European Journal of Computer Science and Information Technology
Vol.10, No.5, pp.41-52, 2022
Print ISSN: 2054-0957 (Print),
Online ISSN: 2054-0965 (Online)

An Investigation of Translation of Text Language to Sign Language Using Machine Learning

Doaa Nasser Alghamdi
Research Scholar, Department of Computer and information Sciences, King Saud University, Riyadh, Kingdom of Saudi Arabia.

Citation: Alghamdi D.N. (2022) An Investigation of Translation of Text Language to Sign Language Using Machine Learning, *European Journal of Computer Science and Information Technology*, Vol.10, No.5, pp.41-52

ABSTRACT: *Among the fastest-growing areas of study today is the translation of sign language, which is the most natural form of communication for those with hearing loss. Deaf individuals may be able to communicate with hearing people directly, without an intermediary, with the use of a hand gesture recognition device. The method was developed to facilitate the automatic translation of American Sign Language into text and sound. The suggested system uses a large data collection to interpret individual words and phrases in traditional American Sign Language, alleviating any fears that the user may have about using a virtual camera. Deaf and mute persons must rely on Sign Language as their only method of communication. However, a large portion of the general public is illiterate in sign language. Therefore, those who sign have a more difficult time communicating with those who don't without the help of an interpreter. The proposed technique employs data collected by Fifth Dimension Technologies (5DT) gloves to try to interpret hand movements as spoken language. The data has been classified into text words using a variety of machine learning techniques, including neural networks, decision tree classifiers, and k-nearest neighbors.*

KEYWORDS: translation, text language, sign language, machine learning, classify, translator.

INTRODUCTION

The grammar of American Sign Language (ASL) is completely separate from the grammar of spoken languages, yet the two have a common etymological ancestry. This allows ASL to be represented via the use of intentional body movements. Native Americans, especially the hearing- and visually-impaired, find it hilarious. Unfortunately, there is no universally recognized or standardized version of sign language [1]. In many domains, various signal languages are making educated guesses. For example, American Sign Language (ASL) is a completely separate language from British Sign Language (BSL), and vice versa. Some countries' sign languages have adapted ASL's features. For those with hearing or speech impairments, sign language provides a means of communication. There are around 360 million people with hearing loss in the world; this includes approximately 328,000000 adults and 32000000 children. Disabling hearing loss is defined as a decrease in hearing threshold of more than 40 dB in the better listening ear. As a result, there is a rising need for interpreters due to the increasing number of persons who are deaf. It becomes important to ensure successful discussion amongst all people by reducing the communication gap between those with hearing impairments and those without [2].

When communicating with a deaf or mute person, a translation into the person's native language is required. It's difficult for them to interact with normal people since they can't constantly be followed by a translator. The deaf and mute people might benefit greatly from a machine translation system that can convert sign language into standard spoken English. The proposed system utilizes 5DT gloves equipped with position sensors and a suite of machine learning algorithms [3] to translate sign language hand gestures into audible speech. Native Australian sign language and its accompanying gestures The suggested technique utilizes a dataset downloaded from the UCI Repository of Machine Learning Databases. Data was gathered using a pair of Fifth Dimension Technologies (5DT) gloves. As there are 10 position sensors per glove in a 5DT set, the user has six degrees of freedom. This means there are 22 fields in each record. Based on the terms that exist elsewhere in the dataset, each record is classified into one of 96 groups. Around 80% of the data was utilized for training, while 20% was used for testing. Each record in the dataset was assigned to one of several categories using a variety of machine learning methods, such as k-Nearest Neighbor, Decision Tree Classifier, and Neural Network.

The United States as a whole does not recognize sign language as a natural language. India has its own sign language, signal 6, and different regions have their own vernaculars much like there are different languages spoken in different parts of the world. American Sign Language has a 90% detection rate compared to the grammatical precision of Indian Sign Language. The Deaf people in India recall that it's much higher than specific sign languages because it's far a natural way for them; they study through the natural interplay with the human beings around them. The grammar is similar at certain levels in both the United States and India. Similar to the development of spoken languages, sign language is learned in stages, with babies starting off by babbling with their hands [4]. There is a lack of information on the part of certain individuals and some institutions that have recommended American Sign Language (ASL) over Indian Sign Language (ISL) because of the lack of infrastructure in India to support the development of ISL.

