FOLK DANCE PATTERN RECOGNITION OVER DEPTH IMAGES ACQUIRED VIA KINECT SENSOR

Eftychios Protopapadakis [a, *] Athina Grammatikopoulou [b], Anastasios Doulamis [a], Nikos Grammalidis [b]

[a] School of Rural and Surveying Engineering, National Technical University of Athens,
9 Iroon Polytechneiou str, 15780 Zografou, Greece – eprotopapadakis@isc.tuc.gr, adoulam@cs.ntua.gr
Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece – {agramm, ngramm} @iti.gr

Commission II

KEY WORDS: Intangible Cultural Heritage, Folk Dances, Depth Camera, Unsupervised Clustering, Skeleton data

ABSTRACT:

The possibility of accurate recognition of folk dance patterns is investigated in this paper. System inputs are raw skeleton data, provided by a low cost sensor. In particular, data were obtained by monitoring three professional dancers, using a Kinect II sensor. A set of six traditional Greek dances (without their variations) consists the investigated data. A two-step process was adopted. At first, the most descriptive skeleton data were selected using a combination of density based and sparse modelling algorithms. Then, the representative data served as training set for a variety of classifiers.

1 INTRODUCTION

Intangible Cultural Heritage (ICH) and its perpetual preservation, is an intriguing domain that attracts both the scientific community and the general public. The main challenges involved are associated with the complex structure of ICH; i.e. its dynamic nature, the interaction among the objects and the environment, as well as emotional elements, such as the dancers' expressions and style.

Folk dances are important to ICH; they are directly connected to local culture and identity (Shay and Sellers-Young, 2016). Hence, the preservation of folk dances is a basic requirement, since the history and style of each folk dance will be readily available to the public through a system that includes descriptive information, videos, movement and 3D modelled data relevant to it.

Different songs of the same music genre are usually represented choreographically by the same set of signature moves and gestures. This holds especially true in case of traditional folk music and dances that abide by a stronger sense of structure. This structure can be exploited in order to index and classify the key elements that compose each dance. A properly designed database of characteristic dance instances to compare against, will allow for a number of useful applications to emerge.

Dance recognition, with semantic information such as genre, provenance, and correlation, symbolism, all the way to difficulty level, tutorials and even pertinent advertising such as museums, upcoming performances, dance studios and discography, will be available through a single snapshot or short video. The technology required to achieve the aforementioned goal is within reach, since depth cameras and classification algorithms that lie in the center of these applications are not only available but capable enough to achieve the aforementioned goals.

Arguably, the prevalent and most widespread depth camera is the Microsoft (MS) Kinect (Microsoft, 2016), however mobile variants of this technology, such as Project Tango by Google (Google, 2016) are already market available, and can be employed.

The focus of this paper is the application of segmentation and classification algorithms to Kinect captured depth images and videos of folkloric dances in order to identify key movements and gestures, compare them against database instances and determine the dance genres they represent, as well as to provide helpful metadata. Then, the goal is to identify the geometric structures of the dance and then to classify the dances with respect to common geometric features they share.

2 RELATED WORK

There are several examples in the literature which present applications that exploit Kinect 3D data of human (dance) movements' for recognition and classification purposes. In the work (Gianaria et al., 2014), gait analysis is performed using skeletal data provided by Microsoft Kinect sensors and a set of physical and behavioural features is defined, in order to identify the more relevant parameters for gait description. The aim of this work is the gait characterization and people recognition using SVM classification.

The authors of (Raptis et al., 2011) describe a gesture classification system for skeletal wireframe motion. A classifier was designed and trained to recognize certain gestures, among several dozen, in real-time and with high accuracy. In another work (Zanfir et al., 2013), a simple non-parametric Moving Pose framework is proposed, for low-latency human action and activity recognition using a modified kNN classifier. Furthermore, a method to recognize individual persons from their walking gait using 3D skeletal data from a MS Kinect device using the k-means algorithm is described in the work (Ball et al., 2012).

Ever since the introduction of the first Kinect, depth cameras have grown in importance and are widely used as low-cost

---

* Corresponding author

peripherals for several applications. The advantages of a depth camera is that it produces dense and reliable depth measurements, albeit over a limited range and offers balance in usability and cost. Kinect is the sensor of choice for such applications and will be employed to capture sets of dance moves and gestures in 3D space and in real time, resulting in a recorded sequence of points in 3D space for each joint at certain moments in time t.

