HAND GESTURE RECOGNITION IN CLASSICAL INDIAN PERFORMING ARTS USING IMAGE PROCESSING TECHNIQUES

Niveditha Parthasarathy,
Research Scholar, Department of Information Science and Technology, College of Engineering, Guindy, Chennai-600025
Niveditha_parthasarathy@yahoo.com

Yogesh Palanichamy
Professor, Department of Information Science and Technology, College of Engineering Guindy, Chennai-600025
yogesh@annauniv.edu

Abstract

Hand gesture recognition plays a vital role in many applications in this Information and Communication Technology (ICT) era. Despite being a widely explored area, only a few research works are available specific to mudra recognition that indicates symbolic hand gestures of Bharathanatyam, an Indian classical dance. The present pandemic situation has made this work very relevant. This attempts to aid the transition to digital learning in the area of performing arts. This poses a lot of challenges as no public datasets are available specific to this study which makes the prospective research a pioneering work in this field that is significantly utilitarian. The research work presented in this paper has used a Gaussian filter for noise removal, canny edge detection for boundary identification, a novel combination of statistical geometrical features with Histogram Oriented Gradient (HOG), and Speeded up Robust Features (SURF) for feature extraction. This system with K-Nearest Neighbours (KNN) for classifying the mudras, has achieved an accuracy of 82.67% that outperforms popular classic Machine learning classifiers considering the dynamic nature of the data set in contrast to existing static sets. Moreover, the framework proposed in this paper is flexible enough to apply different combinations of image processing.

Keywords: Mudra, bharathanatyam, digital learning, feature Extraction, classification

1. Introduction

India being a land of variety and diversity has given to the world many beautiful dance forms reflecting the soul of each region through its unique movements, costumes, languages, and, music both folk and classical. To make the paper easier to understand we introduce a small terminology, Mudra- which indicates a symbolic hand gesture that has the power of producing joy and happiness. Adavu- The movements of hands, legs, head, and eyes using the entire body are categorized into groups based on the similarity of movements. Jathi- Combination of various patterns of movements (adavus) with intricate patterns of footwork and hand gestures with specific non-lexical vocables ending with a geometric pattern or dance sequence which are repeated in odd numbers. Swara -The combination of music notes which are melodious and pleasing to the ear. Ancient India classified hand gestures as 'mudras' in yoga, rituals, and classical dance forms.

This work has focused only on two particular mudras, Thripathakam and Alapadmam, as it is observed their usage is persistent. It observed that in today's world anywhere and everywhere computer has permeated, with a thought of enhancing the quality of life, to put it in other words computers have become an inevitable part of human life. This work proposes a framework that is created to boost the classroom teaching-learning experience. for learning and practicing mudras by applying various techniques of image processing, this work attempts to kindle the interest of participants in these art forms. The closure of regular education centres and the institutes for performing arts have had a drastic impact on teacher-participant engagement.

Digital applications in the area of performing arts have become vital in the present scenario of lockdowns and social distancing The outline of the present work can later be extended to other mudras and also used via the Internet as an e-learning package which will be more useful in terms of economy, accessibility, and evaluation. Performing arts play a decisive role between the academic learning and emotional needs of learners. Since mudra recognition falls under the broad category of hand gesture recognition this work can be applied in various domains besides dance like Human-Computer Interaction (HCI), sign language processing, surveillance, and security applications. Gesture Recognition Systems (GRSs) are broadly classified into two categories namely vision-based recognition and non-vision-based recognition. Non-vision-based methods, also called data-glove methods, demand the users to wear data-gloves made up of sensors and take care of wiring details also whereas vision-based systems, require only the camera and do not require sophisticated hardware. A vision-based technique is applied for mudra recognition in this work since it is economical and provides a natural interaction between learners and computers.

The data was captured using a single mobile camera, keeping both the economical and accessibility aspect in mind. Although there are many literary works on hand signs, when it comes to mudra recognition there are hardly any publicly available datasets pertinent to the present study. It is seen that not much information can be provided about some standard data sets specific to Mudra recognition. Presently it is difficult to find a benchmark dataset for mudra recognition which makes this work although more significant. In the field of hand gesture recognition, there are many established datasets but most of these pertain to sign language, driver hand gestures, pantomime, etc. It is observed these datasets are not explicitly designed for Bharathanatyam mudra recognition. Therefore, generating a suitable dataset although being a complex process becomes a prime requisite. A couple of mudras that are more frequently used are chosen and are graded according to their correctness with the help of a domain expert. Pre-processing techniques like image resizing, converting to grayscale and filtering have been used to enhance the video clarity. Different image segmentation techniques and feature extraction methods have been utilized to improve the accuracy and we have applied K-Nearest Neighbour (KNN) to classify various hand gestures into appropriate mudras.

This paper has been organized with Section 2 consists of related works and the value-added of proposed work to existing literature. Section 3 demonstrates the methodology applied, and Section 4 shows the experimental results and analysis. Finally, Section 5 which is the last part of the paper presents the conclusion and direction for future work.

