ABSTRACT

Progress of computer vision and machine learning technologies in recent years have enabled us to handle the complex information related to human activities in multimodal data. So, now we can analyze and interpret performing arts like dance. The dance is a complex domain associated with the rhythm of the music and storytelling through different body actions synchronizing with the lyric of the music. *Bharatanatyam* is an old legacy of Indian heritage. It comprises musical rhythms, lyrics, tempo, stances/postures, and movements. Like any other classical dance, it tells a story through facial expressions, hand gestures, and different body movements during a performance. In *Bharatanatyam*, *Adavu* is the basic dance step that is guided by the music, called *Sollukattu*s. This thesis aims to analyze and interpret the *Bharatanatyam* dance performances using machine learning technology and demonstrates an end-to-end workflow from data capture to application.

Every machine learning implementation requires labelled or annotated data-sets. So, this thesis first captures 1276 *Adavu*s and 162 *Sollukattu*s. *Adavu*s are recorded using Microsoft Kinect 1.0 sensor. After recording, all these videos and audios are annotated. We use these data and annotations to perform several tasks that are part of this thesis.

Music drives any dance. So, the structural analysis of *Sollukattu*s are done as the first work. In this, we segment the audio and recognize the voiced beats (*bol*s) using Gaussian Mixture Model (GMM). We achieve *bol* recognition accuracy of 94.79%. Next, we recognize *Sollukattu* as a sequence of *bol*s using the Bayesian Network and get an accuracy of 94.4%. We also host a part of the annotated data at `http://hci.cse.iitkgp.ac.in/` for other researchers.

The second task is segmenting the video to identify the key frame (momentarily stationary) and motion frames. In this, we use Support Vector Machine (SVM) and Convolutional Neural Network (CNN) to build the models, which helps in the motion-guided segmentation. Both the ML models achieve $\approx$ 90% accuracy. However, CNN performs better with 92.3% accuracy.

Next, to analyze *Adavu*s' complexity based on the associated motions, we cluster the motions associated with each *Adavu*. The clustering accuracy is 92%. We use the skeletal joint velocity and Motion History Image (MHI) as the features for motion recognition across the *Adavu*s using k Nearest Neighbour (kNN), SVM, and CNN classifiers. We achiever the accuracy of 68.33%, 95.58%, and 93.92% respectively for kNN, SVM, and CNN.

We also analyze the joint trajectory to classify the motion with the help of SVM. In this, SVM achieves an accuracy of 90.55%. CNN and SVM classifiers are used on RGB images and skeletal joint angles respectively to recognize the key postures involved in *Adavu*s. CNN achieves a better accuracy (99.46%) compared to SVM (93.08%). After

successfully recognizing the Key posture and Motions, we classify the *Adavu*s as a sequence of motions and postures using Edit Distance algorithm and achieve a recognition accuracy of 98.66%.

The thesis also extensively surveys various aspects of computer analysis of World as well as Indian Classical Dance to identify several research gaps in a structured manner. Many of these gaps are then addressed to provides solutions as an end-to-end workflow from data capture to application. It also presents a comparative study between the proposed solutions and the earlier work. This work contributes significantly to the state-of-art in computer analysis of Indian Classical Dance.

Keywords: *Bharatanatyam, Adavu, Kinect, Machine Learning, Audio / Video Annotation, Video Segmentation, Motion Clustering, Motion Recognition, Posture Recognition, Dance Sequence Recognition, Tutoring Application.*