Demo: Human Perception-Enhanced Camera System for Web Conferences Leveraging Device Motions

Anish Shrestha, Zeyu Deng, