Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

An Efficient Sparse Gabor Descriptor Framework for Sign Language Recognition

Kiran C. Kulkarni[1], Manoj A. Wakchaure[2]
[1]*Research Scholar Dept. of Computer Engineering METs Institute of Engineering, Nashik Savitribai Phule Pune University, Pune*
Kiranc.kulkarni@gmail.com
[2]*Research Guide Dept. of Computer Engineering METs Institute of Engineering, Nashik Associate Professor, Amrutvahini College of Engineering Sangamner Savitribai Phule Pune University, Pune*
Manoj.wakchaure@avcoe.org

ARTICLE INFO

Received: 30 Dec 2024

Revised: 19 Feb 2025

Accepted: 27 Feb 2025

ABSTRACT

Recognising person sign language for videos through algorithms remains a significant difficulty in computer vision (CV). The challenge of recognising human hand gestures with computer vision might be solved using machine learning techniques, allowing individuals to evaluate visual data and potentially react to our movements. This study investigates the development and implementation of machine learning methods for recognising hand movements in dynamic video data. This study builds on previous advances in machine learning and proposes a novel Sparse Gabor Descriptor (SGD)-based technique. Finally, these features can be classified using a random forest for gesture recognition. Experiments show that the suggested strategy delivers competitive results on gesture recognition datasets while requiring significantly less processing complexity. The comparison with roughly three state-of-the-art methodologies yields competitive results in terms of the system's many activities, such as accuracy and classification precision. The proposed technique achieves a recall rate of 99%, an accuracy rate of 94%, an F1-score of 97%, and a precision rate of 94%, which is higher than the KNN, NB, and LR methods.

Keywords: *Hand Gesture, Sign Language, Gabor Filter, Random Forest.*

I. INTRODUCTION

Gestures are a significant aspect of our communication. A crucial element in developing an automated hand gesture recognition system is its capability to capture hand movements via a single video camera. Hand gesture recognition enables computers to recognise and comprehend human gestures using mathematical algorithms, hence executing commands based on these movements to facilitate human-computer interaction [1][2].

The World Health Organisation (WHO) reports that over 5% of the global population, around 430 million individuals, require rehabilitation for hearing loss. Due to population increase and ageing, it is anticipated that by 2050, around 2.5 billion individuals would experience varied degrees of hearing loss, with at least 700 million requiring rehabilitation. These alarming trends underscore the pressing necessity for improved assistive technologies. Individuals with hearing difficulties depend on manual gestures for communication. Unfortunately, the overwhelming majority of individuals fail to comprehend the importance of these gestures. Effective communication is crucial for all persons, but it poses significant challenges for the deaf community, as sign language serves as their primary mode of communication. The absence of extensive understanding on the significance of these gestures results in a communication divide between the hearing-impaired and hearing populations.

In recent years, considerable attention has been dedicated to gesture recognition. The burgeoning interest in human-computer interaction (HCI) has led to a swift expansion of research in gesture recognition. The increasing prevalence of robots in daily life has positively influenced the advancement of natural human-robot interaction (HRI) in robotics. The significance of automatic hand gesture recognition has escalated for the following reasons: 1) The expansion of the deaf and hard-of-hearing demographics, and

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

2) the increased utilisation of vision-based and touchless applications and devices, including video games, smart TV controls, and virtual reality applications.

Identifying gestures from video streams presents a significant challenge due to the great variability in motion characteristics among individuals. Nonetheless, these algorithms encounter obstacles due to contextual variables such as fluctuations in lighting, intricate backgrounds, heterogeneous hand shapes, and resemblances among various gesture categories. Attaining precise gesture recognition under these conditions is a challenging endeavour, requiring resilient solutions to guarantee dependable performance.

Throughout the years, computer scientists have employed various computational algorithms and techniques to address our challenges and enhance our lives [4]. The utilisation of hand gestures in many software applications has enhanced human-computer interaction. We offer a suitably designed deep architecture for the automatic recognition of hand gestures. This study's primary contributions are as follows:

1) A technique to standardise the spatial dimensions of gesture videos utilising spatiotemporal characteristics for hand movements. The model receives a sequence of RGB frames obtained from a standard camera. It does not necessitate additional input channels, coloured gloves, or a complicated configuration.