2. Literature Review

Hand gesture recognition continues to be a challenging area to work in, considering various difficulties concerning real-time, cluttered background, lighting, and dynamic movements. Kavitha *et al.* [1] created a data set of 60 images by capturing hand gesture images of 10 different artists 20-33 years of age. Applying various rotation, scaling, and translation techniques made it more ready for deeper understanding. Filtering has been done by using both high pass filters and low pass filters. The authors have extracted a Discrete Wavelet Transform (DWT) and Direct Cosine Transform (DCT) features and finally classified them using the neuro-fuzzy classifier. In the work proposed by Saha *et al*. [2] using a single camera, images are captured with a static background, and the RGB image is converted to grayscale for texture-based segmentation, Sobel edge detection is taken into consideration, and classification is achieved using boundary and chain code which are produced with the help of polygon approximation. Ramadoss & Rajkumar [3] utilized both JPEG (Joint Photographic Experts Group). MPEG (Moving Picture Experts Group) recorded material to annotate and also retrieve dance movements.XML (Extensible Markup Language schema) has been used and alternates like MPEG-7 and RDFs (Resource Description Framework schema) have been discussed. It can be observed from the work done by Xie, *et al.* [4] that to capture the movements of hand accelerations the users wear a ring in which a 3-axis accelerometer is used to collect the hand motions which are integrated into the ring which in turn are transferred via Bluetooth. The average filter has been applied to reduce high-frequency noise and Fast Fourier Transformation (FFT) is also applied to get a good filtering effect. Johnson code has been suggested for comparison of similarity at the gesture matching stage. The forearm is segregated from the arm and by hand calibrations, certain key geometric features are obtained by Bhuyan *et al.* [5]. The variations in the finger positions are discerned using Gaussian distribution. To extract the hand after eliminating noise effects maximum area criteria have been performed. The thickness of the contour of the segmented skin region indicates the separation of the hand from the forearm region. The area that is considered is the wrist region where the thickness is the least. A probabilistic model of abduction and adduction movements has been applied by increasing features and flexion angles more finger combinations can be included. Ansari & Harit, [6] successfully applied the Indian sign language recognition system. Most of the signs were two hand gestures mostly of day-to-day words. Microsoft Kinect camera has been used. The authors have worked in this area to help people with speech disabilities to communicate and have gotten encouraging results. Devi & Saharia, [7] worked in the area of single-hand gestures of the Assamese classical dance Sattriya. The classification is done in two levels based on structural similarity in level 1 which extracts Medial Axis Transformation (MAT) from the images and is later individually recognized by the group. The outcome of this work could be the e-learning of Sattriya Mudras. The major Contribution of this research to Existing Literature are as follows;

- The uniqueness of this work is the creation of a dataset specific to the two mudras (hand poses) of art form Bharatanatyam chosen with great discretion, each further according to their level of correctness subdivided into three grades, by domain expert therefore, (2*3) are six classes, and not readily available. It is pioneering work in this field.
- From a global perspective, the present computer age helps knowledge of traditional intangible art forms to transcend boundaries and helps prevent many ancient art forms from becoming extinct. It enlivens the traditional pedagogy of teaching-learning in the sphere of performing arts besides Indian classical arts like Balinese Indonesian, Thai, Cambodian, and other dance forms of the world where gestures play a significant role.
- The purpose of this research work is to come up with a framework that can recognize two popular and useful (from the pedagogical viewpoint of teaching Bharatanatyam)

mudras of the classical dance which is not addressed in existing works. The framework proposed in this paper is flexible enough to apply different combinations of image processing, segmentation, and machine learning techniques which may depend on video capturing scenarios and the amount of data used in training and testing which is a novel attempt in this direction.

- The exclusivity of this work is that an evaluating algorithm created deals with dynamic images where an orientation has to be given high priority and treated accordingly. Therefore, HOG SURF features that are greatly involved in playing a crucial role in handling orientation are combined with various feature combinations of statistical and geometrical features not taken in prior research with the KNN classifier that achieves 82.67% accuracy which outperforms the accuracy of other classic classifiers.
- Interacting with computers kindles the interest of the participants. More so when it comes to online long-distance learning, especially during a pandemic situation where face-to-face teaching was not possible, a system that classifies the mudra according to its level of correctness makes the session interesting and helps retain the attention span of the learner.
- It is observed that higher accuracy has been derived in earlier work only when the dataset happens to be small and poses are static images of only perfect mudras as against real-time classroom situations wherein in the very first attempt it cannot be expected of learners to hold the pose of the hand perfectly.
- In addition, the combinations of these image processing techniques like HAAR had been analysed and found to not fit certain scenarios, since results were not encouraging for this combination the same was discarded. In this work, dynamic images of 40065 were generated and used from 300 video clips. No other existing work has been done in this domain to this magnitude.

3. Methodology

The mudra recognition framework developed in this research work has five stages: data acquisition, pre-processing, segmentation, feature extraction, and classification. The stage of classification includes two sub-stages namely training and testing. The details of the steps carried out in our research work are shown in the flow chart as depicted in Fig. 1.

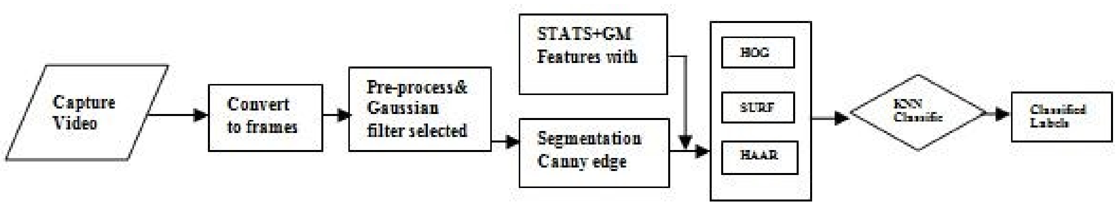


Fig. 1 Flow Chart for Mudra Recognition

3.1 Data Acquisition

Generally, image processing techniques and machine learning algorithms are applied to the existing data set. The proposed work has generated its data set since the data set related to the two mudras of Bharathanatyam was chosen with great discretion, each further subdivided into three grades, therefore, (2*3) are six classes, and are not readily available. The requirement for domain expert assistance needs to be highlighted. One important reason Bharathanatyam mudra recognition is less explored could also be due to the fact there is a shortage of datasets exclusive to this area of application. This work has taken into consideration two mudras of three subgrades each, from video images, and has proposed a system for mudra recognition that falls under the sub-category of hand gesture recognition. Despite the fact literature on hand gesture recognition is adequate, it is observed that there are relatively few works pertinent to this specific domain. The problem is an interesting one and discussed in a detailed fashion in the literature. After careful consideration, two mudras of different levels of correctness have been hand-picked. The two mudras which have been selected are based on the repeated usage of these gestures both during the regular basic movements in dance and also while performing and miming the lyrics. These mudras have been concentrated upon as they have frequent usage and are highly utilitarian. To start with this dataset is designed for two perfect poses of Alapadmam and Thripathakam which are labelled as A grades. For further application in a classroom environment as it cannot be expected that learners keep the pose of the hand perfectly in the very first attempt itself, this dataset is extended in such a manner that each of this Alapadmam mudra and Thripathakam mudra is subdivided according to their correctness for three levels or categories of gradation Good, Average and Not Related Movement.

