SHS Web of Conferences **194**, 03006 (2024) https://doi.org/10.1051/shsconf/202419403006
ETLTC2024

Interactive Game Development and Evaluation Based on Device-Free Pose Estimation Techniques

Toshizumi Abe[1], *Cher Yen* Tan[1], *Takashi* Imamura[2], and *Tatsuya* Yamazaki[2*]

[1]Graduate School of Science and Technology, Niigata University, 8050, Ikarashi 2-Nocho, Nishi-ku Niigata, 950-2181, Japan
[2]Faculty of Engineering, Niigata University, 8050, Ikarashi 2-Nocho, Nishi-ku Niigata, 950-2181, Japan

Abstract. In recent years, significant advancements have been made in pose estimation methods. These methods can be broadly divided into two categories: device-based and device-free. Device-based methods, such as virtual reality data gloves and marker-based motion capture, are known for their accuracy. However, these require specific equipment, making them less accessible for the public. On the other hand, device-free methods need no device to recognize human movement. They usually use cameras and estimation algorithms to recognize the body parts. Owing to evolving artificial intelligence (AI) technology, estimation accuracy has increased and we adopt one of the device-free methods to develop two interactive games, "Brain Wall" and "Touch de Pose". "Brain Wall" challenges players to imitate a silhouette, scoring them based on the accuracy of their pose compared to the silhouette using the Intersection over Union metric. The game encourages competitiveness and participation through a leader board system. "Touch de Pose" allows players to choose a pose theme and the players are required to position specific body parts within the displayed circles on the screen. "Touch de Pose" also includes real-time evaluation of the player's pose and combines their image with themed background images generated by generative AI, some of which demonstrate the limitations of the technology. The games were showcased at a university open campus event, receiving overwhelmingly positive feedback: 95.6% satisfaction for "Brain Wall" and 98.1% for "Touch de Pose". Our study shows that pose estimation-based interactive games show immense potential in generating interest in technology for those who are unfamiliar with information and communication technology. Additionally, the full-body engagement aspect of these games could also play a role in promoting physical activity as a regular habit.

1 Introduction

* Corresponding author: yamazaki.tatsuya@ie.niigata-u.ac.jp

Pose estimation plays an important role in the world of interactive games. Recent advancements in this field have greatly improved the gaming experience. In general, there are two approaches to this technology: device-based and device-free methods.

Device-based methods, as the name suggests, rely on devices to detect human movement. A common example is the use of data gloves in virtual reality (VR) applications. These gloves come equipped with sensors such as accelerometers and gyroscopes. When users wear these gloves along with VR goggles, they can recognize and interact with their virtual hands in the VR space. This setup creates an immersive experience that allows users to interact with objects within the virtual space.

Another device-based technology, known as the motion capture, is also widely used in various fields to capture and record human movements. It involves attaching markers or sensors to specific parts of a person's body, usually at joints. Multiple time-synchronized cameras are placed around the subject to capture their movements simultaneously. By analysing the camera footage, the 3-dimensional (3D) positions of these markers or sensors can be calculated accurately. This enables replication of movements through a 3D avatar. Although precise tracking is possible with this method, it does come with the drawback of requiring a setup of specialized equipment which can be challenging for casual or spontaneous pose estimation. Additionally, the cost associated with motion capture technology tends to be high.

On the other hand, device-free methods offer a different approach by recognizing human movement without the need for individuals to wear any devices. While these methods generally have lower performance compared to device-based approaches, they have a significant advantage in that they eliminate the requirement of attaching sensors or markers to the body. Typically, these methods involve the use of cameras and sensors to detect faces and body parts. For instance, Microsoft's Kinect, which has a depth camera, is an example of such technology [1].

The advancement of machine learning techniques such as deep learning has greatly improved device-free pose estimation capabilities. These advancements have led to systems like OpenPose, a pose estimation framework developed by Carnegie Mellon University [2]. Using a single camera, OpenPose is able to extract human movement information by estimating and connecting main joints of a person in a video using straight lines in real time. The feature points extracted during this process are referred as keypoints.

