Machine Learning Model for Sign Language Interpretation using Webcam Images

Kanchan Dabre
Department of Computer Engineering
Sardar Patel Institute of Technology
Student of M.E.(Computer)
Mumbai, India
kanchandabre@gmail.com

Surekha Dholay
Department of Computer Engineering
Sardar Patel Institute of Technology
Mumbai, India
surekhadholay@yahoo.co.in

***Abstract*— Human beings interact with each other either using a natural language channel such as words, writing, or by body language (gestures) e.g. hand gestures, head gestures, facial expression, lip motion and so on. As understanding natural language is important, understanding sign language is also very important. The sign language is the basic communication method within hearing disable people. People with hearing disabilities face problems in communicating with other hearing people without a translator. For this reason, the implementation of a system that recognize the sign language would have a significant benefit impact on deaf people social live.**

In this paper, we have proposed a marker-free, visual Indian Sign Language recognition system using image processing, computer vision and neural network methodologies, to identify the characteristics of the hand in images taken from a video trough web camera. This approach will convert video of daily frequently used full sentences gesture into a text and then convert it into audio. Identification of hand shape from continuous frames will be done by using series of image processing operations. Interpretation of signs and corresponding meaning will be identified by using Haar Cascade Classifier. Finally displayed text will be converted into speech using speech synthesizer.

Keywords— Indian Sign Language (ISL),Computer Vision(CV).

I. INTRODUCTION

A sign is a form of non verbal communication done with body parts, hand shapes, positions and movements of the hand, arms, facial expressions or movements of the lips and used instead of oral communication. Most people use both words and signs during communication. A sign language is a language that uses signs or action to communicate instead of sounds. According to the above definition, a sign language has three major components [1].

First important component is finger-spelling which means for each letter of the alphabet there is a corresponding sign. This type of communication is used mainly for spelling names sometimes for spelling the location names. Sometimes this can be used for expressing words for which no signs exist or for emphasizing or clarifying a particular word [2].Second vital component of any sign language is word level sign vocabulary which means for each word of the vocabulary there is a corresponding associated sign in the sign language. The most commonly used type of communication between people with hearing disabilities in combination with the facial expression is this type. Third essential component in sign communication is non-manual-features. This type of communication involves facial expressions, tongue, mouth, eyebrows and body position. Among all these components most used form of sign language by deaf community in reality is word level sign language. Hence this paper throws more light on frequently used daily words or sentences and their interpretation by Sign Language Interpreter System.

The most important part of Indian Sign Language (ISL) is it does not include grammar. Like the spoken languages and dialects, the sign language has designed differently depending on the region and culture. Sign Language used in India is different than American Sign Language. A government website www.indiansignlanguage.org has been launched for empowering the deaf, which presents a huge database of Indian Sign Language (ISL) signs [3]. Very less work has been done on Indian Sign Language so far. That is why it would be very useful to have an automatic Indian Sign Language Interpretation system [4].

This system will definitely contribute to society via helping towards physically Indian handicap people. This system will act as an auxiliary tool for a deaf-mute to communicate with ordinary people through computer. Thus the major objective of this system is to make possible the communication between deaf people and the rest of the world in daily life will come into reality. Since there are very few proficient sign language tutors at schools for the deaf, the teaching and learning process is lagging behind. In such cases this system can be used for sign language education purpose, where any person can learn or practice sign language [5].

Sign language recognition interfaces can also be used as a natural communication channel in between humans and machines and give rise to applications such as hardware-free remote controls, human-computer interaction in virtual reality, gaming and other human welfare applications. Other benefits of such human computer interface using gesture applications are replace mouse and keyboard i.e. virtual keyboard or virtual mouse, pointing gestures, navigate in a virtual environment, to catch and manipulate virtual objects, interact with a 3d world , no physical contact for computer interaction, communicate at a distance.

978-1-4799-2494-3/14/$31.00 ©2014 IEEE

II. LITERATURE SURVEY

Vision based hand gesture recognition is proposed by Ilan Steinberg and Tomer London [1] using supervised learning algorithm .This system used multiclass classifier having integrated with several binary classifier such as Support Vector Machine Algorithm(SVM) [5] to train and classify hand gestures. Initially acquired images were preprocessed using color normalization, skin detection, blob analysis filtering and feature calculation techniques. Image acquisition, hand segmentation, feature extraction and then classification based on supervised feed forward backpropogation algorithm [2][4]was used by Adithya, Vinod and Usha Gopalkrishnan for hand feature extraction having average recognition rate of 91.11%.

