A Wearable Device for First Person Vision

Michaël Devyver
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213, USA

Akihiro Tsukada
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213, USA

Takeo Kanade
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213, USA

***Abstract*—We have developed, built and tested a wearable device made to monitor, record and assist people in their daily lives, also known as First Person Vision device. It consists of a scene camera and a non-active lighting eye camera as well as audio and movements sensors. It is built to be worn on any type of eyeglasses and optimized for shape, size and weight. The resulting data are recorded on-board or transmitted to an external computer for further processing. Some images are captured and used successfully in vision algorithms. They show how such a product is useful to improve the quality of life of persons with disabilities.**

I. Introduction

First Person Vision (FPV) is a new concept [1] that augments human cognitive functions. By working alongside patients and users, FPV devices provide them with support in their daily activities. FPV devices can analyze people's intentions by tracking certain signals such as the eye gaze, providing feedback such as information (about someone or something) or helping, for example, by triggering a nurse alert. The ultimate goal of a FPV device is to work side by side with people and understand their behavior in order to improve the quality of life, in the same way as a caregiver. It would be particularly helpful given the increasing number of disabled and elderly people in our society today.

A. Related work and problems

Most non-invasive devices nowadays are using Video-Oculography (VOG) [2], the measurement of the eye position using video, in order to better understand the focus of attention [3]. Fixed systems, specifically made to be placed in front of a computer, for example, are a mature and common technology; however, they dramatically reduce the movements of the subject and are not appropriate for applications where people are mobile. Some portable products are offered on the market, including Tobii[1] or SMI[2] however, they are targeted toward other applications [4] (such as marketing) and do not provide real-time feedback, or an easy-to-use open interface for human-computer applications such as FPV. Some open eye-tracking systems have recently appeared in the scientific community [5] [6], although these devices are shaped in the form of eyeglasses which can be cumbersome when the user already has glasses, and their usability can be questioned. Finally, most of these devices only track the eyes and do not measure other signals such as audio or movements.

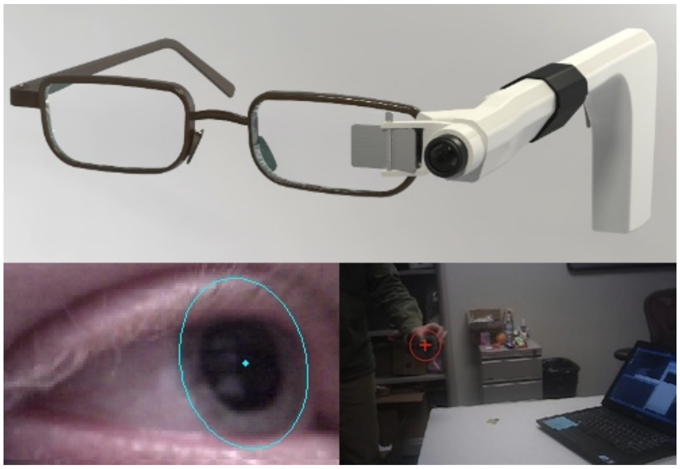


Fig. 1. General overview of the device attached to eyeglasses (up) and gaze tracking example taken with the system (down).

B. Our work

Our goal is to build a wearable system that disabled people would wear continuously during the day for the purpose of assisting them as well as collecting data about their performances during daily activities. In order to achieve that goal, FPV devices rely mostly on tracking the movements of the eye and computing the gaze. By combining gaze (”where I am looking?”) and environmental information, we not only understand people's interests or intentions but also their behavior [7], and thus, try to provide solutions to their needs. The wearable device is part of a larger system that performs the following tasks: captures data on the user and from the surrounding scene, combines and analyze them for specific signals and finally, provides feedback to the user or any other competent person.

