Introducing an Edge-Native Deep Learning Platform for Exergames

Antonis Pardos[1] , Andreas Menychtas[1,2(✉)] , and Ilias Maglogiannis[2]

[1] BioAssist SA, 1 G. Mpakou Street, 11524 Athens, Greece
{antonispardos,amenychtas}@bioassist.gr
[2] Department of Digital Systems, University of Pireas,
80 M. Karaoli & A. Dimitriou Street, 18534 Pireas, Greece
imaglo@unipi.gr

Abstract. The recent advancements in the areas of computer vision and deep learning with the development of convolutional neural networks and the profusion of highly accurate general purpose pre-trained models, create new opportunities for the interaction of humans with systems and facilitate the development of advanced features for all types of platforms and applications. Research, consumer and industrial applications increasingly integrate deep learning frameworks into their operational flow, and as a result of the availability of high performance hardware (Computer Boards, GPUs, TPUs) also for individual consumers and home use, this functionality has been moved closer to the end-users, at the edge of the network. In this work, we exploit the aforementioned approaches and tools for the development of an edge-native platform for exergames, which includes innovative gameplay and features for the users. A prototype game was created using the platform that was deployed in the real-world scenario of a rehabilitation center. The proposed approach provides advanced user experience based on the automated, real-time pose and gesture detection, and in parallel maintains low-cost to enable wide adoption in multiple applications across domains and usage scenarios.

Keywords: Deep learning · Edge computing · Exergames · Computer vision · Convolutional neural networks · Rehabilitation

1   Introduction

The computer games landscape is very rich nowadays and is continuously expanding with new approaches which are based on innovative technologies for human-computer interaction, provide advanced game-play, are available on multiple platforms and devices, and target different user groups. Therefore, the users can find in the market a variety of games, from the traditional video games, in which a player is seated in front of the screen and controls the game with a controller such as keyboard, mouse or gamepads, to the virtual and augmented reality games which require special equipment. In this work, we focus on the *Serious Games* [16] category and particularly the exergames, that has great scientific

value and impact, and serves purposes beyond the game itself and the satisfaction of the player [15]. Active Video Games (AVG) or exergames are based on the technology that monitors body movement or reactions and have been credited with upgrading the game stereotype as a sedentary activity, promoting a more active lifestyle. Golstein et al. [6] showed that there was a significant improvement in the reaction time of 69–90 year old when they were playing video games for 5 h each week, for 5 consecutive weeks.

There several works in the literature highlighting the positive effects of exergames, both generally, regarding the increase of the physical activity and improvement of well-being, and also for specific use case scenarios, such deployments in schools, elder houses, rehabilitation centers, and more, where the benefits in the certain context are better measurable with the direct involvement of scientific personnel and experts [8,13]. In technical level, these exergames are usually based on the use of particular platforms such as Nintendo Wii, and require the use of peripherals that the players hold or wear and act as game controls. The controls (boards, sticks, patches) capture the characteristics of the user's movement and interpret it to events or gestures through which the game is controlled.

The recent technological advancements in the areas of computer vision and deep learning enable the real-time analysis of images and streaming video to solve several classification problems, from object detection and face recognition, to emotion analysis and pose detection, which in turn are applied to real-world scenarios for providing advanced interactivity with systems and creating rich user experiences. Particularly, the detection body pose or face landmarks is applied in several entertainment, medical and business use cases exploiting the increased computational capabilities of the available hardware nowadays and the advanced features of the computer vision and deep learning frameworks and models. In this work we propose an exergame platform which runs on low-cost commodity hardware and utilizes deep learning pose detection based on Convolutional Neural Networks - CNN for human-game interaction. Nowadays, the exegames are implemented using expensive, special purpose systems, which are designed and implemented specifically for the a particular game. The proposed solution, is a modular and extensible platform that is based on commodity hardware and uses generic purpose software and tools, not only allowing for the inexpensive implementation of several different games that exploit computer vision and deep learning on top of the platform, but also enabling its use and the adoption of the overall concept, in other fields and use cases where human-computer interaction is required.

The rest of the document is structured as follows. Section 2 analyses the related works and the technological and scientific baseline for the proposed platform. The platform of the overall system is presented in Sect. 3 along with the hardware elements, the software frameworks and the techniques that were deployed. Section 4 describes the use of the system in a real scenario as well as initial results of its operation. Finally, Sect. 5 concludes the manuscript and presents future extensions and improvements.

