INDIAN SIGN LANGUAGE RECOGNITION

A report on
Computer Vision Lab Project
[CSE-3181]

Submitted By

MIHIR JATINKUMAR PATEL	**210962192**
PEDDIREDDY SIDDHARTH	**210962124**
KAMATHAM PRANAV KUMAR	**210962122**

MANIPAL
ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MANIPAL INSTITUTE OF TECHNOLOGY,
MANIPAL ACADEMY OF HIGHER EDUCATION

NOVEMBER 2023

Indian Sign Language Recognition using Computer Vision Techniques

Mihir Patel
Computer Science Engineering
(Artificial Intelligence
& Machine Learning)
Manipal Institute of Technology,
India
mihirp741@gmail.com

Peddireddy Siddharth
Computer Science Engineering
(Artificial Intelligence
& Machine Learning)
Manipal Institute of Technology,
India
psiddharth413@gmail.com

Kamatham Pranav Kumar
Computer Science Engineering
(Artificial Intelligence
& Machine Learning)
Manipal Institute of Technology,
India
kamathampranav@gmail.com

Abstract*— This investigation delves into the domain of Indian Sign Language (ISL) recognition, a field that is currently in its nascent stages when compared to its counterpart, American Sign Language (ASL). The diversity of India's linguistic landscape presents unique challenges, including the presence of regional variations and the dual-handed nature of ISL, which often results in the occlusion of critical gestural features. These complexities have hindered the development of robust recognition systems. Addressing this gap, our project contributes to the burgeoning research in ISL by curating a novel dataset encompassing ISL alphabets and numerals. We have employed advanced image preprocessing techniques alongside the Bag of Visual Words (BoVW) model to extract and encode distinctive features from the segmented ISL gesture images. Subsequently, these features are represented as histograms that correlate the visual gestural elements with their corresponding linguistic outputs. The culmination of this process involves the application of supervised machine learning algorithms to classify the ISL signs accurately. Our approach not only paves the way for enhanced ISL recognition systems but also provides a foundational dataset that can catalyse further research in this vital area of computational linguistics and accessibility technology.

Keywords*— Indian Sign Language, Bag of Words, Canny Edge Detection, SIFT, SVM

I. Introduction

The recognition of sign language is a critical step towards bridging the communication gap between the hearing and the deaf communities. While American Sign Language (ASL) has been extensively studied and integrated into various technological solutions, Indian Sign Language (ISL) lags in both research and development. This discrepancy is primarily due to the intricate nature of ISL, which utilizes both hands in its gestural communication, introducing complexity in visual recognition due to occlusion and the vast cultural diversity within India that leads to regional variations in sign language.

ISL's unique challenges stem from its dual-handed gesture system, which, unlike the one-handed gestures of ASL, increases the likelihood of feature occlusion where one hand obscures the other during signing. This occlusion poses significant difficulties for accurate gesture recognition using computer vision techniques. Moreover, the lack of a comprehensive and standardized dataset for ISL further hinders the development of robust recognition systems.

To address these challenges, our project presents a novel approach to ISL recognition by constructing a comprehensive dataset that captures the nuances of ISL gestures for alphabets and numerals. We have employed advanced image preprocessing techniques to enhance feature extraction, which is pivotal for the subsequent classification stages. The Bag of Visual Words (BoVW) model serves as the cornerstone of our feature extraction process, transforming raw image data into a quantifiable format that machine learning algorithms can interpret.

The histograms generated from the BoVW model create a visual vocabulary, mapping each gesture to its corresponding ISL symbol. This mapping is crucial for the supervised learning models that follow, which are trained to discern and classify the ISL signs accurately. By leveraging a combination of Support Vector Machines (SVM) and other machine learning paradigms, we aim to achieve high accuracy in ISL recognition, thus contributing a significant tool for the ISL community and researchers alike.

This paper details the journey from data acquisition to the final classification, highlighting the methodologies employed at each stage to overcome the inherent challenges of ISL recognition. Our contributions are twofold: providing a valuable dataset for ISL and proposing a recognition framework that can serve as a benchmark for future research in this vital area of humancomputer interaction.

II. LITERATURE REVIEW

As per our previous discussion, the use of sign languages is essential for communication within the disabled community. In a research paper by Karishma Dixit et al. [4], a technique is presented to perceive the Indian Sign Language (ISL) and convert it into normal text. This process involves three phases, namely preparation, testing, and classification. The proposed technique utilizes a combination of Hu invariant second and basic shape descriptors to create a new feature vector for sign recognition. A multi-class Support Vector Machine (MSVM) model is used for classification. The effectiveness of the proposed technique is demonstrated on a dataset that consists of 720 images. The trial results indicate that the proposed framework can recognize hand signals with a 96% accuracy rate.

A three-step procedure for identifying ISL gestures is described by Subhash Chand Agrawal et al. [1] in their study: segmentation using the Otsu algorithm, feature extraction using SIFT key points and HOG descriptors, and classification with an MSVM, which yields a 93% accuracy rate. This framework is modified by our approach, which combines segmentation with SIFT and clever edge detection.

Bhumika Gupta [5] proposed a model for perceiving static pictures of the marked letters in the Indian Sign Language (ISL) in order. Unlike other sign languages that use gestures such as the American Sign Language and the Chinese Sign Language, ISL letters use both double hands and single hand, making it easier to recognize the gestures by categorizing them into single hand and double-handed signature. Two different features, HOG and SIFT, were used for the two classes, which were separated for a set of training images and consolidated into a single matrix. To classify the test image, the features of HOG and SIFT for the input test images were joined with the feature matrices of HOG and SIFT. The resultant classification of the test image was obtained by feeding a k-Nearest neighbor classifier with computed correlation for the matrices. Your paper must be in single column format.

M. Grif, R. Elakkiya, M. Bakaev, Alexey L. Prikhodko, and Rajalakshmi E.[9] concentrated on using machine learning to recognize Russian and Indian sign languages. They created a system with non-manual markers, movement, localization, configuration, orientation, and other elements. They used LSTM networks to achieve a 95% accuracy rate for individual gestures and the Mediapipe Holistic Library to achieve an 85% accuracy rate for continuous sign language. This study emphasizes how crucial it is to have more labeled data in order to increase recognition accuracy.

In their presentation, Akansha Tyagi and Sandhya Bansal [10] provided an extensive analysis of feature extraction methods applied to vision-based sign language recognition systems. They finished with future directions for feature extraction in ISL recognition systems and gave a taxonomy of existing methods. The performance of different classifiers used in sign language recognition is heavily reliant on feature extraction, as this review highlights.

Using automatic sign language interpretation, Poonam Pawar, Nikita Mandage, Shreya Sasane, Sakshi Ransing, and Prof. Bhosle Swati [11] sought to offer a communication aid for deaf and hard-of-hearing people. They took a series of pictures of ambidextrous ISL, processed them with Python, and then turned the images into text and speech. Their objective was to develop a system that could interpret ISL gestures into readable text and audible speech.

Figure 1. Gestures for Numbers and Letters in ISL

III. METHODOLOGY

This paper presents an algorithm that uses static images to recognize gestures from Indian Sign Language. As shown in Figure 2, the procedure is divided into four primary phases: gathering images, segmenting and extracting features from images, and classifying the images. The Bag of Visual Words (BoW) model, an idea taken from the Natural Language Processing (NLP) domain, helps with the classification. The BoW model is similar to an independent feature representation technique based on histograms in image processing, which enables an image to be read as a document. For the BoW model to effectively represent gestures, this interpretation is essential. In order to define "words" within images, feature descriptions and codebooks—also referred to as visual words—are usually created. It is possible to create histograms for every image using these codebooks. A Support Vector Machine (SVM) model is then used to classify the images.

1. Image Collection

Research on Indian Sign Language (ISL) is still in its infancy, and complete datasets are notably lacking. In order to rectify this disparity, we have created a large ISL dataset of 35 categories, each with 1,200 photos, for a total of 42,000 images. The entire alphabet (A-Z) and numbers (1–9) are included in this dataset. We took the pictures with a regular webcam, recording every frame of the video as a different image file. We captured the movements against a uniform black backdrop to maximize gesture clarity and reduce visual noise. Figure 1 shows the entire set of gestures arranged according to class.

