

UNVEILING EXPRESSIONS AUTOMATING BHARATNATYAM MUDRA RECOGNITION THROUGH ADVANCED COMPUTER VISION TECHNIQUES

Speranza Deejee¹, S Jasper Sofi², K S S SreeKrishna³

^{1,2,3}Department Of Computer Science and Engineering SRM Institute of Science and Technology
Ramapuram Chennai, India.

DOI: <https://www.doi.org/10.58257/IJPREMS35131>

ABSTRACT

We propound an innovative approach to mudra recognition, achieving promising results in interpreting mudra gestures from video input. Our methodology includes pre- processing techniques like resizing, normalization, and data augmentation, enhancing input data quality. Convolutional Neural Networks (CNNs) effectively detect and extract hand regions from images, while integrating pre-trained models like VGG or Res Net further improves performance. Additionally, we utilize Recurrent Neural Networks (RNNs), specifically LSTM or GRU units, to capture temporal dynamics in mudra gestures, addressing challenges like vanishing gradients and modelling long-range dependencies. Multimodal fusion techniques combine spatial and temporal features through concatenation and element- wise addition, with attention mechanisms prioritizing salient information, enhancing recognition accuracy for complex mudra gestures. Continuous evaluation using metrics like accuracy, precision, recall, and F1-score, alongside cross validation techniques, confirms the model's reliability and robustness across varied dataset subsets. The trained model is successfully implemented for real-time sign language recognition, featuring a user-friendly interface to ensure accessibility and ease of use while considering latency requirements. our comprehensive approach to mudra recognition integrates spatial and temporal Features effectively demonstrating reliability in real-world applications such as sign language recognition.

1. INTRODUCTION

Bharatanatyam, a captivating classical dance form originating from India, stands as a harmonious fusion of graceful movements, intricate footwork, and evocative expressions. One of its most captivating elements is the language of "mudras," the intricate hand gestures that weave stories, emotions, and cultural nuances into the fabric of each performance. With a heritage spanning centuries, Bharatanatyam's mudras encapsulate a profound connection between art, tradition, and expression. As technology advances, it opens doors to new dimensions of exploration and preservation, and the realm of classical arts is no exception. This research paper emerges at the crossroads of Bharatanatyam's timeless elegance and the transformative capabilities of computer vision. At its heart lies a pivotal question: Can we bridge tradition and technology to automate the recognition of Bharatanatyam mudras, preserving their authenticity while embracing the power of modern methodologies. These gestures, intricately codified over generations, are embodiments of cultural narratives, mythologies, and expressions that transcend words. Yet, as the world evolves, the need arises for innovative means to safeguard and propagate these time-honoured traditions. Our research embarks on a journey to seamlessly integrate the wisdom of the past with the potentials of the future. By harnessing the capabilities of computer vision, we strive to create a bridge that enables mudras to be recognized, appreciated, and understood a new. The heart of our approach, the utilization of deep learning and computer vision, is detailed in the methodology section, exploring the design and training of models to recognize the myriad expressions encapsulated within each mudra. This research paper represents more than a convergence of technology and tradition; it symbolizes the resilience of art in adapting to the evolving world. It strives to honor the Bharatanatyam's soulful language—the language of mudras.