MediaPipe, developed by Google, is an open source framework that is designed to perform machine learning tasks like pose estimation for live streaming [3]. It also extracts keypoints on the human body and is specifically designed to work efficiently on mobile and edge devices. Compared to OpenPose, MediaPipe requires less computational power. This makes MediaPipe particularly suitable for real-time applications such as interactive gaming, where a heavy computational load is undesirable. Acknowledging its capabilities, Mahajan et al. [4] have introduced six application categories for MediaPipe Pose detection, including entertainment and gaming. However, they did not provide detailed examples in their overview. In [5], Jyotsna et al. suggested a method for physically disabled individuals that used head movements as a controller of a game using MediaPipe Face Mesh. Recently, Kim et al. [6] advanced this field by building a real-time human pose recognition system based on MediaPipe that demonstrated estimation of various types of pose changes. Their experiments showcased the potential of MediaPipe.

As the world of information and communication technology (ICT) continues to expand, perceptions of artificial intelligence (AI) vary greatly among individuals. According to a survey conducted in Japan [7], people generally hold positive views of AI, seeing it as something that "enriches lives" and has a "positive impact on daily life". However, there are also negative perceptions with some "feeling anxious" or finding it "somewhat scary." These mixed opinions suggest that the concept of AI may overshadow its applications, leading to misunderstandings about how it can be used in products and services. In order to establish a connection with AI, it is important to create opportunities for people to interact with AI-powered products and services which can generate interest and understanding towards AI.

The main objective of this study is to develop interactive games using pose estimation technology that require no specialized equipment. By eliminating the need for any devices, these games become more accessible especially for those who are unfamiliar with ICT. This approach not only has the potential to spark interest in AI but also offers a fun and engaging solution to combat physical inactivity by involving full body movement and encouraging physical engagement.

In this paper, we introduce two games that utilize pose detection technology: “Brain Wall” and “Touch de Pose”. The main focus of “Brain Wall” is to provide a gaming experience rather than solely highlighting the AI technology behind it. It challenges players to perform poses against a virtual wall, with a strong emphasis on fun and competition. On the other hand, “Touch de Pose” aims to provide knowledge to players about AI technology in an interactive way. By engaging with the game, players can learn more about AI and its applications. To align with the recent trends in AI, we have integrated Midjourney, an image generative AI model into the game [8]. This integration allows us to notify the latest advancements in AI.

For both games, we selected MediaPipe as our base pose detection model due to its computational performance, which is crucial for real time gaming. Additionally, we present and discuss the insights gathered from surveys completed by participants who had the opportunity to experience these games at an open campus event.

2 System architecture

In this study, we develop two interactive games using a shared system architecture. As shown in Fig. 1, the system consists of a personal computer (PC), a camera, speakers, a projector, and a screen. The games are designed for one single player in front of the screen. We project the visuals from the PC onto the screen from behind. In order to capture the player’s poses effectively, the camera is placed below the screen, facing towards the player. The reason behind using rear-screen projection is to avoid any disruption between the projector and the player, which is crucial considering the body movements required in both games. Minimizing any obstruction or interference with the projector during gameplay is important to provide a smooth gaming experience. Similarly, speakers are placed under the screen, facing towards the player, in order to enhance their gaming experience.

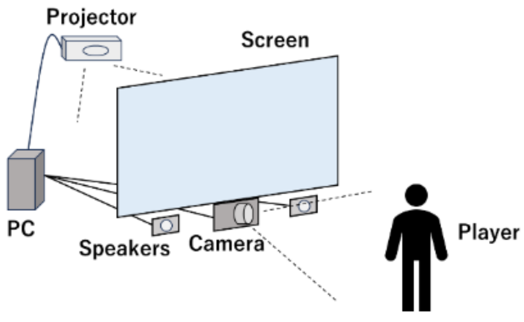

Fig. 1. Configuration of the game system.

3 Game flows

3.1 Brain Wall

"Brain Wall" is an interactive game that challenges players to imitate a displayed silhouette on the screen. In this game, the screen symbolizes a wall, with the silhouette represents a hole in it. The goal for players is to align their pose with the silhouette as closely as possible, creating the illusion of passing through the hole in the wall. We use the semantic segmentation feature of MediaPipe's pose estimation model to capture the player's pose. The score is calculated based on the similarity between the player's pose and the specified silhouette using Intersection over Union (IoU), a common metric in object detection.

The game flow of "Brain Wall" and the opening image are shown in Figs. 2 and 3 respectively. On the opening image, players can choose to start the game in Easy mode by pressing "E", in Hard mode by pressing "H", or start a tutorial by pressing "T". They can also exit the game by pressing the bottom right "Quit" button.

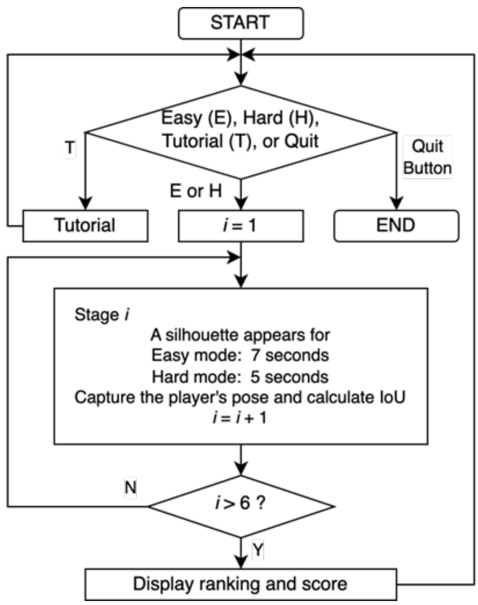

Fig. 2. Game flow of "Brain Wall".

Fig. 3. The opening image of "Brain Wall".

The game mode consists of six stages. In each stage, a silhouette is displayed for a certain duration (7 seconds for the Easy mode and 5 seconds for the Hard mode), and the player moves their body to imitate the pose of the silhouette. Simultaneously, the player's image is captured by the camera, and their pose is extracted using the semantic segmentation feature of MediaPipe Pose Landmark Detection. The process of the player posing according to the displayed silhouette is shown in Fig. 4.

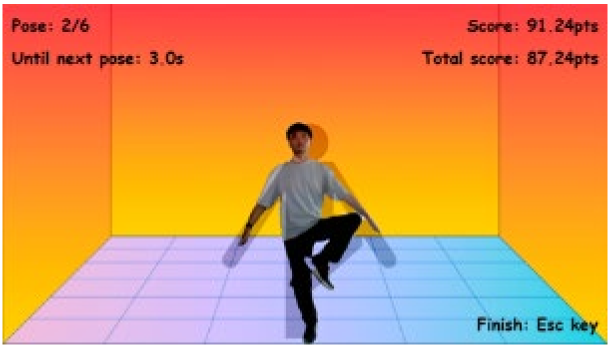

Fig. 4. A posing player in “Brain Wall”.

As for the tutorial mode shown in Fig. 5, the content provided to the player is similar to that in the game mode, with only a single silhouette displayed for 99 seconds. We show the semantic segmentation mask on the top right window allowing the player to grasp the concept of the game before moving to the game mode. During the tutorial session, we briefly introduce the logic of the AI technology used in this game.

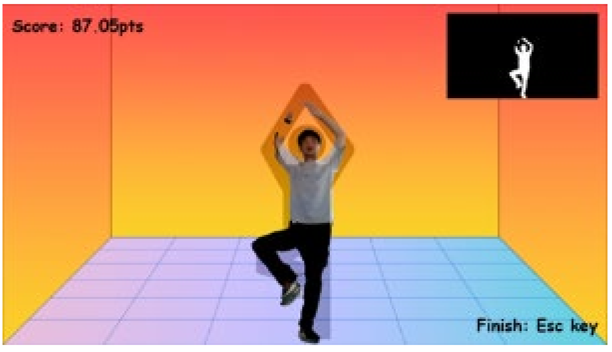

Fig. 5. The tutorial mode of “Brain Wall”.

A score based on IoU is calculated at each stage, and the total score shown is the sum of the scores from all stages. After the 6th stage of the game mode, the final score is displayed in a ranking format, which is called the leader board. It allows players to compare their scores with those of previous players. The leader board is shown in Fig. 6.

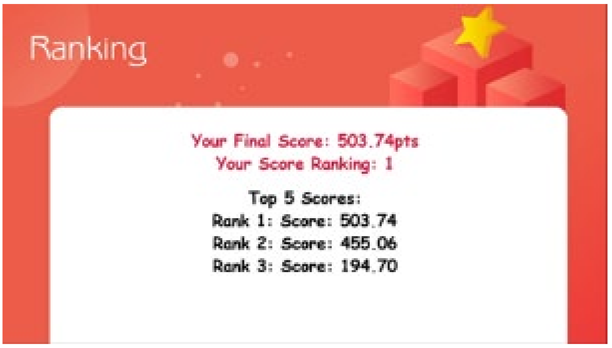

Fig. 6. The leader board of "Brain Wall".

The leader board adds an element of competition to the game, making it not only fun but also encouraging interaction among players. Additionally, by offering two modes (Easy/Hard), players have the flexibility to choose a difficulty level based on the previous players' gameplay.

3.2 Touch de Pose

"Touch de Pose" provides an interactive and creative gaming experience where players can choose from a variety of pose themes, take photos based on these themes, and then view their images combined with background image.

The game flow and the opening image are depicted in Figs. 7 and 8 respectively. On the opening image, players can choose to either go through a tutorial or start the game.

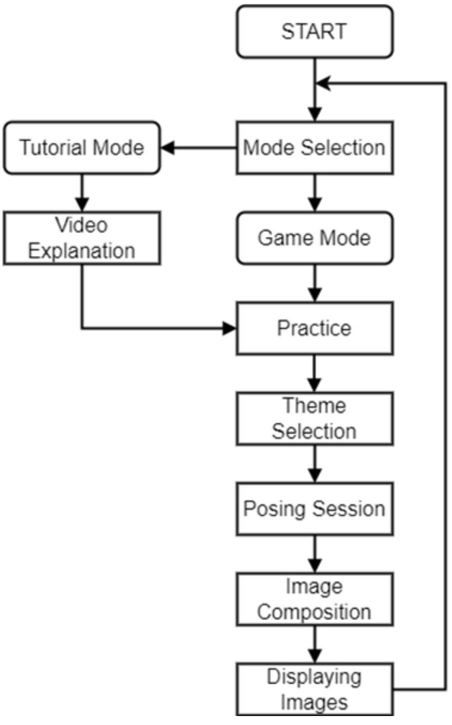

Fig. 7. Game flow of "Touch de Pose".

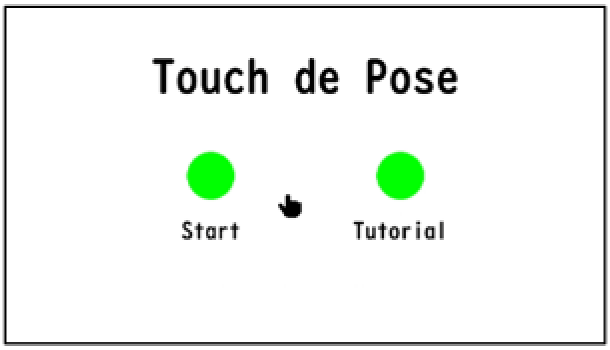

Fig. 8. The opening image of “Touch de Pose”.

In “Touch de Pose”, the players have control over the game’s progression without the need of interfaces such as keyboards or controllers. This is achieved by using the pose estimation model from MediaPipe. The system recognizes the player's right index finger and it is used as a user interface for the game. The player can indicate intention of selection by using the user interface of the right index finger in the game. This method of using the body itself as an input interface allows for a more natural and intuitive gameplay experience. As shown in Fig. 9, players can select a specific circle by aligning their right index finger over it on the screen for 5 seconds. Then it triggers a transition to the next screen. During this selection process, the colour of the chosen circle changes from green to pink, providing a clear visual indication of the player's choice.

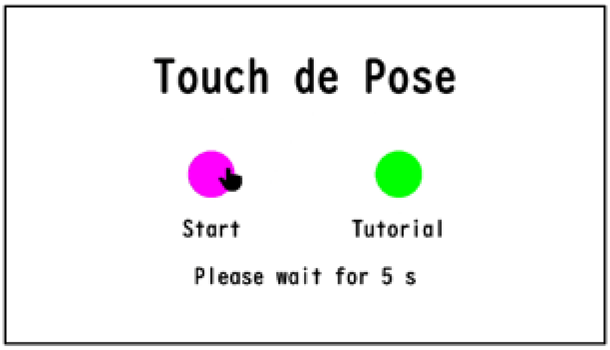

Fig. 9. Selection by the right index finger.

When the player selects the “Tutorial” circle in the opening image, a short video is played to provide the game flow and play instruction. After the video, the player is directed to a practice session, where the player can practice using a very simple pose. This helps the player to get used to the game content. When the player selects the “Start” circle, the player is led to the posing session. Then, in the theme selection image as shown in Fig. 10, the player selects a theme from four different options: “Seaside”, “Mountain”, “Sports”, and “Winter”.

Fig. 10. The theme selection image of "Touch de Pose".

During the posing session, three circles labelled with "H", "R", and "L" appear on the screen, guiding the player in positioning their body parts. The "H" circle indicates where to place the head, "R" is for the right hand, and "L" for the left hand. Achieving a high degree of alignment turns the circles from red to green, this provides a visual confirmation of the player's accurate overlapping with the displayed circles. The player has to keep the pose for 10 seconds, and the pose image is captured as a photo at the moment the countdown reaches zero. The system evaluates how well the player aligns with these positions in real time and the player repeats pose alignment three times. A guidance image of the pose is also shown at the bottom right of the screen. Figs. 11 and 12 show low and high degree alignment cases of the left hand respectively. In both figures, the ten second countdown is shown in the upper left corner and the small image on the bottom right corresponds to the guidance image of the pose.

Fig. 11 A low degree alignment case (see the left hand).

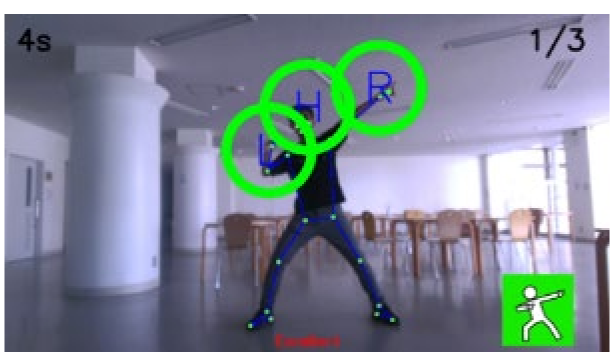

Fig. 12 A high degree alignment case (see the left hand).

After capturing the poses, the player's image is automatically trimmed and combined with a prepared background image that matches the selected theme. The background images are generated by Midjourney, and we intentionally include some unrealistic designs to notify the limitations of AI. For example, within the 'Sports' theme, there are basketball court images like the one shown in Fig. 13, where the number of basketball hoops and court lines are not correctly generated. Fig. 14 shows a composite image with well-generated image, while Fig. 15 shows a composite image with unrealistic generated image. Additionally, since the composition process takes time to generate images, a slideshow is presented to the player to explain the AI technology used in this game.

Fig. 13 A basketball court image generated by Midjourney.

Fig. 14 A composite image with a well-generated background image.

Fig. 15 A composite image with an unrealistic generated background image.

4 Evaluation

These two interactive games based on device-free pose estimation techniques were unveiled to the public at the Niigata University's open campus event, held on August 9th and 10th, 2023. The visitors who played these games were asked to anonymously complete surveys about their experience. The surveys were conducted via Google Forms using mobile devices provided at the event. 45 responses were collected for "Brain Wall" and 53 for "Touch de Pose". Since the target audience for the open campus event is mainly high school students, many of the participants in the game were teenagers, and the ratio of teenagers in the evaluation survey was 93.9%.

From the "Brain Wall" survey, 97.8% of participants reported that they could move the body significantly during the game, indicating the game's effectiveness in encouraging physical activity. In terms of satisfaction, 95.6% were satisfied with the game. The high satisfaction rate is largely due to its fun, engaging gameplay, and the competitive edge provided by the leader board and scoring system, which motivated the players to actively participate and improve their performance.

In the "Touch de Pose" survey, we initially asked if the players had previously used a generative AI tool, to which 57.1% of them answered "No". After the game, we gathered feedback on how well they understood the AI technology used in the game. 96.2% of the participants found the explanation easy to understand, and 98.1% enjoyed the game. These results indicate that the game's objective was successfully attained to provide knowledge about AI technology in an interactive manner.

5 Conclusion

In this study, we developed two interactive games based on pose estimation technology. These games capture the players' poses without having to wear anything on the body, and were designed so that anyone can enjoy them intuitively. The completeness of the games was confirmed by having unspecified visitors try them out at the open campus of Niigata University. The feedback obtained from participants who played these games showed a high level of satisfaction, suggesting that we successfully created engaging and enjoyable gaming experiences.

We also tried to increase interest in AI through the interactive games based on pose estimation technology, and found that it was particularly effective among people who are not familiar with AI. Furthermore, because these games require to use several body parts simultaneously, they may also be useful for promoting the habit of regular physical activity as an easy and light form of exercise.

Mr. Ryotaro Akagawa, Mr. Ryo Ito, Mr. Keisuke Sato, Mr. Koya Takizawa, and Mr. Yuya Mito cooperated with the public exhibition of "Brain Wall" and Mr. Chaowei Luo, Mr. Kosaku Sato, Mr. Kotaro Takano, Ms. Arisa Matsuki, Ms. Kanon Saito, and Mr. Jiahong Gang cooperated with the public exhibition of "Touch de Pose". They also contributed to the evaluation survey of the participants. Furthermore, Mr. Ryo Ito and Mr. Koya Takizawa participated in some of the programming of "Brain Wall". The authors would like to express their sincere gratitude for their generous cooperation.

References

1. T.-Q. Wang, Y. You, K. Osawa, M. Shimodozono, E. Tanaka, A Remote Rehabilitation and Evaluation System Based on Azure Kinect. Journal of Robotics and Mechatronics. **34**, 1371-1382 (2022)

2. Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, Y. Sheikh, OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. IEEE Transactions on Pattern Analysis and Machine Intelligence. **43**, 172-186 (2019)
3. MediaPipe (online), available from 〈https://developers.google.com/mediapipe〉, (accessed 2023-12-16)
4. P. Mahajan, S. Gupta, D.K. Bhanushalii, Body Pose Estimation using Deep Learning. International Journal for Research in Applied Science & Engineering Technology. **11**, 1419-1424 (2023)
5. L.N.B. Jyotsna, N. Jagadeesh, G. V. K. Reddy, S.J. Pradeesh, N.H. Raj, REAL TIME GAME PLAY USING HEAD POSE ESTIMATION. Juni Khyat Journal. **12**, 554-561 (2022)
6. J.-W. Kim, J.-Y. Choi, E.-J. Ha, J.-H. Choi, Human Pose Estimation Using MediaPipe Pose and Optimization Method Based on a Humanoid Model. Applied Sciences. **13**, 2700 (2023)
7. AI Working Group, Study Group on Consumer Response to Digitalization: Results of the first consumer awareness survey (online, in Japanese), available from 〈https://www.caa.go.jp/policies/policy/consumer_policy/meeting_materials/assets/consumer_policy_cms101_20316_03.pdf〉, (accessed 2023-12-16)
8. Midjourney (online), available from 〈https://www.midjourney.com/〉, (accessed 2023-12-16)