IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Signspeak: American Sign Language To Text And Speech Convertor Using Machine Learning

Aishwarya Dinesh Warulkar, Prof. P.N. Deshmukh

Department of MCA (Engg)

Gokhale Education Societies R.H. Sapat College of engineering T. A. Kulkarni Vidyanagar, college road, Nashik, MH India422005

Abstract: Sign language is one of the oldest and most natural form of language for communication, hence we have come up with a real time method using neural networks for finger spelling based American sign language. Automatic human gesture recognition from camera images is an interesting topic for developing vision. We propose a convolution neural network (CNN) method to recognize hand gestures of human actions from an image captured by camera. The purpose is to recognize hand gestures of human task activities from a camera image. The position of hand and orientation are applied to obtain the training and testing data for the CNN. The hand is first passed through a filter and after the filter is applied where the hand is passed through a classifier which predicts the class of the hand gestures. Then the calibrated images are used to train CNN.

Index Terms – American Sign Language, Hand Gesture Recognition, Machine learning, Deep Learning, Hand detection, Convolutional Neural Network (CNN), Computer Vision

INTRODUCTION

American Sign Language (ASL) is a predominant form of communication for the deaf and mute (D&M) community, enabling them to express thoughts and messages through hand gestures rather than spoken language. Communication involves the exchange of ideas and information through various means, such as speech, signals, behaviour's, and visuals. For D&M individuals, sign language serves as the primary method to convey their ideas, utilizing hand gestures understood visually. These gestures, which form the basis of ASL, are essential for their nonverbal communication [1].In our project, we focus on developing a robust model that can recognize fingerspelling-based hand gestures to form complete words by combining each gesture. Fingerspelling is a critical component of ASL, used to spell out words, particularly names and technical terms, that do not have a specific sign[2]. The recognition of these gestures involves capturing images of hand movements, processing these images to identify distinct gestures, and then interpreting these gestures to form words and sentences. We employ convolutional neural networks (CNNs), a class of deep learning algorithms particularly well-suited for image recognition tasks. CNNs have shown remarkable success in various applications, including medical image analysis, facial recognition, and now, sign language recognition[3]. The images of hand gestures are processed through several layers of the CNN to extract and learn features that are essential for accurate gesture recognition. By training the CNN on a large dataset of ASL gestures, the model can achieve high accuracy in real-time gesture recognition.

The importance of this technology cannot be overstated. Accurate and real-time ASL recognition can significantly improve the quality of life for D&M individuals by enhancing their ability to communicate with the hearing population. This can lead to better educational outcomes, more employment opportunities, and a more inclusive society [4]. Furthermore, the ability to automatically recognize ASL gestures can be integrated into various applications, such as virtual assistants, translation services, and educational tools, providing widespread benefits [5]. To achieve these goals, our project involves several key steps. First, we collect a comprehensive dataset of ASL fingerspelling gestures. This dataset includes images of different hand shapes and orientations under various lighting conditions to ensure robustness. Next, we pre-process these images to standardize their size and enhance the features relevant to gesture recognition. The pre-processed images are then used to train the CNN model, which learns to identify the distinct features of each gesture. Finally, we evaluate the model's performance using a separate test set and refine it to improve accuracy and speed. Our project leverages advanced deep learning techniques to develop a real-time ASL fingerspelling recognition system. By accurately recognizing hand gestures, this system can translate ASL into spoken or written language, thereby bridging the communication gap between D&M individuals and the hearing population. This not only empowers the D&M community but also promotes greater inclusivity and accessibility in society.

1. LITERATURE REVIEW

The development of automatic American Sign Language (ASL) recognition systems has been an area of significant research interest, driven by the need to bridge communication gaps between the deaf and mute (D&M) community and the hearing population. This literature review examines various methodologies, technologies, and advancements in the field of sign language recognition, focusing on the use of deep learning techniques, particularly convolutional neural networks (CNNs), for recognizing ASL finger spelling.

