Sign Language to Speech Converter Using Neural Networks

Mansi Gupta, Meha Garg, Prateek Dhawan

Lingaya's Institute of Management & Technology, Faridabad, India
{manasigupta18@gmail.com, mehagarg.be@gmail.com,prateek_3212@yahoo.com}

Abstract*:* The normal community has a limited fluency in sign language and because of this a communication barrier persists between the normal and the hearing-impaired people. This Barrier is diminishing as projects of the past two decades have unfolded. These not only help in interpreting the signs but also ease the communication between deaf and general communities. Through the use of artificial intelligence, researchers are striving to develop hardware and software that will impact the way deaf individuals communicate and learn. In an attempt towards the same, a converter has been proposed in this paper. This converter would act as a medium by recognizing the signed images made by the signer and then convert those into text and subsequently into speech. The signed images are classified to increase the accuracy and efficiency of the algorithm.

Key words: Sign converter, Sign language, neural networks, image processing

1. Introduction

A sign language (also signed language) is a language which, instead of acoustically conveyed sound patterns, uses visually transmitted sign patterns (manual communication, body language and lip patterns) to convey meaning—simultaneously combining hand shapes, orientation and movement of the hands, arms or body, and facial expressions to fluidly express a speaker's thoughts. Sign languages commonly develop in deaf communities, which can include interpreters, friends and families of deaf people as well as people who are deaf or hard of hearing themselves. [8]

Wherever communities of deaf people exist, sign languages develop. [9] In fact, their complex spatial grammars are markedly different from the grammars of spoken languages. Hundreds of sign languages are in use around the world and are at the cores of local deaf cultures. Some sign languages have obtained some form of legal recognition, while others have no status at all. In addition to sign languages, various signed codes of spoken languages have been developed, such as Signed English and Warlpiri Sign Language.[1] These are not to be confused with languages, oral or signed; a signed code of an oral language is simply a signed mode of the language it carries, just as a writing system is a written mode. Signed codes of oral languages can be useful for learning oral languages or for expressing and discussing literal quotations from those languages, but they are generally too awkward and unwieldy for normal discourse. For example, a teacher and deaf student of English in the United States might use Signed English to cite examples of English usage, but the discussion of those examples would be in American Sign Language.

Several culturally well developed sign languages are a medium for stage performances such as sign-language poetry. Many of the poetic mechanisms available to signing poets are not available to a speaking poet.

1.1 List of sign languages

Sign language is not universal. Like spoken languages, sign languages emerge naturally in communities and change through time. The following list is grouped into three sections:

- □ Deaf sign languages, which are the preferred languages of Deaf communities around the world
- ○ Signed modes of spoken languages, also known as Manually Coded Languages;
- ○ Auxiliary sign systems, which are not "native" languages, but signed systems of varying complexity used in addition to native languages.

1.2 British Sign Language

British Sign Language (BSL) is the sign language used in the United Kingdom (UK), and is the first or preferred language of deaf people in the UK; the number of signers has been put at 30,000 to 70,000. The language makes use of space and involves movement of the hands, body, face and head. Many thousands of people who are not deaf also use BSL, as hearing relatives of deaf people, sign language interpreters or as a result of other contact with the British deaf community.

2. Literature review

There are various approaches that have been used for converting sign language images into text or speech.

The threshold model with Conditional Random Field (CRF) is an excellent mechanism for distinguishing between vocabulary signs and non sign patterns (which include out-of vocabulary signs and other movements that do not correspond to signs). A short-sign detector, a hand appearance-based sign verification method, and a sub-sign reasoning method are included to improve sign language spotting accuracy. [2]

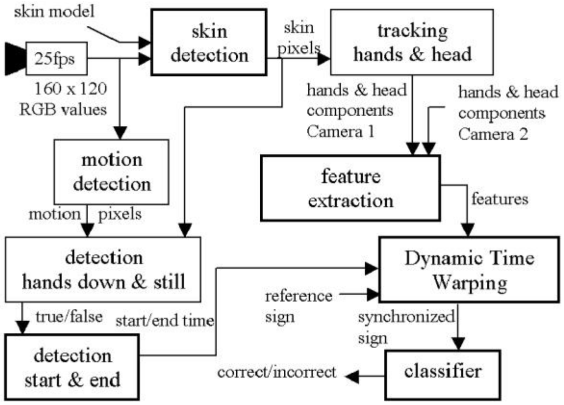

Fig. 1 Block diagram of sign detection [3]