In (Kitsikidis et al., 2014b), a methodology is proposed for dance learning and evaluation using multi-sensor and 3D gaming technology. The learners are captured during dancing, while an avatar visualizes their motion using fused input from multiple sensors. Motion analysis and a two-level Fuzzy Inference System (FIS) are applied, using as input low level skeletal data and high level motion recognition probabilities, for the evaluation of dancer's performance. In (Kitsikidis et al., 2015b), a 3D game environment for dance learning is presented, which is based on the fusion of multiple depth sensors data in order to capture the body movements of the user/learner. In addition, the system automatically assesses the learner's performance, by utilizing a combination of Dynamic Time Warping with a FIS, and provides feedback in a form of a score as well as instructions from a virtual tutor in order to promote self-learning.

In (Kitsikidis et al., 2014a), improved robustness of skeletal tracking is achieved by using sensor data fusion to combine skeletal tracking data from multiple sensors. The fused skeletal data is split into different body parts, which are then transformed to allow view invariant posture recognition. For each part, a posture vocabulary is generated by performing k-means clustering on a large set of unlabelled postures. Finally, body part postures are combined into body posture sequences and Hidden Conditional Random Fields classifier is used to recognize motion patterns, e.g. dance figures. In (Kitsikidis et al., 2015a), a skeletal representation of the dancer is again obtained by using data from multiple depth sensors. Using this information, the dance sequence is partitioned first into periods and subsequently into patterns. Partitioning into periods is based on observing the horizontal displacement of the dancer while each period is subsequently partitioned into patterns by means of training an exemplar-based Hidden Markov Model that classifies frames to exemplars representing HMM states.

In (Dimitropoulos et al., 2016), human action recognition is treated as a special case of the general problem of classifying multidimensional time-evolving data in dynamic scenes. To solve detect correlations between channels, a generalized form of a stabilized higher-order linear dynamical system (sh-LDS) and the multidimensional signal is represented as a third order tensor. Each multidimensional signal is represented as a cloud of points on the Grassmann manifold and a codebook is created by identifying the most representative points to be used in classification by applying a bag-of-systems approach.

3 PROPOSED METHODOLOGY

Dance recognition via pattern identification is the scope of this paper. In particular, we try to identify Greek folk dances by matching recorded sequences to a database of characteristic dance instances; i.e. motion and rotation 3D points plus the time. Therefore, the problem at hand entails to a conventional machine learning paradigm: create a set of appropriate features and train the classifiers.

A meaningful train set should contain data spanning as much as possible the feature space, without redundant information or almost identical entries. In our case, raw data space lies in $\mathbb{R}^{7 \times 1}$,

since we have the 3D position of various body parts, plus the corresponding rotations (four in total), for a set of n consecutive frames.

At first, the structure of the data had to be captured by creating clusters. Then, for each of the identified clusters, find the smaller possible set that can accurately describe the remaining cluster data. Such sparse modelling approach has multiple advantages; non redundant information, reduced storage space, faster pre-processing, etc. All the above steps can be done without supervision. Once the process is complete, the remaining data can either serve as a reference point for future comparison with other motion data, in order to identify the dance patterns or serve as a training data set for a variety of classifiers.

The aforementioned method leads to dance clustering through the comparison of properly captured dance movement sets by a depth sensor and sets already recorded and stored to a relevant motion capture database. The high-level 3D representations of dance movements obtained from the low-cost consumer-level depth sensor, will then be fed to an unsupervised clustering algorithms in order to produce meaningful instances for comparison purposes.

3.1 Sensors used

The Microsoft Kinect II is currently one of the most advanced motion sensing input device that is available to the public. It is a physical device with depth sensing technology, built-in color camera, infrared (IR) emitter, and microphone array, which projects and captures an infrared pattern to estimate depth information. Based on the depth map data, the human skeleton joints are located and tracked via the Microsoft Kinect II for Windows SDK ("Kinect - Windows app development," 2017)

More specifically, the Microsoft Kinect II sensor can achieve real-time 3D skeleton tracking, while at the same time it is relatively cheap and easy to setup and use. The tracked skeleton consists of twenty five joints with each one to include the 3D position coordinates, its rotation and a tracking state property: "Tracked", "Inferred", and "UnTracked" (Webb and Ashley, 2012). Moreover, the sensor can work in dark and bright environments and the capture frame rate is 30fps. On the other hand, there are some limitations that should be taken into account: it is designed to track the front side of the user and as a result the front and back side of the user cannot be distinguished, and that the movement area is limited (approximately 0.7–6 m).
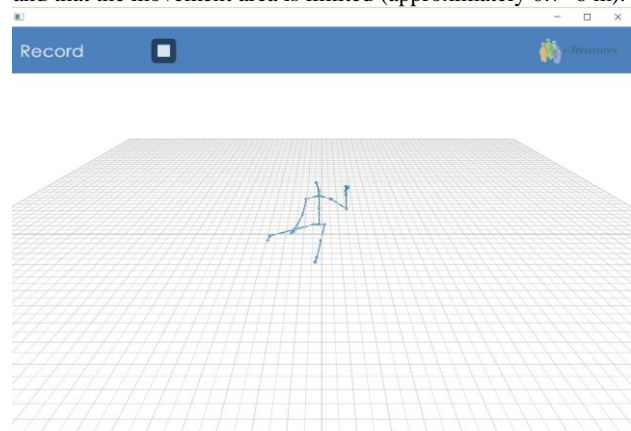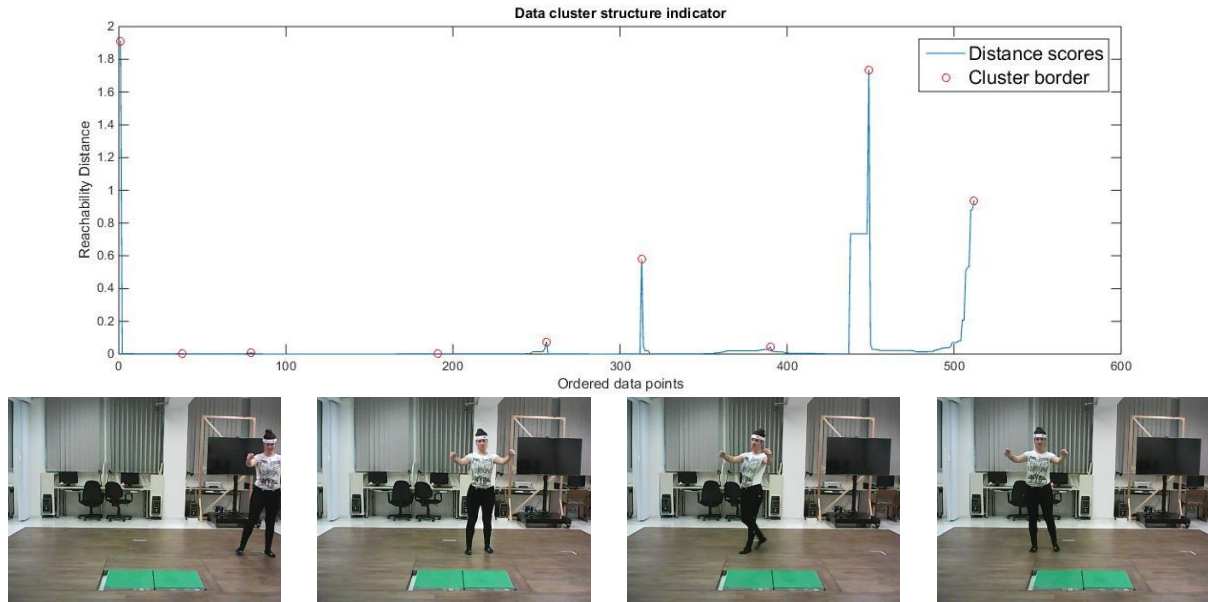

Figure 1. ITGD module's record screen

Figure 2. Identifying clusters within data, using as joint of reference the right foot. Positions and rotation data are compared among all dance frames. Peaks (upper image) are indicators for a new cluster; four cluster border corresponding frames are, also, illustrated (lower image).

3.2 Sampling algorithms

The main purpose of data sampling is the selection of appropriate representative samples in order to provide a good training set and, thus, improve the classification performance of risk assessment models. The main purpose of data sampling is the selection of appropriate representative samples in order to provide a good training set and, thus, improve the classification performance of risk assessment models. The most important factor in data selection is the definition of distance function. For any two given data points x_i and x_j, $x \in \mathbb{R}^m$ let $d(x_i, x_j)$ denote the distance between them. In order to compute the distance, let $A \in R^{m \times m}$ be a symmetric matrix and the distance measure defined as:

$$d_A(x_i, x_j) = \sqrt{(x_i - x_j)^T A (x_i - x_j)} \quad (1)$$

All the proposed approaches are Euclidean based (i.e. $A = I$).

