

Recognizing Bharatnatyam Mudra Using Principles of Gesture Recognition

¹ Shweta Mozarkar, ² Dr.C.S.Warnekar

¹ Department of Computer Science & Engineering, SRCOEM, RTMNU
Nagpur, Maharashtra, India

² Department of Computer Science & Engineering, JIT, RTMNU,
Nagpur, Maharashtra, India

Abstract

A primary goal of gesture recognition research is to create a system which can identify specific human gestures and use them to convey information for the device control. Gesture Recognition is interpreting human gestures via mathematical algorithms. Indian classical Dance uses the expressive gestures called Mudra as a supporting visual mode of communication with the audience. These mudras are expressive meaningful (static or dynamic) positions of body parts. This project attempts to recognize the mudra sequence using Image-processing and Pattern Recognition techniques and link the result to understand the corresponding expressions of the Indian classical dance via interpretation of few static Bharatnatyam Mudras. Here, a novel approach of computer aided recognition of Bharatnatyam Mudras is proposed using the saliency technique which uses the hypercomplex representation (i.e., quaternion Fourier Transform) of the image, to highlight the object from background and in order to get the salient features of the static double hand mudra image. K Nearest Neighbor algorithm is used for classification. The entry giving the minimum difference for all the mudra features is the match for the given input image. Finally emotional description for the recognized mudra image is displayed.

Keywords: Saliency detection technique, Gesture recognition, Image processing, Quaternion Fourier Transform.

1. Introduction

Advances in digital image processing in the last few decades have led to development of various computer aided digital image processing applications. One such significantly researched area is the human gesture recognition. A gesture is a form of non-verbal action-based communication made with a part of the body, and used in combination with verbal communication. It is thus a form of perceptual information (mostly visual) [1]. The language of gesture is rich in ways for individuals to express a variety of feelings and thoughts, from contempt and hostility to approval and affection. We often use different hand and facial gestures to supplement verbal communication. Frequent use of certain gestures has

acquired standard meaning. As these gestures are perceived through vision, it is a subject of great interest for computer vision researchers. It is well known that the classical Indian dance like Bharatnatyam traditionally uses certain hand and facial gestures to convey standard emotions as a supporting visual mode of communication with the audience. Nearly Sixteen types of Indian classical dance like Bharatnatyam, Kathak have been traditionally using over fifty two types of expressive gestures called mudras (like pataka, mayur) to enact the background song (Patham). These mudras are expressive meaningful (static or dynamic) positions of body parts. Mudras may thus be perceived as a body-language based text (Patham) compression technique for the information to be conveyed to the audience. Under ideal situation, the mudra viewer should be able to understand the meaning of dance sequence irrespective of the language of background song. The recognition of Mudra sequence can thus create language independent universal communication environment for the dance drama [2]. This project attempts to interpret such Bharatnatyam mudras through gesture recognition process.

A novel approach of computer aided recognition of Bharatnatyam Mudras is proposed using the hybrid saliency technique which is an amalgamation of both top down & bottom up approach. The system uses the hypercomplex representation (i.e., Quaternion Fourier Transform) to get the salient features of the static mudra image. Now this process is carried out for each Mudra image and is saved in the database along with the calculated features and meaning of the corresponding Mudra. The different features are area, major axis length, minor axis length, centroid and eccentricity of each mudra image [10]. The values of mudra features values are then compared with entries for each Mudra in the database and classification is done using k Nearest Neighbor algorithm. The entry giving the minimum difference for all these values is the match for the given

input image. Finally emotional description for the recognized mudra image is displayed.

2. Gesture Recognition Process

A gesture is a form of non-verbal action-based communication made with a part of the body, and used in combination with verbal communication' [1]. A gesture is categorized into two distinct categories: dynamic gestures and static gesture. A dynamic gesture is intended to change over a period of time (e.g. A waving hand means goodbye), whereas a static gesture is observed at the spurt of time (e.g. stop sign). The project considers static gestures. There are two basic approaches in static gesture recognition:

1. The top-down approach, where a previously created model of collected information about hand configurations is rendered to some feature in the image co-ordinates. Comparing the likelihood of the rendered image with the real gesture image is then used to decide whether the gesture of the real image corresponds to the rendered one.
2. The bottom-up approach, which extracts features from an input image and uses them to query images from a database, where the result is based on a similarity measurement of the database image features and the input features[9].