2) Formulating several fusion methodologies to universalise the localised characteristics acquired by the machine learning model and evaluating their efficacy. The subsequent sections of this work are structured as follows.

Section II examines pertinent literature. Section III delineates the dataset's description. Section IV presents the proposed approach. Section V presents the experimental results and discussions. The research findings are outlined in Section VI.

Mus mauris vitae ultricies leo integer malesuada nunc vel risus. Nec nam aliquam sem et tortor consequat id. Risus nec feugiat in fermentum posuere. A pellentesque sit amet porttitor eget dolor. Nibh tortor id aliquet lectus proin nibh nisl condimentum id. Ultrices dui sapien eget mi proin sed. Amet risus nullam eget felis eget nunc lobortis mattis aliquam. Morbi blandit cursus risus at ultrices mi tempus imperdiet nulla. Vitae turpis massa sed elementum tempus. Diam ut venenatis tellus in metus vulputate eu. Consectetur a erat nam at lectus. Eros donec ac odio tempor orci dapibus. Suspendisse in est ante in nibh mauris cursus. Massa massa ultricies mi quis. Ultricies lacus sed turpis tincidunt id aliquet risus feugiat in. In iaculis nunc sed augue lacus viverra vitae congue eu. Ipsum a arcu cursus vitae congue mauris rhoncus. Ultricies mi quis hendrerit dolor magna eget est lorem. Purus semper eget duis at.

Donec massa sapien faucibus et molestie ac feugiat. Sed risus ultricies tristique nulla aliquet. Nibh tortor id aliquet lectus proin nibh nisl. Ut etiam sit amet nisl purus in. Lectus mauris ultrices eros in cursus turpis massa tincidunt. Pretium vulputate sapien nec sagittis aliquam malesuada. Auctor neque vitae tempus quam. Aenean sed adipiscing diam donec adipiscing. Magnis dis parturient montes nascetur ridiculus mus mauris. Placerat in egestas erat imperdiet sed euismod nisi porta lorem. Vel facilisis volutpat est velit egestas dui. Ultrices gravida dictum fusce ut placerat orci nulla pellentesque dignissim. Egestas tellus rutrum tellus pellentesque eu tincidunt tortor aliquam nulla. Mattis pellentesque id nibh tortor id. Ut venenatis tellus in metus vulputate.

II. LITERATURE SURVEY

There have been a variety of approaches to human hand gesture recognition that have been proposed by previous research. When it was first developed, vision-based gesture recognition could only be used to still pictures.

A skeleton-based architecture for hand gesture recognition was proposed by H. Pengcheng et al. [6]. This architecture would incorporate a spatio-temporal graph convolution network and a transformer that would make use of the Kolmogorov Arnold Network model. The goal of this architecture would be to properly represent the dependency among the hand joints. A novel deep model for hand gesture recognition was presented by Gaikwad S. and colleagues [7]. This model makes use of adaptive thresholding-based region expansion, Canny edge detection for segmentation, and a multiscale, attention-embedded residual DenseNet that is improved using the Modified Tasmanian Devil optimisation. For intelligent cars and modern transportation systems, the detection of traffic police hand signals must be done quickly and accurately. An information enhancement graph convolutional network with dual modules for spatial and temporal information was proposed by Shi P. and colleagues. Specifically, they include

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

the Synergy Attention Module (SAM) and the Keyframe Extraction Module (KEM) into the spatial-temporal graph convolutional network in order to enhance the network's capacity to extract synergistic action characteristics and significant action frames. The complexity of hand movements, which includes differences in spatial and temporal aspects, usually presents challenges for the approaches that are currently in use. Yildiz, A. et al. [9] propose a novel approach that combines a hybrid Spectral-Polynomial Convolutional Neural Network with an Inception module and Long Short-Term Memory architecture. This approach is intended to solve the restrictions that have been mentioned. L. Jianbo and colleagues [10] are working on constructing a model that is extraordinarily lightweight for skeleton-based gesture detection by making use of a module that is solely focused on performing self-attention. In order to collect the characteristics of the joints that are the most informative, the self-attention module makes use of dynamic attention weights. This is accomplished using a shallow network.

