HandReha: Dynamic Hand Gesture Recognition for Game-based Wrist Rehabilitation

Farnaz Farahanipad*
Harish Ram Nambiappan*
The University of Texas at Arlington
Arlington, Texas, USA
farnaz.farahanipad@mavs.uta.edu
harishram.nambiappan@mavs.uta.edu

Ashish Jaiswal
The University of Texas at Arlington
Arlington, Texas, USA
ashish.jaiswal@mavs.uta.edu

Maria Kyrarini
The University of Texas at Arlington
Arlington, Texas, USA
maria.kyrarini@uta.edu

Fillia Makedon
The University of Texas at Arlington
Arlington, Texas, USA
makedon@uta.edu

ABSTRACT

Hand-gesture recognition systems have recently gained more popularity. Moreover, there is a growing interest in building games for other purposes apart from entertainment, such as education and rehabilitation. This paper focuses on developing a novel game-based system for wrist rehabilitation, called HandReha. The idea is to automatically recognize pre-defined hand gestures using a web-camera, so to control an avatar in a three dimensional maze run game. The pre-defined gestures are picked from a pool of well-defined gestures suitable for wrist rehabilitation. Deep learning techniques were utilized to perform real-time hand gesture recognition from the images. To evaluate the performance of the developed wrist rehabilitation system, a preliminary study with 12 healthy participants was conducted. The results showed that the developed wrist rehabilitation system is intuitive and engages the user, which is crucial for rehabilitation purposes.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; • **Applied computing → Computer games**; • **Computing methodologies → Image segmentation**.

*Both of the authors contributed equally and consider as the first author

Woodstock '18, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

KEYWORDS

gesture recognition, convolutional neural networks, hand rehabilitation, gesture-based gaming

ACM Reference Format:
Farnaz Farahanipad, Harish Ram Nambiappan, Ashish Jaiswal, Maria Kyrarini, and Fillia Makedon. 2018. HandReha: Dynamic Hand Gesture Recognition for Game-based Wrist Rehabilitation. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Rehabilitation is the recovery from the lost ability to control or coordinate a body part of the patient through a repetitive task. Based on worldwide statistics annually, more than 15 million people suffer from stroke, which is one of the main causes for people to lose their control over their body-part such as hands or feet [1]. Though physical rehabilitation can help patients regain the ability to control the affected body-parts for many decades, conventional rehabilitation exercises tend to not engage stroke patients due to tedious and repetitive nature [19]. Moreover, it is important to perform these rehabilitation exercises at home to improve their recovery. Therefore, a new area of research focuses on developing gamified approaches for various purposes other than entertainment or recreation. For instance, in [8] they designed and developed an automated Activate Test for Embodied Cognition (ATEC), which measures cognitive skills through physical activity.

Moreover, based on the initial testing in [27], they claimed that the interactivity aspect of the game had the potential which cause more enjoyment by the user and it can be used to engage a diverse population and clinical setting. Recently, Ramirez at al. [29] proved that how the design of an augmented reality (AR) increase the engagement in independent stroke rehabilitation. In 2018, Heng-Yi chen et al. designed a smart suit, Heath Posture Protector (HP2), and 4 serious games for rehabilitation of office workers who usually suffer from neck and back pain [5]. Furthermore, there is a growing interest in research to utilize gesture recognition as a viable interface between humans and computers [10] [15]. Gesture-based interfaces can be an alternative to several physical peripherals that we use with our personal computers [2]. In [6], Chiang Wei Tan et al. introduced a human computer interaction system by using hand gesture recognition to apply into game-based rehabilitation application. However, in most of these studies, the focus was on design and development of 2D games or performing a limited variety of tasks such as going up or down or performing ordered-actions repeatedly.

In this paper, a game-based wrist rehabilitation system was developed, called HandReha. It recognizes pre-defined hand gestures performed in front of a web camera. The recognized gestures are used to control the actions of an avatar in a three-dimensional maze run game. The system runs on a personal computer equipped with a web-camera, which enables patients to perform their hand rehabilitation exercises at the comfort of their home. Additionally, as the hand recognition is vision-based, no additional sensors, such as Armbands [24], special suits [5], gloves, motion sensors or depth-cameras [16][27] are needed. The system uses images from the web-camera to recognize the gestures. The raw images are first pre-processed using image processing techniques followed by a background subtraction for hand segmentation. Subsequently, the segmented hand in the processed image is sent as an input to a three-layered convolutional neural network that classifies the type of gesture performed. Based on the classified gesture type, the avatar in the designed maze run game performs actions such as moving, rotating and shooting hostile drones in the 3D environment. Five different gestures were selected from a well-researched pool of medically-approved gestures suitable for hand therapies[28].

