Chapter 1

Introduction

1.1 Preface

Gesture recognition is one of the important research areas in computer science and technology. Human gestures can be interpreted using mathematical algorithms. Gestures can begin from any real movement or state however generally originates from face or hand. Users can utilize simple motions to control or associate with gadgets without genuinely contacting them. Many computer vision algorithms have been used for gesture recognition including sign languages, posture, gait, and human behaviours. Gesture recognition can be seen as a way for computers to begin to comprehend human non-verbal communication. Subsequently it fabricates a more extravagant scaffold among machines and user interfaces- text or even graphical.

With the advancement of gesture recognition technology, users can communicate with machine naturally. The innovation of gesture recognition technology has been viewed as the exceptionally fruitful innovation as it saves time to open any gadget. An individual does not have to recall the security codes like passwords for its gadget with the assistance of gesture recognition. The individual can open its gadget with next to no work for the most part hand gestures are pertinent. Gesture recognition can be conducted with various techniques including computer vision, machine learning, and image processing.

Gesture recognition is a challenging research area in the present interactive and intelligent computing world. Today, gesture recognition impacts the world diversely, from physically challenged people to robot control to virtual reality environments. Gesture recognition is the manner by which the articulation or motion made by the user is perceived by the receiver [49]. The approaches of gesture recognition can be broadly categorized into two groups: Glove-based and Vision-based approaches. The systems that use extra devices (gloves, sensors) are more intricate contrasted with vision-based systems. Vision-based systems are not difficult to use yet most challenging to implement. The area of vision-based gesture recognition leads to a large body of literature describing different body parts for gesture recognition, including static and dynamic gestures. Recognition of dynamic gestures using vision-based approaches is a confounded task.

1.2 Gesture

The meaningful and expressive body motions are called gestures. It tends to be actual movements of various parts of human body for communicating information or interacting with the environment. Gesture is the only way of communication for persons with hearing disability. In dance, gestures play an important role to convey the meaning of the dance to the audience.

The area of vision-based gesture recognition leads to a large body of literature describing different body parts for gesture recognition. In the field of dance gesture recognition, most of the works are centered round hand gestures. It is to be noted that the present work focuses on full body gesture recognition.

1.2.1 Classification of Gestures

Gestures are classified differently by different authors [49, 73, 76] based on different criteria. The different classification of gestures is depicted in Figure 1.1.

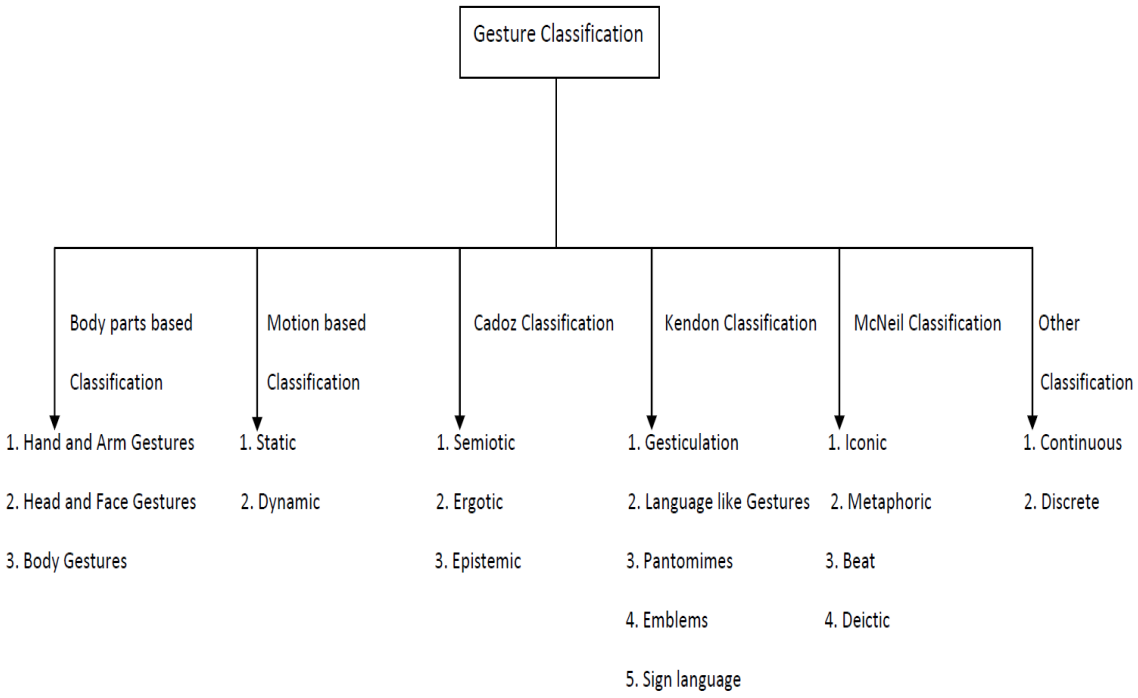

Figure 1.1 Gesture Classifications

1.2.1.1 Classification based on Body Parts

Based on body parts gestures can be classified into the following groups, as shown in Table 1.1 [73].

1. **Hand and arm gestures:** Movement of hand and arm of human body defines the hand and arm gestures. Hand gestures are generally used to recognize hand poses and sign languages. Also, hand and arm gestures are used for interactive game in virtual environments as entertainment applications for kids.
2. **Head and face gestures:** Head shaking, course of eye stare, causing a commotion, a person opens the mouth for talking, winking, erupting the nostrils, looks of shock, joy, disdain, dread, outrage, bitterness, and hatred are different sort of head and face gestures. Head gestures are used in different applications. In Kazakh traditional dance gesture recognition [87], head gestures are used. Recently, head gesture-controlled wheelchair is developed for Quadriplegic Patients [68]. There has been a lot of exploration on computer vision-based face recognition in the last two decades.

3. **Body gestures:** Full body movement is involved in body gestures. For example, movements of interaction of two individuals outside, human gait recognition, and dance step analysis along with music.

Table 1.1 Types of gestures [73]

Gesture Type	**Example**
Hand and Arm Gesture	Hand pose recognition
Head and Face Gestures	Head shaking, course of eye stare, causing a commotion, looks of shock, joy, and bitterness.
Body Gestures	Analysing body movements.

1.2.1.2 Motion based Classification

Based on motion gestures can be classified into two types: Static gestures and Dynamic gestures [73].

1. **Static Gestures**

Static gestures, also known as posture is a static command which is still for a tiny period of time. Posture is simple and needs less computation [39] compare to gestures that are complex and suitable for real-time environments [40, 73]. Some static hand gestures and body gestures are shown in Figure 1.2 and Figure 1.3.

2. **Dynamic Gesture**

Dynamic gesture, simply gesture is a sequence of static gestures. Gestures are dynamic over a period of time, i.e., it considers the moving expressions (e.g. hand waving in upward and downward direction). Sign languages are combination of static and dynamic gestures. A gesture comprises of three phases of motion [55]: preparation, stroke and retraction. The stroke is the salient gestural movement. The preparation phase orients the body parts for the stroke and the retraction phase returns it to rest or orients for the next gestural stroke. The

stroke is differentiable from the other two phases in velocity and acceleration. An example of dynamic gesture is shown in Figure 1.4.

Spatio-temporal variability is the most addressable issue of dynamic gesture recognition since velocity, duration, and shape features can differ for the same gesture. Thus, dynamic gesture recognition is a challenging task than static gesture recognition [73].