We propose a non-invasive hardware system (Fig. 1) that is portable enough so it can be worn in any day-long situation without comprising the user's comfort. The device must be able to record images of the eye and the surrounding scene in real-time, as well as audio and movement data, in a usable manner. Usability is defined here as the device's weight, size and shape optimization and is a very important part of the process. While most eye-tracking devices use infrared to improve the accuracy of the measurements in a large spectrum of lighting situation, this system does not use active lighting. The reason is that it is worn for multiple hours, even days. So it is better not to use prolonged irradiation due to safety,

[1] Tobii Technologies, http://www.tobii.com

[2] Sensomotoric Instruments, http://www.smivision.com

especially for small children [8].

Finally, we validate the system by capturing some images then applying vision algorithms on them. We show how such a vision device, combined with gaze tracking and face recognition, can be helpful to people with face blindness or Alzheimer's disease.

II. System overview

The acquisition system is a key component of the wearable device for First Person Vision and allows us to capture the required data to understand the user's behaviors and intents. To summarize, it basically consists of two cameras: one looking at the eye of the subject and one looking towards the environment. The system consists of a main circuit, an extension and a camera module (Fig. 6). The extension is not essential and the main board can be adapted for various cases and design. Two recording modes are available: local recording or remote transfer. Local recording (on a memory card with wireless updates) is the preferred mode for day-long activities performed using the device collecting data.

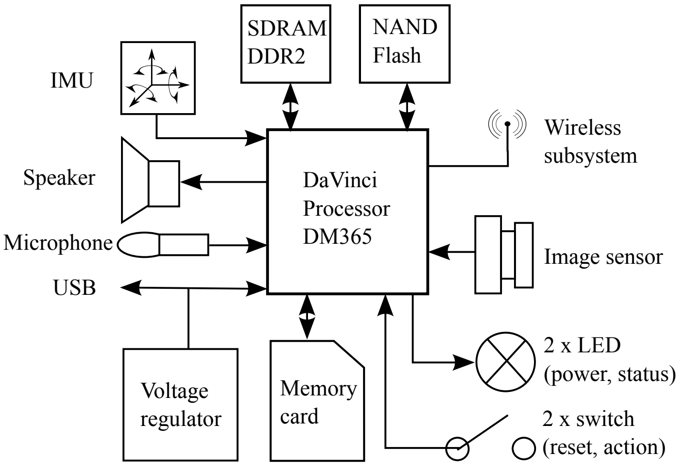


Fig. 2. Functional design of the circuit board.

III. System implementation

In this chapter, we describe how the device is built, starting from the shape choice to the electronics, the sensors, the embedded software, the data transmission and finally, the design.

A. *Design survey*

A small survey was conducted among 57 stakeholders, students and staff at the Quality of Life Technology Center at Carnegie Mellon University. They were asked to choose among four different versions: headset, pendant, eyeglass clip and earbud. The primary focus of the poll was about aesthetics, willingness to wear and privacy issues. The devices were presented and briefly explained in front of the participants so they could see them.

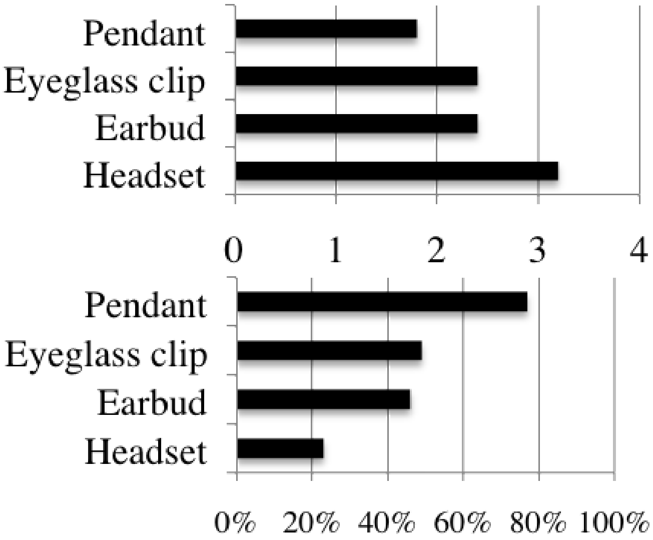


Fig. 3. Design survey (N=57): ranking from 1 (best) to 4 (worst) (up) and willingness to wear in public (down) for every type of design.