There are several established datasets like Sebastien Marcel Static Hand Posture Database [8] the images found here contain thousands of images taken from 10 persons with 6 different hand poses depicting a, b, c, point, five, and v in the PNM format. Cambridge hand-gesture dataset [9] contains 9 different gesture categories with 900 video sequences which are further divided into 5 sets of diverse illumination conditions for 3 primitive movements. Data from both the Kinect and the Leap motion [10; 11] is made up of 10 varied gestures acquired from 14 people with both Leap Motion and the Kinect devices, which helps to build and evaluate hybrid gesture recognition systems. Creative Senz3D dataset [12; 13] comprises 11 assorted static gestures each with a repetition of 30 times by 4 persons with a total of 1320 samples. CVRR-HAND3D data set [14] is a publicly available one wherein the gestures are dynamic with 886 samples performed by 8 subjects which include 19 classes of driver hand gestures. The focal point is more on the illumination artefacts and the diverse effects of the interface on driving become secondary. To the best of our knowledge, these datasets do not serve the purpose of our objective which is to recognize the correct mudra. The above standard data sets often used in research works related to hand gesture recognition are not suitable for recognizing mudras in Bharatanatyam since position, orientation, movement, holding and other related features of the hand images available in these datasets are different from that of Bharatanatyam mudras. Hence the need for the creation of a new dataset.

To start with we acquired data from 15 participants. Video frames acquired from different subjects are shown in Fig. 2. This is logical and relevant since the classification algorithm applied comes under supervised learning. The process is tedious and time-consuming. The volunteers were asked to keep their hands away from the body so that other distractions like the problem of clutter, occlusion, and problems related to the dynamic background are avoided. While capturing video the focus is primarily and exclusively on the hand and finger position against a flat plain surface, to avoid having to consider other edges besides that of the finger positions, and only the edges of the fingers are taken into account when edge detection technique is applied for segmentation.

In real-time, the participant doesn't always hold the gesture in a correct position, therefore, we take the different positions of holding the fingers for the same mudra which is further categorized separately by the Bharathanatyam teacher into a particular class based on his/her expertise on the subject. Categorized into grade A (perfect holding of the mudra by a set of learners) grade B (not very correctly held finger positions for the same mudra) and Not Related Movement(NRM) (A set of learners hold the positions in very low accuracy. positions). This work is intended for two mudras of 3 different categories put into the class of its level of correctness and graded as Grade A, Grade B, and Not Related Movement (NRM). The proposed work is meant for 2 mudras of 3 levels of correctness (2*3) and hence 6 classes. In this way, 131 videos were shot and the total video set has a size of approximately 1 GB, and the video set has 3991 images. This aids in a realistic assessment of the correct manner of holding the mudra. This can further the e-learning content for modules in course of study in Bharathanatyam. This study assists the initiative to equip teachers to transit to an online teaching-learning experience. Participants were requested to make mudras as shown in Figure

2 (Alapadmam) and (Thripathakam). Video from participants for 3 categories namely good, average and non-relevant is shown in Fig. 3.

Filtering techniques are applied to remove noise and enhance images. After the canny edge is applied for the segmentation a combination of feature extraction techniques is applied to the Region of interest, which here is the hand mudra. Videos of size 8MB were shot in a classroom lit with ordinary fluorescent tubes using a single in-built camera in a smartphone Redmi with 3 GB RAM and a capacity of 32 GB. However later we added more videos to the data set as an effort to improve the accuracy of recognition and to deal with all the classes of mudras. Mudras can be shown by both the right and left hand and hence we have acquired a few videos that are about the left hand also. According to our literature survey, Quet & Dahdah [15] dealt with digital studies and how they have influenced people from all walks of life. How different disciplines have made use of cyberspace generated immense interest.

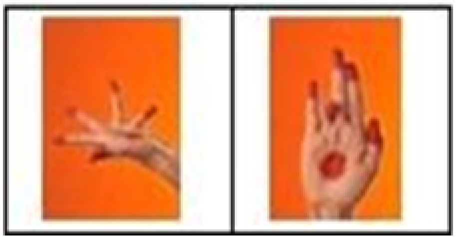
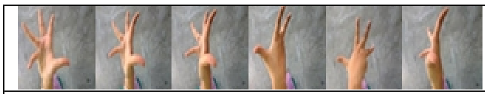
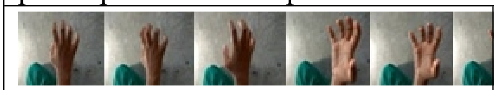
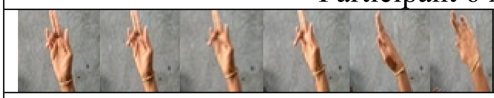
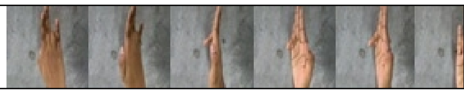
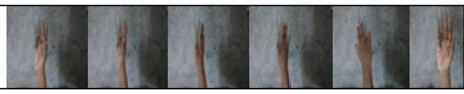


Fig. 2 Displays Alapadmam Thripathakam

	(a)		(b)
3.1.Video frames acquired from different participants (a)-participant 1 (b)-participant 2 for AlapadmamA Grade			
	(c)		(d)
3.2.Video frames acquired from different participants (c)-participant 3 (d)-participant 4 for Alapadmam B Grade			
	(e)		(f)
3.3. Video frames acquired from different participants(e)-Participant 5, (f)-Participant 6 Alapadmam NRM Grade			
	(g)		(h)
3.4 Video frames acquired from different participants (g)-Participant 7, (h)-Participant 8 Thripathakam A Grade			
	(i)		(j)

