Chapter 4

A Two-level Classification Method for Ground Exercises of Sattriya Dance

In this chapter, the SDGE dataset is validated with five states of the art classifiers from the diverse range viz., k-nearest neighbor, Bayesian network, decision tree, Support Vector Machine and Hidden Markov Model. Initially the features- height to width ratio of minimum bounding rectangle, inter-frame energy difference and inter-frame entropy difference are extracted. To improve the classification accuracy, a two-level classification method of ground exercises of Sattriya dance is presented.

The rest of the chapter is organized as follows: A brief description of the features used in this chapter is discussed in Section 4.1. Section 4.2 describes the theory of the state-of-the-art classifiers used in this work. In Section 4.3, dataset is validated using the mentioned classifiers. Section 4.4 presents the new method for classification of the dynamic gestures. In Section 4.5, experimental results are discussed. Finally, in Section 4.6, the summary of this chapter with the scope of future work is discussed.

4.1 Feature Extraction

Feature extraction is one of the most important steps in gesture recognition since it greatly affects the recognition. Feature Extraction aims to reduce the number of features in a dataset by creating new features from the existing ones (and then discarding the original features). These new reduced set of features should then be able to summarize most of the information contained in the original set of features. In this way, a summarized version of the original features can be created from a combination of the original set.

Human body is a highly articulated structure so it is an important issue to extract features that best describes the articulation. Various features used for dynamic gesture recognition are 2D silhouettes [104], 3D visual hull [29], body centroid, body orientation based on the centroid; velocity and acceleration of the subject etc. 2D silhouettes can be extracted from the frames of the video after background subtraction. It is found from the literature that body centroid, body orientation based on the centroid features works well for static gestures. Velocity and acceleration of the subject features perform well while considering a particular body movement. From experimental analysis we have found that the following features perform well in this work, namely height: width ratio of minimum bounding rectangle of each frame, Inter-frame energy difference and Inter-frame entropy difference.

4.1.1 Height: Width ratio of Minimum Bounding Rectangle

From the 2D silhouettes the MBR of each frame has been extracted. Height to width ratio of each frame is calculated which is considered as a feature.

This ratio can be defined as

$$r = \frac{h}{w} \tag{4.1}$$

where,

r is the ratio

h is the height of the MBR, and

w is the width of the MBR.

After finding out the MBR for a video, the height: width ratio is calculated for each frame. It is observed that this ratio is smaller for the group “Facing front” for almost all the image frames and for the ground exercises “Facing Around”, this ratio is found to be larger for the frames where the dancer is facing side since width becomes very small. From this observation we have classified these two groups of ground exercises.

4.1.2 Inter-frame Energy Difference

Inter-frame difference is normally used to determine the object movement. But it is not the best choice for extremely low luminance images. An alternative way can be the use of energy information of the images - the Inter-frame difference of energy. Energy is defined based on a normalized histogram of the image. Energy shows how the gray levels are distributed. When the number of gray levels is low then energy is high. This feature is extracted from the video frames.

Energy can be defined as the square root of Angular Second Moment (ASM). The uniformity of distribution of gray level in the image is known as ASM.

$$ASM = \sum_{m,n=0}^{N-1} p_{m,n}^2 \tag{4.2}$$

where $p_{m,n}$ is the probability of the color intensity at position (m, n). N is the gray level.
and

$$\text{Energy} = \sqrt{ASM} \tag{4.3}$$

4.1.3 Inter-frame Entropy Difference

Entropy is a statistical measure of randomness that can be used to characterize the texture of the input image [123]. It gives the randomness or uncertainty of information to represent an image. Inter-frame entropy difference feature is used in this work. Entropy can be calculated using equation 4.5,

$$H = -\sum_{k=1}^{K} p(e_k) log\, p(e_k) \tag{4.4}$$

Here $\{e_1, e_2, \dots \dots, e_k\}$ is the set of possible events of a random event E with probabilities $\{p(e_1), p(e_2), \dots \dots \dots p(e_k)\}$. K is the total number of events.