3.2.1 OPTICS algorithm
Ordering Points to Identify the Clustering Structure (OPTICS) is an algorithm for finding density-based clusters in spatial data (Ankerst et al., 1999); i.e. detect meaningful clusters in data of varying density. In order to do so, the points of the database are (linearly) ordered such that points which are spatially closest become neighbors in the ordering.

OPTICS requires two parameters: the maximum distance (radius) to consider (ε), and the number of points required to form a cluster ($MinPts$). A point p is a core point if at least $MinPts$ points are found within its ε –neighborhood, $N_\varepsilon(p)$. Once the initial clustering is formed, we may proceed with any sampling approach (e.g. random selection among clusters).

3.2.2 Sparse representative selection
In order to extract the most important, i.e. descriptive, data, the work of (Elhamifar et al., 2012) around sparse modeling, is employed. Sparse representative selection (*Sparse)* focuses on the identification of representative objects. Their work is summarized through the following formulation:

$$\min \lambda \|C\|_{1,q} + \frac{1}{2} \|X - XC\|_F^2 \quad (2)$$
$$s.t. \mathbf{1}^T C = \mathbf{1}^T$$

where X and C refer to data points and coefficient matrix respectively. This optimization problem can also be viewed as a compression scheme, where we want to choose a few representatives that can reconstruct the available data set.

3.3 Utilized classifiers

A set of well-known classifiers were applied in order to evaluate the detection rates, for various sets of input data.

3.3.1 k nearest neighbors
In pattern recognition, the k-nearest neighbors ($k$nn) algorithm is a non-parametric method used for classification (Bhatia and Vandana, 2010). An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors; it is therefore, a type of instance-based learning, where the function is only approximated locally and all computation is deferred until classification.

3.3.2 Classification trees
Decision tree learning uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. In classification tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Each internal (non-leaf) node is labeled with an input feature. The arcs coming from a node labeled with a feature are labeled with each of the possible values of the feature. Each leaf of the tree is labeled with a class or a probability distribution over the classes.

3.3.3 Artificial neural networks
Artificial neural networks are non-linear mapping structures, inspired by biological nervous systems, which are capable of machine learning and pattern recognition (Li et al., 2011). ANNs are universal approximators which however have multiple local minima (i.e. solutions), due to their structure; they are composed

from multiple hierarchical layers of interconnected nodes. Their structure consists of weights, biases and activation functions, imitating the real brain's neurons and synapses.

3.3.4 Support vector machines

Support vector machines (SVMs) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification analysis (Abe, 2010). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap (margin) that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the margin they fall on. The mappings used by SVM schemes are defined through a kernel function $k(x, y)$ selected to suit the problem. In our case we utilized linear and RBF kernels.

4 EXPERIMENTAL RESULTS

In our study in order to capture and record the performers' body motions, we used a motion capture system using one Kinect II depth sensor (Fig. 1) and the ITGD module, developed within the i-Treasures project (Dimitropoulos et al., 2014) by UMONS. The ITGD module enables the user to record and annotate motion capture data received from a Kinect sensor.

Figure 3. The dance capturing process. Image on the left demonstrates the sensor position. On the right, we can see the dancer while acting.

The recording process took place at the School of Physical Education and Sport Science of the Aristotle University of Thessaloniki. Six Greek traditional dances with a different degree of complexity were recorded. Each dance was performed by three dancers twice: The first time in a straight line and the second in a semi-circular curving line.

4.1 Dataset description

Data set consists of six different dances. Their execution was either in straight line or circle (table 1). Every dance is described by a set of consecutive image frames. Every frame, $I_i, i = 1, ..., n$, has a corresponding extensible mark-up language (XML) file with positions, rotations and confidence scores for 25 joints on the body (table 2), in addition to timestamps.

Investigated dances were:
1. Enteka (eleven): A dance, performed by both women and men, which is popular mainly in the large urban centers of Western Macedonia (Grevena, Kozani, Florina, Kastoria, etc.). The dance is performed freely as a street carnival dance, but also around the carnival fires. The dancers' hands during the dance move freely or are placed at the waist.
2. Kalamatianos: It is a popular Greek folkdance throughout Greece, Cyprus and internationally, often performed at many social gatherings worldwide. It is a circle dance performed in a counterclockwise rotation with the dancers holding hands. It's a twelve steps dance and the musical beat is 7/8.

3. Makedonikos: A circle dance, performed by both women and men, with a 7/8 musical beat. The basic pattern of dance is performed in twelve movements / steps. Therefore, it resembles the Kalamatianos dance in a great degree with the difference that it is a more joyous dance. It is popular in the region of Western and Central Macedonia.
4. Syrtos (2 beat): The Syrtos (2 beat) dance is organized in a quick (2 beat) rhythm. It is a circle dance, performed by both women and men mostly in the region of Pogoni of Epirus. In the past, the dance was performed separately by men and women, in one, two or more lines.
5. Syrtos (3 beat): Syrtos is one of the most popular dances throughout Greece and Cyprus. The Syrtos (3 beat) dance is organized in a slow (3 beat) rhythm. It is a line dance and a circle dance, performed by dancers (both women and men) in a curving line holding hands, facing right. It is widespread through Epirus, Western Macedonia, Thessaly, Central Greece and Peloponnese.
6. Trehatos (Running): A circle dance, performed by both women and men, which is danced in the village Neochorouda of Thessaloniki. The kinetic theme of the dance is composed of three different dance patterns. The first one resembles the Syrtos (3 beat) pattern, the second takes place once and connects the first and the second pattern, and the third one is characterized by intense motor activity.

Dance	Variation	Short name	Duration (frames)		
			D1	D2	D3
Enteka	Straight	Syrt_11_Str8	749	807	858
Kalamatianos	Circular	Kal_Circ	655	593	561
	Straight	Kal_Str8	304	378	455
Makedonitikos	Circular	Mak_Circ	424	582	409
	Straight	Mak_Str8	283	367	418
Syrtos 2	Circular	Syrt_2_Circ	608	543	352
	Straight	Syrt_2_Str8	623	639	334
Syrtos 3	Circular	Syrt_3_Circ	608	964	947
	Straight	Syrt_3_Str8	1366	678	511
Trehatos	Circular	Treh_Circ	991	723	443
	Straight	Treh_Str8	315	295	355

Table 1. List of available dances and their variations as well as their duration, depending on the dancer (Dx).

Body part/region	Joints tracked	
Head	Head	
Torso	SpineBase SpineMid SpineShoulder Neck	
Arms	ShoulderLeft ElbowLeft WristLeft HandLeft ShoulderRight ElbowRight	WristRight HandRight HandTipLeft ThumbLeft HandTipRight ThumbRight
Legs	HipLeft KneeLeft AnkleLeft FootLeft	HipRight KneeRight AnkleRight FootRight

Table 2. List of observed nodes (skeletal data) and their corresponding category
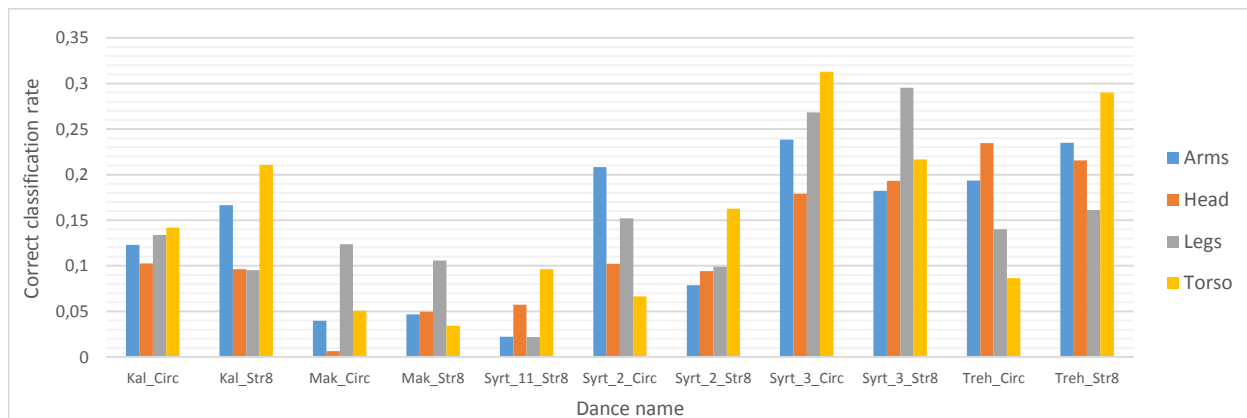
Figure 4. Body region impact on the classification accuracy over the folk dances. Results are average classification percentages for all the employed classifiers.