Research on Indian Sign Language made by Pravin Futane and Dr.Rajiv [3] used feature extraction based on shape and geometry feature and lastly learning by General Purpose Fuzzy MinMax(GFMM)neural network. Unsupervised Feature Learning Algorithm is used in designing softmax classifier to understand American Sign Language, an efforts made by Justin Chen, Debabrata, Rukmani Sundaram[6].This approach uses skin modeling using 2D Gaussian curve over 600 training dataset using kinect camera, which is a special camera that supplies dept information was used to identify hand gestures for 40 iteration to obtain estimate weights for classifier. A Hidden Markov Model (HMM) [7] being designed for data occurring over time was used by Nathan and James to correctly classify a feature vector of unknown sign. Feature vector was calculated using Hue Saturation (HSI) model, centering method and grid extraction method. Peter O'Donovan from Canada applied Restricted Boltzmann Machines (RBM) to model gestures and also provided comparison with classical neural network and k-Nearest Neighbors method [8].

Haar-like Algorithm is used for getting the region of interest from hand image preceding preprocessing techniques involving skin detection and size normalization. Fourier transformations is applied to form the feature vectors is then classified by K- Nearest Neighbor (KNN) algorithm to determine Arabic Sign Language (ArSL)[9].

The Other computer vision methods used for hand gesture recognition include specialized mappings architecture principal component analysis, fourier descriptors, neural networks [10], orientation histograms, particle filters etc. The method proposed in this paper used Haar Cascade Classifier that translates word level sign vocabulary in Indian Sign Language to textual as well as audio form.

The rest of the paper is organized as follows. In section III, the system architecture is discussed which follows the preprocessing and classification phase description. In section IV, development environment and experimental setup is explained. In section V, classification results and statistics of our system is presented. In remaining sections conclusion and future scope are presented.

III. SYSTEM ARCHITECTURE

The Sign Language Interpretation System works in two stages as shown in Figure1. The first is the preprocessing phase i.e. image processing phase, where the hand shape and other distinguishable features are extracted from the image using background subtraction, blob analysis, filtering and noise removal, grayscale conversion, brightness and contrast normalization, scaling and several other image processing techniques.

The second stage involves the classification of an image into given many different possible gestures using Haar Cascade Classifier, where this classifier is trained on a given training set that contains samples of the different gestures. This training sample images are taken from several different angle and captured in different lightening conditions. Training dataset consist of positive, negative as well as test sample database. Positive samples are those image samples which contain perfect hand gesture where as in negative sample images the required gesture is absent or only background details are available, no hand moment is present. These datasets are mostly used in training part of classification phase. The test sample dataset can be used in testing part of classification phase.

After the training of setup is done, the system is now ready to interpret input images from the videos. A database of haar cascade classifier which denotes different signs is then observed. The classifier which produces the highest probability is then chosen as the most possible interpretation of the sign. Classification or ANN phase which follows the text to speech conversion this phase is known as speech synthesis phase.

A. Preprocessing Phase

This phase involves extracting frame from video stream and performing image processing steps to extract features from the image by performing background subtraction, Blob analysis, noise reduction, gray scale conversion, brightness normalization and scaling operation one by one.

1) Background Subtraction: This phase involves removing unwanted background details from captured image frame from video stream. and extracting only hand sign to perform image processing steps.

2) Blob Analysis: A blob is a region having same properties and pixel values which constant or varies within a prescribed range. This step discovers region of interest for further processing by finding all connective parts of the frame and choose the biggest (largest area) amongst them (since the hand is the largest area suspected of being a hand). Blob analysis is applicable in the field of object recognition or object tracking.

3) Noise Reduction: Noise reduction is meant to filter the discontinuity and noise by using smooth Gaussian filter. This filter removes the noise by smoothening operation. The Gaussian kernel size used for this filter is 3.