The three features- inter-frame energy difference (EN), inter-frame entropy difference (EP) and height to width ratio of MBR (HW) are extracted from each frame of the videos and get the feature vector for that image sequence. Each ground exercise video is represented by the computed feature vector,

Feature vector= $[EN, EP, HW]$

The features for all images are extracted and obtain the feature vectors for the entire training dataset and obtain the feature matrix.

The feature matrix, mathematically, can be defined as 2D matrix, where the row indicates features vector of each image sequence and column indicates the number of features. Using this feature matrix, mean feature vector for each person is calculated and this is used as a template for each ground exercises in the testing phase to verify the ground exercise in the recognition phase.

$$featur\text{e\_matrix} = \begin{bmatrix} EN_1 EP_1 \mathrm{HW}_1 \\ EN_2 EP_2 \mathrm{HW}_2 \\ \dots\dots\dots\dots\dots\dots\dots \\ EN_n\, EP_n \mathrm{HW}_\mathrm{n} \end{bmatrix}_N$$

Where EN_i is the energy difference of $(i+1)^{th}$and i^{th} frame, EP_i is the entropy difference of $(i+1)^{th}$and i^{th}frame, HW_i is the height to width ratio of minimum bounding rectangle of the i^{th} frame and N is the number of videos and n is the total number of frames of a video.

We have created four features set with various combinations of the three features- EP, EN and HW. The feature matrices are:

The feature matrix consisting of features $EP$and EN

$$feature\_matrix_1 = \begin{bmatrix} EN_1 EP_1 \\ EN_2 EP_2 \\ \dots\dots\dots\dots \\ EN_n\, EP_n \end{bmatrix}_N$$

The feature matrix consisting of features EP and HW

$$feature\_matrix_2 = \begin{bmatrix} EP_1 HW_1 \\ EP_2 HW_2 \\ \dots \dots . \dots \dots \\ EP_n \ HW_n \end{bmatrix}_N$$

The feature matrix consisting of features EN and HW

$$feature\_matrix_3 = \begin{bmatrix} E\mathrm{N}_1 HW_1 \\ EN_2 HW_2 \\ \dots \dots . \dots \dots \\ EN_n \, HW_n \end{bmatrix}_N$$

The feature matrix consisting of featuresEN, EP and HW

$$feature\_matrix_4 = \begin{bmatrix} EN_1 EP_1 HW_1 \\ EN_2 EP_2 HW_2 \\ \dots \dots . \dots \dots \dots \dots . \\ EN_i \ EP_i \, HW_i \end{bmatrix}_N$$

4.2 Classifiers

The classifiers used in the experiments are KNN, SVM, Bayesian network, Decision tree and HMM. Here different types of classifiers are chosen for classification. Brief introductions of the classifiers are presented in the following subsections.

4.2.1 KNN

K nearest neighbors is a supervised machine learning algorithm. A supervised machine learning algorithm's goal is to learn a function such that $f(M) = N$ where M is the input, and N is the output. KNN is a lazy learning algorithm and a non-parametric method. Lazy learning means that the algorithm takes almost zero time to learn because it only stores the data of the training part. The stored data will then be used for the evaluation of a new query point. The KNN classifier uses distance-based similarity measure to classify an unknown object to one of the known classes. In the training phase, the model will store the data points. In the testing phase, the distance from the query point to the points from the training phase is calculated to classify each point in the test dataset. Various distances can be calculated, but the most popular one is the Euclidean distance for smaller dimension data.

For small K value, the problem that arises is if an outlier is present in the data, the decision surface considers that as a data point. Due to this, KNN will perform exceptionally well on the training dataset but will misclassify many points on the test dataset (unseen data). This is considered as overfitting, and therefore, KNN is sensitive to outliers. As the value of K increases, the surface becomes smooth and will not consider the outliers as data points. This will better generalize the model on the test dataset also. If K value is extremely large, the model will under fit and will be unable to classify the new data point. For example, if K is equal to the total number of data points, no matter where the query point lies, the model will always classify the query point based on the majority class of the whole dataset.

In our experiment, we use $K = 5$ for feature set which gives low error rate on SDGE dataset. We use Euclidean distance as it gives good results compared to other distance. Time complexity and space complexity is very high, which is a major disadvantage of KNN.

4.2.2 Decision Tree

Decision tree builds classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. Leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data.