Early Approaches to Sign Language Recognition

Early attempts at sign language recognition primarily relied on sensor-based methods, which involved using gloves equipped with sensors to capture hand movements and gestures. These systems, while innovative, were often cumbersome and limited in their practicality due to the need for specialized hardware. For example, Zimmerman et al. (1987) developed a glove-based system that could capture the position and movement of the fingers using sensors, but the technology was expensive and not user-friendly for everyday applications.

Vision-Based Recognition Systems

As computer vision technology advanced, researchers began exploring vision-based approaches, which involve using cameras to capture images or videos of hand gestures. This method is more user-friendly and scalable compared to sensor-based systems. Sterner and Pentland (1995) were among the pioneers in this area, developing a real-time hidden Markov model-based system for recognizing American Sign Language using wearable computing . While their system showed promise, it required extensive computational resources and struggled with variations in lighting and background conditions.

Machine Learning and Feature Extraction

The introduction of machine learning techniques brought significant improvements to sign language recognition. Researchers began using feature extraction methods to identify key characteristics of hand gestures, such as shape, orientation, and movement. Thangali et al. (2011) utilized histogram of oriented gradients (HOG) and scale-invariant feature transform (SIFT) to extract features from images of hand gestures and trained support vector machines (SVMs) for classification. These approaches improved accuracy but were still limited by the handcrafted nature of feature extraction.

Deep Learning and Convolutional Neural Networks

The advent of deep learning, particularly convolutional neural networks (CNNs), revolutionized image recognition tasks, including sign language recognition. CNNs automatically learn hierarchical feature representations from raw image data, eliminating the need for manual feature extraction. Liang et al. (2014) applied CNNs to sign language recognition and demonstrated significant improvements in accuracy and

robustness. Their system could handle variations in lighting, background, and hand shapes more effectively than traditional machine learning approaches.

Recent Advancements in ASL Recognition

Recent studies have focused on further refining CNN-based models and integrating them with other deep learning architectures to enhance performance. For instance, Zhang et al. (2019) combined CNNs with recurrent neural networks (RNNs) to capture both spatial and temporal features of hand gestures in video sequences, achieving state-of-the-art results in dynamic sign language recognition. Additionally, researchers have explored data augmentation techniques to increase the diversity of training datasets, thereby improving the generalization capabilities of the models.

Challenges and Future Directions

Despite significant advancements, several challenges remain in developing robust ASL recognition systems. One major challenge is the variability in hand shapes, sizes, and skin tones among different users, which can affect the accuracy of recognition. Moreover, the real-time processing requirements necessitate efficient algorithms that can operate on standard hardware without significant latency.

Future research is likely to focus on addressing these challenges by exploring hybrid models that combine CNNs with other machine learning techniques, such as attention mechanisms and transformer networks, to enhance the contextual understanding of gestures. Additionally, leveraging large-scale datasets and transfer learning could further improve model accuracy and robustness.

2. METHDOLOGY

As shown in figure (Fig-1), the recommended system's approach consists of many blocks.