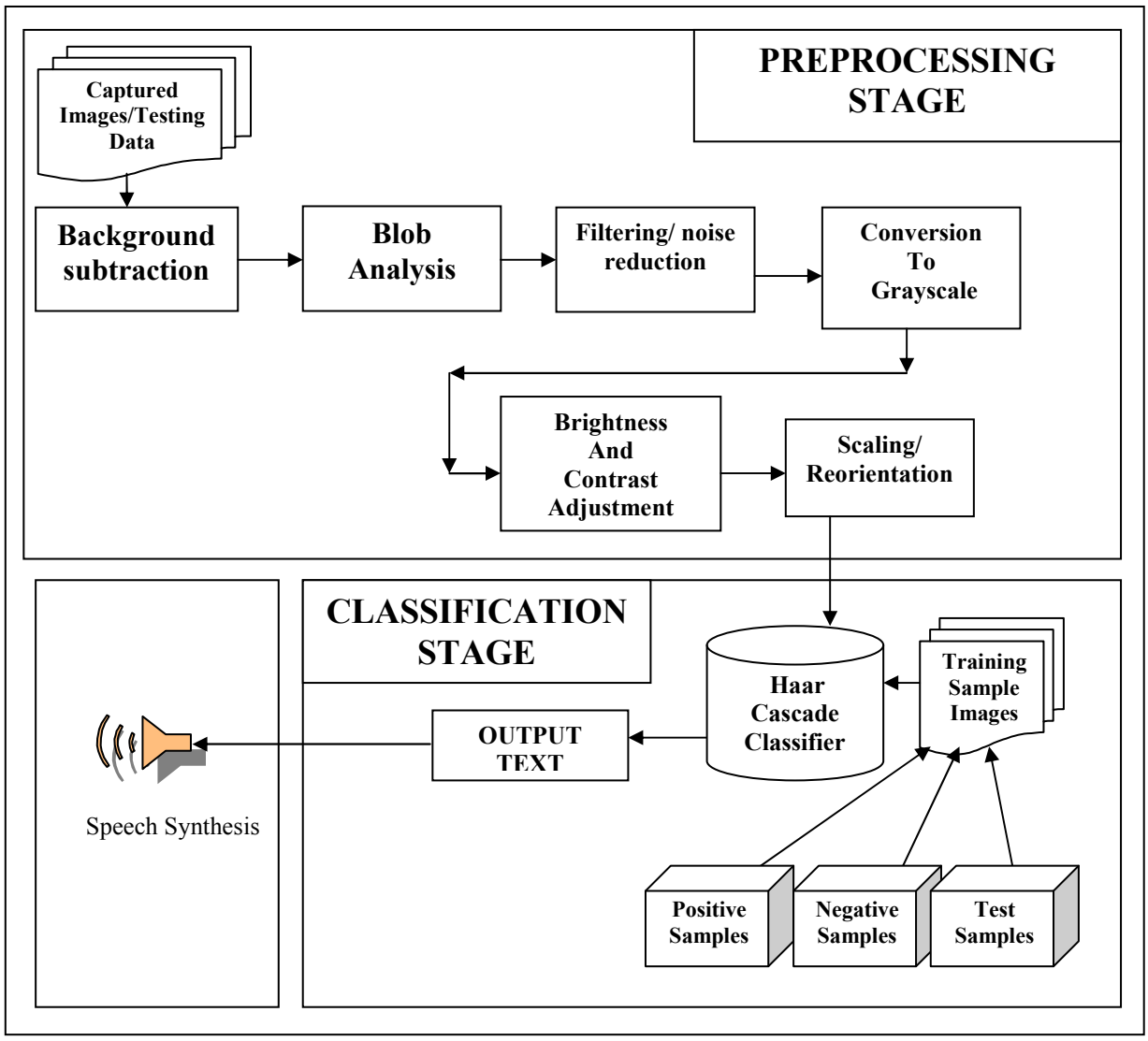

Fig. 1. Sign Language Interpreter Architecture

4) GrayScale Conversion: This step converts color image into grayscale image which helps in further calculations on pixel operations and interrelating signs. Memory space in terms of bits required to store grayscale image are lesser than the bits required storing color image.

5) Brightness and Contrast Normalization: Images acquired in low illumination have close contrast values hence there is a need to adjust pixel intensity values. Histogram equalization is performed in order to adjust and normalize brightness and contrast of processing frame.

6) Image Scaling: Image scaling is done to reduce the computational effort needed for image processing. Every image will be scaled to 45*45 sizes for further processing.

B. Classification phase

This phase involves application of haar cascade algorithm to correctly classify the extracted feature. Input to the segmentation block is processed resized images. Output of this phase is correctly classified word/sentence in textual format. Classification phase is further divided into training and testing stages.

