Navigation of a humanoid robot via head gestures based on global and local live videos on Google Glass*

Zibo Wang[1], Xi Wen[1], Yu Song[1], Xiaoqian Mao[1]

[1]School of Electrical Engineering and Automation
Tianjin University
Tianjin, China
tjuwzb@tju.edu.cn
wenxi203120@tju.edu.cn
songyu@tju.edu.cn
maoxiaoqian@tju.edu.cn

Genshe Chen[2]

[2]Intelligent Fusion Technology, Inc.,
Germantown, MD 20876, USA
gchen@intfusiontech.com

Wei Li[1, 3, 4, *]

[3]Department of Computer & Electrical Engineering and
Computer Science
California State University
Bakersfield, CA 93311 USA
[4]Shenyang Institute of Automation
Chinese Academy of Sciences
Shenyang 110016, China
wli@csub.edu

Abstract—**The navigation of a mobile robot is a central problem in robotics research. A smart wearable device called Google Glass provides a new way to achieve the human and machine interaction. This paper presents a navigation strategy for NAO humanoid robot via head gestures based on global and local live videos displayed on a Google Glass. We develop a module to establish connection from the Google Glass to the robot and detect head gestures by fusing multi-sensor data through a complementary filter to eliminate drift of the head gesture reference. We conduct an obstacle avoidance task to validate the effectiveness of the control system. An operator wearing Google Glass was able to navigate the robot smoothly.**

Keywords—Google Glass; NAO humanoid robot; global live videos; local live videos; sensor fusion

I. INTRODUCTION

Since the 20th century, intelligent robots have been applied into various fields due to the development of human-machine interaction (HMI) technology to effectively apperceive the environment for dynamic decision-making and planning, behavior control and execution [1]. Traditional input devices rely more on hands, a single operator has to adopt multiple cooperated patterns to complete complex operational tasks. So it is necessary to find a natural and flexible interaction manner to improve the control performance.

Google glass, as the representative of wearable devices, provides a new way to HMI. The Google Corporation released Google Glass in April 2012. It is regarded as an innovative wearable computer with functions including processing information, surfing the Internet, sending messages, etc.

Researchers have been engaging in the exploration of Google Glass in varies of fields. In medical fields, Google Glass is applied to observe the ECG (electrocardiogram) [2], to solve surgical education challenges in the operating room [3] or to record operation procedure [4]. An application called gPhysics is developed to perform an educational physics experiment [5]. Google Glass is also used in a guidance system for manual assembling tasks [6]. Our team, which is one of the pioneers, has reported preliminary research outcomes on the robot control via Google Glass [7][8][9].

In this study, the camera mounted on the frame of Google Glass is used as a global visual sensor to provide a broader view over an operation environment and to display the current status of the robot. A desired path for robot navigation can be found by an operator. The camera of NAO robot serves as the primary sense when the global visual feedback is not available. This camera captures images around the NAO robot in real-time and then local visual information is transmitted to the Google Glass through a Wi-Fi network. This local live video feedback is used as telepresence control of NAO robot navigation.

In our previous research [7][8], we calculate the angle change of the rotation matrix via rotation vector sensor. A critical problem which results in the failure of control is that the head reference position is significantly drifting during operations. In this paper, we adopt a complementary filter to fuse data measured by an accelerometer, a geomagnetic sensor and a gyroscope to update the operator's head reference gestures. Consequently, the reference drift is eliminated and the robot control performance is improved.
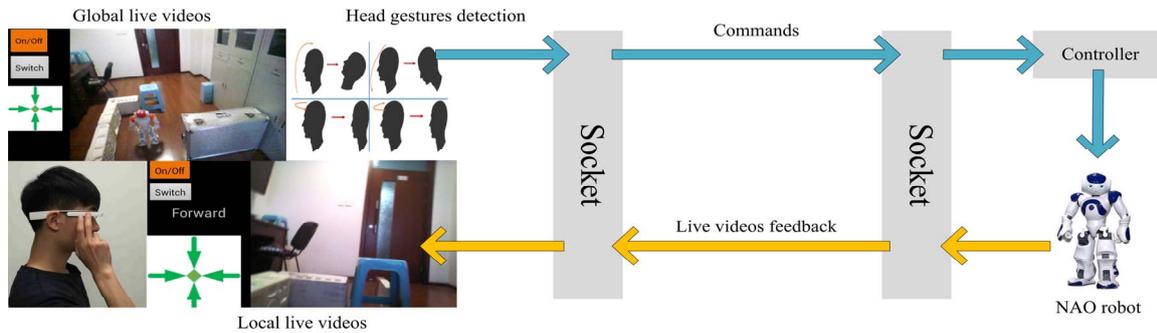
Fig. 1. Stucture of the navigation system

The rest of this paper is organized as follows: Section II describes the structure of the Google Glass-based navigation system. Section III introduces the detection of head gestures based on sensor fusion. Section IV presents the application on Google Glass and the controller of the NAO robot. We design an obstacle avoidance task in Section V to verify the effectiveness of the system. Finally, Section VI discusses the advantages of the Google Glass-based navigation system and the prospect of wearable devices.

II. GOOGLE GLASS-BASED NAVIGATION SYSTEM

The navigation system is consisted of two parts: a Google Glass and a NAO humanoid robot as shown in Fig. 1. We develop a module on Google Glass to achieve the human and robot interaction. More specifically, the head gestures are detected by multiple sensors equipped on the Google Glass, parsed into robot control instructions and then transmitted to the NAO robot through Wi-Fi network. NAO performs actions accordingly and sends back the local live videos to the Google Glass in real-time for teleoperation. Displaying live local and global videos can be switched by clicking the button on a User interface (UI).

The Google Glass-based navigation system provides a feeling of immersion, which is more natural and flexible. The combine of such an innovative device and the NAO robot is capable to accomplish a variety of tasks in complex environments.

A. Google Glass

Google Glass is an interactive wearable device composed of OMAP4430 processor, high-definition camera, bone-conduction transducer, battery, microphone, prism, speaker, and wireless connectivity[10], the structure of Google Glass is shown in Fig. 2. Optical head mounted display (OHMD) technology applied in Google Glass achieves the augment of reality.
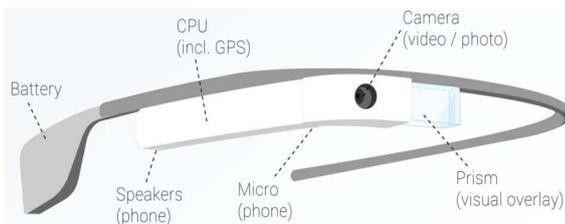

Fig. 2. Components of Google Glass

Following components are involved in the navigation system:

- Prism: demonstrate the live video feedback in the operator's view;

- Touchpad: detect the single or double tap gestures;

- Accelerometer, gyroscope, geomagnetic sensor: detect head gestures performed by the operator;

- Wi-Fi connection: Network for data interchange.

B. NAO Humanoid Robot

NAO is an autonomous, programmable robot developed by Aldebaran Robotics, known as the most widely used humanoid robot in the academic field. The components of NAO are shown in Fig. 3. It is 58cm in height and 4.3kg in weight with 25 degrees of freedom to execute specified actions. Numerous sensors are equipped to perceive the environment. Two HD cameras provide a horizontal field of view of 60.9 degrees, and vertical field of view of 47.6 degrees for recognition of facial expression. The sonars located on torso allow the NAO to detect the obstacles that may encounter. Four microphones enable the NAO to discover the origin of sounds nearby. The inertial sensor in association with four force sensitive resistors under each feet help the NAO keep balance. The NAO is capable of deducing human emotional changes from learning body language and expression. It has been applied in computer science, artificial intelligence, mathematics, physics, psychology, mechanical automation, and communications engineering in the field of academic research.
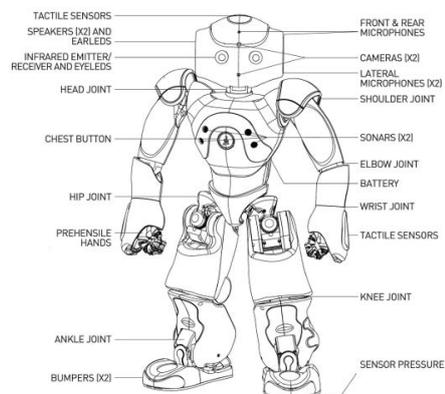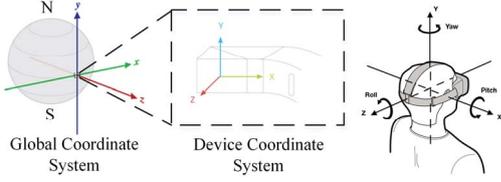