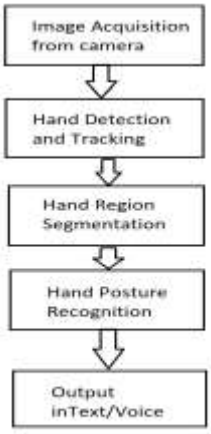

Fig-1: System Flow

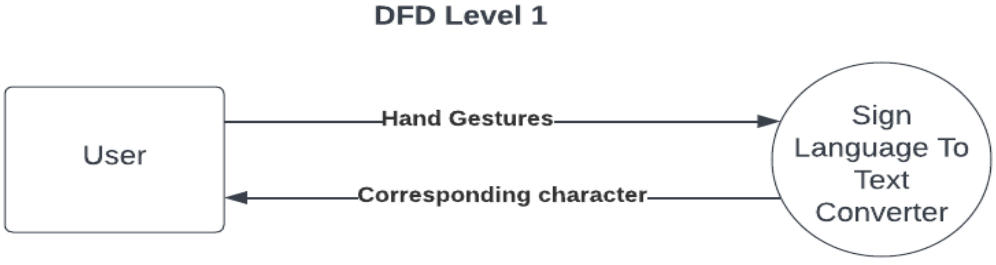

Fig-2: Data Flow Diagram 1

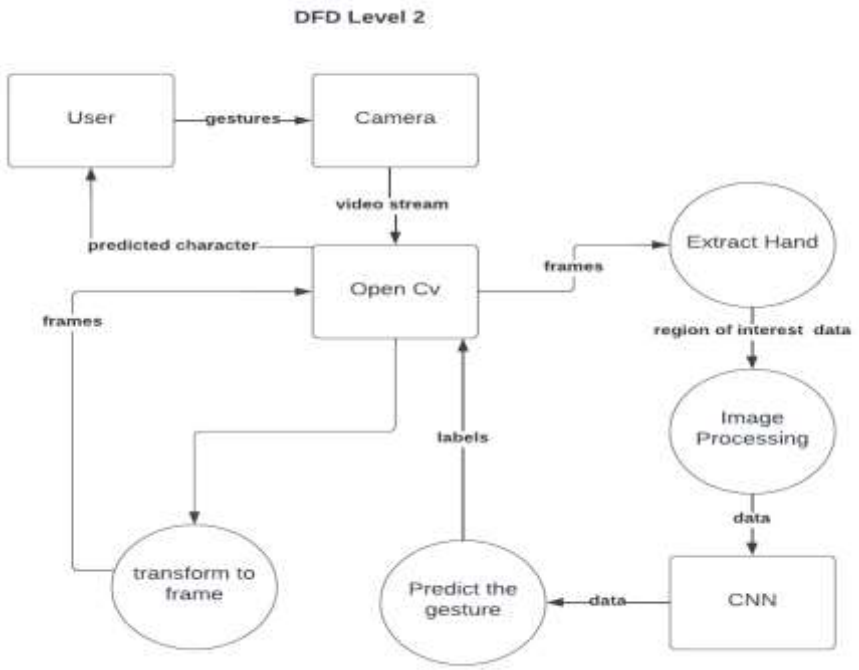

Fig-2: Data Flow Diagram 2

Data Acquisition:

The different approaches to acquire data about the hand gesture can be done in the following ways:
It uses electromechanical devices to provide exact hand configuration, and position. Different glove-based approaches can be used to extract information. But it is expensive and not user friendly. In vision-based methods, the computer webcam is the input device for observing the information of hands and/or fingers. The Vision Based methods require only a camera, thus realizing a natural interaction between humans and computers without the use of any extra devices, thereby reducing costs. The main challenge of vision-based hand detection ranges from coping with the large variability of the human hand's appearance due to a huge number of hand movements, to different skin-color possibilities as well as to the variations in viewpoints, scales, and speed of the camera capturing the scene.