1) Training stage: Haar Cascade Classifier is trained using 500 positive, 500 negative and 50 test image samples of each gesture. These images are stored in their respective folders. These images especially positive samples are collected from different people with different hand shape, size and color and different lightening condition in various angles. Accuracy of recognition can be improved by locating area of interest in each sample images. This can be accomplished by drawing a box around the region of interest i.e. hand shape. The co-ordinates of region of interest are then analyzed to measure the contrast between each of these images. This stage will enable to build required cascade and find thresholds after analyzing each coordinates of hand sign. Classifier uses Haar like feature like

edge, line and center surround features to be trained using simple Haar function given in formula as in (1).

$$H(t)= \begin{cases} 1 & 0 \le t \le 1/2 \\ -1 & 1/2 \le t \le 1 \\ 0 & otherwise \end{cases} \quad (1)$$

To achieve perfection in results it is recommended to use maximum no of samples. Training classifier in order to interpret the different signs based on the features learnt by pre-processing takes longer time. Training procedure executes only once, where Haar Cascade Classifier is trained for particular sign. After training is over system is ready to interpret signs in the video using web camera.

2)Testing Stage: Once training is over, classifier is now well trained to distinguish between different signs. Testing is performed on the live video through web camera. Output of this phase is in text form.

C. Speech Synthesis phase

This phase converts textual information into audio format. Converting speech into text is also called as voice synthesis, which can be accomplished by including "System. Speech" reference in the project. Microsoft .NET framework provides "System.Speech.Synthesis" library. It contains methods such for both speech synthesis and recognition. The input for this stage is text related to recognized sign and output is audio form of text with male or female voice format.

IV. DEVELOPMENT ENVIRONMENT

The input for the Sign Language Interpretation System is video captured through the web camera. The frames from the input video are extracted. The extracted input images will be processed by the system using several image processing techniques. The processed output is then classified using Haar Cascade Classifier to generate segmented output.

The hardware requirement of the system is low-resolution web camera or mobile-integrated cameras or laptop inbuilt camera and the Software Requirement is Emgu CV which is a cross platform .Net wrapper to the OpenCV image processing library.

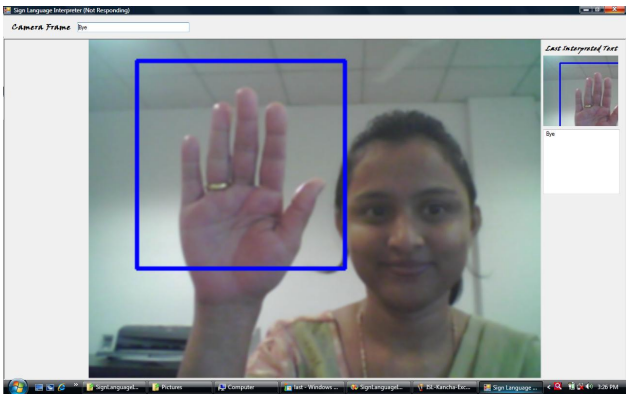

Fig. 1. Sign Language Interpreter Interface reconising sign"Bye".

The Simulation environment is shown in Figure 2 and Figure 3 which identifies sign "Bye" and "Excuse Me" made by user. This interface includes input video, corresponding current frame interpretation in text above the video being played, right hand side shows current recognized frame and history of last interpreted signs same time the sound of text is also audible.

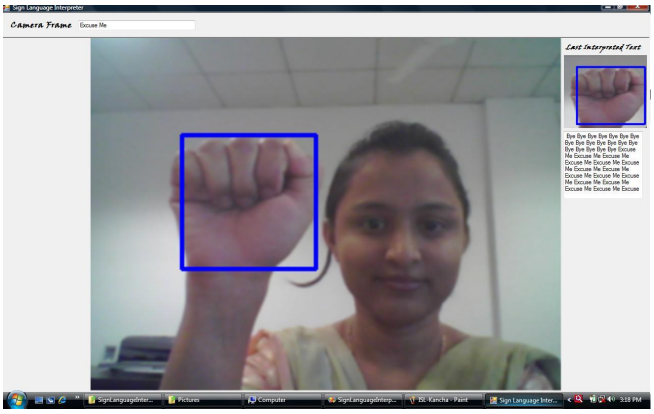

Fig. 2. Sign Language Interpreter Interface recognising sign"Excuse Me"

V. EXPERIMENTAL RESULTS

The expected result of this system is fully segmented words or sentences in text as well as audio format. The Sign Language Interpretation will predict the Indian Sign Language. The Result set for Indian Sign Language is shown in Table I, where every hand gesture depicts different words. This software is expected to train and classified only few sentences listed below. Using a computer-mounted video camera as the sensor, the system can reliably interpret gesture set at 18 to 21 frames per second on a Personal Computer.

TABLE I. TRAINING /RESULT SET

Sr.No	Words/Sentences used for training
1	Bye
2	What is your name?
3	I am going out now.
4	Excuse Me
5	I am Hungry.