Fig. 3. Components of NAO

Fig. 4. (a) Sensor coordinate system of Google Glass. (b) Head movement diagram.

III. HEAD GESTURE DETECTION

A. Raw Sensor Data

In order to describe a head gesture, the method of space vector transformation is adopted. A sensor device coordinate system, called X-Y-Z, and a global coordinate system, called x-y-z, are defined by Android operating system, whose relationship is shown in Fig. 4 (a). We use Euler angles (Pitch, Yaw, and Roll) to describe a head gesture which represents the transform relationship between the device coordinate system and the global coordinate system [11], as shown in Fig.4 (b).

Android API (Application Program Interface) provides an access to monitor and process data from the sensors. A rotation matrix R is calculated by function getRotationMatrix from class SensorManager:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \quad (1)$$

and function SensorManager.getOrientation is adopted to derive the Euler angles [11]. Yaw, Pitch and Roll can be calculated as follows:

$$\begin{cases} Yaw & = & \arctan(r_{12}/r_{22}) \\ Pitch & = & \arcsin(-r_{32}) \\ Roll & = & \arctan(-r_{31}/r_{33}) \end{cases} . \quad (2)$$

B. Fusion of Sensor Data

The head gestures cannot be estimated precisely because of noise and drift from raw sensor data. We adopt a widely used complementary filter to eliminate the angle drift integrated by the gyroscope and the noise delivered by the accelerometer and geomagnetic sensors, as shown in Fig. 5. In order to guarantee that the sensor fusion output is accurately estimated, in this study the high-pass filter processes gyroscope data to remove a reference drift, while the low-pass filter processes accelerometer data to smooth output results. The latest signal data from the accelerometer and magnetometer are stored in *AM(n)*, while the latest gyroscope data are stored in *G(n)*. *dT* is the integral interval for processing the gyroscope data. The output *Y(n)* is calculated by
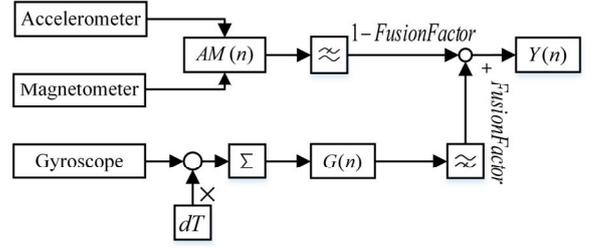

Fig. 5. Process of sensor fusion.

$$\begin{aligned} Y(n) & = & FusionFactor \times [Y(n-1) + G(n) \times dT] \\ & & + (1 - FusionFactor) \times AM(n) \end{aligned}, \quad (3)$$

where the parameter *Fusion Factor* is the weight that affects the response rate and proportion of the sensor data processed by both the high and low frequency filters. It can be calculated by

$$FusionFactor = \frac{\tau}{\tau + dT}, \quad (4)$$

where the parameter τ gives the duration of a signal. A low-pass filter can filter out signals much shorter than τ while a high-pass filter can filter out signals much longer than τ. Here, we select the parameter *Fusion Factor* as 0.9 for purpose of a fast gyroscope measurement time *dT* so the gyroscope doesn't drift more than a couple of degrees before the next measurement is taken [11][12]. *Fusion Factor* plays an important role in achieving a good control quality of the navigation system.

