An Augmented Reality Mirror Exergame using 2D Pose Estimation

Fernando Losilla and Francisca Rosique[a]
Universidad Politécnica de Cartagena, Spain

Keywords: Exergame, Augmented Reality, Body Pose Estimation, OpenPose.

Abstract: Exergames have become very popular for fitness and rehabilitation purposes. They usually rely on RGB-D sensors to estimate human 3D body pose and, therefore, allow users to interact with virtual environments. Currently, a new generation of deep learning techniques enable the estimation of 2D body pose from video sequences. These video sequences could be augmented with the estimated pose and other virtual objects, resulting in augmented reality mirrors where players can see their reflection along with other visual cues that guide them through exercises. The main benefit of using this approach would be replacing RGB-D cameras with simpler and more widely available webcams. This approach is explored in this work with the development of the ExerCam exergame. This application relies on a single webcam and the OpenPose library to allow users to perform exercises where they have to reach virtual targets appearing on the screen. A preliminary study has been performed in order to explore the technical viability and usability of this application, with promising results.

1 INTRODUCTION

Exergames are a new generation of digital gaming systems with an interface that requires physical exertion to play the game (Larsen et al., 2013). Research indicates that they can produce improvements in physical health, cognitive function as well as other non-physical effects (Anderson-Hanley et al., 2012; Larsen et al., 2013; Staiano and Calvert, 2011). In addition, they also have the potential to increase motivation and, therefore, the adherence to exercise or rehabilitation programs.

Typically, exergames are used to construct virtual environments where an avatar or other elements of the game are controlled by the movements of the player in the real world (Synofzik and Ilg, 2014; Zhao et al., 2017). As a result, they rely heavily on different technologies to track human body movement. To date, the most widely adopted of these technologies are RGB-D sensors such as the Microsoft Kinect.

An interesting alternative to the use of the RGB-D sensors can be found in recent research in machine learning techniques that allow the estimation of human body 2D pose from images (Cao et al., 2016). These techniques enable using simple RGB cameras such as webcams in exergames. In opposition to RGB-D sensors, webcams are widely spread among users of computers, easily available and have a smaller size and a much lighter weight. Consequently, they provide a great opportunity to increase the usage of exergames in home exercise and rehabilitation programs.

Due to their nature, 2D body pose estimation technologies create virtual skeletons of people in an image, being able to superimpose them on top of the original image. In that regard, they could be used to create exergames that would behave as augmented mirrors. In this approach, the original image would allow players to see their reflection. Virtual skeletons, even if not displayed, would be used as the interface to control the game or as a reference to other virtual elements. Finally, other visual cues could be superimposed to guide movements. This is the approach that will be considered in this paper.

Augmented reality has been used previously in exergames. It became very popular thanks to mobile games such as Pokemon Go (An and Nigg, 2017). However, current mobile exergames are very different in nature from the one studied in this paper and are aimed at promoting healthy lifestyles rather than specific exercises. Other approach to augmented reality consists in the projection of guidance hints on

[a] https://orcid.org/0000-0002-3311-8414

Losilla, F. and Rosique, F.
An Augmented Reality Mirror Exergame using 2D Pose Estimation.
DOI: 10.5220/0007798906430648
In *Proceedings of the 14th International Conference on Software Technologies (ICSOFT 2019)*, pages 643-648
ISBN: 978-989-758-379-7

top of body parts or other surfaces such as LightGuide (Sodhi et al., 2012) or SleeveAR (Sousa et al., 2016). A problem found with the previous proposals is that they are still too complex and expensive to be deployed in home settings.

There is other research work more in line with the proposal of this paper, in which Augmented Reality mirrors are applied. Physio@Home (Tang et al., 2014) employs them to allow patients to practice physiotherapy at home. It displays visual cues such as arrows and arm traces to guide patients through exercises and movements without the presence of a physiotherapist. MotionMA (Velloso et al., 2013) is aimed at learning complex movements, allowing expert users to specify movements, which can be repeated later by other users while receiving feedback about their performance. YouMove (Anderson et al., 2013) is also conceived as a full-body movement training in which the authors employed an actual mirror for the implementation. Finally, the NeuroR (Klein and Assis, 2013) system applies an augmented reality mirror for patients with paralyzed arms.

