A Game-like Application for Dance Learning using a Natural Human Computer Interface

Alexandros Kitsikidis[1], Kosmas Dimitropoulos[1], Deniz Uğurca[2], Can Bayçay[2], Erdal Yilmaz[2], Filareti Tsalakanidou[1], Stella Douka[3], and Nikos Grammalidis[1]

[1]Information Technologies Institute, ITI-CERTH, 1st Km Thermi-Panorama Rd, Thessaloniki, Greece {ajinchv, dimitrop, filareti, ngramm}@iti.gr

[2]Argedor Information Technologies, Turkey {dugurca, can.baycay, erdlylmz}@gmail.com

[3]Department of Physical Education and Sport Science, Aristotle University of Thessaloniki, Greece sdouka@phed.auth.gr

Abstract. Game-based learning and gamification techniques are recently becoming a popular trend in the field of Technology Enhanced Learning. In this paper, we mainly focus on the use of game design elements for the transmission of Intangible Cultural Heritage (ICH) knowledge and, especially, for the learning of traditional dances. More specifically, we present a 3D game environment that employs an enjoyable natural human computer interface, 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 (DTW) with Fuzzy Inference System (FIS) approach and provides feedback in a form of a score as well as instructions from a virtual tutor in order to promote self-learning. As a pilot use case, a Greek traditional dance, namely Tsamiko, has been selected. Preliminary small-scaled experiments with students of the Department of Physical Education and Sports Science at Aristotle University of Thessaloniki have shown the great potential of the proposed application.

Keywords: dance performance evaluation, natural human computer interface, traditional dances

1 Introduction

Modern games tend to transcend the traditional boundaries of the entertainment domain giving rise to the proliferation of serious and pervasive games, i.e., games that are designed for a primary purpose other than pure entertainment. Exploiting the latest simulation and visualization technologies, serious games are able to contextualize the player's experience in challenging, realistic environments, supporting situated cognition [1][2]. To this end, serious games are gaining an ever increasing interest in various domains such as defense, education, scientific exploration and health care. Especially in the field of education, serious games have the potential to be an important

teaching tool because they can promote training, knowledge acquisition and skills development through interactive, engaging or even immersive activities. Therefore, by combining gaming and learning, serious games have introduced a new area in the educational domain.

In this paper, we focus on a special field of education, which is the preservation and transmission of intangible cultural heritage (ICH). Such expressions include music, dance, singing, theatre, human skills and craftsmanship. The importance of these intangible expressions is not limited to cultural manifestations, but it coexists with the wealth of knowledge, which is transmitted through it from one generation to the next [3]. As the world becomes more interconnected, many different cultures come into contact and communities start losing important elements of their ICH, while the new generation finds it more difficult to maintain the connection with the cultural heritage treasured by their elders. To this end, the advances in serious games technologies are expected to play a crucial role in the safeguarding of ICH by providing more engaging and, hopefully, effective learning environments.

A characteristic case of ICH is dance (either traditional or contemporary), which can convey different messages according to the context e,g,, artistic, cultural, social, spiritual etc. A dominant factor towards this direction is the human body motion. Therefore, the capture, analysis, modelling and evaluation of dancer's motion with the help of ICT technologies could contribute significantly to the preservation and transmission of this knowledge. Dance analysis is an active research topic [4], while commercial products also exist, such as the Harmonix' Dance Central video game series [5], where a player tries to imitate the motion demonstrated by an animated character. Many research projects have been conducted on the topic of dance assistance and evaluation employing various sensor technologies. Saltate! [6] is a wireless prototype system to support beginners of ballroom dancing. It acquires data from force sensors mounted under the dancers' feet, detects steps, and compares their timing to the timing of beats in the music playing, thus detecting mistakes. Sensable project [7] also employs wireless sensor modules, worn at the wrists of dancers, which capture motions in dance ensembles. On the other hand, the VR-Theater project [8] allows choreographers to enter the desired choreography moves with a user-friendly user interface, or even to record the movements of a specific performer using motion capture techniques. Finally, in [9] different kinds of augmented feedback (tactile, video, sound) for learning basic dance choreographies are investigated.

