ART – BASED TRAINING METHODS FOR EMPOWERING ADULTS IN THE DIGITAL ERA

A. Roumana[1], I. Rallis[1], A. Georgopoulos[1], P. Pisimisi[2]

[1] Laboratory of Photogrammetry, Rural, Surveying and Geoinformatics Engineering, NTUA Athens, Greece
[2] Echodrama Cultural Group
antigoniroumana@gmail.com, irallis@central.ntua.gr drag@central.ntua.gr, phaedra-5@hotmail.com

Commission II

KEY WORDS: Kinect, motion capture, choreographic sequences, Virtual Reality, social concepts

ABSTRACT:

Since the dawn of humanity, dancing has been central to life occasions, for celebration, mourning, entertaining or even communicating. Still, it remains a vital practice around the world. With the development of technology, many efforts have been made to adapt specific choreographies and dancing rituals to the modern era, via digital platforms. However, these sequences cannot be construed as realistic when they are viewed through computers, because the actual expressions are lost in the translation. A lot of attempts have been made to provide a solution to this issue, but the most successful are the motion capturing devices and equipment. In this research, we focus on the recording of choreographies, which represent important social concepts, using a Microsoft Kinect sensor as a motion capturing device. Microsoft Kinect is chosen because it is a low-cost device that can deliver adequate results. The outcome of the research is a "handbook" in virtual reality environment, which can be used as an educational tool by anyone. The project aims at being a response to the global need of today's world for creation of inclusive environments, where socially aware people can peacefully co – exist. The digital handbook can enhance the skills and competences of adult trainers in arts and culture and every participant gets in touch with sensitive concepts and develops social awareness. Inspiration for the project is drawn via social and humanistic values stemming from philosophical, sociological and psychological texts and it is innovative not only for its content but also for the way this is developed.
.

1. INTRODUCTION

Performing arts are an essential factor of intuitive communication. Body signals, movements and gestures can punctuate a storyline in an aesthetically pleasing form. The way people dance can reveal whole stories and transmit feelings and continues to have a vital role in societies, wherever humans reside. People use dance for entertainment, exercise, or artistic expressions, and at the same time to form strong bonds with one another. Dancing has a very important role in education as well, since learning through a creative process that involves the whole body is proven to be very efficient.

The best way to share the unique language of dancing in this modern era is through digital platforms and immersive technologies, such as Virtual or Augmented Reality. The main difference between these environments is that the latter amplifies the physical world using digital assets, whereas the former digitally recreates realistic simulations of the physical world that users believe they are existent and interact with them in real time.

However, computers cannot interpret expressions and kinesiology the same way humans can since the detailed sequence of movements cannot be retrieved appropriately. During the past years, many efforts have been made to solve the aforementioned issue, mainly through motion capturing devices and equipment. Motion capture is utilized mostly for animations and video games, but is also applicable in education, engineering, rehabilitation, sports, kinesiology, neuroscience, etc (Rallis, et al., 2018) and can be translated as the process of recording live movements or patterns digitally and converting them to usable mathematics.

Dancing is an important part of Intangible Cultural Heritage (ICH). According to UNESCO (2003), ICH consists of nonphysical intellectual wealth, and it describes traditions or living expressions inherited from our ancestors and passed on to our descendants. Dancing rituals are part of folklore, customs and traditions and the general term also contains language and beliefs. Even though it is not as solid as tangible cultural heritage, it is equally important in maintaining cultural diversity and encouraging mutual respect for other ways of life. Dance heritage is often associated with music, singing and celebration and can either be localised and practised strictly in the country of origin, or be enjoyed internationally.