When they were asked either to rank the different versions from the best to the worst or which one they would be willing to wear in public (Fig. 3), the pendant version came first in both surveys. Eyeglass clip and earbud came almost even in the second place and the headset version came last. Consequently, we chose the eyeglass clip, as the pendant version is not feasible if we want to capture eye movements, since it is too far from the face.

B. *Main circuit board*

The device is made of two identical custom-made Printed Circuit Boards (PCB). See Fig. 2 and Table I. One circuit is being used for eye recording, while the other is for forward scene capture. The core architecture of each circuit is centered around the Texas Instrument's DaVinci video processor. Its dedicated Video Processing Front-End (VPFE), allows image acquisition (up to 30 frames per second at a resolution of 1280x720[px] progressive) and JPEG compression to be performed on-board, therefore reducing the circuitry and the bandwidth during transmission. It also offers all the functionalities needed to efficiently record (audio and Inertial Measurement Unit values) and transmit the data (locally on the memory card or remotely on the computer), therefore reducing the footprint of the circuit and thus, the weight and size for the user. This circuit is powered using the 5[V] USB connector, which can also be used as a battery input.

In order to reduce the size of the board on the user's head, each circuit is made of 6 copper layers produced with a resolution of 0.004[in]. 0402 SMD components were also privileged for their small size. The overall dimensions of each circuit are 25x55[mm] with a thickness of 8[mm] and a weight of 8[g]. Using two circuits only doubles the thickness to 16[mm] and does not change other dimensions.

C. *Video acquisition*

1) Camera: Video is acquired for both eye and scene views through two 5M camera boards [ref. Leopard Imaging LI-

TABLE I
BILL OF MATERIALS (SUMMARY)

Function	Reference	Requirements
DaVinci processor	TMS320DM365	Video, sensors proc.
Accelerometer	ADXL322	Measure acceleration [9]
Gyroscope (X-Y axis)	LPR530AL	Measure head rotation
Gyroscope (Z axis)	LY530ALH	Measure head rotation
Image sensor	MT9P011	Take pictures
Light Emitting Diode	SML-LXT0805	Be visible by user
Memory card	32GB SD card	Data storage (day-long)
Microphone	SPM0408LE5H	Record ambient sound
NAND Flash	MT29F2G08AAD	Program memory (2Gb)
SDRAM DDR2	MT47H64M16	Temporary memory (1Gb)
Speaker	Audio jack	Emit sound to user
USB connector	ZX62-AB	Smallest (microUSB)
Voltage regulator	TPS65053RGE	Power for system
Wireless subsystem	W2CBW003	WiFi 802.11b/g

LBCM5M1]. Each of them consists of an Aptina 1/2.5 CMOS Sensor (see Table I) encased in a plastic lens (with a vertical Field of View of 60.3°) and connected to the main circuit using a flexible printed circuit board. As mentioned in the introduction, the device does not emit infrared lighting towards the eye and only captures the image of the eye. Hence, there is no prolonged irradiation.

2) *Optics:* Various initial tests led us to change the optics on the forward-looking camera. While the purpose of our device is to capture the same images as a first person viewpoint, the characteristics of the camera modules were not good enough given their low field of view. Therefore, in order to mimic the capabilities of the eye, a Fish Eye lens was added to the camera module. The glass lens [ref. Sunex DSL215] has a wide field of view (188° horizontal, 128° vertical) allowing us to capture a full picture of the direct surroundings of the subject. In order to use it, we removed the lens and its holder from the initial camera and put the new optics on a specially made custom thread (size M12) integrated into the casing of the device.

Due to the large focal length of the eye camera, a mirror is placed between the image sensor and the eye in order to increase the distance.