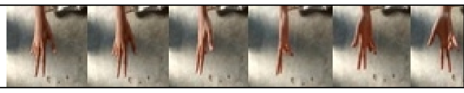
<table>
<tr><td colspan="4">3.5 Video frames acquired from different participants (i)-Participant 9, (j)-Participant 10 Thripathakam B Grade</td></tr>
<tr><td></td><td>(k)</td><td></td><td>(l)</td></tr>
<tr><td colspan="4">3.6.Video frames acquired from different participants(k)-Participant 11,(l)Participant 12Thripathakam NRM Grade</td></tr>
</table>

Fig. 3 Video frames were acquired from participants for Alapadmam and Thripathakam mudras A, B, and NRM Grades

3.2 Pre-Processing

This stage is very much significant to enhance the quality of the image. The videos are converted into frames of size 64x128 to retain the resolution, have uniformity, and enhance the speed of processing. The extracted frames are in RGB form and these frames which are in three dimensions in color are converted into grayscale binary images of two dimensions in which the intensity of the pixel varies from 0 to 253. Most images are affected to some extent by noise, i.e., unexplained variation in data disturbances in image intensity which are either uninterpretable or not of interest. Image processing and analysis are often simplified if this noise can be filtered out. The techniques with which certain frequency components in an image can be chosen or rejected is what we term filtering. Filtering techniques that are applied to remove noise and enhance images are classified into two categories namely global and local techniques.

In certain situations, to focus on the enhancement of a particular portion of the image, local details of an image are to be dealt with and local operations are preferred. There are two types of local or neighbourhood operations: image smoothing which is spatial domain low pass filtering and the other one is image sharpening which is spatial domain high pass filtering. The research work proposed in this paper is to recognize mudra in Bharathanatyam which is a type of hand gesture. Hence the system has to focus on a particular portion of an image which is nothing but the hand. Hence local techniques are preferred in our work rather than global techniques. From various related works available in the literature, we have observed that Gaussian filters are widely used in recognizing mudras of various classical dances in particular and hand gestures in general [5; 7; 16]. Moving filter and radius outlier removal filters have been used widely in hand gesture recognition [4; 17; 18]. Mudra recognition, a special type of hand gesture recognition requires operations, like removal of noise that follows a normal distribution, edge enhancement, and adding motion blur. Hence Gaussian, Laplacian, and

Motion-blur filters are a few suitable candidates for hand gesture recognition and we have considered these techniques in our work for mudra recognition in Bharathanatyam.

Peak Signal to Noise Ratio (PSNR) and Mean Square Error (MSE) are the parameters that are used regularly to measure the quality of the images. MSE is the cumulative squared error between the compressed and original images. PSNR is the reciprocal of MSE that represents the similarity ratio between two different images and is measured in decibels. However, in terms of PSNR and MSE, we have found that the Gaussian filter performs better than other filtering techniques in our work. Moreover, the Gaussian filter is linear and can blur edges and reduces contrast, and is faster than other filters since it is based on addition and multiplication and sorting is not involved.

3.3 Segmentation

The process of segmentation can be performed in two different ways. In one method the image is dealt with in its entirety as one whereas the other method considers the image as a collection of small objects. The first one is called global segmentation and the second one is called local segmentation and the parameters considered for segmentation vary according to the composition of the image like the regions present or objects present in the image. The various segmentation techniques are threshold-based, Edge-based, ANN-based, Region-based, PDE Kernel-based, and Fuzzy kernel-based. Among these various segmentation algorithms, we have chosen edge-based segmentation. Canny edge has been found to give a better visual representation over Sobel and Prewitt edge and has been applied in this work. Features are extracted from detected edges as edges are important features and aid to separate the regions and regions of interest which in this work is the hand mudra that is segmented, based on edge detection. Every single video clip is converted to several frames which vary depending on the duration of the clip. For every single movement of the position of the finger and the orientation of the hand in every frame, a combination of HOG, SURF, and Statistical and geometrical features are extracted and trained. The aspect of the different orientations of the hand movement is taken care of during training and testing. Moreover, HOG and SURF features play a major role in dealing with orientation. The Fig. 4 shows the output when a canny edge is applied.

	(a)		(b)

4.1 Canny edge segmentation being applied for (a)-participant1,(b)-participant2
Alapadmam A Grade

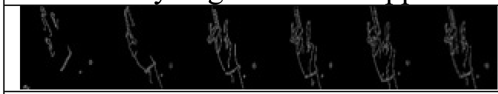
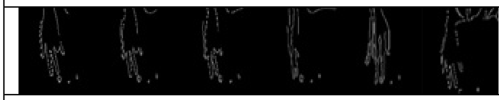
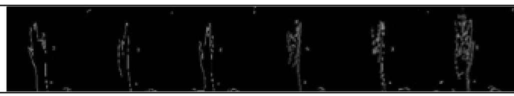
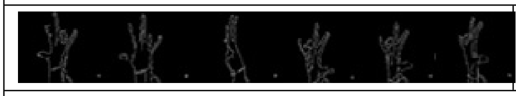
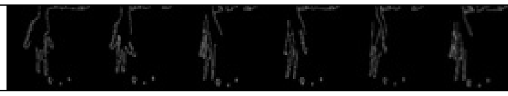
	(c)		(d)
4.2 Canny edge segmentation being applied for (c)-participant 3,(d)-participant 4 Alapadmam B Grade			
	(e)		(f)
4.3 Canny edge detection applied for (e)-participant 5,(f)-participant 6 Alapadmam NRM			
	(g)		(h)
4.4. Canny edge detection applied for (g)-participant 7,(h)-participant 8 Thripathakam A Grade			
	(i)		(j)
4.5 Canny edge detection applied for (i)-participant 9,(j)-participant 10 Thripathakam B Grade			
	(k)		(l)
4.6. Canny edge detection applied for (k)-participant 11,(l)-participant 12 Thripathakam NRM			

Fig. 4 Canny edge detection was applied for Thripathakam and Alapadmam mudras Grade A, B & NRM.