Despite the works in [6] and [32], where they restricted their focus on using gestures for simulating mouse and certain keyboard events for system interaction, we focused on using gestures to directly interact with the system and navigate the avatar in the 3D game. Besides, in many studies, they try to collect hand gestures through a variety of sensors, including gloves, electromagnetic or optical position and orientation sensors for the wrist. Yet wearing gloves or trackers, as well as other mentioned sensors, is uncomfortable and time-consuming approaches due to setup time or calibration process. Finally, due to our computer-vision-based interfaces, there are several notable advantages such as providing a non-intrusive approach where the hardware is commercially available at a lower cost compared to other approaches. Additionally, by applying the background subtraction method to segment hand we were able to achieve a robust gesture classifier model that can classify the performed gesture in a real-time manner with the highest accuracy of 98.8, which is in the same range of the state of the art methods.

The following sections of the paper are organized as follows. Section 2 describes the background and related work is done in gesture recognition. Section 3 explains the methodology and architecture of the entire system. Section 4 describes the gesture classifier,the data-set obtained for this system, followed by the achieved result in detail. Section 5 describes the game design and development. Section 6 explains the experimental user study and the devices used for experimentation and finally the results obtained through pre and post-survey results for our experiment. Section 7 discusses the conclusion and future work to be done in this system respectively.

2 BACKGROUND AND RELATED WORK

In recent years, as the percentage of older adults increased, the need for medical and rehabilitation dramatically raised. Furthermore, motivating game-based training improved therapy for people with physical impairments. As a result, research works that are performed in the area of developing games for rehabilitation purposes became extremely popular. Moreover, some research has been done in integrating gestures with a human-computer interaction system. [32] focused on integrating human gesture recognition in a human-computer interaction system by using gestures for certain mouse events such as mouse hold, mouse drag and mouse click along with certain keyboard events such as up and left keyboard press and used those gestures for certain scenarios such as playing angry birds game and working in Robot Operating System (ROS). In [4], Yassine Bouteraa et al. developed a customized augmented reality system for stroke rehabilitation.

We were inspired by the work done in [12] that implemented the use of gesture recognition for therapy. There has been compelling research on hand gestures that help in increasing a joint's range of motion or lengthen the muscle and tendons via stretching. Some of the popular motions like wrist ulnar and radial deviation gestures that are effective for improvements in hand motion are included as part of gestures in our system. This study has a focus on assisting patients to carry out rehabilitation with hand gesture recognition [27]. The process is supported by the patient playing a game on their computer to make the process interesting and make it as a rehabilitation session.

As noted in the introduction, the proposed model has two main parts for the vision-based hand gesture recognition; background subtraction along with hand detection and gesture classification. For the image processing part, we apply a background removal technique by taking a continuous average of the background at the start of the game, which then acts as a threshold throughout the game. Any object or obstacle that forms a contour in the image after the background is subtracted is detected by the system. This detected contour is passed to our classifier which predicts one of the five possible gestures.

We utilized some naive approaches explained in [20][3] to apply background subtraction and extract relevant information about the hand. By doing so, we were able to remove unnecessary noise from the background making it easier for the neural network to perform the classification. However, hand detection and background removal using only an RGB camera can be relatively tough depending on the lighting conditions and objects present in the background. So, we used a fixed bounded box that would enclose the person's hand. Therefore, the recognition task is done faster, processing the smaller area, which is a key factor to have a real-time hand gesture recognition.

In deep learning, a convolution neural network (CNN, or ConvNet) is mostly applied to analyze visual imagery[22]. Ever since CNNs gained popularity back in 2012 after the revolutionary success of AlexNet on the public dataset ImageNet, they have come into the hype and been implemented in a lot of research areas. We wanted to apply similar methods to the datasets we collected after applying the processing techniques. Since CNNs are mostly applicable for image data, we first started with a simple 3-layer convolutional neural network in PyTorch [26] (a popular open-source deep-learning library). Besides, We implemented transfer learning with pre-trained models like ResNet50 and VGG16 [17] [30]. On comparing the results, the CNNs trained from scratch performed better than other methods for our use-case.

3 METHODOLOGY

In this paper, we design and develop a unique Maze run game such that the user can perform specific hand movements (according to their therapy treatment) in front of the web-camera and control the character in the game through the maze. As a result, doing the therapy through the game can motivate and interest the patient to continue their treatment without getting bored, all the user needs is a personal computer with web-camera. As soon as the user starts performing the predefined gestures in front of the web-camera, the avatar in the game will perform several actions such as starting the game, moving forward, rotating right/left and shooting hostile drones in the maze. The conceptual model for our proposed method is shown in figure 1.

In summary, the network consists of two stages, one for hand detection and the second stage for gesture classification. Once the gesture is shown by the candidate the hand is detected and then the gesture classification process is performed on the hand image. The gesture classification is explained in detail in the next section. Once the gesture classification is done, the classified gesture is sent to the interface module that interconnects the gesture classifier and the game. Based on the gesture, the corresponding action is performed by the avatar in the game.