3) *Extension:* A design choice was made that consists of decoupling the camera from the main circuit board. Because the size of the latter and our desire to leverage the use of the ear as a support, the heavier and larger component, in this case the main circuit, was placed behind the ear, in the same fashion as hearing aids, while the cameras were brought forward, near the eye. This allowed the non-essential parts to be hidden. A solution had to be found in order to create that extension, linking the camera to the main board. Because flexible circuits can be very expensive (around $5,000 or more), a custom rigid circuit board (PCB) was produced to fit all along the side of the glasses. Each extension weighs around 2[g].

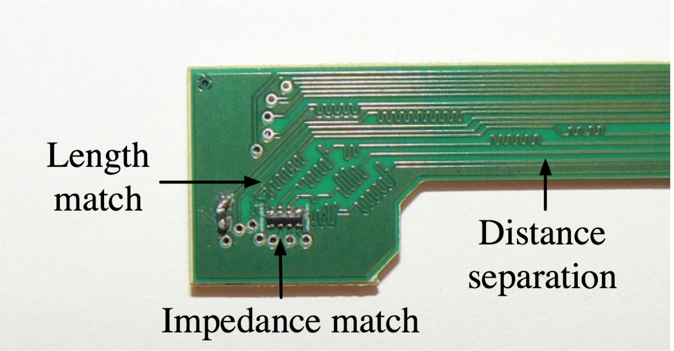


Fig. 4. Detailed view of the high-speed extension board with the principal features to prevent high-speed signal noise on the camera lines.

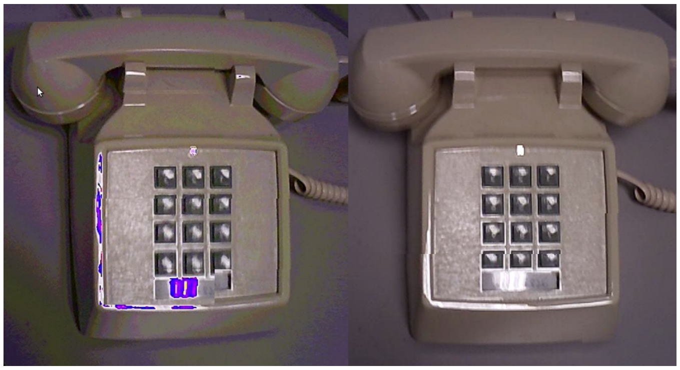


Fig. 5. Before (left) and after the high-speed extension board (right). Notice the color saturation (purple noise) on the left image and how it disappears with the new high-speed extension circuit.

While testing this new configuration, the resulting image quality was poor and noisy, with some purple noise lines appearing randomly on the image. We discovered that CMOS sensors are to be placed close to the processor. Due to the high-speed line (up to 96[Mhz]), the circuit must be optimized for speed on longer lines (around 200[mm] in this case). A solution was found by creating a custom Printed Circuit Board (PCB) with special features to prevent noise at high frequencies [10]:

1) Length match: set the same distance for all lines,
2) Impedance match: place serial resistors,
3) Distance separation: separate every line with 3 times its width,
4) PCB material: choose one with a small loss factor.

The resulting circuit is shown on Fig. 4. These small adjustments helped to mitigate the noise on the line and the resulting image was crystal clear (Fig. 5).

D. Sensors

Sensors measuring the movements of the user, as well as recording ambient sound, are present on the wearable device in order to get a full idea of the patient's state. Using an Inertial Measurement Unit (combining accelerometers and gyroscopes), we are able to get a full idea of the anatomical, physiological, mechanical, environmental, sociological and psychological behavior of the subject [9]. In addition to that, audio is being recorded for the purpose of analyzing interactions that the user might have with other people. Two Light Emitting Diodes (LED) were added to provide some visual status of the device activity (power ON, recording ON) and two switches are available for the user to change these statuses (reset, record ON/OFF).

E. Embedded software

The DaVinci processor runs on an Embedded Linux version provided by Texas Instruments. It offers all the convenience of the kernel with the necessary tools and drivers for programs to use. GStreamer is being used to stream video and audio to the memory card or Wireless. JPEG compression is done in real-time with the same program which interacts with the DSP side of the processor and compresses images on the fly. Besides the video streaming, a custom-made program has been made to read the value of the switches and, if necessary, pause the recording. It also captures the data coming from accelerometers and gyroscopes.

One of the main challenge, when using two cameras is the synchronization of the images. Every image taken with each camera must be timestamped in order to recombine the eye and the forward looking camera together for later processing. The idea here is to use the processor clock, with a resolution of 1 millisecond, as the basis for the time for every image. The only condition to make it happen is to power up both circuits at the same time so that their clock counter starts at equal times. This can be done if both power supplies are coming from the same source, which is the case in this device.

F. Data recording & transmission

The video, audio and movement data can be transmitted or recorded (thanks to the capabilities of the DaVinci processor), through different options: remotely through USB or Wireless, or locally on a memory card. The USB mode is useful if the device is used near a computer and the wearer wants real-time feedback. The Wireless mode, although not offering enough bandwidth for all the data, is good for a punctual verification of the data but does not replace USB or memory card for recordings. The last mode, using the memory card, is good for local recordings in places where there are no computers (for example outside) and provides a greater mobility for the subject.

G. Casing design

The function of the casing (Fig. 6) is two-fold: protect the circuit from external threats (fingers, objects, liquids) and attach the device to the user's head. The casing fits all along the side of the glasses and clips itself on the eyeglasses of the user, assuming the user wears them. If that is not the case, eyeglasses without lenses are provided as a support.

The casing was drawn using a Computer-aided Design system with size, weight and comfort in mind. It was printed using StereoLithoGraphy.

The total weight of the system, including the casing, the circuits and the cameras is around 50[g].

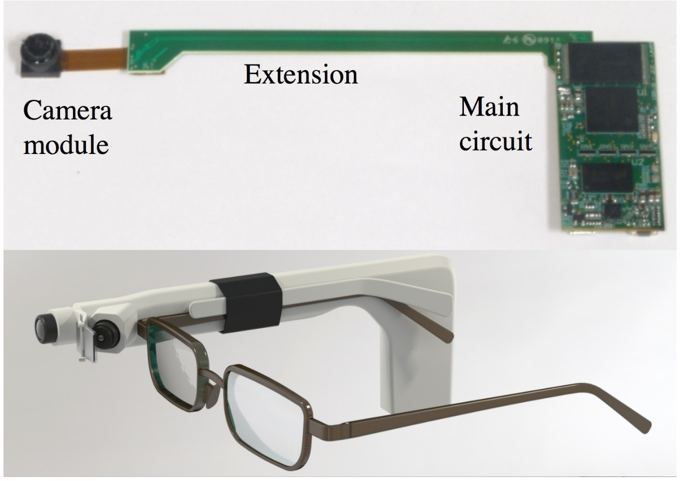


Fig. 6. Resulting circuit (up) and casing attached to some eyeglasses (down).

IV. System experimentation

The device was tested in one of the multiple applications offered in the field of First Person Vision. This scenario involves one person wearing the device and looking at people's faces. The computer algorithm recognizes the faces and gives their names as feedback. See Fig. 7. Such an application is useful to assist people with face blindness or Alzheimer's disease to help them recall the names of people they know.

A. Methodology

Images captured from the wearable device are streamed live at 30 frames per second in High Definition 720p format (without Fish Eye lens) and sent to a laptop using the USB connection. The computer being used is a Dell with a Core i5 (2.4GHz) processor, 4GB of memory and Windows 7. The images are processed using custom-made software running on OpenCV 2.1 to estimate the user's gaze.

The gaze tracking software consists of an edge detection algorithm combined with ellipse fitting (Fig. 7(a)). Using a look-up table generated during calibration, the position of the eye is then linked to the position on the scene camera, creating a heatmap (Fig. 7(b)). The PittPatt Face Recognition API[3] is applied to detect and identify the faces on the picture (Fig. 7(c)).