A novel framework for real-time hand gesture detection from video was presented by Uke N. and colleagues [11]. This framework makes use of optimum video and soft computing algorithms. Real-time video capture, video normalisation, Hand Frames Feature Extraction (HFFE), and classification are all things that are utilised by this system. Using RGB and depth data, F. Farid and colleagues propose the application of deep learning algorithms for the purpose of recognising automated hand gestures. Following the acquisition of RGB video and depth data with the Kinect sensor, hand tracking was performed with a single-shot detector Convolutional Neural Network. In order to recognise dynamic hand gestures that have been collected in actual scenarios, Hax D. et al. [13] propose a hybrid architecture that incorporates a recurrent neural network (RNN) that has a lengthy short-term memory layer on top of a convolutional neural network. HandSense is a unique system for multi-modal hand gesture identification that was developed by Zhang, Z., and colleagues [14]. It makes use of a mix of RGB and depth cameras to improve fine-grained action descriptions while yet retaining the ability to recognise generic actions. A video-based hand gesture identification system that makes use of Random Forest was proposed by J. Himasree and colleagues [15] in order to capture temporal changes in sign language gestures. A video-based hand gesture identification system that makes use of a depth camera and a lightweight convolutional neural network (CNN) model is presented by D. León and colleagues[16].

For the purpose of gesture recognition, Mujahid, A. et al. [17] offer a lightweight model that makes use of YOLO (You Only Look Once) v3 and DarkNet-53 convolutional neural networks. This model eliminates the requirement for further preprocessing, picture filtering, and augmentation. A dynamic hand gesture recognition approach is presented by Gao et al., which makes use of 3D hand posture estimation in conjunction with a 3DCNN + ConvLSTM architecture. This method is designed to address the obstacles that are provided by complex settings and dynamic perceptual situations. An inquiry into hand gesture recognition is presented by Qi, J. et al. [19]. This study makes use of monocular and RGB-D cameras, and it includes data collection, hand gesture detection and segmentation, feature extraction, and gesture categorisation.

At the moment, there are two distinct types of approaches that may be used to recognise human gestures. These are the traditional handmade feature engineering techniques that make use of machine learning and the deep learning techniques. Due to the fact that typical machine learning approaches are heavily dependent on human knowledge and experience, the results that are obtained from these techniques are not entirely satisfying. As a consequence, the effectiveness of a recognition system is limited by the expertise that humans possess in the relevant area. In order to be successful, deep learning requires a significant amount of highly tagged training data. There is a possibility that the performance of the model will be hindered by a limited or uneven dataset. When applied to the training data, the model has a tendency to overfit, particularly in situations when the architecture is extremely complicated or the dataset does not include an adequate amount of variety.

One major drawback of the literature is its dependence on manually created elements, which could not always adequately convey the intricacy of sign language motions. In real-world SLR, where the connection between many variables (such hand position and movement) might be critical, traditional approaches like Naive Bayes, which are based on a probabilistic model, presume feature independence. However, SVM and KNN, which categorise data according to closeness to the closest neighbours, may have issues with efficiency and scalability, especially when the dataset is big or has a high dimensionality. Additionally, these techniques could be sensitive to noise since even little changes in the way the sign is executed or the surrounding environment might have a big impact on the categorisation

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

accuracy. Furthermore, it might be difficult to acquire big, varied datasets for training Deep Learning models, which are often needed for under-represented sign languages or particular gestures.

III. DATASET

World Level American Sign Language (WLASL) is the biggest video dataset for Word-Level American Sign Language (ASL) recognition, including 2,000 commonly used ASL terms. L. Dongxu et al. [20] provide a new large-scale Word-Level American Sign Language (WLASL) video collection with over 2000 words performed by over 100 signers.

IV. METHODOLOGY

Over the course of the past several years, numerous strategies for the enhancement of holistic techniques for feature extraction have been published. One of the methods that has shown to be more successful is the utilisation of Gabor image representation. High-definition features are an expansion of the description of the integrated features. A band-pass direct channel with a motivating reaction that is described by a symphonious capacity multiplied by a Gaussian capacity is what the Gabor channel, also known as the Gabor Wavelet, corresponds to. To continue along similar lines, a bi-dimensional Gabor channel generates a complex sinusoidal plane that is controlled by a Gaussian envelope and possesses accurate recurrence and direction. It has been demonstrated that Gabor characteristics are relevant to the field of representation. However, only a few methods make use of the phase feature, and the performance of these methods is frequently inferior to that of methods that make use of the magnitude feature. Because of this, the magnitudes of the Gabor coefficients are the only pieces of information that are deemed meaningful for feature extraction. In both the spatial and recurring domains, it accomplishes the goal in the best possible way.

Through the use of our method, a two-dimensional odd-symmetric Gabor filter is constructed, which has the following structure:

$$HD(F)_{\theta_k, f_i, \sigma_x, \sigma_y}(x, y) = exp\left(-\left[\frac{x_{\theta_k}^2}{\sigma_x^2} + \frac{y_{\theta_k}^2}{\sigma_x^2}\right]\right).\cos(2\pi f_i x_{\theta_k} + \varphi) \quad (1)$$

This 2D gabor filter is used to all datasets to extract sparse features.

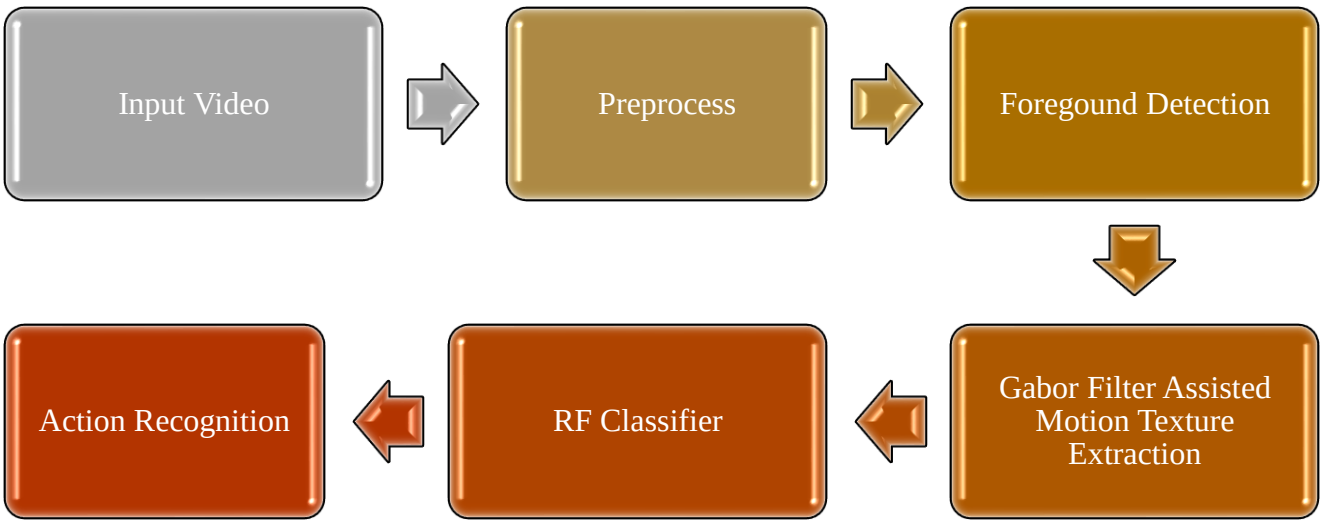

Fig 1: Sign Language Recognition using Gabor Filter

V. DISCUSSION

Gabor filters are often used for texture analysis and edge recognition because of their ability to gather spatial and frequency information at a wide range of scales and orientations. On the other hand, the standard Gabor filter may be computationally expensive and inefficient in some particular circumstances. By utilising a sparse representation of the filter's response, the Sparse Gabor Filter is able to address these problems. This representation dramatically decreases the amount of processing burden while maintaining the high-quality feature extraction that it provides.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Even under challenging circumstances, such as shifting illumination, occlusions, and background clutter, SGFs are very excellent at identifying fine-grained patterns and edges in hand shapes and motions. This is the case even when the conditions are not ideal. Through the process of capturing the multi-scale spatial and frequency components of hand gestures, the SGF enhances the system's ability to differentiate between subtle variations in gestures and boosts the algorithms that are charged with recognition. By recovering discriminative spatial and frequency information from hand movements in an effective manner, the Sparse Gabor filter is able to maintain essential features while simultaneously reducing the amount of computational complexity involved. After that, these characteristics are input into the Random Forest classifier, which employs a collection of decision trees in order to categorise the movements.

