Text To Sign Language Conversion & Vice Versa

[1]S.Menaga, [2]B.Gayathri,[3] K.Yogisha, [4]T.Poovarasan, [5]K.Karthikeyan
[1]Assistant Professor, [2,3,4,5]Students, Department of Electronics & Communication Engineering
Jai Shriram Engineering College,Tirupur, India

ABSTRACT—Recognition of sign language is a rapidly advancing area of research, offering a natural means of communication for an individual's who experiencing hearing difficulties. A system for recognizing hand gestures presents an opportunity for deaf and dump individuals to communicate directly with those who can speak, eliminating the necessity for an interpreter. The primary objective of this proposed system is to recognize fundamental components of sign language and facilitate translation to and from text. This proposed system is designed specifically for automatic recognition of Indian sign language, aiming to provide instructional tool to educate deaf and dump users in Indian sign language. The system is geared towards training new users who may be unfamiliar with sign language, offering offline lessons facilitated by a database featuring predefined sign language alphabets and words. Using a collection of samples captured via a camera, the system compares the samples with the words in the database then identifies isolated words. By using neural network algorithm, the images are converted into proper text format and vice versa. The proposes app is developed by using python with front end languages and the result produce 96% of accurate identification.
Keywords: ISL- Indian Sign Language, Fingerspelling, CNN- convolutional neural network, FSDC - Frame Stream Density Compression, SSIM - Structural Similarity Index Measure

I. INTRODUCTION

In recent years around 430 million people, or over 5% of the world's population (432 million adults and 34 million children), need rehabilitation to treat their disabling hearing loss, according to the World Health Organization (WHO). Presently, about 72 million individuals who are deaf or hard of hearing across the globe speak more than 300 different sign languages. Deaf and hard-of-hearing communities in the US now primarily use American Sign Language (ASL). In this field, there is great promise for the creation of a hand gesture detection system that would provide deaf people with an interpreter-free direct line of communication. It is employed as teaching aids for those learning sign language, as well as tools for information retrieval and querying from pictures in signed language, real-time automatic interpretations of sign language, and personal assistants that respond to sign language. This Web tool has the features of converting text to sign language and also converting sign language to text. Presently, the majority of Sign Language Processing (SLP) research has been conducted by the Computer Vision (CV) community, primarily concentrating on the visual aspects of signed languages, with minimal engagement from the NLP community.

II. LITERATURE SURVEY:

According to M. Mohandas [1] in this research, sign language is still the most popular way for the deaf and hard of hearing to communicate. This report also identifies the primary difficulties associated with Arabic sign language and suggests possibilities for future research.

In the present research, G. Khurana, G. Joshi, and J. Kaur [2] offer a simple and clear method for shape-based recognition of Indian Sign Language (ISL). The minimum Euclidian distance is employed in the recognition of sign language. There are 26 (ISL) alphabet databases, of which 19 different alphabets are taken into consideration. ISL database generated with a webcam in the background. In both METLAB and WEKA, 14 alphabets provide 100% accuracy. The ISL recognition algorithm is implemented with Windows 7 (64 bit) and METLAB. A two-megapixel web camera is used to take pictures.

In the present analysis, V. Kulkarni [3] proposes a sign language recognition system. Indian Sign Language (ISL) database building is completed at the initial stage. The following phase involves segmentation and hand tracking. This article presents the outcomes of the successful implementation of the system. The outcomes show how motion tracking, edge detection, and skin colour detection function both separately and in concert. Alphabets A through Z and numbers 1 through 10 are included in the gesture in the database.

In this article, D. Jain, A. Saxena, and A. Singhal [4] introduce principal component analysis, a quick and effective method for identifying sign gestures. Three frames per second are captured from the video stream by the proposed technique. To determine their significance, these pictures are compared to a database. With a roughly 90% matching rate, this system has been successfully tested and developed in a real-world setting. A webcam and an

Android tablet were used in a lab to generate a database of ten sign gestures that are derived from Indian sign language. The 60X80 pixel sign gesture in the system database reduces processing time and memory requirements. The achieved recognition rate is within an acceptable range, ranging from 70 to 80 percent. Black background is required by the system.