In this paper, we propose a novel serious game aiming to enhance the learning of a Greek traditional dance through multi-sensing and 3D game technologies and evaluate the dancer's performance by means of sensorimotor learning. For the capturing of dancer's motion, multiple depth sensors are used to address the problems occurring due to self-occlusions by parts of dancer's body. Subsequently, the skeletal data from different sensors are fused into a single, more robust skeletal representation, which is used for both driving the user's avatar in the 3D environment and for his/her performance evaluation. The proposed evaluation algorithm is based on the estimation of motion similarity between the learner's movements and an expert's recording through a DTW/ FIS-based approach in conjunction with an appropriate set of metrics. The 3D environment was developed based on Unity 3D engine [10], which is a popular

multiple-platform gaming solution, and supports a set of activities and exercises designed in close cooperation with dance experts, in order to teach different variations of the dance.

2 The System Architecture

The architecture of the proposed game application consists of three modules, which communicate bilaterally with each other: the Body Capture Module, the 3D Game Module and the Web Platform, as illustrated in Fig. 1. Specifically, the Body Capture Module is used for the capture and analysis of motion data, which are subsequently transmitted to the 3D Game Module for visualization purposes. For motion capture, Kinect depth sensors are used [11]. Although it is possible to play the game using just a single Kinect sensor, multiple Kinect sensors are supported, which leads to improved motion robustness. Each sensor provides the 3D position and orientation of 20 predefined skeletal joints of the human body. A skeletal fusion procedure is performed to combine the data obtained from multiple sensors onto a single fused skeleton as described in Section 3. This fused skeletal animation data are transmitted via TCP/IP in the form of XML messages to the 3D Game Module, where they are used for animating the 3D avatar as well as for the evaluation of user performance. The 3D Game Module is implemented in the Unity game engine and is described in detail in Section 4. The game has bilateral communication with the Web Platform, which is responsible for user profile management and game analytics storage.

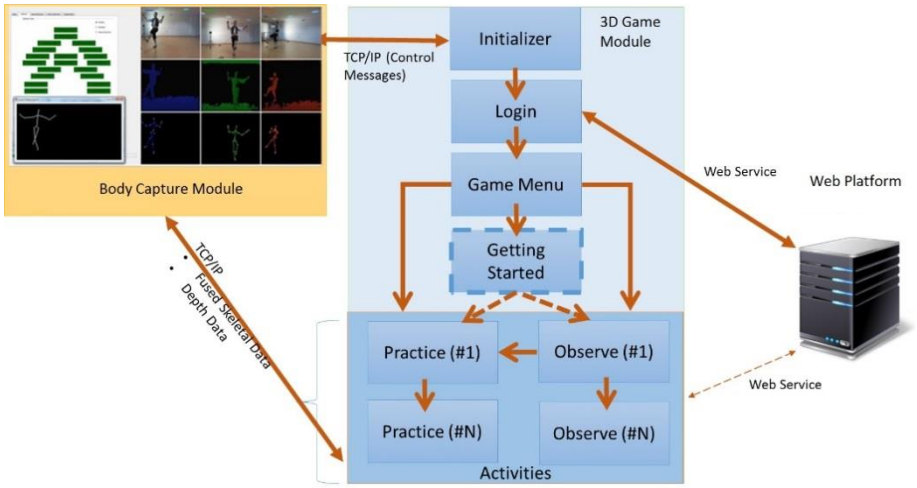


Fig. 1. Tsamiko game architecture

3 Natural Human Computer Interface

To improve the performance of the markerless body motion capture, we apply skeleton fusion, which is the process of combining skeletal data captured by multiple sensors into a single, more robust skeletal representation [12]. This helps address problems occurring due to occlusions, in particular self-occlusions by parts of the dancer's body, e.g. when one raised leg hides other parts of the body from the field of view of a specific sensor. In addition, using multiple sensors results in a larger capturing area since the total field of view can be increased, depending on the placement of the sensors. We use three depth sensors placed on an arc of a circle in front of the dancer, thus allowing the dancer to move in a larger area. In addition, skeleton fusion decreases the noise inherent in skeletal tracking data of depth sensors.

