Chapter 2

Literature Review

In the area of computer vision, gesture recognition plays an important role that includes dance gesture recognition. The emerging growth of technology has contributed a lot to this domain. However, recognizing dynamic gestures using vision-based approaches is a very challenging task, but this approach is more convenient to the user.

The gesture is the most primitive way of communication among human beings. The meaningful expressions made by human body parts for communication with others are called gestures. The identification of significant expression of human motion is called gesture recognition [73]. Today in modern technology, gesture recognition influences the world very diversely, from physically challenged people to robot control to virtual reality environments.

This chapter presents the state-of-the-art regarding different approaches used for dynamic gesture recognition in dance forms.

The organization of this chapter is as follows. Section 2.1 illustrates the different application areas of gesture recognition. Section 2.2 presents a brief description of the approaches of dynamic gesture recognition. Section 2.3 describes the related work on

dynamic gesture recognition. A review of literature for dance gesture recognition is presented in Section 2.4. Section 2.5 gives a brief description of various video datasets. Section 2.6 concludes this chapter.

2.1 Applications of Dynamic Gesture Recognition

Today, dynamic gestures are used everywhere in the real world. Dynamic gesture recognition has wide range of applications in today's world. Sign languages can be recognized using different gesture recognition techniques [2, 27, 28, 30, 37, 52]. Gesture recognition is also a tool for entertainment that includes computer games [3, 56, 122], music systems [26, 99], and televisions [66]. The field of medical science is enhanced with the help of gesture recognition technology [40, 41, 42, 43, 44]. Today, this advanced technology becomes a part of the education system [3] also. Gesture recognition is widely used in intelligent driving [34], where gestures are used to control the car. Dance gesture recognition [49, 58, 87], which is a particular application of gesture recognition technology, has many applications recently. Gesture recognition technology becomes part and parcel in almost all the application domains. Here we have discussed some of the crucial applications of dynamic gesture recognition.

2.1.1 Sign language recognition

Sign and gestures are the most convenient way to convey messages among people through body movements [2, 27, 28]. Several systems have been developed on different sign languages to recognize gestures. Some remarkable works are found on American Sign Language (ASL) Recognition by different researchers. Starner et al. [105] developed an HMM-based real-time sign language recognition system to interpret ASL. Chong et al. [28] developed a machine learning approach for ASL recognition using a leap motion controller. Rahman et al. [93] put a new benchmark on ASL recognition using a Convolutional Neural Network. Aly et al. [2] have developed a method for user-independent American sign language alphabet recognition based on depth image and PCANet features. Japanese Sign Language (JSL) interpreter is proposed by Ito et al. [52] using Recurrent Neural Network

(RNN). A recent work [11] on the classification of JSL using gathered images and CNN. Another significant work on JSL is carried out by Chu et al. [12] using sensor-based hand gesture recognition methods. An Arabic Sign Language (ArSL) recognition system is proposed in [13] using two different neural networks, Partially and Fully Recurrent neural networks. Elpeltagy et al. [37] proposed a multimodality-based ArSL system. Mustafa et al. [84] carried out a significant review on ArSL. Korean Sign Language (KSL) is recognized by Shin et al. [103]. A new gesture recognition algorithm is defined in [57] for Korean scripts. Na, Y. [85] developed a classification system of the KSL alphabet using an accelerometer with a SVM. A CNN-based method is developed for recognizing KSL by Shin, H. [103]. A lexicon of 250 vocabularies in Taiwanese Sign Language (TSL) is recognized by [20]. Huang et al. [51] proposed a method for Video-Based Sign Language Recognition without Temporal Segmentation from a German sign language dataset using a Convolutional Neural Network. A comprehensive review is carried out on sign language recognition by Cheok et al. [24]. Several significant works are found in the literature on sign language recognition.

2.1.2 Virtual Reality

Gesture recognition technology contributes a lot to the field of Artificial Intelligence (AI). Several remarkable works are found in the literature in different applications of AI. To control the robot using gestures is one of the interesting applications in this area [10, 31]. Ghobadi et al. [41] propose a system that uses the numbering to count the five fingers for controlling a robot using hand pose signs. For a robot manipulator, pointing can specify positions on a 2D workspace as guidance [31]. Gestures are considered as one of the effective means for virtual environments [48]. Virtual reality interaction uses the hand gesture to manipulate the virtual movements using one or two hands for 2D and 3D interactions display [91]. Bertsch et al. [10], provide 3D pointing gesture recognition for natural human-computer Interaction (HCI) in real-time from binocular views. Some virtual reality applications are available in [45, 92]. Mo et al. [74] summarized the key issues of gesture recognition in the AI field. A wearable biosensing system is developed for hand gesture recognition by Moin et al. [80] with in-sensor adaptive machine learning capability.