Decision trees are very fast at classifying unknown records. These are easy to interpret for small-sized trees. Large trees can be difficult to interpret and the decisions they make may seem counter intuitive which is a drawback of decision tree classifier.

4.2.3 Bayesian Network

Bayesian classification is based on Bayes' Theorem. Bayesian classifiers can predict class membership probabilities such as the probability that a given tuple belongs to a particular class. A Bayesian network also known as a Bayes network, belief network, or decision network is a probabilistic graphical model that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG). Bayesian networks are ideal for taking an event that occurred and predicting the likelihood that any one of several possible known causes is the contributing factor. For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases.

Doing full Bayesian learning is extremely computationally expensive. This even holds true when the network structure is already given. Furthermore, Bayesian networks tend to perform poorly on high dimensional data. Finally, Bayesian network models can be hard to interpret, and require Copula functions to separate out effects between different parts of the network.

4.2.4 Support Vector Machine

Support vector machines are a set of supervised learning methods used for classification, regression and outliers detection. The advantages of support vector machines are effective in high dimensional spaces. Still, it is effective in cases where number of dimensions is greater than the number of samples. SVM uses a subset of training points in the decision function (called support vectors), so it is also memory efficient. The disadvantages of support vector machines include if the number of features is much greater than the number of samples, avoid over-fitting in choosing Kernel functions and regularization term is crucial.

SVMs do not directly provide probability estimates, these are calculated using an expensive k-fold cross-validation. Classification in Machine Learning is the task of learning to distinguish points that belong to two or more categories in a dataset. In geometrical terms, associating a set of points to some category involves finding the best possible separation between these.

4.2.5 Hidden Markov Model

HMM is a statistical model which is widely used in dynamic gesture recognition because of its ability to handle spatio-temporal identity of gesture [6]. HMM is advantageous for dynamic gesture recognition because it is analogous to human performance. The HMM is a doubly stochastic process that involves a hidden immeasurable human mental state and a measurable, observable human action [13]. It has been proven that HMMs are effective in sign language recognition and other complex hand gesture recognition processes [61]. Besides, HMM performs well in [45, 48, 61] for recognizing full body gestures. HMM is capable of modeling spatiotemporal time series of gestures effectively and can handle non-gesture patterns [2]. There are three major issues in HMM: evaluation, estimation, and decoding. These problems are solved by using Forward algorithm, Viterbi algorithm, and Baum-Welch algorithm, respectively [13].

4.3 Dataset Validation

The SDGE dataset is validated with five benchmarking machine learning classifiers, namely KNN, SVM, Decision tree, Bayesian Network and HMM. The Classification rate with these classifiers is shown in Table 4.1.

Table 4.1: Classification rate with different classifiers

Classifier	Total Classes	Correctly Classified	Classification Rate
KNN	28	19	67.85
SVM		20	71.4
Decision Tree		19	67.85
Bayesian Network		19	67.85
HMM		22	78.57

The classification rates are not satisfactory. To improve the classification rates, we have proposed the two-level classification method.

4.4 The Proposed Method

The two-level classification system is proposed for classification of ground exercises of Sattriya dance. This method classifies all the 28 ground exercises into two groups and then classifies the individual ground exercises of each group.

The total 28 ground exercises are classified into two groups in level 1. These are:

- Group 1: The ground exercises in which only front movement of dancer occurs, fall into Group 1, named as Facing Front (FF) group.
- Group 2: The ground exercises, in which the dancer dances in all the directions, fall into Group 2, named as Facing Around (FA) group.

In first level of classification, an unknown ground exercise of Sattriya dance is classified into one of the two groups using a method based on HW ratio of MBRs. The method for level 1 classification is briefly described here. Initially we have taken one image sequence for each type of ground exercise. Then 2d silhouettes are extracted for each image sequence. The HW ratio of the MBRs of each frame is found out. Key frames are extracted using a simple algorithm explained in chapter 3. Heights of the MBRs are almost same for all ground exercises. For the Group 1 (FF), the HW ratio is smaller for almost all the image frames since width is larger. For the Group 2 (FA), this ratio is found to be larger for the frames where the dancer is facing side since width becomes very smaller. Using this feature 19 ground exercises belong to group 1 and 9 ground exercises belong to group 2. At this level, 100% accuracy is achieved.