Sign Languages

In Germany alone, there are an estimated 80,000 persons who are deaf and an additional 16 million who have some degree of hearing loss. Numerous people who are deaf or hard of hearing rely only on sign language to communicate [5]. Studies in linguistics have established that sign languages have many of the same characteristics as spoken languages, including phonology, morphology, and syntax. Similar to how words are constructed from a variety of phonological units, signs are made up of a variety of hand shapes, hand positions, and hand movements. Similar to how syllables and words form the basis of spoken language, such components are used to build signs.

Some of the functional differences between sign languages, using American Sign Language as an example, are explored below. Instead of using suffixes on verbs, as is done in spoken languages,

Vol.10, No.5, pp.41-52, 2022
Print ISSN: 2054-0957 (Print),
Online ISSN: 2054-0965 (Online)

ASL tenses are formed by adding words to the beginning or end of a sentence. American Sign Language (ASL) verbs of motion not only describe the speed and direction of movement, but also the precise location and circumstances under which it occurs. More than 300 kinds are in use, according to the World Federation of the Deaf [6].

Static Hand Sign Translation

Based on fixed hand gestures, automated sign language translation may be possible. Natural language alphabets have clear correspondences to hand forms. An AI classification algorithm can enable the identification of these immobile indicators. The most difficult part is extracting features such as finger identification, hand rotation, and orientation. In recent years, several vision-based strategies have been explored with the aim of resolving the issue of hand sign identification [7]. For their 2019 study, Fayyaz and Ayaz used a static sign language picture dataset to compare and contrast the effectiveness of several classifier designs. When using SURF, the authors improved the accuracy of their Support Vector Machine (SVM) from 56% to 83%. Using SURF features improved Multilayer Perceptron (MLP) accuracy to about 86%, whereas manually generated features only improved it to around 58%. investigated depth-based hand detection and tracking for real-time static hand form identification. Random forest models were used as classifiers. As a result of their hard work, we have 500 samples to analyze. The graphical user interface developed by the authors worked well on commercially available computers when compared to "the combined vector" (mean accuracy 75%), "the appearance" (mean precision 73%), and "the depth" (mean precision 69%).

Dynamic Sign Language Translation

Since the input data for translating dynamic sign language are time-based sequences, more advanced network topologies like RNNs are required [9]. With the advancement of machine learning and natural language processing techniques, the potential for sign language translation has grown as well. The fundamental difficulty is translating the unbroken sign language sequences into the vocabulary and syntax of a spoken language. In particular, individual actions and gestures cannot be transposed onto the spoken word. In this case, it could be helpful to utilize a gloss notation [10] [11]. Here, we'll go through several techniques for capturing live signers' movements for later real-time recognition. To translate sign language to spoken language from sign video, we developed attention-based encoder-decoder RNN and Convolutional Neural Network (CNN) architecture in 2018. The authors of this work developed the "RWTH PHOENIX-weather 2014" dataset, the first of its type to be made publicly accessible, in order to offer continuous sign language translation. To solve the problem of one-to-one mapping between words and symbols, an attention-equipped CNN was placed in front of the RNN that was used to represent probabilities [12]. Without include specifics like dates, hours, or places, they discover that the networks really perform rather well.

Literature Survey

Rhythm Shahriar (2017) method for digitally translating Bangla voice to Bangla sign language and another system for translating text to speech were designed to guarantee bidirectional communication. When users provide input in Bangla, the system translates that voice into written Bangla. It's broken down into words so you may study it more closely. The database contains maps between the words and pictures of signs for each of the terms. The primary problem with this strategy is that it will be difficult for hearing-impaired people to grasp static images.

Jordi Porta (2014) Jordi Porta has created a system to translate Spanish text into Spanish sign language. It follows a set of guidelines developed for the translation of Spanish glosses into Spanish Sign Language. This system was found to have a Translation Error Rate of 42% and a BiLingual Evaluation under Study Value of 0.30.