The main purpose of this research is to create a digital educational tool for dancing, using the low – cost Kinect sensor as a motion capturing device. Five actors are used to record fifty-five choreographies, each one for different concepts. Combining 3D motion data from Kinect and Unity Game Engine (Hou - Ju, et al., 2017) to create Virtual Reality scenes, the vocabulary of kinetic methodology is visualized with 3D graphic display methods, enhanced with animations and graphical environments. The final result is a digital "handbook" which can be used as an educational tool by dance trainers anywhere and enjoyed by any participant, regardless of their educational background.

https://doi.org/10.5194/isprs-archives-XLVIII-2-W1-2022-221-2022

2. RELATED WORK

2.1 Kinect as a motion capturing device

Microsoft's Kinect has been used in many applications through the years, successfully replacing high – priced motion capturing equipment. In detail, in the early 2010s, when the Kinect sensor was released, Hsu (2011) examined the potential of using Kinect as an educational tool, by taking advantage of the kinesthetic and gesture – based interaction it provides. The research emphasizes the importance of interactivity in education and suggests that it can offer an innovative educational development, in alignment with bodily intelligence, making it a proper tool for teaching and learning.

Later on, in 2013, a method was developed (Ren, et al., 2013) which extracted hand gestures from Kinect recordings, since the sensor itself cannot recognize them, due to low resolution depth map (640 x 480). Using color images and the depth map, the hand position was accurately calculated.

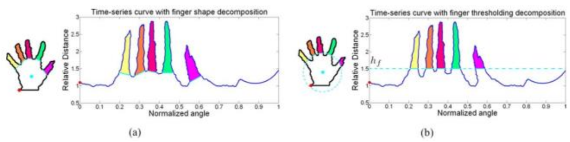

Figure 1. Hand gesture recognition (Ren, et al., 2013)

A few years later, Napoli, et al. (2017) made a thorough research on how Kinect can be used as a motion capturing device, focusing on its exploitation in biomechanical research and clinical applications. They used the real-time captured skeleton data from Kinect to study joint pathology, ergonomics and athletic performance. For each body entry, the system captured two data streams: joint location as 3D coordinates and body segment orientation. The team used two Kinect sensors to get the best results, which were adequate, with high precision and acceptable accuracy levels, marking Kinect as a viable alternative to motion capturing.

2.2 Motion – capture technologies for choreographies

Choreographic sequences and kinesiology have often been examined closely using motion capturing devices. In 2017, a research team from National Technical University of Athens (Rallis, et al., 2017), realising the difficulty computers have in recognising body signals, movements, and gestures, used another motion capturing device, Vicon, to extract 3D information of a dance sequence. The subject of the research were Greek folklore dances and the data extracted consisted of skeleton and joints that represented the human body movement. To make the results better, noise removal methods were also applied.

The following year, similar research was conducted which extracted salient 3D human motion data from real – world kinematic sequences (Voulodimos et al., 2018). The purpose of this research was to amplify the importance of dancing to intangible cultural heritage and how digitizing it can facilitate that. The same goal was approached by another research team a year later, which resulted in an interactive framework that enhances the learning procedure of folklore dances (Rallis, et al., 2019). For this project, Kinect was selected as the motion capturing device, as it was easy to capture the human skeleton joints and extract dance kinematics. The outcome was an in – home learning tool.

2.3 Algorithms for performance recordings

In [Aristidou, et. al., 2015] the authors present a motion analysis algorithm based on ad-hoc quantitative metrics, to provide insights on style qualities of a performance. In this work, a framework based on the principles of Laban Movement Analysis (LMA) is presented. In [Kico 2019] a prototype mobile AR interface is presented for assisting the process of learning folk dances. In [Aristidou, et. al., 2021] the authors have designed and developed a virtual dance museum to provide the technological tools that allow for widely educating the public, most specifically the youngest generations, about the story, costumes, music, and history of our dances. In [Aristidou, et. al., 2022] the authors describe a choreographic framework, which exploits emerging technologies to digitize, analyze, and holistically document choreographic performances. In [Kico, et., 2022], the authors present an enhanced framework for supporting choreographic learning using AR and non-invasive brain stimulation.

