# TraDance: A dataset for assessing the performance of Greek folk dances recognition models

Thanasis Balafoutis*, Vasiliki Balaska†and Antonios Gasteratos‡
Department of Production and Management Engineering, Democritus University of Thrace
Xanthi, Greece
Email: *abalafou@pme.duth.gr, †vbalaska@pme.duth.gr, ‡agaster@pme.duth.gr

***Abstract*—In computer vision, datasets and benchmarks are widely used to compare algorithms and boost scientific progress. Especially in the human action recognition research field, extracting dance poses from video sequences for fragmentation and recognition of dancer movements is a challenging task, and new datasets are always important. This paper presents a new benchmark dataset and evaluation methodology for dancing analysis and step movements classification. We focus on traditional Greek dances, which are characterized by the repetition of simple steps throughout the duration of the dance track. The dataset, named TraDance, contains 1.4 hours of 3D dance motion videos from 5 different experienced dancers and 3494 annotated dance steps, covering 5 traditional dance genres with known RGB-D camera settings. In addition, to evaluate the quality of TraDance, we provide extensive experiments using existing machine learning algorithms that are commonly used in computer vision. The results show that TraDance is feasible for research-purpose studies and reliable. The dataset can be found online at: https://robotics.pme.duth.gr/tradance.**

***Index Terms*—Skeleton-based action recognition, Folk Greek Dances, Dataset, Classification, Intangible Cultural Heritage.**

## I. Introduction

The field of computer vision has enjoyed remarkable advances in recent years, largely thanks to the development of deep learning models and the availability of large-scale datasets. One of the most active research areas in computer vision is human action recognition, which seeks to automatically classify and understand the movements and gestures of human subjects in video data [1]. Action recognition has numerous applications, including video surveillance, human-computer interaction, and entertainment. However, it remains challenging due to the complexity and variability of human actions, the presence of occlusions and clutter in the data, and the need for large and diverse datasets for training and evaluation.

Dance is a particularly interesting domain for action recognition research, as it involves highly stylized and expressive movements that can convey rich cultural and emotional meanings. Dance analysis and recognition can provide insights into dance forms' history, aesthetics, and social significance and support their preservation and dissemination [2]. Moreover, dance can be challenging for computer vision due to the complex and dynamic nature of dance movements, the variability and idiosyncrasy of individual dancers, and the need to capture both spatial and temporal information in the data.

To address these challenges and promote progress in dance analysis and recognition, several benchmark datasets and evaluation protocols have been proposed in recent years [3]–[5]. These datasets typically involve many video clips featuring dancers performing various actions or dance styles, along with annotations and labels indicating the action or style being performed. They have been used to develop and evaluate a variety of machine learning and computer vision algorithms.

However, most existing datasets have focused on Western dance styles or contemporary dance forms, and there is a need for more diverse and culturally specific datasets that can capture the richness and diversity of traditional dance forms from different regions and cultures [6]. This paper introduces TraDance, a publicly available benchmark dataset specifically curated for traditional Greek dances. TraDance comprises 1.4 hours of 3D dance motion videos featuring the performances of five experienced dancers. The dataset encompasses a total of 3494 annotated dance steps, representing five distinct traditional dance genres. To ensure accurate capture, TraDance was recorded using RGB-D camera.

The TraDance dataset is unique in several respects. First, it covers a range of traditional Greek dances that are less well-known outside Greece, including the Tapinos, Gikna and Dilbera from the area of Thrace, Sta Dyo from Epirus and Mazemenos from the Lesvos island. These dances are characterized by the repetition of simple steps throughout the duration of the dance track, making them ideal for analysis and recognition tasks. Second, the TraDance dataset includes data from multiple dancers, which increases its diversity and impartiality. Third, the dataset was captured using an RGB-D camera, which provides richer visual cues and can help overcome some of the limitations of 2D cameras, such as occlusion and lighting changes. Finally, the dataset includes manual annotations of dance steps, which were defined based on expert knowledge and analysis of the videos.