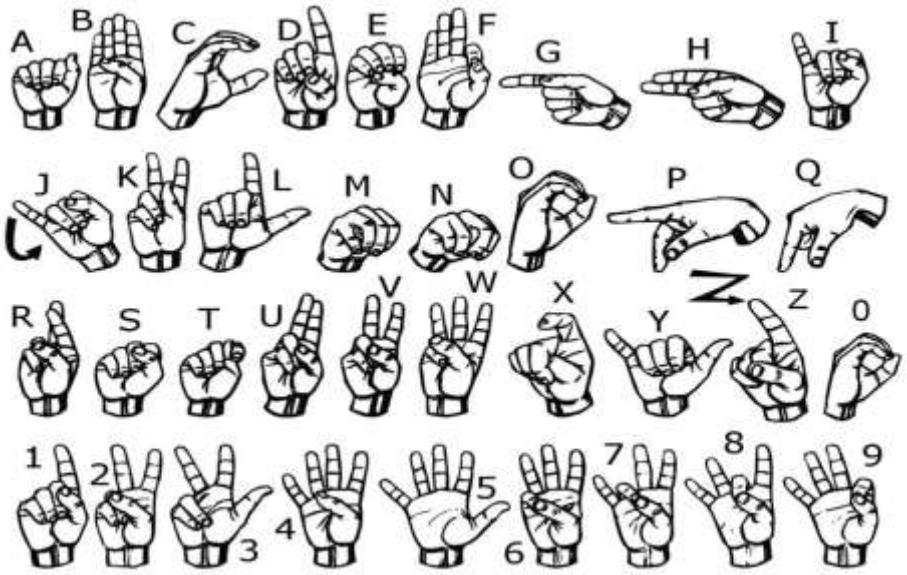

Fig 4: American Sign Language symbols

Data pre-processing and Feature extraction:

In this approach for hand detection, firstly we detect hand from image that is acquired by webcam and for detecting a hand we used media pipe library which is used for image processing. So, after finding the hand from image we get the region of interest (Roi) then we cropped that image and convert the image to Gray image using OpenCV library after we applied the gaussian blur. The filter can be easily applied using open computer vision library also known as OpenCV. Then we converted the Gray image to binary image using threshold and Adaptive threshold methods.

We have collected images of different signs of different angles for sign letter A to Z. in this method there are many loop holes like your hand must be ahead of clean soft background and that is in proper lightning condition then only this method will give good accurate results but in real world we don't get good background everywhere and we don't get good lightning conditions too.

So to overcome this situation we try different approaches then we reached at one interesting solution in which firstly we detect hand from frame using Mediapipe and get the hand landmarks of hand present in that image then we draw and connect those landmarks in simple white image

Mediapipe Landmark System:

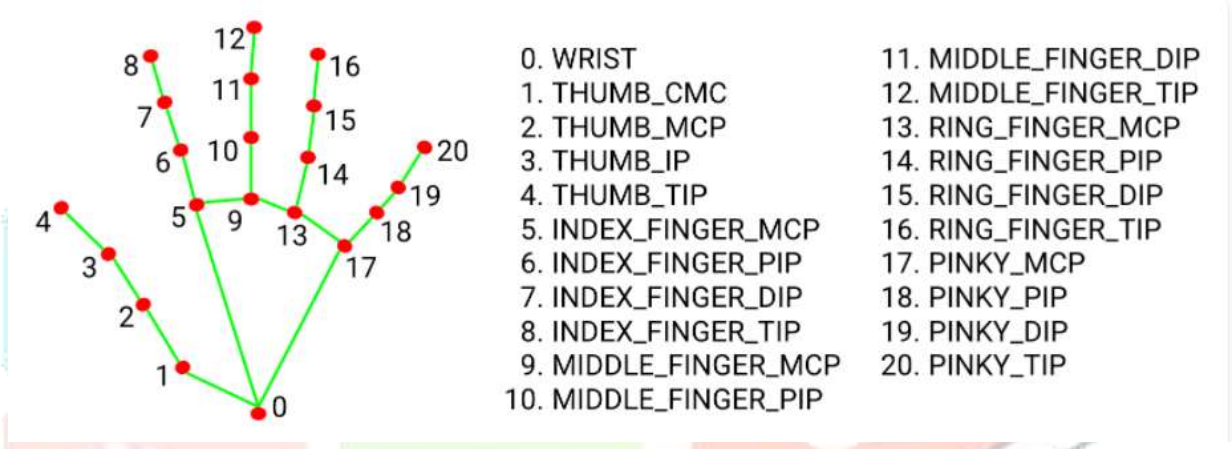

Fig 5: Mediapipe Landmark System

Fig 5.1: A

Fig 5.2: B

-we have collected 200+ skeleton images of Alphabets from A to Z

Gesture Classification:

CONVOLUTIONAL NEURAL NETWORK (CNN)