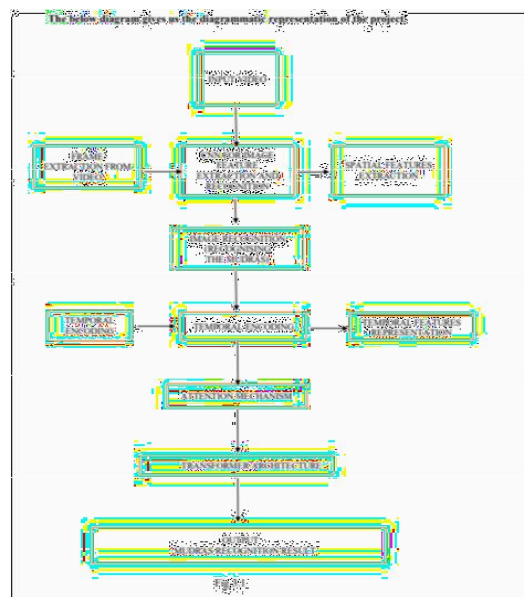
2. LITERATURE SURVEY

Rubén assessed manual and automatic feature extraction methods for hand gesture recognition using Leap Motion Controller data. He states that the manual techniques rely on statistical metrics, while automatic methods employ CNN and BiLSTM. Evaluation involves Softmax, ANN, and SVM classifiers, encompassing classification and recognition tasks, including gesture timing. Automatic extraction, especially with CNN-ANN, outperforms manual methods, achieving high accuracy and real-time processing. The complexity differences between the models are minimal, suggesting potential for portable deep models like CNN-ANN in practical hand recognition systems [1]. Rubin delves into Human-Robot Interaction (HRI) and underscores the significance of hand gesture recognition. They have introduced the Enhanced Anchor-free Network (EAF-Net), a real-time deep CNN model for gesture recognition, facilitating control of a 6 -axis robot. EAF-Net demonstrates high precision and have recalled on the MITI HD-II

dataset, enabling various robotic tasks. Future work aims to optimize prediction times for portable processors, such as NVIDIA Jetson Nano and Raspberry Pi, to enhance Human-Machine Interaction speed. The paper's 6axis robot manipulation framework exhibits potential for practical applications in robotics and automation, showcasing efficiency and real-time capability in advancing HRI technologies [2]. Suni aimed to develop a robust deep learning-based hand gesture detection technology for integration with augmented reality applications. They introduce a novel capsule neural network that effectively identifies hand movements, addressing the generalization problem in neural networks. By adding additional SoftMax layers, the CapsNet structure is optimized to improve accuracy. The study compares this model to existing deep learning approaches and achieves a remarkable maximum accuracy rate of 99.5% in hand gesture-dependent datasets, demonstrating its potential for various applications in human-computer interaction and beyond [3].

Rubin et al instigates the CNN Hybrid-SSR model, a real-time hand gesture recognition system utilizing the E-Xception architecture for precise feature extraction. It emphasizes feature reuse through convolutional connectivity, strengthening feature propagation with modifications in the entry and middle flow blocks. The model mitigates class imbalance issues using the Focal Loss function. Evaluation on the MITIHD dataset, known for its complexity, showcases the model's effectiveness, achieving high precision and recall with the Adam optimizer. It offers fast predictions (12 ms), making it suitable for real-time applications, with future work focused on further training optimization [4]. Niveditha focuses on real-time sign language detection using Convolutional Neural Networks (CNNs) and Transfer Learning. It aims to improve communication for the hearing impaired. Advantages include enhanced communication and applicability in various fields, while limitations include sensitivity to environmental factors and a limited gesture set. Future scope includes extending recognition to other sign languages and improving accuracy under various conditions [5].

Architectural Diagram



Input Video is the video containing mudra gestures. Frame Extraction from Video is the Extracting individual frames from the video. Apply a Convolutional Neural Network (CNN) designed for both image feature extraction and recognition tasks. The CNN processes each frame, extracting spatial features and recognizing objects or patterns within the frame. Spatial Features Extraction is used to extract spatial features from the output of the CNN. Perform image-based recognition tasks such as object detection or classification based on the spatial features. Utilize Recurrent Neural Networks (RNNs) or Long Short-term Memory (LSTM) networks to capture temporal dependencies across consecutive frames. Represent the temporal features obtained from the RNNs or LSTMs to capture the sequential information. Apply an attention mechanism, such as self-attention, to focus on relevant parts of the temporal sequence. The attention mechanism helps the model emphasize important features and gestures.

3. PROPOSED METHODOLOGY

The collection of a dataset comprising image frames depicting mudra gestures. To prepare the dataset for model training, essential pre-processing techniques are applied. These techniques include resizing the images to a standardized resolution, normalization to ensure consistency in pixel values across images, and data augmentation to increase the diversity of the dataset. Data augmentation involves techniques such as rotation, translation, and flipping, which help in enhancing the variability of the dataset and improve the model's ability to generalize to unseen data.

