2017 2nd International Conference on Advanced Robotics and Mechatronics (ICARM)

Real-time Human-in-the-loop Remote Control for a Life-size Traffic Police Robot with Multiple Augmented Reality Aided Display Terminals

Liang Gong[1], Changyang Gong[1], Zhao Ma[1], Lujie Zhao[1],Zhenyu Wang[1], Xudong Li[1], Xiaolong Jing[1] ， Haozhe Yang[1], Chengliang Liu[1]

1. School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
E-mail: gongliang_mi @sjtu.edu.cn

Abstract: Policing of road traffic is listed in the most hazardous tasks since many traffic police personnel injured in a series of accidents at intersections. A life-size traffic cop robot called “IWI” is invented as an alternative to human traffic police in the street for directing traffic on point duty, and this paper proposes a human-robot cooperation scheme with wearable augmented glasses as an HMI to obtain a more immersing teleoperation experience with the aid of wearable augmented reality glasses. First, a humanoid robot is tailored for mobile surveillance with raspberry pi camera based eyes and omnidirectional Mecanum wheels. Second, the real-time video stream acquired by the on-site robot is distributed to multiple terminals, such as the central sever, the wearable augmented reality glasses and the mobile control tablet. Third, with a depth of focus estimation the augmented indications are displayed to aid to understand the remote scenario. Finally, the human policing results, such as STOP, PULL OVER, TURN-LEFT, are compiled and programmed as simplified patterns to control the robot body/hand pantomime. Experimental results show that the proposed methodology and control scheme is feasible in real-time application with high real-time performance of less than 0.5s latency, and open possibilities of easing the traffic jam via simultaneously scheduling multiple traffic cop robots.

Key Words: Traffic police robot, augmented reality glasses, wearable robotics, human-in-the-loop control, life-size humanoid robot, human-robot cooperation

1 INTRODUCTION

With fast development of the social economy the volume of road traffic keeps growing rapidly，which poses a hidden threat to safety of the traffic police. In addition, due to traffic noise and automobile exhaust pollution, many traffic officers are weary and exhausted under inferior working environment. Therefore, the traffic police are also known as one of the “three high” professions, i.e., high pollution, high work intensity and high risk. Traffic police are listed worldwide as one of the most hazardous occupations since many traffic police personnel injured in a series of accidents on crossroads. In view of this case, a robotic traffic police system plays an alternative role to conventional traffic control. A robot cop replaces the traffic police in the street and the traffic officers remotely monitor the traffic. Notably, in such a security-critical scenario it is still necessary to maintain a real-time optimal decision-making from traffic police, which desires a human-intervening technique for manipulating the robot with wearable surveillance facilities. This project aims to develop a human-robot cooperative system to conduct the traffic management, that is, to design a robotic system using human-in-loop control technology and multiple display terminals for remotely regulating the traffic flow. When the police staff in a police station monitors the real-time traffic situation via wearing augmented glasses, the corresponding signals such as Left -turn, Right-turn are sent to the robot cop to implement the corresponding hand gestures. This design saves a lot of human resources and opens the possibilities for communication with the future autonomously driving cars. The proposed application can not only improve the current working situation of traffic police, but also greatly improve the efficiency of traffic guidance. The operator can directly control the robot through wireless module in the command center to complete the traffic operation according to the situation of the traffic scheduling center. The system can achieve multi-terminal remote control, real-time high-definition video capture and transmission, real-time human-in-the-loop remote police control and multiple augmented reality aided display functions.

The basic design philosophy is to seamlessly merge a life-size robot into the human community and the unstructured field environment, especially in such scenarios that the humanoid robot plays an essential role in interacting with human being. Specifically, automatic traffic flow management and semi-autonomous traffic commanding are dominating parts in urban traffic research, which necessitates a remote monitoring and commanding system to exempt the traffic officers from being on duty in the street. Hence a ground-supported mobile traffic control system, which mainly consists of a humanoid robot and a mobile platform, is constructed and deployed to replace a traffic police. The traffic police robot is under remote control of the officers in the police box, using human-like body and hand gesture to signal the traffic commands. This robotic traffic signaling system can be used to reduce the need of deploying workers at various construction sites [1] [2]. Referring to a system and apparatus to control the movement of vehicles such as highway traffic and control a traffic flow nourishes the robotic system proposed in this paper [3] [4].

Since the traffic control process requires the robotic system to communicate with in-car drivers, human-in-loop and augmented reality technologies are adopted to boost the teleoperation and telepresence. Related results are achieved in robotics research fields. Jean Scholtz outlined the theory

This work is supported by National Key Technology R&D Program under Grant No. 2015BAF13B02

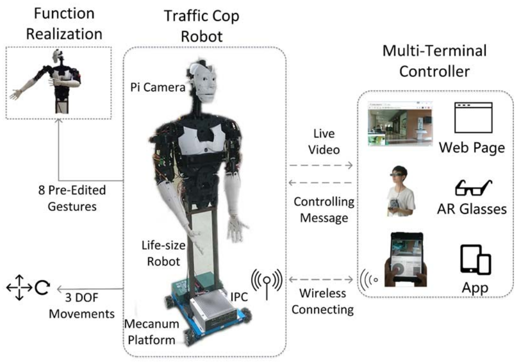

Fig 1. Control Scheme of an AR-aided traffic cop robot

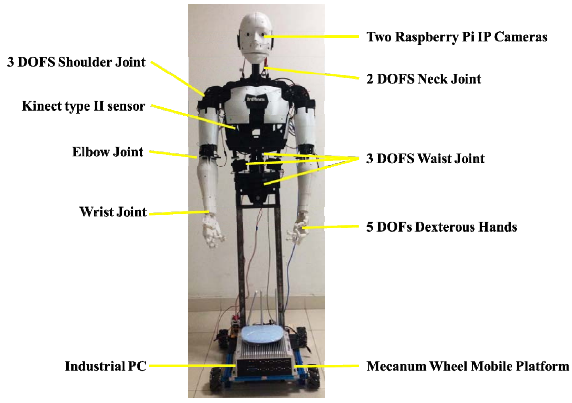

Fig 2. Mechanical structure of the traffic cop robot

and evaluation of human robot interactions and proposed the interactions and information needed by both humans and robots for the different levels of interaction[5]. Leeper[6] proposed a strategy for human-in-loop robotic grasping. KM Tsui presented a complete system [7] having the ability of retrieving a desired object from a shelf in order to operate robotic arms using human-in-the-loop control, which we can learn from. Ronald T. Azuma surveys the field of augmented reality [8], in which 3D virtual objects are integrated into a 3D real environment in real time describing manufacturing, visualization, path planning and so on. D.W.F van Krevelen researched the technologies, applications and limitations of augmented reality. O Bimber introduces in great details spatial augmented reality and how to merge real and virtual worlds [9]. Obviously, the augmented reality is an efficient means for the operator to communicate with the robot visual system in an intuitive means. It paves a new way for human-in-the-loop interaction and manipulation.

This paper organizes as follows. In section2, the setup of the traffic cop robot systems is given. In section3, video and directive transmission channels are described. Section 4 highlights the techniques of augmented reality for field depth calculation, which enhances the experiences of traffic police for more precise manipulations. Section 5 gives the details of programming the hand gestures into the robot cop. Finally, conclusions are given to enlighten more promising work in the future.

2 SETUP OF THE ROBOT COP SYSTEMS

The whole control scheme of an AR-aided traffic cop robot is shown in Fig 1. The cameras in robot's eyes record the live video and then transmit it via WIFI to remote terminals. The traffic police officials supervise and oversees the general traffic conditions in the city from multiple screens in the remote monitoring center. And a pointsman wearing the augmented reality glasses "sees" the on-road objects in a policebox, meanwhile if needed a partner police personnel can monitor the traffic condition with a tablet. These 3 terminals are able to receive the live video simultaneously with ignorable network latency. Further, both the pointsman and the remote monitoring center might control a traffic cop robot on the crossroad to direct the traffic flow. With clicks on the screen or taps on the tablet, the hand gesture commands or the movement directives are sent to the robot, hereby the robot moves to a position and makes a hand gesture to guide the traffic flow in the street.

2.1 Humanoid Robot for the Design

Humanoid robots have human-like design and are able to mimic human motions [10]. It is ideal for the humanoid robots to be similar with human in appearances and functions, especially in circumstances where robots need to interact directly with human beings. However, due to traditional manufacturing method constraints, most of the humanoid robots have high power consumption, but lack of friendly appearance. Thanks to the rapid development of the 3D printing, 3D-printed humanoid robot like Flobi [11], iCub [12] and InMoov [13] have been designed in recent years. These robots have vivid human-like appearance and can be printed in common desktop 3D printers with inexpensive resin materials. These robots have been widely used as the development platforms for Universities and Laboratories [14-16].

In this paper, the robot is based on the humanoid robot InMoov, initialed by Gael Langevin, a French sculptor in 2012 is the first Open Source 3D printed life-size humanoid robot. The traffic cop robot based on modified InMoov robot is printed with PC(Polycarbonate) material. As is shown in Fig 2, the whole robot driven by 32 servos has 29 DOFs: 10 DOFs for left and right fingers, 2 DOFs for left and right wrists; 8 DOFs for left and right arms; 3 DOFs for eyes, 1 DOF for mouth; 2 DOFs for neck; and 3 DOFs for waist. The abundant DOFs made it possible for the robot to pose various traffic command gestures.

2.2 Raspberry Pi Camera Remote Video Stream

A commercial-on-the-shelf device Raspberry Pi is selected as the processing unit to drive the camera for remote live video monitoring. The application programming interfaces provided by Raspberry Pi, like Raspistill for image capture, Raspivivid for video capture, are available for instant uasge. At the same time, the operating system Raspbian is derived from Linux operating system, so v4l2 library (video for Linux 2) is able to be easily integrated in the video frame, which is used to drive the camera and acquire the video stream from Pi camera module. In Linux operating system, camera device is regarded as a specific file located in /dev folder. So v4l2 library will track the camera device and

Table1. Comparative Test on PI

Item	Web node	live555 node
Latency	Less than 0.5s	Near 2s
Terminals	modern browsers	VLC
Resolution	720p 30fps	1080p
Remark	Occupation of computational resource	With bugs on soft decoding process

power it on, then read the camera status. When it's needed, camera configuration can be edited to fit different environment. Generally, v4l2 library applies two methods to capture the video stream, including memory mapping or reading the stream directly from the device [17]. Here we inherited the FrameSource class from v4l2 as the video capture interface, and used x264 encoder to package the stream captured as x264 encoding video stream for further processing based on Live555 library. Notably Live555 library is an open-sourced program library based on C++, mainly focusing on live stream solution, especially for multiple platforms. It provides excellent supports for various kinds of standard live stream transferring protocols, such as RTP/RTCP, and RTSP. Besides, it's compatible with multiple mainstream audio/video stream encoding forms. One typical situation where Live555 is applied to provide function realization is VLC player, which supports Live555 library.

In this project, we applied Live555 library to process the x.264 encoding video stream captured from Raspberry Pi camera module driven by v4l2 library. Live555 library made it possible for us to process video stream of 1080P resolution catering to the calculating ability of Raspberry Pi, a microchip computer. With help of this library, we transformed the video stream into a RTSP (real time streaming protocol) stream, thus we could locate the video stream with the IP address and the port and watch live video stream remotely on different terminals like PC, mobile phone, as long as it's compatible with live555 library and RTSP protocols. However, the problem with live555 library and RTSP stream is that there exists latency with the video frame. This solution works well with PC running VLC and Exacqvision software provided by Tyco company, while android device fails to watch the video stream smoothly.

Meanwhile, the other way of live video stream is a web page node situated at the Raspberry Pi. It is based on the JSMpeg-an excellent MPEG1 video and MP2 audio decoder in JavaScript. JSMpeg can decode 30fps 720P video at most. Because it is based on JavaScript, it works in any modern browser (Firefox IE10 Chrome etc.). What's more, the decoder has a low latency via WebSockets. Low latency is one of the most important features in our application.

The results of the comparative test on PI are as table 1.

In our application, we need to remote control the robot by live video, so the latency is very important. Two seconds of latency might be too long and Android APP can't decode RTSP protocol from live555. As a result, we use web node in our Android APP. However, the JSMpeg need lots of CPU resource, the augmented glasses have a stutter (5fps) when get a 720P video. The appropriate resolution is as low as 480×272, while pad runs smoothly under such settings.

3 MULTIPLE DISPLAY TERMINALS

We have different terminals of video stream owning to the two completely different stream nodes. Because of the diversity of terminals, the stream can be watched by almost electrical devices conveniently nowadays.

3.1 AR Glasses Terminal

The type of AR glasses we use is MOVERIO BT-200. These binocular, transparent smart glasses can give users more feeling of the field. With them, users can get the view from a "floating" 80-inch perceived screen. Interactive track pad and smart navigation menus put Android-based apps and content at user's fingertips. The glasses have access to Wi-Fi, which means users can watch the stream and get the situation from anywhere. Since the glasses are based on Android, all the useful features can be transferred to the glasses terminal.

All videos are in 2-Dimensions. It is inevitable that some information will be lost by watching the screen. So, it is important to use the AR glasses and some AR assistance. It can help the user to get more information and feel like being at the working field themselves.

3.2 Webpage Terminal

One of our stream nodes is based on the decoder called JSMpeg, it is a decoder using JavaScript. As a result, any devices which installed a modern browser (Firefox Chrome IE10) can have access to the low latency live video stream easily. For the users, the only things they need to do is type the URL of Raspberry Pi in their browser (like 192.168.1.155:8082) and then link them. The technical things need to be concerned are the bandwidth of the Raspberry Pi and the Internet conditions, which is mostly affected by wireless router signal.

By inserting some JavaScript HTML5 plugs in the webpage file, the webpage can even get the gravity accelerometer data. Thorough the data the web itself can act like a controller.

3.3 Android APP Terminal

In our application, the main controller is the Android APP, as is shown in Fig 3. Using the Apps we can control the movement of the robot, watch the live video stream, make the robot execute specific traffic command motion and get the distance of the exact point.

Android Apps can also provide some features by network communication with the robot. Such as make the robot to do police officer's motion, move the robot and ask the specific point on the screen. All the interface on the Pad screen has a Java thread behind it. The thread will be called as soon as the interface is touched. After called, the corresponding thread will encode the message and cyclic redundancy check numbers into a UDP package. Then the package will be sent to the target IP which represents the robot. Some

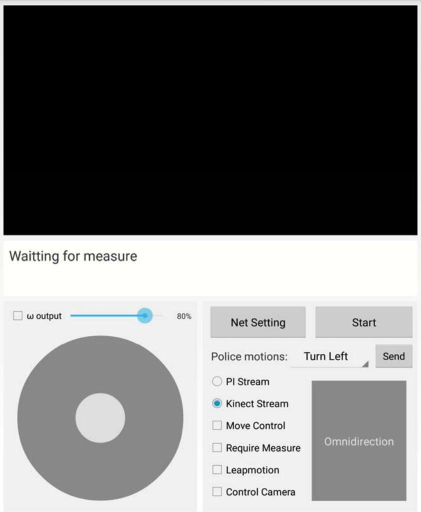

Fig 3. Android Apps GUI

features will need the feedback from the robot. There is a thread running at the background to do the specialized UDP receive work all the time.

In order to take full advantage of the Touchpad, the APP uses a touch lever for the user to control the movement. A touch lever only outputs two degrees of freedom. However, we need three for the use of Mecanum wheel. Considering a user may not need to move in all three degrees of freedom at same time, we add a square touch area on the right. When the square area is touched, the output of x-axel (horizontal) at the touch lever will be considered as left or right translation. What's more, taking everyone's habits of driving a car into account, the angular velocity will output opposite numbers when the car is reversing. At last, benefit from the help of a speed bar and the low latency, the user can control the movement very easily via the Android Apps.

The video stream source node we choose in the Apps is the Web node. There are two reasons for the choice. First of all, the latency is important in our application. RTSP has two seconds of latency which may not be acceptable. Secondly, JSMpeg is able to provide 720P 30fps HD video even though it costs more CPU resource and the live555's RTSP protocol is not compatible with the MediaPlayer class in Android.

There is another video stream, which is essentially an image stream from the Kinect part on the robot. In some situations we may need to do some work about the distance, so we need the first person view of the Kinect. However Kinect is not designed for video transmission. We can only get the images via IAI Kinect2 package from ROS topics. Then we use the file sharing protocol --- SMB via the Java package of JCIFS to get the real-time images form the robot to our touchpad. The frame rate is a little low but there is almost no latency.

4 Augmented Reality Aided Display

All users will work in front of the screens. However, people can't get the exact information from the 2D video, especially the depth of the field. The depth of field may be a very important factor for the user to evaluate the traffic situation. Errors might be made without the actual feeling of

Fig 4. The Color Depth Picture as a means for AR

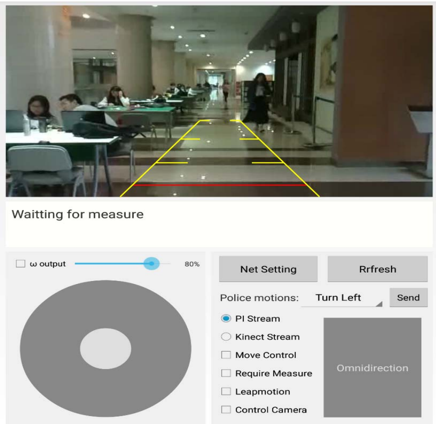

Fig 5. Equidistant Line-indicated View Depth

distance. To solve the problem, the features of augmented reality assisted technology are highlighted.

One means to get the distance information is using Kinect toolkit. Kinect has two cameras, one is normal colored camera, and the other one is an infrared camera. The infrared camera works with the infrared projector. After calculating the difference in the light phase, we can get the distance information. Using the IAI Kinect2 package in ROS, the point cloud containing the distance will be available. The depth of field renders a colorful picture where the color difference indicates the distances, as is shown in Fig 4.

Although the depth colorful pictures are very useful, the targeted objects have chances to be occluded. The second way is to draw equidistant lines (Fig 5) as an augmented reality indication for the users. Equidistant line is an intuitive way for the users to get the distance information. Comparing the object on the ground with the equidistant line, people can easily tell where the object is.

In our application, the height of the camera is fixed. As a result, the equidistant Line doesn't change very often. We can draw the lines at our terminals. In Android APP, we make our class extends from View class. By overriding the OnDraw method and using invalidate method, we are able to draw the equidistant line above the screen field.

We find it is very similar to the auxiliary lines in automobile reversing radar. The best color for the equidistant line is yellow and the danger line is red. We

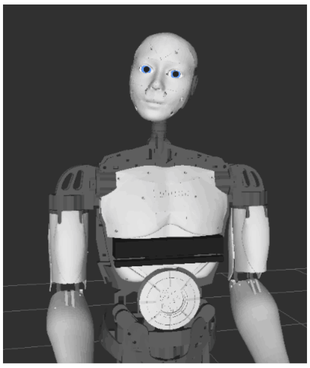

Fig 6. Urdf model of Traffic Cop Robot

adjust the width of the line field carefully and repeatedly to find the best width to compatible with the 16:9 resolution. Using two raspberry Pi cameras as robot eyes and making full use of the glasses features, we can display different cameras at different lens. After installing the cameras at suitable positions, we are able to get a pseudo-3D vision and get the distance information directly.

5 DECISION STRATEGIES AND ACTION EDITTING

5.1 Decision Strategies

Generally, the traffic environment is complex and changeable. Since it is difficult for the robot to complete the fully-autonomous judgment and make decisions, the traffic cop robot uses HITL (human-in-the-loop control) to improve system capacity and reduce costs [18]. People with traffic management experience can give action decisions through multiple display terminals including server real-time video and augmented glasses.

However, action decisions usually consist of two categories， real time quick response actions without systematic planning, and continuous complex set of actions. Quick response actions include moving, turning, rotating head and moving eyeballs (cameras). This kind of actions provides the robot with the most basic position, movement, and visual field acquisition capabilities like motion transition or altering the angle of view. The robot sends the action data to the slave controller directly and the slave controller perform the desired movement. Continuous complex movements include a set of traffic police directing actions such as the pantomimes signaling the stopping, straight forward, left turn, right turn, lane change, deceleration, pull over. Robot directs traffic and performs the functions of the traffic police through such actions. This kind of the actions is more complex and will take a relatively longer span of time to execute. The robot uses the ROS system to plan the action by constructing a simulated URDF model (Fig 6) of the actual robot and importing it into the RVIZ. By giving the initial attitude and the target pose, RVIZ will give a better collision free trajectory solution and be discretized into a series of postures that contain position and time information. The robot moves to corresponding positions in accordance with time to perform complex movements.

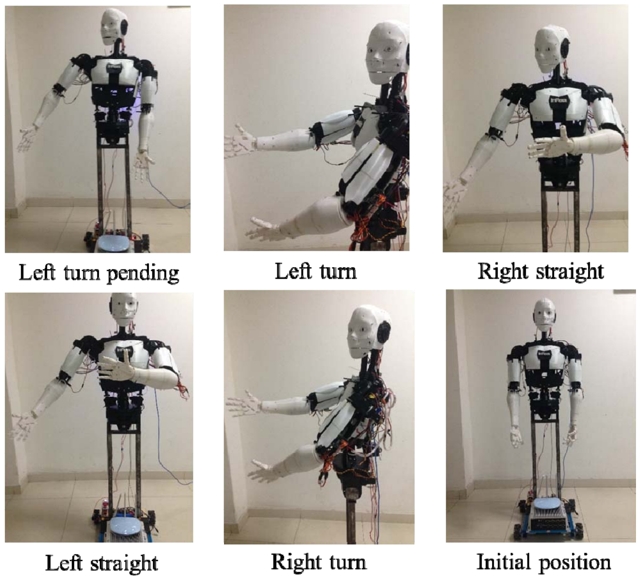

Fig 7. Typical hand gestures of traffic police commands

5.2 Action Editing

The traffic cop robot uses a kind of motion editing storage systems, and uses the system to edit a common set of command actions in advance. By configuring the action editor interface, the position parameters of each degree of freedom of the robot are modified, and the corresponding robot simulation posture is obtained. After the robot is adjusted to the key posture in the continuous operation, the key positions are kept in the system. Since the traffic police command is fixed and contains certain repetitive movements, the stored reusable key positions can avoid repeated planning and simplifies the motion planning process. The planned action data that contains location and time information will be stored by number.

Due to the memory limitation of the mega board, only a small amount of movement data can be stored in the board. The robot is equipped with an SD card outside the board and stores the motion data in a predefined format according to different operations. IPC calls the movement instructions when hand gestures and locomotion are required Mega will read the corresponding TXT file and parse to the memory. The action data package issued to NANO lower embedded processors which take in charge of the corresponding DoF. And the lower embedded processors will drive the servos to the corresponding positions with predefined time stamps. In the case of pre-editing actions, remote scheduling requires no temporary planning and improves the speed of complex action response. We plan a number of traffic command actions(as shown in Fig 7) based on the above systems. Robot body can’t move to the limit due to the problem of servo. The traffic cop robot is under test on the road, as shown in Fig.8.

6 CONCLUSIONS

The contribution of this work is threefold. First, an AR-aided monitoring and control scheme is provided to manipulate a life-size robot in a human-in-the-loop fashion.

Fig 8. Traffic police on duty

Second, 3 remote terminals are simultaneously provided with real-time site live video, which facilitates potential hierarchical multiple personnel control for traffic planning and scheduling. Third, a preset embedded gesture program exempts the robot from computationally-intensive online operation, which ensures a reliable real-world application of this traffic cop robot.

ACKNOWLEDGEMENTS

This work is supported by the National Key Technology R&D Program under Grant No. 2015BAF13B02. Tyco International Ltd provides the video-acquiring hardware in IWI eyes and the Server software Exacqvision. This work is also financially supported by a joint international capstone project program co-sponsored in Shanghai Jiao Tong university and Pennsylvania State University

REFERENCES

[1] Scovin S. Portable traffic control device: U.S. Patent 5,001,475[P]. 1991-3-19.

[2] Jones D G. Robotic traffic signalling device: U.S. Patent 6,448,905[P]. 2002-9-10.

[3] Wheaton B. Remotely-controlled traffic control system: U.S. Patent 7,902,998[P]. 2011-3-8.

[4] Entrup R N. Traffic Control Apparatus: U.S. Patent Application 11/779,279[P]. 2007-7-18.

[5] Scholtz J. Theory and evaluation of human robot interactions[C]//System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on. IEEE, 2003: 10 pp.

[6] Leeper A E, Hsiao K, Ciocarlie M, et al. Strategies for human-in-the-loop robotic grasping[C]//Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction. ACM, 2012: 1-8.

[7] Tsui K M, Kim D J, Behal A, et al. “I want that”: Human-in-the-loop control of a wheelchair-mounted robotic arm[J]. Applied Bionics and Biomechanics, 2011, 8(1): 127-147.

[8] Azuma R T. A survey of augmented reality[J]. Presence: Teleoperators and virtual environments, 1997, 6(4): 355-385.

[9] Bimber O, Raskar R. Spatial augmented reality: merging real and virtual worlds[M]. CRC press, 2005.

[10] N. N. Rodriguez, G. Carbone, and M. Ceccarelli, "Antropomorphic design and operation of a new low-cost humanoid robot," in Biomedical Robotics and Biomechatronics, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on, 2006, pp. 933-938.

[11] S. Schulz, I. Lütkebohle, and S. Wachsmuth, "An affordable, 3D-printable camera eye with two active degrees of freedom for an anthropomorphic robot," in Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on, 2012, pp. 764-771.

[12] S. Davis, N. G. Tsagarakis, and D. G. Caldwell, "The initial design and manufacturing process of a low cost hand for the robot icub," in Humanoid Robots, 2008. Humanoids 2008. 8th IEEE-RAS International Conference on, 2008, pp. 40-45

[13] G. Langevin, "Inmoov," URL http://www. inmoov. fr/project, 2014.

[14] Y. Pititeeraphab and M. Sangworasil, "Design and construction of system to control the movement of the robot arm," in Biomedical Engineering International Conference, 2015.

[15] K. A. Alkhulayfi, "Vision-Based Motion for a Humanoid Robot," 2016.

[16] M. R. Abid, P. E. Meszaros, R. F. D. Silva, and E. M. Petriu, Dynamic hand gesture recognition for human-robot and inter-robot communication, 2014.

[17] Xu J, Chen Q, Camera drive development based on V4L2[J], Computer engineering and design, 2010 (16):3569-3572.

[18] Human in the loop control in robotics for manufacturing . A. Sena, C. McGinn, A. McCreevey, C. Donovan and K. Kelly. Department of Mechanical and Manufacturing Engineering, Trinity College, University of Dublin, Dublin 2, Ireland