International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 10 | Oct - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Indian Sign Language (ISL) Recognition: An Economical Vision-Based Approach for Hearing & Speech Impaired People

Miss. Aparna Patil[1], Dr. Kishor Pandyaji[2]

[1]Student, [2]Professor

Département Of Electronics Engineering

Dr. V.P.S.S.Ms Padmabhooshan Vasantraodada Patil Institute of Technology, Budhgaon, Sangli, Maharashtra 416304 India.

Abstract - *The communication gap between the hearing and speech impaired community and the general population remains a significant social challenge. This paper presents an economical and non-intrusive vision-based system for Indian Sign Language (ISL) recognition, designed to translate static hand gestures into text and speech. The proposed methodology eliminates the need for expensive hardware like data gloves by leveraging a simple webcam for image acquisition. The core of the system involves a robust image processing pipeline: skin color segmentation in the CIEL*a*b* color space, Otsu's thresholding for binarization, and feature extraction using centroid calculation and the Convex Hull algorithm. These extracted features are classified using a Learning Vector Quantization (LVQ) neural network, chosen for its efficacy in handling complex image datasets with poorly defined cluster boundaries. The system was trained on a dataset of 24 ISL alphabets and several word gestures, achieving accurate real-time recognition. The output is presented both as text and synthesized speech, providing a bidirectional communication aid. This work demonstrates a practical, software-centric solution that enhances accessibility and independence for the deaf and dumb community.*

Key Words: Indian Sign Language (ISL), Hand Gesture Recognition, Convex Hull, Learning Vector Quantization (LVQ), Image Processing, Human-Computer Interaction (HCI), Assistive Technology.

1. Introduction

Hand gesture recognition stands as a powerful and intuitive form of non-verbal communication, providing a separate and complementary modality to speech for the expression of ideas and commands. Its flexibility and naturalism make it an ideal candidate for creating more seamless and effective interactions between humans and computing systems. For individuals with hearing and speech impairments, this technology transcends mere convenience; it is a vital bridge to the wider world, enabling communication where spoken language is not an option. The development of robust gesture recognition systems thus represents a significant stride in assistive technology, with the potential to foster greater social inclusion and independence for a often-marginalized community.

The challenge is particularly acute in a country like India, where an estimated 4.482 million people are hearing impaired. This community frequently faces dispossession from various social activities and is often underestimated by society at large due to persistent communication barriers. A functioning sign language recognition system could directly address this issue by providing an opportunity for the deaf to communicate with non-signing people without the constant need for a human interpreter. By translating signs into speech and text, such a system can empower individuals, making them more self-reliant and integrated into daily social and professional life.

The field of automated sign language interpretation has seen extensive research, primarily evolving along two trajectories: glove-based techniques and vision-based techniques. Glove-based methods involve users wearing sensor-embedded gloves that provide precise data on hand position and finger flexion. While accurate, these systems are often intrusive, expensive, and impractical for everyday use. In contrast, vision-based techniques leverage cameras to capture and interpret gestures, offering a non-intrusive, more natural, and economically viable alternative. This approach aligns with the goal of creating an accessible technology that can be widely adopted.

This paper presents a vision-based system specifically designed for the recognition of Indian Sign Language (ISL). The proposed methodology centers on a robust image processing pipeline that begins with acquiring images via a standard webcam. The core of the approach involves sophisticated feature extraction using the Convex Hull algorithm, which provides a geometrically invariant representation of the hand shape. These features are then classified using a Learning Vector Quantization (LVQ) neural network, a method chosen for its robustness in handling complex image datasets with poorly defined boundaries between different gesture clusters.

The primary objective of this dissertation is to develop a software system that can accurately detect, track, and recognize hand gestures representing ISL alphabets, numbers, and special signs. The system processes this input and computes a result that is displayed both as text and converted into a voice output. This dual-output mechanism ensures that communication is facilitated in both directions: from the signer to the non-signing person via

text and speech, and vice-versa through the non-signer's spoken language.