Color-based and texture-based segmentation are other available alternatives. However, these two segmentation techniques are not suitable for our work since the dancer applies color (altha) during a performance on the tips of their fingers. The fundamental thrust in image processing is the edges and they represent the boundaries between the regions effectively. Moreover, edge-based segmentation can reduce the position of the image to be analysed considerably by eliminating the unwanted regions in the image. At the same time, edge-based segmentation can retain the structural properties of the image which are important for processing and analysis.

3.4 Feature Extraction

Image feature is a simple image pattern in terms of which it is possible to describe the content of an image to some extent. Edges, corners, shapes, and interest points are a few features that are often used to obtain useful information about the image content. The main role of features in image processing and pattern recognition is to transform the visual information into vector spaces. Feature extraction is the process of extracting the relevant features that are required for tasks like classifying the objects in an image, recognizing the objects or a few portions of the image, etc. It is a dimensionality reduction technique since it reduces the dimension of the image by selecting only the relevant information from the ROI that has been obtained from the segmentation of objects. This process makes the job of classifiers easy and effective since it chooses only relevant information from the image without losing the quality. The process of classification results in a loss of accuracy if some vital features are lost in the process of segmentation. Feature extraction techniques consider unique features that can be easily tracked and can be easily compared. Corners are the versions in images that have large variations in intensity in all directions. Harris & Stephens, [19] found the difference in intensity for a displacement of (u, v) in all directions as shown in equ (1)

$$E\,(u,v) = \sum w(x, y) + [I(x+u,y+v) - I(x,y)]xy \qquad (1)$$

where $w(x,y)$is the window function; $I(x+u,y+v)$ is the shifted intensity and $I(x,y)$is the intensity. This function is maximized to find the corners in an image. Later Shi, [20] made a minor modification to the Harris corner detector. They modified the scoring function, $R = \lambda_1\lambda_2 - K(\lambda_1 + \lambda_2)$ as $R = \min\,(\lambda_1,\lambda_2)$, where λ_1,λ_2 are the eigenvalues computed in terms of image derivatives and directions. These techniques are rotation invariant, however, they are not scaling invariant. Scale Invariant Feature Transform (SIFT) is a technique proposed by Lowe [21] that is both rotation-invariant and scale-invariant.

Five steps involved in the SIFT algorithm are scale-space extreme detection, keypoint localization, orientation assignment, keypoint descriptor, and keypoint matching. However, SIFT takes more time to extract the features. Speeded-Up Robust Features (SURF) is a speedup version of SIFT.SIFT approximates Laplacian of Gaussian (LOG) with Difference of Gaussian (DOG) to find the scale space and surf approximates LOG with box filter SURF relies on the determinant of Hessian matrix for both scale and location. SURF uses wavelet responses in horizontal and vertical directions and a neighbourhood of size 20x20 is taken around the key point where S is the size and is separated into 4x4 sub-regions. For each sub-region, horizontal and vertical wavelet responses are obtained and represented in the form of a vector. This vector gives SURF feature descriptor with 64 dimensions. SURF employs the sign of Laplacian for underlying interest points without additional computation cost. It is three times faster than SIFT and the performance is comparable to SIFT. SURF is preferred over SIFT since the mudra is to be recognized in a reasonable time and the mudra recognition does not have any dangerous consequences if the recognition includes any false positive.

Histogram of Oriented Gradients (HOG) is based on the distribution of directions of gradients, i.e. the histograms of oriented gradients are used as features. An image gradient refers to the directional change in the intensity or colour of an image and this is useful since the magnitude of gradients is large around edges and corners. Five steps involved in HOG are pre-processing, calculating the gradient images, calculating the histogram of gradients, block normalization, and HOG feature vector calculation. HOG feature descriptor is successful in target detection and recognition and it represents the surface and shape of the local area in an image. HOG descriptor is highly useful for pedestrian or animal detection in video or images since it describes contour and edge features exceptionally in various objects such as cars, bikes, and animals. Moreover, since it operates on local parts it is invariant to geometric and photometric transformations. This is a desirable feature in the context of mudra recognition since it allows different body movements to be overlooked if a participant maintains a roughly upright position.

Most of the feature extraction techniques are based on the intensities which are computed in terms of RGB values of every pixel in the image. The alternate factor that can be used to extract the features is the HAAR wavelets. The major advantage of HAAR-like features is the speed of calculation. The outcome of the experiments based on the combination of HAAR features on the dataset proposed in this work was dismal and therefore not considered further. In the research work proposed in this paper, we had taken 90 videos which are 70% of the total video clips and trained the mudra recognition system based on the statistical features. Later we enhanced the mudra recognition system by adding geometric features and HOG features to it. We have also tried to improve the system by adding geometrical features to SURF features.7 features have been. distinctly mentioned as statistical features with their names and for further training, additional 15 geometrical features have also been specified. Table 1 shows the 7 statistical features used in this work are mean, standard deviation, entropy, variance, skewness, smoothness, and kurtosis, and the 15 geometrical features are area, centroid, Bounding Box, Major Axis Length, Minor Axis Length, Eccentricity, Orientation, convex hull, convex image, convex area, filled area, extrema, equivdiameter, solidity and extent along with 1*3780 HOG features [22] and 26*2 SURF features [23].

Table 1 statistical features used

Category of features	**Constituent Features**	**No of features**
Statistical Features	Mean, standard deviation, entropy, variance, smoothness, kurtosis, and skewness	7

Geometrical Features	Area, centroid, Bounding Box, Major Axis Length, Minor Axis length Eccentricity, Orientation, convex hull, convex image, convex area, filled area, extrema, equivdiameter, and solidityextent	15
HOG features	HOG	1*3780
SURF features	SURF	26*2

3.5 Classification

Dealing with large datasets is a complex process and conventional data processing techniques are not suitable. Mudra recognition also involves a large data set since we have to make the computer recognize the hand gesture of a participant and decide if it is a correct mudra or not. The process of making the computer recognize the hand gestures as mudras involves mining or digging deep into a dataset for identifying patterns, establishing the relationship among the patterns, and analysing the patterns. The data analysis part of problem-solving needs to find an appropriate model that can describe and distinguish different data available in the dataset to different data classes and concepts. Even if we go for machine learning techniques to classify the video into the appropriate category of the mudra, we need a domain expert, who labels the video data to the correct mudra based on his expertise. Classification is the process that does this task and it segregates the total population (given a large dataset) into a set of sub-population (a set of categories). The classification includes two major stages namely learning (training) and predicting class labels and evaluating the developed model (classifying and testing).

Due to exponential developments that are taking place in the fields of machine learning and data mining, the veracity of efficient algorithms is available in the literature and the selection of an appropriate algorithm is a difficult task. The selection of algorithms, by and large, depends on the dataset and the task to be accomplished. Another way of deciding the algorithm is to explore the literature in the relevant field. K-Nearest Neighbour (KNN) algorithm is a frequently used algorithm in the field of hand gesture recognition. It is a non-parametric algorithm and its computational complexity is less. Since we expect the mudra recognition system to be available at a reduced cost we want our system to be simple in terms of both hardware and software components. Hence KNN algorithm has been applied to classify various frames of the video dataset and it is seen that it has a better recognition rate.

4. Experimental Results

The dataset for two mudras that are significantly oft used - Thripathakam and Alapadmam, consisting of 131 videos (3991 images) have been acquired from around 15 different participants. With the help of a domain expert each mudra is classified into three classes, namely grade A (good quality), grade B (average quality), and NRM (Non-Related Mudra). Hence two mudras of three sub-categories we get six different classes for recognition. The quality of the image is enhanced using certain pre-processing techniques. Using the edge detection technique of Canny from amongst the edge detectors the hand gesture is segmented, and feature vectors such as statistical features and geometrical features combined with HOG, SURF, and HAAR features are extracted for these six different classes.

The images are trained and tested by our system. Training has been done once and for testing, we are giving both already trained data along with freshly given data. Viola & Jones [24]. proposed an object detection scheme based on HAAR feature-based cascade classifiers. Since results using HAAR are not encouraging for the proposed dataset it was discarded. Accuracy, specificity, and sensitivity have been chosen to represent the performance metrics for training and testing 300 videos. A comparison of various classifiers on a data set is made and the results are presented in Table 2.

Table. 2 Performance of KNN in Comparison against few Classic Machine Learning Classifiers

Feature extracted and No. of videos	Classifier	Accuracy	Sensitivity	Specificity
SURF +7 statistical + 14 geometrical features 102 videos	KNN	0.8244	0.9524	0.9636
HOG +7 statistical + 14 geometrical features 102 videos	KNN	0.8244	0.9524	0.9636
SURF +7 statistical + 14 geometrical features 102 videos	Random Forest	0.8015	0.9048	0.9364
HOG +7 statistical + 14 geometrical features 102 videos	Random Forest	0.8168	0.8333	0.9065

SURF +7 statistical + 14 geometrical features 102 videos	Discriminant Analysis	0.4351	0.4762	0.8000
HOG +7 statistical + 14 geometrical features 102 videos	Discriminant Analysis	0.4733	0.5000	0.8131
SURF +7 statistical + 14 geometrical features 102 videos	Decision tree	0.5191	0.5238	0.7909
HOG +7 statistical + 14 geometrical features 102 videos	Decision tree	0.5496	0.5833	0.8131

KNN classification algorithm has been applied to classify the mudras into their respective class. The process of classification includes two major phases namely training and testing. The training phase fits the data to the model and the testing phase tests the model to check its suitability of the model for the problem chosen. The experiment was tried with two proportions. The First proportion is 70% for training and 30% for testing and the second proportion is 80% for training and 20% for testing. It is observed from the classification result that the proportion of 80% for training and 20% for testing gives relatively better results. We have applied the KNN classification algorithm for mudra recognition by creating and applying an algorithm with different combinations of statistical features geometrical features, HOG features, and SURF features with a varying number of videos. The results we obtained for training 80% of 131 videos that is 102 videos against comparison with a few classic Machine Learning Classifiers Random forest, Discriminant Analysis, and Decision tree. It is observed empirically KNN classification gives better accuracy for this dataset.

To see, if we can get better results when we generate more data we have increased the dataset to 300 videos (around double). Table 3. shows the results for further increase in data and the Graph to represent it is shown in Figure 3. We have discarded HAAR as the results were not encouraging. Thereafter when we increase the data set significantly we directly take 80% as the proportion to be trained we find that there is only a slight increase in the accuracy. A few major observations we have found that the mismatch between look-alike hand gestures shall be attributed to different hand shapes, sizes, and bigger orientations.

Table. 3 Performance of KNN on 6 Classes Alapadmam A Grade, B Grade, and NRM, Thripathakam A Grade, B Grade, and NRM with a Combination Of Statistical,

Geometrical Features with Hog and Surf Features with an Increased (300) Number of Video.

Feature extracted and No. of videos	Accuracy	Sensitivity	Specificity
7statistical +14 geometrical and HOG of 210 videos and test 300 videos	0.7333	0.7400	0.9400
7 statistical +14 geometrical and HOG of 240 videos and test 300 videos	0.8267	0.8400	0.9600
7 statistical +14 geometrical and SURF of 210 videos and 300 videos	0.7333	0.7400	0.9400
7 statistical +14 geometrical and SURF of 240 videos and test 300 videos	0.8267	0.8400	0.9600

The comparative analysis of existing approaches and the proposed work for hand mudra recognition is done in terms of various classifiers and a combination of features extraction techniques with existing relatively smaller datasets is illustrated in Table 4. From the analysis, it is found that our proposed work uses video clips, capturing only the hand mudra during movement which not only captures the perfect mudra but also mudras of varying correctness where each level of correctness refers to a class. This is done since our work is used in a classroom for teaching-learning purposes and not merely to recognize the perfect mudra.