In the second level of classification, the considered feature vector is $[EN, EP, HW]$. These values of feature vector are then compared with those in the database. The group with which the input image frames match the most is returned as the output of this step. For classification of each gesture of these groups we have chosen five state of the art classifiers - KNN, SVM, Bayesian network, Decision tree and HMM.

4.5 Experimental Results

The experimental description and the obtained outcome for the proposed method on Sattriya dance ground exercise dataset is discussed in the following sub-sections.

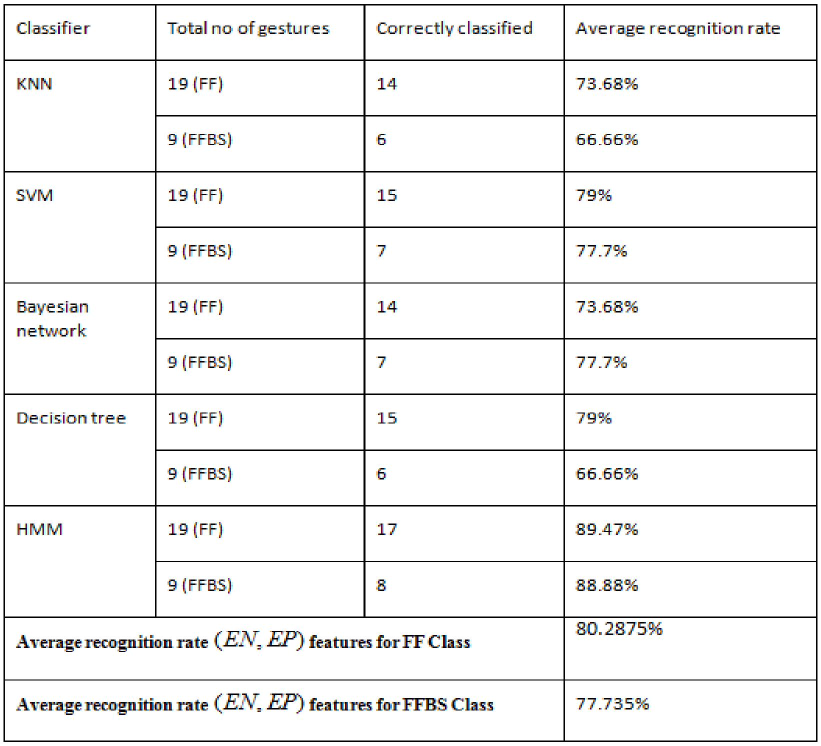
4.5.1 Dataset Description

In this experiment, the feature dataset containing the three features- EN, EP and HW is used from the SDGE dataset.

4.5.2 Results and Discussion

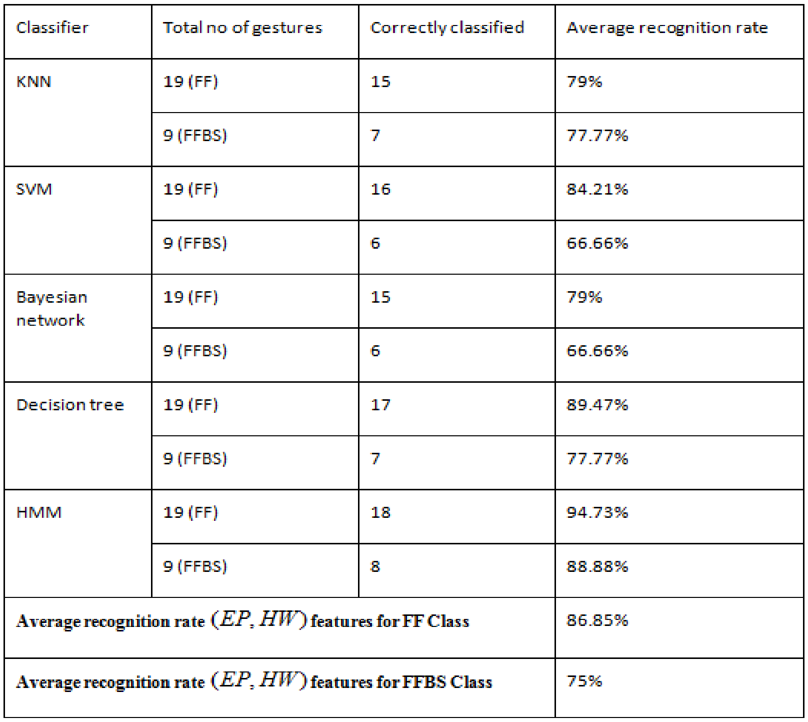
The average classification accuracy of the different ground exercises that are achieved are shown in Table 4.2 – Table 4.5 consecutively according to the four feature sets. A comparative analysis is shown among the five states of the art classifiers for classification of the ground exercises of Sattriya dance since.