To evaluate the quality of the TraDance dataset, we conducted extensive experiments using the popular FeedForward Neural Networks (FFNNs) as our machine learning algorithm that is commonly used in computer vision. We evaluated different feature representations, including raw RGB-D data, skeleton data, and a combination of both. Our results show that TraDance is a reliable and challenging dataset that can be used to develop and compare different recognition and analysis methods. We also provide baseline performance for this algorithm and feature representation, which can serve as a reference for future studies.

The remainder of this paper is organized as follows. In Section II, we review related work in the area of dance recognition and analysis. Section III describes the TraDance dataset, including its acquisition, processing and annotation. In Section IV, we outline our evaluation methodology and report the results of our experiments. Finally, in Section V, we conclude the paper and discuss directions for future research.

## II. Related Work

Our present approach focuses on dataset creation to recognition of the human skeleton for the various dance movements. Initially, the work [7] presents a system that uses machine learning to generate 3D dance animations that are synchronized with music. The system uses a dataset called AIST++, which contains motion capture data of professional dancers performing in various musical genres. Then, the machine learning model is trained on this dataset and can generate new dance sequences in real-time based on the input music. Finally, the system is evaluated through user studies, and the results show that it can generate visually appealing and musically synchronized dance animations. Subsequently, authors of [8] explore the application of convolutional neural networks (CNN) to studying morphological transitions in advanced dance movements. In order to study dance motions and pinpoint the essential parts of the body that contribute to the execution of certain movements, the authors suggest a method that blends biological image visualization technology with CNN. The proposed method was tested on a group of expert dancers, and the results demonstrate that it is a useful instrument for assessing complex dance moves and pinpointing the corresponding bodily modifications.

Moreover, based on spatiotemporal features extracted from motion capture data, the authors of [9] recognised different dance styles. The authors utilize the Laban Movement Analysis framework to extract features related to the dancers' movements, such as the quality and effort of the movements. Furthermore, the authors of [10] introduced a novel methodology outlined in their work for performing 3D scans of intangible cultural heritage. Their research focused specifically on the Lazgi dance, a traditional dance originating from Dagestan, Russia. The authors describe the steps involved in capturing the dance in 3D using motion capture technology and explain how the resulting data can be used to preserve and promote the cultural heritage of the dance. The proposed methodology can potentially be applied to other forms of intangible cultural heritage, such as music, theatre, and oral traditions, to document and share these valuable cultural assets for future generations. Similarly, the authors [11] explore using immersive technologies, such as virtual reality, to protect and promote intangible cultural heritage. The paper presents various case studies of immersive experiences based on intangible cultural heritage, such as a virtual tour of a historical site, a virtual reality dance performance, and an augmented reality mobile app for exploring cultural landscapes. Ultimately, the authors conclude that immersive technologies have the potential to play a significant role in the preservation and promotion of intangible cultural heritage.

Regarding the recognition of Greek dances, which is an interesting issue of our work, some related works research this dance. Thus, the paper [12] presents a computer vision-based system for recognizing Greek folk dances using a Bag of Words (BoW) approach. The proposed system uses a set of visual features extracted from video frames, such as color and texture, and applies clustering and classification techniques to recognize different dance types. The system was evaluated on a dataset of Greek folk dance videos and achieved a high accuracy rate in recognizing different dance types. The proposed system can be used for automated dance recognition in cultural events, competitions, and therapy applications. Additionally, the research approach [13] describes a comparative study of various deep learning techniques for automatically recognising Greek folk dances from video recordings. The authors used a dataset consisting of video recordings of five different Greek folk dances and evaluated the performance of different deep learning models, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and combinations of both. The results showed that a combination of CNNs and RNNs performed an accuracy of over 95%, providing the ability of automated systems development for dance analysis and education.

Finally, the approach [14], which follows similar techniques to our methodology, proposes a system for automatically identifying and classifying dance poses in choreography using convolutional neural networks (CNNs). The system uses motion capture data to extract the 3D coordinates of the dancers' body joints and creates a dataset of choreographic poses. The CNN is trained on this dataset to learn the visual features of each pose and then used to classify new poses in real-time. The proposed system has the potential to assist choreographers in creating and refining dance routines and to aid in the analysis and comparison of different dance styles.

## III. The TraDance Dataset

The main objective of this work is to design a high-density, scalable and balanced dataset depicting the dance movements of Greek traditional dances. The TraDance dataset consists of 1.4 hours RGB-D video recordings, corresponding to 3.494 dance steps. Videos were captured at the Philoproodi Union in Xanthi, Greece. Five different traditional genres were recorded. Five dancers with varying levels of experience performed each dance multiple times. The camera's position was fixed during recordings, and the dancers performed their movements in straight and curving lines.

### A. Dances Description

Greece has a wide spectrum of folk dances which is characterized by diversity [15]. The majority of dances are circular. In this dataset, we have worked with the following five traditional dances:

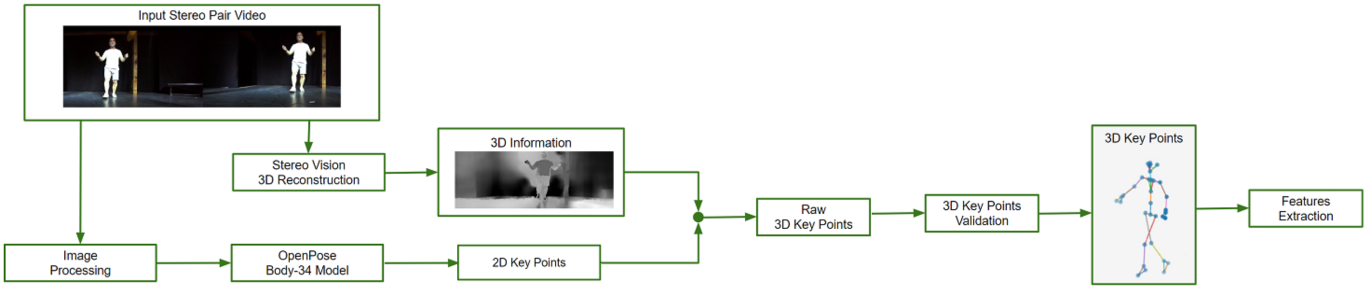


Fig. 1. Input video processing to obtain human body features.

**Mazomenos**: This dance comes from Lesvos Island. It is performed in six steps; (1) Open Right Leg; (2) Cross Left Leg over Right; (3) Uncross Legs; (4) Left Leg Up; (5) Left Leg Down; (6) Right Leg Up.

**Gikna**: It is a wedding dance celebration that occurs within the expansive region of Thrace [16]. The grip of the hands is from the palms, stretched down, and they perform back and forth movements. It has the same six steps as the Mazomenos dance.

**Dilbera**: A Thracian dance from the village of Metaxades in the prefecture of Evros. It is one of the main erotic dances that were said in a row, every Sunday, in the village square during the winter season, until Easter. It consists of eight steps; (1) Open Right Leg; (2) Cross Left Leg; (3) Uncross Legs; (4) Cross Left Leg; (5) Uncross Legs; (6) Close Left Leg; (7) Left Leg Forward; (8) Left Leg Back.

**Tapinos**: This dance also comes from the region of Thrace. It was the first dance done after the wedding ceremony, led by the bride. It is made up of six slow and modest steps with the arms swung forth and back; (1) Open Right Leg; (2) Cross Left Leg; (3) Uncross Legs; (4) Left Leg Forward; (5) Left Leg Back; (6) Right Leg Back.

**Sta Dyo**: Dance from the region of Epirus in eight steps. The grip of the hands is with the elbows bent; (1) Cross Left Leg; (2) Uncross Legs; (3) Cross Left Leg; (4) Uncross Legs; (5) Left Leg Back; (6) Open Right Leg.