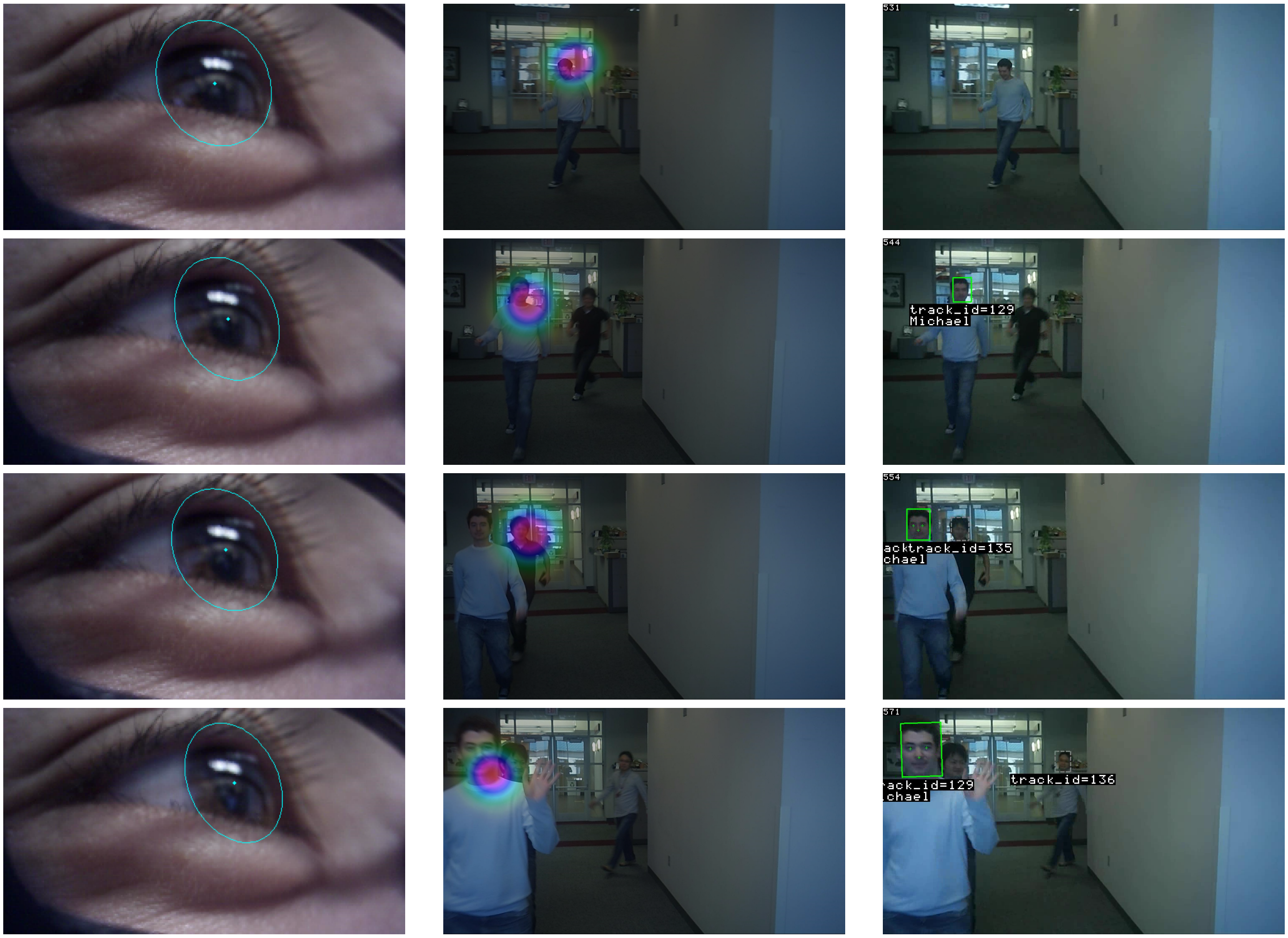
This example shows how easy it is to use the device and how gaze information is useful to deduce a person's interests by estimating where they are looking.

