A Survey on the Techniques Applied to the Recognition and Conversion of Indian Sign Language

Kaushal Kishore Rao Mangalore; Nikhitha Pradeep; Bhawesh Rajpal; Nitin Prasad; Ravi Shastri

1. Introduction

The enhancement of science and technology leads to make the life more comfortable than older days. The emerging technologies like neutrosophic shortest path (1,2,3,4,5), transportation problem (6,7,8), uncertainty problem (9,10,11,12,13,14), fuzzy shortest path (15,16,17,18,19), powershell (20), wireless sensor network (21,22,23,24,25,26,27,28), computer language (29,30), neural network (31), routing (32), image processing (33) making the products more intelligent and self-healing based. The smart city applications like smart water (34,35), smart grid, smart parking, smart resource management, etc. are based on IoT and IoE (36,37,38,39) technologies. In this manuscript, a survey and an assessment of the techniques applied to the recognition and conversion of Indian Sign Language are performed. In this section, these sign languages along with some of their recognize techniques are discussed in below.

1.1. Sign Languages

According to its latest report on deafness and hearing loss, the World Health Organization (WHO) estimates that some form of hearing impairment impacts around 6% of the world’s population (40). It is likely that over 466 million people currently have hearing disabilities, and that by 2050 as many as 900 million individuals may have them (41).
In communities with individuals that suffer from hearing disability, especially those with deafness from a young age, sign languages provide a unifying form of communication. Sign languages exist independently of a spoken language and as such regions that share a single spoken language may have multiple independent sign languages. Sources vary on the number of sign languages in use with estimates ranging from 144 (42) to 196 (43) (including non-deaf community sign languages).

A sign language can only successfully bridge the gap between speakers (users of spoken languages) and signers (users of signed languages) if there are sufficient speakers in the community educated in its use. However, as most speakers have no need for sign languages for their participation in society, this is seldom the case. This leaves the gap open to a constructed solution, which has initiated the creation of machine-aided conversion of a sign language to its locally used spoken language.

1.2. Sign Language Recognition

Sign language recognition (SLR) is the generic terminology used to describe the use of machine learning models to identify and classify signs from sign languages around the world. SLR has also been extended to use in real-time video streams and is called continuous SLR (CSLR).

As early as 1986, the identification of Japanese Sign Language (JSL) was conducted, using the classification of image features (44) (an English-only version was later published as (45)). Proceeding approaches to JSL recognition included the use of recurrent neural networks (RNN) to process the dynamic nature of gestures (46).

Similarly, image processing systems that tracked hand motion and adaptive clustering were devised for the recognition of American Sign Language (ASL) (47). Hidden Markov models (HMM) were used to identify signs from ASL with good results on independent test sets in (48).

Researchers around the world have applied SLR to the dominant sign languages in their region, such as Korean Sign Language (KSL) (49), British Sign Language (BSL) (50) and French Sign Language (FSL) (51).

1.3. Indian Sign Language

According to The World Bank, India’s population is estimated to be 1.366 billion as of 2019 (52). If we use the WHO approximation of 6% (41), it can be estimated that as many as 81.96 million Indians suffer from some form of hearing loss (hearing disability is a subset of hearing loss). If we use the 2011 census data for India for our estimation (53), we can approximate that as of 2019, there are 2.25 million Indians suffer from speech disabilities, and 5.72 million Indians suffer from hearing disabilities. But, as these estimations show, there is a huge variation in reporting which makes it hard to get an accurate picture.

Localized sign languages have existed in India for a long time, though they were not studied much until the authors of (54) studied sign languages from major cities around the country. They found that there were common signs across the sign languages in different cities, and some variations as well.

Indian Sign Language (ISL) likely has a shared history with the sign languages in other nations of the Indian subcontinent, according to (55). But, unlike other national or formal sign languages, ISL is still in the process of being standardized and has yet to achieve both a nationalized education approach and
a recognized status as an official language (56). Its development has also been significantly different from other sign languages around the world.

Attempts to create a standardized ISL have been ongoing for several decades, with attempts from different institutions, including the Ramakrishna Mission and CBM International to create an unofficial 2500-word ISL dictionary in 2001 (57), and the Indian Sign Language Research and Training Centre (ISLRTC) to create the first and second official ISL dictionaries in recent years (56). These efforts have brought the number of standardized signs in ISL to nearly ten thousand, but its lack of official status has severely diminished its dissemination and education (56).

These compilations of signs laid the foundations for standardization and efforts are still ongoing to make this standardized ISL the educational norm for children with speech and hearing impairments. It should be noted that ISL is only the standard and that the actual sign languages of deaf communities in India have local dialects and slang that make some signs unintelligible elsewhere in the country (58).

1.4. Recognition of ISL

Research into the development of recognition systems for ISL signs began around 2010, when in their paper (59), the authors proposed a method for processing video into frames, extract features from them and then use Euclidean distance or K-nearest neighbour (kNN) for recognition of the signs used.

The same authors took ISL conversion into the field of human-robot interaction (HRI) in (60). They developed WEBOTS, a system for using signs as gesture commands for the HOAP-2 robot (developed by Fujitsu Automation in 2004 (61)).

In (62), its authors use the YCbCr skin colour model reference to extract hand features and utilize kNN, support vector machine (SVM) and multi-class SVM for the classification and identification of signs.

The authors of (63) used an SVM to recognize images and focus on comparing feature extraction methods to identify which ones were best suited for sign language recognition.

1.5. Outline

This paper is aimed at providing a comprehensive introduction to machine learning and deep learning techniques that have been applied to the sign language recognition problems within the context of ISL. It reviews some recent image-based and sensor-based techniques. From this point on, this paper is organized as follows.

In section 2, the purpose, the motivation and the objectives of this paper are stated. The section also highlights the significance of this paper to the research community.

The techniques applied to ISL recognition and conversion are examined in section 3.

A comparison of the techniques discussed is presented in section 4, followed by a look at the challenges in ISL research and the limitations of the surveyed techniques in section 5.
Finally, section 6 concludes the survey and presents the scope for future research in ISL recognition and conversion.

### 2. Purpose and Importance of this Survey

#### 2.1. Motivation

Recognizing signs and converting them to text has been done for several decades now. The recognition of ISL signs through computer vision techniques is a relatively recent development which, given the right impetus, could change communication for signing communities in India.

With this motivating factor in mind, this paper was written with the aim of:

- Compiling ISL sign recognition techniques,
- Analysing the proposed methods and the results of these techniques,
- Evaluating the utility of these techniques for wider application, and,
- Examining the scope for expanding these techniques.

#### 2.2. Purpose and Objectives

This survey is aimed at researchers in the computer vision field who are looking to apply image and gesture recognition techniques to the identification of signs in any sign language, but especially those looking to promote their use on ISL signs.

The objectives of this survey are:

- To provide a comprehensive reference for future research into sign language recognition (specifically ISL),
- To evaluate the validity and utility of various techniques, within the context of the papers surveyed,
- To explain shortcomings in current research, and,
- To highlight focus areas for future research.

#### 2.3. Significance of this Survey

This survey paper aims to be a comprehensive starting point for researchers in the fields of ISL recognition and conversion. It surveys key techniques used to this end by researchers in recent years and compares them for a better look at their utility and application in the wider field.

### 3. Techniques for Indian Sign Language Recognition

This section presents the techniques that have been applied to Indian Sign Language recognition, whether through static images or videos.
3.1. K–Nearest Neighbours

In their paper (64), the authors propose the recognition of ISL alphabets through static images using the K-nearest neighbours (kNN) algorithm (65) on the correlation of features extracted from the images.

The authors’ method involves multiple steps that begin with the classification of the images into single-handed gestures (SHG) and double-handed gestures (DHG), using histograms of oriented gradients (HOG) features (66) and a support vector machine (SVM) as the classifier.

Following this initial classification, the images are processed with HOG features and scale-invariant feature transform (SIFT) (67), as they did not perform as well individually as they did when combined (fused-features). The correlation of HOG and SIFT features are then calculated to reduce the dimensionality of the features before the application of the kNN algorithm on the correlation matrix for the classification of images.

Overall, their model using the kNN algorithm had an accuracy of 84.23% on HOG feature images, 81.92% for SIFT feature images, and 91% for the fused-feature images (64). Their approach also worked better on SHG, with an accuracy of 98.33%, than on DHG, where the accuracy was 89.5% (64).

3.2. Multilayer Perceptron

In (68) the authors propose the use of a Leap Motion Controller (LMC) to collect hand and finger position data, and a multilayer perceptron (MLP) neural network (69) to recognize ISL numerals.

The authors’ proposed method uses an LMC to collect data on the positions of hands and fingers and their movement. The features extracted from this data include the Euclidean distances between each fingertip and the palm, and between each fingertip and the adjacent fingertips. The data is labelled with the corresponding ISL numerals for training the MLP.

This data is normalized and then passed on to the MLP. The MLP used by the authors is a simple supervised artificial neural network (ANN) with a single hidden layer that uses backpropagation to update its weights.

The network was tested on a different set of data, processed the same way, and achieved an accuracy of 100% (68). The author’s noted that the high accuracy was due to a lack of any similarity among signs for ISL numerals.

3.3. Convolutional Neural Network

The authors of (70) propose the application of a convolutional neural network (CNN) (71) to identify ISL alphabets and convert them to the corresponding English alphabets.

The images of ISL alphabets were collected, resized, and their features extracted. Additionally, they flipped the images along the vertical axis. The images were then passed through MobileNet (72), a pre-trained CNN classifier, to train their ISL converter.
The CNN model was tested on two sets of data – test images presented like the training images where the model achieved an accuracy of 96%, and a set of real-time cropped images where the model achieved an accuracy of 87.69% (70).

3.4. Long Short-term Memory

The application of a long short-term memory (LSTM) network is proposed in (73) for the classification of gesture data collected through a sensor glove.

The authors proposed the use of a sensor glove designed by them to collect finger and hand movement data from the wearer. The glove employs flex sensors and an inertial measurement unit (IMU) to sense the position of fingers and the movement of the fingers and palm. Labelled data is then passed on to the LSTM network (74) for training the model over 30 epochs.

The model that consisted of an LSTM layer, a dense layer, a dropout layer and a softmax classifier, accurately classified 98% of the dataset on which it was tested (73).

3.5. Eigenvalue-weighted Euclidean Distance

In (75) the authors present a technique that involves the use of Euclidean distance (ED) weighted by eigenvalues (ev) computed from the images, which will be referred to as ‘ev-ED’ technique hereafter.

The authors have used a four-step approach starting with skin detection. Based on skin detection, the images are cropped, and eigenvectors (EV) are calculated. The Euclidean distances for these images are derived and multiplied with five selected eigenvectors. The data based on these values is then labelled.

Testing images undergo the same processing, and their eigenvalues are compared with those of the training images. The difference in eigenvalues is multiplied by the Euclidean distance to compute the ev-ED values; this can be understood from these formulas:

- The ED for an image can be calculated using, $\sqrt{\sum_{n=1}^{m}(EV1(n) - EV2(n))^2}$, where EV1 and EV2 are the eigenvectors (75).
- The weighted ED for an image calculated using, $g_{ht}ed\times ED \times |ev_T - ev_L|$, where $ev_T$ is eigenvalues of the test image, and $ev_L$ is eigenvalues of the labelled image (75).

The image is then classified as belonging to that class where the sum of the weighted Euclidean distances is minimum. The ev-ED technique proposed by the authors had an accuracy of 97% in the classification of ISL alphabets (75).

The same authors, in (76), extended this technique to the classification of signs in a live video sequence.

Initially, the samples are processed using skin filtering techniques. Following this, the authors proposed using histogram matching across consecutive images to track similarity and classify a sequence based on whether it appears to contain a sign. The sequences deemed to contain signs are passed on for feature extraction and the computation of eigenvector, eigenvalues and Euclidean distances.
The ev-ED technique was then applied to recognize the signs in the video snippets and had an accuracy of 96.25% (76).

3.6. Neuro-fuzzy System

The authors of (77) propose the use of a neuro-fuzzy algorithm to classify routine ISL signs.

In this paper, the authors have collected a set of images for routine ISL signs and used skin detection to segment them. Using training and testing images, the authors compute the correlation coefficient of their contrasts. These extracted features are then run through a neuro-fuzzy network (78) comprised of two hidden layers – one for the application of fuzzy logic and another for the classification. With this technique, the authors were able to achieve an accuracy of 92% on their testing data (77).

4. Comparison of Techniques

This section looks at the techniques surveyed and compares them to understand their utility.

4.1. Nature of Samples Collected

As shown in Table 1, there is some variation in the data used by the authors of the surveyed papers. The authors of three of the six surveyed papers have collected data on ISL alphabets (64,70,75), the authors of two papers collected routine ISL signs, i.e. words or concepts, and the authors of one paper collected ISL numerals (68). Also note that other than (70), the other papers surveyed limited their data per sign to between ten and forty samples per sign.

4.2. Splitting of Sample Data

It should also be noted, as observed from Table 1, there was no common convention followed in the splitting of data into training and validation samples, with the validation sample ranging from 10% (a 90-10 split) to 1000% (a 9-91 split) of the training sample size. However, some papers weren’t entirely clear on the training-validation split of the samples collected (70,75,77).

4.3. Data Processing Methods Utilized

In Table 1, it is noted that the authors of each paper surveyed had a set of processing frameworks to get the samples ready for training or validating the model. The most common methods used across frameworks were skin filtering (73,75,77) and image cropping (70,77).

4.4. Accuracy of Models

The models used across the papers surveyed proved to be accurate for the data they were presented with, as seen in Table 1. Although in some papers, the performance of the model is unclear due to the scarcity of information on the exact training-validation split of the samples (70,75,77).
**Table 1. Comparison of ISL recognition techniques**

<table>
<thead>
<tr>
<th>Technique</th>
<th>Data Description</th>
<th>Sample Size</th>
<th>Training Size</th>
<th>Validation Size</th>
<th>Pre-processing Methods</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>kNN (64)</td>
<td>ISL Alphabets (26), images</td>
<td>26 * 30 = 780 samples</td>
<td>520 samples</td>
<td>260 samples</td>
<td>HOG features through SVM classifier, HOG and SIFT feature extraction, and feature fusion</td>
<td>91%</td>
</tr>
<tr>
<td>MLP (68)</td>
<td>ISL Numerals (10), images</td>
<td>10 * 20 = 200 samples</td>
<td>100 samples</td>
<td>100 samples</td>
<td>LMC hand and finger features extraction</td>
<td>100%</td>
</tr>
<tr>
<td>CNN (70)</td>
<td>ISL Alphabets (26), images</td>
<td>26 * 2000 = 52000 samples</td>
<td>Part of 52000 samples</td>
<td>The remainder of 52000 samples</td>
<td>Image cropping, resizing and flipping</td>
<td>96%</td>
</tr>
<tr>
<td>LSTM (73)</td>
<td>ISL Signs (26), sensor measurements</td>
<td>26 * 40 = 1040 samples</td>
<td>936 samples</td>
<td>104 samples</td>
<td>Filtering, and normalization</td>
<td>98%</td>
</tr>
<tr>
<td>ev-ED (75,76)</td>
<td>ISL Alphabets - [except H and J] (24), images (75)</td>
<td>24 * 10 = 240 samples</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Skin filtering, hand cropping, eigenvalue and eigenvector computation and selection, and Euclidean distance computation</td>
<td>97%</td>
</tr>
<tr>
<td></td>
<td>ISL Alphabets - [except H and J] (24), videos (76)</td>
<td>24 * 10 = 240 samples</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Skin filtering, histogram matching, feature extraction, eigenvalue and eigenvector computation and selection, and Euclidean distance computation</td>
<td>96.25%</td>
</tr>
<tr>
<td>NFS (77)</td>
<td>ISL Signs (10), images</td>
<td>Unclear, likely 10 samples</td>
<td>10 * 10 = 100 samples</td>
<td>Unclear, likely 10 samples</td>
<td>Skin filtering, and correlation computation</td>
<td>92%</td>
</tr>
</tbody>
</table>
5. Challenges and Limitations

5.1. Challenges in ISL research

There are common challenges that the authors of the surveyed papers have likely faced; these include:

- **Standardized ISL signs**
  Although by recent count there are close to 10,000 signs in standardized ISL (56), they are yet to permeate the signing community in India. As such, many researchers likely found it a challenge to locate and involve local signers who could conform to standard ISL.

- **Absence of a public ISL dataset**
  Unlike corpora and datasets compiled for ASL (79), BSL (80), and other sign languages around the world (81,82,83), there’s no corpus or dataset for ISL easily available to researchers. Although there are authors who identified the problems and created ISL datasets, such as (84), these datasets are not easily available or accessible to the research community.

- **Limited knowledge of ISL**
  The researchers likely had limited resources for the assimilation of ISL beyond its alphabets. This was probably a contributing factor to the challenges they faced in data creation and collection.

5.2. Limitations of current techniques

In surveying and comparing techniques for this paper, the following limitations were noted:

- **Small sample sizes**
  As shown in Table 1, the sample sizes in most datasets were extremely limited, and appear to be limited to around 40 samples per sign, except for (70).

- **Minimal sign variety**
  With data collected mostly of ISL alphabets, numerals, or common signs, and never more than 26 unique signs, the ability to accurately evaluate the utility of these techniques for wider application is limited.

- **Limited signer involvement**
  The datasets were created by the authors for their research, and there is no explicit mention in any of the papers of the involvement of native signers or qualified interpreters. Thus, there is no way to evaluate whether the proposed models would perform as well on data from native signers or interpreters.

6. Conclusion

This paper reviewed some of the most recent research on the recognition of ISL signs. The papers were surveyed for the techniques they proposed and their authors’ approaches to creating a reliable and usable ISL conversion system. Specifically, recent papers that applied k-nearest neighbour, multilayer perceptron, convolutional neural networks, long short-term memory, eigenvalue-weighted Euclidean distance, and neuro-fuzzy system algorithms were discussed. The challenges faced by researchers in ISL recognition and the limitations of current research in this field were brought to light. We encourage future researchers in this field to focus on the creation of a public corpus or dataset of
ISL signs and contemplate the modelling of speech-to-sign conversion approaches. To achieve this, we feel it is vital to collaborate with deaf and signing communities in India.

References


Global, pp 213-259.


10.1049/pbce128e_ch3


University Press: Cambridge, Chapter 2, pp 8-32.


©2020 by the authors. Licensee Journal of Applied Research on Industrial Engineering. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).