4.2 Feature extraction

At first the positions and rotation values for each frame, I_i, $i = 1, \dots, n$ of a dance, with n consecutive frames, are extracted. Thus, the dance is described by a matrix, $\boldsymbol{D}_i$, of size $b \times m \times n$, where b is the number of body joints (i.e. 25), m is the number of feature vectors (i.e. 3 coordinates plus 4 rotations), and n is the duration of the dance.

However, the dance is not a static act; the time dimension should be also considered. Therefore, we utilized the information of two consecutive frames, I_i and I_{i+1}. In the end, each dance, $\boldsymbol{D}_i$, was of size $b \times (2m) \times n - 1$. Prior to the representative selection step, data were normalized using minmax normalization.

4.3 Representative samples selection

A combination of Combination of OPTICS and SMRS algorithms was adopted. In this case, SMRS is performed to the sub-clusters obtained through the OPTICS algorithm. This approach is similar to the work of (Protopapadakis et al., 2014). It creates a small subset of representative samples from each cluster formed through the OPTICS algorithm.

The number of clusters, k, was defined using the rule: $k = \lceil \sqrt{N/2} \rceil$, where N denotes the number of available samples. The minimum number of data within a cluster, required by OPTICS, m_c, was defined as: $m_c = \lfloor N/k \rfloor$.

4.4 Algorithms setup

All algorithms were implemented in MATLAB except for the SVMs (Chang and Lin, 2011). In our case the $k$nn parameterization process considers the number of k nearest points, which was set as $k = 5$. Classification trees required no further parameterization. A feed forward network of two hidden layers was utilized. All activation functions were hyperbolic tangent and the training method was the back propagation method. Parameters d and σ where defined according to cross-validation accuracy scores.

4.5 Experimental results

It appears that the exploitation of raw data from two consecutive frames does not suffice for accurate folk dance recognition; neither body joint nor classifier appear dominant behaviour in terms of detection rates. All results correspond to frame detection rates; i.e. given two consecutive frames of any dance, we try to identify the dance, using joints position.

At first, the impact of the joint location is investigated (see fig. 4). In order to facilitate the illustration the 25 body joints were classified into body regions, as shown in table 2. Although torso or leg provide better average detection rates, it appears that no body parts / regions are dominant for all dance types.

Secondly, the impact of the classifier selection has been assessed (see fig. 5); for illustration purposes results are shown only for
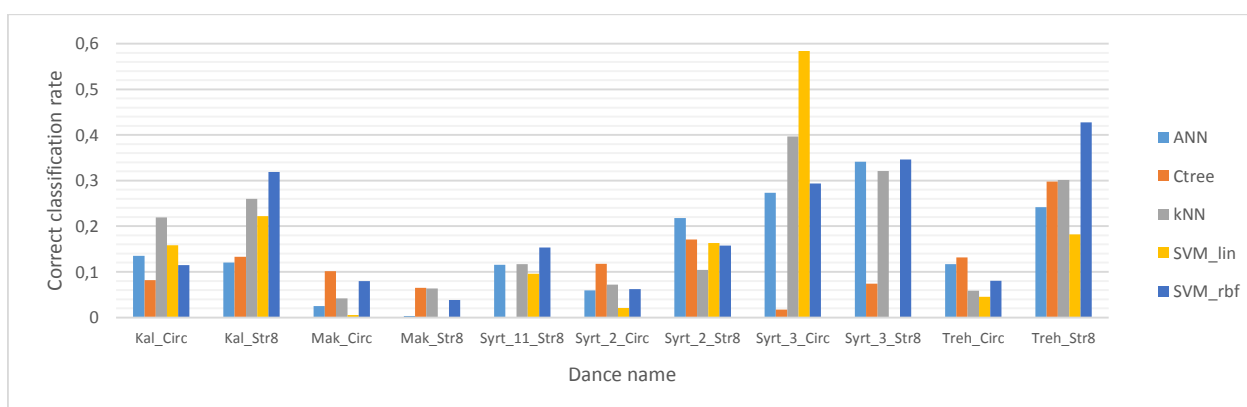

Figure 5. Classifier classification impact on dance identification, for torso joints. Results correspond to average correct classification percentages over torso region joints, as described in table 2.
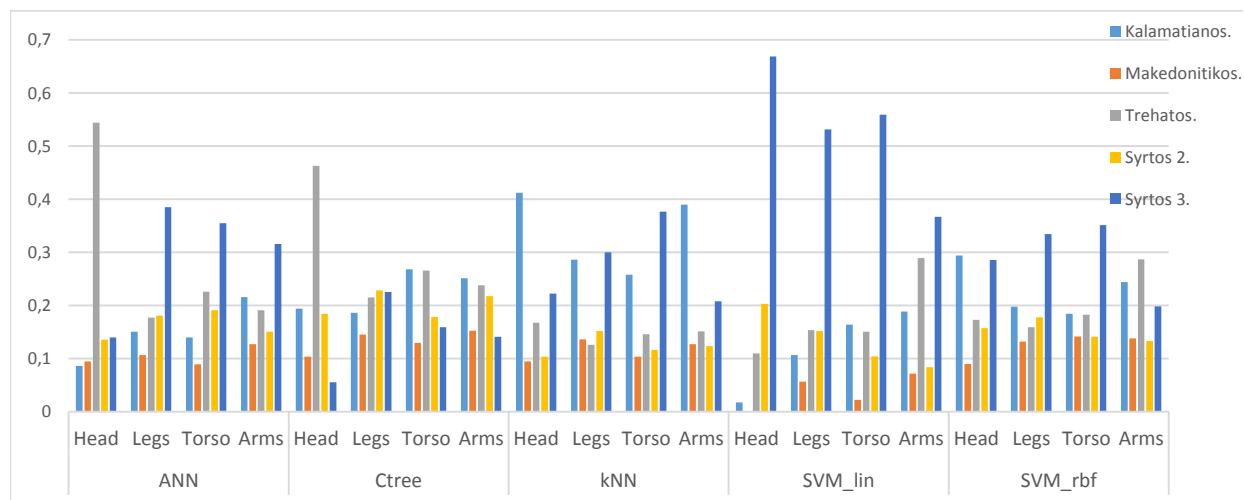
Figure 6. Dance category identification percentages for different joint groups. Results correspond to both straight line and semi-circular curving line trajectories.

torso joints group. No classifier achieved better performance than the rest in the dance recognition activity.

The combinatory detection rates, for dance categories, were, also, investigated (fig. 6). Depending on the dance, the importance of body region for the feature extraction varies. It is also intriguing that the arm and head related joints position contains meaningful information for the clustering. For instance, the Syrtos (3 beat) detection accuracy exceeds 65% using head joints and SVMs with linear kernel. Yet, for the Syrtos (2 beat) the same classifier-joints combination performs poorly; i.e. below 20%.

5 CONCLUSIONS

In this paper an investigation has been conducted, regarding the identification abilities of well-known classifiers, over folk dance identification. The impact of the body joint regions was also investigated. Analysis was based on raw data provided by a single Kinect II sensor. In total there was six Greek folk dances; most of them had two variations. The feature space was the coordinates and the rotations of the body joints, in pairs of consecutive frames, in order to incorporate the time dimension. Future work will focus on the exploitation of more complex feature extraction processes and classifiers.

ACKNOWLEDGEMENTS

REFERENCES

Abe, S., 2010. Support Vector Machines for Pattern Classification. Springer.

Ankerst, M., Breunig, M.M., Kriegel, H.-P., Sander, J., 1999. OPTICS: Ordering Points to Identify the Clustering Structure, in: Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data, SIGMOD '99. ACM, New York, NY, USA, pp. 49–60. doi:10.1145/304182.304187

Ball, A., Rye, D., Ramos, F., Velonaki, M., 2012. Unsupervised Clustering of People from "Skeleton" Data, in: Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction, HRI '12. ACM, New York, NY, USA, pp. 225–226. doi:10.1145/2157689.2157767

Bhatia, N., Vandana, 2010. Survey of Nearest Neighbor Techniques. ArXiv10070085 Cs.

Chang, C.-C., Lin, C.-J., 2011. {LIBSVM}: A library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 1–27.

Dimitropoulos, K., Barmpoutis, P., Kitsikidis, A., Grammalidis, N., 2016. Classification of Multidimensional Time-Evolving Data using Histograms of Grassmannian Points. IEEE Trans. Circuits Syst. Video Technol. PP, 1–1. doi:10.1109/TCSVT.2016.2631719

Dimitropoulos, K., Manitsaris, S., Tsalakanidou, F., Nikolopoulos, S., Denby, B., Kork, S.A., Crevier-Buchman, L., Pillot-Loiseau, C., Adda-Decker, M., Dupont, S., Tilmanne, J., Ott, M., Alivizatou, M., Yilmaz, E., Hadjileontiadis, L., Charisis, V., Deroo, O., Manitsaris, A., Kompatsiaris, I., Grammalidis, N., 2014. Capturing the intangible an introduction to the i-Treasures project, in: 2014 International Conference on Computer Vision Theory and Applications (VISAPP). Presented at the 2014 International Conference on Computer Vision Theory and Applications (VISAPP), pp. 773–781.