Shiv Naresh & Saritha Khethawat (2012) utilizes the emotion ontology to determine how to categorize feelings. To identify the tone of the text, they employed a list of predetermined terms or keywords. Next, the emotion is categorized using the emotion ontology and the key term.Mikhail Grif et al. (2011) using a method based on semantic translation. They did the translation in phases, first analyzing the text's morphology, then examining the syntactic and semantic aspects of the Russian text and Russian sign language. Avatar animation is then used to visually represent the motions.

Achraf Othman & Mohamed Jemni (2011) strategy for translating English into American Sign Language based on statistics. The translation corpus was created initially, and then the probability and jaro-winkler distance metric were used to determine the degree of similarity between two texts in order to produce the appropriate ASL motions.

Ameera Amasoud & Hend Al-Khalifa (2011) Describes the process of translating from Arabic to Arabic Sign Language using a predetermined set of rules and underlying domain ontology. In the realm of prayer, they developed a sign language translation system. The semantic analysis of a sentence was used to examine the grammatical transformation used in the development of SignWriting.

Sara Morrissey & Andy Way (2009) strategy wherein a collection of sample phrases is taken into account was offered. After being matched with their counterparts in the target language, these phrases are then saved in a database. There are three stages to the translating process. After the first input is examined, the closest possible match is identified. The translation is first broken down into sub-sentences and then put back together again. Establishing a similarity score dependent on word occurrences is an important part of the search for optimal matches. Due to the small size of the dataset, this approach can only be used for a limited number of translations.

Purushottamkar et al. (2007) INGIT, a system based on a hybrid-formulaic method, was presented to translate Hindi text into ISL. The system was developed specifically for the railway sector, and it only has a small corpus of 230 phrases to work with. Rail ticket counter talk, if you will. The syntactic and semantic similarities between the two languages provide the basis for the rules that govern their interaction. They've designed and tested the system on a limited dataset, proving its viability for future expansion.

Francois-Regis Chaumartin (2007) put out a linguistically-informed rule-based procedure. The foundation of their work is identifying sentiments behind major news stories. They provide guidelines for identifying feelings like surprise and identify additional data like "compassion for persons in need of protection." As an added bonus, they attempted to spot any tech-related acronyms. Lexical databases WordNet, SentiWordNet, and WordNet-Affect have all been used.
Changhua Yang et al. (2007) look at how well Support Vector Machine (SVM) and Condition Random Field (CRF) machine learning approaches can classify the sentiment of online blog corpus. Sentence-level training is used to create emotion classifiers that can then be applied to documents. When classifying a sentence's mood, their systems also consider the surrounding material. The CRF classifier has been shown to outperform SVM classifiers in experiments. When using document-level emotion categorization on a blog, the sentiment of the last phrase is very relevant. Words were utilized in place of physical characteristics.

RESEARCH METHODOLOGY

Image Based Sign Language Translator

The picture of the sign is converted into the word it represents using methods from Image Processing and machine learning. Image processing is used to extract the image's features, and the resulting feature vector is then categorized with the use of machine learning methods. There is no specialized hardware required other than a Smartphone. There are a few problems with this system:

- They have trouble functioning in dim lighting.
- It's not uncommon for modern Smartphones to go through batteries quickly.
- They have to be put somewhere out of the way from the person making the signs, which isn't always possible.

Glove Based Sign Language Translator

One way to convert sign language movements into text is via a pair of "smart gloves" that incorporates a number of sensors and a low-power printed circuit board. Each ASL letter has its own unique Bluetooth transmission sequence that may be received on a Smartphone or PC. Because of its tailor-made nature, the circuit board can only handle the standard set of 26 letters.

Sign Language Translator Using Sensor Gloves and Neural Networks
The training process involves the employment of sensor gloves, each of which has seven sensors: five on each finger to detect the force applied between the knuckle and the first joint of the finger, and two more to track the hand's orientation and movement as it is used. After a neural network has been built, it may be used to transform numeric input into character values. The system is able to decipher the 24 letters of the English alphabet, as well as two punctuation marks. The technology can convert hand motions into the letters or numbers they represent instantly.