2   Related Work and Background Information

Studies have shown that exergames can offer significant benefits to the mental and physical health of people who play especially for the elderly, children, adolescents and people with disabilities [2,3,14], such as the significant decrease in BMI (Body Mass Index) in some cases. Games that enhance physical activity have been developed in the past. Some examples are: "Dance Dance Revolution" [17], introduced in Japan in 1998. Players standing on a "dance platform" or stage and hit colored arrows laid out in a cross with their feet to musical and visual cues. Players are judged by how well they time their dance to the patterns presented to them and are allowed to choose more music to play to if they receive a passing score. "Active Life Outdoor Challenge" [5], a video game for the Wii platform, where players use a mat in conjunction with the Wii remote in order to complete a variety of mini games. "The Think & Learn Smart Cycle" [1] is a stationary bike that hooks up to a tablet via Bluetooth for preschoolers to play different learning games. There are games about Letters & Phonics, Spelling & Vocabulary, and Reading & Rhyming. In addition to learning, this game also keeps them active because the faster they pedal, the faster the on-screen action. "Wii Fit" [5], It is an exercise game with several activities using the Wii Balance Board peripheral, a device that tracks the user's center of balance.

In order to be effective, a game of this type must always take into account the user himself. Children, adolescents, elderly or people with disabilities have different needs, which if properly defined will greatly contribute to the success of the game. Another key component to the success of these games after identifying users' needs, is that the game should entertain and stimulate the user's interest, while convincing him for more exercise. A game that captures the attention of the user is often the impulse for a longer training period, examples like a fixed bike on which a user controls the flow of a game while cycling, resulting in more power being consumed by another user who did the same without controlling a game. Graphics are also an important factor in the success of the game. Seniors need games with simple graphics, easy information and more time available for understanding processes [3]. On the other side are the younger users who are expecting games with rich graphics and speed. These games may be adapted in accordance with user's physical condition improvement, or user needs as mentioned above. Through this process significant benefits have been observed in restoring the development of children and young people (5–25 years old) with mobility problem [11].

Studies have shown that after a few weeks the frequency the user plays is decreased and as a result there is no significant improvement in the user's physical condition [14]. This is why it is prominent to be the right evaluation of the target group in order the main strategy and design of the game to be drawn. Another crucial factor that limits the scope of these games is the hardware that must be applied. Examples of the exergames we saw above make use of special peripherals and this makes the game more complicated in terms of both architecture and cost. However, there should be a balance in the length and frequency the user is involved with the game, and extra care must be taken to avoid overuse

as there is possibility of injury [14]. To avoid such situations, clear instructions should be given to the user, as well as a game play program. A key factor in using these games is their design and their strategic goal. For example, some games require the user to carry on-screen objects with their hands, while others, such as the "Fish'n Steps" project encourage the community to walk around in order to grow a virtual aquarium fish. In all cases the player is evaluated according to the strategy of the game and the result he wants to achieve. For example he can be scored according to the accuracy, speed, or the "Quantity" of movement.

3   System Architecture and Implementation

The objective of this work is to deliver a stand-alone exergame solution, exploiting computer vision and deep learning techniques for the human - game interaction. The proposed system combines state-of-the-art hardware and software technologies and tools in order to deliver advanced exergames features and performance, without compromising the requirements for cost and user experience. An overview of the system is depicted in Fig. 1.
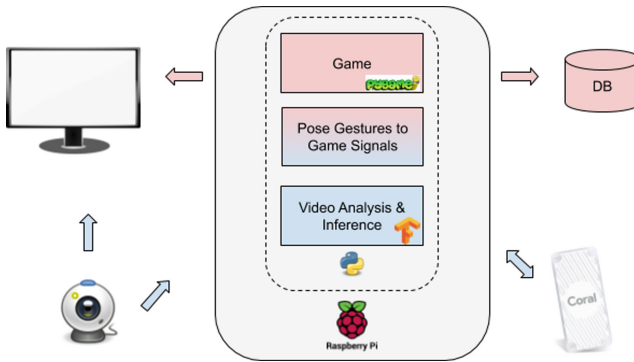
Fig. 1. Overview of the system.

The main components of the system include a) the core game implementation, b) the video analysis and inference mechanism, and c) a custom mapper for interpreting the users' pose and gestures to events and signals for controlling the game. In addition, a database for storing the game results and managing the users has been integrated into the system. The game implementation is based on the popular Python library PyGame while the video analysis and inference on OpenCV and Tensorflow framework.