In [Rallis, et al., 2018], a very important conclusion was reached for performing arts. The research elaborates the importance of automatic extraction of choreographic patterns, because computer systems cannot interpret human body signals the same way as humans themselves and uses up – to- date devices to capture video content and depth. The 3D motion data formed the geometry of dance trajectory, and they offered the main structural component of a dance, they assisted trainees towards a proper learning and improved dance experts as well.

3. METHODOLOGY

In this section, the development of the application is analysed step by step from the equipment and software that was used to the way the movements were selected and created and finally to the final product.

3.1 Software and Equipment

The equipment selected for the completion of this research is a Kinect sensor 2.0. Because it is low-cost, Kinect stands as an attractive alternative to expensive laser scanner devices (Khoshelham, 2011). The sensor captures depth and colour images at 30 fps with a resolution of 300000 points per frame. It consists of a depth sensor, working with infrared laser, an RGB camera and a four – microphone array which provides full – body 3D motion capture, without using marks.

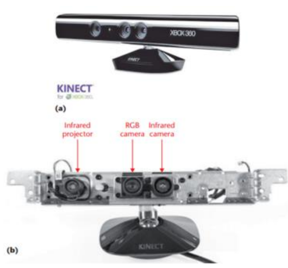

Figure 2. Kinect 2.0 sensor parts

To work successfully when used for motion capturing, the sensor requires environment 3x3 meters (10 feet by 10 feet), because on a smaller scale, actors will not fit into camera view. Moreover, even though there are not restrictions at clothing, it is suggested that actors chose slim clothes, to reduce noise in the recording.

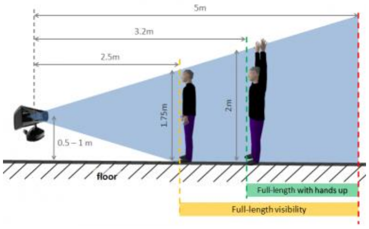

Figure 3. Recording area for Kinect 2.0 *(Source: https://wiki.ipisoft.com)*

The extracted skeleton consists of twenty – five joints, each including the 3D coordinates, the rotation parameters, and a tracking state property. The points replace the head, the neck, all the joints in human arms and legs, and bones for hips and spine. Generally, the center of the rig is considered to be in the hips.

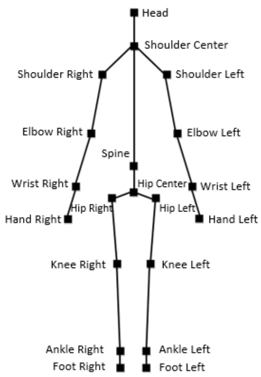

Figure 4. Depiction of human skeleton in Kinect 2.0 (*Source: https://subscription.packtpub.com)*

On the downside, Kinect does not create separate joints for fingers and since the head is depicted with only one joint, the head movements are imperfect. Also, it is designed to track only the front side of the subject, causing some limitations when it comes to capturing a 360º or profile view.

The software used is completely open – source. The capturing is achieved via the SDK (Software Development Kit) that Kinect offers, and more specifically the Kinect studio and the visual gesture builder, and the skeleton is extracted using Kinect animation studio. The visual gesture builder is created to fix the issue of not tracking hand gestures. However, in order to be used, all the hand gestures have to be recorded separately beforehand, and they cannot be extracted from an existing recording from Kinect Studio. Consequently, the hand gestures cannot be attached to other body movements.

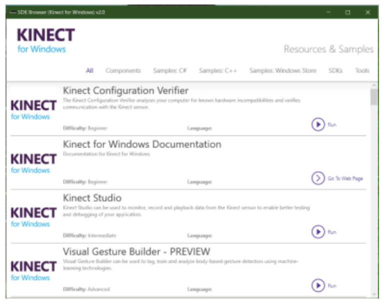

Figure 5. Kinect SDK

Some data is captured using iPi Recorder 4 instead of Kinect Studio, because it additionally ameliorates depth. The recording process is similar, except iPi Recorder requires background evaluation. This recorder does not depict actors with their skeletons, but as outlines with color (Figure 10). The skeleton is extracted during further editing in iPi MoCap Studio.

