Cristina Manresa, Javier Varona, Ramon Mas and Francisco J. Perales, “Real –Time Hand Tracking and Gesture Recognition for Human Computer Interaction”, Electronic Letters on Computer Vision and Image Analysis 0(0):1-7, 2000.

[7] Xiaolong Teng, Bian Wu, Weiwei Yu, Chongqing Liu, “A hand gesture recognition system based on local linear embedding” , Journal of Visual Languages and Computing 2005 Published byElsevier Ltd.

[8] Mahmoud M. Zaki , Samir I. Shaheen, “Sign language recognition using a combination of new vision based features”, Pattern Recognition Letters 32 (2011) 572–577, journal homepage: www.elsevier.com/locate/patrec.