In this paper a new exergame that makes use of recently available 2D pose estimation libraries, called ExerCam, is introduced. Consequently, it requires a webcam instead of a RGB-D sensor. ExerCam is conceived as a general-purpose exergame which can be customized to perform simple exercises in an augmented reality mirror. A preliminary study with healthy volunteers to assess its technical viability and usability is also presented.

2 ExerCam

For the purpose of this research, the ExerCam application has been developed. It works as an augmented reality mirror that uses a screen to reflect the image of the player and augments it with virtual objects. The reflection on the mirror allows players to have visual feedback of their own body. The virtual components on top of the reflection provide additional feedback for the proper execution of the exercises.

ExerCam operation is based on the appearance of virtual objects on the screen that players have to reach with their joints. Exercises can be performed by sequentially reproducing previously programmed sequences of objects. In this way, it provides with means to stimulate physical activity, by inducing movement, or gross motor learning, as it able to provide the three key principles in motor learning: repetition, feedback and motivation (Pereira et al., 2014). However, at this stage, ExerCam has only be used to perform exercises not aimed at treating specific medical conditions, but to gain more knowledge about its technical viability.

ExerCam has three types of virtual objects (see Figure 1). First, a virtual skeleton is superimposed on top of the user's body, highlighting body joints. The joints that can be used by the exergame correspond to the keypoints used in the COCO (Lin et al., 2014) image dataset, with 17 keypoints for each medium and large sized person in the images. Superimposed joints let players know which body parts they have to use to reach target objects. When targets can be reached by any joint, virtual joints are drawn as small circles, whereas, when a particular joint has to be used, a larger circle with a color associated to the joint is displayed.

The second type of virtual objects are the target objects, currently depicted by bullseyes, that players have to reach with their joints. When they are hit, a simple sound is played in order to give additional positive feedback and increase the feeling of immersion (Jørgensen, 2008) and engagement in the game. If targets are assigned to a joint, i.e. the player has to use a particular joint for the target, a colored circle will indicate the joint that has to be used. They can be in a fixed position or move along predefined trajectories. Besides, they can be configured to appear or disappear at a certain time or in response to events such as timers being fired or previous targets being hit. For example, it is possible to schedule a certain target to appear a few seconds after all the previous targets have disappeared and thus let the player rest.

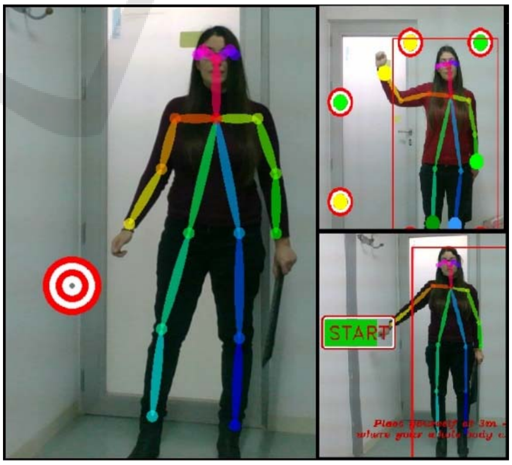

Figure 1: ExerCam virtual objects.

Finally, visual cues and instructions are the last type of virtual objects in the game. They can be further subdivided into different kinds of text (instructions, score, time elapsed, etc.), progress bar-like buttons to start an exercise and geometric shapes that help position the user and calibrate the system.

One key differentiator from most of other exergames that require specific devices (body markers) or a RGB-D sensor is that it only needs a webcam and the processing power of a graphics card. These devices are mainstream and owned by many computer users. The main limitation of the presented software, though, is that it requires a relatively powerful graphics card to operate properly. However, as the processing power of graphics cards increases every year and their price drop, this is less of a problem.

The previous requirements are imposed by the use of a novel library, OpenPose (Cao et al., 2018), the first open-source realtime library for multi-person 2D pose detection. OpenPose takes an image as the input and outputs 2D keypoints for the different body parts of the people in the image. Its use has also conditioned the usage of other programming languages and libraries in the initial implementation of ExerCam, using those in which OpenPose relies on. In particular, the C++ language has been used along with the OpenCV (Bradski and Kaehler, 2008) computer vision library. In the future, porting ExerCam to a game development platform such as Unity will be studied, as it will help in the development of more sophisticated graphical interfaces for the game.

3 EVALUATION

3.1 Outline of the Study Protocol

In order to check the system viability, a preliminary study was carried out on 28 healthy subjects with ages between 25 and 75 years. A laptop with an NVIDA GTX 1070 graphics card was used due to its mobility and computing capability to run ExerCam. An instructor guided and supervised participants from beginning to end. The protocol describing the study is detailed in Figure 2.

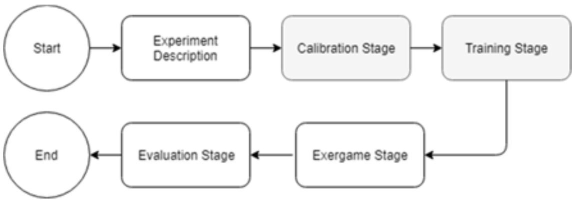

Figure 2: Stages of the study protocol.

Initially, the experiment was described to the participants. Then, in the calibration stage, the screen, the camera and the participants were positioned, with a distance between the camera and the players ranging from 2.5m to 3.5m, depending on the player height. In the Training Stage, the instructor helped subjects to perform low difficulty tasks with ExerCam. In the Exergame Stage, each of the subjects performed a set of experiments based on the execution of a task of incremental difficulty. The participants were asked to reach diverse targets (static, mobile as well as targets assigned to specific joints) as they appeared on the screen. Finally, in the Evaluation Stage, the instructor checked the results achieved by the subjects in order to determine future changes to the task.

Three different modalities of the protocol had to be used, called m1, m2 and m3. Initially, in the m1 modality, the exergame had no visual aids to help the user position during the calibration stage and the participants performed only a single repetition of the task. It was observed that some kind of aid was necessary to allow participants to place themselves properly and return to the original position during the game. This aid was introduced in the m2 modality and consisted in a rectangle on the screen where the participants had to place their body. In case that players moved from their original position during the exercise, they had to move back to it. Finally, in the m3 modality, the players were asked to perform several repetitions of the task instead of one (with a maximum of 5 repetitions, less if a plateau was observed in the results). In this way, it could be observed how their performance improved after familiarizing with the game.

In order to correctly assess the performance of the participants, some metrics related to the performance of the users were stored. These included, among other data: the response time, measured as the time elapsed from the moment that a target appears to the time when the player hits it; the duration of the execution of each task; whether targets are reached or not; the total score for each task, obtained adding the points associated with each target that was hit.

In addition, the SUS (Brooke, 1995) test was chosen to assess the usability of the system. It is a test with a reasonable number of questions (10), easily understandable and widely validated and adopted for technological-based. Each of its questions is rated on a Likert scale between 1 and 5 (from 1 for “strongly disagree” to 5 for “strongly agree”). The results of the test provides a usability score ranging from 0 to 100.

Table 1: Target hit times and percentage of misses for each modality of the study and each type of target.

Target Type	% Misses m1	% Misses m2	% Misses m3	Hit Time m1	Hit Time m2	Hit Time m3
Static	9.62%	8.92%	1.2%	2196.88ms	1856.78ms	1068.3ms
Mobile	20.06%	17.85%	2.76%	1713ms	1022.2ms	856.9ms
Joint-assigned	19.68%	15.23%	4.75%	3284.5ms	2873.4ms	1387ms
Joint-assigned Mobile	79.56%	52.14%	7.6%	3925.3ms	3568.8ms	1573.25ms
All	21.85%	16.82%	3.25%	2471ms	2282ms	1370ms

Table 2: Average execution time and information for the targets with longest and shortest response times.

Mod	Task duration	Shortest time (target)			Longest time (target)		
		Time	Position	Type	Time	Position	Type
M1	124575ms	1609ms	(600,300)	mobile	9757ms	(750,250)	Joint-assigned Mobile
M2	120929ms	445ms	(400,500)	mobile	9269ms	(750,250)	Joint-assigned Mobile
M3	61464.5ms	93ms	(600,300)	static	4635ms	(800,200)	Joint-assigned

Table 3: SUS test results.

	Answer					
Questionnaire item	**1**	**2**	**3**	**4**	**5**	$\bar{X}$
I1: I think that I would like to use this system frequently	0	1	5	11	3	3.75
I2: I found the system unnecessarily complex	14	5	1	0	0	1.35
I3: I thought the system was easy to use	0	1	3	5	11	4.3
I4: I think that I would need the support of a technical person to be able to use this system	2	8	2	2	6	2.48
I5: I found the various functions in this system were well integrated	0	0	0	2	18	4.9
I6: I thought there was too much inconsistency in this system	17	3	0	0	0	1.15
I7 I would imagine that most people would learn to use this system very quickly	0	0	0	11	9	4.45
I8: I found the system very cumbersome to use	12	6	2	0	0	1.5
I9: I felt very confident using the system	0	0	1	16	3	4.1
I10: I needed to learn a lot of things before I could get going with this system	18	2	0	0	0	1.1

3.2 Results

The protocol presented in Section 3.1 has been applied to acquire and analyze data. Table 1 shows the percentage of missed targets in each modality of the study and the average time required by the participants to hit each of the targets in the different modalities. In both cases targets have been differentiated by their type (static, mobile, joint-assigned, joint-assigned mobile).

Table 2 shows the average time (in milliseconds) required by the participants to finish the defined task (Task duration) in each of the modalities of the study. The time, position and type of the target that was hit in the shortest time, as well as the same information for the target that took the longest time, are also shown. The position of the targets have been expressed in pixels in a 1280x720 screen, where (0, 0) corresponds to the upper-left corner. Missed targets have not been taken into account for the calculations.

The test questions for the subjective assessment usability of the system by the participants in the study are shown in Table 3. The SUS score of the system is 84.8 out of 100.

4 DISCUSSION

The results in the previous section show that developing a mirror like exergame using state of the art technologies for 2D body pose is technically viable. The developed exergame received a SUS score of 84.8 from the participants in the study, which is a good result. The times that the participants needed to hit the targets, as well as the percentage of targets not missed were also satisfactory in the m3 modality of the study protocol (see Table 1), where the participants were able to perform several repetitions of the task. During the first repetition, they were not still familiarized with the interface of the game. After several repetitions using ExerCam, participants were able to perform exercises reaching almost all the targets, with less than a 4% of misses. In addition, they finished the tasks earlier, they were able to react more quickly to targets and they spent less time in the calibration stage.

It can also be observed in Table 1 that the type of virtual objects, in all of the modalities of the protocol study, is a relevant factor affecting both the percentages of misses as well as the time needed to reach the object itself. As expected, static objects subject to no joint constraint had a smaller percentage of misses. On the contrary, objects subject to both movement and assigned to a particular joint had a higher percentage of misses and, in addition, they required more time to be reached. It is surprising, though, that the objects that required the shortest time to be reached are the moving objects, as shown in Table 2. However, it has to be taken into account that these objects had a large amount of misses and that their starting position was easy to reach.

The introduction of visual aids in m2 also led to better results. Firstly, the use of aids to help participants go back to their original position was helpful in performing exercises as they were originally intended. That is, targets appeared at the expected position relative to the user. In addition, other aids such as the use of a superimposed virtual button to start the game was also helpful. In the m1 modality, without this button, participants required some time to react at the beginning of the game and missed the first or firsts targets. Using a button activated by their left wrist, which has to be held over the button until a progress bar is full, let users to remain focused at the beginning of the task.

There are also some considerations and limitations that have to be remarked about the software and the study. First, and more important, a relatively powerful graphics card is required. An NVIDIA GTX 1070 had to be used because of this. The computers available with lower end graphics cards offered an unacceptable delay between the acquisition of images from the webcam and the finalization of their processing by the OpenPose library. In addition, it could be observed that using the standard settings in OpenPose also resulted in large delays (even with the GTX 1070 card). We found that dropping the number of lines of the images to 128 (228 x 128 resolution using a 16:9 aspect ratio) still provided good tracking with a barely noticeable delay. This is the resolution of the images processed by OpenPose in the study. Further reductions in the resolution resulted in poor tracking. It was possible, though, to perform partial body estimation (for example, upper limbs) using 96 lines of input resolution, but it required the user to be close to the camera in order to be detected and showed some instability in the estimations.

The size of the screen may also be a concern with older adults. Most of the participants of the study were young and, therefore, the study did not focus on age. However, it was observed that older players tended to get close to the screen in order to see the screen better, which can affect their execution of tasks. This group of users also reported in the SUS questionnaire that they would need the support of a technical person in order to be able to use the system.

Finally, it has to noted that ExerCam only requires determining the position of the joints in order to operate. This is not problematic using 2D body pose estimation libraries. However, the development of exergames that require more complex analysis of the pose, for example for the recognition of gestures or movements could be more challenging than with 3D poses from RGB-D cameras.

5 CONCLUSSIONS

This paper studies the feasibility of applying new 2D body pose estimation techniques to the development of exergames based on augmented reality mirrors. To that end, the ExerCam application was developed. It works as an augmented reality mirror that overlays virtual objects on images captured by a webcam. It can be used for different purposes, allowing trainers or therapists to define tasks based on the appearance of virtual targets on the screen, which the player has to reach. A preliminary study has been carried out to assess the viability of this approach, with satisfactory results. Its use in specific medical conditions has not been studied. However, the application allows the creation of tasks that could promote physical exercise and gross motor skills.

In spite of some limitations, emerging 2D body pose estimation methods based on deep learning are a promising solution for augmented reality exergames, particularly for those based on augmented reality mirrors. They only require a computer with enough processing power and a webcam. In addition, webcams are cheap and widely available devices that can be easily installed due to their small size and weight.

ACKNOWLEDGEMENTS

This research has been supported by the AEI/FEDER, UE project grant TEC2016-76465-C2-1-R (AIM) and the “Research Program for Groups of Scientific Excellence in the Region of Murcia" of the Seneca

Foundation (Agency for Science and Technology in the Region of Murcia) [19895/GERM/15], Spain.

REFERENCES

An, J.-Y., Nigg, C.R., 2017. The promise of an augmented reality game—Pokémon GO. *Ann. Transl. Med. 5*. https://doi.org/10.21037/atm.2017.03.12

Anderson, F., Grossman, T., Matejka, J., Fitzmaurice, G., 2013. YouMove: Enhancing Movement Training with an Augmented Reality Mirror, in: *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, UIST '13*. ACM, New York, NY, USA, pp. 311–320. https://doi.org/10.1145/2501988.2502045

Anderson-Hanley, C., Arciero, P.J., Brickman, A.M., Nimon, J.P., Okuma, N., Westen, S.C., Merz, M.E., Pence, B.D., Woods, J.A., Kramer, A.F., Zimmerman, E.A., 2012. Exergaming and older adult cognition: a cluster randomized clinical trial. *Am. J. Prev. Med. 42*, 109–119. https://doi.org/10.1016/j.amepre.2011.10.016

Bradski, G., Kaehler, A., 2008. Learning OpenCV: Computer Vision with the OpenCV Library, Edición: 1. *ed. O'Reilly Media*, Beijing.

Brooke, J., 1995. SUS: A quick and dirty usability scale. *Usability Eval Ind 189*.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., Sheikh, Y., 2018. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *ArXiv181208008* Cs.

Cao, Z., Simon, T., Wei, S.-E., Sheikh, Y., 2016. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *ArXiv161108050 Cs*.

Jørgensen, K., 2008. Left in the dark: playing computer games with the sound turned off. *Ashgate*.

Klein, A., Assis, G.A. d, 2013. A Markeless Augmented Reality Tracking for Enhancing the User Interaction during Virtual Rehabilitation, in: *2013 XV Symposium on Virtual and Augmented Reality. Presented at the 2013 XV Symposium on Virtual and Augmented Reality*, pp. 117–124. https://doi.org/10.1109/SVR.2013.43

Larsen, L.H., Schou, L., Lund, H.H., Langberg, H., 2013. The Physical Effect of Exergames in Healthy Elderly-A Systematic Review. *Games Health J. 2*, 205–212. https://doi.org/10.1089/g4h.2013.0036

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollár, P., 2014. Microsoft COCO: Common Objects in Context. *ArXiv14050312 Cs*.

Monge Pereira, E., Molina Rueda, F., Alguacil Diego, I.M., Cano De La Cuerda, R., De Mauro, A., Miangolarra Page, J.C., 2014. Use of virtual reality systems as proprioception method in cerebral palsy: clinical practice guideline. *Neurol. Engl. Ed. 29*, 550–559. https://doi.org/10.1016/j.nrleng.2011.12.011

Sodhi, R., Benko, H., Wilson, A., 2012. LightGuide: Projected Visualizations for Hand Movement Guidance, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12*. ACM, New York, NY, USA, pp. 179–188. https://doi.org/10.1145/2207676.2207702

Sousa, M., Vieira, J., Medeiros, D., Arsenio, A., Jorge, J., 2016. SleeveAR: Augmented Reality for Rehabilitation Using Realtime Feedback, in: *Proceedings of the 21st International Conference on Intelligent User Interfaces, IUI '16*. ACM, New York, NY, USA, pp. 175–185. https://doi.org/10.1145/2856767.2856773

Staiano, A.E., Calvert, S.L., 2011. Exergames for Physical Education Courses: Physical, Social, and Cognitive Benefits. *Child Dev. Perspect. 5*, 93–98. https://doi.org/10.1111/j.1750-8606.2011.00162.x

Synofzik, M., Ilg, W., 2014. Motor Training in Degenerative Spinocerebellar Disease: Ataxia-Specific Improvements by Intensive Physiotherapy and Exergames [WWW Document]. *BioMed Res. Int.* https://doi.org/10.1155/2014/583507

Tang, R., Alizadeh, H., Tang, A., Bateman, S., Jorge, J.A.P., 2014. Physio@Home: Design Explorations to Support Movement Guidance, in: *CHI '14 Extended Abstracts on Human Factors in Computing Systems, CHI EA '14*. ACM, New York, NY, USA, pp. 1651–1656. https://doi.org/10.1145/2559206.2581197

Velloso, E., Bulling, A., Gellersen, H., 2013. MotionMA: Motion Modelling and Analysis by Demonstration, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13*. ACM, New York, NY, USA, pp. 1309–1318. https://doi.org/10.1145/2470654.2466171

Zhao, W., Reinthal, M.A., Espy, D.D., Luo, X., 2017. Rule-Based Human Motion Tracking for Rehabilitation Exercises: Realtime Assessment, Feedback, and Guidance. *IEEE Access 5*, 21382–21394. https://doi.org/10.1109/ACCESS.2017.2759801.