In summary, this work details the development and implementation of an economical, non-intrusive solution for ISL recognition. By leveraging advanced image processing and neural network techniques, the proposed system aims to reduce the communication gap for the hearing and speech impaired, offering a practical tool to enhance their interaction with the world around them. The following sections will elaborate on the literature survey, detailed methodology, experimental results, and conclusions drawn from this research.

2. Literature Survey

A substantial and diverse body of research has been established in the domain of sign language recognition, charting a clear evolution in methodological approaches. Early systems heavily relied on data gloves equipped with embedded sensors to precisely track hand position, orientation, and finger flexion [4, 5]. While these glove-based methods were effective and provided high-accuracy data, their intrusive nature, high cost, and lack of practicality for everyday use rendered them prohibitive for widespread personal adoption. Consequently, the field has witnessed a significant pivot towards vision-based techniques, which offer a more natural and user-friendly experience by using cameras as the primary input device. A principal challenge within this vision-based paradigm is achieving robust hand segmentation under varying environmental conditions, a problem that researchers have addressed by exploring various color spaces—such as RGB, HSV, and YCbCr—for effective skin color modeling to reliably isolate the hand region from complex backgrounds [6, 7]. For the subsequent binarization step, Otsu's method [8] has emerged as a widely adopted and powerful technique for automatic, non-parametric image thresholding, and it is this method that we employ in our system to create a clear binary representation of the hand.

The process of feature extraction is equally critical, as the chosen features directly determine the discriminative power of the recognition system. A variety of techniques have been investigated for this purpose, including statistical methods like Principal Component Analysis (PCA) for dimensionality reduction [9], moment invariants for capturing shape descriptors that are resilient to transformations [10], and the analysis of fundamental geometric properties. In our work, we specifically utilize the Convex Hull algorithm [11] to derive a simplified, geometrically invariant representation of the hand shape. This algorithm effectively identifies the smallest convex polygon that encloses all points of the hand contour, and the features derived from its vertices and defects provide a robust description that is largely invariant to rotation and scale, thereby enhancing the system's generalization capability.

Finally, the stage of classification has been tackled using a spectrum of machine learning models, each with its own strengths. For recognizing dynamic gestures that unfold over time, Hidden Markov Models (HMMs) have been a popular choice due to their ability to model temporal sequences [12]. For static postures, Artificial Neural Networks (ANNs) have been extensively applied for their powerful pattern recognition capabilities [13]. Within this landscape, we selected the Learning Vector Quantization (LVQ) network, a specific type of supervised neural network, for its particular robustness in clustering complex datasets where class boundaries may not be well-defined, as well as for its efficient and stable learning mechanism [14]. A comparative analysis summarizing the characteristics of these various recognition methods is provided in Table 1, highlighting the distinct advantages of our chosen approach.

Table -1: ***Comparison of Gesture Recognition Methods***

Primary Method of Recognition	Number of Gestures Recognized	Background	Additional Markers Required	Training Images
Artificial Neural Network (ANN) [13]	34	General	No	270
Hidden Markov Models (HMM) [12]	97	General	Multicolored Gloves	400
Linear Approximation Models	26	Blue Screen	No	7441
Proposed System (LVQ + Convex Hull)	**24 Alphabets + Words**	**General**	**No**	**72**

3. Proposed Methodology

The system architecture for the Indian Sign Language recognition follows a structured pipeline from image acquisition to final output generation. The complete process can be visualized through the system block diagram and flow chart that outline the sequential stages of operation.

3.1 System Architecture Overview

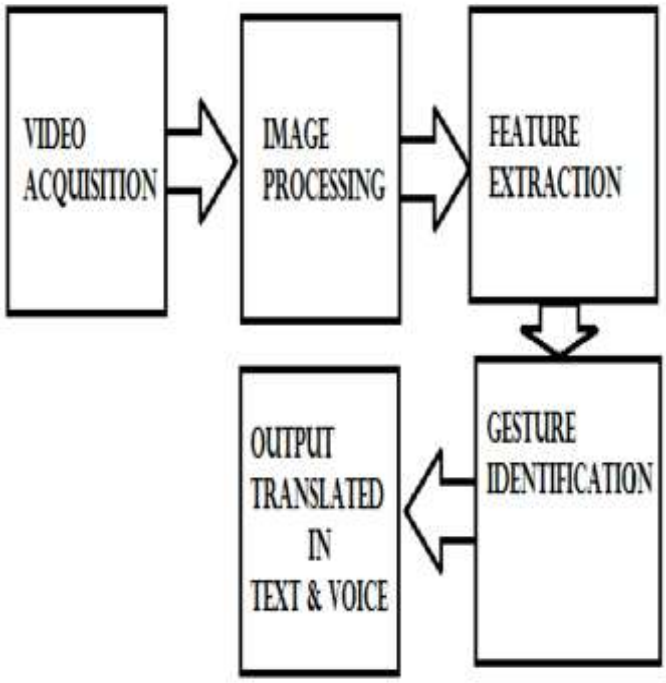

Fig 1:- General Block Diagram of Proposed System

The proposed system implements a sophisticated modular architecture specifically engineered for robust Indian Sign Language recognition, where each distinct component is meticulously designed to execute specialized functions within the comprehensive recognition pipeline. The hardware foundation incorporates a standard VGA webcam carefully selected for its capability to capture high-quality images at sufficient resolution to facilitate accurate and detailed gesture analysis, while simultaneously maintaining the cost-effectiveness crucial for widespread accessibility. Complementing this hardware foundation, the software component is comprehensively developed within the MATLAB environment, leveraging its powerful image processing and neural network toolboxes to efficiently manage all computational stages—from the initial image acquisition through the complex sequence of preprocessing, feature extraction, and classification operations, ultimately culminating in the generation of multi-modal outputs. This deliberate separation of hardware and software concerns creates a flexible and scalable system architecture that not only permits straightforward maintenance and troubleshooting but also provides a solid foundation for future enhancements and feature integrations, all while ensuring consistently reliable performance across diverse operating conditions and user environments.

The inter-module communication framework follows a rigorously sequential and deterministic pathway, where each processing stage systematically consumes its designated input, executes its specialized computational tasks, and precisely delivers the processed results to the subsequent stage in the pipeline. This carefully engineered linear data flow architecture is instrumental in minimizing processing latency and computational overhead, thereby making the entire system exceptionally well-suited for real-time interactive applications where immediate feedback is paramount for effective user engagement. The comprehensive block diagram visually demonstrates the precise propagation of data through the entire system architecture, clearly illustrating the journey from initial raw image capture through the complex transformation processes, and finally culminating in the generation of multi-modal outputs that seamlessly integrate both visual text displays and synthesized speech, thereby creating a complete and accessible communication solution for end-users.

.

3.2 Detailed System Workflow

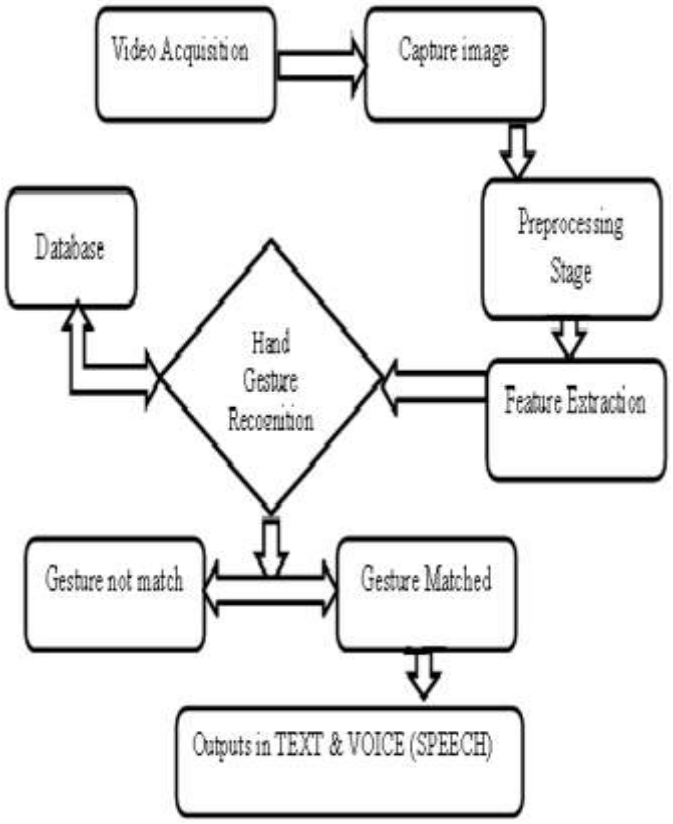

Fig 2:- System Flow Chart

The operational workflow initiates with the crucial image acquisition phase, where the system systematically captures continuous frame sequences from the webcam input, ensuring comprehensive coverage of the dynamic gesture performance. Each individual captured frame immediately undergoes a sophisticated color space transformation process, converting from the native RGB color model to the advanced CIEL*a*b* color space, which is specifically chosen for its superior perceptual uniformity characteristics and significantly enhanced capabilities in maintaining color consistency across varying illumination scenarios. This meticulous conversion process establishes a standardized color representation framework that effectively compensates for the inconsistencies typically introduced by different imaging devices and environmental lighting conditions, thereby substantially improving the system's adaptability and reliability when deployed across diverse real-world scenarios, from controlled laboratory settings to challenging practical

environments with fluctuating light sources and background variations.

Following the essential color space conversion, the system strategically implements Otsu's automated thresholding methodology to perform optimal clustering-based image binarization, leveraging its sophisticated computational approach that maximizes between-class variance to achieve highly effective separation of the hand region from complex and potentially distracting background elements. The resulting binary image subsequently undergoes a series of advanced morphological processing operations, including carefully calibrated opening and closing procedures, which work in concert to systematically eliminate residual noise artifacts, smooth irregular boundaries, and fill inconsequential gaps within the hand mask, thereby producing a refined and clean binary representation ideally prepared for the subsequent feature extraction phases. This comprehensive preprocessing pipeline represents a critical foundational stage that ensures only the most relevant and geometrically pure hand shape information progresses forward in the processing chain, effectively normalizing input variability while dramatically enhancing the signal-to-noise ratio in the raw image data before it undergoes further detailed analytical processing.

3.3 Feature Extraction Process

The feature extraction phase transforms the pre-processed binary image into a compact numerical representation suitable for classification. This stage involves three fundamental operations: centroid calculation, contour detection, and convex hull computation. The centroid, representing the geometric center of the hand, is calculated using image moments and serves as a stable reference point for spatial normalization and subsequent processing steps.

Contour detection identifies the external boundary of the hand using advanced border following algorithms. The system intelligently selects the largest contour, assuming it corresponds to the hand region, and applies the convex hull algorithm to determine the smallest convex polygon that contains all contour points. This convex hull computation generates geometrically invariant features that maintain robustness against rotation and scaling variations, while the topological defects between the original contour and the convex hull provide distinctive features corresponding to specific finger arrangements and hand orientations.

The comprehensive feature vector comprises normalized coordinates of convex hull vertices, defect depths and angles, spatial relationships with the centroid, and various moment invariants. This multi-dimensional feature set effectively captures both global shape characteristics and local distinctive attributes of the hand gesture, providing the classification algorithm with rich discriminative information while maintaining computational efficiency necessary for real-time operation.

3.4 Classification and Output Generation

The extracted feature vector is classified using a Learning Vector Quantization neural network, which employs a sophisticated competitive learning strategy for pattern recognition. The network architecture consists of two distinct layers: a competitive layer that learns to cluster similar input patterns, and a linear layer that intelligently maps these clusters to target classes representing different ISL gestures. During the training phase, the LVQ network adjusts its weight vectors through an iterative process that reinforces correct classifications while penalizing misclassifications.

During operational deployment, the LVQ network computes the Euclidean distance between the input feature vector and all prototype vectors in the competitive layer. The winning neuron, corresponding to the smallest calculated distance, activates its associated output class. The classification decision undergoes validation against pre-defined confidence thresholds, and when sufficient confidence is achieved, the recognized gesture is immediately displayed as text on the screen in an intuitive user interface.

The final stage involves sophisticated speech synthesis, where the system converts the recognized text into natural audio output using advanced text-to-speech capabilities. This carefully designed dual-output approach ensures comprehensive accessibility for both hearing and visually impaired users, making the system remarkably versatile for different usage scenarios and interaction patterns. The entire integrated process, from initial image capture to final speech output, completes within a tightly constrained timeframe, enabling seamless near real-time interaction between signers and non-signers.

4. Experimental Results and Discussion

The system was implemented using MATLAB R2009b. A database of 72 images was created, comprising 3 samples each for 24 ISL alphabets (letters 'M' and 'N' were excluded due to high similarity). Additional gestures for words like "FRIEND," "NAMASKAR," and "SORRY" were also included.

The LVQ network was trained on this dataset. The confusion matrix (Fig. 3) generated after training shows a high rate of correct classification along the diagonal, with minimal misclassifications, indicating the model's robustness.

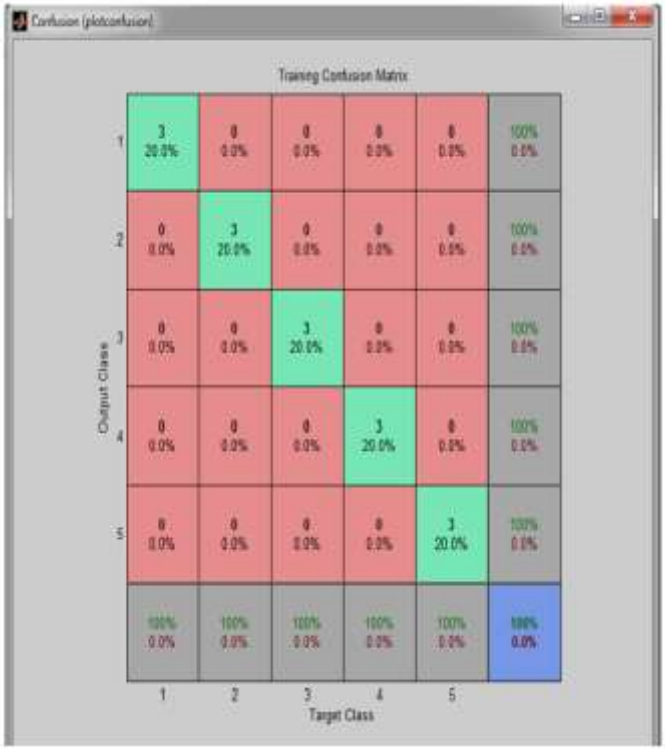

Fig 3:- Confusion Matrix showing classification performance of the LVQ network.

The system was tested in real-time using a GUI. Fig. 4 shows the system correctly recognizing the sequence of gestures for "WE ARE GOOD FRIEND" and displaying the result in the GUI while also speaking it out. The system successfully recognized isolated alphabets (e.g., 'A', 'B') and word gestures with high accuracy.