Table 4.2: Accuracy of EN and EP feature



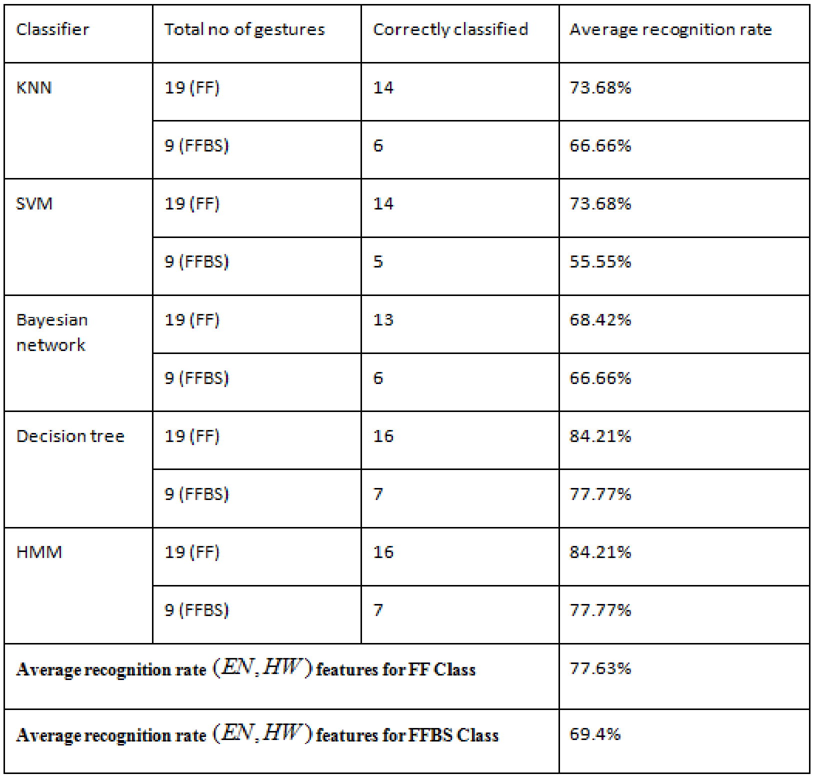
Classifier	Total no of gestures	Correctly classified	Average recognition rate
KNN	19 (FF)	14	73.68%
	9 (FFBS)	6	66.66%
SVM	19 (FF)	15	79%
	9 (FFBS)	7	77.7%
Bayesian network	19 (FF)	14	73.68%
	9 (FFBS)	7	77.7%
Decision tree	19 (FF)	15	79%
	9 (FFBS)	6	66.66%
HMM	19 (FF)	17	89.47%
	9 (FFBS)	8	88.88%
Average recognition rate (EN, EP) **features for FF Class**			80.2875%
Average recognition rate (EN, EP) **features for FFBS Class**			77.735%