Due to the fact that the Random Forest model is able to handle the high-dimensional feature space of the Sparse Gabor filter while also being immune to perturbations such as noise, illumination variations, and background clutter, the combination provides exceptionally precise detection. The fact that this method makes use of the rich textural information that is captured by the Sparse Gabor features as well as the ensemble strength of the Random Forest makes it an excellent choice for solving challenges involving real-time gesture identification. Because this sparse technique produces more efficient processing of real-time video analysis, it is ideally suited for applications such as gesture-based control in interactive systems, translation of sign language, and interaction between humans and robots.

Random Forest is used to classify the extracted features.

Let $D = \{(x_1, y_1), \dots (x_N, y_N)\}$ denote the extracted features with $X_i = (x_{i1}, \dots, x_{ip})$

For $j = 1{:}j$

1. Take a bootstrap sample D_j of size N from D.
2. Using the bootstrap sample D_j as the training data, fit a tree using binary

Recursive Partitioning:

   a. Start with all observations in a single node.
   b. Repeat the following steps recursively for each unsplit node until the
   c. stopping criterion is met:

3. Select m predictors at random from the p available predictors.
4. Find the best binary split among all binary splits on the m predictors
5. from step i.
6. Split the node into two descendant nodes using the split from step

To make a prediction at a new point x

$\hat{f}(x) = \frac{1}{j}\sum_j^J \widehat{h_j}(x)$ for regression (2)

$\hat{f}(x) = argmax_y \sum_{j=1}^{J} I\left(\widehat{h_j}(x) = y\right)$ for classification (3)

where $\widehat{h_j}(x)$ is the prediction of the response variable at x using the j^{th}tree

VI. EXPERIMENT & RESULTS

The proposed DSN is built in Python on a laptop with an Intel Core i52.50GHz CPU and 16GB of RAM. To decrease computational costs, the picture dimension of the supplied data is reduced to 512*512 via binear interpolation.

Performance Analysis:

Additionally, a confusion matrix, accuracy, sensitivity, specificity, and positive prediction rate are utilised in this study in order to conduct a more precise analysis and investigation of the categorisation consequences of each

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

model. Accuracy refers to the rate at which the actual condition of a pose is identified; specificity refers to the rate at which different poses are differentiated from one another; sensitivity refers to the rate at which the absence of a specific pose is identified; and the positive predication rate refers to the rate at which accurate positive identifications are carried out. An overview of the method's performance evaluation is provided in the following paragraphs. An F1 score of 97%, a recall score of 98%, a precision score of 94%, and an accuracy score of 95% are all achieved thanks to the suggested model. The results are broken down into each of the three groups below according to the approach that used.

$$\text{Accuracy} = [\text{TP} + \text{TN}]/[\text{P} + \text{N}] \quad (4)$$

$$F1\ Score = [2 * \text{TP}] / [2 * \text{TP} + \text{FP} + \text{FN}] \quad (5)$$

$$\text{Recall} = [\text{TP}] / [\text{P}] \quad (6)$$

$$Pecision = [\text{TP}] / [\text{TP} + \text{FP}] \quad (7)$$

As can be seen in the figure, the strategy that was recommended performs better than others when it comes to sign recognition. This is evidenced by the fact that all of the criteria have been improved. The reason for this is most likely due to the fact that the frequency resolution of the approach that was presented is more adaptive in terms of telling the difference between activities. To put it another way, this strategy is superior to the conventional approaches in terms of the feature representation it provides for the recognition of gestures.

Table:1 Proposed model compare with State of the art methods.

Method	Accuracy	Recall	F1-Score	Precision
RF	0.943	0.997	0.974	0.947
KNN	0.927	0.976	0.948	0.91
NB	0.901	0.973	0.933	0.903
LR	0.909	0.953	0.913	0.892

The proposed model outperforms existing classifiers, indicating that it would provide excellent recognition. We also find that out of these three classifiers, the proposed model demonstrates more than 96% accuracy in recognition, as shown in Fig. 2.