The main difference between the two programs, apart from the way skeleton data is extracted, is that iPi Recorder displays the image mirrored, after the necessary background evaluation.

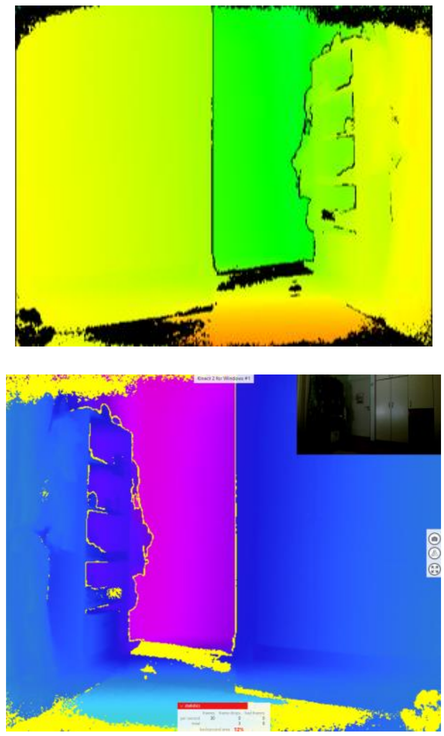

Figure 6: The same background displayed in Kinect Studio (top) and iPi Recorder (bottom)

The second software, iPi Recorder combined with iPi MoCap Studio, is an advanced software used for recording animations with a single or multiple Kinect sensors or simple RGB cameras. It provides specific settings that can be changed according to the body shape of the actor – height, body mass, legs and arms' length, shoulder width etc. What makes this particular application important is that, recognising the inability of Kinect to capture finger movements, it offers the opportunity to manually add actions on both hands. More specifically, it allows the users to select a first and last keyframe, between which a specific action of either hand takes place. The actions it provides at the moment are a) default, b) relaxed or c) fist.

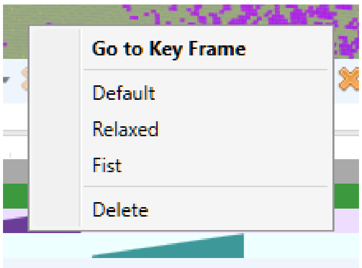

Figure 7. Hand gestures in iPi MoCap Studio

The programm also allows users to manually edit movements that are not translated smoothly. It offers the ability to move or rotate every bone separately, to match it with the actions of the actor. Moreover, head tracking can be enabled or disabled according to the user's desire and certain bones can be removed from the movement altogether or tracked separately from the other body to correct the result. Finally, it provides real-time feedback on the success of the actor tracking by generating a mismatch number. The lowest the number, the better the match.

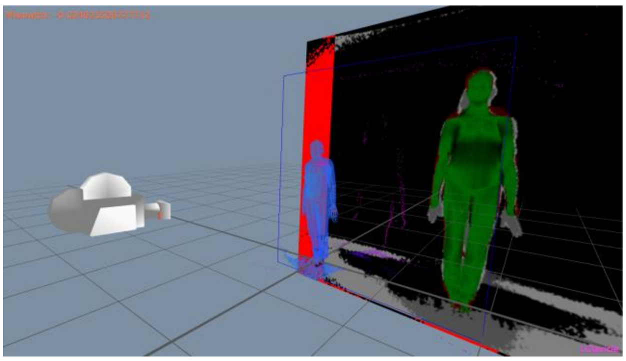

Figure 8. Pose mismatch in iPi MoCap Studio

It is worth mentioning that many export options are available, because iPi Recorder is directly compatible with Unreal Engine and Autodesk 3DS Max. The developers of the programm have provided a very insightful wiki, that clarifies the procedure of recording with Kinect and also proposes advices to achieve the best result.

As final step, the Virtual Reality application is developed in Unity Game Engine. The 3D models that overwrite the actors are drawn from Mixamo, an online database which has 3D models available to download free of charge.