The project uses suitable amalgamation of the two approaches called as hybrid approach.

The whole process of static gesture recognition can be coarsely divided into four phases:

- Image capturing
- Pre-processing
- Feature extraction
- Classification

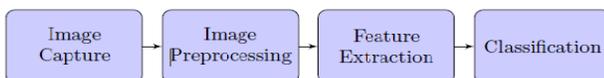


Fig. 1: Schematic view of gesture recognition process

2.1 Image Capturing

The task of this phase is to acquire an image, which is then processed in the next phases. The capturing is mostly done using a single camera with a frontal view of the person's hand, which performs the gestures. However, there also exist systems that use two or more cameras in order to acquire more information about the hand posture. The advantage of such a system is that it allows recognition of the gesture, even if the hand is occluded for

example by the body of the person that performs the gesture, since the other camera captures the scene from another perspective.

In general, the following phases of the recognition process are less complex if the captured images do not have cluttered backgrounds, although several recognition systems seem to work reliably even on cluttered images. Therefore, the image capturing is often performed in a cleaned up environment having a uniform background. It is also desirable to have an equalized distribution of luminosity in order to gather images without shadowy regions.

2.2 Pre-processing

The basic aim of this phase is to optimally prepare the image obtained from the previous phase in order to extract the features in the next phase. How an optimal result looks like depends mainly on the next step, since some approaches only need an approximate bounding box of the hand, whereas others need a properly segmented hand region in order to get the hand silhouette. In general, some regions of interest, that will be subject of further analysis in the next phase.

2.3 Feature extraction

The aim of this phase is to find and extract features that can be used to determine the meaning of a given gesture. Ideally such a feature, or a set of such features, should uniquely describe the gesture in order to achieve a reliable recognition. Therefore, different gestures should result in different, good discriminable features.

2.4 Classification

The classification represents the task of assigning a feature vector or a set of features to some predefined classes in order to recognize the hand gesture. In previous years several classification methods have been proposed and successfully tested in different recognition systems. In general, a class is defined as a set of reference features that were obtained during the training phase of the system or by manual feature extraction, using a set of training images. Therefore, the classification mainly consists of finding the best matching reference features for the features extracted in the previous phase [4].

3. The Proposed System

Mudra recognition is carried out using Image-processing and Pattern Recognition techniques. The Mudra image is captured, processed, Pattern recognized, decoded to Text

form and linked with the dance-sequence. The sequence of text of the background story thus generated by machine is compared with that interpreted by section of audience.

The proposed gesture recognition system consists of two major stages:

1. Training the system
2. Testing

Training is where the system database is created as shown in Fig. 2. In this we take images of the entire Bharatnatyam Mudras one at a time. We perform some initial pre-processing on the image which is simple filtering to remove noise if any. Then we apply the proposed hybrid saliency detection technique. The output of this stage highlight the object from background and get the salient features of the static double hand mudra image properties of the desired region of the Mudra in the image. These properties are area, major axis length, minor axis length and eccentricity. Now this process is carried out for each Mudra image and is saved in the database along with the calculated properties and meaning of the Mudra.

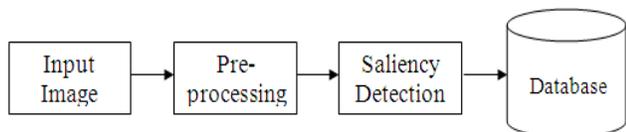


Fig. 2 Training the system

Fig. 3 below shows the block diagram for Mudra recognition and testing part. The input is any Mudra image, which is pre-processed and given to the saliency detection block of stage 1. The output of this block gives the area, major axis length, minor axis length and eccentricity of each mudra image. These values are then compared with entries for each Mudra in the database by using knn classifier. The entry giving the minimum difference for all these values is the match for the given input image [10].