Elhamifar, E., Sapiro, G., Vidal, R., 2012. See all by looking at a few: Sparse modeling for finding representative objects, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1600–1607. doi:10.1109/CVPR.2012.6247852

Gianaria, E., Grangetto, M., Lucenteforte, M., Balossino, N., 2014. Human Classification Using Gait Features, in: Cantoni, V., Dimov, D., Tistarelli, M. (Eds.), Biometric Authentication, Lecture Notes in Computer Science. Presented at the International Workshop on Biometric Authentication, Springer International Publishing, pp. 16–27.

Kinect - Windows app development [WWW Document], 2017. URL https://developer.microsoft.com/en-us/windows/kinect (accessed 1.20.17).

Kitsikidis, A., Boulgouris, N.V., Dimitropoulos, K., Grammalidis, N., 2015a. Unsupervised Dance Motion Patterns

Classification from Fused Skeletal Data Using Exemplar-Based HMMs. Int. J. Herit. Digit. Era 4, 209–220. doi:10.1260/2047-4970.4.2.209

Kitsikidis, A., Dimitropoulos, K., Douka, S., Grammalidis, N., 2014a. Dance analysis using multiple Kinect sensors, in: 2014 International Conference on Computer Vision Theory and Applications (VISAPP). Presented at the 2014 International Conference on Computer Vision Theory and Applications (VISAPP), pp. 789–795.

Kitsikidis, A., Dimitropoulos, K., Uğurca, D., Bayçay, C., Yilmaz, E., Tsalakanidou, F., Douka, S., Grammalidis, N., 2015b. A Game-like Application for Dance Learning Using a Natural Human Computer Interface, in: Antona, M., Stephanidis, C. (Eds.), Universal Access in Human-Computer Interaction. Access to Learning, Health and Well-Being, Lecture Notes in Computer Science. Presented at the International Conference on Universal Access in Human-Computer Interaction, Springer International Publishing, pp. 472–482.

Kitsikidis, A., Dimitropoulos, K., Yilmaz, E., Douka, S., Grammalidis, N., 2014b. Multi-sensor Technology and Fuzzy Logic for Dancer's Motion Analysis and Performance Evaluation within a 3D Virtual Environment, in: Stephanidis, C., Antona, M. (Eds.), Universal Access in Human-Computer Interaction. Design and Development Methods for Universal Access, Lecture Notes in Computer Science. Presented at the International Conference on Universal Access in Human-Computer Interaction, Springer International Publishing, pp. 379–390.

Li, H., Liao, X., Li, C., Huang, H., Li, C., 2011. Edge detection of noisy images based on cellular neural networks. Commun. Nonlinear Sci. Numer. Simul. 16, 3746–3759. doi:10.1016/j.cnsns.2010.12.017

Protopapadakis, E., Doulamis, A., Matsatsinis, N., 2014. Semi-supervised Image Meta-filtering in Cultural Heritage Applications, in: Ioannides, M., Magnenat-Thalmann, N., Fink, E., Žarnić, R., Yen, A.-Y., Quak, E. (Eds.), Digital Heritage. Progress in Cultural Heritaage: Documentation, Preservation, and Protection, Lecture Notes in Computer Science. Springer International Publishing, pp. 102–110.

Raptis, M., Kirovski, D., Hoppe, H., 2011. Real-time Classification of Dance Gestures from Skeleton Animation, in: Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA '11. ACM, New York, NY, USA, pp. 147–156. doi:10.1145/2019406.2019426

Shay, A., Sellers-Young, B., 2016. Dance and Ethnicity. doi:10.1093/oxfordhb/9780199754281.013.38

Webb, J., Ashley, J., 2012. Beginning Kinect Programming with the Microsoft Kinect SDK. Apress.

Zanfir, M., Leordeanu, M., Sminchisescu, C., 2013. The Moving Pose: An Efficient 3D Kinematics Descriptor for Low-Latency Action Recognition and Detection. Presented at the Proceedings of the IEEE International Conference on Computer Vision, pp. 2752–2759.