3.2 Movements

The source of the body vocabulary is words that are found in anti – racist and anti- discrimination texts, in an aim to use dance as a way to promote solidarity, tolerance and anti – racism. The group who conducted the research for the project involved nine dancers, one cultural pedagogue - educator, and one theatre director from 8 different nationalities (Greece, Germany, France, Togo, Nicaragua, Venezuela, Spain and Albania). The EU funded Erasmus+ project is conducted with the collaboration of five companies and institutes, and more specifically Bridgeworks e.V. from Germany, Lab of Photogrammetry of the National Technical University of Athens from Greece, Echodrama Cultural Group from Greece, Zakk (Zentrum für Aktion, Kultur und Kommunikation) from Germany and Association Act' Dem from France.

The new international body language which is created, is formed with words, where every choreographical movement has a specific meaning (words = movements). Among the relevant litareture researched, it was decided to use as best source for these words the book "Exit Racism", (Tupoka Ogette 2018). The words that are extracted from the book are these related to social meanings, as displayed below (Table 1), both positive (equality, upliftment) and negative (fear, oppression). A total of 56 words are selected, based on the frequency they are encountered in anti – racist, feministic and anti – discrimination texts.

CONCEPT	DEFINITION
Power	Authority that is given to a person or body
Discrimination	The unjust or prejudicial treatment of different categories of people
Silence	The fact or state abstaining from speech
Contempt	The feeling that a person is worthless or beneath consideration
Taboo	A social or religious custom prohibiting or restricting a particular practice
Equality	The state of being equal in status, rights or opportunities
Patriarchy	A system of society or government in which men hold the power and women are largely excluded from it
Freedom	The power or right to act, speak or think as one wants
Solidarity	Unity or agreement of feeling or action, especially among individuals with a common interest

Table 1: Examples of concepts described and their definition (*Source of the definitions: OxfordLanguages*)

The decision of the movement selection for every word was a long process. Each dancer worked individually at first to make a movement proposal for each word. After many discussions and try – outs and often by combining ideas, the definitive results were formed, based on methods of contemporary dance. With the help of the cultural pedagogue and the theatre director, the movements were finalized.

Some factors that were taken into consideration for the selection of the movements was that everyone can do them – children, schoolteachers, individuals who are not dancers – and that these movements express in a clear but artistic way the meaning of each word and the feeling they ought to transmit.

3.3 Application Description

3.3.1 Data Recollection: The choreographic sequences have been recorded using a Microsoft Kinect sensor, with a plain white background to avoid noise collection from irrelevant items. To complete the recording, some videos were also captured using a simple camera.

Figure 9. One of the actors, right before recording a movement in the plain white background

Each actor begins their movement in T- pose, the default pose animators and character artists use, because this is the only way skeleton data from Kinect can be matched to a 3D model. This specific pose is chosen because it allows artists to see every part of the character, including those that are typically hidden by other things on camera, such as the armpits, between the fingers and so on. This rule is followed whether the data are collected through Kinect Studio or iPi Recorder.

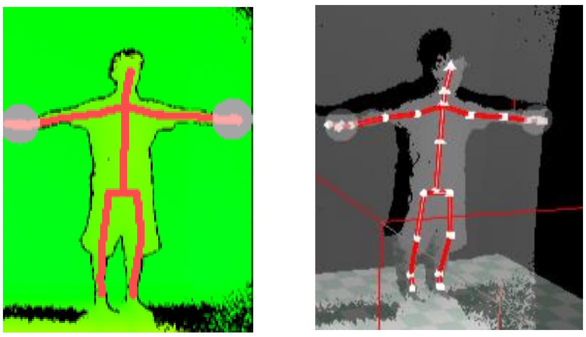

Figure 10. T – Pose in Kinect Studio, in Playback 3D view (left) and in Playback 2D view (right)

Figure 11. T- Pose in iPi Recorder

Kinect Studio captures motion data in real time, but in order to convert them to skeleton data, Kinect animation studio is needed. Through that, every human skeleton joint is visualized as a point and a simple skeleton is formed with 25 points. As it is mentioned, fingers are not depicted.