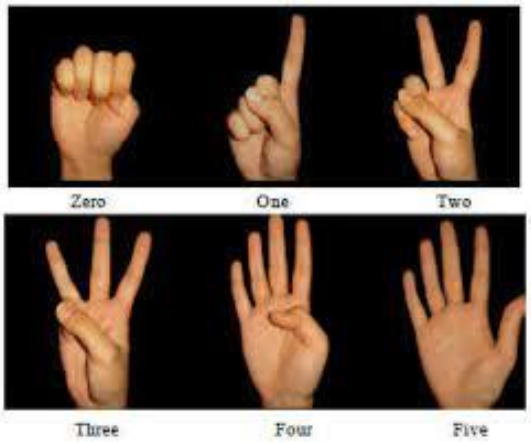

Figure 1.2: Static Hand Gesture

Figure 1.3: Static Body Gesture

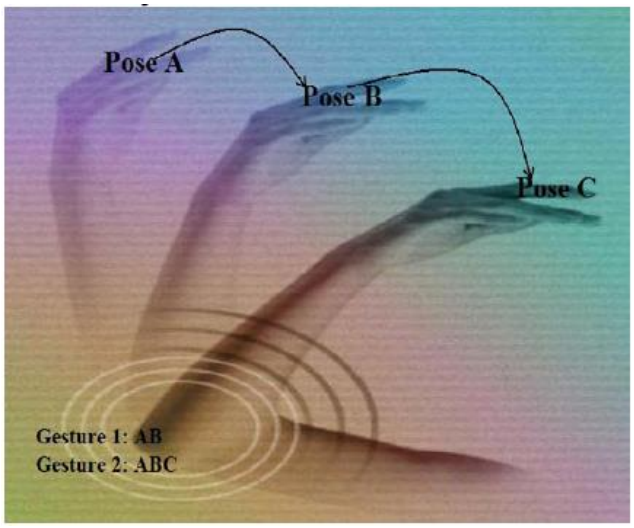

Figure 1.4: Dynamic Gesture

1.2.1.3 Cadoz Classification

Three functional roles of human gestures are described by Cadoz [18]. The different types of gestures according to Cadoz are described below.

1. Semiotic

Semiotic gestures communicate meaningful information. Semiotic is basically an investigation of creation of significant information. It is the scholastic investigation of hidden meaning in signs and images. It is an unconscious way of interpreting the world and the culture that we live. Semiotic gesture has a massive impact on all of us in day today life. Human is a semiotician, because everyone is constantly unconsciously interpreting the meaning of signs around them – from traffic signals to shades of banners, the different structures of vehicles, the engineering structures, and the package designing. The main purpose of semiotic gesture is to convey significant message around the environment. The semiotic gesture structure is typical and basically results from shared cultural experience.

2. Ergotic

The purpose of ergotic gesture is to manipulate the environment. The ergotic gesture is related with the idea of work. These gestures can be the manipulated directly from the real world or the environment.

3. Epistemic

Human can experience knowledge from the environment which is the function of epistemic gesture. We can assume the material of a structure by touching the object with our hand.

Some examples of different types of gestures according to Cadoz classification are presented in Table 1.2.

Table 1.2: Cadoz Classification

Gesture Type	Example
Semiotic	Various signs of different sign languages are some semiotic gestures.
Ergotic	Clay shaping, dust wiping.
Epistemic	By moving the hand over an object, one can tell the material of the object that is made of.

1.2.1.4 Kendon Classification

According to Kendon [113], there are five types of gestures:

1. **Gesticulation:** The impulsive movements of the hands and arms that accompany speech are called gesticulation.
2. **Language like gestures:** Language like gestures is some spontaneous movement of hands and arms during verbal communication. It is a kind of gesticulation that is used to put back a specific word or clause.

3. **Pantomimes:** The gestures that express any object with action without any word are known as pantomimes.
4. **Emblems:** Emblems are some culturally specific gestures. The usual emblems can be victory, thumbs up gestures.
5. **Sign languages**: Sign languages are some standard meanings of gestures of a language. American Sign Language, Japanese Sign Languages are example of sign languages.

Some examples of gestures according to Kendon are given in Table 1.3.

Table 1.3: Kendon Classification

Gesture Type	Example
Gesticulation	The good-bye gesture, the American sign language, the operational gestures used to guide airplanes on the ground, and even the vulgar "finger", each illustrates the semiotic function of gesture
Language like gestures	Shaping pottery from clay, wiping dust, etc. result from ergotic gestures.
Pantomimes	An actor is knocking into the air, pretending to be knocking on a door.
Emblems	"V" for victory, thumbs up.
Sign languages	American Sign Language (ASL)

1.2.1.5 McNeil Classification

According to McNeill [71], gestures are of four types:

1. **Iconic:** Iconic gestures are symbolic gestures that describe some features of the object, operation, or incident.

2. **Metaphoric:** The metaphoric gestures can represent a common symbol, rather than the object or event directly.
3. **Beat:** Beats are tiny gestures without any specific form. These are often associated to give emphasis on word.
4. **Deictic:** Deictic gestures are used to point to people, objects, or events in space or time.

Some examples of different types of gestures according to McNeil classification is presented in Table 1.4.

Table 1.4: McNeil Classification

Gesture type	Example
Iconic	Gesture made by a person with thumb and his little finger showing thumb up and little finger down to mean "Call you later."
Metaphoric	Holding or weighting a large object suggests the importance of an idea.
Beat	Up and down or back and forth hand movements that comes with spoken language.
Deictic	Pointing gestures, such as index finger pointing.

1.2.1.6 Continuous and discrete

Gestures can be classified as continuous and discrete gestures especially for multi-touch gestures.

1. **Continuous:** The continuous gestures must be known by the user in advance.

2. **Discrete:** The discrete gestures are those that are used to convey meaning with provided additional information.