For result analysis testing is performed on two distinct signs for only one second and the results are evaluated to calculate the performance and accuracy of system.

TABLE II. EXPERIMENTAL RESULT

Signs/ Classes	Frames per Second	Correctly Classified Signs	Not Classified Signs	Percentage Accuracy
Bye	21	21	0	100%
Excuse Me	20	17	3	85%
Average Accuracy	41	38	3	92.68%

Experimental results of Sign Language Interpretation System are shown in Table II , where it shows that few frames are not classified it happens if hand is in very high motion or sign is not made correctly system does not classify sign. Experimental result clearly shows the success of Indian Sign Language Interpreter System having average accuracy rate is more than 92.68%. System gives correct output even in the presence of multiple hands and different lightening situations. System efficiently outlines each correct hand gesture if it detects multiple signs in single frame. This result is satisfactory enough for real time use of this system.

VI. CONCLUSION

The prediction using Haar Cascade Classifier for Indian Sign Language recognition is presented in this paper. The system interprets signs made in front of computer in textual and audio format. This method proves to be much faster than any other former methods like support vector machine, hidden markov model, backpropagation algorithm and k-nearest neighbor method. The convergence rate is also faster it improves the speed and accuracy rate of sign language interpretation within system model for real time system.

The biggest benefit of this system is that the training setup can be done before the real use of system. Therefore it reduces processing power and increases efficiency by interpreting faster during testing time. According to the statistics system processes almost all frames correctly with the average accuracy speed of 92.68%.This system can be expanded further by adding whole dataset of ISL gestures.

VII. FUTURE SCOPE

Introduction of facial expression and whole body gesture in identifications of sign made by deaf people will contribute significantly to this Sign Language Interpretation System.

Designing a large vocabulary for sign language interpretation system for a sign language is challenging task as training every sign is time consuming considering every aspect of moment, hand shape, size ,hand color, background illumination etc.

System can be made more versatile working on mobile platform where person making sign while moving or walking and also implementation on mobile environment where person and system both are in motion for example travelling through train.

The speech synthesis phase of sign recognition process sometimes gives delayed response as speech synthesizer object requires time to complete the speech where as making signs are faster than speaking. This happens only in case of signing big sentences, can be eliminated by slowing down the speed of signing and has a future scope of improvement.

REFERENCES

[1] Ilan Steinberg, Tomer M. London, Dotan Di Castro, " Hand Gesture Recognition in Images and Video", Technion-Israel Institute of Technology, 2003.

[2] Adithya V., Vinod P., Usha Gopalakrishnan, "Artificial Neural Network Based Method for Indian Sign Language Recognition", IEEE Conference on Information and Communication Technologies (ICT 2013), JeJu Island, pp. 1080-1085, April 2013.

[3] P. R.Futane, Dr. R. V. Dharaskar, "Video Gestures Identification And Recognition Using Fourier Descriptor And General Fuzzy Minmax Neural Network For Subset Of Indian Sign Language," IEEE Conference on Hybrid Intelligent Systems (HIS), 12th International Conference, Pune, pp. 525-530, Dec 2012.

[4] Corneliu Lungociu, "Real Time Sign Language Recognition Using Artificial Neural Networks," Studia Univ. Babes_Bolyai, Informatica, Volume LVI, Number 2011.

[5] J. Jones. (1991, May 10). Networks (2nd ed.) [Online]. Available: http://www.atm.com R. Kurdyumov, P. Ho, J. Ng. (2011, December16) *Sign Language Classification Using Webcam Images* [Online]. Available:http://cs229.stanford.edu/.../KurdyumovHoNg-SignLanguageClassificationUsi...

[6] J. K. Chen, D. Sengupta, R. R. Sundaram, IT University of Copenhagen *Sign Language Gesture Recognition with Unsupervised Feature Learning*[Online].Available:http://cs229.stanford.edu/.../ChenSenguptaSundaram-SignLanguageGestureRe...

[7] Nathan L. Naidoo, James Connan(2009), *Gesture Recognition Using Feature Vectors*, University of Western Cape, Available:http://connan.co.za/jconnan/publications.

[8] Peter O'Donovan, "Static Gesture Recognition with RestrictedBoltzmann Machines,"University of Toronto, Canada, 2007

[9] Nadia R. Albelwi, Yasser M. Alginahi, "Real-Time Arabic Sign Language (ArSL) Recognition," Taibah University, © ICCIT 2012, pp. 497-501.

[10] Jonathan C. Rupe, "Vision-Based Hand Shape Identification for Sign Language Recognition," MS in CE Thesis. RIT. 2005.