### *B. Capturing postures*

In order to collect data from the dancers, we used a Stereolabs ZEDs camera [17], which is a stereoscopic device equipped with a pair of cameras that are mounted with co-planar optical planes and co-linear sensor bases, to gather data from the dances. Our recording setup (Figure 1) enabled us to capture stereoscopic video and process it to generate a 3D model of the dancer. The two cameras were separated horizontally by a distance of 12 cm, which allowed them to capture high-quality 3D video of the dancer and calculate depth and motion by analyzing the displacement of pixels between the left and right images. Based on the stereo matching principle, given a pair of stereo images, the pixels in the left image are matched with their corresponding pixels in the right image, and the depth value of each pixel is calculated using the camera parameters [18]. The ZED2 camera employs neural networks to obtain depth information and convert it into a 3D point cloud. A point cloud is a set of 3D points that represent the exterior surface and color information of a scene, and it stores data in four channels using a 32-bit float for each channel. The first three channels are used for the 3D point coordinates (x, y, and z), while the fourth channel is used to store color information. The camera uses the SVO proprietary file format to save data on a disk.

ZED2 camera has a depth range of 0.2–20 meters, a maximum resolution of 4416 x 1242 pixels, and can capture footage at a maximum frame rate of 100 Hz. The ZED2 software development kit (SDK) employs the NVIDIA library CUDA-Compute Unified Device Architecture to perform rapid artificial intelligence and computer vision tasks on a graphics processing unit, which allows the camera to extract depth information quickly and efficiently. The SDK also allows the detection and tracking of human body joints in real-time, similar to other time-of-flight cameras like Kinect [19] but with higher resolution (up to 2K) and better skeleton tracking.

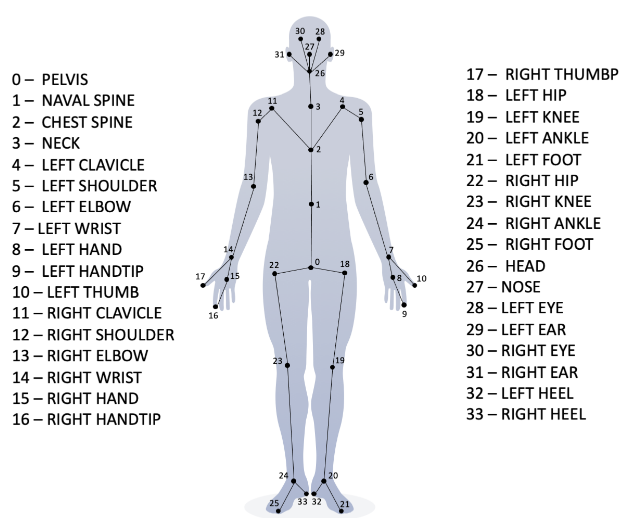


Fig. 2. The selected body format contains 34 keypoints following this configuration.

In an effort to capture the rich visual information and depth cues of the dance movements in the TraDance dataset, we utilized a stereoscopic camera setup with specific settings. The cameras were configured to record the dance performances at a frame rate of 60 frames per second (fps), ensuring smooth and detailed motion capture. The resolution setting for each camera was set to 1280x720 pixels, resulting in high-definition video capture. The stereoscopic setup allowed for the acquisition of synchronized videos from two different viewpoints, simulating the binocular vision of human observers. This setup enhanced the spatial perception of the dance movements, enabling the capture of fine details and depth information.

### C. Data processing

The 3D geometric information was obtained by processing the recorded SVO files. Through the ZED2 SDK interface, we extracted the 3D joints of the human skeleton, which were then utilized to create the choreographic representation. We selected to use the 34 keypoint formats, which were arranged based on the configuration depicted in Figure 2.

In the following, let us denote $J_k = (x_k, y_k, z_k)$ the $k$-th body joint out of $M$ available. In our case, $M = 34$. Variables $x_k, y_k, z_k$ indicate the $x, y, z$ coordinates with respect to a reference origin set by the ZED2 camera. Before transforming the coordinates $J_k$ to a new local coordination system, we had to separate the choreography into distinct dance steps. To achieve this, we should find the time points where each dance step starts and ends. The music beat of the song that is playing along with the choreography is a reliable tool for detecting these time points.