Some examples of continuous and discrete gestures are given in Table 1.5.

Table 1.5: Continuous and Discrete Gesture

Gesture type	Example
Continuous	Pinching and flicking gesture.
Discrete	Tapping on a button of smartphone.

Full-body motion is needed to recognize large-scale gestures and to recognize or analyze human activity [40, 80]. Activity may be defined over a much longer period of time than what is normally considered a gesture; for example, jogging, walking is recognizable activity.

Bobick [15] proposed the following category of motion:

• Movement- the atomic elements of motion.

• Activity- a sequence of movements or static configurations.

• Action- high-level description of what is happening in context.

1.3 Dance Gesture Recognition

The diversely influenced areas of gesture recognition are Sign language recognition [110], robot control [41, 69, 91], virtual environments [48], games [56, 122], television control [66], dance [87, 100] etc. One of the special applications of gesture recognition is dance gesture recognition. Recognition of dynamic dance gestures is a demanding research area in computer vision. Dance gesture recognition has applications in performance-evaluation of dances [49], e-learning [53], and dance form recognition [14] and so on.

Many approaches have been proposed to recognize static gestures of different Indian classical dance forms, specifically hand gestures of Sattriya dance [35], Kathakali hand gestures [12], and Bharatnatyam double-handed gestures [81]. But very less works on

dynamic dance gesture recognition on Indian classical dance forms are reported in the literature. Some significant works on Indian classical dance forms are [11, 12, 14, 81, 100]. Sattriya dance is an important ICD form of Assam. It is a classical dance of India since the 15^{th} century by the great saint and reformer of Assam, Mahapurusha Sankaradeva [101]. Sattriya dance is originated in the eastern state of Assam and has remained a living tradition since its creation. It is one of the eight principal Indian classical dances. Ground exercises are the basic requirement to learn Sattriya dance, and it is vital to preserve and archive digitally these which lead to the overall preservation of Indian cultural heritage. Ground exercises are dynamic gestures of Sattriya dance. However, classification of ground exercises of Sattriya dance is not addressed yet.

1.4 Motivation

Many approaches have been proposed to recognize static gestures of various classical dance forms of India. But very less works on recognizing dynamic dance gestures of Indian classical dance forms is reported in the literature. Sattriya dance is a 500-year-old living dance tradition of Assam. Since ground exercises are the basic dance unit to learn Sattriya dance, there is scope to research in this dance form. Specifically ground exercises are the dynamic gestures of this dance. The primary motivation of this research work is classification of ground exercises of Sattriya dance that is not addressed in the literature yet. The principal objective of this research is to develop a vision-based method for classification of ground exercises of Sattriya dance. Though this work mainly focuses on recognizing ground exercises of Sattriya dance, it can also be applicable to recognize other dance forms. Some of the ground exercises are similar to certain yogic postures or asanas based on yoga Shastra. So, it will also be helpful for the recognition of yogic postures. It is also observed that there is no publicly available dataset of Ground exercises of Sattriya dance, and the creation of such a dataset will be a great help to the research community in pattern recognition and related areas. This is the prime motivation behind this work. It is worth noting that no significant research work has been discernible for recognition of ground exercises of Sattriya dance. It is the lack of such a method that has motivated the present research work.

Dynamic dance gesture recognition has numerous applications, such as analyzing dance gesture [79], performance evaluation of dance [49], dance form recognition [14], learning dance from online videos [19], etc. Moreover, dynamic gesture recognition based on computer vision is more convenient than glove-based systems, yet challenging. The human body is highly flexible and there is a need to focus on extracting the features that best describes the articulation. Using the best feature set, it is aimed to enhance the phase of recognition.

These problems and their wide range of applications motivate this work towards aiming to develop an effective vision-based dynamic gesture recognition system.

1.5 Objectives

This work aims at developing methods for classification of the dynamic gestures of Sattriya dance. Classification of 28 commonly used ground exercises of Sattriya dance is the main focus of this work.

It proposes to achieve the following objectives:

- To develop a video dataset of ground exercises of Sattriya dance.
- Extraction of features to distinguish between ground exercises of Sattriya dance.
- To develop a classification method for ground exercises of Sattriya dance using machine learning approach.

1.6 Conceptual Framework

Figure 1.5 shows the conceptual framework of the proposed system. The different phases of the proposed framework are described below.