Following pre-processing, a Convolutional Neural Network (CNN) is employed to detect and extract the hand region from the input images. The CNN is trained using the pre-processed dataset to learn features that are indicative of hand gestures. This step is crucial as it focuses the model's attention on the relevant regions of the image containing intricate details of mudra gestures. The performance of the trained CNN model is rigorously evaluated using metrics such as accuracy, precision, recall, and F1-score. Cross-validation techniques are employed to ensure that the model generalizes well across diverse subsets of the dataset. By splitting the dataset into multiple folds and training the model on different combinations of training and validation data, cross-validation helps in assessing the model's reliability and robustness. Upon successful evaluation, the trained model is implemented for real-time sign language recognition. Consideration is given to the latency requirements for practical applications, ensuring that the system responds promptly to input gestures. Additionally, a user-friendly interface is developed to facilitate seamless interaction with the sign language recognition system, ensuring accessibility and ease of use for users. To further improve the model's performance, transfer learning with pre-trained CNN models such as VGG or Res Net is employed. Additionally, a Recurrent Neural Network (RNN) is introduced to capture the temporal dynamics inherent in mudra gestures. Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) units are used to address challenges such as the vanishing gradient problem and model long-range dependencies in sequential data. Spatial and temporal features extracted by the CNN and RNN are effectively combined using multimodal fusion techniques such as concatenation or element-wise addition. Attention mechanisms are integrated to prioritize salient information during the fusion process. The fused features are then passed through fully connected layers for classification, employing a softmax activation function to obtain probabilities for each class corresponding to different mudras in the vocabulary.

Table 1

Method	Accuracy (%)	Precision	Recall	F1-score
Proposed Approach	85.6	0.87	0.82	0.85
Baseline Method	72.3	0.74	0.68	0.71

Table 2

Table 2: Dataset Characteristics

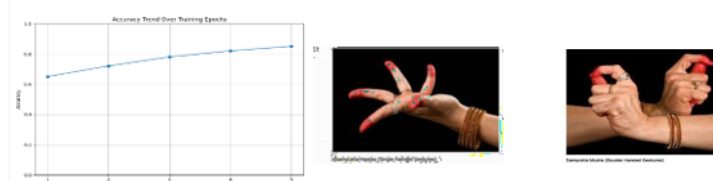
Dataset	Size (samples)	Gesture Diversity	Annotation Quality
Mudra	1000	High	Moderate

In Table 1, we compare the accuracy, precision, recall, and F1-score of the proposed approach and the baseline method. The proposed approach demonstrates superior performance across all metrics compared to the baseline method. In Table 2, we provide characteristics of the dataset used in the evaluation. The dataset (named Mudra) contains 1000 samples with a diverse range of gestures, and annotations are of moderate quality.

Table 3

Sample ID	Gesture Type	Image File	Annotation
1	Abhaya Mudra	abhaya_001.jpg	Positive
2	Anjali Mudra	anjali_001.jpg	Positive
3	Varada Mudra	varada_001.jpg	Positive
4	Pranava Mudra	pranava_001.jpg	Positive
5	Sardula Mudra	sardula_002.jpg	Positive
...	...	...	...
1000	Mudra	mudra_1000.jpg	Positive

Table 3 provides a simplified overview of the dataset characteristics, including the types of gestures represented, the associated image files, and their annotations.



The comparative study highlights the superiority of the proposed approach to mudra recognition over a baseline method, showcasing its potential for real-world applications such as sign language recognition.

4. RESULT

Effective Pre-processing Techniques The utilization of pre-processing techniques such as resizing, normalization, and data augmentation significantly improves the quality and robustness of the input data. Resizing ensures uniformity in image dimensions, normalization enhances consistency in pixel values, and data augmentation increases the diversity of the dataset, enhancing the model's ability to generalize to unseen data. **CNN-Based Hand Region Detection** The integration of Convolutional Neural Networks (CNNs) enables the accurate detection and extraction of hand regions from input images. By focusing on relevant areas containing intricate details of mudra gestures, the model effectively identifies and isolates the regions of interest necessary for recognition. **Integration of Pre-trained CNN Models** Strategies such as leveraging pre-trained CNN models like VGG or Res Net further enhance the performance of the