V. Conclusion

This paper has presented challenges associated with building a device for people to wear in First Person Vision applications. Not only should it satisfy the users needs (size, weight), but it also needs to provide good images and data to the computer for processing.

We have conceived and manufactured a device that is optimized for size and weight, comparable to commercial

[3] Pittsburgh Pattern Recognition, Inc. http://www.pittpatt.com



(a) Eye tracking (b) Gaze tracking (heatmap) (c) Face recognition

Fig. 7. The user is wearing the device and looking at people's faces. Our eye tracking algorithm is processing the eye image using ellipse fitting (left) and estimating the gaze of the user (center). The PittPatt API is then recognizing the faces on the picture and displaying names (right).

products, and that captures video images in a delayed or real-time fashion to any computer. The generated images have sufficient quality, frame rate and resolution to be processed by common vision algorithms and used in automated applications. We successfully showed such an example with gaze tracking and Face Recognition.

The device is currently being tested in various applications such as a study involving older adults performing daily activities. It is also being used in sport, security and piloting applications.

ACKNOWLEDGMENT

We would like to thank our colleagues who helped us write this paper. This includes Jason May, Zachary Mason, Scott Beach, Masafumi Takimoto, Martial Hebert, Myung Hwangbo and Mark Baskinger. We also would like to thank PittPatt and Altium limited for providing us with their respective software.

REFERENCES

[1] T. Kanade, "First-person, inside-out vision," 2009, keynote speech, Proceedings of 2009 IEEE Computer Vision and Pattern Recognition Workshops: First Workshop on Egocentric Vision.

[2] L. Young and D. Sheena, "Survey of eye movement recording methods," *Behavior Research Methods*, vol. 7, pp. 397–429, 1975, 10.3758/BF03201553. [Online]. Available: http://dx.doi.org/10.3758/BF03201553

[3] M. A. Just and P. A. Carpenter, "Eye fixations and cognitive processes," *Cognitive Psychology*, vol. 8, pp. 441–480, 1976.

[4] A. Duchowski, "A breadth-first survey of eye-tracking applications," *Behavior Research Methods*, vol. 34, pp. 455–470, 2002, 10.3758/BF03195475. [Online]. Available: http://dx.doi.org/10.3758/BF03195475

[5] D. Li, J. Babcock, and D. J. Parkhurst, "openeyes: a low-cost head-mounted eye-tracking solution," in *Proceedings of the 2006 symposium on Eye tracking research & applications*, ser. ETRA '06. New York, NY, USA: ACM, 2006, pp. 95–100. [Online]. Available: http://doi.acm.org/10.1145/1117309.1117350

[6] S.-H. Yang, H.-W. Kim, and M. Y. Kim, "Human visual augmentation using wearable glasses with multiple cameras and information fusion of human eye tracking and scene understanding," in *Proceedings of the 6th international conference on Human-robot interaction*, ser. HRI '11.

New York, NY, USA: ACM, 2011, pp. 287–288. [Online]. Available: http://doi.acm.org/10.1145/1957656.1957774

[7] M. Hayhoe and D. Ballard, "Eye movements in natural behavior," *Trends in Cognitive Sciences*, vol. 9, no. 4, pp. 188 – 194, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/B6VH9-4FM9MR7-1/2/04d487b0857b320b1edc8de807bf906f

[8] B. Noris, J.-B. Keller, and A. Billard, "A wearable gaze tracking system for children in unconstrained environments," *Computer Vision and Image Understanding*, vol. 115, no. 4, pp. 476 – 486, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/B6WCX-51M60JB-1/2/f2ad47670de8b54c0b8be10ba460a4ad

[9] A. Godfrey, R. Conway, D. Meagher, and G. ÓLaighin, "Direct measurement of human movement by accelerometry," *Medical Engineering and Physics*, vol. 30, no. 10, pp. 1364 – 1386, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/B6T9K-4TW53FN-1/2/e70fb09eba74af2d82757fb46e7b3ebd

[10] H. W. Johnson and M. Graham, *High-speed digital design: a handbook of black magic*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.