To synchronize the music with the dancer's movements captured by the video recording, we used a beat alignment metric proposed in [7]. The extraction of music beats was carried out using the Librosa audio processing toolbox [20], while the kinematic beats were computed by identifying the local minima of the kinetic velocity, as illustrated in Figure 3. The Beat Alignment score was calculated, as the average distance between every kinematic beat and its nearest music beat, by using the following equation:

$$BeatAlign = \frac{1}{n}\sum_{i=1}^{n} exp(-\frac{min_{\forall t_j^y \in B^y} \|t_i^x - t_j^y\|^2}{2\sigma^2}), \quad (1)$$

where $B^x = \{t_i^x\}$ is the kinematic beat and $B^y = \{t_j^y\}$ the music beats. We also set the normalisation parameter $\sigma = 3$, since our FPS during video recording was 60. By sliding the kinematic beats over the music beats, the best matching was found at frames where the Beat Alignment value had the lowest value.

### D. Pose Normalization

Differences in the width and height of different dancers can cause inconsistent results predictions. To ensure that our analysis structure is consistent and predictable, the resulting keypoints for each person can be represented as a vector in high dimensional space. We then use $L2$ normalisation to scale the vector to have unit magnitude [21]. Moreover, we used translation and rotation transformations to place the coordinates of keypoints for each step at the origin of the 3D axis reference system. We selected the middle point between hips as the $(0, 0, 0)$ point.

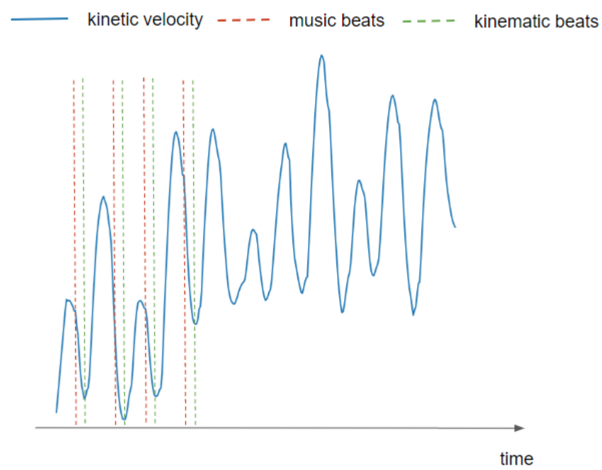


Fig. 3. Beats alignment between music and dancer.

These transformations ensure that all the keypoints during detection are consistent with each other, independent of the width and height of the human. After carrying out the normalization process, we are left with vectors of pose data that are of unit norm. This will be useful when we compare pose data from various body shapes and sizes. The 34-Dimensional normalized pose vector can now be compared with a pre-recorded normalized reference vector to obtain the matchings.

### E. Dataset annotation

The dance steps in the TraDance dataset were manually annotated by an experienced dance teacher with extensive knowledge of traditional Greek dances. The annotation process involved carefully analyzing each video sequence and assigning a unique identifier to every distinct dance step performed by the dancers. The dance teacher was well-versed in the nuances and intricacies of traditional Greek dances, allowing for accurate identification and annotation of the various steps.

During the annotation process, the dance teacher closely examined the movements and gestures exhibited by the dancers in the videos. A distinct identifier was assigned for each step observed, capturing its unique characteristics and defining its boundaries within the dance sequence. This manual annotation approach ensured the accuracy and reliability of the dance step labels in the TraDance dataset, as it leveraged the expertise and domain knowledge of the dance teacher.

By incorporating manual annotation, we aimed to provide a comprehensive and detailed understanding of the dance steps performed in the traditional Greek dances captured in the dataset. These annotations serve as a valuable resource for researchers and practitioners, enabling them to analyze and classify the specific movements and sequences within the dances accurately. The manual annotation process reinforces the reliability and authenticity of the TraDance dataset, contributing to its usefulness in advancing research and devel-

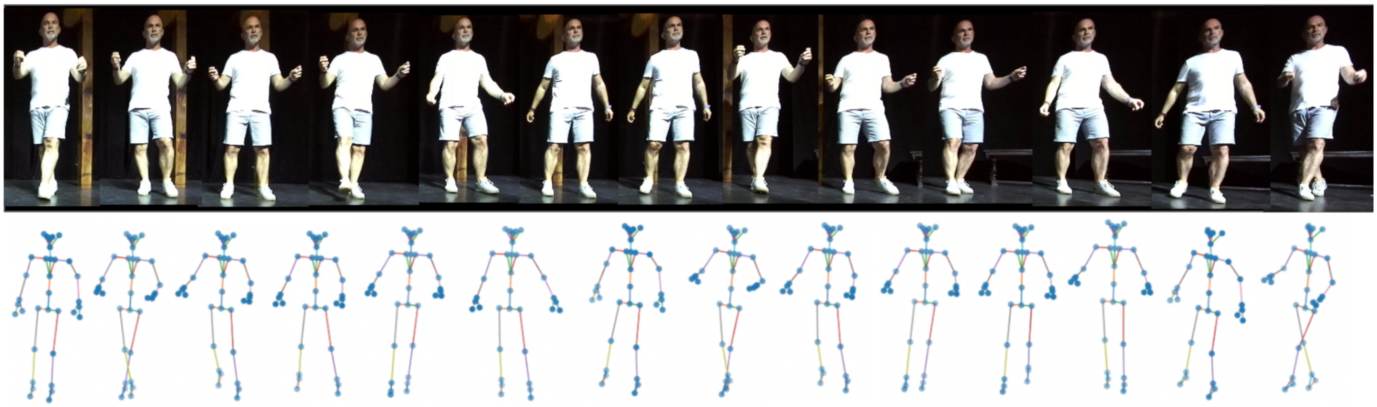

Fig. 4. An example of key primitives of the Greek folk dance Gikna

opment in the field of traditional Greek dance analysis and recognition.

All raw data files containing video information, keypoints coordinations, and annotation files are readily available for download on our dedicated web page: https://robotics.pme.duth.gr/tradance. We have also provided comprehensive technical description details to facilitate a thorough understanding of the dataset's structure, format, and organization.

## IV. Experimental Validation

To evaluate the performance of the TraDance dataset and its suitability for dance recognition tasks, we conducted experimental validation using a Feed-Forward Neural Network (FFNN) approach. The experiments involved training the NN model using a cross-validation strategy to ensure robustness and reliable performance assessment.

Our experimental setup divided the dataset into five subsets, each corresponding to a different dancer. For each iteration of the cross-validation process, we trained the NN model using data from four dancers and reserved the data from the remaining fifth dancer for testing. This approach allowed us to assess the generalization capability of the trained model by evaluating its performance on unseen dancers.

During training, we employed various architectural configurations of the FFNN, optimizing hyperparameters such as the number of hidden layers, the number of neurons in each layer, and the learning rate. We utilized standard training techniques, such as backpropagation and stochastic gradient descent, to optimize the model parameters and minimize the classification error.

The results of our experimental validation were highly promising, demonstrating the efficacy of the TraDance dataset for dance recognition. The accuracy of the NN model consistently surpassed 95% across the different cross-validation iterations, indicating the dataset's ability to capture distinctive features and patterns of traditional Greek dance steps. The high accuracy percentage underscores the dataset's quality and the potential for developing robust and reliable dance recognition algorithms using the TraDance dataset.

Table I summarizes the experimental results, showcasing the achieved accuracy percentages for each cross-validation iteration. The consistently high accuracy values validate the effectiveness of the TraDance dataset and its suitability for dance recognition tasks. These results highlight the dataset's potential for advancing research in traditional Greek dance analysis, facilitating the development of innovative algorithms, and fostering deeper insights into the recognition and classification of intricate dance movements.