In the present investigation, T. Ayshee, S. Raka, Q. Hasib, Md. Hossain, R. Rahman [5] build an intelligent system that can function as a translator between spoken and sign language in Bangali by using image processing and a fuzzy rule-based system. In the beginning, raw image data is analyzed, and angles are measured to identify the rules. The technique is tested using just two Bangali signs at first. There were 22 input photos in total for the two gestures in the test dataset. METLAB was used in the system's implementation.

III. BLOCK DIAGRAM:

The Figure 1 shows the working model of the Text / Sign language conversion. The input is given with the text or Sign language and it is detected by the Language Detection and then it is processed with the Quasi helical and neural network algorithm then the input is divided into segments and these data are matched with the database then the appropriate output is produced in the text/sign language.

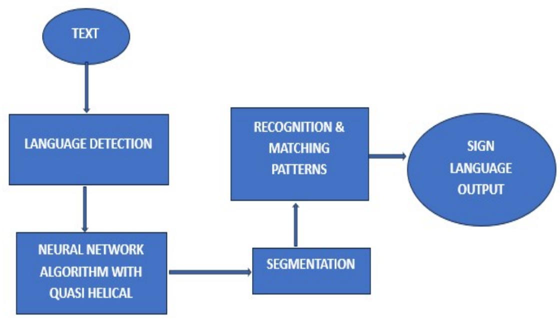


Figure 1

IV. WORKING FLOW

The Figure 2 shows that the input is given in the form of Text. This text is further processed then it is segmented and matched with the database. According to the database the Sign Language Output is produced.

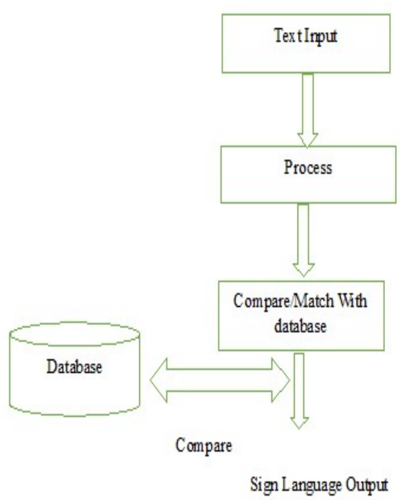


Figure 2

V.WORKING PROCESS

A. INPUT SIGN IMAGE:

An alphabet image, a word image, or a phrase image can be the input image. Every process that is carried out makes use of a trained and preconfigured database. The input image will be saved in the database in the appropriate format.

(ii) INPUT TEXT:
Text will be saved in an alphabetic, word, or sentence format, similar to the input image. The keyboard is used for entering text into the system.
The Figure 3 shows that the input is given in the form of Sign Language. This Sign Language is further processed with the Gray Scale Image and Edge Detection then it is segmented and matched with the database. According to the database the Text Output is produced.

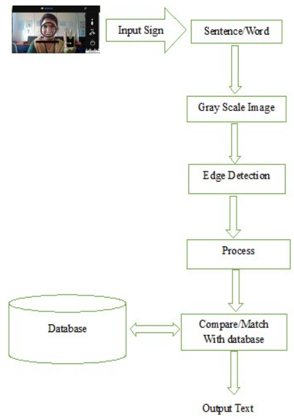


Figure 3

(iii) SIGN LANGUAGE DETECTION:
Finding the existence of signing activity in a video frame is known as "sign language detection." To encode frames and identify signature actions, a variety of techniques have been used, including spatial-temporal analysis with Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). For real-time detection, recent techniques additionally take motion analysis and human posture estimate into account. Since sign language is a unique language with its own set of linguistic conventions, the Sign Language Typology (SLT) model adheres to the Neural Machine Translation (NMT) framework. Assume for the moment that y= (y_1, y_2, y_3, …., y_{Ty}) the output sentence matches the training set's sign video frame sequence x= (x_1, x_2, x_3, …., x_{Tx}). We first optimize the frame-level input words using the unsupervised Frame Stream Density Compression (FSDC) algorithm module. After that, frames are convoluted using a spatial Convolutional Neural Network (CNN) to create a tokenization layer, which is then sent into the NMT module for encoding and decoding.