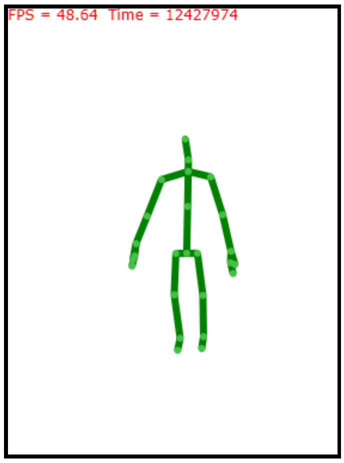

Figure 12. Skeleton from Kinect Animation Studio

Since iPi Recorder captures depth map, iPi MoCap Studio is required to convert this to skeleton data. The skeleton created in iPi MoCap Studio uses the same number of joints with the one from Kinect Studio, but it depicts with more detail the hands. Automatic hand recognition is still not an option, but the joints allow the actions to be imported manually. The studio offers many options to export the animated rig, but in this case the simple FBX format is used.

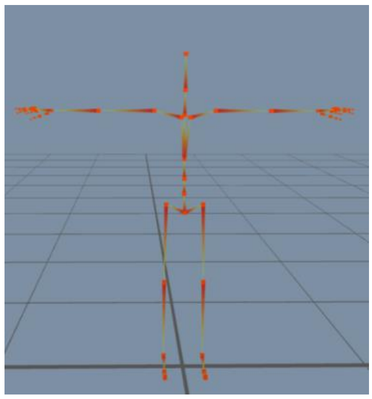

Figure 13. Skeleton from iPi MoCap Studio

The skeleton data are then edited accordingly and imported to Unity Game Engine as animations. Using Mixamo, 3D models are selected that represent the actors. The animations are added to the skeleton of these models, and now they appear as they do the choreographies.

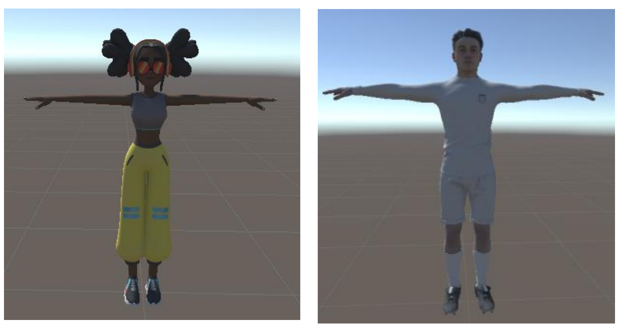

Figure 14. Examples of the Mixamo Models

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-2-W1-2022-221-2022 | © Author(s) 2022. CC BY 4.0 License.

Animating in Unity is a very straightforward process, since the only prerequisites are standard tools. For the specific application the tools are , animation data, which are the skeleton movements extracted from Kinect, and a 3D model. The match of the model and the animation is achieved via an animator controller.

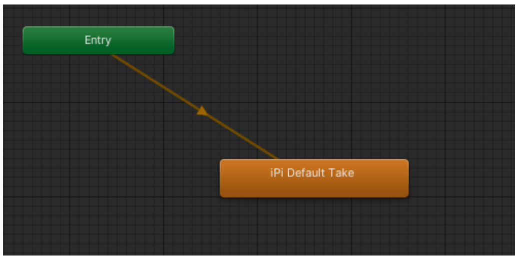

Figure 15. Unity's Animator Controller

Some experiments are conducted, in which two actors performed in sync the same action instead of one, to offer the chance to practice the choreographies in couples. For this, Kinect Studio is used because it enables multiple skeleton tracking, something that does not stand for iPi Recorder.

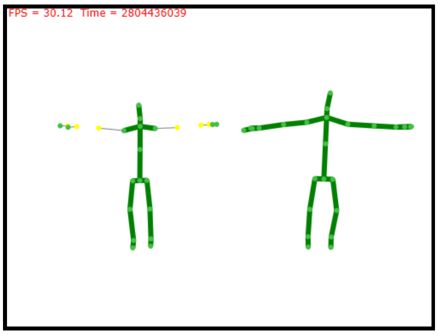

Figure 16. Couples recordings in Kinect Studio

The results are satisfying, for the programm behaves similarly to having two different recordings joined together, instead of two people performing. However, if the actors are too close to one another, the sensor gets confused and misperceives hand actions. The experiment is also transferred in Unity, using two different models in the same scene.

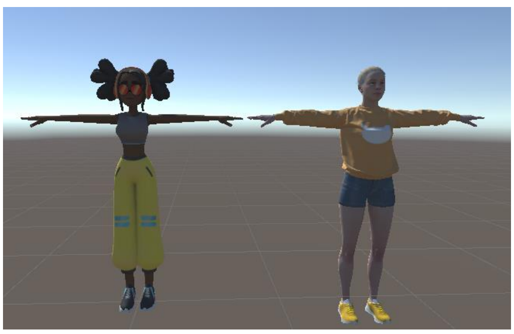

Figure 17. Couples in Unity

3.3.2 Virtual Reality Application: In order to achieve the best learning experience possible, the application is developed in Virtual Reality Environment using Unity's fitting plug – ins. Every movement is developed as a separate file, so in the final handbook the user can select which choreography they want to learn without having to load a whole application. Of course, the issues that come along with Virtual Reality applications, such as dizziness, have to be avoided, so the time the user spends in the VR environment is limited.

The handbook briefly explains the meaning of each movement before proceeding to the choreography. Because of this, users have a chance to dive in the concepts and figure out themselves the relation between the value and the kinematics that describe it.

4. CONCLUSION AND EVALUATION

This research helps to recognise how low- cost motion capturing devices such as Kinect 2.0 can be used to create educational applications. The project in particular is a reply to today's need of inclusive environments, because it teaches social awareness the universal language of dancing, via an application accessible to everyone.

Even though Kinect offers an adequate replacement to expensive motion- capturing devices, the few issues that were faced cannot be ignored. First of all, the hand and feet recognition is not ideal, since the sensor could not capture detailed or fast – paced movements, making the result look awkward and unrealistic. This issue can be partially solved using Kinect's visual gesture builder, but the issue with the feet remains. Of course, external applications such as iPi MoCap Studio or Blender can be used to improve rigging, but it requires a certain amount of experience to rectify the problem. Secondly, it is common among low – cost sensors that are used for motion capturing that they do not capture face expressions. This is a logical result, because as it is mentioned Kinect translates movements using skeleton data and does not provide any information about the actual model. This is not a very critical issue, as the application fulfilled its goal anyhow, but the experience would have been better otherwise. Finally, the sensor could not capture movements that occurred in profile, albeit this could be solved by using two Kinect sensors. This way, a 360° view of the actor can be captured and the result will be far more accurate. All in all, Kinect offers an ample alternative to motion capturing.

The final result is convincing, and generally wins over the users. The Virtual Reality environment is the hardest part to get used to, but the issue is solved by making each session short.

To conclude, this project is innovative not only for its content, but for the way that is transmitted as well. Using 3D motion data, captured with low – cost sensors, the vocabulary of a specific kinetic methodology is visualized and the result is enhanced with animations and graphics.

As an improvement, and to take this research one step further, the words that are recorder can be combined in various ways, to form sentences. In the future, the same application can be developed for generalized dancing lessons, replacing the conventional way of learning.

ACKNOWLEDGEMENTS

This project has received funding by Erasmus +KA2. Project code: 2021-1-DE02-KA220-ADU-000029772. Project acronym: ATMEADE

REFERENCES

Aristidou, Andreas, et al. "Virtual Dance Museum: The case of greek/cypriot folk dancing." Proceedings of the Eurographics Workshop on Graphics and Cultural Heritage (Aire-la-Ville, Switzerland, Switzerland, 2021), Hulusic V., Chalmers A.,(Eds.), GCH. Vol. 21. 2021.

Aristidou, A., Chalmers, A., Chrysanthou, Y., Loscos, C., Multon, F., Parkins, J., ... & Stavrakis, E. (2022, April). Safeguarding our Dance Cultural Heritage. In Eurographics' 2022.

Aristidou, A., Stavrakis, E., Charalambous, P., Chrysanthou, Y., & Himona, S. L. (2015). Folk dance evaluation using laban movement analysis. Journal on Computing and Cultural Heritage (JOCCH), 8(4), 1-19.

Hou - Ju, Z., Xiu - Ying, S., Jian - Jun, P., Ji - Ping, L., & Rong - Li, G. (2017). Interaction between virtual scene based on Kinect and Unity 3D. *19th International Conference on e-Health Networking, Applications and Services (Healthcom).* IEEE. doi:10.1109/HealthCom.2017.8210835

Hsu, J. H.-M., 2011. The potential of Kinect in Education. *International Journal of Information and Education Technology*, 1(5).

Khoshelham, K., 2011. Accuracy analysis of Kinect Depth Data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII. Calgary.

Kico, I., & Liarokapis, F. (2019, November). A mobile augmented reality interface for teaching folk dances. In 25th ACM symposium on virtual reality software and technology (pp. 1-2).

Kico, I., & Liarokapis, F. (2022). Enhancing the learning process of folk dances using augmented reality and non-invasive brain stimulation. Entertainment Computing, 40, 100455.

Napoli, A., Glass, S., Ward, C., Carole, T., & Iyad, O., 2017. Performance analysis of a generalized motion capturing system using microsoft kinect 2.0. *Biomedical Signal Processing and Control*, 38, 265-280 (2017).

Ogette, T., 2018. Exit Racism: Rassismuskritisch denken lernen. Unrast Verlag ISBN: 978-3897712300

Rallis, I., Doulamis, N., Doulamis, A., Voulodimos, A., & Vescoukis, V. (2018,). Spatio- temporal summarizarion of dance choreographies. *Computers & Graphics*, pp. 88-101.

Rallis, I., Georgoulas, I., Doulamis, N., Voulodimos, A., & Terzopoulos, P. (2017). Extraction of key postures from 3D human mostion data for choreography summarization. *2017 9th International Conference on virtual worlds and games for serious applications (VS-Games)* (pp. 94-101). IEEE.

Rallis, I., Langis, A., Georgoulas, I., Voulodimos, A., Doulamis, N., & Doulamis, A. (2018). An embodied learning game using kinect and labanotation for analysis and visualization of dance kinesiology. *2018 10th international conference on virtual worlds and games for serious applications (VS-Games)* (pp. 1-8). IEEE.

Rallis, I., Protopapadakis, E., Voulodimos, A., Doulamis, N., Doulamis, A., & Bardis, G. (2019). Choreographic pattern analysis from heterogeneous motion capture systems using dynamic time warping. *Technologies*, 56.

Ren, Z., Yuan, J., Meng, J., & Zhang, Z. (2013). Robust Part – Based Hand Gesture Recognition Using Kinect Sensor. *IEEE Transactions on Multimedia*, *15*, 1110-1120.

Unesco Institute of Statistics, 2009. The 2009 Unesco Framework for cultural statistics (FCS). UNESCO ISBN: 978-92-9189-075-0

Unesco. 2003. Basic texts of the 2003 Convention for the Safeguarding of Intangible Cultural Heritage, 2020 Edition https://ich.unesco.org/doc/src/2003_Convention_Basic_Texts-_2020_version-EN.pdf

Voulodimos, A., Doulamis, N., Doulamis, A., & Rallis, I. (2018). Kinematics - based Extraction of Salient 3D Human Motion Data for Summarization of Choreographic Sequences. *24th International Conference on Pattern Recognition (ICPR).* Beijing.