TABLE I
EXPERIMENTAL CROSS VALIDATION RESULTS FOR EACH DANCE BY USING THE FFNN

| Dance | Averaged Accuracy (%) |
|---|---|
| Mazomenos | 96.5 |
| Gikna | 92.6 |
| Dilbera | 96.4 |
| Tapinos | 99.1 |
| Sta Dyo | 89.9 |
| Average | 94.9 |

Furthermore, we generated confusion matrices for each dance genre to gain deeper insights into the classification performance of the FFNN model. The confusion matrices visually represent the classification results, showing the frequency of correct and incorrect predictions for each dance class. Please refer to Figure 5 for the confusion matrices corresponding to the five dance genres present in the TraDance dataset.

The experimental validation of the TraDance dataset using a FFNN model with cross-validation demonstrated its reliability and effectiveness for dance recognition tasks. The dataset's meticulous manual annotation of dance steps and stereoscopic camera settings provided a rich and comprehensive resource for developing accurate and robust dance recognition algorithms. The high accuracy percentages achieved through the experiments, along with the detailed confusion matrices, validate the dataset's quality and its potential for advancing research in traditional Greek dance analysis and recognition. The TraDance dataset, along with the accompanying confusion matrices, stands as a valuable tool for researchers and practitioners seeking to explore the fascinating world of traditional

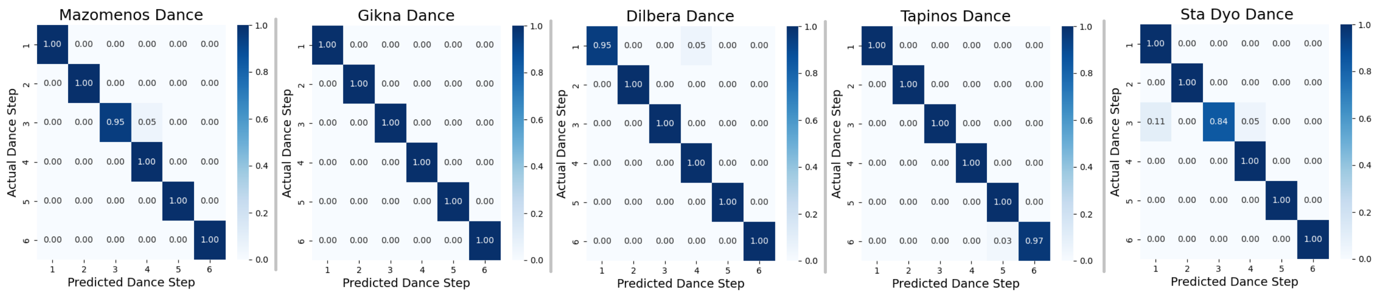


Fig. 5. Confusion Matrices results for the five dances, tested with a FF-NN.

Greek dances and leverage the power of machine learning for their analysis and preservation.

## V. Conclusion

This paper introduces TraDance, a novel benchmark dataset for traditional Greek folk dances. The dataset offers a comprehensive collection of 3D dance motion videos and detailed annotations, specifically recorded for research purposes. It represents the first dataset of its kind for traditional Greek folk dances, providing researchers with a valuable resource for analysis, recognition, and preservation efforts. Through extensive experimental evaluations, we have demonstrated the quality and reliability of the TraDance dataset. Our evaluations using established machine learning algorithms consistently achieved high accuracy percentages affirming the dataset's feasibility and robustness for research purposes. TraDance opens up new possibilities for researchers and practitioners interested in traditional Greek dances. The dataset enables the development and evaluation of innovative algorithms, advancing our understanding of dance forms, and contributing to the preservation of cultural heritage.

## Acknowledgment

This research was supported by European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation under the call Research - Create - Innovate project T2EDK-04800 "TraMIR: Traditional Musical Instruments Room".

## References

[1] R. Poppe, "A survey on vision-based human action recognition," *Image and vision computing*, vol. 28, no. 6, pp. 976–990, 2010.

[2] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal, "Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions," in *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009, pp. 1932–1939.

[3] Z. Luo, S. A. Golestaneh, and K. M. Kitani, "3d human motion estimation via motion compression and refinement," in *Proceedings of the Asian Conference on Computer Vision*, 2020.