Prior to fusion, sensor calibration procedure must take place in order to estimate the rigid transformation between the coordinate systems of each sensor and a reference sensor. For each sensor, we use the Iterative Closest Point algorithm [13] implementation from the Point Cloud Library (PCL, http://pointclouds.org/) [14] to estimate a rigid (Rotation-Translation) transformation, which is subsequently used to register the skeleton data captured by the sensor in the reference coordinate system. Registered skeletons corresponding to each sensors are then combined into a single skeleton representation using a skeletal fusion procedure, based on positional data (Fig. 2).

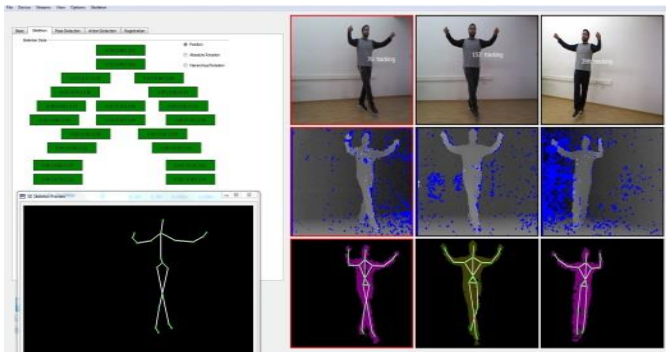


Fig. 2. Skeletal fusion from three Kinect sensors

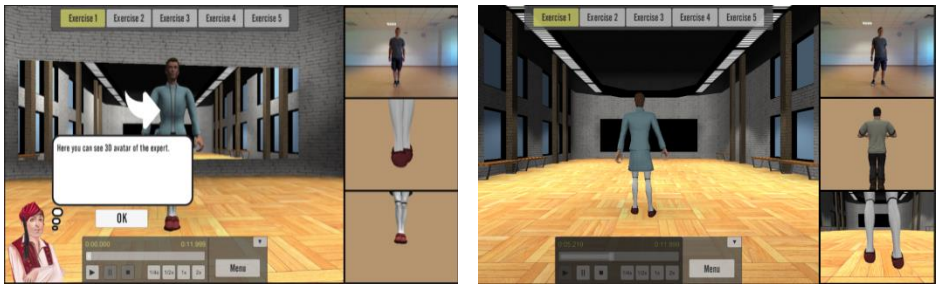
Initially, the sum of all joint confidence levels of each (registered) skeleton is computed and the skeleton with the highest total is selected as the most accurate representation of the person's posture (base skeleton). Based on the joints of this skeleton, we enrich it with data provided by the remaining skeletons. Specifically, if the confidence of a joint of the base skeleton is medium or low, the joint position is corrected by taking into account the position of this joint in the remaining skeletons: if corresponding joints with high confidence are found in any of the remaining skeletons, their average position is used to replace the position value of the joint; otherwise, the same procedure is repeated first for joints containing medium confidence values and then for joints containing low confidence values. Finally, a stabilization

filtering step is applied in order to overcome problems due to rapid changes in joint position from frame to frame which may occur because of the use of joint position averaging in our fusion strategy. The final result of this procedure is a more robust animation of the 3D virtual avatar of the user.

4 The Game-like application

The main objective of the proposed game-like application is to teach two variations of the Greek traditional dance "Tsamiko" (i.e., the single style consisting of 10 steps and the double style consisting of 16 steps). Towards this end, two learning activities were designed and implemented, in close cooperation with dance experts, each one consisting of a number of specific exercises. The learner has to perform all of the exercises successfully to achieve the activity objective. Each exercise consists of several dance steps, which are presented to the learner one by one. In order to proceed to the next exercise, the learner must repeat the current exercise at least 3 to 5 times correctly. At the beginning of each exercise, a 3D animation of the virtual avatar of the expert performing the specific moves is shown to the learner. Afterwards, the learner is expected to imitate the same moves correctly. If the imitation is successful enough, the game progresses to the next exercise. Otherwise, the learner is expected to repeat the same exercise until she/he performs the moves correctly.

More specifically, the game consists of two *Activities related to the* two variations of the dance and a *Final Challenge* with two different options each: "Observe" and "Practice" in order either to observe the exercises or to start practicing the dance routines respectively. There is also a tutorial aiming at explaining to the learner the basics of how to play the game. In order to teach how to use each user interface element, a 2D virtual tutor presents both the sensors and the GUI to the learner. More specifically, the virtual tutor explains a) the basics of Tsamiko dance, b) the sensing technology used in this game, c) the Observe screen sub components and d) the Practice screen sub components. When tutorial finishes, the learner is expected to continue with the observe phase of the first activity, followed by the practice phase.