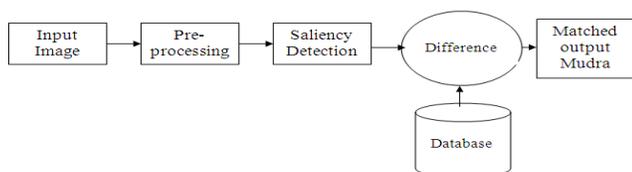


Fig. 3 Testing

3.1 Objectives

1. To apply image processing and pattern recognition techniques for static mudra recognition.
2. To interpret the emotions embedded in certain Mudra of Indian classical dance Bharatnatyam, using Gesture Recognition process.

3. The recognition of Mudra sequence can thus create language independent universal communication environment for dance drama.
4. Goal is to create a system which can identify specific human gestures and use them to convey information.

4. Implementation of Phases

Fig. 4 below shows the flowchart representation of the proposed system. The first step is to capture the images of static double hand gesture with the help of standard camera system. All the images are captured with the black background and are noise free. In second stage, Gaussian filter is applied to remove the noise if any. After this the saliency detection technique is used to extract the object from the background. In the next stage features are extracted and these features are carried to the next stage to create database of each image. In the last stage, images are classified and emotional description of the particular image is displayed.

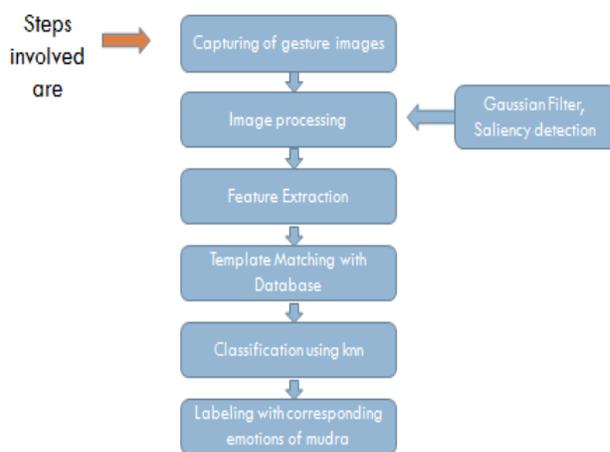


Fig.4. Flowchart for program execution

4.1 Selected Static Double Hand Gesture

In our system we have considered 13 types of static double hand mudra of the Indian classical dance known as Bharatnatyam, for recognition. Fig. 5 below shows 13 types of the images and their corresponding description used to training the system.

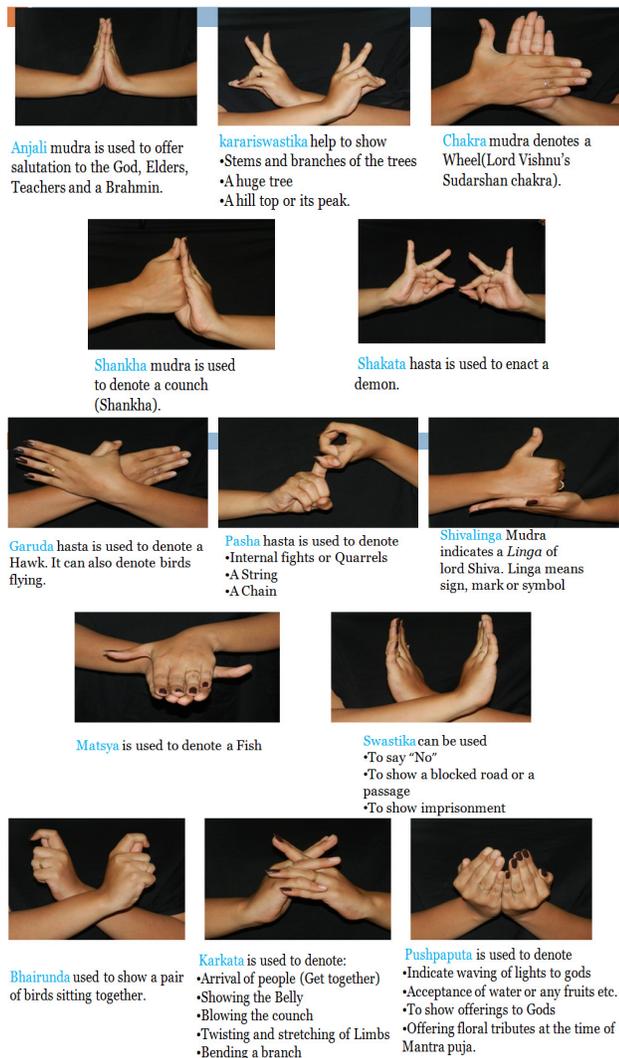


Fig. 5 Images used to train the system

4.2 Phases

The proposed system consists of 4 phases:

1. Object detection
2. Feature extraction
3. Database creation
4. Emotion recognition or classification

4.2.1 Object Detection

The first step towards object recognition is object detection. Object detection aims at extracting an object from its background before recognition. A Saliency detection technique, which aims at detecting the image regions that represent the scene, is used to detect the object. Given an image, human can detect salient regions from the image extraordinarily fast and reliable. The

salient regions in the image may contain foreground, parts of the background, interesting patterns and so on. Saliency detection is useful in many image-processing tasks including image segmentation and object recognition. It can be used in the pre-processing step to reduce the search space. One important application of saliency detection is image retargeting, in which we would like to keep the salient regions the same but remove pixels, which are not salient. Different from traditional image statistical models, this technique analyze the log spectrum of each image and obtain the spectral residual. Then the spectral residual is transformed to spatial domain to obtain the saliency map, which suggests the positions of proto-objects. Some initial pre-processing on the image is done which is simple filtering to remove noise if any, a Gaussian filter is used. Then saliency detection technique is applied. It uses the hypercomplex representation (i.e., Quaternion Fourier Transform) of the image.

4.2.2 Feature Extraction & Database Creation

Feature extraction is the process of transforming the input data into the set of features (called feature vector). Find and extract features that can be used to determine the meaning of a given gesture. These features describe the gesture in order to achieve a reliable recognition. Different features or properties considered for this system are area, major axis length, minor axis length, centroid and eccentricity. These features are calculated by using the feature vector with region props technique which measures properties of image regions. This process is carried out for each Mudra image and the relevant features are saved in the database along with the calculated properties and meaning of the each Mudra.

4.2.3 Emotion Recognition or Classification

Classification is the task of assigning a feature vector or a set of features to some predefined classes in order to recognize the hand gesture. Classification here depicts that identifying the type of the mudra in the input image & displaying the emotional description behind it. There are number of methods available for classification. In this project we have implemented k Nearest Neighbor algorithm for classification.knn is most often used for classification, although it can also be used for estimation and prediction. K-Nearest neighbour is an example of *instance-based learning*, in which the training image set is stored, so that classification for a new unclassified image may be found simply by comparing it to the most similar records in the training set. It takes all the features and computes the difference between each feature from the database. Here we have used KNN classifier for the classification task. Then the database entry which has

maximum features matching (means minimum difference), is selected as output, indicating the emotions corresponding to recognized image.

5. Experimental Results

5.1 Output of Object detection stage

Output for two images is shown below:



Fig 6(a) Input Swastika Mudra

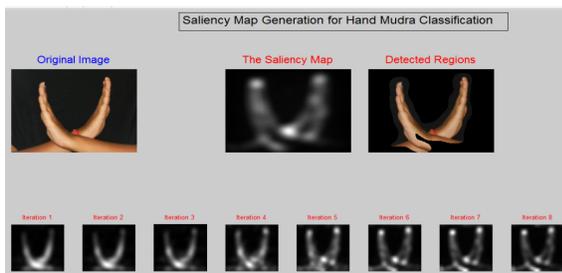


Fig 6(b) Saliency map for swastika mudra

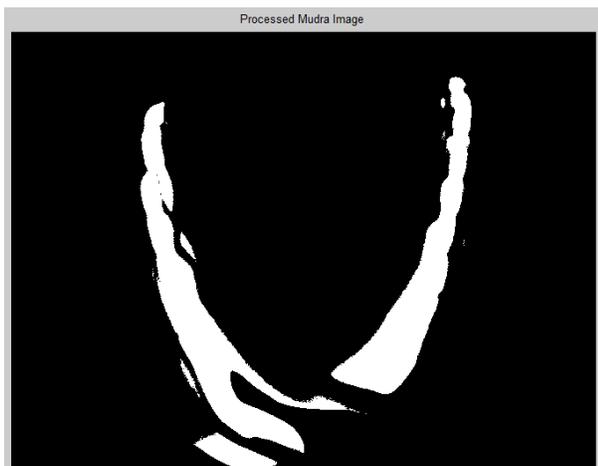


Fig. 6(c) processed mudra images for both the mudra

Fig. 6(a) shows the input images, Fig. 6(b) shows the graph for saliency map, Fig 6(c) shows the processed image which will be used in the next phase.

5.2 Output of Feature Extraction Stage

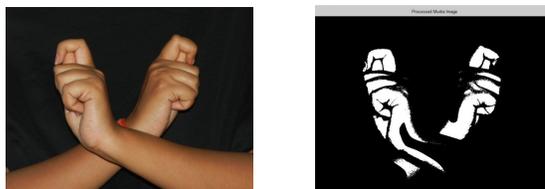


Fig.7.(a) Input image

7.(b)Processed mudra image



Fig.7. (c) Evaluated mudra features

Figure 7(a) shows the input mudra image, 7(b) shows the processed mudra from previous stage which is used in this stage for calculating the mudra features and 7(c) shows the actual mudra features calculated.

5.3 Output of Database Creation Stage

In this stage, features of each input mudra are stored in the database with the corresponding emotional description of it. This database is used in the next stage for classifying the input mudra image in the class to which it is related.

5.4 Output of Emotion Recognition stage



Fig8 (a) Input Anjali mudra

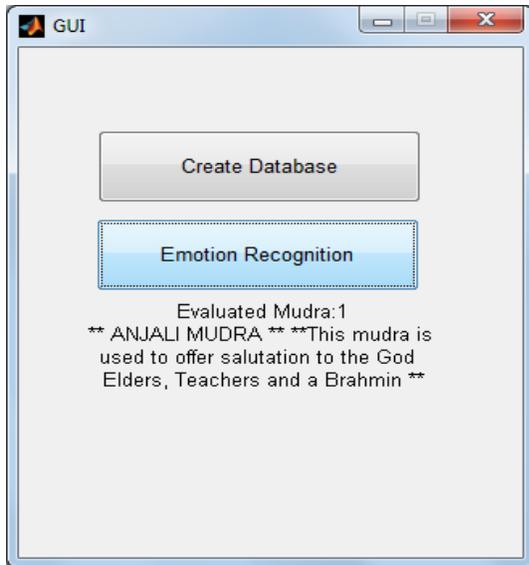


Fig.8 (b) Emotion description of Anjali mudra

Fig. 8(a) shows the input Anjali mudra image and Fig. 8(b) shows the emotional description of the Anjali mudra.

6. Result Analysis

We have collected more than 150 sample images, minimum 10 images for each mudra type. We classified these images in training set and testing set. The training set consists of 102 images and testing set contains the 68 images. Firstly, with the help of training set we trained the database for each type of selected mudra and after that with the help of testing set we tested the system accuracy. We have calculated the system accuracy with the formula of precision.

For classification tasks, the terms true positives, true negatives, false positives, and false negatives compare the results of the classifier under test with trusted external judgments. The terms *positive* and *negative* refer to the classifier's prediction (known as the expectation), and the expressions *true* and *false* states that whether the prediction corresponds to the external judgment (known as the *observation*). This is illustrated by the table shown below:

Predicted class (expectation)	Actual class (observation)	
	tp (true positive) Correctly classified	fp (false positive) Unexpected classification

Precision is then defined as:

$$\text{Precision} = \frac{tp}{tp + fp}$$

In our system, we have considered two values from above table for result analysis i.e. *tp* (true positive) and *fp* (false positive). In *tp*, images which are correctly classified are enlisted and in *fp*, misclassified images are considered. After applying the precision formula, the system accuracy is 85.29%. The below table shows the result analysis in a tabular form consisting of number of samples images available in the testing set, number of samples correctly classified, number of samples misclassified and the accuracy of the system.