[4] J. Wang, Z. Liu, and Y. Wu, *Human action recognition with depth cameras*. Springer, 2014.

[5] M. Raptis, I. Kokkinos, and S. Soatto, "Discovering discriminative action parts from mid-level video representations," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 1242–1249.

[6] D. Weinland, R. Ronfard, and E. Boyer, "A survey of vision-based methods for action representation, segmentation and recognition," *Computer vision and image understanding*, vol. 115, no. 2, pp. 224–241, 2011.

[7] R. Li, S. Yang, D. A. Ross, and A. Kanazawa, "Ai choreographer: Music conditioned 3d dance generation with aist++," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 13 401–13 412.

[8] R. Zhang, "Analyzing body changes of high-level dance movements through biological image visualization technology by convolutional neural network," *The Journal of Supercomputing*, vol. 78, no. 8, pp. 10 521–10 541, 2022.

[9] S. Dewan, S. Agarwal, and N. Singh, "Spatio-temporal laban features for dance style recognition," in *2018 24th international conference on pattern recognition (ICPR)*. IEEE, 2018, pp. 2911–2916.

[10] M. Skublewska-Paszkowska, P. Powroznik, J. Smolka, M. Milosz, E. Lukasik, D. Mukhamedova, and E. Milosz, "Methodology of 3d scanning of intangible cultural heritage—the example of lazgi dance," *Applied Sciences*, vol. 11, no. 23, p. 11568, 2021.

[11] T. Pistola, S. Diplaris, C. Stentoumis, E. A. Stathopoulos, G. Loupas, T. Mandilaras, G. Kalantzis, I. Kalisperakis, A. Tellios, D. Zavraka *et al.*, "Creating immersive experiences based on intangible cultural heritage," in *2021 IEEE International Conference on Intelligent Reality (ICIR)*. IEEE, 2021, pp. 17–24.

[12] E. Fotiadou, I. Kapsouras, N. Nikolaidis, and A. Tefas, "A bag of words approach for recognition of greek folk dances," in *Proceedings of the 9th Hellenic Conference on Artificial Intelligence*, 2016, pp. 1–4.

[13] G. Loupas, T. Pistola, S. Diplaris, K. Ioannidis, S. Vrochidis, and I. Kompatsiaris, "Comparison of deep learning techniques for video-based automatic recognition of greek folk dances," in *MultiMedia Modeling: 29th International Conference, MMM 2023, Bergen, Norway, January 9–12, 2023, Proceedings, Part II*. Springer, 2023, pp. 325–336.

[14] N. Bakalos, I. Rallis, N. Doulamis, A. Doulamis, E. Protopapadakis, and A. Voulodimos, "Choreographic pose identification using convolutional neural networks," in *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*. IEEE, 2019, pp. 1–7.

[15] K. Stiga and E. Kopsalidou, "Music and traditions of thrace (greece): a trans-cultural teaching tool," *DEDiCA Revista de Educação e Humanidades (dreh)*, no. 3, pp. 145–164, 2012.

[16] E. Filippidou, M. Koutsouba, V. Lalioti, and V. Lantzos, "Cybernetic approach of the" k'na" dance: The construction of ethnic and national identity in nea vyssa, thrace, greece," *Mediterranean Journal of Social Sciences*, vol. 9, no. 2, p. 17, 2018.

[17] "Stereolabs zed2 camera," https://www.stereolabs.com/, accessed: 2023-06-01.

[18] F. Wang, V. Mohan, A. Thompson, and R. Dudley, "Dimension fitting of wheat spikes in dense 3d point clouds based on the adaptive k-means algorithm with dynamic perspectives," in *2020 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*. IEEE, 2020, pp. 144–148.

[19] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *CVPR 2011*. Ieee, 2011, pp. 1297–1304.

[20] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," in *Proceedings of the 14th python in science conference*, vol. 8, 2015, pp. 18–25.

[21] A. Y. Ng, "Feature selection, l 1 vs. l 2 regularization, and rotational invariance," in *Proceedings of the twenty-first international conference on Machine learning*, 2004, p. 78.