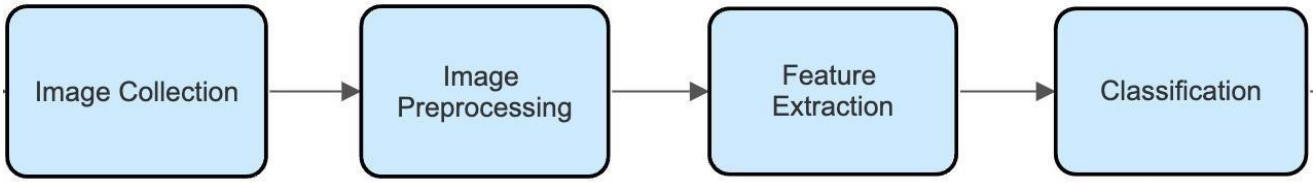
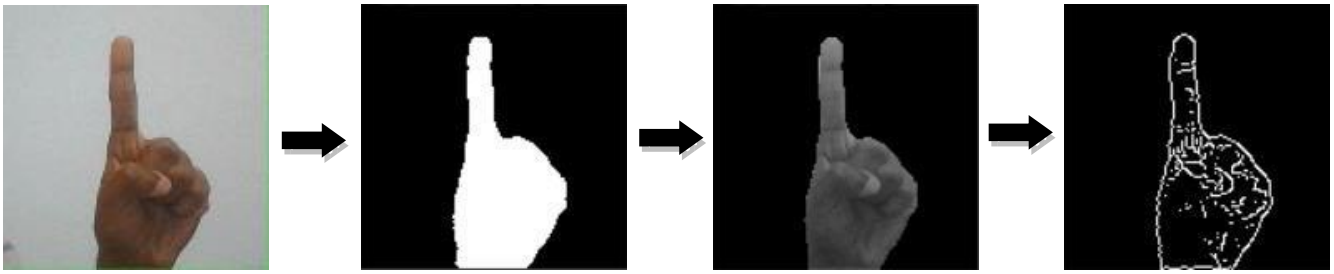


Figure 2. Overview of the project

2. Image Pre-processing

We perform a series of processing steps on the images during the preprocessing phase to prepare them for feature extraction. Initially, we employ skin masking to segment the image and isolate skin regions. This involves converting the image to the HSV color space and identifying pixels within the (H, S, V) range of (0, 40, 30) to (43, 255, 254) as skin. The resulting skin mask enables the clear delineation of skin areas in the image. Subsequently, we apply the Canny Edge Detection method, which identifies areas within the image that exhibit significant intensity contrasts, thereby detecting the image's edges. This edge detection process is essential for discerning the various outlines and shapes within the image.



a) Raw Image b) Skin Mask(Grayscale) c) Skin (Grayscale) d) Canny Edge Detection

Figure 3. Image Pre-processing

3. Feature Extraction

The feature extraction phase is critical and involves three integral steps: feature detection, clustering, and the creation of a codebook for the Bag of Words (BoW) model. Initially, we considered various algorithms and decided to employ the Scale-Invariant Feature Transform (SIFT) algorithm due to its ability to detect image features that remain consistent despite changes in scale and rotation. The SIFT algorithm is robust against variations in viewpoint and occlusion, making it highly reliable for our purposes. Figure 4 illustrates the SIFT features extracted from an image.

The subsequent step involves clustering similar SIFT features to construct a visual vocabulary. Manual identification of similar feature descriptors is impractical, hence we utilize the K-Means clustering algorithm. This unsupervised algorithm is renowned for its ability to partition 'n' features into 'k' clusters and assign new features to these clusters based on the mean (centroid) of each cluster. Given the extensive dataset of SIFT features from 42,000 images, employing the traditional K-Means clustering would be computationally demanding. Therefore, we opt for the mini-batch K-Means approach, which offers the benefits of reduced memory consumption and faster processing time. After training the SIFT features with mini-batch K-Means, the algorithm clusters them into 'bags', with the number of clusters (visual words) equating to 'k'. In order to support the classification of 35 different classes, we selected a k value of 280. We can predict visual words for every image using this model.

Compute the histograms for these anticipated visual words as the last stage of this phase. The frequency of each visual word connected to an image is tallied against the total number of visual words to create histograms. Figure 5 shows the entire procedure for this stage.

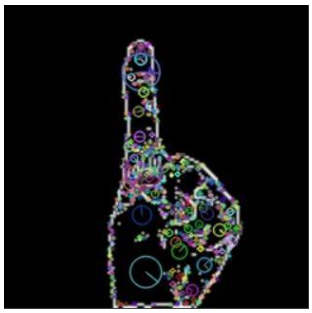


Figure 4. SIFT Key points and Feature Descriptors for an image

4. Classification

Once the histograms for the entire dataset are generated, we advance to the classification phase. The dataset must first be partitioned into training and testing subsets. We allocate 80% of the data for training and 20% for testing, ensuring that each class is represented by 960 images for training and 240 for testing. After organizing the dataset, we introduce the training data to a machine learning model. In this project, we utilize a Support Vector Machine with a linear kernel for training and making predictions. Other potential models for prediction include Logistic Regression classifiers, Convolutional Neural Networks, and K-Nearest Neighbours. Furthermore, we have developed a real-time recognition system that enables users to predict gestures through a live video feed.

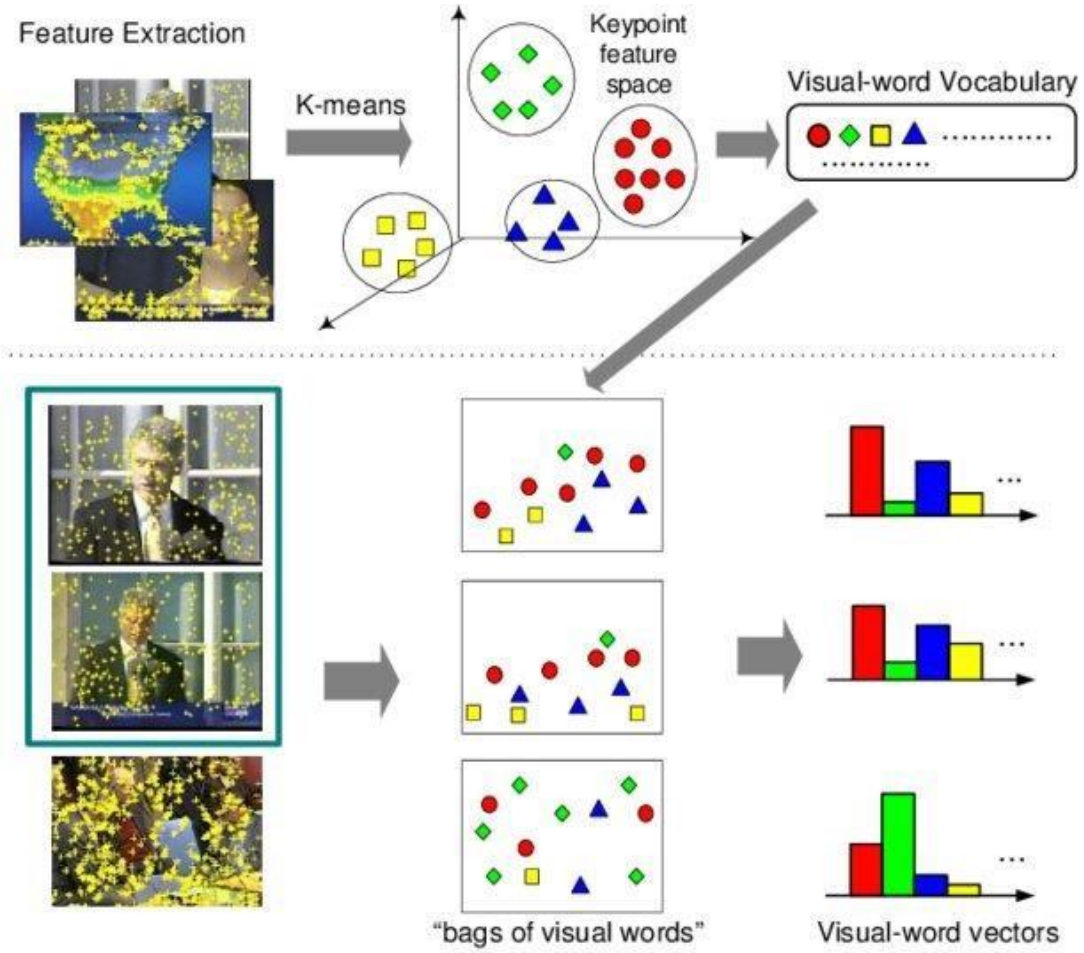


Figure 5. Bag of visual words representation. **Top:** K-means Model Training with Comprehensive SIFT Feature Set **Bottom:** Visual Word Prediction and Histogram Representation for Individual Images. [6]]

IV. EXPERIMENTAL SETUP

This project's experimental setup was created to take pictures of Indian Sign Language (ISL) gestures, process them, extract pertinent features, and then classify the images using machine learning methods. The tools utilized, the procedure for gathering data, and the overall design of the experiment are described in detail in the ensuing subsections.

A. Instruments and Detectors

A standard webcam was the main tool used to collect the data; it recorded the hand gestures against a black background consistently to ensure contrast and reduce noise. The experiments were carried out on a computer system with enough processing power to handle the computationally demanding tasks of training machine learning models and processing images.

B. Data Collection Procedure

The process of gathering data comprised taking pictures of ISL motions that matched the letters (A–Z) and numbers (1–9). There were 1200 photos in each class, for a total of 42,000 images in the dataset. The steps taken to gather the data were as follows:

1. **Image Capture:** In a controlled setting, gestures were captured using the imageCapture.py script. Within the designated Region of Interest (ROI) on the webcam's live feed, participants were instructed to make gestures. The system recorded the gesture images and stored them in the appropriate class directory when a designated key was pressed.

2. **Image Segmentation:** The imagePreprocessingUtils.py script was used to process the raw images. It used skin masking techniques to separate the hand gestures from the background.

3. **Data Organization:** To make it easier to access the images during the feature extraction and model training stages, the images were arranged into directories according to their class labels.

C. Experimental Procedure

The following steps were essential to the development of the ISL recognition system and were part of the experimental procedure:

1. **Image preprocessing:** To bring attention to the hand gestures, a skin mask was applied after each image was converted to grayscale. The gestures were then highlighted using clever edge detection.

2. **Feature Extraction:** From the preprocessed images, important features were identified and described using the SIFT algorithm. The mini-batch K-Means algorithm was then used to cluster these features in order to produce a visual vocabulary.

3. **Model Training:** Using the histogram data obtained from the visual vocabulary, a Support Vector Machine (SVM) with a linear kernel was trained. To guarantee a balanced representation of each class, the dataset was divided into training and testing portions in an 80-20 ratio.

4. **Real-Time Recognition Development:** To test the model's performance in real-time using live video input from the webcam, a separate script called recogniseGesture.py was created.

5. **Testing and Validation:** Using the test dataset, the trained model was verified, and accuracy scores and confusion matrices were used to examine the performance metrics.

V. RESULTS AND DISCUSSION

Utilizing the powerful SIFT feature descriptors in conjunction with the Bag of Words model, the system achieved an impressive accuracy rate of 97%. The model's real-time recognition capabilities are illustrated in Figure 6. The recall, F1 score, and precision of the model were remarkably high, each measuring at 98.4%. However, it is important to note that there may be a slight bias in the model's predictions due to a lack of diversity in the dataset, particularly in aspects such as skin tones and lighting conditions. To enhance the robustness and applicability of this approach for diverse real-world applications, it is recommended to expand the dataset to include a wider variety of images.

The high accuracy rate shows how well the Bag of Words model and SIFT feature descriptors capture the subtleties of ISL gestures. A more diverse dataset is necessary, though, as indicated by the slight bias in the model's predictions that was noticed. A wider range of images could be used to address the limitations of the current dataset with regard to variations in skin tones and lighting conditions. This would likely increase the model's robustness and generalizability.

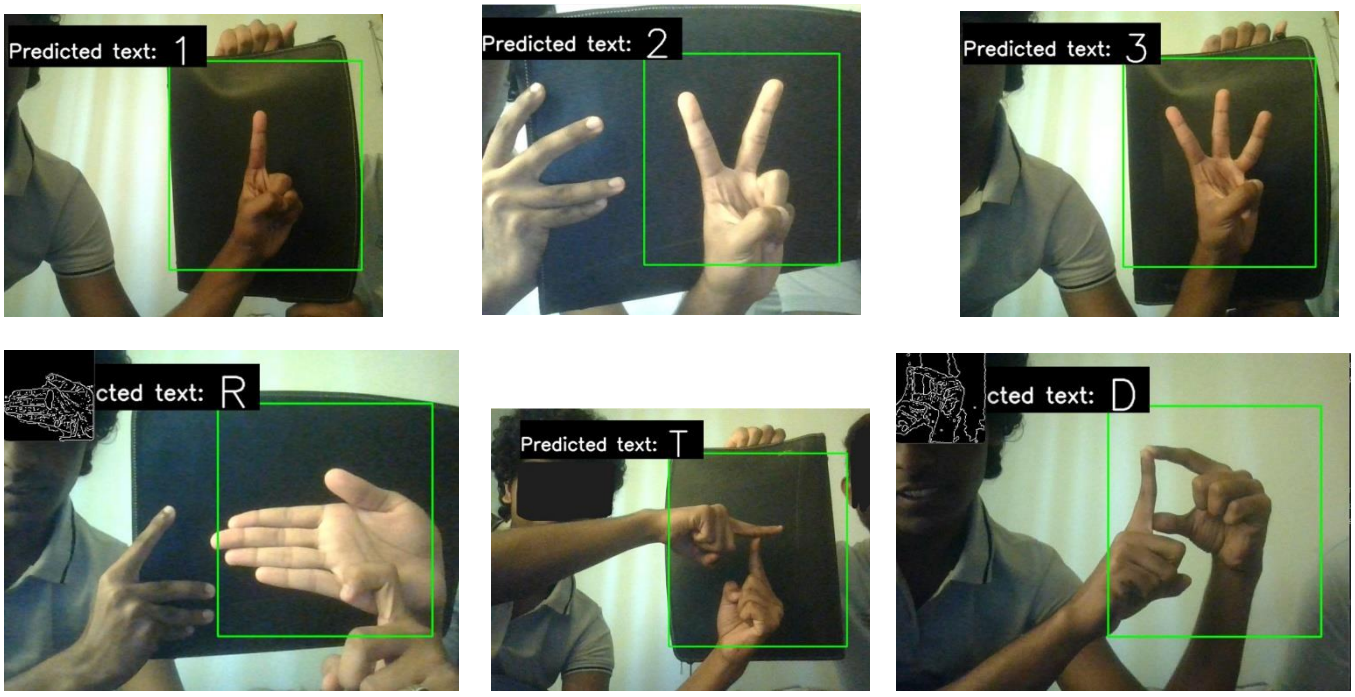


Figure 6. Real time recognition of gestures results

VI. CONCLUSIONS

The research project set out to improve the comparatively undeveloped field of Indian Sign Language (ISL) recognition by applying machine learning techniques to classify ISL gestures. The high accuracy rates in gesture classification demonstrate the project's notable success, which was attained through the creation of a substantial dataset and the application of image processing and machine learning algorithms.

The accurate prediction of ISL gestures was made possible by the effective combination of the Bag of Words model for classification and the SIFT algorithm for feature extraction. The research's practical potential was further demonstrated by the creation of a real-time recognition system, which raises the possibility that such technology could be incorporated into everyday applications to improve communication for the deaf and hard-of-hearing community.

Still, the study made clear how crucial diverse datasets are. A greater variety of lighting conditions, backgrounds, and skin tones should be included in the dataset to ensure more comprehensive data collection, as evidenced by the slight bias in model predictions resulting from the lack of variation in the dataset. This would enhance both the model's resilience and its adaptability to various user groups and environments.

To sum this up,the project marks a major advancement in machine learning-based ISL gesture recognition. It establishes the framework for further investigation that may result in the creation of more accessible and inclusive communication resources for the ISL community. The encouraging outcomes encourage further research and development of the employed methodologies in order to ultimately develop a flexible and dependable ISL recognition system that can be implemented across a range of technological platforms.

VII. FUTUREWORK

1. Deep Learning Enhancements

Deep learning models are being developed to increase ISL gesture recognition accuracy. Convolutional neural networks (CNNs) are one method for addressing this complexity, as they are capable of recognizing non-manual signals and the facial expressions that go along with hand gestures.

2. Temporal Sequence Analysis

Developing algorithms that can comprehend sign language's temporal and sequential structure. Accurate interpretation requires not only the recognition of individual signs but also an understanding of how those signs flow together in natural signing.

3. 3D Motion Capture

Incorporating 3D motion capture technology to offer a deeper comprehension of the subtleties and depth of sign language movements. To fully capture the range of motion in ISL, this may entail the use of stereo cameras or depth sensors.

4. Dataset Expansion and Diversification

Constructing more extensive and varied datasets that represent the differences in environments, signer traits, and signing styles. This is essential for creating reliable models that function well in a variety of environments and for a range of users.

5. Real-Time Translation Systems

Developing ISL translation systems in real-time that are compatible with commonplace devices like AR glasses and smartphones. This would remove communication barriers by enabling smooth communication between ISL users and non-users of sign language.

ACKNOWLEDGEMENT

We express our profound gratitude to our esteemed guide, Professor Murali Krishnan S. N., whose expertise and insights have been invaluable to the progression and success of this project. His unwavering support and constructive criticism have been pivotal in steering this research in the right direction.

We are also immensely thankful to the Department of Computer Science and Engineering at Manipal Institute of Technology for providing the necessary facilities and an enriching environment that fostered our intellectual growth and facilitated our research endeavors.

Our appreciation extends to our peers and the technical staff whose assistance and contributions have been instrumental throughout this journey. Their willingness to give their time so generously has been very much appreciated.

REFERENCES

[1] S. C. Agrawal, A. S. Jalal, and C. Bhatnagar. Recognition of indian sign language using feature fusion. In 2012 4th International Conference on Intelligent Human Computer Interaction (IHCI), pages 1–5, 2012.

[2] Herbert Bay, TinneTuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In AleˇsLeonardis, Horst Bischof, and Axel Pinz, editors, Computer Vision – ECCV 2006, pages 404–417, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg

[3] J. Canny. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI8(6):679–698, 1986

[4] K. Dixit and A. S. Jalal. Automatic indian sign language recognition system. In 2013 3rd IEEE International Advance Computing Conference (IACC), pages 883–887, 2013.

[5] B. Gupta, P. Shukla, and A. Mittal. K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion. In 2016 International Conference on Computer Communication and Informatics (ICCCI), pages 1–5, 2016.

[6] Yu-Gang Jiang, Jun Yang, Chong-Wah Ngo, and Alexander Hauptmann. Representations of keypoint-based semantic concept detection: A comprehensive study. Multimedia, IEEE Transactions on, 12:42 – 53, 02 2010.
[7] David G. Lowe. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision, 60(2):91–110, November 2004
[8] Dutta, Kusumika& K, Satheesh& S, Anil & Sunny, Breeze. (2015). Double handed Indian Sign Language to speech and text. 374-377. 10.1109/ICIIP.2015.7414799.
[9] M. Grif, R. Elakkiya, Alexey L. Prikhodko, M. Bakaev, Rajalakshmi E. Recognition of Russian and Indian Sign Languages Based on Machine Learning. Procedia Computer Science, 192: 2856-2865, 2021
[10] Akansha Tyagi, Sandhya Bansal. Feature Extraction Technique for Vision-Based Indian Sign Language Recognition System: A Review. Advances in Intelligent Systems and Computing, 1131: 33-42, 2021
[11] Poonam Pawar, Nikita Mandage, Shreya Sasane, Sakshi Ransing, Prof. Bhosle Swati. Sign Language Recognition System Using Machine Learning. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 10(5): 3174-3180, 2022

MEMBER CONTRIBUTION

1. Mihir Patel
   - Responsible for the creation of the image gathering procedure, guaranteeing a varied and extensive dataset of ISL gestures.
   - Supervised the application of the pipeline for preprocessing images, which included edge detection and skin masking methods.
   - Took responsibility for the methodology and results documentation for the research paper.

2. Peddireddy Siddharth
   - Led the feature extraction process, concentrating on the SIFT algorithm's application and optimization.
   - Carried out thorough clustering algorithm testing and validation, optimizing the mini-batch K-means procedure.
   - Controlled the creation and upkeep of the project's code repository in order to facilitate cooperation.

3. Kamatham Pranav Kumar
   - Oversaw the entire classification process, including the choice, optimization, and performance assessment of the SVM model.
   - Created the real-time recognition system by incorporating the user-friendly interface with the trained model.
   - Led the conversation about the conclusions and their implications for further research while contributing to the results analysis.