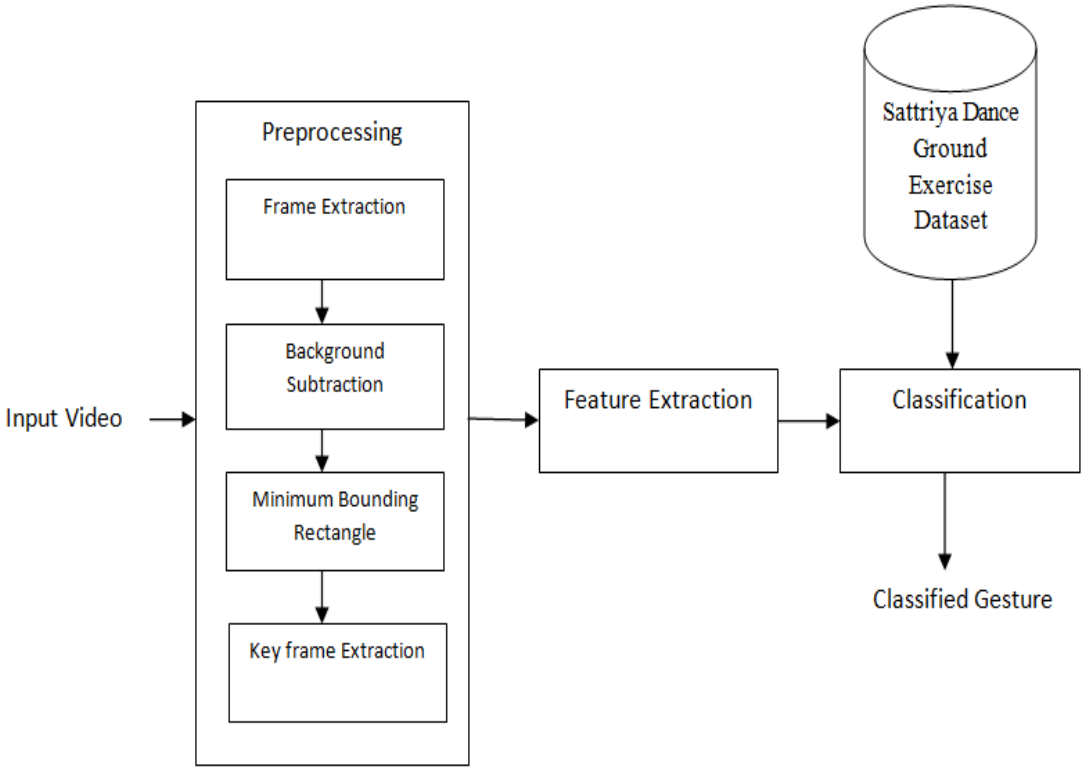

Figure 1.5: Conceptual framework of the proposed system

1.6.1 Preprocessing

The inputs to the system are the video of unknown ground exercise of Sattriya dance. The first and foremost phase of any gesture recognition system is data preprocessing. Preprocessing plays a vital role in any pattern recognition task. Data preprocessing plays a vital role in the knowledge discovery from massive operational data. It includes some techniques for enhancing raw data quality, including noise removal, scaling, data reduction, transformation, and missing value imputation.

From each frame of a video, the Minimum Bounding Rectangle (MBR) has been carried out. Then key frame extraction is performed. A key frame extraction algorithm is proposed to extract the key frames.

1.6.2 Feature extraction

Recognition is highly influenced by feature extraction. Features play a vital role in any recognition problem. Here the features are extracted from the image frames that are suitable for classification.

1.6.3 Classification

Classification is the final phase of gesture recognition. Here we have proposed three classification methods-

Two-Level Classification: In first level of classification, an unknown ground exercise of Sattriya dance is classified into two groups using an algorithm based on height:width ratio of minimum bounding rectangles (MBR). At this level, 100% accuracy is achieved. At second level, the ground exercise is individually recognized from the database within the group.

Classification using Ensemble methods: To combine the output of simple classifiers and make the final decision, the well-known ensemble learning methods- Random Forest, AdaBoost, Gradient Boost and XGBoost have been chosen.

Classification using Deep learning approach: For better performance of the system, we have gone through deep learning approaches. Here the VGG16 architecture with Imagenet is selected. It has 2 additional dense layers. The activation function is the Softmax for our multi-class classification problem. In this work we have 28 classes. The dataset that is used for training is our own created dataset.