(a) (b)

Fig. 3. a) Observe and b) Practice screen

Both learning activities have an Observe screen (Fig. 3a), which shows the performance of the expert. In this screen, the learner can watch the moves repeatedly in order to learn them. The 3D animations, video recordings and dance music are presented to the user. The moves are all controlled via the animation player, which is located at the bottom center of the screen. The Observe screen consists of different visual and functional elements, such as the 3D expert view (central view), the animation player (bottom center), the exercise indicator (top center), a video of the expert performance (top-right) and two close-up camera views (center/bottom-right). Subsequently, the Practice screen is used to enable the learner to practice the moves that he/she observed in the Observe screen. The Practice screen (Fig. 3b) shows the expert's avatar at the central window from a backward view, so as to facilitate the learner to repeat his moves. It also contains: a) the animation controller (central-bottom), used both for recording of the performance of the learner as well as for its playback so that the learner can examine the mistakes he/she made, b)The exercise indicator (central-top), i.e. a panel where the learner can see the current exercise and the remaining exercises to complete the activity, c) a video of the expert performance (top-right), synchronized with the animation of the expert, d) the avatar of the learner (center-right) from a backward view and e) close-up view (bottom-right) of the expert avatar's legs from a backward view.

The practice starts with an introduction from the virtual tutor. After providing some basic instructions, the virtual tutor asks the learner to get ready and a counter starts counting down. Then, the learner is expected to imitate the moves of the expert displayed on the screen. After the performance is completed, the evaluation process starts and the virtual tutor presents the evaluation score along with an appropriate message, e.g. "Outstanding performance! You are ready for the next exercise/activity!". If the learner goes beyond a pre-defined success threshold, the virtual tutor asks him/her to get ready for the next exercise. If the performance was not as good as expected, the learner is asked to repeat the exercise. Although the exercises progress sequentially, in some cases the learner has to repeat not only the previous exercise but also some of the previous ones. The flow graph of dance exercises for each activity is shown in Fig. 4, where each exercise is labeled (a)-(h). If all of the exercises are completed successfully, the activity is considered as completed and the next activity unlocks.

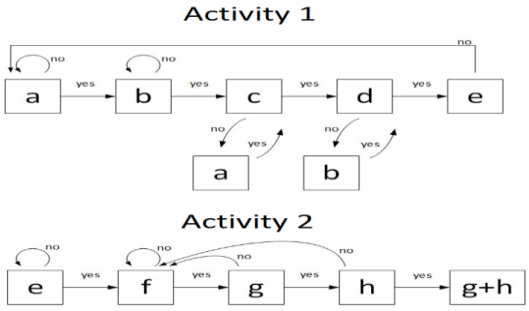


Fig. 4. The flow graph of dance exercises in each activity

Two 3D environments were created for the Tsamiko dance game. For the first and second activity, the 3D environment is a modern dance studio with parquet floor and a big mirror at the back wall (Fig. 5a). The mirror has the same functionality as in real dance studios and provides a reflected image of the dancer. For the final challenge activity, a 3D model of the famous "Odeon of Herodes Atticus" in Athens is designed, as shown in Fig. 5b. This scene is provided as a motivation factor and can be unlocked only when the learner has completed the previous activities.

Fig. 5. (a) Dance studio (left) and (b) Odeon of Herodes Atticus 3D environments (right).

4.1 Evaluation metrics and score calculation

One of the most important elements of the game is the evaluation of the performance of the learner, which is based on the degree of similarity between his body movement and the moves performed by the expert. In order to perform this comparison, specific features are extracted from the learner and expert motion. Those features constitute two time series, which are compared by applying the Dynamic Time Warping (DTW). DTW is a well-known technique for measuring similarity between two temporal sequences that may vary in time or speed. The DTW algorithm calculates the distance between each possible pair of points/features of the two time series, constructing a cumulative distance matrix. It is used to find the least expensive path through this matrix, which represents the ideal warp, the alignment of the two time-series which causes the feature distance between their points to be minimized [15]. This is considered as a distance between the time series under comparison, and provides a good metric of their similarity.