4 BUILDING THE GESTURE CLASSIFIER

To build a real-time gesture recognition model, we propose a two-stage model. In the first stage, the user's hand region is segmented from the static background and then, in the second stage, these segmented hand gestures will be used as input to the Convolutional Neural Network (CNN), to be classified.

4.1. Hand detection and tracking

One of the main challenges and essential processes for information extraction in many computer vision applications is the detection and tracking of the moving objects in video streams or image sequences [14]. Generally, there are three approaches for this moving object detection task; Background subtraction, temporal referencing and optical flow [7]. Among these, we chose a background subtraction method because it is one of the most popular ones in motion object detection and it takes less computational time and space. In this method, the foreground mask for every frame is generated by subtracting the background frame from the current frame. In other words, it has two major steps. First, constructing a good statistical representation for the background which is robust to noise; second, building a statistical model for foreground object which represents the changes that take place in the current frame [25].

To build a background frame which is less affected by noise, we applied some pre-processing steps. First, all the captured frames were resized to 128 by 128, then flipped to avoid mirroring problem. Afterward, to simplify the process and decrease the processing time, we cropped the images from the pre-defined Region Of Interest (ROI window) around the hand in the original frames. Figure 2 displays the output images for this step.

Then, the cropped images are converted to grayscale to avoid the long processing time for color image analyses. The converted images are then followed by Gaussian blurring filter[13] to remove noise. Afterward, to create a smooth background image, we calculate the average frame for the first k stationary background frames where k equals 30 and it is selected empirically(clean plates).

$$AvBg = 1/n \sum_{i=1}^{n} BG(i) \tag{1}$$

where n is total number of frames, *AvBg* represents the smooth-average background image(clean plate) and *BG(i)* is i-th background frame. As a result, noise is suppressed as