Multiclass Classification
In multiclass classification, a record is sorted into one of several categories based on its similarities to those defined for each class. The goal of the algorithm is to classify the given record into one of the 96 words in the Australian Sign Language Signs Data Set. Several machine learning techniques, such as k-Nearest Neighbors, Decision Tree Classifier, and Neural Network, have been used to complete this classification.

K-Nearest Neighbors (k-NN):
Since the input record is likely to share characteristics with its neighbours, k-NN can provide an accurate prediction about the input record's class. By calculating the Euclidean distance between each sample and the full collection of samples, this method finds the k closest training instances. The evaluation of the k closest neighbours is performed, and the sample's class is determined by whichever of the two classes has the most members.

Decision Tree Classifier:
The Decision Tree Classifier is one kind of classifier; it produces a decision tree where each node represents an attribute test and the path to be followed is decided by the attribute test's outcome and each leaf represents a label. By applying a sequence of tests provided in the tree nodes, the resulting decision tree can place a particular test record into the appropriate class (leaf). A decision tree classifier is built using the following operations:

Entropy: It determines how similar the data is.

Information Gain: It measures how much the entropy of a dataset has changed after being divided along a particular criterion.
The node at which the data set is divided is determined by the property that provides the most information gain in this approach. Iterate this process until further splitting is impossible. The relevant class's leaf node is made here.

Neural Networks:
In a Neural Network, a perceptron serves as a stand-in for a natural neuron. Each perceptron takes in a number of weighted inputs and sends the result through an activation function. Neural networks undergo two stages:

Training: The neural network is currently in development. The training data is initially sent to the input layer and the final result is collected in the output layer. This process is called "forward propagation" Just take the projected number and take it away from the actual amount to get the error. As a result of the error, the weights are adjusted via back propagation. When making adjustments to the weights in Backpropagation, the gradient descent optimization technique is used, which is a kind of hill-climbing optimization (wij). Find the weight that minimizes the mistake by computing the error's partial derivative with respect to wij. This is the starting point for the derivation of the error gradient, which is obtained by subtracting the weight wij from the starting point. The process is continued until the amount of error has been minimized to an acceptable level.

Testing: After the test data has been sent into the input layer, we are now receiving results from the output layer. Using the perceptron with the greatest value for prediction purposes will provide the best results in multiclass classification.

Analysis

In the suggested method, training and testing are separated into two distinct steps. To determine the system's efficacy, it must first be taught using a variety of machine learning techniques and then tested. The various methods in the system have been implemented with the help of scikit-learn, a python machine learning package.

k-NN:

Each test record is analyzed using the k-NN algorithm.

Let (Xi, Ci) where i = 1,2…, n be data points. Values at index Xi represent features, whereas those at index Ci represent labels for Xi. Ci ∈ {0,1, 2, 3, ……, 95}.

Suppose x is a location that needs a label.

1. Calculate d(x,Xi) for i = 1, 2 …, n; where d is the distance between the two locations in Euclidean space.
2. Arrange the n Euclidean distances found in the calculation by ascending order.
3. Determine which Classes the first k values in the ordered set belong to.
4. Let S_j count how many of the k points belong to the Cj.
5. The class with the highest Sj is the one chosen as the output.

Table 1 summarizes the accuracy at various k values.

K	Accuracy
4	97.0574123849
5	96.8093180872
7	96.5950351714
10	95.5659038405
15	93.9186361633

Table 1: k-NN Accuracy

It is necessary to settle on a value for k. When k is low, noise is more noticeable, while increasing k lengthens the time it takes to cycle through each test record.

Decision Tree Classifier:
As part of the training process, the decision tree is built using a decision tree classifier. In this example, the approach splits the training set into smaller and smaller subsets until each subset has data from exactly one class, or at least a majority of data from one class. Each category is subdivided until it is small enough or can be classified in its entirety. The tests at the internal nodes guide the test record along the appropriate route, while those at the leaf node establish the record's class. The following are the basic building blocks of the Decision Tree Classifier Algorithm:

- The first step is to determine the initial Entropy E.
- Find the information acquired for each property separately.
- Divide the data using the property that yields the most useful insights.
- After partitioning a dataset, step 4 is to repeat processes 1-3 for each new dataset.