We used three feature sets which are used as input to DTW separately in order to obtain three distinct distance measures. Taking into account that in Tsamiko dance the leg movements constitute the key elements of the choreography, knee and ankle joint positions were used. The first feature set consists of local ankle positions (relative to the waist). By taking the 3D positional coordinates of both left and right ankle relative to the waist, we obtain a 6-dimensional time series. A variant of DTW called multi-dimensional DTW is applied to these time series in order to provide a distance measure. For the 2nd and 3rd feature sets we derive the normalized knee and ankle distances [16]. Specifically, the relative (to waist) 3D position of knee (K_R, K_L) and ankle (A_R, A_L) joints for both left and right feet are used to measure the knee-distance D_K and the ankle distance D_A in each frame:

$$D_K = |K_L - K_R| \tag{1}$$

$$D_A = |A_L - A_R| \tag{2}$$

However, these distances both depend on the height of the dancer, so to ensure invariance to dancer's height a specific normalization process is applied. More specifically, normalized distances are calculated by dividing them by the distance of a "body path" connecting these joints. Thus, the normalized knee-distance $\widehat{D}_K$ and ankle distance $\widehat{D}_A$ are:

$$\widehat{D}_K = \frac{D_K}{|K_L - H_L| + |H_L - R| + |R - H_R| + |H_R - K_R|} \quad \text{and} \tag{3}$$

$$\widehat{D}_A = \frac{D_A}{|A_L - K_L| + |K_L - H_L| + |H_L - R| + |R - H_R| + |H_R - K_R| + |K_R - A_R|} \tag{4}$$

respectively, where H_L, H_R are the left/right Hip positions and R is the root(waist) position.

For each feature set, DTW provides a distance measure between the time-series of the user and the expert. These distances, one per feature set, are subsequently fed to a Fuzzy Inference System (FIS), which computes the final evaluation score (Fig. 6). The proposed FIS system is based on Mamdani method [17], which is widely accepted for capturing expert's knowledge and allows the description of the domain knowledge in a more intuitive, human like manner. The evaluation function produces a normalized scalar value between 0 and 1, which can also be translated into the appropriate text to be displayed by the virtual tutor.

