A Translation Device for the Vision Based Sign Language

M.Akilan[1] K. Ganesh Kumar[2] R. Arvind[3] B. N. Ajai Vikash[4] P. Dass[5]

[1,2,3,4,5] Department of Electronics and Communication Engineering

[1,2,3,4,5] Saveetha School of engineering, Saveetha University, Chennai, India

Abstract— The Sign language is very important for people who have hearing and speaking deficiency generally called Deaf and Mute. It is the only mode of communication for such people to convey their messages and it becomes very important for people to understand their language. This paper proposes the method or algorithm for an application which would help in recognizing the different signs which is called Indian Sign Language. The images are of the palm side of right and left hand and are loaded at runtime. The method has been developed with respect to single user. The real time images will be captured first and then stored in directory and on recently captured image and feature extraction will take place to identify which sign has been articulated by the user through SIFT(scale invariance Fourier transform) algorithm. The comparisons will be performed in arrears and then after comparison the result will be produced in accordance through matched key points from the input image to the image stored for a specific letter already in the directory or the database the outputs for the following can be seen in below sections. There are 26 signs in Indian Sign Language corresponding to each alphabet out which the proposed algorithm provided with 95% accurate results for 9 alphabets with their images captured at every possible angle and distance.

Key words: SIFT (scale invariance Fourier transform), Deaf, Mute, Sign Language.

I. INTRODUCTION

One can create life but no one has the rights to destroy it. The saying goes like this. Even though some human are physically challenged, but this doesn't mean they have to be devoid of all worldly pleasures, fun, etc., Most of the current sign language recognition system use specialized hardware like hand gloves with sensors. Here in our project we are going to overcome the restrictions and convert action into text without using gloves. It has the capability of capturing human hand signals and produces text output accordingly. Our project has the capability of detection, capturing and processing of hand gestures irrespective of the background. This project is sure to create revolution in the field of image processing and is guaranteed to enlighten the lives of dumb people. Although deaf, hard of hearing and hearing signers can fully communicate among themselves by sign language, there is a large communication barrier between signers and hearing people without signing skills. The manual features obtained from the tracking are passed to the statistical machine translation system to improve its accuracy. In this work the problem is tackled by means of the typical speech decoder in a serial architecture with an action-to-text translation model in order to evaluate all the alternatives. The community of the hearing marred people is very large in every corner of the world. This community cannot understand the normal speaking language. Signed communication is characterized by the use of hand gestures. Sign language translator makes it easier to understand the sign language. For the users, the translator should be easy and natural to use.

II. SCOPE OF THE PROJECT

Our project aims to bridge this gap by introducing an inexpensive computer in the communication path so that the sign language can be automatically captured, recognized and translated to text for the benefit of deaf/dumb people. Hand signals must be analyzed and converted to textual display on the screen for the benefit of the hearing impaired.

III. EXISTING METHODS

The author Adithya published his paper "Artificial Neural Network based method for Indian Sign Language Recognition" on 2013 in IEEE Conference and he proposes that, "Sign Language is a language which uses hand gestures, facial expressions and body movements for communication. A sign language consists of either word level signs or finger spelling. It is the only communication mean for the deaf-dumb community. But the hearing people never try to learn the sign language. So the deaf people cannot interact with the normal people without a sign language interpreter. This causes the isolation of deaf people in the society. So a system that automatically recognizes the sign language is necessary. The implementation of such a system provides a platform for the interaction of hearing disabled people with the rest of the world without an interpreter. In this paper, we propose a method for the automatic recognition of finger spelling in Indian sign language. The proposed method uses digital image processing techniques and artificial neural network for recognizing different signs".

The below given figures were some of the actions used for some of the existing methods.

Fig. 3.1: Hand Gesture

All rights reserved by *www.ijsrd.com*

The first sign should be the same for at least three consecutive partial sign sequences. With these restrictions, a 40% delay reduction is achieved without affecting the translation process performance. One important aspect to be considered in a speech to sign language translation system is the delay between the spoken utterance and the animation of the sign sequence. This delay is around 1–2 s and it slows down the interaction.

IV. PROPOSED METHOD

The proposed method introduces the Sign Language (SL) recognition, although has been explored for many years, is still a challenging problem for real practice. The complex background and illumination conditions affect the hand tracking and make the Sign Language recognition very intricate.

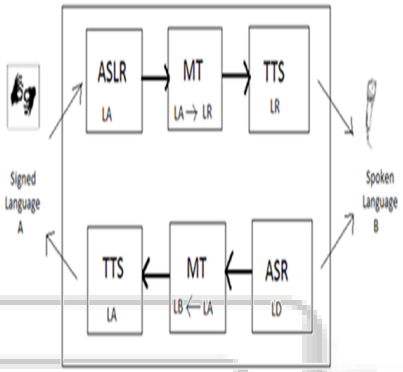

Fig. 4.1: Translation

The Sign performed by human were sensed first through the sign language recognition device installed with MATLAB coding. They were then converted to Machine language through Machine Translation and then finally to text by Sign to Text.