The decision tree classification algorithm system gets an accuracy of 76.376%. When there are a lot of classes to sort through and the environment is noisy, Decision Tree Classier struggles to do its job. Overfitting occurs as a result of the algorithm, which reduces training error but has a negative impact on accuracy.

Neural Network:
A neural network is employed for multiclass classification in the suggested approach. Training then culminates in a final examination.

Creation of Neural Network:
Multilayer perceptrons (MLPs) are very deep artificial neural networks in which the signal is received in the input layer and the class of the input is predicted in the output layer. Between these two layers is the MLP's true computational engine, which may have any number of hidden layers. Six discrete layers, each with 100 perceptrons, are used in the proposed system. The Backpropagation Algorithm is used to fine-tune the weights and biases of a neural network to ensure that training errors are kept to a minimum. Given the wide variety of categories, multiclass classification is achieved by using a whopping 96 perceptrons in the final "output" layer. Using the softmax function, this stage of the output chain further improves accuracy. Using a vector of log its or feature variables as input, the softmax function calculates a posterior probability distribution. As a result, the aggregate perceptron output is 1. When the classes are incompatible, this operation is helpful. The formula for the softmax function is:

$$y_c = \frac{e^{z_c}}{\sum_{d=1}^{n} e^{z_d}}$$ For c= 1, 2…, 96.

Where n = 96 and z_c is the value of output layer cth perception.

Training phase:
The following procedures are carried out during this stage:

- **Pre-processing of data:** Here, the algorithm labels each word with a certain category. One hundred and eight words make up the training dataset. In this way, the 96 courses are numbered from zero to ninety-five.
- **Forward Propagation:** The neural network's output is calculated after the training data is supplied into the network's input layer.
- **Error calculation:** Here we see the discrepancy between predicted and actual results. The extent of this mistake must be reduced.
- **Back propagation:** Based on the estimated mistake, the weights are updated using the back propagation technique.
- **Steps:** The steps 2-4 are performed again and again until the mistake is reduced to a minimum.

Figure 1 depicts the cost of mistake during system training. In order to maximize efficiency, we've decided to set the learning rate at 0.001. As the iteration count rises, the error rate drops.

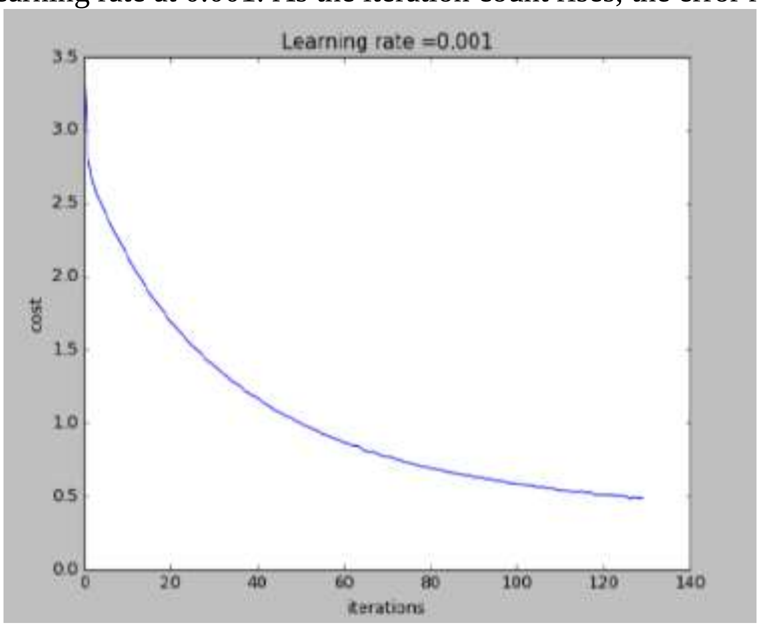

Figure 1: Cost Graph for Neural Network Error

Testing phase:

We feed the test data into the trained neural network and use the network's output to label the data. During this stage, we determine that the accuracy is 76.027 percent.

Methods of Machine Learning: An Analytical Comparison
Table 2 provides an overview of the precision of several methods.

Technique	Accuracy
Neural Network	76.0374873514
Decision Tree Classifier	76.0374873452
KNN	76.0374873431

Table 2: Accuracy Comparison

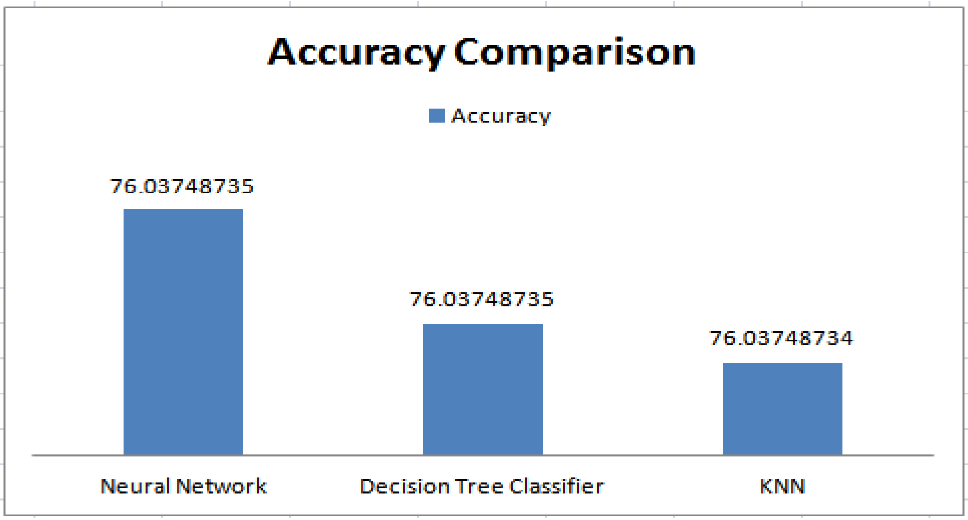

Figure 2: Accuracy Comparison

CONCLUSION

k-NN, a Neural Network, and a Decision Tree Classifier were used to create the suggested system, and k-NN was determined to have the greatest accuracy (97%) of the three. A combined 76% accuracy from a decision tree and neural network. When there are a lot of classes and not a lot of data, the Decision Tree Classifier struggles. As a result of noise during the decision tree creation process, many inconsistencies are inserted into the branches, leading to overfitting. Over fitting happens when an algorithm repeatedly decreases an error in the training set while simultaneously increasing an error in the test set. As a result, the success rate of this method has been lower than expected. Overfitting occurs in Neural Networks when just a little amount of data is used to train the model. However, without a large and varied training dataset, the resulting Neural Network is only able to accurately categorize records inside the training dataset, and cannot be used to correctly categorize test records. Since the system is tuned for a particular data set, it cannot provide a broad answer. This means that the precision suffers while working with a smaller dataset.

Classification may be achieved using the k-NN method by comparing the classes of the nearest neighbours. Because it prevents overfitting, this method also excels at multiclass classification. K-Nearest Neighbors has shown to be the most successful of the three approaches. This research work can be extended in future using the deep learning to achieve great accuracy.

1. References

[1]. Achraf Othman & Mohamed Jemni 2011, 'Statistical Sign Language Machine Translation; From English Written to American Sign Language Gloss. International Journal of Computer Science,vol.8, no.5, pp. 65-73.

[2]. Achraf, O., Jemni, M.: Designing high accuracy statistical machine translation for sign language using parallel corpus: case study English and American Sign Language. J. Inf. Technol. Res. 12(2), 134–158 (2019)

[3]. Almohimeed, A. et. al.: Arabic text to Arabic sign language translation system for the deaf and hearing-impaired community. In: Proceedings of the second workshop on speech and language processing for assistive technologies, pp 101–109 (2011)

[4]. Ameera M Almasoud, Hend, S & Al-Khalifa, A 2011, 'Proposed Semantic Machine Translation System for translating Arabic text to Arabic sign language', proceeding of Second Kuwait conference on Eservices and E- Systems, pp. 79-84.