Table. 4 Comparison of various approaches to hand gesture recognition

References	Number of datasets	Feature extraction technique	Recognition rate (%)		Classifier used
Kumar *et al.* 2017 [25]	4 dance videos from 5 dancers for 2		Same dance Exp-1	Diff dance Exp-2	MCMLAB classifier

	songs in the dance styles of Kuchipudi and Bharathanatyam	HOG	84	67	
		SIFT	82	65	
		SURF	80	59	
Anami & Bhandage [26]	75static images trained and 25 static images tested for each class of 28 mudras	Hu moments, and of the grid lines with the Eigenvalues	98		ANN Classifier
		Intersections	97.1		
Cote-Allard, *et al*. [27]	7 gestures 17 participants	Informative features learned from the sizeable amount of data generated by the accrual of the signals of multiple users	98.31		CWT-based ConvNet
	18 gestures 10 participants		68.98		Raw EMG-based ConvNet
Benitez-Garcia, *et al.* [28]	4000 static and dynamic gestures from 50 subjects	Semantic segmentation	82.2		ResNext 101 with RGB
Upadhye, *et al.* 2021[29]	For PSO optimization approach, dataset consisting of 65000 images of Kannada language	For enhancing feature extraction encoder-decoder is set up.	91.66		CNNPSO

	numerals are utilized for training and for testing 5000 images are used			
Proposed	300 video clips consists of 40065 dynamic frames	7 statistical + 14 geometrical for both HOG and SURF of 240 videos and test 300 videos	82.67	KNN

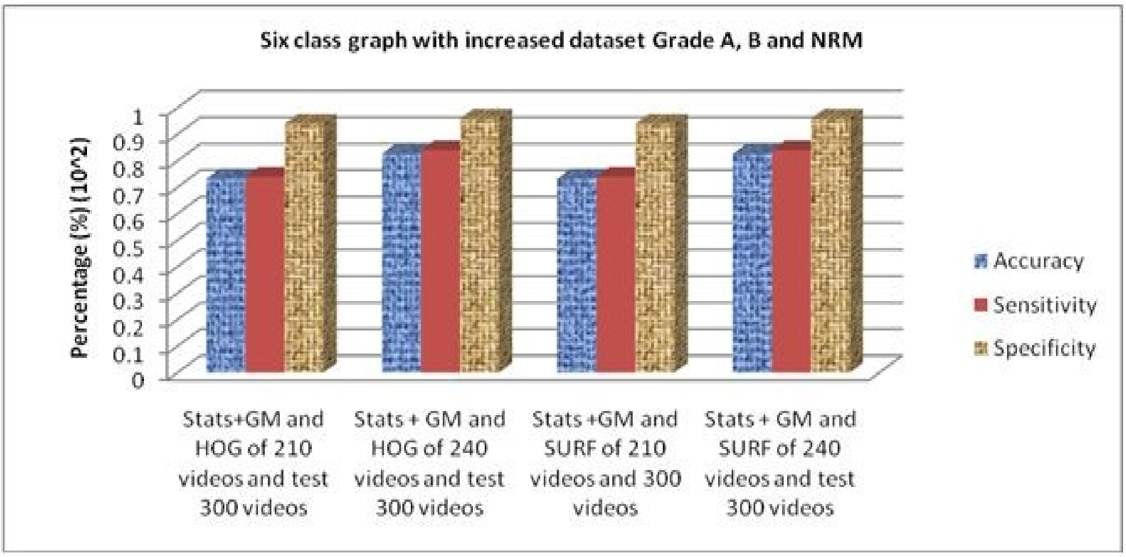


Fig. 5 The graph represents the Dataset of Six classes for Grade A, B & NRM with a combination of features and performance metrics.

5. Conclusion and Future Work

With evolution taking place and the aesthetic creativity of art forms joining hands, hand gestures have been codified for sign languages and secret codes. Indian classical dance has classified hand gestures as single hand and double hand mudras. The research work that has been reported in this paper is to match single hand gestures present in the video to the appropriate class which is one among the six classes we have identified for recognizing Alapadmam and Thripatakam mudras of Bharathanatyam. A hand gesture is classified as Class

A (good), class B (average), and Non-Related Movement (NRM) of both Alapadmam and Thripathakam Alapadmam and Thripathakam as in real-time it cannot be expected of participants to hold the mudra in a perfect pose in the very first time of learning.

A set of image processing operations has been carried out to extract the features of the hand gestures which are combined with the KNN classification algorithm. Different combinations of statistical, geometrical, SURF, HOG, and HAAR features have been used in the KNN algorithm to classify hand gestures. The thought of this work is to kindle the interest of the millennial tech-savvy generation towards the art form of Bharathanatyam and to enhance their learning experience using modern-day technology. This helps propel digital distance learning techniques. It supports the learning journey of participants in their artistic pursuits. To have a better understanding, as the next step, we can explore and use different segmentation techniques and make a comparative analysis of the results. The recognition scheme designed in this paper is to be an application of techniques for a specific task which can be enhanced further by using different combinations of image processing techniques, features and classification algorithms to improve the accuracy of the mudra recognition process. We have planned to come up with a mudra recognition system that can recognize and classify all 28 single-hand mudras in our subsequent research work. The ultimate goal of our research work is to culminate with an e- tutorial package using which the aspirants can learn most of Bharathanatyam mudras with less human intervention.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The author expresses the sincere thanks to College of Engineering, Anna University, Chennai for completing this manuscript.

