

# Robot Avatar: A Virtual Tourism Robot for People with Disabilities

Chong Wing Cheung, Tai Ip Tsang, and Kin Hong Wong

**Abstract**—Developing a robot avatar for disabled individuals can help them to reduce the difficulties in their daily lives. In this project, we propose to build a robot that takes pictures or videos from a faraway place and feeds the data to the user in real time. The user may be wearing a Head Mount Device HMD to view the video sent from the robot to enjoy the virtual presence experience. We also propose to design a hand gesture recognition system that the user can use very simple gestures to control the motion of the robot to select his/her favorite views. Currently the robot and finger gesture recognition system have been built and tested successfully. Data analysis of the video latency from the robot to the user has also been carried out. The result will be used to guide us to improve the system for further development. In the next step, we will continue to develop the system including adding the HMD to our system and solve the networking problem to reduce latency. It is hoped that disabled people can enjoy more freedom with the help of the ever advancing computer technologies. Moreover, the same idea has also great potential in.

**Index Terms**—Robot control, virtual reality systems, robot avatar, finger gesture.

## I. INTRODUCTION

The target of this project is to develop a system that helps people with disability to extend their social interactions with the society. When one is sick in bed or unable to move out of a restricted environment this system can help the user to observe and experience the world outside. We propose that it can be achieved by using a head mount display HMD so he/she can see the outside world through the displaying device. And the video stream can be coming from a remote robot that is equipped with a stereo camera pair. Moreover we can also design a vision based system that the user can use a set of very minimal finger gestures to control the robot movement, it helps people with disability to control the robot to move round.

The idea is illustrated in Fig. 1. The recent computer revolution has brought us many useful gadgets such as the mobile phones, wearable intelligent watches [1] and head mount displays. They give designers the tools to build better virtual reality systems. In this project we would like to explore the possibilities of enabling Virtual Tourism for people, especially those with disabilities. The idea of using robots to

Manuscript received August 16, 2016; revised December 9, 2016. This work is supported by a direct grant (Project Code: 4055045) from the Faculty of Engineering of The Chinese University of Hong Kong.

Chong Wing Cheung, Tai Ip Tsang, and Kin Hong Wong are with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, HSH Engineering Building, CUHK, Shatin, Hong Kong, China (e-mail: cwcheun4@cse.cuhk.edu.hk, titsang4@cse.cuhk.edu.hk, khwong@cse.cuhk.edu.hk).

enhance human communication has been studied in the past, e.g. [2], [3]. In [4], it explores the design and social acceptance of sociable autonomous robots that interact with people in a human-like way. In [5], the project introduces an emotion interactive robot to make people feel that the robot is not a mere machine, but reliable living assistant in the home. Stiehl et al, discusses a huggable robot human for communication and education [6]. The idea of using a robot to represent the body of a human has been mentioned in some research work, e.g. [7]. In [8] a wearable avatar mounted on a person's shoulder allows the users to share a vision remotely. More recently, the use of robot as a substitute to human's body for various tasks such as space exploration, inspection and surveillance activities *et al.* have been investigated [9]. The research opportunity and challenges of robot-mediated communication, where at least one party is communicating through telepresence has been discussed in [10]. More recent experiments on robot avatar designs and concepts are reported in [11] and [12].

Disabilities are always a group of people who need the society cares and assistances. In accordance with the survey of Hong Kong Census and Statistics Department in 2013, there are around 4.5% of Hong Kong people suffering certain degrees of disabilities that restrict their body movement [13]. As reported in the World Health Organization (WHO) survey, it is estimated that up to 15% or 1 billion of people in the world are disabled, of which 5.1% are having difficulties in mobility [14]. In other words, it implies 1 out of 20 people are physically disabled of a certain degree. In addition to the significant number of disabled, the rate of disabilities is also shown to be increasing in Hong Kong, partly because of the number of aged people are increasing. From 2007 the total number of mobility disabled is 187,800, which is an increase of 2/3 of the total to a number of 320,500 in 5 years.

As a result, we would like to prototype a system which can encourage disabled people to keep in touch with what is happening outside their living environment. Virtual Reality (VR) is a popular research area especially suitable to be applied to the gaming industry. Recently, many companies are creating new products to be used in this field. For example, the company Oculus VR is developing a HMD product for the gaming industry. Moreover, there are already researchers using this for various applications. For example, an experiment which streams what the user sees to the Oculus Rift device is reported in [15]. And a third person perspective experiment is described in [16]. However, there is no report of work utilizing the Oculus Rift to be an accessibility assistive device. The objective of our system is to allow the disabled to navigate the vision system to any places without leaving the attached medical equipment or sick bed. Therefore, a robot

agent should be developed to be the avatar of the user, this enables the user to see through the robot's cameras and perform different actions. To achieve this, the robot agent has to fulfill a numbers of requirements. First, the robot agent should be able to move around and be controlled remotely. Second, the robot agent should provide some degrees of flexibility adapting to the users' physical profile and preferences. For example, our control input should be able to track the user's hand and finds out the finger gesture for robot control. In this project we explore these directions and prototype a system that fulfills the above requirements.

This paper is divided into several parts: in Section II we will explain the background of our project. In Section III the theory and design methodology will be discussed, the implementation and testing result are explained in Section IV. Finally, discussion and conclusion are found in Section V and VI, respectively.

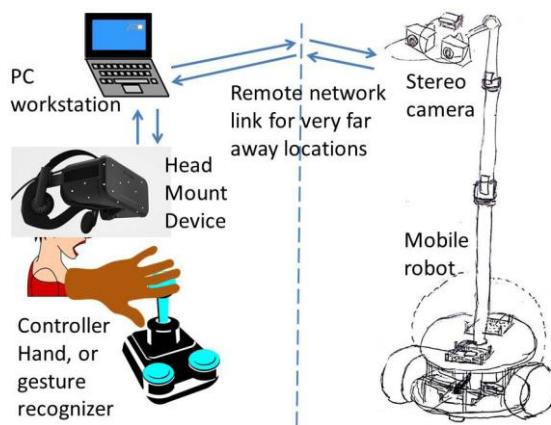


Fig. 1. The overall design of the robot avatar system.

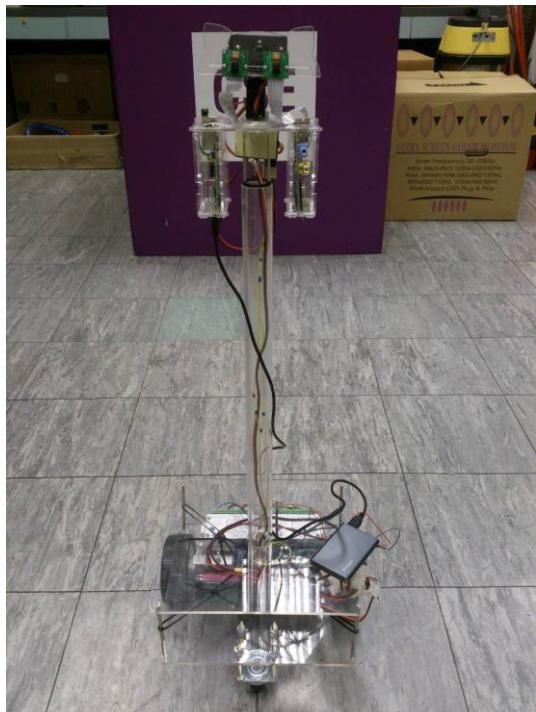


Fig. 2. The robot with a stereo camera pair.

## II. BACKGROUND

In this project, we will be using three subsystems. First, it is

the head mount display device to be worn by the user, and the second one is a computer vision system that interfaces the user and the machine. Finally we need a robot that carries the camera around. Background of these systems and technology are discussed in this section.

### A. The Head Mount Device HMD

In order to make virtual tour possible, a user is required to wear a head mounted display (HMD). It is a pair of stereo displays mounted in front of the user's eyes to provide a 3-D viewing experience. The HMD is also equipped with sensors to track the movements of the user's head to make corresponding changes in the display so that the user can obtain an immersed reality experience. There are a few requirements for choosing the HMD for this project: 1) Low communication latency: Since the image is retrieved from the robot agent through the network and the robot agent is moving in response to the user control, the latency of video sending from the camera of the robot to the display for the user should be as low as possible. 2) Low cost: So that it can be affordable by most people. There are a number of companies already developing their own virtual reality head mounted display devices for the consumer market such as Oculus VR and Google – Cardboard. Google Cardboard [17] is the cheapest solution in the market (from about USD10 to 25, phone excluded) and is totally portable as it is driven by an Android phone. However, its performance is poor especially its latency is high that limits its application and popularity. The system by HoloLens [18] has all the advantages: holographic computer (HPU) embedded, high performance, wireless and great technical support. Unfortunately, since its price tag is as high as USD3,000 each, so it will not be considered in the project. The Oculus Rift [19] is a suitable choice for our project, since the price is set a reasonable level at around USD350. Although it is wired, however, our target user with disabilities is expected to sit at a fixed place; hence it is not a major hurdle. On the other hand, the wired communication allows lower data communication latency.

### B. The OpenCV Library

The OpenCV (Open Source Computer Vision) library is a cross-platform, free and open source library aimed at providing real-time computer vision programming functions for applications [20]. It is written in C++ but it supports other common programming languages such as C, Java, Python and Matlab. The library provides convenient functions that enable developers to build applications in computer vision. For example, it provides mathematical functions such as contour and convex hull calculation, which are essential in building hand gesture detection procedures. It also provides functions for image processing and graphical interface building, which can help to build user interfaces or augment reality applications.

## III. THEORY AND DESIGN

The overall system is shown in Fig. 1. The user is wearing the Head Mount Device HMD, which is linked up a local PC via WiFi. The PC is connected to the Internet that sends commands to the remote robot which may be situated at a

faraway location. The user can use a keyboard or a joystick for the control of the robot. To ease the control of the robot by disabled people, we also designed a gesture control vision system that the user can use simple finger gestures to issue control commands. The robot is driven by two large motors and capturing stereo images by its stereo camera pair. In future we will develop a system to stream stereo videos back to the head mount device through the Internet. So the user can have the experience of virtual presence as if he is in the remote environment. Currently only one video stream is provided. The main workflow of hand gesture detection using a color camera is as follows:

- 1) Get Sample Hand Color: First, the color of the hand has to be obtained in order to perform background subtraction. To obtain this information the user is required to put the hand on a blue background to complete this step.
- 2) Capture RGB Image: After hand color calibration is complete, the system should start capturing images for recognition. An OpenCV [20] algorithm is responsible for capturing the image through the webcam and stores the data in RGB format.
- 3) Background Subtraction: In order to recognize the hand accurately, the background should be eliminated. It can be done by setting a threshold for the color image, and then a binary image with the hand will be produced. Furthermore, the RGB image can be converted to HLS color spaces (hue, lightness and saturation) in advance to enhance recognition results. The related OpenCV functions are cv::cvtColor() and cv::inRange().
- 4) Find Biggest Contour: By using the binary image, we can locate the contour of the hand. It is assumed that the bigger the contour, the better the performance. By comparing the size of each contour, the biggest contour is found and the index of the contour will be marked. The OpenCV cv::findContour() function is used in this step.
- 5) Find Convex Hull: Using the biggest contour, the convex hull can be found using the OpenCV functions cv::convexHull() and cv::approxPolyDP().
- 6) Find Convexity Defects: Based on the convex hull points and the contour, the convexity defects can be calculated along with the farthest point and depth of the defect. Defects between two fingers can be determined by setting a threshold of depth and a threshold of the angle between two fingertips. The related OpenCV function is: cv::convexityDefects().
- 7) Finger Gesture Recognition and Output: The number of fingertips and their positions are obtained from the number of convexity defects, and the positions of the start and end points of the defects. The information is being displayed on the output image for inspection. In this project, the functions for hand gesture detection are modified from the open source program by Simen Andresen [21]. The program performs a similar algorithm mentioned above using OpenCV and achieved good result with the use of some image filtering techniques such as blurring. We make use of it to obtain the number of fingers detected, and use the positions of the detected fingers to judge the hand gesture and make corresponding controls.
- 8) Hand Gesture for Robot Control: The program is

designed to recognize some predefined hand gestures for controlling the robot. The user setup of the system is illustrated in Fig. 3. And Fig. 4 shows how to use two fingers to control the movements of the robot with respect to the angle made by the two fingertips. Detection of the hand is illustrated in Fig. 5. The algorithm for finger gesture recognition is shown in Algorithm1.

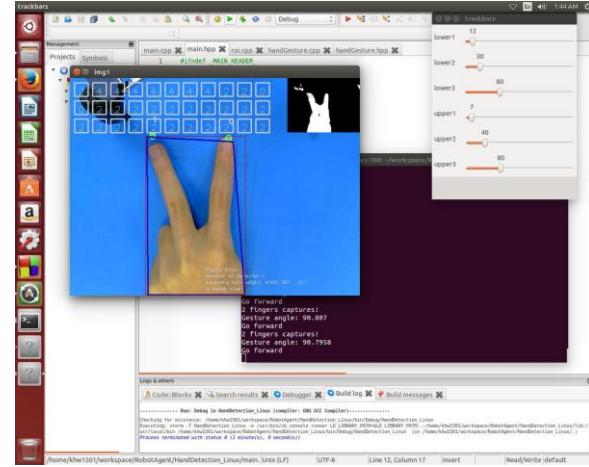


Fig. 3. User setup of the finger gesture detection system.

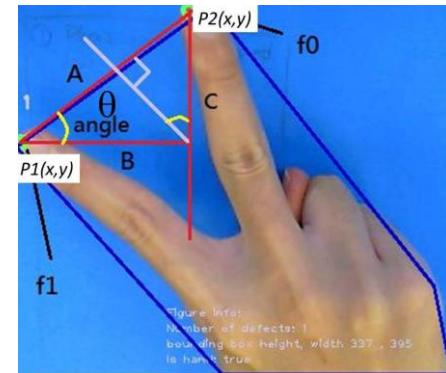


Fig. 4. Detection of the finger gesture using the angle between two fingers.



Fig. 5. Hand detection and gesture detection.

#### IV. IMPLEMENTATION, EXPERIMENTS AND RESULTS

##### A. Issues in Implementation

There are two major software modules in our system. (1)

An Arduino program which is responsible for driving the robot. (2) A Windows/ Linux C++ program for hand gesture detection and forms communication between the robot and the computer through the Bluetooth serial communication link. In fact, the Arduino can be connected through any serial communication tools such as Moba Xterm and Android's Bluetooth Serial. The overall software system hierarchy is shown in Fig. 6. The hardware architecture of the system is also shown in Fig. 7.

### B. Motor and Control

The two motors are driving by the micro-controller onboard through two power electronic circuits. The system is shown in Fig. 8.

**Algorithm 1** Finger Control Interpretation Interface

```

if NumberOfFinger == 2 then
    Read angle( $\theta$ ) as defined in Figure 4
     $\theta \leftarrow \arctan\left(\frac{P_1(y)-P_2(y)}{P_1(x)-P_2(x)}\right)$ ;
    if  $\theta > -20.0^\circ$  and  $\theta < 20.0^\circ$  then
        Robot Moves Forward
    end if
    if  $\theta < -20.0^\circ$  then
        Robot Moves Left
    end if
    if  $\theta > 20.0^\circ$  then
        Robot Moves Right
    end if
end if

```

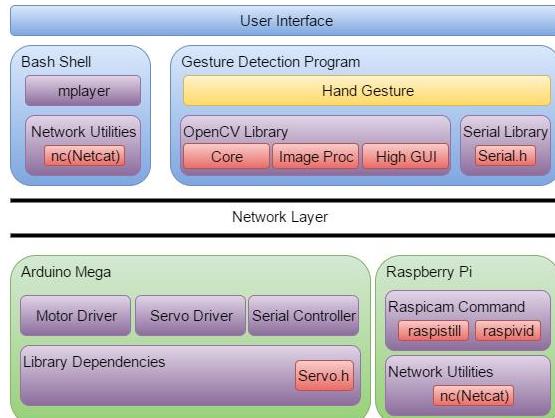


Fig. 6. Software modules of the system.

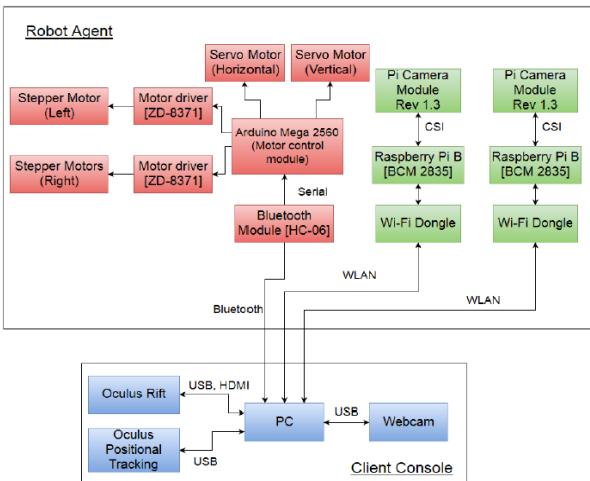


Fig. 7. Hardware architecture of the system.

### C. Tests of the System

#### 1) Video streaming latency test

Since the performance of system depends on the

responsiveness of the video playback, latency in the video stream from the robot to the client PC becomes very important. So, a test to find out the latency is performed. The experiment was set up by the following procedures. A timer was set in front of the robot for capturing the time. The screen which was directly connected to the robot-camera module and the screen on the PC will be captured and used to calculate the latency. Fig. 9 and Fig. 10 show the Latency Test result. The result shows that time delay is acceptable but it is more serious when streaming occurs.

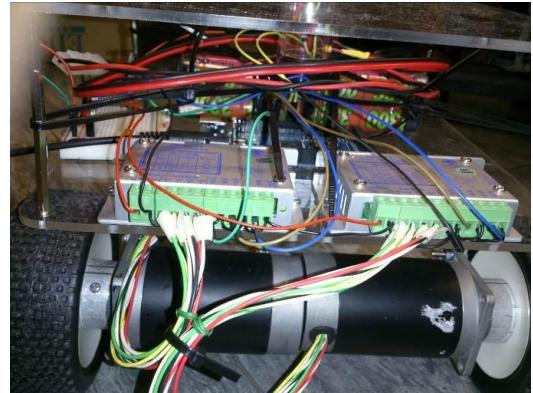


Fig. 8. The motors and interfacing circuits.

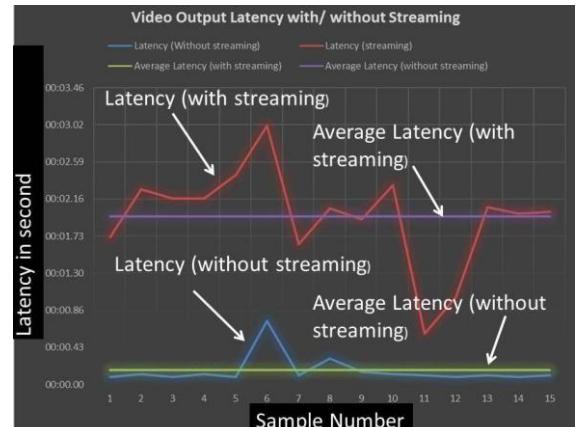


Fig. 9. Video streaming latency test result.

	A Data	B Local Time	C Client Time (without streaming)	D Client Time (With streaming)	E Latency (Without streaming)	F Latency (streaming)
1						
2	1	08:50.5	08:50.5	08:48.8	00:00.1	00:01.7
3	2	09:01.2	09:01.1	08:58.9	00:00.1	00:02.3
4	3	09:10.5	09:10.4	09:08.3	00:00.1	00:02.2
5	4	09:20.4	09:20.3	09:18.3	00:00.1	00:02.2
6	5	09:30.4	09:30.3	09:28.0	00:00.1	00:02.4
7	6	09:40.4	09:39.7	09:37.4	00:00.7	00:03.0
8	7	09:49.6	09:49.5	09:48.0	00:00.1	00:01.6
9	8	10:00.0	09:59.7	09:57.9	00:00.3	00:02.0
10	9	10:10.5	10:10.4	10:08.6	00:00.2	00:01.9
11	10	10:20.2	10:20.1	10:17.9	00:00.1	00:02.3
12	11	10:30.2	10:30.1	10:29.6	00:00.1	00:00.6
13	12	10:40.2	10:40.1	10:39.2	00:00.1	00:01.0
14	13	10:50.9	10:50.8	10:48.9	00:00.1	00:02.1
15	14	11:00.5	11:00.4	10:58.5	00:00.1	00:02.0
16	15	11:11.9	11:11.8	11:09.9	00:00.1	00:02.0
17				Average Latency	0.16s	1.96s

Fig. 10. Data for video streaming latency test result.

## 2) Overall test

An integrated test is performed in order to check the compatibility of the components. The test includes various possible motions of the robot under different finger gesture control commands. A video of the robot in operation during the test is recorded and uploaded to YouTube for viewing. The video link is <https://youtu.be/smDG6nV49yc>. The test shows the operating of the system is satisfactory and the finger gestures can control the robot effectively.

## V. DISCUSSION

One of the major problems in our application is latency in video Streaming. Generally, there are several sources that may cause communication latency in video streaming:

- Video processing and forwarding time in the microcontroller (Raspberry Pi).
- Latency in network transmission.
- Network package processing in router, Raspberry Pi and PC.
- Network stability.
- Efficiencies of network protocols used.
- Processing time in PC and media player.
- Buffer size.

While little can be done on the network quality, there should be some methods to suppress latency. For example, the following strategies can be explored:

- Lowering the bitrate: It may result in lowering video quality, so less data will be transmitted to ensure the transmitting packages will be received on time.
- Lowing buffer size: Frames that fall too behind will be skipped, better synchronization can be achieved but greater stutter may occur. More possible solutions are to be explored in future.

Displaying the image captured by the robot and controlling the robot by the HMD motion are the goals to be achieved in the next step. The Oculus SDK will be studied in order to complete these tasks. As the robot agent is aimed to be working outdoor, networking between client workstation and the robot through LAN or even WAN have to be studied. In order to simplify the circuit design and shorten the circuit connections within the robot agent, centralizing the robot control modules would be a solution. An additional Raspberry Pi 2 computing module will be added to enable better communication with the client workstation. Regarding the latency issue, we will research on minimizing the latency caused by network transmission and computation. Moreover, additional useful features on the robot can be implemented. For example, a Collision Avoidance system by using Ultrasonic sensors to prevent irrational movements and a two way voice communication system are also interesting options to enhance the robot agent, making it more robust and interactive.

## VI. CONCLUSION

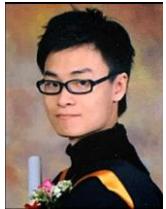
Our target is to build a robot avatar system to help disabled people to perform virtual tour to places where they cannot reach because of their physical disabilities. In this paper we discuss the work that we have successfully built. It includes a

remote control robot and a visual recognition system that the user can control the robot by using simple hand gestures. Moreover, the camera on board of the remote robot can stream back video data to the user almost in real time. We will continue to complete the whole system by adding 3-D stereo streaming to enable full virtual presence functionalities. Performance issues such as streaming latency, communication between devices and remote control through LAN and WAN will be handled in future.

## REFERENCES

- [1] L. Whitney. Apple watch clocks the competition with 7 million in sales. [Online]. Available: <http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm>
- [2] D. Sekiguchi, M. Inami, and S. Tachi, "Robotphone: Rui for interpersonal communication. In CHI'01 extended abstracts on human factors in computing systems," *ACM*, pp. 277–278., 2001.
- [3] T. Nishiyama, H. Hoshino, K. Sawada, Y. Tokunaga, H. Shinomiya, M. Yoneda, I. Takeuchi, Y. Ichige, S. Hattori, and A. Takanishi. 'Development of user interface for humanoid service robot system,' in *Proc. IEEE International Conference on Robotics and Automation*, 2003, vol. 3, pp. 2979–2984.
- [4] C. Breazeal, "Social interactions in hri: the robot view," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 34, no. 2, pp. 181–186, 2004.
- [5] D.-S. Kwon, Y. K. Kwak *et al.*, "Emotion interaction system for a service robot," in *Proc. The 16th IEEE International Symposium on Robot and Human interactive Communication*, 2007, pp. 351–356.
- [6] W. D. Stiehl, J. K. Lee, C. Breazeal, M. Nalin, A. Morandi, and A. Sanna. "The huggable: a platform for research in robotic companions for pediatric care," in *Proc. the 8th International Conference on Interaction Design and Children*, 2009, pp. 317–320.
- [7] S. Maeyama, S. Yuta, and A. Harada, "Experiments on a remote appreciation robot in an art museum," in *Proc. 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2000, vol. 2, pp. 1008–1013.
- [8] T. Kashiwabara, H. Osawa, K. Shinozawa, and M. Imai. "Teroos: a wearable avatar to enhance joint activities," in *Proc. the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 2001–2004.
- [9] A. R. Azli Ali, M. H. M. Saad, R. A. Shahad, and A. Hussain. "Design and development of an indoor avatar robot," in *Proc. 2015 International Electronics Symposium*, 2015, pp. 89–93.
- [10] S. C. Herring, "New frontiers in interactive multimodal communication," *The Routledge Handbook of Language and Digital Communication*. London: Routledge, pp. 398–402, 2015.
- [11] H. Mamode and H. Ziyaad, "Robot mediated communication: Enhancing tele-presence using an avatar," PhD thesis, University of the West of England, 2015.
- [12] S. de la Rosa, M. Lubkull, S. Stephan, A. Saulton, T. Meilinger, H. Büthhoff, and R. Cañal-Bruland, "Motor planning and control: Humans interact faster with a human than a robot avatar," *Journal of Vision*, vol. 15, no. 12, pp. 52–52, 2015.
- [13] Census and Statistics Department, Hong Kong Special Administrative Region. (2014). Social Data Collected via the General Household Survey: Special Topics Report - Report no. 62. [Online]. Available: <http://www.unicef.org/protection/> World report on disability eng.pdf
- [14] World Health Organization, The World Bank. (2015). World Report on Disability. [Online]. Available: <http://www.statistics.gov.hk/pub/B11301622014XXXXB0100>
- [15] Umeaenergi. (2014). Living with lag — an oculus rift experiment. [Online]. Available: <https://www.youtube.com/watch?v=fNp37zFn9Q>
- [16] P. Mepi. (2014). Real world third person perspective vr ar experiment. [Online]. Available: <https://www.youtube.com/watch?v=RgBeRP4dUGo>
- [17] Google, google cardboard. (2015). [Online]. Available: <https://www.google.com/get/cardboard/>
- [18] Microsoft, Microsoft hololens. (2015). [Online]. Available: <https://www.microsoft.com/microsoft-hololens/en-us>
- [19] Oculus. (2015). Oculus rift development kit 2 (dk2). [Online]. Available: <https://www.oculus.com/en-us/dk2/>
- [20] Itseez. (2015). OpenCV. [Online]. Available: <http://opencv.org/>

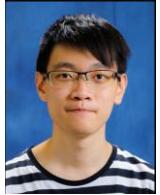
- [20] Simen Andresen. (2013). Hand tracking and recognition with opencv. [Online]. Available: <http://simena86.github.io/blog/2013/08/12/hand-tracking-and-recognition-with-opencv/>



**Chong Wing Cheung** is now a Ph.D. student at the Department of Computer Science and Engineering of the Chinese University of Hong Kong (CUHK). He obtained a first class BEng. Degree in Computer Engineering from CUHK in 2016. His research interests include VLSI Engineering, optimization techniques, computer vision and virtual reality.



**Kin Hong Wong** obtained his Ph.D. from the University of Cambridge, UK and is now an Associate Professor of the Computer Science and Engineering Department, the Chinese University of Hong Kong. His research interests include computer vision, signal processing and virtual reality.



**Tai Ip Tsang** was born in Hong Kong in 1993. He is currently an undergraduate student studying Computer Engineering at the Computer Science and Engineering Dept. of The Chinese University of Hong Kong (CUHK). He was an exchange student at the Technical University of Denmark (DTU) in 2015, studying wireless electronic design. His research interests include embedded system, computer vision and virtual reality.