(iv) Unsupervised FSDC Module:

The geographical CNN is mostly used to decrease the amount of fine-grained video frame input. The video frame is the most fundamental input unit in SLT. The quality of the tokenization layer and the processing performance of CNN are directly impacted by the compression of video frames. Consequently, optimizing the tokenization layer entails optimizing the number of frames as well.

The video has to be framed in a dataset at a given Frames Per Second (FPS), which results in a large number of similar redundant frames in the temporal vicinity. An example would be someone signing the same sign language quickly and slowly, respectively.

The Figure 4 explains the concept of reducing the amount of video frame input. To minimize the frames of the input sign language conversion. The Figure 5 shows that the FSDC algorithm to avoid the duplicated frames in the sign language input.

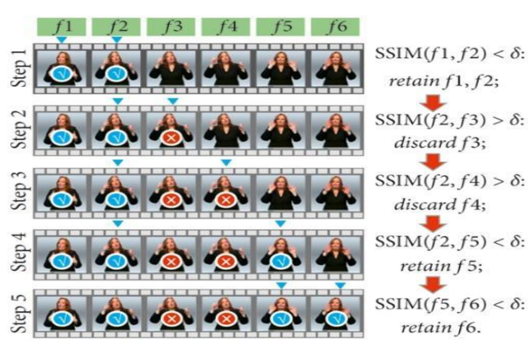


Figure 4

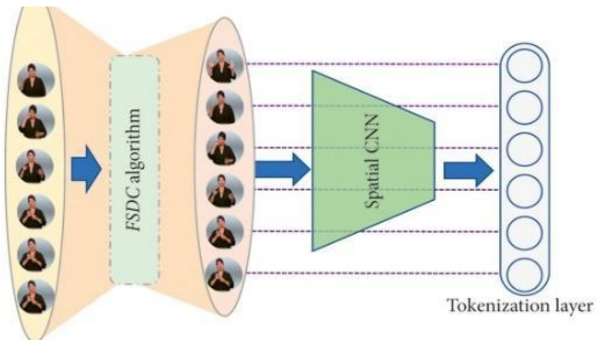


Figure 5

Despite having the same message, the two create videos that are different in length. It goes without saying that a slower-moving video will include more duplicated frames in the same temporal region.

B.PRE-PROCESSING:

Choosing an image or text input is a part of pre-processing. Then follows edge detection, grayscale convergence, picture array creation, and database comparison or matching. These features are all included in preprocessing. Preprocessing will be carried out correctly and without difficulty if the system receives input that is accurate or entered in an appropriate way.

(i) SIGN LANGUAGE RECOGNITION, TRANSLATION, AND PRODUCTION:

Translation of spoken language to sign language (SLT), identification of signs in films (SLR), and creation of sign language movies from spoken language text are all included in sign language recognition, translation, and production. In recognition tasks, signs are identified and labelled in isolated or continuous films; in translation and production jobs, computer vision and natural language processing techniques are used to convert between various data representations. Processing visual information in signed language movies is made easier by combining linguistic theories with computer vision techniques like CNNs. However, present methods do not adequately handle higher-level linguistic aspects, such as lexical and structural distinctions, and future research into these areas is necessary to address the linguistic complexity of signed languages. Furthermore, for more accurate evaluations and better performance in real-world circumstances, full datasets including distractions and interferences must be developed.

C. EDGE DETECTION:

The Canny edge detection algorithm is implemented for edge detection. The process requires less time while providing superior accuracy. Canny's algorithm performs better at eliminating noise and identifying accurate and unambiguous input that the system needs. Low error rate, localized edge point response, and single edge point response are provided by Canny's algorithm. It first applies a Gaussian filter, which causes the image to become slightly blurry. Then, using the gradient on the x and y axes, identify edges. Following edge detection, an array of the input image is created using the edges that were found. This array is kept in the database, which will be utilized in the future for identification.

(i) SIGN LANGUAGE IDENTIFICATION:

Determining the precise signed language utilized in a particular movie entail examining visual indicators like finger spelling patterns and distinctive linguistic elements seen in various signed languages. While more recent methods use activity maps and distinguishing factors to distinguish across signed languages, earlier attempts used basic classifiers.

(ii) SEGMENTING SIGN LANGUAGE:

Because there is no linear modelling and signs occur simultaneously, sign language segmentation is more difficult than that of spoken languages. Present techniques divide up units that roughly correspond to signed language units, but they don't specifically take prosody or other linguistic variables into account. Sequence alignment, using classifiers based on spatiotemporal features, graph convolutional networks, and deep learning architectures are some of the methods used to find the temporal boundaries between signs.

D. DATA BASES:

The Figure 6 represents the databases inserted in the backend server of the Text to Sign language Learning Tool. The input (Text/Sign Language) is converted and processed matched with these data bases and produce an appropriate output corresponding to the input.

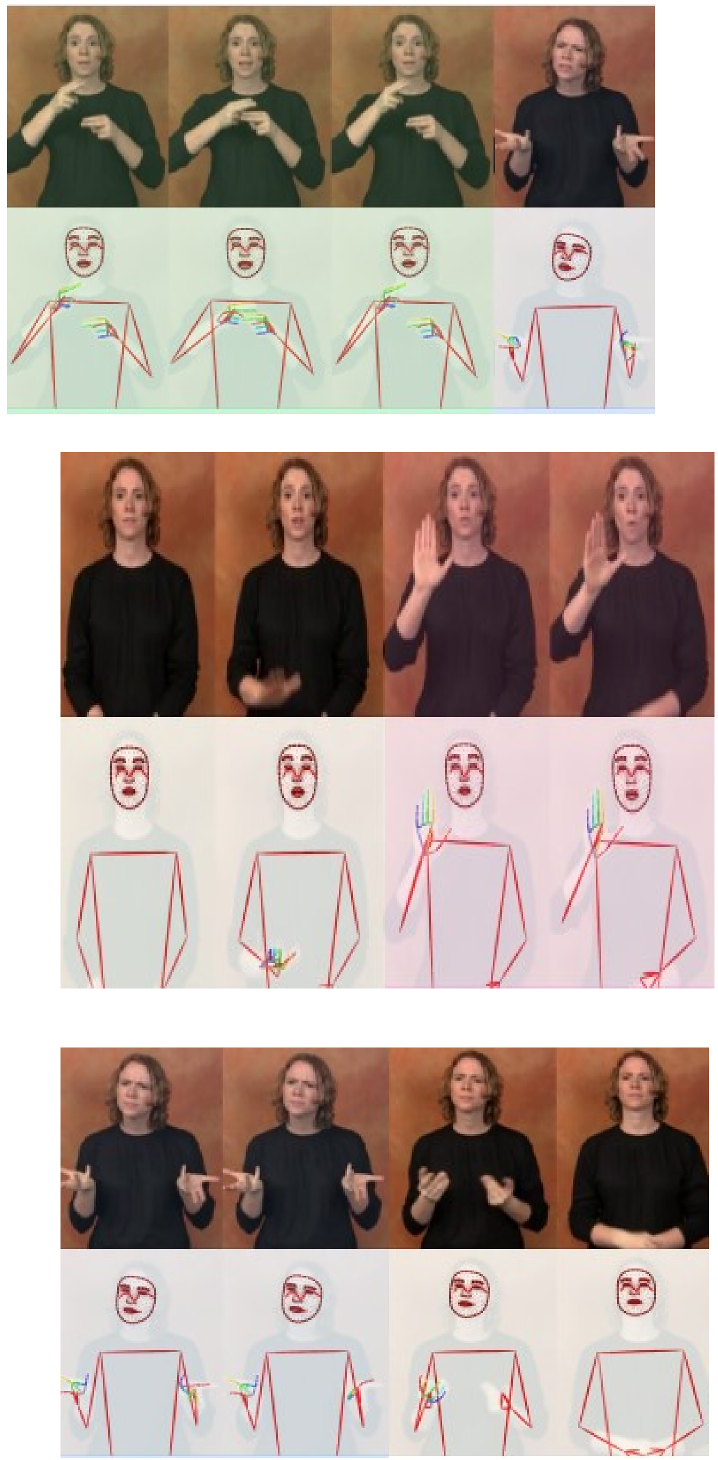
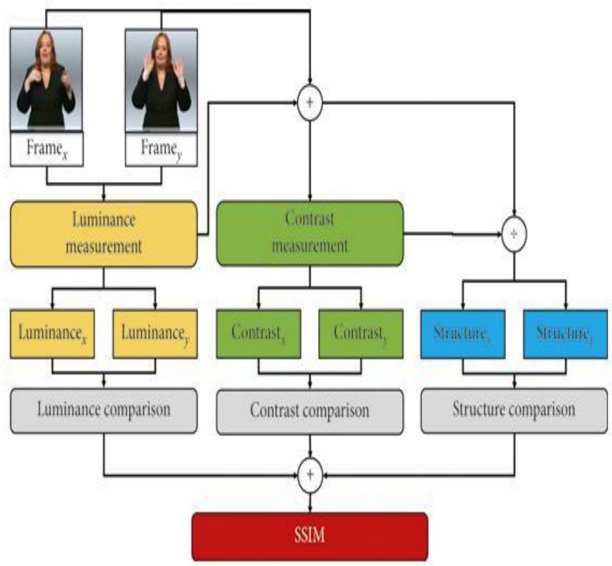


Figure 6

(i) PATTERN RECOGNITION/MATCHING:

The parameters that are extracted from the text or image input are compared to the database. The corresponding result appears once the right data points have been matched.

The FSDC algorithm is suggested as a way to lessen this impact. By comparing the similarity index, we eliminate the less significant frames while maintaining the fixed frame sequence. Theoretically, it aids in lowering the quantity of training data and errors produced on account of sign speed and FPS. To determine how similar two photos are, we employ the Structural Similarity Index Measure (SSIM) method [11], which approximates the visual sense that humans have. The associated computation flow chart is displayed in Figure 3 for determining the



structural similarity of frame and frame. The following is the SSIM algorithm's formula:

$$SSIM(fi,fj) = [L((fi,fj))]^{x} \,.\, [C((fi,fj))]^{y} \,.\, [S((fi,fj))]^{z}$$

Figure 7

The SSIM indices are determined by the FSDC for every frame as well as every frame in the neighbourhood. The remaining frames will be removed as redundant if the SSIM index is higher than a predetermined threshold, leaving only one of them. Figure 7 displays an example of Algorithm 1 in action.

INPUT: input F; threshold Δ ($0 < \Delta < 1$); number of video frames N.
OUTPUT: F
Initialize x=0, i = 1
for x + i < = N, do
    if SSIM (f_x, f_x + i) > Δ, then
        Retain x + i, discard f_x + i, update i= i + 1
    Else if SSIM (f_x, f_x + i) <= Δ, then
        Retain f_x, f_x + i, update x = x + i, i=1
End if
End for

Formally, as seen in Figure 5, we investigate frame-level input tokenization and translate the feature vectors to the tokenization layer as

$$\Gamma = SpatialCNN(FSDC(x))$$

E. TEXT/SIGN OUTPUT:

In the event that the system receives input in the form of a sign, the result will be text representing the meaning of the sign language image.

F. OUTPUT:

In this project, if the input is given as text, the output occurs in sign language with the help of our pose stream. In this picture (Figure 8), we give the input as prayer; the actual output is shown in the pose stream. The input is also given by the sentence pattern this input is accepted, segmented, matched with the pre-processed databases, and given the sign language output by the pose stream.

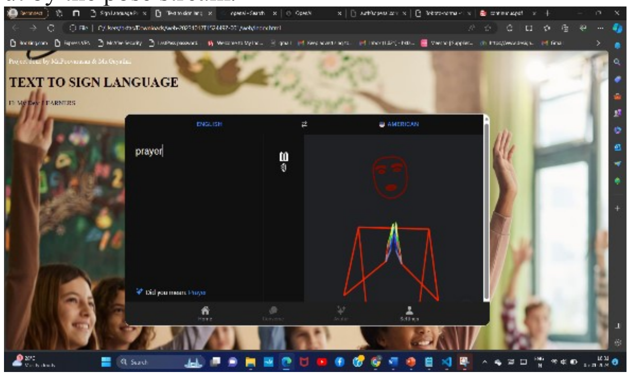


Figure 8

The Figure 9 is an example for a Sign language input and it shows the alphabet "w", the facial expressions and the hand gestures are recognized and the output is generated according to the input.

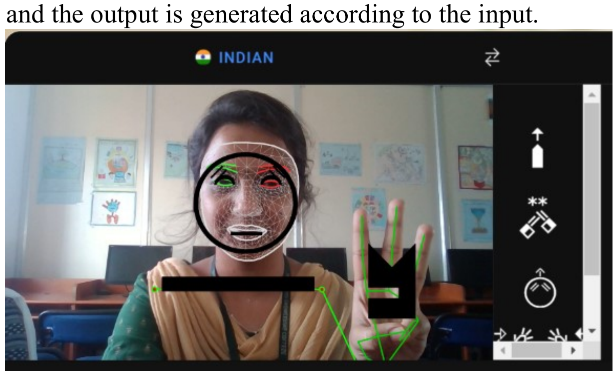
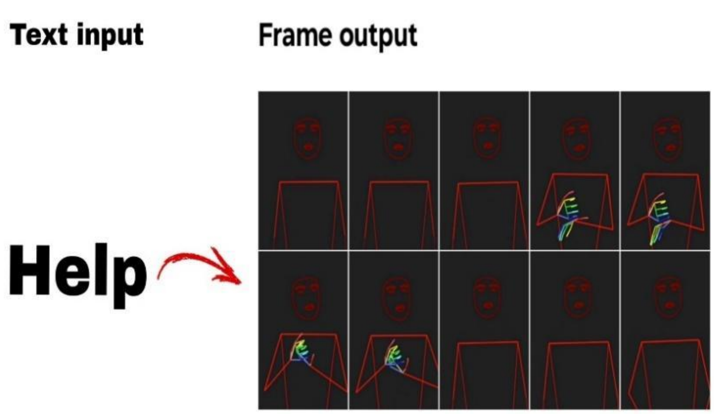


Figure 9

RESULT & ANALYSIS:

The following Figure (10,11,12) shows that some examples for the text to sign language conversion. The some of the examples are Prayer, Help & sorry.



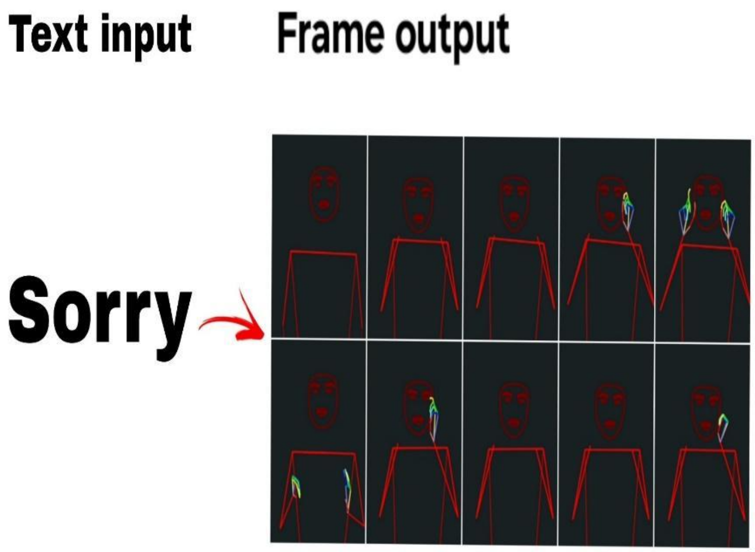


Figure 12

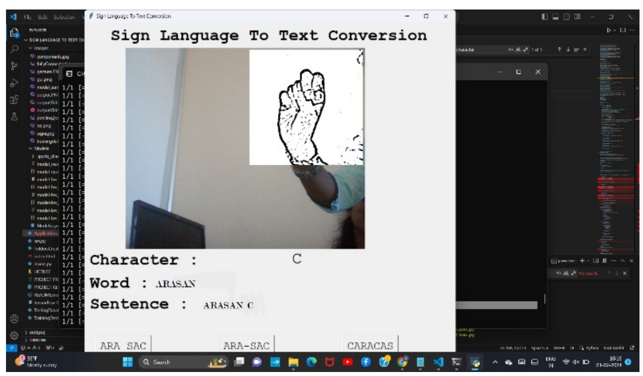


Figure 13

The following Figure 13 shows the sign language to text conversion.

G. CONCLUSION:

In summary, technology that translates text to sign language is essential to overcome barriers to communication between the deaf and hearing cultures. This technology makes it feasible for deaf people to access information and communicate more effectively in their native language by converting written text into sign language with creative algorithms and advances in machine learning. High accuracy, naturalness, and cultural sensitivity in sign language production remain obstacles despite impressive advances. Further investigation, cooperation, and interaction with the deaf community are essential to improving these systems and guaranteeing their applicability, inclusiveness, and relevance in a variety of language and cultural settings. The development of Text-to-Sign Language technology has promise for increasing the deaf community's accessibility, equality, and integration in a variety of contexts.

REFERENCES

[1] Mohandes, M., Deriche, M. and Liu, J., 2014. Image-based and sensor-based approaches to Arabic sign language recognition. IEEE ransactions on human-machine systems, 44(4), pp.551-557.