C. Head Gesture Detection based on Sensor Fusion

Four control instructions correspond to the four head gestures, respectively: head-up, head-turning-left, head-turning-right, and head-down. Because these head movements mainly change the value of the Pitch and Yaw, only these two values are investigated. We select the gesture of looking horizontally ahead as the reference position, denoted as *referPitch* and *referYaw* while other head gestures as *newPitch* and *newYaw*.

All the head movements can be distinguished via selecting a threshold *T* properly according to the discrimination rule below:

$$\begin{cases} newPitch - referPitch & > & T & \Rightarrow & Head\_Up \\ newPitch - referPitch & < & -T & \Rightarrow & Head\_Down \\ newYaw - referYaw & > & T & \Rightarrow & Head\_Left \\ newYaw - referYaw & < & -T & \Rightarrow & Head\_Right \end{cases} . \quad (5)$$

IV. CONTROLLER DESIGN

A. Android Application on Google Glass

Applications can be developed base on the GDK (Glass development kit). The target SDK (Software development kit) version must be set appropriately to assure the project is compatible with operating system Android 4.0.4.

The application presented in this paper consists two modules: user interface (UI, allows the operator to interact with the system) and Wi-Fi communication module (with functions of control instructions sending and the real-time videos receiving).

Development software like Eclipse or Android Studio provides varies of widgets to design the user interface. One of the widgets called ImageView is responsible for displaying real-time videos.

Wi-Fi network is the pathway for data exchange between the robot and Google Glass. Two most popular protocols are TCP (Transmission Control Protocol) and UDP (UserDatagram Protocol). TCP provides reliable and ordered delivery of a stream of bytes through three-times handshaking dialogue. UDP is a connectionless transmission model with no guarantee of order or duplicate protection. UDP is usually used in time-sensitive applications because dropping packets is preferable to waiting for delayed packets, which may not be an option in a real-time system [13]. Due to the rapidity of UDP, it is applied in the transmission of the real-time video in the Google Glass based navigation system.

B. Controller of Nao

NAO behaviors can be programmed by a Linux-based operating system (OS) called NAOqi, as shown in Fig. 6. The broker loads libraries defined in a preference file called "autoload.ini". Each library includes the modules that advertise their methods by the broker. The broker provides network access which allows the methods of attached modules to be called from outside the process[14]. The ALMotion module is responsible for controlling the NAO movements. The ALVideoDevice module is in charge of providing images from the camera to all modules processing them.
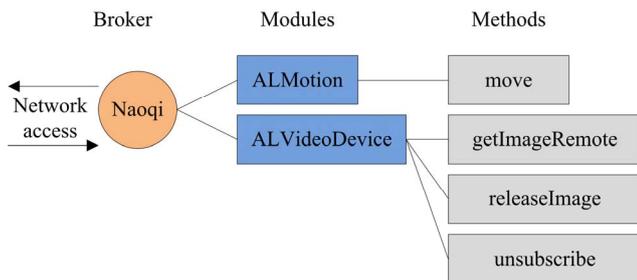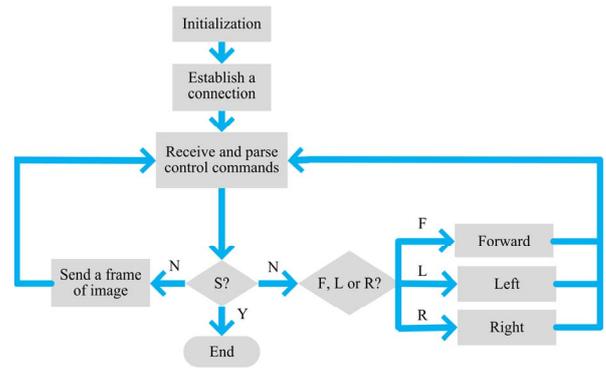

Fig. 6. Architecture of NAOqi

Fig. 7. Flowchart of the controller

The controller is mainly responsible for the execution of control commands and the transmission of live videos. The program is implemented in C++ according to the API listed in the user manual. The flowchart of the controller is shown in Fig. 7. The descriptions of the commands are listed in TABLE I.