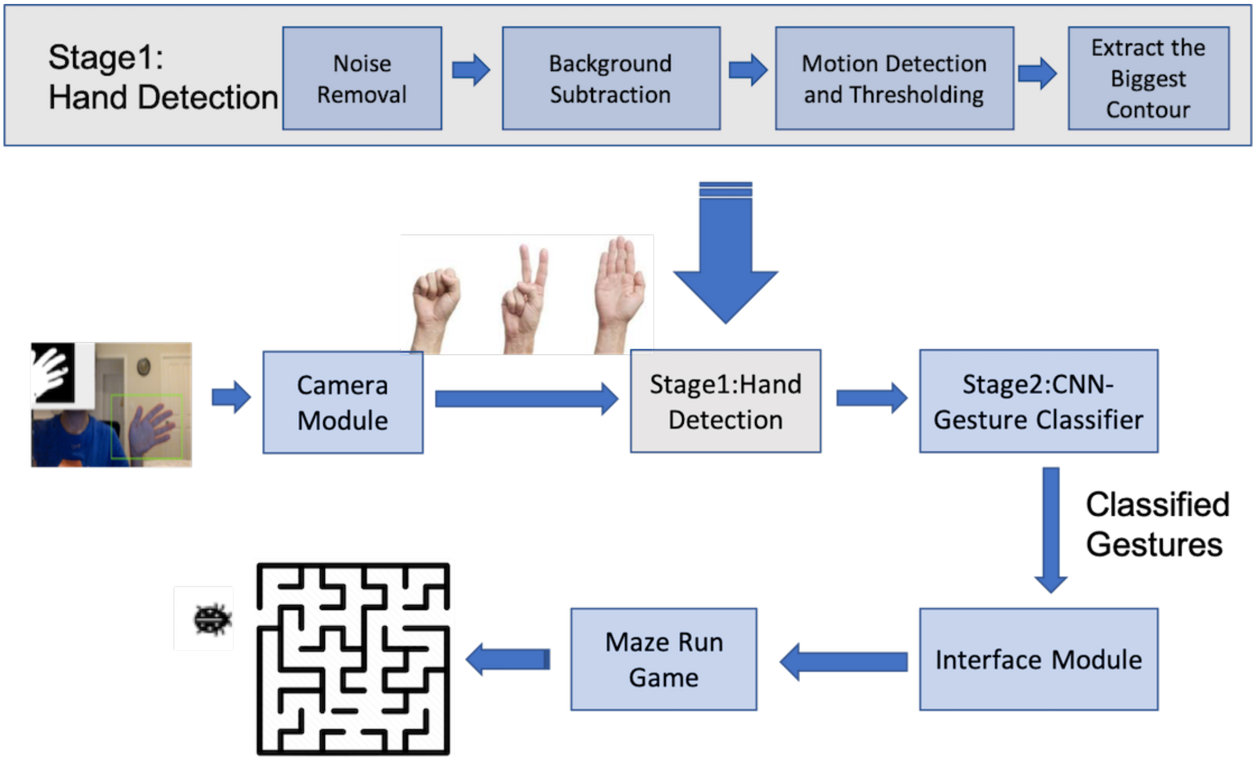

Figure 1: Overview of HandReha System

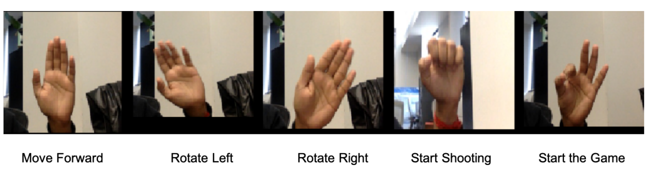

Figure 2: Input frames after applying resize, flip and crop operations

much as possible and the model will become more robust to changes in lightening. Second, to build a proper representation of the foreground object, we need to subtract the background and segment the image such that it has two components; hand as foreground segment and stationary background segment. To eliminate background there are two approaches: one with a known background called a clean plate and the other one is without known background[23]. In this study, we consider the first approach as we already created a clean plate in the first step.

Then, as the candidate starts performing desired gestures the absolute value of frames and clean plates are calculated and saved as different images.

$$diff(i) = |I(i) - AvBg| \tag{2}$$

Where i represents the current frame with and $AvBg$ as define in equation (1) is the clean plate image as a background frame. In other words, *diff* is the absolute difference between these two frames. Afterward, we applied a threshold value to get the foreground object (hand) for the difference image.

$$segmented = \begin{cases} foreground, & \text{if } diff(x,y)(i) > \tau \\ background, & \text{otherwise} \end{cases} \tag{3}$$

τ is the threshold value that was selected empirically as 25. It means that all the pixels in *diff*(i-th) frame which has a value larger than τ will be assigned a value equal to one and the remaining pixels belong to the background segment with a value equal to zero. Although the extracted hand in our case, will be segmented from a black background after applying the threshold, the final segmented hand might have some missing pixels. This can happen due to many reasons such as using an average of the stationary frames for the background. However, in our case, this does not affect the segmented images. The resulting image is a black and white segmented image with a hand segment as white and background as black. Finally, these segmented images are resized into a fixed size keeping the same aspect of ratio and ready to be used as input to our CNN classifier in the second stage.

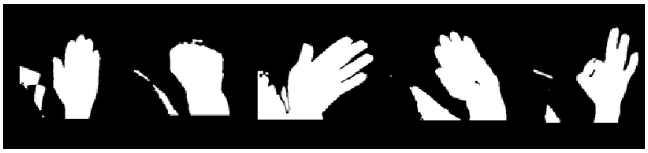

Figure 3: Output of the Hand Detection Stage

In figure 3 the final processed frames are depicted. Besides, we get the contours in these segmented images and consider the contour with maximum length as hand's contours. These contours are then displayed around the user's hands in each frame to evaluate the performance of the hand detection system.

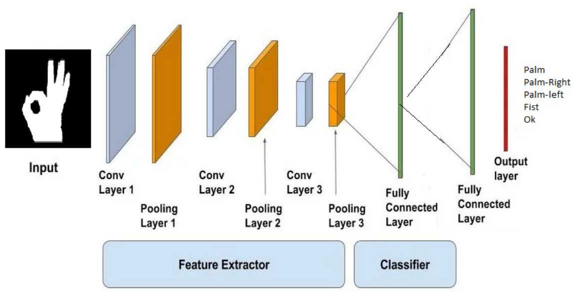

Figure 4: Overview of the CNN classifier

4.2. Hand gesture classifier model

In the second stage, to build the gesture classifier model we propose a 2D Convolution Neural Network (CNN) as a gesture recognition model. The architecture of the model is depicted in figure 4. The CNN feeds on binary images so that the color features do not affect the classifier. All the frames are first pre-processed, as explained in section 4.2. The pre-processing steps include resizing, converting to grayscale and applying Gaussian blurring, background subtraction, thresholding, and contour extraction. The model includes three 2D-convolutional layers for feature extraction, each followed by a max-pooling layer, two fully connected layers and a softmax layer for predicting gestures probabilities. The model classifies the gesture as the one with the highest probability. The input to this classifier is the processed frames from the hand detection stage which have the size of 128 by 128. Compared to other architecture such as RESNET50 [17]and VGG16 [30], the proposed CNN model only has 3 layers which makes the model much faster compared to above architectures.

4.3. HandReha dataset

After researching on which types of gesture have been used in hand therapy, we create our dataset of five gestures; "Fist", "Ok", "Open Palm", "PalmRight" and "PalmLeft" that are widely used in hand therapy procedure especially for wrist hand injuries[28]. A total of 7405 images are taken from three persons (two males, one female). We ask the participants to perform the gestures at a distance of 40 to 50 centimeters from the web-camera. Each class has around 1400 images as we tried to create a balanced data set. While collecting data all the mentioned pre-processing steps are applied to the captured images and finally, the processed images are saved into the separated folders. In our designed game, we assigned different tasks to each of these gesture classes; for instance, the character in the game should go straight if the model recognizes the "Palm" gesture or it should start shooting after recognizing "Fist" gesture.

4.4. Training the CNN gesture classifier

For the training process, we used a HAND-REHA dataset which includes 7405 images. We will explain the reason why we choose these special gestures in the Data collecting section. For each gesture, we used 0.1 percent of each class as the validation set and the remaining as the training set. The model uses input images of size 128 by 128. The loss function set as MSE and ADAM[21] as our optimizer. We trained the model for 5 epochs with processed images with various values of the learning rate, α, and realized α= 0.001 and image size as 128*128 provides the best accuracy for the classification task. Figure 5 and figure 6 show the loss and accuracy on training and validation set. The model achieves 100 and 98.8 percent accuracy for training and test set respectively.

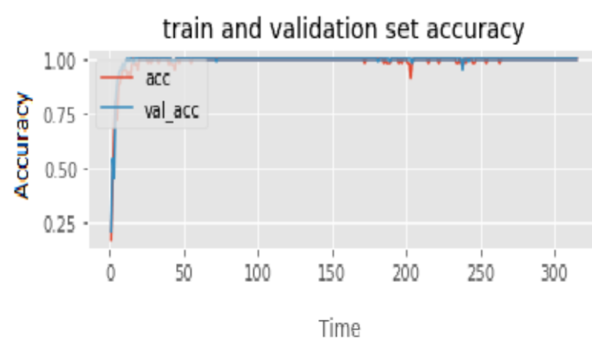

Figure 5: Train and validation accuracy

5 GAME DESIGN AND DEVELOPMENT

We developed a 3D maze run game where an avatar navigates inside the maze surrounded by hostile drones that shoot once the avatar is near to its location. The design for the main avatar, drones and the surrounding environment was done using Blender 2.8[11]. The game engine used for developing the game is Godot 3.1[9]. Godot is a free and open-source

Figure 6: Train and validation loss

Figure 7: Navigation of avatar in the Game

game engine under the MIT License. The avatar, drones and several other aspects of the game designed in Blender were then imported into the Godot game engine. The stage for the game along with the avatar and drone navigation was implemented in the Godot game engine.

We built two levels in the game. The first level is a normal maze without any drones and the avatar only navigates around the maze. This level is built for the purpose of familiarizing the users with hand gestures by using them to navigate the avatar around the maze. The next level is the main level where the avatar navigates around the maze while drones are present in it. The avatar has to navigate around the maze and shoot the drones when they encounter them.

Figure 7 and 8 shows the navigation and shooting performed by the avatar in the game. Figure 9 shows the game played with gestures.

A video of the gesture-based game can be found in the following link: https://www.youtube.com/watch?v=V7X4CCbExmc.

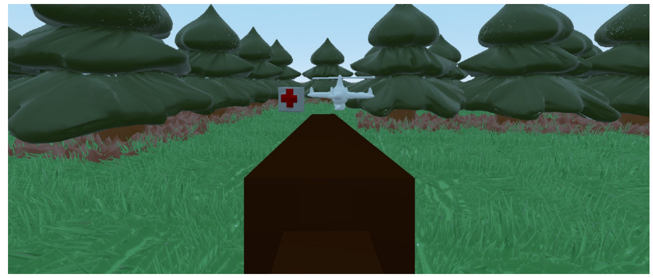

Figure 8: Shooting in the Game

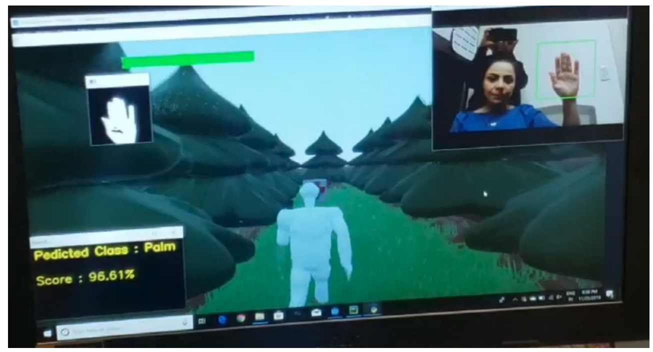

Figure 9: Game played with gestures

6 EVALUATION OF THE HAND-REHA SYSTEM

Twelve healthy participants from the Computer Science and Engineering department at the University of Texas, Arlington participated in the user study to evaluate the Hand-Reha system. . Of those twelve participants, six were male and six were female participants. Eight participants were aged between 25 and 30 years. Two participants were aged between 31 and 50 years. Two participants were aged between 18 and 24 years.

6.1 Hardware

For the experimentation, we used an Acer NITRO 5 Laptop with a Windows 10 64-bit Operating system. The laptop has an 8GB RAM along with an NVIDIA Geforce GTX 1050 Ti GPU. The laptop has a built-in camera which was used for gesture recognition. The Game runs on Godot Game Engine and the gesture recognition model runs on PyCharm IDE. Both the game and the model run on the same laptop.

6.2 Results of vision-based gesture classification

As mentioned in the previous section the accuracy for the validation set after 5 epochs reaches 98 percent. To test and evaluate the model we tried both testings offline and in real-time prediction. Although the model is not 100 percent accurate and has some flaws detecting gestures performed at far distance compared to training data, or in some cases the size of the hand matters(it defines how far the hand is located from the camera), our model still shows promising results while doing real-time predictions compared to other studies. Considering that we only use a shallow CNN model(only 3 layers) compared to other states of the art models and the fact that we only have around 1400 images per class will support our efficiency of the proposed model. Besides, the other issue was how perfect the participants perform the gestures, as some of them were focused on the game so they did not perform the gestures well, especially in the first couple of minutes starting to play the game. Yet considering all the mentioned reasons, we proved that our gesture classification model has acceptable performance in terms of accuracy and processing time.

6.3 User study methodology

Before the study begins, each participant filled a pre-study survey form. The form contained questions that asked whether the participant had any previous experiences of hand pain or difficulty in hand movement and their preferences in the kind of therapy if they had any such pain.

After filling the pre-study survey form, we explained the user study process of our system in detail. Afterward, the participants played the first level of the game, a plain maze where the avatar can only move around, for making the participants to get familiarized with the gestures. Then they played the main game that included the avatar and the drones along with shooting capability for both drones and the avatar.

After 8 minutes of playing the game, the participants filled the post-study survey form. In one of the questions, participants are asked to compare the difficulty and feeling of pain playing the proposed 3D Maze game through gestures with playing similar Maze game with controllers. The form was used to get feedback from the participants regarding the efficiency of the system along with their opinion about using gestures for gaming and suggestions for future work.

6.4 Pre-study survey responses

In the pre-study survey form, apart from the name, age, and gender of participants we also asked a few questions regarding whether they faced any problem with hand movements and their preferences with treatment if they had any such problems. According to the participants' responses, no one faced any problem with respect to hand movement in the past. With respect to the type of therapy, in case of any pain or difficulty in hand movement, we gave two options in the survey questionnaire: Individual therapy and Group therapy. Everyone chose Individual therapy as their preferred type of therapy.

Table 1 shows the response for the survey question regarding the preferred place for receiving treatment in case of any pain. For this question, 5 participants prefer treatment either

In case of pain where do you prefer to receive treatment?	
Options	Number of responses
Clinic	2
Home/Private Assistant	5
Hospital	4
Rehabilitation Center	0
Any Place is fine	3

Table 1: Table focusing on response regarding preferred place for receiving treatment

at home or through a private assistant, 4 of the participants prefer to receive treatment in hospitals, 2 of the participants prefer clinic treatment and 3 of the participants were fine to receive treatment in any place.

6.5 Post-study survey responses

After conducting the experimentation we had a post-study survey that had a set of questions regarding comfort, difficulty, effectiveness, and excitement while playing the game with hand gestures. The answers are based on the 5-point Likert scale rating with a range from 1 ("least") to 5 ("most"). As shown in Figure 10 and 11 most of the participants gave positive responses in each categories.

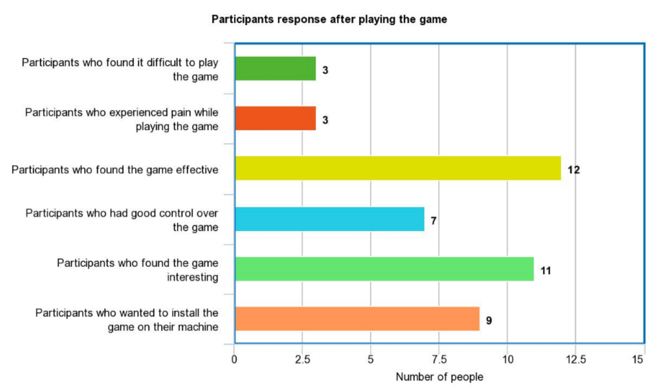

Figure 10: Graph representing participants feedback on HandReha

6.6 Discussion of the user study results

Overall, there was a good response from the participants where most of them reported that the hand gesture-based game was effective, exciting and comfortable. Most of the participants reported that they felt less pain and difficulty in controlling the game with gestures compared to traditional method such as using a controller. However, some participants reported that they had moderate pain and difficulty while playing the game. Three main reasons may have caused

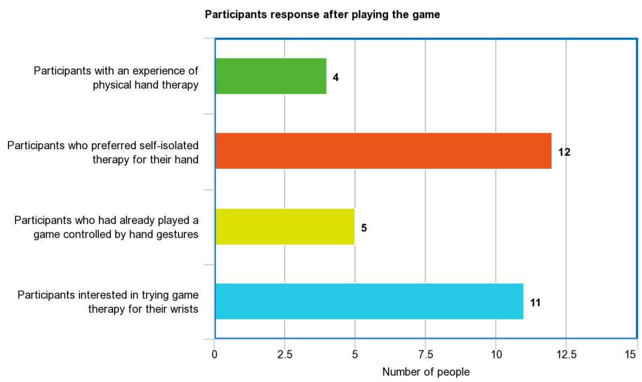

Figure 11: Graph indicating responses from the participants before using HandReha

such an issue for some participants. First reason is that in the classifier, for some participants, there was some difficulty in differentiating between "Open Palm" and "Ok" gestures, at certain situations, since both the gestures are shown at the same angle and after final processing "Open palm" and "Ok" gestures seem to be a bit similar. The second reason is that gesture recognition worked with better accuracy for participants with smaller hands when compared to participants with bigger hands which might be because of a shortage of dataset containing gestures from bigger hands. These two issues cause some difficulties for participants while navigating the avatar in the game. The third reason is that there was a time lag between gesture recognition and character movement in the game which is probably because of lower RAM capacity (8 GB) on the laptop. The third reason has been the cause of moderate pain for some participants as the time lag (a few seconds) between gesture recognition and character movement must have made the participants hold a particular gesture for an additional period at certain points in the game.

7 CONCLUSION AND FUTURE WORK

We have designed a game-based wrist rehabilitation system, which enables the user to control an avatar in a three-dimensional maze-run game using hand gestures. This is a unique and novel approach because the gestures are selected from a set of human gestures suitable for wrist rehabilitation and implemented to control a game built in a 3D environment as compared to previous works where most of the games designed for rehabilitation purposes are built in a 2D environment. Moreover, the game is built with an avatar performing more than one action and the gestures are assigned for every action performed by the avatar in the game. The gestures are implemented in such a way that they directly interact with the system. This is a different approach than the previous works where the game is built with an avatar performing only a single action and the gestures are implemented to simulate mouse or keyboard events through which the interaction with the system takes place. A user study was conducted to evaluate the developed system where the participants played the game with gestures. For now, the participants surveyed were from a healthy group of people with no problem with their hands and wrists. In future, we plan to include actual patients to test the efficiency of the system and to assess its feasibility in people with real injuries. This would be an extension of our current research. The results from the user study showed a good and favorable outcome where almost all the participants provided moderate to high ratings in terms of effectiveness and interest in playing the game with gestures.

Regardless, future research could continue to explore in improving the overall efficiency, accuracy, functionality, and usability. We plan to extend the HandReha system to be compatible with other everyday devices such as smartphones and tablets using Mobilenetes(a family of mobile-first computer vision models for TensorFlow) [18] or Inception-V3 [31]. Furthermore, the complexity of the game could be improved and better incentives for the players can be incorporated in the game rather than shooting a drone. Moreover, we plan to extend the HandReha dataset by including additional gestures and capturing data from a larger number of people who have problems in moving their hands and wrists to improve accuracy during gesture recognition and investigate the possible benefits of home rehabilitation.

REFERENCES

[1] 2008. The Internet Stroke Center. http://www.strokecenter.org/patients/about-stroke/stroke-statistics/

[2] Ashwin Ramesh Babu, Mohammad Zakizadeh, James Robert Brady, Diane Calderon, and Fillia Makedon. 2019. An Intelligent Action Recognition System to assess Cognitive Behavior for Executive Function Disorder. In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*. IEEE, 164–169.

[3] Khawla Benabderrahim and Med Bouhlel. 2014. Detecting and tracking the hand to create an augmented reality system. *IOSR Journal of Computer Engineering* 16 (01 2014), 31–35. https://doi.org/10.9790/0661-16173135

[4] Yassine Bouteraa, Ismail Ben Abdallah, and Ahmed M Elmogy. 2019. Training of hand rehabilitation using low cost exoskeleton and vision-based game interface. *Journal of Intelligent & Robotic Systems* 96, 1 (2019), 31–47.

[5] Heng-Yi Chen, Tse-Yu Lin, Li-Yang Huang, An-Chun Chen, Yu-Chen Zheng, Hsing-Mang Wang, Shi-Yao Wei, and Yin-Yu Chou. 2018. HP2: Using Machine Learning Model to Play Serious Game with IMU Smart Suit. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia (MUM 2018)*. Association for Computing Machinery, New York, NY, USA, 397–402. https://doi.org/10.1145/3282894.3289731

[6] Chiang Wei Tan, Siew Wen Chin, and Wai Xiang Lim. 2013. Game-based human computer interaction using gesture recognition for rehabilitation. In *2013 IEEE International Conference on Control System,*

Computing and Engineering. 344–349. https://doi.org/10.1109/ICCSCE.2013.6719987

[7] Rita Cucchiara, Costantino Grana, Massimo Piccardi, and Andrea Prati. 2003. Detecting moving objects, ghosts, and shadows in video streams. *IEEE transactions on pattern analysis and machine intelligence* 25, 10 (2003), 1337–1342.

[8] Alex Dillhoff, Konstantinos Tsiakas, Ashwin Ramesh Babu, Mohammad Zakizadehghariehali, Benjamin Buchanan, Morris Bell, Vassilis Athitsos, and Fillia Makedon. 2019. An automated assessment system for embodied cognition in children: from motion data to executive functioning. In *Proceedings of the 6th international Workshop on Sensor-based Activity Recognition and Interaction*. 1–6.

[9] Godot Engine. 2019. Free and open source 2D and 3D game engine. https://godotengine.org/

[10] Sergio Escalera, Vassilis Athitsos, and Isabelle Guyon. 2017. Challenges in multi-modal gesture recognition. In *Gesture Recognition*. Springer, 1–60.

[11] Blender Foundation. 2019. Home of the Blender project - Free and Open 3D Creation Software. https://www.blender.org/

[12] Alana Da Gama, Thiago Menezes Chaves, Lucas Silva Figueiredo, Adriana Baltar, Meng Ma, Nassir Navab, Veronica Teichrieb, and Pascal Fallavollita. 2016. MirrARbilitation: A clinically-related gesture recognition interactive tool for an AR rehabilitation system. *Computer methods and programs in biomedicine* 135 (2016), 105–14.

[13] E. S. Gedraite and M. Hadad. 2011. Investigation on the effect of a Gaussian Blur in image filtering and segmentation. In *Proceedings ELMAR-2011*. 393–396.

[14] Sakher Ghanem, Ashiq Imran, and Vassilis Athitsos. 2019. Analysis of hand segmentation on challenging hand over face scenario. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. 236–242.

[15] Shawn Gieser, Angie Boisselle, and Fillia Makedon. 2015. Real-Time Static Gesture Recognition for Upper Extremity Rehabilitation Using the Leap Motion. 144–154. https://doi.org/10.1007/978-3-319-21070-4_15

[16] R. Han, Z. Feng, T. Xu, C. Ai, W. Xie, K. Zhang, and J. Li. 2017. Multi-sensors Based 3D Gesture Recognition and Interaction in Virtual Block Game. In *2017 International Conference on Virtual Reality and Visualization (ICVRV)*. 391–392. https://doi.org/10.1109/ICVRV.2017.00091

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. *CoRR* abs/1512.03385 (2015). arXiv:1512.03385 http://arxiv.org/abs/1512.03385

[18] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861* (2017).

[19] Ya-Xuan Hung, Pei-Chen Huang, Kuan-Ta Chen, and Woei-Chyn Chu. 2016. What do stroke patients look for in game-based rehabilitation: a survey study. *Medicine* 95, 11 (2016).

[20] SN Karishma and V Lathasree. 2014. Fusion of skin color detection and background subtraction for hand gesture segmentation. *International Journal of Engineering Research and Technology* 3, 2 (2014), 1835–1839.

[21] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. http://arxiv.org/abs/1412.6980 cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.

[22] Alex Krizhevsky. 2012. *AlexNet: Convolutional Neural Networks*. https://en.wikipedia.org/wiki/AlexNet

[23] Hsiang-Yueh Lai, Hao-Yuan Ke, and Yu-Chun Hsu. 2018. Real-time Hand Gesture Recognition System and Application. *Sensors and Materials* 30, 4 (2018), 869–884.

[24] Sih-Pin Lai, Cheng-An Hsieh, Teepob Harutaipree, Shih-Chin Lin, Yi-Hao Peng, Lung-Pan Cheng, and Mike Y. Chen. 2019. FitBird: Improving Free-Weight Training Experience Using Wearable Sensors for Game Control. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts (CHI PLAY '19 Extended Abstracts)*. Association for Computing Machinery, New York, NY, USA, 475–481. https://doi.org/10.1145/3341215.3356258

[25] Mahmoud A Mofaddel and Walaa M Abd-Elhafiez. 2011. Fast and accurate approaches for image and moving object segmentation. In *The 2011 International Conference on Computer Engineering & Systems*. IEEE, 252–259.

[26] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in pytorch. (2017).

[27] Rachel Proffitt, Marisa Sevick, Chien-Yen Chang, and Belinda Lange. 2015. User-centered design of a controller-free game for hand rehabilitation. *Games for health journal* 4, 4 (2015), 259–264.

[28] Harvard Health Publishing. 2019. 5 exercises to improve hand mobility. https://www.health.harvard.edu/pain/5-exercises-to-improve-hand-mobility-and-reduce-pain

[29] Edgar Rodríguez Ramírez, Regan Petrie, Kah Chan, and Nada Signal. 2018. A tangible interface and augmented reality game for facilitating sit-to-stand exercises for stroke rehabilitation. In *Proceedings of the 8th International Conference on the Internet of Things*. 1–4.

[30] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[31] Xiaoling Xia, Cui Xu, and Bing Nan. 2017. Inception-v3 for flower classification. In *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*. IEEE, 783–787.

[32] Pei Xu. 2017. A Real-time Hand Gesture Recognition and Human-Computer Interaction System. *CoRR* abs/1704.07296 (2017). arXiv:1704.07296 http://arxiv.org/abs/1704.07296