[2] Khurana, G., Joshi, G. and Kaur, J., 2014, March. Static hand gestures recognition system using shape-based features. In 2014 Recent Advances in Engineering and Computational Sciences (RAECS) (pp. 1-4). IEEE.
[3] Nanivadekar, P.A. and Kulkarni, V., 2014, April. Indian sign language recognition: database creation, hand tracking and segmentation. In 2014 International conference on circuits, systems, communication and information technology applications (CSCITA) (pp. 358-363). IEEE.
[4] Saxena, A., Jain, D.K. and Singhal, A., 2014, April. Sign language recognition using principal component analysis. In 2014 Fourth International Conference on Communication Systems and Network Technologies (pp. 810-813). IEEE.
[5] Ayshee, T.F., Raka, S.A., Hasib, Q.R., Hossain, M. and Rahman, R.M., 2014, February. Fuzzy rule-based hand gesture recognition for Bengali characters. In 2014 IEEE International Advance Computing Conference (IACC) (pp. 484-489). IEEE.
[6] El-Bendary, N., Zawbaa, H.M., Daoud, M.S., Hassanien, A.E. and Nakamatsu, K., 2010, October. Arslat: Arabic sign language alphabets translator. In 2010 international conference on computer information systems and industrial management applications (CISIM) (pp. 590-595). IEEE.
[7] Baowidan, S., 2021. Improving realism in automated fingerspelling of American sign language. Machine Translation, 35(3), pp.387-404.
[8] Berndt, D.J. and Clifford, J., 1994, July. Using dynamic time warping to find patterns in time series. In Proceedings of the 3rd international conference on knowledge discovery and data mining (pp. 359-370).
[9] Blattmann, A., Rombach, R., Ling, H., Dockhorn, T., Kim, S.W., Fidler, S. and Kreis, K., 2023. Align your latents: High-resolution video synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 22563-22575).
[10] Bragg, D., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., Caselli, N., Huenerfauth, M., Kacorri, H., Verhoef, T. and Vogler, C., 2019, October. Sign language recognition, generation, and translation: An interdisciplinary perspective. In Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility (pp. 16-31).
[11] Zwitserlood, I., Verlinden, M., Ros, J., Van Der Schoot, S. and Netherlands, T., 2004, June. Synthetic signing for the deaf: Esign. In Proceedings of the conference and workshop on assistive technologies for vision and hearing impairment (CVHI).
[12] Kausar, S., Javed, M.Y., Tehsin, S. and Anjum, A., 2016. A novel mathematical modeling and parameterization for sign language classification. International Journal of Pattern Recognition and Artificial Intelligence, 30(05), p.1650009.
[13] de Avellar Sarmento, A.H. and Ponti, M.A., 2023. A Cross-Dataset Study on the Brazilian Sign Language Translation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 2816-2820).
[14] Adaloglou, N., Chatzis, T., Papastratis, I., Stergioulas, A., Papadopoulos, G.T., Zacharopoulou, V., Xydopoulos, G.J., Atzakas, K., Papazachariou, D. and Daras, P., 2021. A comprehensive study on deep learning-based methods for sign language recognition. IEEE Transactions on Multimedia, 24, pp.1750-1762.
[15] Albanie, S., Varol, G., Momeni, L., Bull, H., Afouras, T., Chowdhury, H., Fox, N., Woll, B., Cooper, R., McParland, A. and Zisserman, A., 2021. Bbc-oxford british sign language dataset. arXiv preprint arXiv:2111.03635.
[16]. C.Nagarajan and M.Madheswaran - 'Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter' - Journal of ELECTRICAL ENGINEERING, Vol.63 (6), pp.365-372, Dec.2012.
[17].C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis'- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011.
[18].C.Nagarajan and M.Madheswaran - 'Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques'- Taylor & Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011.
[19].C.Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.
[20] Nagarajan C., Neelakrishnan G., Akila P., Fathima U., Sneha S. "Performance Analysis and Implementation of 89C51 Controller Based Solar Tracking System with Boost Converter" Journal of VLSI Design Tools & Technology. 2022; 12(2): 34–41p.
[21].C. Nagarajan, G.Neelakrishnan, R. Janani, S.Maithili, G. Ramya "Investigation on Fault Analysis for Power Transformers Using Adaptive Differential Relay" Asian Journal of Electrical Science, Vol.11 No.1, pp: 1-8, 2022.
[22].G.Neelakrishnan, K.Anandhakumar, A.Prathap, S.Prakash "Performance Estimation of cascaded h-bridge MLI for HEV using SVPWM" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:750-756
[23].G.Neelakrishnan, S.N.Pruthika, P.T.Shalini, S.Soniya, "Perfromance Investigation of T-Source Inverter fed with Solar Cell" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:744-749
[24].C.Nagarajan and M.Madheswaran, "Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation" has been presented in ICTES'08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007
[25].M Suganthi, N Ramesh, "Treatment of water using natural zeolite as membrane filter", Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022
[26].M Suganthi, N Ramesh, CT Sivakumar, K Vidhya, "Physiochemical Analysis of Ground Water used for Domestic needs in the Area of Perundurai in Erode District", International Research Journal of Multidisciplinary Technovation, pp: 630-635, 2019