Another method is automatic sign recognition. Its unique features are an adaptive skin model, DTW on a reference sign for synchronization, robust recognition method which is real-time and person-independent statistics, automatic feature selection for finding the best sign representation and a tolerance parameter TF that changes the behaviour of the base classifiers instead of the threshold on the total likelihood. DTW was used only for finding the best path, to synchronize the signal as in Figure 1. The method is able to generalise well over different persons, which is troublesome for many other systems. [3]

In another technique computer vision method has been used for recognizing sequences of human-hand gestures within a gloved environment. Vectors are utilized for representing the direction and displacement of the fingertips for the gesture. Modeling gestures as a set of vectors with a motion key allows the reduction of complexity in modern form and matching, which may otherwise contain multiple and lengthy datasets. [4]

The hand shape was used in recognizing people with high accuracy. It was believed that the scorecard of the hand geometry modality could be promoted to “high” in the distinctiveness and performance attributes of person recognition in that the interface is user-friendly and it is not subject to variability to the extent faces are under confounding factors of accessories, illumination effects and expression. [10] Preliminary tests indicate that hand biometric accuracy is maintained over a span of time. For any hand-based recognition scheme, it is imperative, however, that the hand image be pre processed for normalization so that hand attitude in general, and fingers in particular be aligned to standard positions[5]

Another comprehensive approach to robust visual sign language recognition system aims to signer-independent operation and utilizes a single video camera for data acquisition to ensure user friendliness. In order to cover all aspects of sign languages, sophisticated algorithms were developed that robustly extract manual and facial features, also in uncontrolled environments. The classification stage is designed for recognition of isolated signs as well as of continuous sign language. For statistical modelling of reference models, a single sign can be represented either as a whole or as a composition of smaller subunits—similar to phonemes in spoken languages. In order to overcome the problem of high interpersonal variance, dedicated adaptation methods known from speech recognition were implemented and modified to consider the specifics of sign languages. [6]

A novel algorithm to extract signemes, i.e. the common pattern representing a sign, from multiple long video sequences of American Sign Language was implemented. A signeme is a part of the sign that is robust to the variations of the adjacent signs and the associated movement epenthesis. Iterative Conditional Modes (ICM) to sample the parameters, i.e. the starting location and width of the signeme in each sentence in a sequential manner were used. In order to overcome the local convergence problem of ICM, it was run repetitively with uniformly and independently sampled initialization vectors. The results on ASL video sequences that do not involve any magnetic trackers or gloves, and also on a corresponding audio dataset were shown. [7]

Yet in another approach, an application’s speech and audio output is translated into text using existing speech-to-text conversion programs. The system translates key text words

or phrases into the appropriate sign language. For this translation, pre captured gesture database and Java 3D were used to construct the simple 3D hand model, achieving a rich, interactive, animated environment focusing more on the hand's degrees of freedom (DOF) rather than texture. However this approach focuses on communicating by hand gestures that can be captured by only fingers and palms. For gestures that require other hand motions, such as wrist rotations and hand translations or facial, expressions, more data needs to be incorporated into the system [11].
A demonstrator for generating VRML animation sequences from Sign Language notation, based on MPEG-4 Body Animation has been developed. The system is able to convert almost all hand symbols as well as the associated movement, contact and movement dynamics symbols contained in any ASL sign-box. [12]

3. Problem definition

Sign language is a non-verbal language used by the hearing impaired people for everyday communication among themselves. It is not just a random collection of gestures; it is a full-blown language in its own right, complete with its own grammatical rules.

Linguistic research has to find easy but efficient strategies for the real-time adaptation of the wording in order to make a message understandable also for an audience with limited language proficiency. In order to improve communication between deaf and hearing people, more exhaustive research in automatic sign language recognition is needed. Research on human–computer interaction could also benefit from gesture and mimic analysis algorithms, originally developed for sign language recognition systems.

Euclidean distance is the "ordinary" distance between two points that one would measure with a ruler, and is given by the Pythagorean formula. On the basis of this distance between images of various signs, they are recognized for the correct output. In order to simplify the computation of this distance, the images for signs are converted into a binary format.

The pattern of the sign in binary form is carefully observed and they are then classified on the basis of Multi Layer Perceptron architecture which is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate output as in Figure 2. It uses ANN which is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. They are usually used to model complex relationships between inputs and outputs or to find patterns in data.

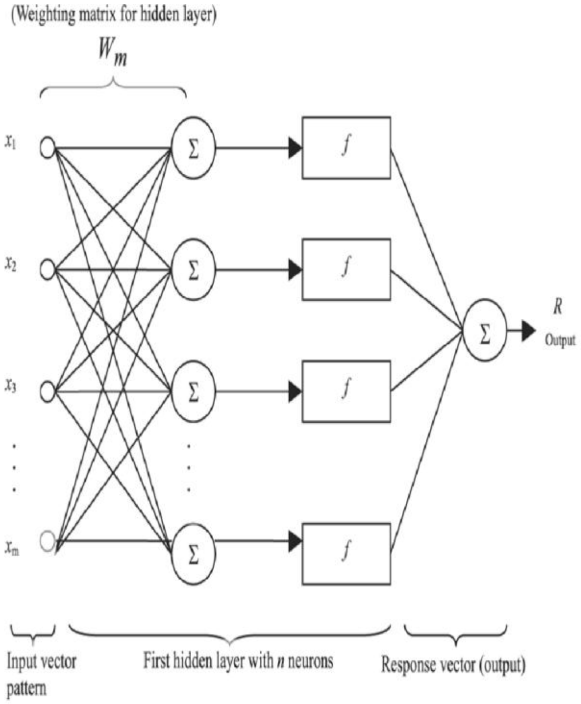

Figure 2: Multi Layer Perceptron Architecture

4. Proposed methodology

This converter recognizes the signed images made by the signer and converts them into text as well as speech. The process followed is described in Figure 3.

In this approach, five sample images per alphabet were taken in a controlled environment. These images were stored in a database. After that they were converted into LAB format as it is considered the most accurate format and can be used as an intermediary for color space conversions. Then the images were converted in binary form and 10X10 blocks were imposed on each binary image. After this the number of black pixels was found in each block. Based on this, an average number of black pixels for all the blocks from all the samples were calculated for each sign.

The signed image is then captured and it undergoes a same process of conversion from RGB to LAB to binary form. Then the number of black pixels for each block is computed and saved.

After this the Euclidean distance between the signed input image and those in the database is calculated. And on the basis of this the image is matched for a particular sign which is then displayed and converted into speech.

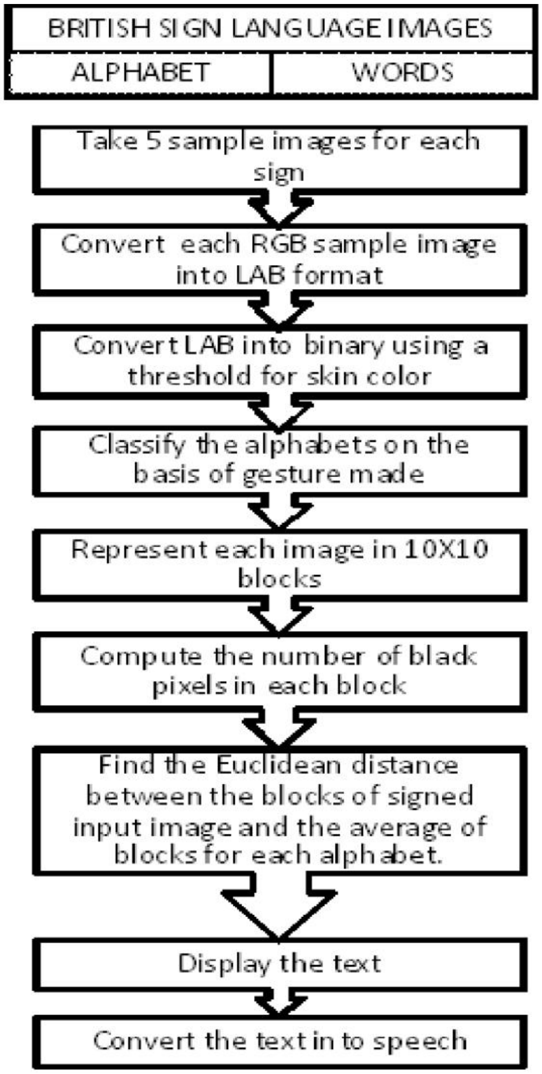
BRITISH SIGN LANGUAGE IMAGES
ALPHABET
WORDS
Take 5 sample images for each sign
Convert each RGB sample image into LAB format
Convert LAB into binary using a threshold for skin color
Classify the alphabets on the basis of gesture made
Represent each image in 10X10 blocks
Compute the number of black pixels in each block
Find the Euclidean distance between the blocks of signed input image and the average of blocks for each alphabet.
Display the text
Convert the text in to speech

Fig. 3 Steps Followed to convert signed image into speech.

The above algorithm was implemented under the following constraints:

1. The camera is at a fixed position and at a fixed distance from the signer.
2. The signs are made in a controlled environment keeping a fixed background.
3. The RGB images are first converted into LAB and then into binary so as to reduce the distortion.
4. Only static signs have been used, i.e., there should be no movement of hands to depict a sign.
5. The size of image is kept constant.

5. Conclusions

Sign language recognition and translation is an active area of research. People with limited fluency in sign language can easily communicate with hearing impaired people with the converter that has been proposed in this paper. As this converter recognizes the signed images made by the signer and converts them into text as well as speech without any use of data gloves or other equipment. Thus, interaction gets simplified between people with or without hearing or speech impairments.

For further work, videos of hand gesture could be captured and recognized through the implementation of the same algorithm.

References

[1] Sign Language Recognition and Translation: A Multidiscipline Approach from the Field of Artificial Intelligence, Becky Sue Parton .

[2] Sign Language Spotting with a Threshold Model Based on Conditional Random Fields Hee-Deok Yang, Member, IEEE, Stan Sclaroff, Senior Member, IEEE, and Seong-Whan Lee, Senior Member, IEEE

[3] SIGN LANGUAGE DETECTION USING 3D VISUAL CUES J.F. Lichtenauer G.A. ten Holt E.A. Hendriks M.J.T. Reinders Information and Communication Theory Group Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS) Delft University of Technology, The Netherlands

[4] Visual Gesture Recognition by J. Davis and M. Shah

[5] Shape-Based Hand Recognition by Erdem Yörük, Ender Konukoˇglu, Bülent Sankur, Senior Member, IEEE, and Jérôme Darbon

[6] Recent developments in visual sign language recognition by Ulrich von Agris Æ Jo¨rg Zieren Æ Ulrich Canzler Æ Britta Bauer Æ Karl-Friedrich Kraiss

[7] Automated Extraction of Signs from Continuous Sign Language Sentences using Iterated Conditional Modes by Sunita Nayak, Sudeep Sarkar, Barbara Loeding

[8] Sign Tutor: An Interactive System for Sign Language Tutoring Oya Aran, Ismail Ari, Lale Akarun, and Bu¨lent Sankur

[9] A Multimodal Framework For The Communication Of The Disabled by Savvas Argyropoulos , Konstantinos Moustakas , Alexey A. Karpov , Oya Aran, Dimitrios Tzovaras , Thanos Tsakiris , Giovanna Varni , Byungjun Kwon

[10] A New Color Image Database for Benchmarking of Automatic Face Detection and Human Skin Segmentation Techniques

Abdallah S. Abdallah, Mohamad A bou El-Nasr, and A. Lynn Abbott

[11] Hand-Gesture Computing for the Hearing and Speech Impaired Gaurav Pradhan and Balakrishnan Prabhakaran University of Texas at Dallas Chuanjun Li Brown University

[12] TEXT-TO-SIGN LANGUAGE SYNTHESIS TOOL

Maria Papadogiorgaki, Nikos Grammalidis, Dimitrios Tzovaras and Michael G. Strintzis Informatics and Telematics Institute,

Centre for Research and Technology Hellas 1st Km Thermi Panorama str., 57001, Thermi Thessaloniki, Greece phone: +(30) 2310464160, fax: +(30) 2310464164, email: mpapad@iti.gr;ngramm@iti.gr web: http://www.iti.gr/

Author Biographies

Mansi Gupta, a final year student at Lingaya's Institute of Mgt. & Tech., Faridabad, Haryana, India. Her areas of interest include Image processing, Artificial Neural Networks, Computer organization and Operating System.

Meha Garg, a final year student at Lingaya's Institute of Mgt. & Tech., Faridabad, Haryana, India. Her areas of interest include Image processing, Artificial Neural Networks, Computer organization and Operating System.

Prateek Dhawan, a final year student at Lingaya's Institute of Mgt. & Tech., Faridabad, Haryana, India. His areas of interest include Image processing, Artificial Neural Networks, Computer organization and Operating