2.1.3 Games

Gesture recognition technology has various game-based applications. In games, hand gestures are used instead of pressing the keys of the keyboard or moving the mouse cursor. Today, gesture-controlled games have an increasing demand. Gestures are used to control the interaction between the player and the computer. In virtual game applications [3] gesture recognition is widely used. Rautaray et al. [96] have developed a hierarchical recognition system of human gestures for sports video annotation. Using computer vision and gesture recognition techniques, Mozarkar et al. [81] have developed a vision-based low-cost input device for controlling the VLC player through gestures. A racing video game is developed by Zhu et al. [122] using real-time hand gesture recognition with Kinect. A novel approach [77] is developed to communicate with video game characters using a cascade classifier. A comparative study of hand gesture recognition devices for games [56] is carried out.

2.1.4 Medical Science

In medical science, gesture recognition contributes a lot. Several works of gesture recognition are found in the literature that is helpful to handle patients. Hand gesture recognition systems can help doctors in a surgical environment. Digital images can be manipulated during medical procedures using hand gestures instead of touch screens or computer keyboards [110]. It helps in medically monitoring the emotional states or stress levels of patients. Gestures can help medical practitioners when the hand is not suitable to touch some sensitive object. Bargellesi et al. [7] propose a hand gesture recognition method with a wearable motion capture sensor using random forest. Another hand gesture recognition system was developed by Zhao et al. [119] for healthcare using a convolutional neural network. Li et al. [63] present spatial fuzzy matching (SFM) in leap motion to improve hand gesture recognition. A continuous hand gesture detection system is implemented by Tai et al. [107] using long short-term memory (LSTM).

2.1.5 Dance

Dance gesture recognition has a wide range of applications. It is used for performance evaluation of dance [49], e-learning of dance [19, 100], and dance form recognition [14].

Gesture recognition task is carried out in different dance forms of the world. Dance gesture recognition in different Indian Classical Dance (ICD) forms is found in the literature. However, several works on different dance forms [49, 68, 100] other than ICD are also reported. The related works on dynamic dance gesture recognition are discussed in Section 2.3.

2.2 Approaches of Gesture Recognition

We have discussed different application domains of dynamic gesture recognition in the previous section. Most of the approaches of gesture recognition are applicable in virtual environments. The techniques of dynamic gesture recognition can be broadly categorized into two groups:

2.2.1 Device-based Approaches

Initially, a 2-D input device such as a pen or mouse was used for gesture recognition. In 1963, light-pen gestures were used in the Sketchpad system [53]. Pen-based gesture recognition systems were developed in the 1970s. These pen-based systems were used for various applications like document editing, controlling air traffic, and design tasks such as editing splines. In the later time, pen and touch-based gesture recognition systems become relevant in different applications.

The hand-related data are more accurate using wearable devices and thus achieve higher recognition accuracy. However, this system has some significant drawbacks also. In a device-based system, calibration can be complex. Here, tethered gloves reduce the range of motion and comfort. Data from inexpensive systems can be very noisy in the device-based system. However, accurate systems are very costly and not user friendly. The user needs to wear a cumbersome device. Thus, device-based systems have many drawbacks due to the need for the wearable device.

2.2.2 Vision-based Approaches

Though the device-based systems are accurate, these are not user friendly. The most significant disadvantage of device-based systems is that they are cumbersome. Computer vision techniques can provide real-time data useful for analysing and recognizing human motion passively and unobtrusively.

The camera is the only device for collecting the user's hand gesture movement in computer vision. Vision-based approaches are more user-friendly and do not require any extra devices for capturing gestures. It is the most natural way of user interaction as human perceives information from their surroundings. These approaches deal with some properties such as texture and color for analysing gestures while tracking devices cannot. In this method [50], the input images or videos are captured using camera(s). Although these approaches are simple, many challenges are raised, such as the complex background, lighting variation, noisy videos as specified in [21, 106]. Compared to the systems that use different devices (gloves, sensors), vision-based systems are more user-friendly and simpler. Vision-based systems are easy to use but challenging to implement.

In vision-based systems, one or more cameras are used to capture images. Typically frame rate of 30 Hz or more is considered for frame extraction, and features are extracted to recognize gestures [73, 116]. The cameras are kept fixed, although they may also be mounted on moving platforms or other people. A significant amount of computer vision-based research has been found in the literature on face detection, speech recognition, human activity recognition, and gesture recognition.

Vision-based gesture recognition systems need to observe several parameters, including several cameras, speed and latency, user requirements, primary features, and representation of time.

2.3 Dynamic Gesture Recognition

This section presents the state-of-the-art in the field of dynamic gesture recognition.

Dynamic gesture recognition is a research field of increasing interest in recent years. Several previous works have focused on developing a more intuitive interface to interact with machines and other devices. Sequences of images that are captured over a continuous period are used for dynamic gesture recognition. The recognition result is related to the temporal features of the object describing the trajectory in the sequence. However, recognizing a dynamic gesture is a challenging issue because a gesture can be represented in different ways within the same context. Besides, the way gestures are performed depends not only on the sequence of body movements but also on the cultural aspect of the people who make the gestures. Consequently, there is still much work to obtain an interface capable of providing effective communication between humans and machines based on the gesture.

2.3.1 Feature Extraction

Feature extraction refers to the process of transforming raw data into numerical features that can be processed while preserving the information in the original data set. It yields better results than applying machine learning directly to the raw data. Features are the vital information that is extracted from a large dataset. Feature extraction involves when the input data to an algorithm is too large to be processed and suspected to be redundant. The reduced extracted meaningful information from a large dataset is named as features or feature vectors.

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. 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.

Inter-frame difference is normally used to calculate the object's movements. Energy information can also be extracted from the images. Inter-frame difference of energy and phase energy spectrums [16] can be used as energy features. The phase energy spectrum is a product of partial derivatives in the spatial phase-frequency spectrum over their spatial frequencies. It provides complete information about motion infinite frames [16]. The edges

in an image play a significant role. The modeling of inter-frame differences of frequency responses is based on analyzing the boundary pixels. It increases the probability of movement detection. A deformation of a moving object's shape, movement characteristics, and several moving objects is defined from analyzing inter-frame differences. The inter-frame differences of frequency responses always lead to the best results than the differences of video signals in the spatial domain. The changes of the energetic indexes in static images determine the efficiency function as a dependence of output and input energies of the 2D filter. This function is defined on a whole set of impulse responses of a filter. The efficiency function is a positively certain quadratic form with certain coefficients. These coefficients are obtained from the energy spectrum decomposition of the input frame into the 2D Fourier series over the cosines. The analysis of stationary points and their efficiency function allows synthesizing the optimum and the quasi-optimum 2D filters.

For online gesture recognition systems [49], the used features are- center of motion, mean absolute deviation from the centre of motion, and motion intensity. However, it works well only for continuous recognition of isolated gestures. Handshape and motion pattern are the features used for remote robot control [55]. Here continuous deformation of hand shapes is not permitted. For robot control [69], distance, velocity and acceleration information, energy measurements, or angle information features were used. It gives satisfactory performance only for a limited alphabet under uniform background. The features 1D binary signal based on the centre of gravity are extracted for robot control by [48]. It considers a limited number of gestures. Hand motion is used by [64] for sign language recognition. It works well for isolated gestures compared to continuous gestures. Hand movement, hand shape, motion trajectory features are used for Taiwan sign language recognition [63]. Here no appropriate fixed threshold to distinguish non-sign segments. Hand position, velocity, size, and shape features are used for virtual reality [67]. This method can handle only single-handed gestures. The joint of interest feature is used for sign language recognition by [73]. The system is capable of detecting gestures that do not involve specific finger movements. Scale Invariant Feature Transform (SIFT) feature is used for ASL recognition by [75]. In this system, error sources can be the viewing angle of the camera, lighting conditions, and different clothing for users. In [59], pose vectors from 2D silhouettes are used for interactive environments with full-body gestures as the control signal. This system is independent of the

view angle of the camera. 2D silhouette and visual hull features are used for multi-person gesture-driven human-computer interactions [81]. It works for isolated gestures only. The Skeleton feature is used by [49] for the performance evaluation of Bali traditional dance. This method can be used as an alternative to dance gesture recognition. Structured streaming skeletons are used by [82] in interactive systems. Skeleton of the body is used for Kazakh dance gesture recognition by [87]. Here recognition is based on head movement. These are some significant features used for various applications of dynamic gesture recognition.

From the experimental analysis, the features that perform well in this work are height to width ratio MBR of each frame, inter-frame energy difference, and inter-frame entropy difference.

2.3.2 Classification

The final phase of gesture recognition is classification. Various vision-based dynamic gesture recognition systems, including machine learning approaches and deep learning approaches are discussed in this section.

Various classifiers used for gesture recognition are - HMM [39, 46], principal component analysis (PCA) [1, 32], finite state machine (FSM) [47, 113], dynamic time warping (DTW) [50, 90], connectionist approaches (time-delay neural network (TDNN) [38, 118], multilayer perceptron (MLP) [78], and RNN [9]) and soft computing (fuzzy systems, genetic algorithm) methods. Standard pattern recognition techniques, template matching, and neural networks can be used in static gesture recognition. In contrast, techniques such as HMMs, FSM, DTW, time compressing templates, and TDNN can be used in dynamic gesture recognition [73].

The dynamic time warping algorithm [50, 90] and the hidden Markov model [39, 46] are the classical methods that handle the temporal and spatial information of dynamic gestures. Initially, DTW was used for speech recognition. It measures the similarity of two time series of different lengths. Corradini et al. [33] first used a dynamic time regularization-based algorithm to recognize dynamic gestures. Initially, preprocessing is performed on the training dataset, and then features are extracted and normalize them. Now, the identical steps are performed on the test set repeatedly. Eventually, the generated results are matched with

each template in the training set, and the gesture recognition result is the category with the smallest difference from the template. The DTW method does not use the statistical model framework for training. However, it is challenging to use temporal information in an image recognition algorithm. Therefore, there are specific difficulties when dealing with large data volumes and complex gestures.

HMM is a statistical model widely used in dynamic gesture recognition because of its ability to handle the Spatio-temporal identity of gesture [13]. 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 [22]. It has been proven that HMMs are effective in sign language recognition and other complex hand gesture recognition processes [108]. Besides, HMM performs well in [108, 89, 94] for recognizing full-body gestures. HMM can efficiently model spatiotemporal time series of gestures effectively and can handle non-gesture patterns [86]. There are three significant issues in HMM: evaluation, estimation, and decoding. These problems are solved using the Forward algorithm, Viterbi algorithm, and Baum-Welch algorithm, respectively [22].

A view-invariant gesture recognition framework using voxel data obtained through visual hull reconstruction from multiple cameras is presented in [81]. View-invariant pose descriptors are extracted using multilinear analysis. Gestures are then treated as sequences of pose descriptors, and HMM are used for gesture recognition. The recognition rate is suitable for isolated gestures. A view-invariant video-based whole-body gesture recognition system is proposed in [89]. The multilinear analysis is performed on the static poses' silhouette images, making up the gestures by tensor decomposition and projection. The pose vectors are the inputs to the HMM for gesture recognition. This is the first system that addresses full-body human gesture recognition from video without the recovery of body kinematics or 3D volumetric reconstruction. A new effective and efficient feature extraction method is presented in [120] for online human gesture recognition. They solve the problem of recognizing gestures from unsegmented streams continuously and differentiating different styles of the same gesture from other types of gestures. This method is applicable in interactive systems. Full body gesture recognition has application in the field of dance. It is used for dance gesture recognition, performance evaluation of dance. Heryadi, et al. [49]

presents an approach for dance gesture performance evaluation of Bali traditional dance. This method can be used as an alternative to dance gesture recognition.

Some of the notable classification approaches of dynamic gesture recognition include K-nearest neighbor, support vector machine, Decision tree, and Bayesian network. Wang and Popovic [114] propose a real-time hand tracking application using colored gloves. They used the KNN approach to recognize the color pattern of the gloves. However, their system required only continuous hand streams. Also, KNN is compared with support vector machine [6, 61, 97, 109], and all these systems have proved that the SVM based approaches outperform the KNN approaches. However, the SVM-based approaches will classify only the static gestures, and it requires more features to train the systems.

Today, deep learning is used everywhere including dynamic gesture recognition. An approach for combining traditional hand-crafted features with a CNN is presented by Chevtchenko et al. [25]. They worked on depth and grayscale images for evaluation where the background has already been removed using depth data. CNNs are used as feature extractors from point clouds captured by a depth sensor by Liang et al. [67]. Finally, hand gesture recognition is performed using SVM. Oyedotun and Khashman [88] propose a method to recognize 24 ASL hand gestures with CNNs and stacked denoising autoencoders (SDAEs). A soft attention mechanism is proposed by Li et al. [65] to localize and classify gestures using a single network automatically. To generate proposals, a sliding window approach was used. Then image features are extracted using a pre-trained CNN architecture. The attention network was applied to acquire the weight of each proposal to locate hands. A softmax function layer was used for gesture recognition.

2.4 Dynamic Dance Gesture Recognition

Dance gestures are some meaningful expressions that differ dance-wise. Automatic recognition of dance gestures can help to create a universal communication environment for a dance drama, independent of the language used in the associated song [81]. Dance gesture recognition systems help audiences understand the meaning of the dance sequence without the knowledge of the background song.

Research on dynamic dance gesture recognition is reported in several dance forms. Typical works in full-body dance gesture recognition are found in [49, 57, 92, 110, 117]. Heryadi et al. [49] proposed syntactical modeling and classification for Bali traditional dance performance evaluation that considers the full-body gestures. Bali dance gestures are represented as a set of skeleton feature descriptors that are extracted from images captured by using a Kinect depth sensor. A set of rules is learned from the training examples to capture the structure of the gesture motion using the grammar inference method. The empirical results show that the elbow and foot of the performer are the most discriminative features of Bali traditional dance. Probabilistic and deterministic grammars have achieved 0.92 and 0.95 of average precision for recognizing the tested dance gestures.

Another full-body gesture recognition system was demonstrated by Nussipbekov et al. [87] for recognizing Kazakh traditional dance gestures in which a Microsoft Kinect camera is used to obtain human skeleton and depth information. This method considers the movement of the head. It uses tree-structured Bayesian network and an Expectation-Maximization algorithm with K-means clustering to calculate conditional linear Gaussians for classifying poses. Furthermore, they use Hidden Markov Model to detect dance gestures.

A work on Ballet dance is found in Konar [58]. They have considered twenty primitive postures of Ballet for the identification of an unknown dance posture. They have measured the proximity of an unknown dance posture to a known primitive simultaneously. A six-stage algorithm is proposed to achieve the intended objective. They have performed skin color segmentation on the dance postures, which are dilated and processed to generate skeletons of the original postures. These are some significant works of gesture recognition on international dances.

In recent years, dance gesture recognition has achieved remarkable success in different Indian Classical Dance forms. There are several significant works on dance gesture recognition on ICD are found in the literature. Most research works in dance gesture recognition have been carried out on different types of Indian classical dances like Bharat Natyam [14], Odissi [100], and Kathak [111].

A work on dance gesture recognition has been reported in [100] on Odissi dance that uses Kinect sensor. The recognition rate is 86.8% using Support Vector Machine. In [14] and [111], a real-time gesture classification system for skeletal wireframe motion has been

developed. They have used the XBOX Kinect platform. The classifier has an average accuracy of 96.9% for approximately 4-second skeletal motion recording.

Another work has been found in [108] for Indian classical dance form recognition using a deep learning approach. The classification accuracy of the work is 75.83%.

Tiwari [111] proposed a pose recognition method on three ICD forms, namely Odissi, Kathak, and Bharatnatyam. They have created a database of 100 images and split them into training and testing datasets. Hu moments are taken as features to describe the shape context of an image since they are scale, translation, and rotation invariant. The foreground and background must be separated to extract Hu moments from the images. SVM is used using one vs. one approach and one vs. all approach since it is a multiclass classification problem. Linear and RBF kernels are used for comparison for both approaches.

Mohanty et al. [79] proposed work on understanding small video sequences of Bharatanatyam, which comprises hand gestures, facial expressions, and dynamic body postures called Adavus. Initially, they recognize static and dynamic hand gestures with 2D and 3D CNNs. Subsequently, the basic dance steps in Bharatanatyam, i.e., Adavu, are interpreted. They propose datasets for static and dynamic hand gestures captured under controlled laboratory settings and from real-world videos. They also create a separate dataset for Adavus. For detecting static hand gestures in video frames, an adaptive boosting (AdaBoost) [102] approach and region-based CNN (R-CNN) [42] are used. Furthermore, a skin detector is used for refining the localization of hands in the images. Finally, they demonstrate the possibility of using the proposed algorithms to understand a video of a Shloka enacted in Bharatanatyam.

Indian classical dance has existed for over 5000 years and is widely practiced and performed worldwide. Every dance conveys some meaningful information. However, the common people rarely understand the information conveyed by the dancer, and the dance experts can only appreciate it. Mohanty et al. [79] address this highly challenging problem and propose a deep learning-based method to identify the meaning of the gestures performed by the dancer. They propose a convolutional neural network and validate its performance on standard datasets for poses and hand gestures and on constrained and real-world datasets of classical dance. They use transfer learning to show that the pre-trained deep networks can

reduce the time taken during training and improve accuracy. They succeed in their intended objective and show with experiments performed using Kinect in constrained laboratory settings and data from YouTube.

Indian classical dance such as Kathakali is composed of complex hand gestures, body moments, facial expressions, and background music [12]. Due to the complexities involved in its hand-gesture language, it is often challenging to understand kathakali mudras. They have generated a dataset for Kathakali hand gestures and explore ways to recognize Kathakali dance mudras performed by artists with the help of machine learning and deep learning techniques. There are 24 classes of hand gestures that are used to convey the story by the performer in Kathakali. They propose a Support Vector Machine model and Convolutional Neural Network model, classifying the images into 24 different classes. They have compared the performance of machine learning algorithms and deep learning algorithms. Classification accuracy of their method is 74%. This is the first attempt to generate a dataset of Kathakali hand gestures, explore data preprocessing techniques for machine learning techniques, and apply deep learning techniques to classify Kathakali hand gestures.

Sattriya is one of the eight principal Indian classical dances. The dynamic gestures of Sattriya dances are known as Ground exercises. These are the basic requirement to learn Sattriya dance. It is observed from the literature that no research works on recognition of dynamic gestures of Sattriya dance are reported so far. This work is added in the research area of dance gesture recognition as a new task.

2.5 Datasets

The two promising databases, Hollywood [62] and UCF50, contain videos captured from cinemas and YouTube. They consider several conditions, namely camera movement, dancer occurrence, point of view of the dancer, and complex background. The activity database, UCF50, contains the ten activity classes from the UCF YouTube dataset. The sports datasets, the UCF Sports dataset [98], and the Olympic Sports dataset [86] include sports videos from YouTube. The identification accuracy is 98% using the leave-one-clip-out method. Based on the proposed HMDB51, they have done a straightforward experiment where they took ten activity classes the same as UCF50 and done manual labeling for the 14 joint positions over

1,100 motion clips. The classification of features represented from the joint positions at one frame resulted in only 35%. However, it is 54% while using movement features. The classification rate is increased to 66% by applying movement features on ten pose classes. Another activity dataset, the HMDB51, contains 51 different activity classes whose pose classes have different movements.

The Let's dance dataset [19] is provided by the Georgia tech college of computing and is a direct result of the Let's Dance: Learning from Online Dance Videos research paper. It's the first dataset encountered that is focused completely on dance. According to the original paper, the dataset consists of 1000 videos, containing 10 dynamic and visual overlapping dances which include; Ballet, Flamenco, Latin, Square, Tango, Breakdancing, Foxtrot, Quickstep, Swing and Waltz. There are meant to be 100 videos per class which are approximately 10 seconds long and at 3o fps.

However, there is no publicly available video dataset of ICD in the literature till date. The researchers in this domain use their own data for their respective work. YouTube videos are also used by many for their work [62, 86, 98]. Due to this limitation, we have created a video dataset of this work that can be used as a benchmark for vision-based dance gesture recognition. We have presented our dataset in Chapter 3.

2.6 Research Issues and Challenges

The following research issues and challenges are identified based on the literature survey-

- Recognition of full-body gestures is an issue since the human body is a highly articulated structure.
- Selecting suitable features is a difficult task in dynamic gesture recognition that significantly affects good recognition.
- Recognition of dynamic gestures itself is challenging work because of its spatio-temporal behaviour.

- Gesture recognition systems usually give satisfactory performance in a controlled environment only. Developing a robust gesture recognition system that works in an uncontrolled environment is still a challenging task.
- There is no video dataset available for Ground exercises of Sattriya dance which is a major issue for the researcher to work in this dance form.

2.7 Conclusion

This chapter attempts to present a comprehensive survey on different research works on gesture recognition based on computer vision. The survey provides a detailed understanding of the state-of-the-art in this area. We also find out the research issues in dynamic gesture recognition. Although dynamic gesture recognition research has made significant progress in recent years, there are still many prospects for improvement. This is a new research work towards dance gesture recognition in Sattriya dance form. The research contributions based on this understanding are presented in the subsequent chapters.