Total no of samples available	No of samples correctly classified (tp)	No of samples misclassified(fp)	Accuracy of system in %
68	58	10	85.29%

7. Conclusion

In this project novel approach of application of gesture recognition to Bharatnatyam Mudras has been presented. The proposed method employs the use of Saliency detection technique to get the detected regions of the Mudra along with its features, which are then compared with the in-built database to match the input image with Mudra from the database using knn and determine its meaning. The proposed method works well with both the single hand Mudras and double hand Mudras. The proposed model can also be applied to other Indian classical dance forms and thus can be used to get a hybrid gesture recognition system for interpreting the symbols, postures and Mudras of various Indian classical dance forms. Such a system can then also be used teach and correct the young dancers.

8. Future Scope

The avenues for further work in this area point to use gesture recognition in the field of multi-touch gestures which are predefined motions used to interact with multi-touch devices, optical imaging and sixth sense which is a wearable gestural interface that augments the physical world around us with digital information and lets us use natural hand gestures to interact with that information.

Gesture Recognition can also be used during business meetings and in robotics also. Sixth Sense can be used with gesture recognition which is a wearable gestural interface that augments the physical world around us with digital information and lets us use natural hand gestures to interact with that information.

References

- [1] C. S. Warnekar, Chetana Gavankar, 'Algorithmic Analysis of Gesture Pattern', June 2008.
- [2] C. S. Warnekar, Deshpande and Chetana Gavankar, 'Mudra Recognition using Image-processing and Pattern Recognition approach'.
- [3] Shweta Mozarkar and C.S. Warnekar, ' Interpretation of Emotions of Certain Bharatnatyam Mudra using Gesture Recognition Process', International conference on cloud computing and computer science, pp 47-50.
- [4] Mitra and S Acharya, 'Gesture Recognition: A Survey', May 2007.
- [5] Thierry Messery, Department of Informatics, university of Fribourg, Switzerland., 'Static hand gesture recognition: Report'.
- [6] Prateem Chakraborty, Prashant Sarawgi, Ankit Mehrotra, Gaurav Agarwal, Ratika Pradhan., 'Hand Gesture Recognition: A Comparative Study'.
- [7] Xiaodi Hou and Liqing Zhang, Department of Computer Science, Shanghai Jiao Tong University, 'Saliency Detection: A Spectral Residual Approach'.
- [8] *Zheshen Wang, Baoxin Li*, Dept. of Computer Science & Engineering, Arizona State University, 'A two-stage approach to saliency detection in images'.
- [9] Henrik Birk and Thomas Baltzer Moeslund, 'Recognizing Gestures From the Hand Alphabet Using Principal Component Analysis', Master's Thesis, Laboratory of Image Analysis, Aalborg University, Denmark, 1996.
- [10] Jesús Angulo, "From Scalar-Valued Images to Hypercomplex Representations and Derived Total Orderings for Morphological Operators", ISMM 2009, LNCS 5720, pp. 238–249, 2009. © Springer-Verlag Berlin Heidelberg 2009.
- [11] T. Liu, J. Sun, N. Zheng, X. Tang and H. Shum, "Learning to Detect A Salient Object", CVPR, 2007.
- [12] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency", In Advances in Neural Information Processing Systems 19, pages 545–552. MIT Press, 2007.
- [13] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. PAMI, 20(11), 1998.
- [14] Andrew Wilson and Aaron Bobick, "Learning visual behaviour for gesture analysis," In Proceedings of the IEEE Symposium on Computer Vision, Coral Gables, Florida, pp. 19-21, November 1995.

First Author. Shweta Mozarkar has received her B.E.degree in Computer Science & Engineering from Anjuman College of Engineering Nagpur, RTMNU University in 2009. She is pursuing M.Tech in Computer Science and Engineering from Shri Ramdeobaba College of Engineering and Management (Autonomous), Nagpur. Her research interests include Image processing and Pattern recognition.

First Author. Dr. C. S. Warnekar . Sr. Professor in CSE @ JIT, Nagpur & Former Principal Cummins College, Pune.