And the reverse process is carried such that, the text typed by another person will be recognized and then translated to machine language through machine translation (MT). Finally it reaches them as Sign with the help of Text to Sign Translation (TTS). But the reverse process will be our future work while we implement for Mobile to Mobile Communication for Deaf and normal people.

V. BLOCK DIAGRAM

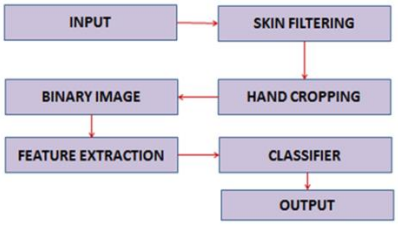

Fig. 3.2: Process Involved

VI. EXPLANATIONS OF THE BLOCK DIAGRAM

The input image or actions was sensed and given to skin filtering there the sensed images will undergo filtering to remove unwanted features. Then the output of skin filtering was given to hand cropping section where the needed portion alone cropped and that was sent to the binary image section where the analog signals are converted to digital pulses(binary form) i.e., continuous signals are converted into discrete signals. Then the output of binary image was given to feature extraction block where the needed features alone get extracted and the output was given to classifier. It classifies the output based upon its recognition and display the text based on its recognition.

The general categories of tools required are classified into two and they are as follows:

- Hardware Requirements, and
- Software Requirements.

VII. HARDWARE REQUIREMENTS

The hardware requirements are as follows

- Web camera
- Vision based computer

A. Web Camera

The term web camera is a combination of "web" and "video camera". The purpose of webcam is, not surprisingly to broadcast video on the web, webcams are typically small cameras that either attach to a user's monitor or sit on a desk or through wireless. Most webcams connect to the computer through USB, though some use a Fire wire connection. Webcams typically come with software that allows the user to record video or stream the video on the web. Since the streaming video over the internet requires a lot of bandwidth, the video stream is typically compressed to reduce the "choppiness" of the video.

The use of webcam is limitless. The webcam basically works by capturing a series of digital images that are transferred by the computer to a server and then displayed to the hosting page. Webcams vary in their capabilities and features and the variance are reflected in price. The maximum resolution of a webcam is also lower than that of handheld cameras, since higher resolution would be reduced anyway. For this reason the webcams are relatively inexpensive.

Fig. 3.3: Webcam

A webcam is a video camera that feeds its image in real time to a computer or a computer network. . They have also become a source of security and privacy issues, as some built-in webcams can be remotely activated through spyware.

1) Early Development

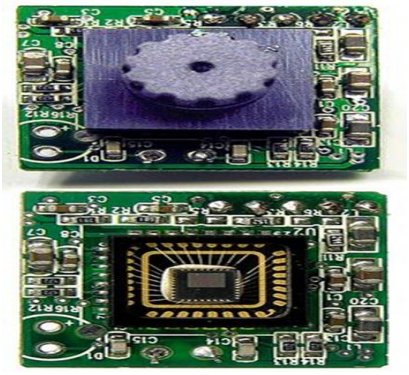

Fig. 3.4: Webcam Circuit

All rights reserved by www.ijsrd.com

First developed in 1991, the webcam was pointed at the Trojan Room Coffee Pot in the Cambridge University Computer Science Department. The oldest webcam still operating is Fogcam at San Francisco University which has been running continuously.

Webcam typically include a lens (shown at top), an image sensor (shown at bottom) and supporting circuitry and may also include a microphone for sound. Various lenses are available. The most common in consumer grade webcam is being a plastic lens that can be screwed in and out to focus the camera. Fixed focus lens are also available in which the adjustment cannot be made. Image sensor can be CMOS or CCD. The support electronics is used to read the image from the sensor and transmit it to the host compilers.

2) Circuit for Web Camera

Fig. 3.5: Webcam Connections

B. Vision Based Computer

Computer vision is a field that includes methods for acquiring, processing, analyzing and understanding images and, in general, high-dimensional data from the real world in order to produce numerical or symbolic information. Computer vision has also been described as the enterprise of automating and integrating a wide range of processes and representation for vision based perception. As a Scientific discipline, Computer Vision is concerned with the theory behind artificial system that extracts information from the images.

1) Applications of Computer Vision

In most practical Vision Based Computers are pre-programmed to solve a particular task and some of the application areas for computer vision are as follows:

- Controlling Process. For example industrial robot
- Navigation. Examples are autonomous vehicle or mobile robot and etc.,

Sub-domains of computer vision include Scene Reconstruction, Event Video Tracking, Object Recognition, Motion estimation to image Restoration.

The field most closely related to Computer Vision is Image Processing. Image Processing and Image Analysis tends to focus on 2D images, how to transform one image to another. This characterization implies that the image processing/analysis neither require assumptions nor produce interpretation about the image content.

2) Typical Characteristics of Computer Vision

Each of the application areas employs a range of computer vision tasks; more or less well-defined measurement problems or processing problems, which can be solved by variety of methods and some of them are

C. Recognition:

The classical problem in computer vision, image processing is that of determining whether or not the image data contains some specific object feature or activity.

Different varieties of recognition problem are object recognition, identification and detection. Those can be explained as follows:

- Object Recognition: One or several pre-specified or learned objects or object classes can be recognized, usually together with their 2D position.
- Identification: An individual instance of an object is recognized. For example object includes identification of a specific person's face or finger print.
- Detection: The image data are scanned for a specific condition. It is based on relatively simple and fast computation is some time used for finding smaller regions of interesting, image data which can be further analyzed by more computationally demanding techniques to produce a correct interpretation.

D. Image Restoration:

The aim of image restoration is the removal of noise (sensor noise, motion blur, etc.,) from images. The simplest possible approach for noise removal is the use of various types of filters like low pass filters, median filters. Most sophisticated methods assume a model of how the local image structures look like, a model which distinguishes them from the noise. By first analyzing the image data in terms of the local image structures, such as lines or edges and controlling the filtering based on local information from the analysis step, a better level of noise removal is usually obtained when compare to the simplest approach.

VIII. SOFTWARE REQUIREMENTS

The software which is required for Vision Based Sign Language Translation Devices is MATLAB. In this chapter we are going to briefly explain about this software.

A. Introduction

The term MATLAB stands for Matrix Laboratory. It is a numerical computing environment and fourth-generation programming language. MATLAB was developed by Math Works. MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces and interfacing with programs written in other languages including C, C++, java and FORTRAN. The file extension for MATLAB is .m and the operating system of MATLAB is cross-platform. Although MATLAB is intended primarily for numerical computing, an optional tool box uses the MuPAD symbolic engine, allowing access to symbolic computing capabilities. An additional package Simulink adds graphical multi-domain simulation and Model-Based Design for Dynamic.

B. History

Cleve Moler, the chairman of the computer science department at the university of New Mexico, started

All rights reserved by www.ijsrd.com

developing MATLAB in the late 1970's.He designed it to give his student's access to LINPACK and EISPACK without them having to learn FORTRAN. It soon spread to other universities and found a strong audience within the applied mathematics community. Jack Little, an engineer, was exposed to it during a visit Moler made to Stanford University in 1983.Recognising its commercial, potential, he joined with Moler and Steve Bangert. They rewrote MATLAB in C and founded Math Works in 1984 to continue its development. These rewritten libraries were known as JACKPAC. In 2000, MATLAB was rewritten to use a newer set of libraries for matrix manipulation LAPACK.

MATLAB was first adopted by researchers and practitioners in control engineering, little specialty, but quickly spread to many other domains. It is now also used in education, in particular the teaching of linear algebra and numerical analysis and is popular amongst scientists involved in image processing.

C. Block Procedures

1) Image Characteristics

Images may be two dimensional, such as a photograph, screen display, and as well as a three dimensional, such as statue or hologram. They may be captured by optical devices like cameras, mirrors, lenses, telescopes, microscopes, etc., and natural objects and phenomena, such as human eye or water surfaces. Those captured images can be rendered manually like by drawing, painting or computer graphics technology.

A volatile image is one that exists only for a short period of time. This may reflection of an object by a mirror, a projection of camera obscures or a scene displayed on cathode ray tube. A fixed image, also called a hardcopy, is one that has been recorded on a material object such as paper or textile by photography or any other digital process.

A mental image exists in an individual's mind. Like something one remembers or imagines. The subject of an image need not be real. The development of synthetic acoustic technologies and creation of sound art have led to a consideration of the possibilities of a sound-image made up of irreducible phonic substance beyond linguistic or musicological analysis.

A still image is a single static image. A film still is a photograph taken on the set of a movie or television program during production used for promotional purposes.

2) Image Fusion

In computer vision, Multi-sensor Image Fusion is the process of combining relevant information from two or more images into a single image. The resulting image will be more informative than any of the input images.

Several situations in Image Processing require high spatial and high spectral resolution in a single image. Most of the available equipment is not capable of providing such data convincingly. Image techniques allow the integration of different information sources. The fused image can have complementary spatial and spectral resolution characteristics. Many methods exist to perform image fusion. The very basic one is the High Pass Filtering Techniques. Later techniques are based on Discrete Wavelet Transform, Uniform Rational Filters bank, and Laplacian Pyramid.

3) Image Processing

In imaging science, image processing is any form of signal processing for which the input is an image, such as a photograph or a video frame; the output of image processing may be either an image or set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as two dimensional signals and applying standard signal processing techniques to it. Image processing usually refers to digital image processing, but optical and analog processing are also possible. The acquisition of images (producing the input image in the first place) is referred to as image as image processing. An image may be considered to contain sub-images, sometimes referred as Region-of-interest or ROIs or simply regions. This concept reflects the fact that images frequently contains collections of objects each of which can be the basis for a region. In a sophisticated image processing system it should be possible to apply specific image processing operations to selected regions. Thus one part of an image (region) might be processed to suppress motion blur while another part might be processed to improve color rendition. Before going to process an image, it is converted into a digital form. Digitization includes sampling of image and quantization of sampled values. This processing technique may be Image Enhancement, Image Restoration and Image Compression.

- Image Enhancement: It refers to accentuation, or sharpening, of image features such as boundaries or contrast to make a graphic display more useful for display and analysis. This process does not increase the inherent information content in data. It includes gray level and contrast manipulation, noise reduction, edge crispening and sharpening, filtering, interpolation and magnification, pseudo-coloring and so on.
- Image Restoration: It is concerned with filtering the observed image to minimize the effect of degradations. Effectiveness of image restoration depends on the extend and accuracy of the knowledge of degradation process as well as on the filter design. Image restoration differs from image enhancement in that the latter is concerned with more extraction or accentuation of image features.
- Image Compression: It is concerned with minimizing the number of bits required to represent an image. Applications of compression are broadcasting in TV, remote sensing via satellite and so on,
  - Text compression-CCITT GROUP3 and GROUP4.
  - Still image compression-JPEG.
  - Video image compression-MPEG.

4) Image Segmentation

Image segmentation is the process of partitioning a digital image into multiple segments (set of pixels, also known as super pixels). The goal of segmentation is to simplify and change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (line, curves, etc.,) in images. The result of image segmentation is a set of segments that collectively cover the entire image or

All rights reserved by *www.ijsrd.com*

a set of contour extracted from the image. The above are the process involved in Image Segmentation.

IX. FUTURE WORK

Due to the many simultaneous 'channels' of signed languages (two hands, Face, head, upper body), the system will extract information not only from the dominant hand, but also from the non-dominant hand and from the facial expression and body position (shoulders, elbows and chest). Sign Speak seeks to explicitly exploit the complementarities and redundancies between these communication channels, especially in terms of boundary detection. This will allow a self-assessment of its own performance, which should yield a high level of robustness to subsystem and also going to implement in mobile phones and the procedure involved is the actions will be primarily converted to text messages through sensors and the another person can receive the text as voice. Their voice is again converted into text which can be read by deaf-dumb people. The general signs for deaf-dumb people were studied. The basic MATLAB coding for the alphabets and numbers are referred. As there are many existing s methods available for Sign Language Translations, their algorithms, calculations, their advantages and their restrictions were noted. In order to expose our project to outside environment in an effective way.

Automatic analysis SL has come a long way from its initial beginnings in merely classifying static signs and alphabets. Current work can successfully deal with dynamic a sign which involve in movement and also appears in continuous sequences. Much attention has also been focused on building large vocabulary recognition system. In this respect, Vision Based System lags behind those that acquire gesture data with direct measure devices.

REFERENCES

[1] Aleem Khalid Alvi, M. Yousuf Bin Azhar, MehmoodUsman, SulemanMumtaz, Sameer Rafiq, Razi Ur Rehman, Israr Ahmed T, "Pakistan Sign Language Recognition Using Statistical Template Matching", World Academy of Science, Engineering and Technology, 2005.
[2] Byung-woo min, Ho-sub yoon, Jung soh, Takeshi ohashi and Toshiaki jima, "Visual Recognition of Static/Dynamic Gesture: Gesture-Driven Editing System", Journal of Visual Languages & Computing Volume10,Issue3, June 1999, Pages 291-309
[3] J. S. Kim,W. Jang, and Z. Bien, "A dynamic gesture recognition system for the Korean sign language (KSL)", IEEE Trans. Syst., Man, Cybern. B, vol. 26, pp. 354-359, Apr. 1996. (Pubitemid 126777790)
[4] T. Shanableh, K. Assaleh, "Arabic sign language recognition in userindependent mode", IEEE International Conference on Intelligent and Advanced Systems 2007.
[5] T. Starner, J. Weaver, and A. Pentland, "Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video', IEEE Trans. Pattern Analysis Machine Intelligence, vol.20, no. 12, pp. 1371-1375, Dec. 1998.
[6] .http://www.bu.edu/asllrp
[7] www.mathworks.com

All rights reserved by *www.ijsrd.com*