Convolutional Neural Networks (CNNs) are a powerful tool for machine learning, especially in tasks related to computer vision. Convolutional Neural Networks, or CNNs, are a specialized class of neural networks designed to effectively process grid-like data, such as images**.**

A Convolutional Neural Network (CNN) is a type of deep learning algorithm that is particularly well-suited for image recognition and processing tasks. It is made up of multiple layers, including convolutional layers, pooling layers, and fully connected layers. The architecture of CNNs is inspired by the visual

processing in the human brain, and they are well-suited for capturing hierarchical patterns and spatial dependencies within images. Key components of a Convolutional Neural Network include:

Convolutional Layers: These layers apply convolutional operations to input images, using filters (also known as kernels) to detect features such as edges, textures, and more complex patterns. Convolutional operations help preserve the spatial relationships between pixels.

Pooling Layers: Pooling layers down example the spatial components of the info, diminishing the computational intricacy and the quantity of boundaries in the organization. Max pooling is a typical pooling activity, choosing the greatest worth from a gathering of adjoining pixels.

Activation Functions: Non-linear activation functions, such as Rectified Linear Unit (ReLU), introduce nonlinearity to the model, allowing it to learn more complex relationships in the data.

Fully Connected Layers: These layers are liable for making expectations in view of the great level elements advanced by the past layers. They associate each neuron in one layer to each neuron in the following layer.

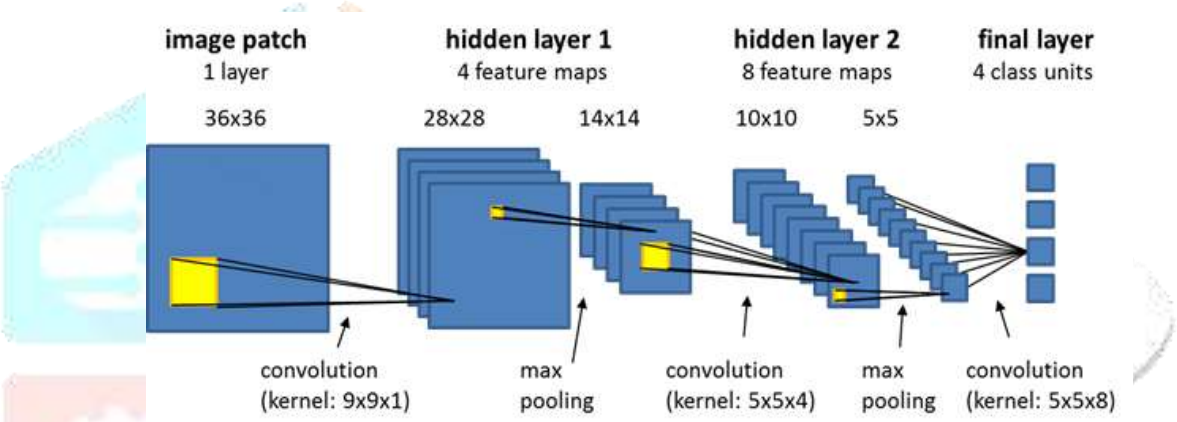

Fig 6: CNN Layers

-Finally, we got 97% Accuracy (with and without clean background and proper lightning conditions) through our method. And if the background is clear and there is good lightning condition then we got even 99% accurate results

4. APPLICATIONS

1**. Education**: Enhancing accessibility to educational materials for deaf or hard of hearing students.

2. **Customer Service**: Facilitating communication between deaf customers and service representatives in various industries.

3. **Meetings**: Enabling participation and inclusion of deaf individuals in meetings and group discussions.

4. **Healthcare**: Improving doctor-patient communication in medical settings for deaf patients.

5. **Public Transportation**: Providing real-time announcements and information for deaf passengers.

6. **Emergency Services**: Ensuring effective communication during emergency situations for deaf individuals.

7. **Legal Proceedings**: Facilitating communication between deaf individuals and legal professionals in courtrooms.