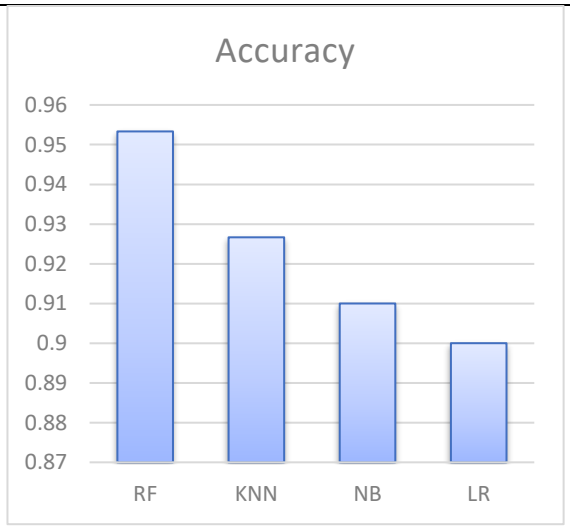

Fig 2: Comparison of Accuracy

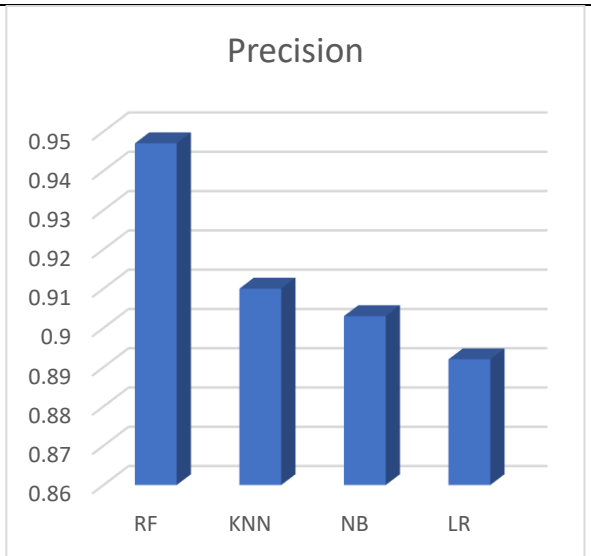

Fig 3: Comparison of Precision

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

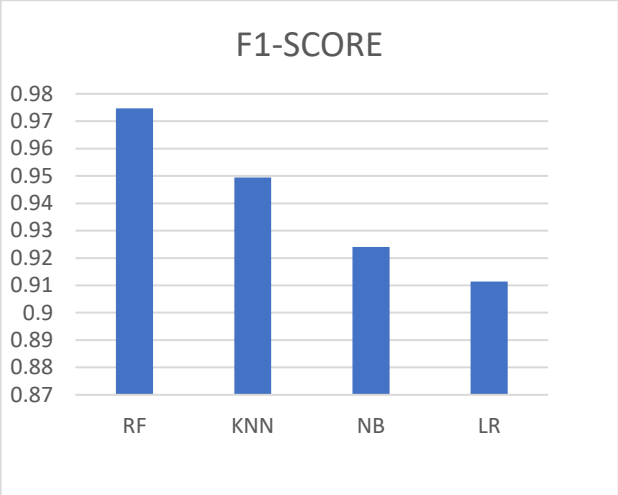

Fig 4: Comparison of F1-Score

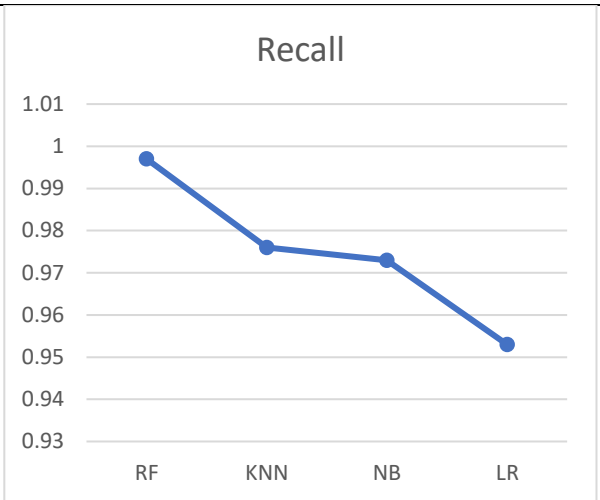

Fig 5: Comparison of Recall

It is apparent that the suggested framework outperforms all current techniques. As a result, we offer our suggested framework for posture classification utilising an Efficient Sparse Gabor Descriptor, which improves the performance of existing machine learning models in stance analysis.

Experiments on Sign Language Recognition utilising various categorisation techniques reveal intriguing differences in performance across measures such as accuracy, recall, F1-score, and precision n. With an accuracy of 94.3%, recall of 0.997, F1-score of 0.974, and precision of 0.947, Random Forest (RF) outperformed the other classifiers. This suggests that RF was a strong option for this assignment because it not only demonstrated excellent overall accuracy but also demonstrated exceptional precision in detecting the right indicators and memory in reducing false negatives.

With a 92.7% accuracy rate, the K-Nearest Neighbours (KNN) classifier also demonstrated good recall (0.976) and F1-score (0.948). Although it accurately detected the majority of the signs, its accuracy (0.91) was somewhat lower than RF's, indicating that it could have been prone to some false positives. Compared to RF and KNN, Naive Bayes (NB) performed well but had a little more space for improvement, achieving a lower accuracy of 90.1%, recall of 0.973, and F1-score of 0.933. It may have had a significant amount of false positives, as shown by its accuracy of 0.903, which was comparable to that of KNN.

Among the investigated approaches, the Logistic Regression (LR) classifier had the lowest precision (0.892) and recall (0.953), with an accuracy of 90.9%. Although LR was acceptable, its F1-score of 0.913 was still somewhat lower than that of the other models, suggesting that it was not as good at identifying signals as RF or KNN.

All things considered, Random Forest was the most well-rounded and effective approach, outperforming KNN in every important statistic. While still feasible, Naive Bayes and Logistic Regression performed somewhat worse in comparison, especially in terms of accuracy and recall.

VII. CONCLUSION

This work introduces the study topic of Computer-Vision-based human gesture analysis, investigates related methodologies, and proposes a human gesture analysis system that parses photos or videos (image sequences) to accomplish human gestures. Automatic hand gesture identification remains a difficult job, mostly due to the variation in how people produce the motions. This paper offers a comprehensive understanding of human gesture estimation. This study builds on previous advances in machine learning and proposes a novel Sparse Gabor Descriptor (SGD)-based technique. Finally, these features can be classified using a random forest for gesture recognition. Experiments show that the suggested strategy delivers competitive results on gesture recognition datasets while requiring significantly less processing complexity. The comparison with roughly three state-of-the-art methodologies yields competitive results in terms of the system's many activities, such as accuracy and classification precision. The

Research Article

proposed technique achieves a recall rate of 99%, an accuracy rate of 94%, an F1-score of 97%, and a precision rate of 94%, which is higher than the KNN, NB, and LR methods.

Experiments on Sparse Gabor Descriptors and Random Forest classifiers for Sign Language Recognition (SLR) show encouraging improvements in accuracy and processing economy. According to the findings, this combination may maintain a high level of identification accuracy despite noise and gesture variances, particularly in solitary sign language tasks. A useful method for SLR is the combination of Sparse Gabor Descriptors and Random Forests, which effectively balances feature extraction and classification. This technique has the potential to be further optimised using large and more varied datasets.

REFERENCES

[1] Kim, M.; Cho, J.; Lee, S.; Jung, Y. IMU sensor-based hand gesture recognition for human-machine interfaces. Sensors 2019,19, 3827.

[2] Linqin, C.; Shuangjie, C.; Min, X.; Jimin, Y.; Jianrong, Z. Dynamic hand gesture recognition using RGB-D data for natural human-computer interaction. J. Intell. Fuzzy Syst. 2017, 32, 3495–3507.

[3] Deafness and Hearing Loss. Available online: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss (accessed on 30 December 2024).

[4] Liu, Z.;Wang, Y.; Vaidya, S.; Ruehle, F.; Halverson, J.; Solja ˇci´c, M.; Hou, T.Y.; Tegmark, M. Kan: Kolmogorov-arnold networks.arXiv 2024, arXiv:2404.19756.

[5] De Smedt, Q.; Wannous, H.; Vandeborre, J.P.; Guerry, J.; Le Saux, B.; Filliat, D. Shrec'17 track: 3d hand gesture recognition using a depth and skeletal dataset. In Proceedings of the 3DOR-10th EurographicsWorkshop on 3D Object Retrieval, Lyon, France, 23–24 April 2017; pp. 1–6.

[6] Han, Pengcheng, Xin He, Takafumi Matsumaru, and Vibekananda Dutta. 2025. "Spatio-Temporal Transformer with Kolmogorov–Arnold Network for Skeleton-Based Hand Gesture Recognition" Sensors 25, no. 3: 702. https://doi.org/10.3390/s25030702

[7] Gaikwad, S.A., Shete, V. User adaptive hand gesture recognition for ISL using multiscale and attention embedded residual densenet with adaptive gesture segmentation framework. SIViP 19, 210 (2025). https://doi.org/10.1007/s11760-024-03668-2

[8] Shi, P., Zhang, Q. & Yang, A. Dual-module spatial temporal information enhancement graph convolutional network for recognizing traffic police command gestures. SIViP 19, 92 (2025). https://doi.org/10.1007/s11760-024-03729-6

[9] Yildiz, Ali & Adar, Nurettin & Mert, Ahmet. (2023). Convolutional Neural Network Based Hand Gesture Recognition in Sophisticated Background for Humanoid Robot Control. The International Arab Journal of Information Technology. 20. 10.34028/iajit/20/3/9.

[10] Liu, Jianbo & Wang, Ying & Xiang, Shiming & Pan, Chunhong. (2021). HAN: An Efficient Hierarchical Self-Attention Network for Skeleton-Based Gesture Recognition. 10.48550/arXiv.2106.13391.

[11] Uke, S.N., Zade, A. Optimal video processing and soft computing algorithms for human hand gesture recognition from real-time video. Multimed Tools Appl 83, 50425–50447 (2024). https://doi.org/10.1007/s11042-023-17608-8

[12] F. A. Farid et al., "Single Shot Detector CNN and Deep Dilated Masks for Vision-Based Hand Gesture Recognition From Video Sequences," in IEEE Access, vol. 12, pp. 28564-28574, 2024, doi: 10.1109/ACCESS.2024.3360857.

[13] D. R. T. Hax, P. Penava, S. Krodel, L. Razova and R. Buettner, "A Novel Hybrid Deep Learning Architecture for Dynamic Hand Gesture Recognition," in IEEE Access, vol. 12, pp. 28761-28774, 2024, doi: 10.1109/ACCESS.2024.3365274.

[14] Z., Tian, Z. & Zhou, M. HandSense: smart multimodal hand gesture recognition based on deep neural networks. J Ambient Intell Human Comput 15, 1557–1572 (2024). https://doi.org/10.1007/s12652-018-0989-7

[15] J. Himasree, J. P L, D. K, A. Kolisetty and S. Naveen, "Video-based Hand Gesture Recognition using Random Forest for Sign Language Interpretation," 2024 Asia Pacific Conference on Innovation in Technology (APCIT), MYSORE, India, 2024, pp. 1-6, doi: 10.1109/APCIT62007.2024.10673591.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[16] D. G. León et al., "Video Hand Gestures Recognition Using Depth Camera and Lightweight CNN," in IEEE Sensors Journal, vol. 22, no. 14, pp. 14610-14619, 15 July15, 2022, doi: 10.1109/JSEN.2022.3181518.

[17] Mujahid, A.; Awan, M.J.; Yasin, A.; Mohammed, M.A.; Damaševičius, R.; Maskeliūnas, R.; Abdulkareem, K.H. Real-Time Hand Gesture Recognition Based on Deep Learning YOLOv3 Model. Appl. Sci. 2021, 11, 4164. https://doi.org/10.3390/app11094164

[18] Q. Gao, Y. Chen, Z. Ju and Y. Liang, "Dynamic Hand Gesture Recognition Based on 3D Hand Pose Estimation for Human–Robot Interaction," in IEEE Sensors Journal, vol. 22, no. 18, pp. 17421-17430, 15 Sept.15, 2022, doi: 10.1109/JSEN.2021.3059685.

[19] Qi, J., Ma, L., Cui, Z. et al. Computer vision-based hand gesture recognition for human-robot interaction: a review. Complex Intell. Syst. 10, 1581–1606 (2024). https://doi.org/10.1007/s40747-023-01173-6

[20] Li, Dongxu & Rodríguez, Cristian & Yu, Xin & Li, Hongdong. (2019). Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison. 10.48550/arXiv.1910.11006.