1.6.4 Dataset

The Sattriya dance ground exercise dataset consists of 560 videos of 28 ground exercises. There is also a feature dataset containing three features namely- HW ratio of MBRs, inter-frame energy difference, and inter-frame entropy difference feature.

1.7 Dissertation Contribution

The notable contributions of the dissertation are presented in following subsections. Here a brief outline of the significant contributions of the dissertation is given:

1.7.1 Creation of Sattriya Dance Ground Exercise (SDGE) Video Dataset

In this dataset, 28 commonly used ground exercises have been considered. Here, ground exercise videos are recorded from 20 individuals. The SDGE dataset consists of 560 (28X20) videos. The videos have been recorded using a digital camera keeping at a fixed position. The motion of the dancers is captured under uniform background. Also, there should be a fixed distance between the camera and dancers. After capturing the videos these are saved in uncompressed AVI file format. The frames are extracted from each video and stored in separate folder. Then background subtraction has been performed to obtain the 2D silhouettes. Minimum bounding rectangle (MBR)s are obtained from the 2D silhouettes.

1.7.2 Developing a classification method for ground exercises of Sattriya dance using a machine learning approach

In this part of the work, empirical analysis has been conducted to evaluate the performance of our developed dataset with three features- HW ratio of MBR, inter-frame entropy difference, and inter-frame energy difference. We have selected five popular machine learning classifiers that are: K- Nearest Neighbour (KNN), Support Vector Machine (SVM), Bayesian Network, Decision tree, and Hidden Markov Model (HMM) to classify ground exercises.

Here a two-level classification method for recognizing ground exercises of Sattriya dance on the extracted features dataset is proposed. In the first level, twenty-eight classes of ground exercises are categorized into two groups based on the height to width ratio (HW) of MBRs. In the second phase, the considered feature vector is inter-frame energy difference, inter-frame entropy difference and HW i.e. [EN, EP, HW]. These values of the feature vectors

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. From the experimental results, it has been observed that HMM gives better classification accuracy.

1.7.3 Classification of Ground Exercises of Sattriya Dance using ensemble classifier

For classification of ground exercises of Sattriya dance, a new method using ensemble classifier is proposed. The proposed method uses the results of the simple classifiers- SVM, K-NN, Decision tree, Bayesian Network and HMM. To combine the output of classifiers and make final decision, the well-known ensemble learning methods- Random Forest, AdaBoost, Gradient Boost and XGBoost have been chosen.

1.7.4 Sattriya Dance Gesture Classification from Video with Convolutional Neural Network

Convolutional Neural Network (CNN) is very popular deep learning architectures for learning features automatically without any human supervision. These are used in different domains including computer vision, image processing, natural language processing, and pattern recognition. Deep learning is more effective when the dataset is very large. For smaller dataset it may cause over-fitting problem. Transfer learning can be used to handle this problem. In this work, we achieve the classification rate of 95.52%. This strategy can be added as a step for better implementation of e-learning techniques for the Indian classical dance, Sattriya dance.

1.8 Dissertation Organization

The organization of the dissertation is as follows:

Chapter 1: Introduction

It is an introductory chapter that gives a brief introduction of gesture recognition, some terms related with gesture recognition, basic framework of the work, explaining the motivation, objectives and summarizes the main contributions of the thesis.

Chapter 2: Literature Survey

In this chapter, an extensive literature survey on dynamic gesture recognition, applications, and existing methods of dynamic gestures recognition using machine learning to deep learning approaches with special emphasis on dance gesture recognition is presented.

Chapter 3: Video Dataset creation of Ground Exercises of Sattriya dance

This chapter presents the development procedures of Sattriya Dance Ground Exercise dataset and how the videos of these dataset are organized. Also, an empirical analysis of 3 types of features on the developed Sattriya dance ground exercise video dataset with five bench mark classifiers has been presented.

Chapter 4: A Classification method for ground exercises of Sattriya Dance using machine learning approach

In this chapter a classification method for ground exercises of Sattriya dance using machine learning approaches has been developed. A two-level classification method is developed for better performance.

Chapter 5: An Ensemble Learning based method for Classification of Ground Exercises of Sattriya Dance

In this chapter, ground exercises are classified using four well known ensemble learning based classifier. A comparative analysis is given among the classifiers.

Chapter 6: Sattriya Dance Gesture Classification from Video using Convolutional Neural Network

In this chapter, ground exercises are classified using Convolutional Neural Network.

Chapter 7: Conclusion and Future Direction

This chapter concludes the dissertation by summarizing the works done and also listing the possible further research in this area.