REFERENCES:

[1] Kavitha J, Lili N L,Razali Y and Jaganathan M (2015), Single hand gesture recognition based on DWT and DCT feature extraction and neuro-fuzzy classifier. *In: E-Proceedings of the International Conference on Social Science Research, ICSSR*, 2015 pages 92-103.

[2] Saha, S., Konar, A., Gupta, D., Ray, A., Sarkar, A., Chatterjee, P., & Janarthanan, R. (2014). Bharatanatyam hand gesture recognition using polygon representation.

In *Proceedings of the 2014 International Conference on Control, Instrumentation, Energy and Communication (CIEC)* (pp. 563-567). IEEE.

[3] Ramadoss, B., & Rajkumar, K. (2007). Semi-automated annotation and retrieval of dance media objects. *Cybernetics and Systems: An International Journal*, *38*(4), 349-379.

[4] Xie, R., Sun, X., Xia, X., & Cao, J. (2015). Similarity matching-based extensible hand gesture recognition. *IEEE sensors journal*, *15*(6), 3475-3483

[5] Bhuyan, M. K., Neog, D. R., & Kar, M. K. (2012). Fingertip detection for hand pose recognition. *International Journal on Computer Science and Engineering*, *4*(3), 501.

[6] Ansari, Z. A., & Harit, G. (2016). Nearest neighbour classification of Indian sign language gestures using kinect camera. *Sadhana*, *41*(2), 161-182.

[7] Devi, M., & Saharia, S. (2016, December). A two-level classification scheme for single-hand gestures of Sattriya dance. In *2016 International Conference on Accessibility to Digital World (ICADW)* (pp. 193-196). IEEE.

[8] Marcel, S. (1999, May). Hand posture recognition in a body-face centered space. In *CHI'99 Extended Abstracts on Human Factors in Computing Systems* (pp. 302-303).

[9] Kim, T. K., & Cipolla, R. (2008). Canonical correlation analysis of video volume tensors for action categorization and detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(8), 1415-1428.

[10] Marin, G., Dominio, F., & Zanuttigh, P. (2014, October). Hand gesture recognition with leap motion and kinect devices. In *2014 IEEE International conference on image processing (ICIP)* (pp. 1565-1569). IEEE.

[11] Marin, G., Dominio, F., & Zanuttigh, P. (2016). Hand gesture recognition with jointly calibrated leap motion and depth sensor. *Multimedia Tools and Applications*, *75*(22), 14991-15015.

[12] Memo, A., Minto, L., & Zanuttigh, P. (2015). Exploiting silhouette descriptors and synthetic data for hand gesture recognition.

[13] Memo, A., & Zanuttigh, P. (2018). Head-mounted gesture controlled interface for human-computer interaction. *Multimedia Tools and Applications*, *77*(1), 27-53.

[14] Ohn-Bar, E., & Trivedi, M. M. (2014). Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations. *IEEE transactions on intelligent transportation systems*, *15*(6), 2368-2377.

[15] Quet, M., & Dahdah, M. A. (2020). Technologies without borders? The digitization of society in a postcolonial world. *Science, Technology and Society*, *25*(3), 363-367.

[16] Kumar, K. V. V., & Kishore, P. V. V. (2018). Indian classical dance mudra classification using HOG features and SVM classifier. In *Smart Computing and Informatics* (pp. 659-668). Springer, Singapore.

[17] Ofli, F., Demir, Y., Erzin, E., Yemez, Y., & Tekalp, A. M. (2007, July). Multicamera audio-visual analysis of dance figures. In *2007 IEEE International Conference on Multimedia and Expo* (pp. 1703-1706). IEEE.

[18] Kapuscinski, T., Oszust, M., Wysocki, M., & Warchol, D. (2015). Recognition of hand gestures observed by depth cameras. *International Journal of Advanced Robotic Systems*, *12*(4), 36.

[19] Harris, C., & Stephens, M. (1988, August). A combined corner and edge detector. In *Alvey vision conference* (Vol. 15, No. 50, pp. 10-5244).

[20] Shi, J. (1994, June). Good features to track. In *1994 Proceedings of IEEE conference on computer vision and pattern recognition* (pp. 593-600). IEEE.

[21] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, *60*(2), 91-110.

[22] HOG feature information is available from ¡https://www.analyticsvidhya.com/blog/2019/09/feature-engineering- images-introduction-hog-feature-descriptor/

[23] SURF feature information is available from ¡https://in.mathworks.com/help/vision/ref/surfpoints.html.

[24] Viola, P., & Jones, M. (2001, December). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001* (Vol. 1, pp. I-I). Ieee.

[25] Kumar, K. V. V., Kishore, P. V. V., & Anil Kumar, D. (2017). Indian classical dance classification with adaboost multiclass classifier on multifeature fusion. *Mathematical Problems in Engineering*, *2017*.

[26] Anami, B. S., & Bhandage, V. A. (2019). A comparative study of suitability of certain features in classification of Bharatanatyam mudra images using artificial neural network. *Neural Processing Letters*, *50*(1), 741-769.

[27] Côté-Allard, U., Fall, C. L., Drouin, A., Campeau-Lecours, A., Gosselin, C., Glette, K., ... & Gosselin, B. (2019). Deep learning for electromyographic hand gesture signal classification using transfer learning. *IEEE transactions on neural systems and rehabilitation engineering*, *27*(4), 760-771.

[28] Benitez-Garcia, G., Olivares-Mercado, J., Sanchez-Perez, G., & Yanai, K. (2021, January). IPN hand: A video dataset and benchmark for real-time continuous hand gesture recognition. In *2020 25th International Conference on Pattern Recognition (ICPR)* (pp. 4340-4347). IEEE.

[29] Upadhye, G. D., Kulkarni, U. V., & Mane, D. T. (2021). Improved Model Configuration Strategies for Kannada Handwritten Numeral Recognition. *Image Analysis & Stereology*, *40*(3), 181-191.