8. **Entertainment**: Making movies, TV shows, and live performances accessible to deaf audiences through captions and translations.

9. **social media**: Enhancing accessibility of social media platforms for deaf users by converting ASL to text or speech.

10.**Remote Communication**: Enabling seamless communication between deaf and hearing individuals in remote or virtual environments.

1. ADVANTAGES

1. **Accessibility**: ASL converters provide a means for individuals who are deaf or hard of hearing to communicate with those who do not understand sign language. By translating ASL gestures into spoken or written language, these converters promote inclusivity and accessibility in various social and professional settings.
2. **Efficiency**: ASL converters streamline communication by converting sign language gestures into text or speech in real time. This allows for more efficient and effective communication between individuals who use sign language and those who do not, reducing the need for intermediaries or interpreters.
3. **Independence**: ASL converters empower individuals who are deaf or hard of hearing to communicate independently without relying on others to interpret for them. This promotes autonomy and self-reliance, enabling individuals to express themselves more freely in diverse situations.
4. **Education**: ASL converters can facilitate learning and comprehension of sign language for individuals who are not fluent in ASL. By providing real-time translations of ASL gestures, these converters serve as educational tools for both deaf and hearing individuals, fostering greater understanding and appreciation of sign language.
5. **Versatility**: ASL converters can be integrated into various devices and platforms, including mobile apps, computers, and assistive technology devices. This versatility allows for widespread adoption and usage of ASL converters across different contexts, from everyday communication to educational and professional settings.

6. SYSTEM REQUIREMENTS

Software Requirement:

1. Operating System: Windows 8 and Above
2. IDE: PyCharm
3. Programming Language: Python 3.9 5
4. Python libraries: OpenCV, NumPy, Keras model, Mediapipe, TensorFlow.

Hardware Requirements:

1. Operating system - windows10
2. Processor: Intel(R) Core (TM) i3-7020U CPU @ 2.30GHz .
3. Installed RAM:4GB
4. System Type: 64-bit operating system, x64-based processor
5. Webcam

7. RESULTS AND DISCUSSION

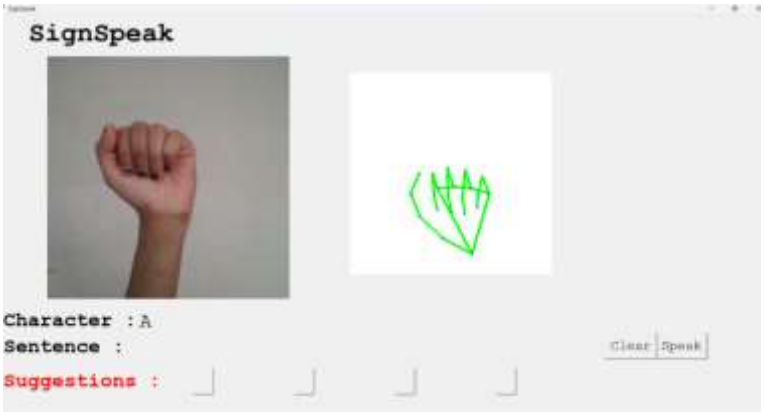

Fig 7: A

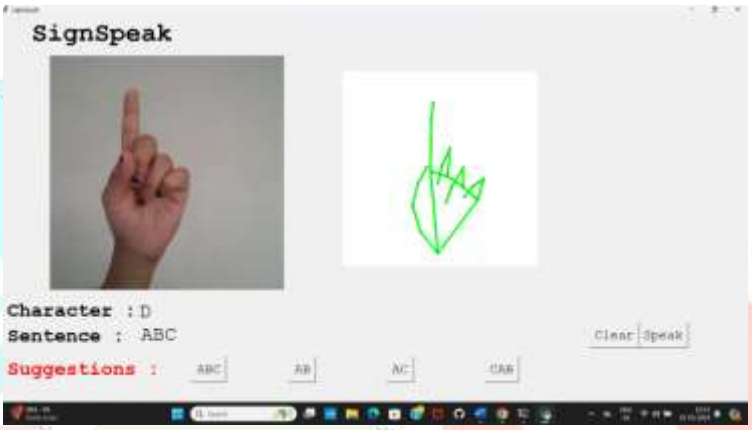

Fig 8: D

Fig 9: ABCD

9. FUTURE SCOPE

1. **Improved Accuracy**: Future advancements aim to enhance the accuracy of ASL converters through refined algorithms and better training data.
2. **Gesture Recognition**: Research focuses on expanding ASL converters to recognize a wider range of gestures and expressions for more nuanced communication.
3. **Multimodal Integration**: Integration with other modalities such as facial expressions and body movements can enrich ASL communication.
4. **Real-Time Feedback**: Development of systems providing real-time feedback and correction to improve user proficiency in ASL.
5. **Mobile Applications**: Increasing availability of ASL converters as mobile apps for on-the-go communication and accessibility.
6. **Customization**: Personalized settings and preferences to tailor ASL converters to individual user needs and preferences.
7. **Inclusive Design**: Designing ASL converters with accessibility and inclusivity in mind to cater to diverse user groups effectively.
8. **Global Adoption**: Promoting adoption and localization of ASL converters worldwide to facilitate communication across different languages and cultures.
9. **Research Collaboration**: Collaborative efforts between academia, industry, and the deaf community to continually advance ASL converter technology.

10. CONCLUSION

In conclusion, the development and implementation of an American Sign Language (ASL) converter hold immense potential to transform communication and accessibility for individuals who are deaf or hard of hearing. By bridging the gap between ASL and spoken or written language, ASL converters offer numerous benefits, including increased accessibility, efficiency, independence, and versatility in various aspects of daily life. From education and customer service to healthcare and emergency situations, ASL converters can revolutionize communication dynamics, fostering inclusivity and empowerment for the deaf community. As technology continues to advance, the continued refinement and integration of ASL converters into diverse applications promise to further enhance accessibility and promote equal opportunities for individuals with hearing disabilities.

11. REFERENCES

[1] T. Yang, Y. Xu, and "A., Hidden Markov Model for Gesture Recognition", CMU-RI-TR-94 10, Robotics Institute, Carnegie Mellon Univ., Pittsburgh, PA, May 1994.

[2] Pujan Ziaie, Thomas M uller, Mary Ellen Foster, and Alois Knoll "A Na¨ıve Bayes Munich, Dept. of Informatics VI, Robotics and Embedded Systems, Boltzmannstr. 3, DE-85748 Garching, Germany.

[3]https://docs.opencv.org/2.4/doc/tutorials/imgproc/gausian_median_blur_bilateral_filter/gausian_median_blur_bilateral_filter.html

[4] Mohammed Waleed Kalous, Machine recognition of Auslan signs using PowerGloves: Towards large-lexicon recognition of sign language.

[5]aeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural Networks-Part-2/

[6] http://www-i6.informatik.rwth-aachen.de/~dreuw/database.php

[7] 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

[8] Zaki, M.M., Shaheen, S.I.: Sign language recognition using a combination of new vision-based features. Pattern Recognition Letters 32(4), 572–577 (2011).

[9] N. Mukai, N. Harada and Y. Chang, "Japanese Fingerspelling Recognition Based on Classification Tree and Machine Learning," *2017 Nicograph International (NicoInt)*, Kyoto, Japan, 2017, pp. 19-24. doi:10.1109/NICOInt.2017.9

[10] Byeongkeun Kang, Subarna Tripathi, Truong Q. Nguyen" Real-time sign language fingerspelling recognition using convolutional neural networks from depth map" 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)

[11] Number System Recognition (https://github.com/chasinginfinity/number-sign-recognition)

[12] https://opencv.org/

[13] https://en.wikipedia.org/wiki/TensorFlow

[14] https://en.wikipedia.org/wiki/Convolutional_neural_nework

[15] http://hunspell.github.io/