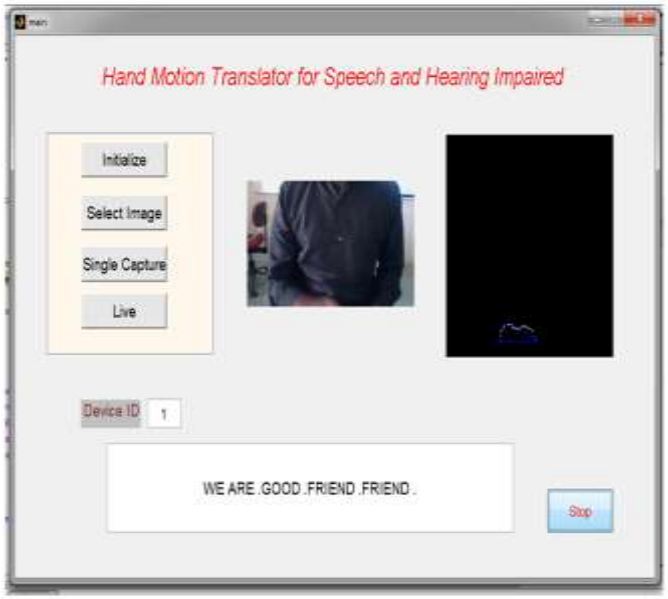

Fig 4:- GUI showing the recognized sentence "WE ARE GOOD FRIEND".

The system's performance is dependent on consistent lighting and background conditions. Furthermore, the current implementation focuses on static gestures. Future work will involve extending the system to recognize dynamic gestures and sentences.

4. Conclusion

This paper presented a functional and economical vision-based system for Indian Sign Language recognition. By integrating CIEL*a*b* color space segmentation, Otsu's thresholding, Convex Hull-based feature extraction, and LVQ neural network classification, the system effectively translates static hand gestures into text and speech. This non-intrusive, software-based solution provides a practical communication aid for the hearing and speech impaired community, promoting greater social integration and independence. The system demonstrates that high-cost hardware is not a prerequisite for building effective assistive technology, paving the way for future work on more complex, dynamic gesture recognition.

References

[1] Patil and G. V. Lohar, "Hand Motion Translator for Speech and Hearing Impaired," in *International Conference on Convergence of Technology*, 2014.

[2] T. S. Hunang and V. I. Pavloic, "Hand Gesture Modeling, Analysis, and Synthesis," in *Proc. of International Workshop on Automatic Face and Gesture Recognition*, Zurich, 1995.

[3] F. Ullah, "American Sign Language recognition system for hearing impaired people using Cartesian Genetic Programming," in *5th International Conference on Automation, Robotics and Applications (ICARA)*, 2011.

[4] Y. F. Admasu and K. Raimond, "Ethiopian sign language recognition using Artificial Neural Network," in *10th International Conference on Intelligent Systems Design and Applications (ISDA)*, 2010.

[5] G. Fang, W. Gao, and J. Ma, "Signer-independent sign language recognition based on SOFM/HMM," in *IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, 2001.

[6] Sachin S. K. et al., "Novel Segmentation Algorithm for Hand Gesture Recognition," in *IEEE Conference on Information and Communication Technologies (ICT)*, 2013.

[7] M. M. Hasan and P. K. Mishra, "HSV brightness factor matching for gesture recognition system," *International Journal of Image Processing (IJIP)*, vol. 4, no. 5, 2010.

[8] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62-66, Jan. 1979.

[9] Divya Deora and Nikesh Bajaj, "Indian sign language recognition," in *1st International Conference on Emerging Technology Trends in Electronics, Communication and Networking*, 2012.

[10] Zhongliang, Q. and Wenjun, W., "Automatic Ship Classification by Superstructure Moment Invariants and Two-stage Classifier," in *ICCS/ISITA '92. Communications on the Move*, 1992.

[11] R. Minhas and J. Wu, "Invariant Feature Set in Convex Hull for Fast Image Registration," in *IEEE International Conference*, 2007.

[12] P. Vamplew, "Recognition of sign language gestures using neural networks," in *European Conference on Disabilities, Virtual Reality and Associated Technologies*, Maidenhead, U.K., 1996.

[13] Adithya V., Vinod P. R., and Usha Gopalakrishnan, "Artificial Neural Network Based Method for Indian Sign Language Recognition," in *IEEE Conference on Information and Communication Technologies (ICT)*, 2013.

[14] E. S. Amin, "Application of artificial neural networks to evaluate weld defects of nuclear components," *Journal of Nuclear and Radiation Physics*, vol. 3, no. 2, pp. 83-92, 2008.