One of the main challenges while designing the system was to address the contradicting requirements considering on one hand the low-cost and one the other hand the nature of the exergame which required satisfying game speed, graphics and control. The most common computer board that could be used in

order to meet the requirements for low-cost, since custom hardware was not an option, is Raspberry Pi Model 3 [12]. There were two technical difficulties following with this approach though: a) to perform the video analysis and deep learning inference operations in such hardware, like in any edge computing architecture and b) the remote management and maintenance of the edge device. The first one was addressed by using the Coral USB [4] Tensor Processing Unit (TPU), a special circuit designed to achieve high computational speeds in neural network applications which is compliant with the Tensorflow tools and the deep learning models applied. The TPU is designed for the performance phase, when systems with compiled models are presented with real-world data and are expected to behave appropriately using a version of TensorFlow called TensorFlow Lite. In order to facilitate remote management and maintenance, the deployment and operation was based on the Balena.io remote management solution which allows for dockerization of the software elements and their instantiation through well-defined DevOps processes [9].

Additional hardware peripherals were integrated for the interaction with the users; a Raspberry Pi Camera Module v2 is used for capturing images as the input sensor for controlling the game as well as a TV monitor connected via a HDMI. On software level, the proposed solution was based on Linux Debian 10 (codename: buster) which allowed for the smooth and flawless delivery of the complete software stack through the balena.io services.

3.1   CNN-Based Pose and Gesture Detection

For the implementation of the pose and gesture detection system, the TensorFlow framework has been used, as an end-to-end open source solution that has a comprehensive, flexible ecosystem of tools, libraries and community resources. This framework allowed for using state-of-the-art convoluntional neural networks and models like Posenet [10].

The model was trained on COCO dataset on top of MobileNet V1 network architecture. MobileNet architecture differs in the convolution process. Standard convolution filters and computes a new set of outputs in one step which costs in speed and time. In MobileNet, this is a two step procedure which is factorized in a depthwise convolution for filtering and a 1×1 convolution for combining. This factorization reduce model size and computation drastically [7]. Posenet provides an average keypoint precision of 0.665 to 0.687 depending on single or multi-scale inference. For person detection and pose estimation, Posenet adopts the bottom-up approach by localizing identity-free semantic entities (landmarks) and grouping them into person instances. The model learns to predict the relative displacement between each pair of keypoints starting from the most confident detection of a distinguished landmark such as nose. Totally seventeen (17) heatmaps have been developed, each one corresponding to a keypoint along with offset vectors. In the proposed approach, this procedure runs on the TPU module that is connected on Raspberry pi. Each keypoint is related to a heatmap which can be decoded and give the highest confidence areas in the image, corresponding to the keypoint. The offset vector is a 3D tensor with size
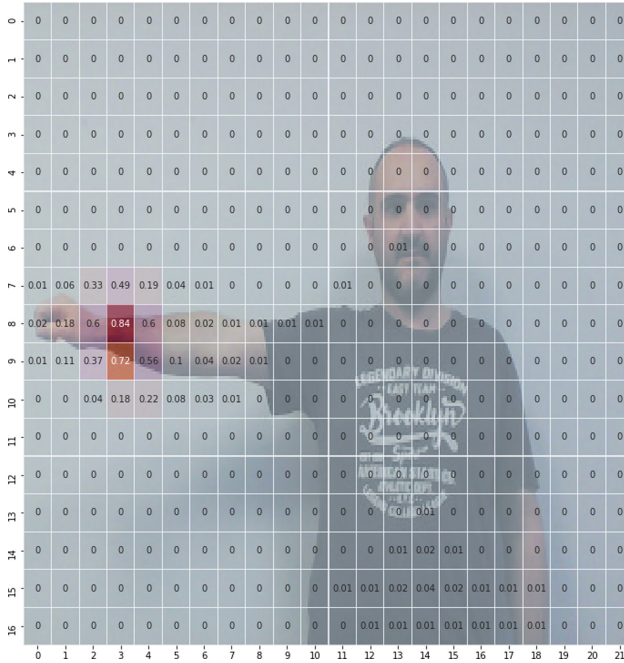
Fig. 2. Example of heatmap #10 that represents keypoint 'Right Wrist'.

Width × Height × 34 (two coordinates for every keypoint) and used to predict the exact location of each keypoint in the image. Models varying in the number of operations per layer have been obtained by the Posenet. The larger the size of ops in layers the more accurate the model at the cost of speed. The keypoints we get from Posenet are: (1) Nose, (2) Left Eye, (3) Right Eye, (4) Left Ear, (5) Right Ear, (6) Left Shoulder, (7) Right Shoulder, (8) Left Elbow, (9) Right Elbow, (10) Left Wrist, (11) Right Wrist, (12) Left Hip, (13) Right Hip, (14) Left Knee, (15) Right Knee, (16) Left Ankle, (17) Right Ankle. An example of Right Wrist landmark detection heatmap is shown in Fig. 2.

3.2   Exergame Engine Implementation

As already mentioned the exergame engine is based on PyGame, which is a cross platform set of modules that is used for writing video games in Python and includes computer graphics and sound libraries. The various software components were designed in order to connect neural network landmarks output with the control of the game. The deep learning framework provides for each analysis a dictionary in the following format: $\{keypoint_1 : (x, y, score), ..., keypoint_{17} : (x, y, score)\}$, where x, y are the pixel coordinates of input image. Following the analysis, set of rules are applied to the result to inference the user's gesture from landmarks. Every gesture is an event associated with a signal which in turn
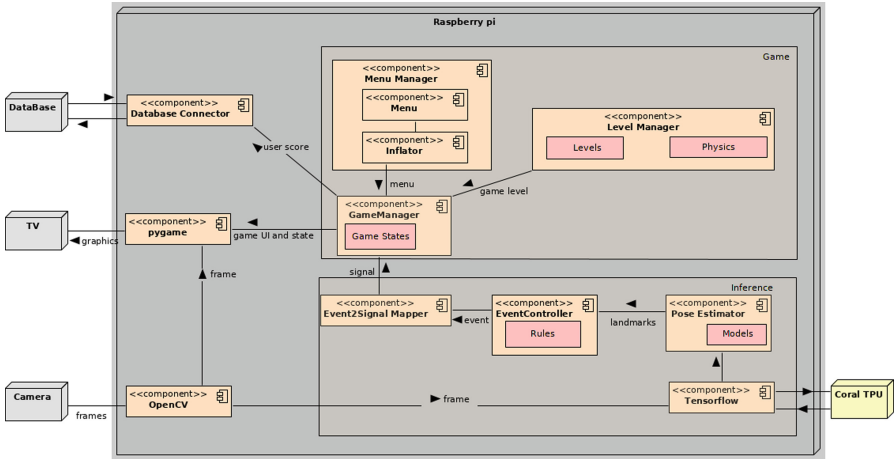
Fig. 3. Exergame platform component diagram

controls the game's functions and menu options. With rules we can identify the pose or gesture of the user and with signals we can control the game flow. More specifically, we assign specific poses to game signals for each control function like *stop, pause, next button, previous button* and we set the rules for recognizing that poses. When a pose is estimated a related signal broadcasts and the function that is related with that signal is executed. Signals control the game logic through the *Game Manager* component. This is the key component that listens to the signals produced after the gesture recognition process as is shown in the component diagram illustrated in Fig. 3. What happens in practise is that a gesture is recognized and the event that is associated with this gesture is passed through *Event2SignalMapper* and a signal is generated. *Game Manager* receives the signal and updates Menu's selected button through the *Inflator* update. The *Inflator* updates the container and displays it on the screen surface.

4   The Game in Practice

In order to assess the capabilities of the game engine and the overall system a prototype game has been developed. The story of the game is as follows: The player tries to pop bubbles with one part of his body, which in this case was the nose for two main reasons: a) it is part of the body that achieves a high score of confidence and b) it forces the player to move his whole body to pop the bubbles. The deep learning framework returns the nose coordinates and a gun-sight icon is drawn on the screen to these coordinates.

In Fig. 4 two sample levels of the game are presented, each one of which has a time limit of two minutes. In the first level the player has to move the gun-sight in a bubble and hold it there for 5 s.
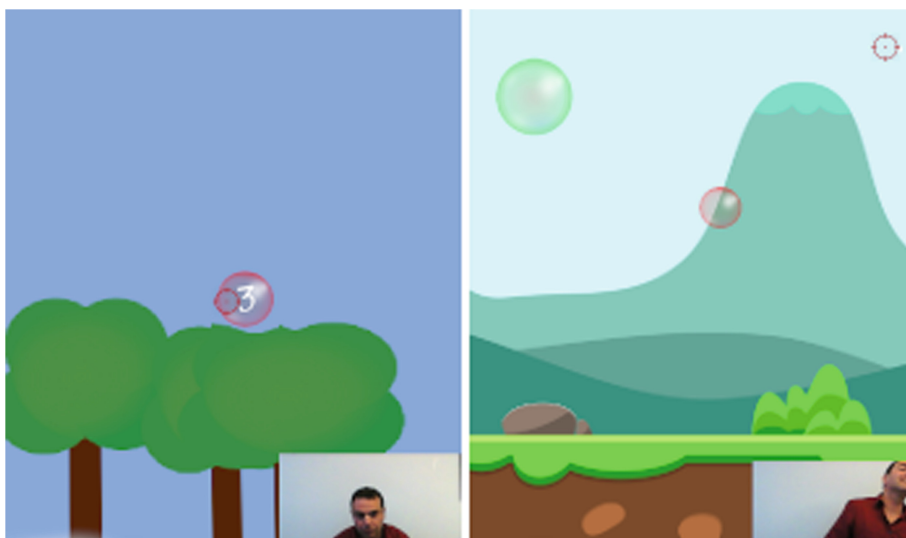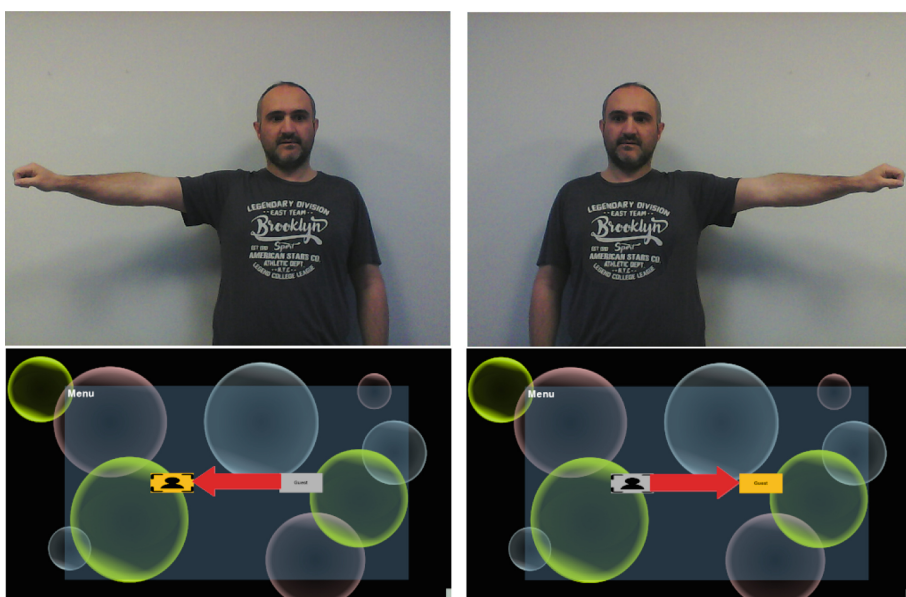
Fig. 4. (a) The user interface of the prototype game.

Fig. 5. Navigation on menu buttons: (a) Previous, (b) Next.

Following this approach, the player experiences how the gun-sight can be moved through the game and how the game perceives the player's movements. At the next level the player has to pop a number bubbles which are appearing gradually. Every effort of the player and the score of each level is kept in a database. The game also includes a main menu which is also controlled in as similar way. More specifically, the user can navigate right or left as presented in Fig. 5 and with other similar gestures, the *OK* and *Cancel* events are triggered.

As part of the system evaluation, extensive measurements have been performed on different aspects of the proposed implementation. During the game different models have been tested in respect to input size Table 1, taking into consideration the input sizes that are accepted in each model, which considerable affects the inference time of the deep learning framework. In Table 2, we measure the total cycle time (in milliseconds) in respect to frame sampling in a model with image input size 480×350. Finally, in Table 3 we depicts the frame rate of the game with 200 ms frame sampling in respect to different video input sizes.

Table 1. Inference time for different inputs with 200 ms frame sampling.

Image input size	Inference time (ms)
480×350	51
640×480	97
1200×720	272

Table 2. Total processing time for input size 480×350 for different frame samplings.

Frame sampling to model (ms)	Total processing time (ms)
300	80–85
200	90–100
150	120

Table 3. Frame rate for different inputs and sampling rate 5 Hz.

Image input size	Achieved game frame rate (fps)
480×350	10
640×480	5
1200×720	2

5   Conclusions and Future Work

The research presented in this work focused on the implementation of an innovative exergame engine exploiting state-of-the-art software technologies and the use of low-cost, commodity hardware. This was achieved by making use of methodologies and frameworks which are based on edge computing and computer vision using deep learning and convolutional neural networks. Concerning the deployment on user's environment, a Raspberry Pi device was used, bundled with a USB tensor accelerator enabling the execution of the deep learning models. The experimentation with a prototype in a real scenario presented significant results, both for the user's perception and for the system performance. The future plans include the use of more effective game control approaches based on better models for pose detection and video analysis techniques, and the support of multi-user gameplay. In addition, given the fact that the proposed exergame engine belongs to the family of serious games and targets specific user groups and use cases, more aspects of the users' physical and mental condition will be taken into consideration such as their biosignal measurements and their emotions during the game. This is expected to increase the effectiveness of the games allowing for better analysis of their health status and providing at the same time more incentives through the advanced gameplay and the enhanced game experience.

References

1. Al-Hrathi, R., Karime, A., Al-Osman, H., El Saddik, A.: Exerlearn bike: an exergaming system for children's educational and physical well-being. In: 2012 IEEE International Conference on Multimedia and Expo Workshops, pp. 489–494. IEEE (2012)
2. Benzing, V., Schmidt, M.: Exergaming for children and adolescents: strengths, weaknesses, opportunities and threats. J. Clin. Med. **7**(11), 422 (2018)
3. Brox, E., Fernandez-Luque, L., Tøllefsen, T.: Healthy gaming-video game design to promote health. Appl. Clin. Inform. **2**(2), 128–142 (2011)
4. Cass, S.: Taking AI to the edge: Google's TPU now comes in a maker-friendly package. IEEE Spectr. **56**(5), 16–17 (2019)
5. Deutsch, J.E., et al.: Nintendo wii sports and wii fit game analysis, validation, and application to stroke rehabilitation. Top. Stroke Rehabil. **18**(6), 701–719 (2011)
6. Goldstein, J., Cajko, L., Oosterbroek, M., Michielsen, M., Van Houten, O., Salverda, F.: Video games and the elderly. Soc. Behav. Pers. Int. J. **25**(4), 345–352 (1997)
7. Howard, A.G., et al.: MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)

8. Matallaoui, A., Koivisto, J., Hamari, J., Zarnekow, R.: How effective is "exergam-ification"? A systematic review on the effectiveness of gamification features in exergames. In: Proceedings of the 50th Hawaii International Conference on System Sciences (2017)
9. Menychtas, A., Doukas, C., Tsanakas, P., Maglogiannis, I.: A versatile architecture for building IoT quantified-self applications. In: 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS), pp. 500–505. IEEE (2017)
10. Papandreou, G., Zhu, T., Chen, L.C., Gidaris, S., Tompson, J., Murphy, K.: PersonLab: person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 269–286 (2018)
11. Pope, Z., Zeng, N., Gao, Z.: The effects of active video games on patients' rehabilitative outcomes: a meta-analysis. Prev. Med. **95**, 38–46 (2017)
12. Senthilkumar, G., Gopalakrishnan, K., Kumar, V.S.: Embedded image capturing system using Raspberry Pi system. Int. J. Emerg. Trends Technol. Comput. Sci. **3**(2), 213–215 (2014)
13. Skjæret, N., Nawaz, A., Morat, T., Schoene, D., Helbostad, J.L., Vereijken, B.: Exercise and rehabilitation delivered through exergames in older adults: an integrative review of technologies, safety and efficacy. Int. J. Med. Inform. **85**(1), 1–16 (2016)
14. Street, T.D., Lacey, S.J., Langdon, R.R.: Gaming your way to health: a systematic review of exergaming programs to increase health and exercise behaviors in adults. Games Health J. **6**(3), 136–146 (2017)
15. Styliadis, C., Konstantinidis, E., Billis, A., Bamidis, P.: Employing affection in elderly healthcare serious games interventions. In: Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments, pp. 1–4 (2014)
16. Susi, T., Johannesson, M., Backlund, P.: Serious games: an overview (2007)
17. Trout, J., Zamora, K.: Using dance dance revolution in physical education. Teach. Elementary Phys. Educ. **16**(5), 22–25 (2005)