Socket is the basis of network communication, serves as the handle provided to the programmer. The IP address of the destination must be set correctly before the establishing of Socket communication. The video stream captured by the NAO is transmitted in forms of data packets. Meanwhile, the NAO robot performs actions corresponding to the control commands received from the Google Glass.

V. OBSTACLE AVOIDANCE TASK

An obstacle avoidance task is designed to validate the effectiveness of the navigation system. The start procedure includes three steps:

1) Press the button on the NAO chest for a few seconds to start it. The NAO IP address is identified when pressing the button again. Ensure that the NAO and Google Glass are connected to the same Wi-Fi network.

2) Launch the NAO controller by Linux terminal.

3) Set the NAO IP address on Google Glass. Live videos sent by the NAO are displayed on the UI when the connection is established successfully.

TABLE I.   DEFINITION OF COMMANDS.

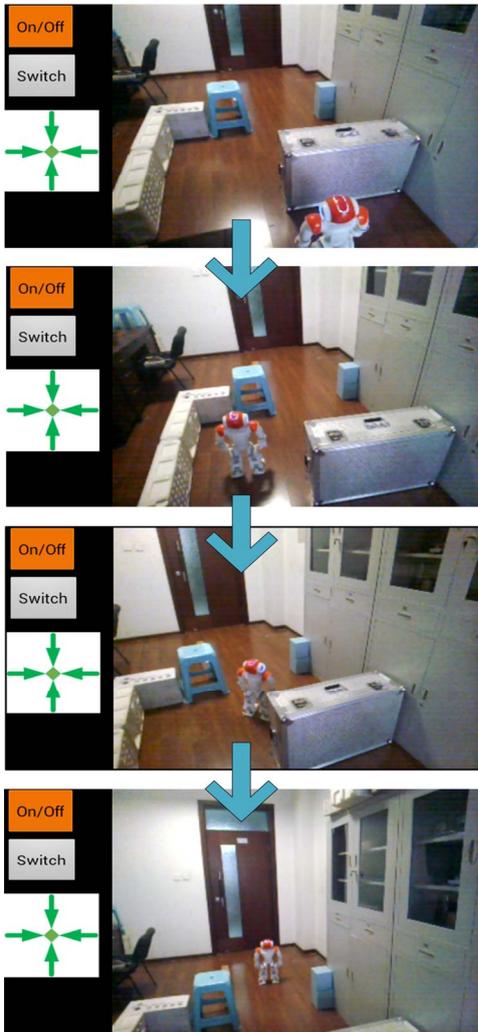| Command	Definition
F	Moving forward
L	Turning left
R	Turning right
S	Stop

Fig. 8. Navigation based on global live videos

An operator navigates the NAO via head gestures from an initial position to a destination by passing through narrow paths and avoiding obstacles. The head gestures detected by Google Glass are parsed into control instructions and transmitted to NAO robot.

As shown in Fig. 8, the global live videos deliver the NAO robot current status and its surrounding information so that it is easy for the operator to find proper paths for the NAO robot.