[5]. Bhargav Hegde, Dayananda P, Mahesh Hegde, Chetan C, "Deep Learning Technique for Detecting NSCLC", International Journal of Recent Technology and Engineering (IJRTE), Volume-8 Issue-3, September 2019, pp. 7841-7843

[6]. Changhua Yang, Kevin Hsin-Yin Lin & Hsin-His chen, 2007 'Emotion Classification Using Web Blog Corpora' proceeding of International Conference on Web intelligence, pp. 275-278.

[7]. Chaumartin, FR 2007, 'UPAR7: A Knowledge-based System for headline sentiment tagging', proceeding of Fourth International workshop on Semantic Evaluations, pp. 422-425.

[8]. Davidson, M.J.: PAULA: a computer-based sign language tutor for hearing adults. In: Intelligent Tutoring Systems Workshop on Teaching with Robots, Agents, and Natural Language Processing, pp. 66–72 (2006)

[9]. Elmahgiubi, M., Ennajar, M., Drawil, N., Elbuni, M.S.: Sign language translator and gesture recognition. In: Proceedings of the Global Summit on Computer & Information Technology. pp. 1-6 (2015)

[10]. Jin, C.M., Omar, Z., Jaward, M.H.: A mobile application of American Sign Language translation via image processing algorithms. In Proceedings of the IEEE Region 10 Symposium. pp. 104-109 (2016)

[11]. Jordi Porta, Fernando López-Colino , Javier Tejedor & José Colás 2014, 'A Rule-based translation from written Spanish to Spanish Sign Language glosses, proceeding of computer speech and language, vol. 28, pp. 788-811.

[12]. Madhuri, Y., Anitha, G., Anburajan, M.: Vision-based sign language translation device. In: Proceedings of the International Conference on Information Communication and Embedded Systems. pp. 565-568 (2013)

[13]. Mikhail G Grif, Olga O Korolkova, Yelena A Demyanenko & Yevgeny B Tsoy, 2011, ‘Development of Computer Sign Langugae Translayion Technology for Deaf People’, proceeding of Sixth International Forum on strategic Technology, pp. 674-677.
[14]. Nikam, A.S., Ambekar, A.G., Sign language recognition using image based hand gesture recognition techniques. In: Proceedings of the Online International Conference on Green Engineering and Technologies. pp. 1-5 (2016)
[15]. Pigou L., Dieleman S., Kindermans PJ., Schrauwen B. (2015) Sign Language Recognition Using Convolutional Neural Networks. In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science, vol 8925. Springer, Cham.
[16]. Purushottamkar, Madhusudhan Reddy, Amutabha Mukerjee & Achla Raina, M 2007 , ‘INGIT: Limited Domain Formulaic Translation from Hindi Strings to Indian Sign Language’, Multilingual Europe Technology Alliance.
[17]. Rhythm Shahriar, Zaman A.G.M, Tanvir Ahmed, Saqib Mahtab Khan, Maruf H.M 2017, ‘A Communication Platform Between Bangla and Sign Language’, IEEE Conference on Region 10 Humanitarian Technology, pp. 1-4.
[18]. Sara Morrissey & Andy Way 2009, ‘An Example-Based Approach to Translate Sign Language’, proceeding of Second Workshop on Example-Based machine Translation, pp. 109-116.
[19]. Shiv Naresh Shivhare & Saritha Khethawar, 2012, ‘Emotion Detection from text’, International Journal of Data Mining and Knowledge Management Process,vol.4, no.6, pp.51-55.
[20]. Sze, V., Chen, Y.H., Yang, T.J., Emer, J.S.: Efficient processing of deep neural networks: A tutorial and survey. In: Proceedings of the IEEE, 105(12), pp.2295-2329 (2017)
[21]. Tolentino, Lean Karlo S. et al. “Static Sign Language Recognition Using Deep Learning.” International Journal of Machine Learning and Computing 9 (2019): 821-827.
[22]. Umang Patel and Aarti G. Ambedkar, “Moment Based Sign Language Recognition for Indian Language” 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA).