NatyaSet: Classical Dance Dataset for Bharatanatyam

Vishakha Hegde[1], Rayudu Srishti[2], Vidisha Chandra[3], Vibha Murthy[4] and Prof. K S Srinivas[5]

[1,2,3,4,5]*PES University, Bengaluru*

Abstract: *Indian classical dances hold deep cultural and historical significance, yet the availability of specialized datasets catering to these art forms remains limited. To address this gap, we present a comprehensive dataset focusing on Bharatanatyam, one of India's prominent classical dance forms. The dataset is meticulously curated, capturing both 2D and 3D pose information, enabling a three-dimensional analysis of dance movements. The need for specialized classical Indian dance datasets arises from the distinct characteristics of these art forms, which may not be accurately represented in conventional datasets focusing on Western dances or general human activities. Our curated dataset aims to bridge this gap by providing a dedicated resource for studying the complexities and nuances of Bharatanatyam and its movements. The integration of audio features further enhances the dataset's utility, opening up new avenues for cross-modal analysis and multimodal research tasks.*

Keywords: **Bharatanatyam, keypoints, dataset, OpenPose, MediaPipe, 3D, 2D, SMPL, pose estimation.**

1. Introduction

Indian classical dances are renowned for their rich cultural heritage, intricate movements, and profound expressions that have been passed down through generations. Bharatanatyam, a prominent classical dance form from South India, is a captivating art that combines precise gestures, rhythmic footwork, and emotive storytelling. As the world embraces advancements in computer vision, motion analysis, and cultural preservation, the need for specialized datasets dedicated to classical Indian dances becomes increasingly evident. However, the availability of such datasets, particularly focusing on Bharatanatyam, remains limited.

To address this gap and contribute to the study and preservation of classical Indian dance forms, we present a novel and comprehensive dataset dedicated to Bharatanatyam. The primary objective of our dataset is to provide researchers and enthusiasts with a valuable resource that captures the intricacies and subtleties of Bharatanatyam performances. Through the inclusion of both 2D and 3D pose information, our dataset enables a three-dimensional analysis of dance movements, enhancing the understanding of the dancers' artistic expressions.

To achieve accurate pose estimation and comprehensive keypoints representation, we employ state-of-the-art techniques such as OpenPose and MediaPipe. This dual approach allows us to extract a diverse range of keypoints, accurately capturing the unique postures, mudras (hand gestures), and body movements inherent in Bharatanatyam. By integrating synchronized audio features, including music beats per minute (BPM) and rhythm patterns, we further enrich the dataset, enabling the study of the inseparable connection between dance movements and the accompanying music.

In the following sections, we provide a detailed description of our data collection process, video preprocessing techniques, and pose estimation methodologies, showcasing the efficacy and uniqueness of our dataset. Additionally, we discuss the diverse choreographies, music synchronization, and non-overlapping train and validation subsets, ensuring unbiased evaluations and encouraging varied approaches in multimodal sequence-to-sequence generation tasks.

Literature Review and Existing Datasets

2.1. AIST++ [1]

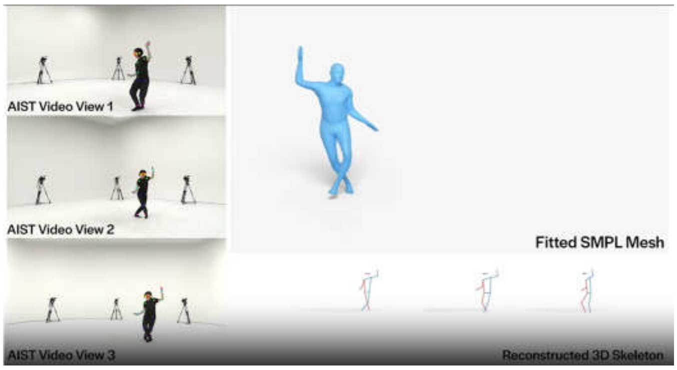

Figure 1. AIST++ Video Multi-views

The AIST++ Dance Motion Dataset is constructed from the AIST Dance Video DB. With multi-view videos, an elaborate pipeline is designed to estimate the camera parameters, 3D human keypoints and 3D human dance motion sequences. It provides 3D human keypoint annotations and camera parameters for 10.1M images, covering 30 different subjects in 9 views. These attributes make it the largest and richest existing dataset with 3D human keypoint annotations.

It also contains 1,408 sequences of 3D human dance motion, represented as joint rotations along with root trajectories. The dance motions are equally distributed among 10 dance genres with hundreds of choreographies. Motion durations vary from 7.4 sec. to 48.0 sec. All the dance motions have corresponding music. With those annotations, AIST++ is designed to support tasks including: Multi-view Human Keypoints Estimation, Human Motion Prediction/Generation, Cross-modal Analysis between Human Motion and Music.

2.2. Everybody Dance Now [2]

This paper presents a two-part dataset. First, the authors present a repository of long single-dancer videos. The deliberate emphasis on single-dancer videos is a strategic choice, enabling researchers to focus on the intricacies of motion transfer and video generation while simplifying the dynamics that can arise from multiple interacting dancers. They specifically designated the single-dancer data to be high-resolution open-source data for training motion transfer and video generation methods. The second facet of the dataset entails a vast collection of short YouTube videos. These videos play a pivotal role in facilitating transfer learning and fake detection.

The dataset comprises self-filmed long target videos spanning 8 to 17 minutes, including 4 videos at 1920 × 1080 resolution and 1 at 1280 × 720. Static backgrounds were ensured in each frame through the use of stationary cameras. Maintaining frame quality was a priority, hence, the target subjects were recorded for durations of 8 to 30 minutes in real-time footage, employing modern cellphone cameras at 120 frames per second.

2.3. ActivityNet [3]

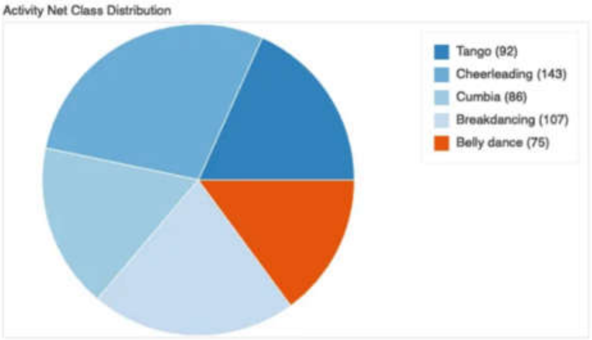

Figure 2. ActivityNet Dance Styles

The ActivityNet dataset is organized in a JSON file with different sections. In the 1.3 version, there are 503 videos specifically focused on dance out of a much larger total of 19994 videos. This means that dance-related content accounts for only around 2.52% of the entire dataset. Within the dataset's collection of 200 activity categories, a mere 5 categories pertain to various forms of dance, making up just 2.5% of the total categories. However, this limited representation of dance content doesn't provide the depth and variety needed for effective deep learning.

2.4. IIT Kharagpur's Annotated Bharatnatyam Data Set [4]

This dataset consists of annotated data sets of Bharatanatyam Adavus (basic unit of Bharatanatyam) and Sollukattus (the corresponding audio) for Human-Computer Interaction research, and was one of the only existing datasets we found for Bharatanatyam dance. The dataset consists of both the audio (Sollakattu) and video (Adavu) data along with the annotated files. There are 23 unique sollakattus and 15 Adavus with 58 variations.

The video data was captured using the sensor Microsoft Kinect nuiCapture is given as MAT Data Files. Each MAT file contains 3 streams data -- RGB frames, Depth frames, and Skeleton frames. Each video has around 700-1000 frames. The motions in each video were annotated with duration (start frame number to end frame number), and type of motion (motion class). These annotations are available as csv files.

3. Methodology

3.1. Data Collection

We have curated a dataset that includes both 2D and 3D information, enabling the analysis of Bharatanatyam movements in a three-dimensional space. We compiled a dataset of over 200 publicly available YouTube videos showcasing Bharatanatyam performances by various dancers and choreographies for data collection. We covered as many styles and variations of Bharatanatyam possible, including both Classical music dances as well as fusion dances. We covered all aspects of the dance, including dances with more movement (Nrithya) and less expressions (Abhinaya) and vice versa. This adds variety to our dataset for a wider range of Bharatanatyam dances to train any future model on.

3.2. Video Preprocessing

The collected videos undergo preprocessing steps to optimize their quality for pose estimation. This involves techniques such as video stabilization, resolution enhancement, and noise reduction. By enhancing the visual clarity and consistency of the videos, we improve the accuracy of subsequent pose estimation algorithms. We used the below filtering techniques to enhance the video quality.

The Gaussian filter method for noise reduction is:

$$G(x',y') = (1/2\pi\sigma{}^\wedge 2) * exp(-(x'2 + y'{}^\wedge 2)/(2\sigma{}^\wedge 2)) \tag{1}$$

where (x', y') are the coordinates of the filter mask, σ is the standard deviation of the Gaussian distribution, and G(x', y') represents the weights assigned to the neighboring pixels.

We used the following bilinear interpolation for resolution enhancement:

$$I(x',y') = (1 - \alpha)(1 - \beta) * I(x,y) + \alpha(1 - \beta) * I(x + 1,y)$$
$$+ (1 - \alpha)\beta * I(x,y + 1) + \alpha\beta * I(x + 1,y + 1) \tag{2}$$

where (x, y) are the original pixel coordinates, (x', y') are the interpolated pixel coordinates, I(x, y) represents the original pixel value, and α and β are the fractional parts of the coordinates.

We estimate the global motion parameters and apply image warping to compensate for camera movements.

The formula for affine transformation can be used for image warping:

$$(x',y') = (ax + by + tx, cx + dy + ty) \tag{3}$$

where (x, y) are the original pixel coordinates, (x', y') are the transformed pixel coordinates, (a, b, c, d) are the affine transformation parameters, and (tx, ty) are the translation parameters.

3.3. Pose Estimation

Utilizing state-of-the-art pose estimation techniques, we extract 2D joint positions from each video. Both OpenPose, a widely adopted computer vision framework, and MediaPipe are employed to detect and localize the key body joints of Bharatanatyam dancers. This dual approach allows us to obtain a more diverse range of keypoints, enhancing the accuracy and comprehensiveness of the captured intricate movements and positions of the dancers. The combination of these two powerful pose estimation methods contributes to a richer analysis of the dancers' performances.

Each video has around 600-1000 frames depending on the duration of the dance video. The MediaPipe pose landmarker model we used tracks 33 body landmark locations, and the OpenPose model consists of 25 keypoints.

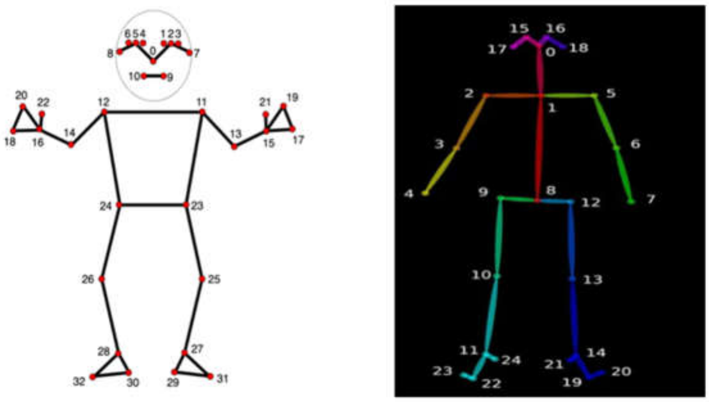

Figure 3. MediaPipe and OpenPose points respectively

The points retrieved from the videos are stored in the form of a JSON file which contains structured data representing landmarks in a three-dimensional space, pertaining to human body poses or keypoints for each frame of the video. Each landmark has three coordinates (x, y, and z), denoting its position in the 3D space. Additionally, there is a visibility value indicating the certainty of the landmark's detection. These JSON files can then be easily fed into any required model to train and further process the data and use it for prediction and dance generation purposes.

3.4. 3D Reconstruction

We recover the 3D motion of the dancers in terms of SMPL (Skinned Multi-Person Linear) parameters which are visualized appropriately. The 3D mesh figure generated is in the form of SMPL points and hence can be used to train various dance models. Furthermore, our dataset includes synchronized music with varying beats per minute (BPM), capturing the interplay between music and dance in Bharatanatyam performances.

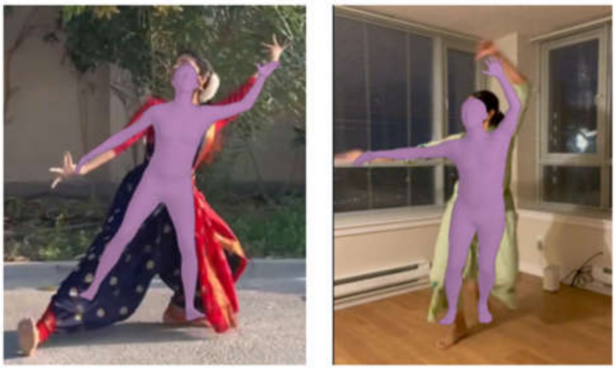

Figure 4: 3D reconstruction with SMPL points

4. Dataset Description

The dataset consists of the following data for each frame:

1. **Bharatanatyam Pose Frames:** Sequential frames extracted from the recorded dance performances, capturing the dancer's body postures and movements.
2. **Mediapipe Keypoints:** Keypoints obtained using the Mediapipe library, providing a set of 2D keypoints for the body, hands, and face with corresponding confidence values.
3. **Openpose Keypoints:** Keypoints obtained using the Openpose framework, providing 2D keypoints for multiple people in the frame with associated confidence scores.
4. **SMPL Keypoints:** Skinned Multi-Person Linear (SMPL)*[5]* keypoints, with 24 pose parameters, representing the 3D pose and shape of the dancer, if available.
5. **Audio Features:** A set of audio features extracted from the accompanying audio clips, including pitch, tempo, and Mel-frequency cepstral coefficients (MFCCs) to represent the musical aspects of the dance performance.

5. Comparison Results

Unlike most existing dance datasets that solely provide video recordings, our dataset stands out by offering accompanying JSON files containing annotated keypoints. This inclusion of keypoints enhances the utility of the dataset for dance generation applications. Unlike video-only datasets, our JSON-enhanced dataset enriches the training process by providing essential pose-related information.

Furthermore, typical dance video datasets often exhibit limitations in terms of the number of dance styles and the diversity of dances captured. Our dataset addresses this constraint by offering a wider range of dances and styles, ensuring a more comprehensive and versatile resource for training dance generation models.

Notably, existing dance generation models predominantly rely on Western dance datasets, overlooking the inclusion of Indian dance forms. Our dataset bridges this gap by providing a valuable resource for incorporating Indian dance styles into dance generation research. The integration of OpenPose, MediaPipe and SMPL keypoints in our dataset contributes to heightened accuracy in predicted pose points. This dual approach enhances the quality and reliability of the keypoints, bolstering their usability for various research applications.

In contrast to the IITKGP Bharatanatyam dataset, which primarily covers basic Adavus without comprehensive dance items, our dataset encompasses a spectrum of dances, ranging from beginner to advanced levels. Moreover, our dataset surpasses the limitations of missing keypoints annotation present in the IITKGP Bharatanatyam dataset, thereby providing a more detailed and complete resource for dance-related research endeavors.

6. Conclusion

Our novel Bharatanatyam dance dataset with annotated keypoints marks a significant advancement in dance research. Going beyond conventional video-only datasets, our inclusion of JSON files enhances the dataset's value for dance generation applications. By incorporating OpenPose, MediaPipe and SMPL keypoints, we ensure accuracy in predicted pose points, offering reliable data for diverse research needs. Moreover, our dataset's focus on Bharatanatyam contributes to the broader dance research landscape by providing a resource to incorporate Indian dance forms. Our contribution aims to bridge the gap between the domain of computer vision and the world of cultural arts, fostering a deeper appreciation for the timeless artistry of Bharatanatyam and its profound impact on India's cultural heritage.

Acknowledgments

We would like to thank to the participants who generously shared their time and expertise, making our Bharatanatyam dance dataset possible. We acknowledge PES University for the opportunity to work on this project. This research journey has been enriched by the collaboration and encouragement from our peers and the wider dance community.

REFERENCES

7.1. Conference Proceedings

[1] Li R, Yang S, Ross DA, Kanazawa A. “AI Choreographer: Music Conditioned 3D Dance Generation with AIST++”. vis IEEE/CVF International Conference on Computer Vision (ICCV) 2021, pp. 13401-13412.

[2] Chan, C., Ginosar, S., Zhou, T., & Efros, A. A. (2019). Everybody Dance Now. In IEEE International Conference on Computer Vision (ICCV).

[3] Fabian Caba Heilbron, Victor Escorcia1, Bernard Ghanem and Juan Carlos Niebles. “ActivityNet: A Large-Scale Video Benchmark for Human Activity Understanding”

[5] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: a skinned multi-person linear model. ACM Trans. Graph. 34, 6, Article 248 (November 2015), 16 pages.

7.2. Links

[4] http://hci.cse.iitkgp.ac.in/