In case of that the robot is out of sight and distance away from the operator, local live videos are selected to display on the Google Glass and the telepresence control module is switched on to navigate the robot. Local live videos provide a feeling of immersion, as shown in Fig. 9. Details of the environment show more specifically but the field of vision is limited by the angle of the camera mounted on the NAO robot. This may lead to the disorientation of the operator and additional commands to adjust the deviation. Thus, the operator has to perform head gestures promptly according to the local live videos feedback. The amounts of the control commands are shown in Fig. 10 which indicate that more commands are needed in the navigation experiment based on local live videos.
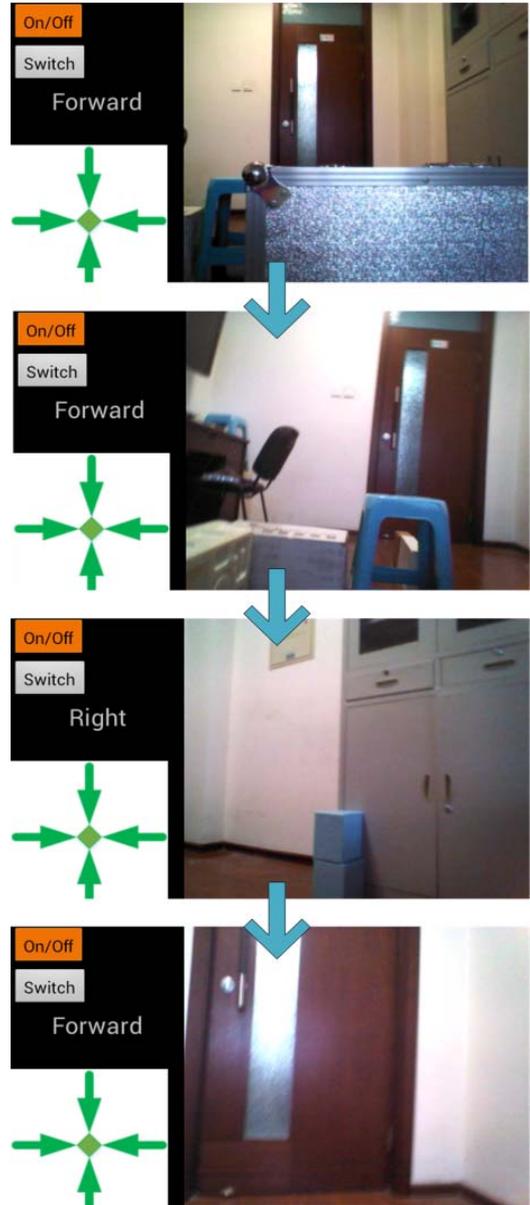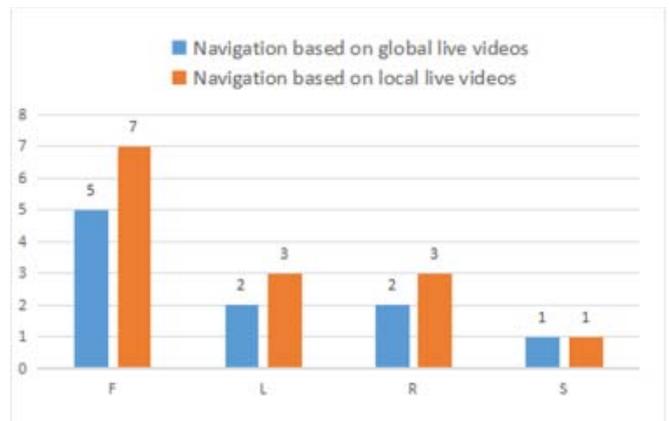

Fig. 9. Navigation based on local live videos

Fig. 10. The amounts of the control commands

VI. Discussion

Our future life is more likely to develop a closer relationship with wearable devices due to the fact that flexible materials with high ductility like graphene has been applied in electronic field. Google Glass, as the pioneer and the representative of augmented reality wearable devices, has more potentials to be explored. The promotion of processors with powerful data computing ability may bring about the prevalence of wearable technology, which changes the way people interact with the environment.

The method proposed herein is a novel scheme for the robot navigation based on global and local live videos achieved by integrating the Google Glass and the NAO.

1) Compared with inertial measurement unit (IMU) with Bluetooth system, Google Glass is equipped with more types of sensors, such as a video camera and Wi-Fi communication capabilities. Especially, the Google Glass-based navigation system is portable and very impact as it doesn't need a personal computer (PC) or laptop for operations.