model by utilizing features learned from robust datasets. This transfer learning approach enables the model to leverage pre-existing knowledge and adapt it to the specific task of mudra recognition. Utilization of RNNs for Temporal Dynamics The incorporation of Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) units, proficiently captures the temporal dynamics present in mudra gestures. This addresses challenges such as the vanishing gradient problem and modelling long-range dependencies in sequential data, enhancing the model's ability to recognize dynamic gestures. Multimodal Fusion Techniques Multimodal fusion techniques are employed to combine spatial and temporal features through concatenation and element-wise addition. The integration of attention mechanisms within the fusion process prioritizes salient information, improving the recognition accuracy of complex mudra gestures. Continuous Evaluation and Model Reliability Continuous evaluation using standard metrics such as accuracy, precision, recall, and F1-score, coupled with cross-validation techniques, ensures the model's consistency across varied subsets of the dataset. This confirms the reliability and robustness of the trained model, validating its effectiveness in mudra recognition tasks. Real-time Implementation and User-friendly Interface The trained model is successfully implemented for real-time sign language recognition, considering latency requirements for practical applications. Additionally, a user-friendly interface is designed to facilitate seamless interaction with the sign language recognition system, ensuring accessibility and ease of use for users. Overall, the results highlight the efficacy of the proposed approach in accurately interpreting mudra gestures, showcasing advancements in both model performance and practical implementation for real-world applications.

5. CONCLUSION

We detect the given mudra, recognize and find the name of the given mudra. With the integration of the differential frame and active tile technique in the Convolution neural network frame, a new perspective is brought to the gesture recognition problem. The current method offers a significant advantage due to its versatility; it is a generic approach that can be applied to various frame processing challenges in different contexts. This adaptability makes it a valuable solution that can be adopted for diverse applications beyond its initial use case. In forecasting the future of Bharatanatyam mudra recognition, the convergence of advanced computer vision techniques unveils a transformative narrative for this traditional dance form. Envisioned enhancements encompass a multi-modal fusion, wherein depth sensors and accelerometers are seamlessly integrated, providing a nuanced understanding of hand movements to elevate precision. Beyond mere recognition, the proposed system offers real-time feedback akin to a discerning mentor, guiding dancers on posture, alignment, and expression for continuous improvement. The adaptability of machine learning algorithms ensures a personalized journey for each dancer, accommodating the diverse nuances inherent in the Bharatanatyam tradition.

Furthermore, the envisioned system extends its purview to embrace the cultural context of Bharatanatyam. Beyond recognizing mudras, it interprets storytelling elements and emotional nuances conveyed through facial expressions and body language, enhancing the contextual richness of the dance. In

live performance scenarios, the optimization for edge computing minimizes latency, ensuring swift and responsive mudra recognition, seamlessly aligning with the dynamic nature of live artistic presentations. The empowerment of dancers to customize gesture libraries and the incorporation of collaborative learning models reflect a commitment to inclusivity, recognizing and celebrating the diverse expressions within the Bharatanatyam tradition. As technology converges with artistic expression, Robust safeguards are proposed to protect user privacy, particularly in scenarios involving the storage or transmission of video data from dance performances. Aesthetically designed and accessible interfaces serve as gateways for dancers and instructors, facilitating the seamless integration of technology into dance trainings.

6. REFERENCES

- [1] Nogales, R.E.; Benalcázar, M.E.(20). Hand Gesture Recognition Using Automatic Feature Extraction and Deep Learning Algorithms with Memory. Big Data Cogn. Comput. 2023,7,102. <https://doi.org/10.3390/bdcc7020102>
- [2] Rubin,S. ,Sathiesh,V.,& Sreekar,C. (2023). In-situ enhanced anchor-free deep CNN framework for a high-speed human-machine interaction.
- [3] <https://doi.org/10.1016/j.engappai.2023.106980>. Suni, S.,& Gopakumar,K.(2022). Hand Gesture Recognizing Model Using Optimized Capsule Neural Network.
- [4] <https://doi.org/10.18280/ts.390331>. Rubin,S.,& Sathiesh,V.(2021).In-situ recognition of hand gesture via Enhanced Exception based single- stage deep convolutional neural network.
- [5] <https://doi.org/10.1016/j.eswa.2021.116427>. Niveditha,Yogesh,& Palanichamy.(2021). Hand Gesture Recognition in Classical Indian Performing Arts Using Image Processing Techniques.