Table 4.3: Accuracy of EN and HW feature



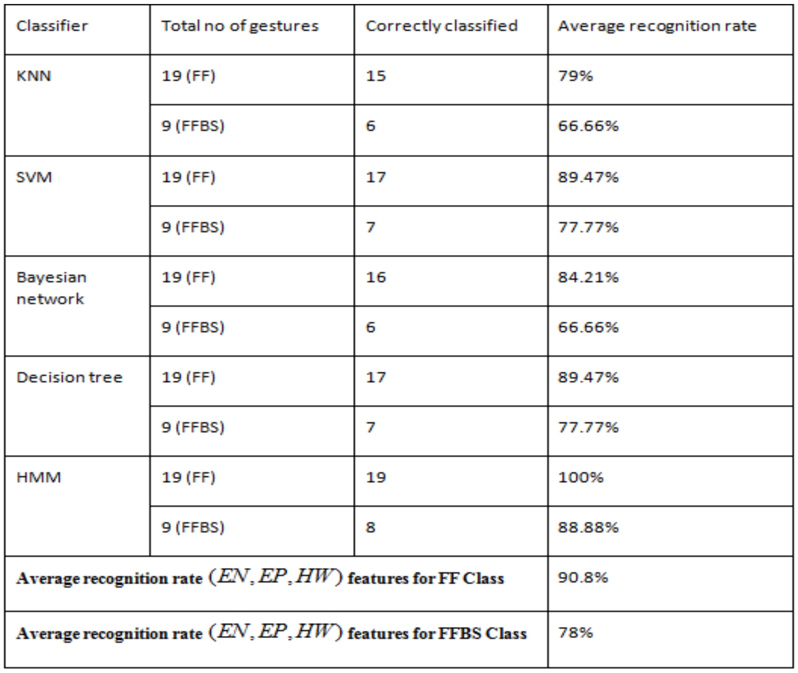
Classifier	Total no of gestures	Correctly classified	Average recognition rate
KNN	19 (FF)	15	79%
	9 (FFBS)	7	77.77%
SVM	19 (FF)	16	84.21%
	9 (FFBS)	6	66.66%
Bayesian network	19 (FF)	15	79%
	9 (FFBS)	6	66.66%
Decision tree	19 (FF)	17	89.47%
	9 (FFBS)	7	77.77%
HMM	19 (FF)	18	94.73%
	9 (FFBS)	8	88.88%
Average recognition rate (EP, HW) **features for FF Class**			86.85%
Average recognition rate (EP, HW) **features for FFBS Class**			75%

Table 4.4: Accuracy of EP and HW feature



Classifier	Total no of gestures	Correctly classified	Average recognition rate
KNN	19 (FF)	14	73.68%
	9 (FFBS)	6	66.66%
SVM	19 (FF)	14	73.68%
	9 (FFBS)	5	55.55%
Bayesian network	19 (FF)	13	68.42%
	9 (FFBS)	6	66.66%
Decision tree	19 (FF)	16	84.21%
	9 (FFBS)	7	77.77%
HMM	19 (FF)	16	84.21%
	9 (FFBS)	7	77.77%
Average recognition rate (EN, HW) **features for FF Class**			77.63%
Average recognition rate (EN, HW) **features for FFBS Class**			69.4%

Table 4.5: Accuracy of EN, EP and HW feature



Classifier	Total no of gestures	Correctly classified	Average recognition rate
KNN	19 (FF)	15	79%
	9 (FFBS)	6	66.66%
SVM	19 (FF)	17	89.47%
	9 (FFBS)	7	77.77%
Bayesian network	19 (FF)	16	84.21%
	9 (FFBS)	6	66.66%
Decision tree	19 (FF)	17	89.47%
	9 (FFBS)	7	77.77%
HMM	19 (FF)	19	100%
	9 (FFBS)	8	88.88%
Average recognition rate (EN, EP, HW) **features for FF Class**			90.8%
Average recognition rate (EN, EP, HW) **features for FFBS Class**			78%

4.6 Summary

In this chapter, for classification of ground exercises of Sattriya dance a two-level classification method is proposed. At the first level, an unknown ground exercise sequence is classified into one of the two groups using a method based on HW ratio of the MBRs. At the first level, 100% accuracy is achieved. At the next level, five state of the art classifiers are chosen to classify the ground exercises of each group. At this level, the accuracy obtained for Group 1 is 90.8% and for Group 2 is 78% using the (EN, EP, HW) feature set. The average accuracy for second level classification is 86.68%. The ground exercises in Group 2 involve the movement of the dancer in all the direction. To catch up the right gesture of Group 2 might be difficult using this method of classification. For better classification accuracy ensemble classifiers are used which are presented in the next Chapter.