2) Google Glass provides an innovative and natural way to achieve the human and robot interaction. The operator gets a strong feeling of immersion and enjoys the novel control experience due to the head-mounted-display technology.

3) The global and local live videos can be switched depending on different circumstances. The global visual sensor is a preferable choice when the robot is in sight because the explicit position of the robot provides a strong sense of direction. The local visual sensor is more applicable to teleoperation, for instance, inspection of areas with limitations of safety or structural architecture in which global information about environments are unavailable.

4) Removing the drift of head gesture reference greatly improves the control quality.

This research intends not only to demonstrate the navigation of a humanoid robot based on global and local visual sensors, but also to inspire the design of a wearable device-based robot control system. In further research, we will explore the Google Glass-based robot control system by integrating additional interaction methods, such as voice input, to make robot control more convenient and more natural.

References

[1] G.B. Xu, Y.X. Yin, and M.K. Zhou, "The present situation and prospect of the technologies of intelligent mobile robots," Robot Technique and Application, no. 2, pp. 29-34, Jun. 2007.

[2] O.M. Jeroudi, G. Christakopoulos, G. Christakopoulos, A. Kotsia, M.A. Kypreos, B.V. Rangan, S. Banerjee, and E.S. Brilakis, "Accuracy of remote electrocardiogram interpretation with the use of Google Glass technology," The American journal of cardiology, vol. 115, no. 3, pp. 374-377, Feb. 2015.

[3] O. Moshtaghi, K.S. Kelley, W.B. Armstrong, Y. Ghavami, J. Gu and H.R. Djalilian, "Using google glass to solve communication and surgical education challenges in the operating room," Laryngoscope, 2015, vol. 125, no. 10, pp. 2295-2297.

[4] E. Rahimy, and S.J. Garg, "Google Glass for Recording Scleral Buckling Surgery," Jama Ophthalmology, vol. 133, no. 6, pp. 710-711, Mar. 2015.

[5] J. Weppner, M. Hirth, P. Lukowicz, and J. Kuhn, "Physics Education with Google," in Proc. ACM International Joint Conference on Pervasive & Ubiquitous Computing: Adjunct Publication, 2014, pp. 279-282.

[6] M. Wille, P.M. Scholl, S. Wischniewski and K.V. Laerhoven, "Comparing Google Glass with Tablet-PC as Guidance System for Assembling Tasks," in Proc. International Conference on Wearable and Implantable Body Sensor Networks Workshops, 2014, pp. 38-41.

[7] X. Wen, Y. Song, W. Li, and G.S. Chen, "Rotation Vector Sensor-Based Remote Control of a Mobile Robot via Google Glass," in Proc. Robot Intelligence Technology and Applications 4, 2017, pp. 581-588.

[8] X. Wen, Y. Song, W. Li, G.S. Chen, and B. Xian, "Rotation Vector Sensor-based Remote Control of a Humanoid Robot through a Google Glass," in Proc. International Workshop on Advanced Motion Control, 2016.

[9] Y. Song, X. Wen, W. Li, and G.S. Chen, "Google Glass-based remote control of a mobile robot," in Proc. Sensors and Systems for Space Applications IX, 2016.

[10] O.J. Muensterer, M. Lacher, C. Zoeller and J Kübler, "Google Glass in pediatric surgery: an exploratory study," International Journal of Surgery, 2014, vol. 12, no. 4, pp. 281-289.

[11] G. Milette, and A. Stroud, "Professional Android Sensor Programming," in John Wiley & Sons, Inc. 2012.

[12] T. Wang, and Y.F. Niu, "Sensor Fusion Technology to Achieve by Complementary Filter," Computer Knowledge and Technology, vol. 10, no. 25, pp. 5950-5954, Sep. 2014.

[13] K.W. Ross and J.F. Kurose, "Computer Networking: A Top-Down Approach Featuring the Internet," Network, 2001, vol. 18, no. 142, pp. 1–11.

[14] http://doc.aldebaran.com/2-1/dev/naoqi/index.html