WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Application to Detect Real Time Hand Gestures

Vidya H A [1], Sahana A [2], Sanjana S[3], Shamanth S [4]

Information Science and Engineering Department, Sai Vidya Institute of Technology Bengaluru, Karnataka

ABSTRACT

Technological advancement is all about making life more comfortable and efficient, but what about those for whom the very basics of communication present a relentless struggle? The differently abled, particularly those who can't speak, often find themselves cut off from the simplest of exchanges. While sign language offers a critical bridge, it still has its limits—after all, not everyone understands the nuanced gestures and movements that make up this complex system.

Enter Sign Language Recognition. It's a fast- growing field where tech meets human expression, blending body language with code to create a new kind of interaction. This isn't just about translating signs into text; it's about unlocking a whole new way for those with speech disabilities to connect with the wider world. Computers can now recognize various signs and interpret them into text or audio, breaking down barriers that once seemed insurmountable.

But that's not all. The beauty of this technology is its versatility. It's not just for those with speech impairments; anyone, regardless of technical background, can harness its potential. From tech applications to daily interactions, the possibilities are vast. The idea is simple: someone performs a sign in front of a camera, and the system translates it into something everyone can understand—whether it's spoken words or on- screen text.

The implications are profound. No longer is sign language just a niche skill; it's becoming a gateway to inclusion, a tool for bridging gaps between the differently abled and everyone else. This paper aims to explore how these frameworks work, how they can be optimized for speed and accuracy, and how they might transform the way we think about human- computer interaction.

To further enhance accessibility, our system incorporates a text-to-speech (TTS) engine that converts the interpreted text into spoken words. This dual conversion, from sign language to text and then to speech, enables individuals with hearing impairments to communicate effectively with those who may not be proficient in sign language. The proposed Sign Language Detection and Conversion to Text and Speech system holds significant promise in fostering inclusivity and equal participation in various social and professional spheres.

Keywords **Differently abled, Hand Gesture Recognition, Sign Language, Text to Speech.**

1. Introduction

Sign language, a vibrant and fluid mode of communication, is the key to connecting with the deaf and speech-impaired community. However, most people without hearing impairments seldom make an effort to learn it, which creates a divide. This gap leads to isolation for those who rely on sign language to communicate. In India alone, it's estimated that there are between 0.9 and 14 million individuals with hearing impairments. That might not mean much at first glance, but consider this: one out of every five people who are deaf globally is in India. This makes it the country with the highest population of deaf individuals, and consequently, a large user base of sign language.

The solution lies in a system that can take sign language and turn it into readable text or spoken words. It would cut out the middleman—those interpreters who are often needed to bridge the communication gap. Think about it: a seamless and user- friendly setup where a simple hand gesture translates into clear text or speech output. It could open doors for the deaf community, allowing them to interact with the hearing world without barriers.

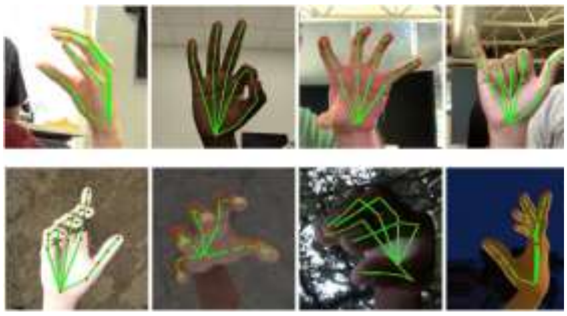
Technology is evolving rapidly, and yet creating an equitable environment for those with disabilities remains a daunting challenge. This sign-to-text/speech conversion framework could be the breakthrough needed to offer equal opportunities for the hearing-impaired. It’s a bold step toward inclusivity in a fast-paced world where everyone deserves a fair chance.

By harnessing the power of artificial intelligence, this system aims to seamlessly translate sign language gestures into written text and spoken words, thereby enabling individuals with hearing impairments to communicate with a broader audience.

The development of this project is grounded in the convergence of interdisciplinary expertise, encompassing computer vision for accurate gesture recognition, natural language processing for text conversion, and text-to-speech technology for vocalization. The integration of these components forms a holistic approach that not only detects and interprets sign language but also ensures the conveyed information is accessible to a wider audience through both written and spoken mediums.

As we embark on this exploration of Sign Language Detection and Conversion to Text and Speech, we delve into the technical intricacies, challenges, and transformative potential of this innovative technology. By addressing the communication barriers faced by individuals with hearing impairments, we envision a future where technology plays a pivotal role in fostering a more inclusive and interconnected global society.

2. Background/Literature Survey



The system revolves around a camera that captures the intricate dance of hand gestures made by users. These hand images become the raw input for a sophisticated algorithm that does the heavy lifting in deciphering these gestures. This algorithm, though straightforward in concept, is a tapestry of four crucial steps: segmentation, orientation detection, feature extraction, and classification.

In the segmentation phase, the goal is simple but critical—separate the hand from the rest of the image. It's a task that requires precision, especially when hands can be in various positions and backgrounds can be distracting. Following segmentation, orientation detection comes into play, where the algorithm determines the spatial alignment of the hand—an essential step, as a slight tilt can change the whole meaning of a gesture.

Feature extraction is the point where things get interesting. Here, the system digs deep to identify the unique characteristics of each hand gesture. Is it a tight fist, an open palm, or something more complex? These features become the building blocks for the final stage: classification. In classification, the system assigns a specific meaning or action to the recognized gesture, transforming raw hand movements into intelligible communication.

One of the most fascinating aspects of this algorithm is its adaptability. It doesn't care who you are or how you gesticulate; it’s designed to be independent of user characteristics. This universality means there's no need for special training to accommodate different users. Unlike some other gesture recognition systems that demand extensive calibration and user- specific tuning, this model is refreshingly flexible, making it more user-friendly.

Hand gestures are captured from live video using a webcam, serving as input for the system. Feature extraction techniques are applied to the captured hand gestures. Feature extraction likely involves identifying and isolating key characteristics of the gestures that are relevant for translation. The system predicts the corresponding symbol or sign based on the extracted features. The results of the prediction are translated into voice format. This implies that the system generates audible output that can be understood by everyone. The voice output serves as a means of communication, allowing dumb individuals to express themselves verbally. This system is highlighted as a valuable tool for improving communication among dumb individuals and the general population. By translating sign language into voice, the system aims to facilitate proper interaction between individuals who are unable to speak and those who can.[2]

A non-parametric method for skin color detection within the context of recognizing sign language symbols using static images. Leveraging the Image Processing Toolbox in MATLAB for implementation, their approach focuses on detecting skin color as a vital step in isolating relevant body parts for subsequent analysis. However, a notable limitation of their work is the reliance on static images for sign language symbol recognition, implying a restricted vocabulary and potential challenges in handling dynamic or continuous signing gestures. The authors' emphasis on a non- parametric approach underscores adaptability to varying data characteristics. While this research contributes to the understanding of skin color detection in sign language

The primary goal of the proposed system is to facilitate seamless communication for individuals who are deaf and mute, allowing them to connect more easily with society.

Image processing forms the core of the system, where hand gestures are captured and analyzed for communication. Angle and peak calculations are employed to extract meaningful information from hand gestures. These calculations likely involve determining the orientation and key points of the hand movements. MATLAB, a widely used numerical computing environment, is utilized for speech synthesis. The authors leverage MATLAB's inbuilt commands to convert the extracted hand gestures into speech. The exact speech synthesis process is not detailed, but it involves using the calculated hand gesture information to generate corresponding vocal output. The system also supports a two-way communication approach by converting speech into gestures. This is achieved through the use of Mel-frequency cepstral coefficients (MFCC).

MFCCs are a representation of the short-term power spectrum of a sound signal and are commonly used in speech and audio processing for feature extraction.[3]

Imagine a system that can take your hand gestures and turn them into speech, all in real time, using vision-based techniques. Sounds like science fiction, right? Yet, this is exactly what this innovative system does. It deploys a variety of algorithms and methods to detect and recognize single-hand gestures, using the hand itself as the primary structure and centering on the centroid for tracking patterns. It doesn't just look at your hand; it studies the intricate movements formed by your fingers and thumb. And here's where it gets fascinating: these patterns are converted into a 5-digit code representation, allowing the system to understand the motion in a way that's almost eerily human.

To achieve this, the system employs advanced techniques like K-means clustering and thresholding to eliminate background noise. It doesn't stop there—text-to-speech APIs come into play for peak detection, and Convex Hull algorithms are utilized to translate gestures into corresponding words or sentences, which are then converted into speech.

Moreover, the system relies on a relatively small dataset, raising concerns about its ability to generalize across a broader range of gestures and users. In other words, what works for one might not work for another.

Despite these limitations, the potential is undeniable. By implementing vision-based techniques and using a range of algorithms, this system is opening doors for people who rely on gestures to communicate.

To supercharge hand gesture recognition for numerical sign language. It starts with capturing video frames from a webcam, but the real magic happens during extraction. Techniques like discrete wavelet transformation (DWT) and singular value decomposition (SVD) break down spatial and temporal components, revealing the crucial features hidden in the gestures.

The paper introduces a fitness function algorithm that trims the excess, reducing the feature set's dimensionality. This slimming down is key for boosting performance. With a more streamlined dataset, support vector machines (SVM), renowned for their accuracy in classification, step in to identify the numerical hand gestures in American Sign Language.

It's all non-touch, meaning it works from a distance—no need for physical contact or special hardware. The paper focuses on numerical sign language, but its implications extend beyond, showing a glimpse of what hand gesture recognition could become with the right approach and technology.

The algorithm operates in the YCbCr colour space, which separates luminance and chrominance components. The key steps involve extracting these components, defining threshold values empirically based on observed skin color characteristics, and creating a binary mask to identify potential skin regions. The mask is generated by applying logical operations to the Y, Cb, and Cr channels, effectively isolating pixels that likely belong to the skin. The original image is then processed using the binary mask, resulting in an output image where non-skin regions are suppressed. The code concludes by displaying both the original and the processed images, providing a visual representation of the skin color detection results. Experimentation with threshold values and color spaces is encouraged to adapt the method to different imaging conditions and skin tones. The constrained vocabulary poses a limitation, prompting consideration for further advancements in dynamic gesture recognition systems.[6]

The proposed system employs gloves equipped with sensors and utilizes Arduino for receiving sensor readings, presenting ahardware- oriented approach to sign language interpretation. However, a notable limitation is highlighted—the system struggles to accurately detect the curves in the fingers. This deficiency has implications for the accuracy of predicting sign language gestures, as finger curvature is integral to forming specific signs. Addressing this limitation could involve exploring advanced sensor technologies or integrating machine learning algorithms to enhance the system's predictive capabilities. The authors may benefit from user feedback to pinpoint instances of less accurate predictions, ultimately guiding iterative improvements in both hardware and software components of the sign language interpretation system.[7]

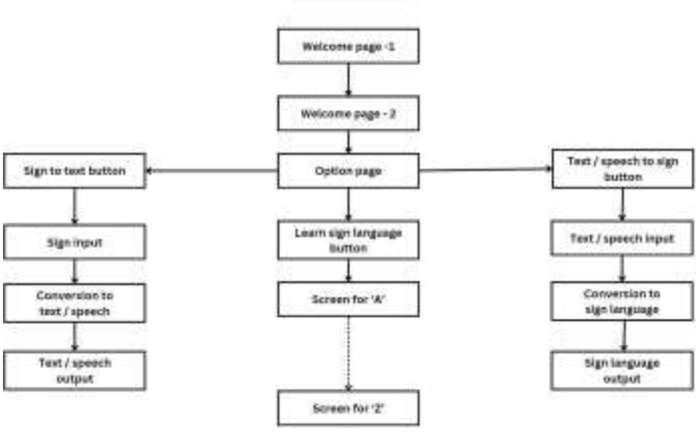
The novel and real time method is shown to distinguish object in affordances from RGBD pictures. The technique employs a Deep Convolution Neural Network with an encoder- decoder design, enabling end-to-end learning of profound features from input data for smooth label prediction. The system leverages multiple modalities to enhance learning efficiency. Notably, this approach sets a new benchmark in identifying the classification of object affordances, demonstrating a 20% improvement in precision compared to state- of-the-art methods using hand-designed geometric features. The method is further applied to a full-size humanoid robot using manipulation strategies. Recognizing the significance of human vision in object detection, the authors emphasize the importance of enabling robots to detect and interact with objects safely. While conventional methods rely on RGB-D images or point cloud data for successful grasping actions, this novel approach treats object affordance detection as a pixel-wise labeling task, employing CNNs to learn deep features from RGBD images. The introduced methodology introduces new data representations, such as Horizontal Disparity, Height Above Ground, and Angle Between Each Pixel's Surface and Normal (HHA), offering advantages in terms of improved results over existing methods for object detection. However, limitations exist, particularly in the grasping method, which is constrained to surfaces that fit the predefined region.[8]

3. Methodology

The proposed system consists of two main stages: (1) Segmentation of hand (2) Recognition of hand sign. The block diagram shows the working of the proposed system. The features of a hand are an important criterion for the classifier to differentiate between the hand gestures. These characteristics must be able to adapt to different hand and gestures by different people. In this system, we have used histograms of oriented gradients (HOG) as a feature descriptor. It is better than other descriptors because it can adapt to changing illuminations and rotation of objectives. It does not consider an image as a whole, but divides it into smaller cells and then for the pixels within the cells edge or gradient direction histogram is calculated. This approach creates a bin, and clubs the histograms of different samples based on magnitude and angle. In the proposed system, we are first segmenting the hand using YCbCr color space and then processing the image through HOG and then provide it to the model. We trained the SVM classifier using 5000 images and developed a model.Besides the above-mentioned neurological graphing methods, the following are a few tasks that help detect SLDs, they are as follows:

1. Line Orientation (L) Judgment Task: This test checks how well someone can

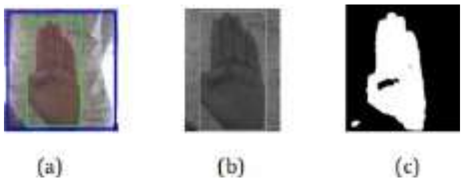
Preparing the dataset We have created a dataset of 26 English language alphabets in Indian Sign Language. Each sign gesture is performed by two different individuals with different hand structure in varied lighting conditions. The videos were recorded on a camera and then each video was broken down frame by frame to images and adjusted to 100 frames and then augmented to get about 250 images for each sign. The data was then divided into 4800 images for training and 1200 images for testing



Sign language interpreter flowchart

Image pre-processing and edge detection

Before performing feature extraction, the images must be processed in such a way that only the useful information is considered and the redundant, distracting noise and superficial data are neglected. The images are first converted to 100 * 100 pixel size for faster computations. The image is then converted to grayscale and finally transform into a binary image. Simultaneously, skin color is detected using YCbCr model. Finally, edge detection is performed using Canny edge detector. The process is illustrated in figure below.



Output of skin colour detection: a) Original cropped image, b) Grey scale converted image, c) Skin colour detection output

Feature extraction

Feature extraction is a crucial aspect of any object detection system, and there are various techniques to achieve it. You might encounter methods like Fourier Descriptor, Scale Invariant Feature Transform (SIFT), or Principal Component Analysis (PCA). Another widely used approach is the Histogram of Oriented Gradients (HOG), which is the focus of this paper. HOG operates on a simple yet effective principle: it represents objects in images by analyzing gradients and edge orientations. The process involves dividing the image into small cells and calculating a histogram of gradient directions for each one. These histograms are then combined to create a feature vector representing the entire image. To ensure accuracy, all cells are normalized to reduce the impact of lighting variations. The result is a robust feature extraction technique that captures the structure and patterns within an image, making it invaluable for object detection systems.

Template matching and sign recognition

The feature vector produced in the above step is fed into an image classifier. In this paper, we have used Support Vector Machine (SVM) for classification. By using SVM classifier we can maximize accuracy and avoid overfitting of data. In SVM the data items are plotted in an n-dimensional space where n is the number of features, each feature is associated with a coordinate value. Then it finds a hyperplane that differentiates the classes. The model is saved for real-time sign language recognition

Text to Speech

We have used Google’s Text to Speech API for transforming the sign language into audio. It is one of the best text to speech API available. Unlike other TTS APIs, this API generates human-like voice. The sign language is interpreted using the above steps and then the result is fed to text to speech function which converts it to audio. In this system, we can see and hear the sign language translation at the same time which makes it very convenient to use.

Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) represent a groundbreaking paradigm in deep learning algorithms, specifically designed for image processing and computer vision tasks. At the core of their functionality is the convolution operation, a process in which a filter or kernel is convolved with an input image. This filter, typically a small learnable matrix, is systematically slid across the entire image, facilitating the extraction of distinctive features. For instance, a filter might identify edges, textures, or intricate patterns. The sliding window mechanism ensures that the network learns to recognize these features across different positions in the image.[9]

The outcome of this convolution operation is a set of feature maps, each capturing the local receptive field and corresponding to a specific feature detected by the filters. Notably, CNNs leverage shared weights and biases throughout the convolutional layers, promoting parameter sharing for increased efficiency and generalization. This characteristic enables the network to learn hierarchical features by stacking multiple layers. Lower layers focus on elementary features, such as edges, while higher layers amalgamate these to recognize more complex patterns or objects. This hierarchical learning process has proven exceptionally effective in tasks like image classification, object detection, and image segmentation, establishing CNNs as indispensable tools in the realm of computer vision.[10]

4. Conclusion

In conclusion, the referenced technologies showcase a diverse array of innovative approaches leveraging image processing, machine learning, and deep learning for the betterment of human-computer interaction and communication. The real-time two-way communication system for the hearing impaired [1] demonstrates the application of image processing in facilitating seamless communication. Vision-based techniques for converting hand gestures into speech [2] underscore the significance of real-time gesture recognition in bridging communication gaps. The reduction of gesture feature dimensions for numerical sign language [3] exemplifies the intersection of numerical recognition and sign language, enhancing the performance of hand gesture recognition. The gesture recognition system for human-computer interaction [4] employs a shape-based approach, demonstrating its effectiveness in recognizing and classifying hand gestures. Translation of sign language to voice for the deaf [5] exemplifies the application of machine learning and Python in providing a voice to sign language users. These technologies collectively highlight the transformative impact of computational intelligence, machine learning algorithms, and image processing in fostering inclusivity and improving the quality of life for individuals with communication challenges. Additionally, the integration of convolutional neural networks (CNNs) for object affordance detection [8] underscores the role of deep learning in enhancing a robot's ability to perceive and interact with its environment. The cited references [6, 7, 9, 10] further contribute to the broader understanding and exploration of CNNs and their applications in computer vision. These advancements collectively propel the field forward, offering promising avenues for future research and development in assistive technologies and human-computer interfaces.

In conclusion, the discussions cover a range of innovative technologies and methodologies aimed at enhancing communication and interaction, particularly for individuals with hearing and speech impairments. Various approaches, such as utilizing Convolutional Neural Networks for object affordance detection and implementing real-time hand gesture conversion systems, showcase the potential of cutting-edge technologies. These advancements not only contribute to the accessibility of information but also promote the integration of individuals with different abilities into society. The mentioned studies demonstrate the continuous evolution of artificial intelligence and computer vision in addressing real-world challenges, fostering a more inclusive and connected global community.

References

[1] Rajesh M. Autee , Shweta S. Shinde, Vitthal K. Bhosale "Real Time Two Way Communication Approach for Hearing Impaired and Dumb Person based on Image Processing", IEEE International Conference on Computational Intelligence and Computing Research in 2016.

[2] G. Mundada, K. Khurana, A. Bagora. S," Real Time Conversion of Hand Gestures to Speech using Vision Based Technique", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278- 3075, Volume-8 Issue-9, July, 2019.

[3] Md Rashedul Islam, Akm Ashiquzzaman, Jungpil Shin, Abdul Kawsar Tushar, and Rasel Ahmed Bhuiyan," Reduction of Gesture Feature Dimension for Improving Hand Gesture Recognition Performance of Numerical Sign Language", 20th International Conference of Computer and Information Technology(ICCIT), 22-24 December of 2017.

[4] Pawan Singh Mehra, Meenakshi Panwar, "Hand Gesture Recognition for the Human Computer Interaction", 2011 International Conference based on Image Information Processing (ICIIP 2011).

[5] Silji Simon C, Stephy Paul, Rosemarry Antony, Scaria Alex," Sign Language Translation to Voice for Dumb People", International Journal of Scientific Research and Engineering Trends . , April 2020.

[6] Deepika Pahuja and Sarika Jain, “Non- parametric approach for skin colour detection, image processing toolbox using MATLAB," IEEE paper on International Conference on Computational Intelligence and Computing Research 2020.

[7] Salma A, Essam El-Din and Abd El- Ghany, “Sign language interpreter system: An alternative for Machine Learning " International Conference on Computational Intelligence and Computing Research in 2020.

[8] A. Nguyen, D. Kanoulas, G. Caldwell, and N. Tsagarakis, “Detecting Object Affordances with Convolutional Neural Networks”, 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016

[9] Christian szegedy, Wei liu, Yangqing Jia et al., “Going Deeper with Convolutions”, Conference on Computer Vision and Pattern Recognition (CPVR) , IEEE explorer, Boston, MA, USA, 2015

[10] Zeiler, M. D. and Fergus, “Visualizing and understanding convolutional networks”. European Conference on Computer Vision, vol 8689. Springer, Cham, pp. 818- 833, 2014.