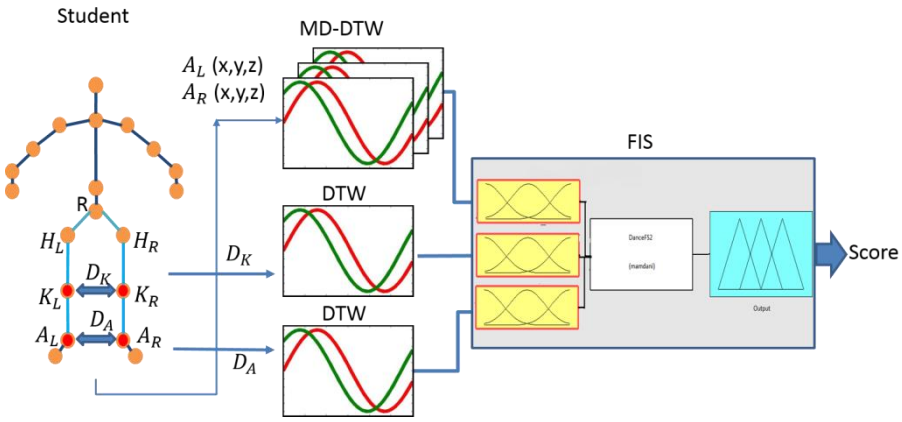


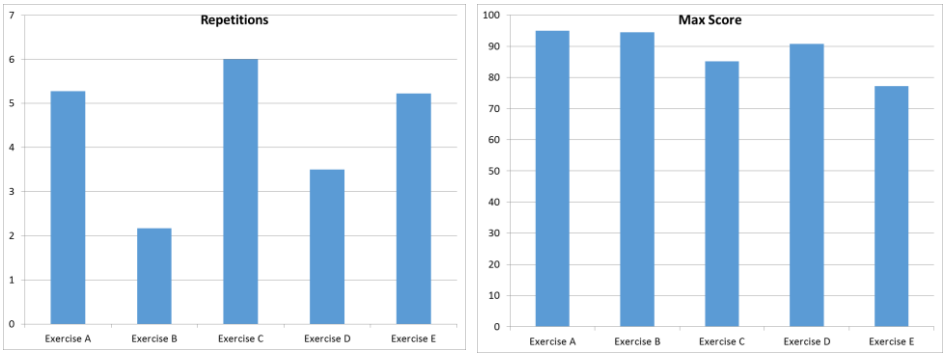
Fig. 6. Estimation of the evaluation score using the proposed DTW/FIS approach.

5 Experimental Results

The proposed game-like application was evaluated by a class of students of the Department of Physical Education and Sports Science at Aristotle University of Thessaloniki. More specifically, 18 students, both male and female, practiced Activity 1, i.e., the "single-step" style, consisting of five exercises. All students were beginners in

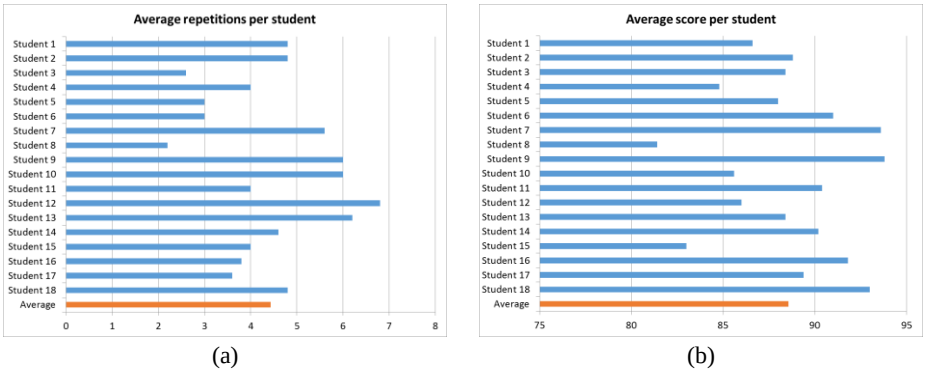
Tsamiko dance and they were initially shown the “Getting Started” and “Observe” modes of the game. During the practice session, analytics were automatically gathered, such as the time of practice for each student, the number of repetitions for each exercise and the intermediate and final scores of students.

Fig. 7 presents the average number of repetitions and the maximum average score per exercise. More specifically, as shown in Fig. 7a, students had less difficulty in passing exercises B and D. This is mainly due to the fact that these exercises have many similarities with their previous ones, i.e. exercise A and C respectively. As a consequence, students who had already practiced these exercises were familiarized with motion patterns, i.e. dance figures, appeared also in exercises B and D and, thus, it was easier for them to pass. As we can see in Fig. 7b, students achieved higher scores in exercise A and B, since these exercises are shorter and easier than the other three ones. It is worth mentioning that although exercise D has a great degree of difficulty, students achieved higher scores than in exercise C due to the reason described above. The lowest scores are appeared in exercise E, which is the most difficult one consisting of the dance figures of both exercise C and D.



(a) (b)

Fig. 7. a) The average number of repetitions and b) the maximum average score per exercise.



(a) (b)

Fig. 8. a) The The average number of repetitions and b) the average maximum score per student

Moreover, it was noticed that a high number of repetitions lead many times to an increased score, which is justified by the fact that students gradually approached the movements of the expert, i.e. they learned to perform correctly the dance figures. Fig. 8 presents the average number of repetitions for all exercises per student as well as the average score of each student. More specifically, the average score for all students was 85.56%, while the number of repetitions for each student in all exercises was 4.4, which corresponds to an average practice time of 7.1 minutes per student.

6 Conclusions

This paper presents a serious game application for transmitting ICH knowledge and specifically the Greek traditional dance "Tsamiko". The game is structured as a set of activities, each consisting of several exercises, aiming to teach different variations of the dance. One of the most important elements of the proposed game-like application is the evaluation of the performance of the learner, in order to provide meaningful feedback. To this end, the learner's movements are captured using a markerless motion capture approach, which aims to fuse skeletal data from multiple sensors into a single, more robust skeletal representation. Subsequently, the motion similarity between the learner's movements and an expert recording is performed through a set of appropriate performance metrics and a DTW/ FIS-based approach. Preliminary small-scale experiments with students of the Department of Physical Education and Sports Science at Aristotle University of Thessaloniki have shown the great potential of the proposed application.

7 Acknowledgement

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7-ICT-2011-9) under grant agreement no FP7-ICT-600676 "i-Treasures: Intangible Treasures - Capturing the Intangible Cultural Heritage and Learning the Rare Know-How of Living Human Treasures".

References

1. De Gloria, A., Bellotti, F., Berta, R., & Lavagnino, E. (2014). "Serious Games for education and training". *International Journal of Serious Game*, 1(1).
2. Watkins, R., Leigh, D., Foshay, R. and Kaufman, R., "Kirkpatrick Plus: Evaluation and Continuous Improvement with a Community Focus", *Educational Technology Research & & Development*, vol. 46, no. 4, 1998.
3. K. Dimitropoulos, S. Manitsaris, F. Tsalakanidou, S. Nikolopoulos, B. Denby, S. Al Kork, L. Crevier-Buchman, C. Pillot-Loiseau, S. Dupont, J. Tilmanne, M. Ott, M. Alivizatou, E. Yilmaz, L. Hadjileontiadis, V. Charisis, O. Deroo, A. Manitsaris, I. Kompatsiaris, and N. Grammalidis, "Capturing the Intangible: An Introduction to the i-Treasures Project", in *Proc. 9th International Conference on Computer Vision Theory and Applications* (VISAPP2014), Lisbon, Portugal, 5-8 January 2014.

4. Raptis, M., Kirovski, D., Hoppe, H., Real-time classification of dance gestures from skeleton animation, Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, August 05-07, 2011, Vancouver, British Columbia, Canada
5. Dance central http://www.dancecentral.com/
6. Drobny, D., Weiss, M. and Borchers, J., Saltate! - A Sensor-Based System to Support Dance Beginners. In CHI '09: *Extended Abstracts on Human Factors in Computing Systems*, pages 3943-3948, New York, NY, USA, 2009. ACM.
7. Aylward, R., "Sensemble: A Wireless Inertial Sensor System for InteractiveDance and Collective Motion Analysis", Masters of Science in Media Arts and Sciences, Massachusetts Institute of Technology, 2006
8. VR-Theater project http://avrlab.iti.gr/HTML/Projects/current/VRTHEATER.htm
9. Drobny D. and Borchers, J., Learning Basic Dance Choreographies with different Augmented Feedback Modalities. In CHI '10: *Extended Abstracts on Human Factors in Computing Systems*, New York, NY, USA, 2010. ACM Press
10. Unity. http://unity3d.com.
11. Kinect for Windows | Voice, Movement & Gesture Recognition Technology. 2013. [ONLINE] Available at: http://www.microsoft.com/en-us/kinectforwindows/.
12. Kitsikidis, A., Dimitropoulos, K., Douka, S., Grammalidis, N., "Dance Analysis using Multiple Kinect Sensors", VISAPP2014, Lisbon, Portugal, 5-8 January 2014
13. Besl, P., McKay, N., A Method for Registration of 3-D Shapes. IEEE *Trans. on Pattern Analysis and Machine Intelligence* (Los Alamitos, CA, USA: IEEE Computer Society) 14 (2): 239–256, 1992
14. Rusu, B., Cousins, S., 3D is here: Point Cloud Library (PCL), *In IEEE International Conference on Robotics and Automation*, 9-13 May 2011
15. G. ten Holt, M. Reinders, and E. Hendriks. Multi-dimensional dynamic time warping for gesture recognition. In *Thirteenth annual conference of the Advanced School for Computing and Imaging*, 2007.
16. Kitsikidis, A., Dimitropoulos, K., Yilmaz, E., Douka, S., Grammalidis, N., "Multi-sensor technology and fuzzy logic for dancer's motion analysis and performance evaluation within a 3D virtual environment", HCI International 2014, Heraklion, Greece, 22-27 June 2014.
17. Mamdani, E.H. and Assilian, S., "An experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, Vol. 7, No. 1, pp. 1-13, 1975.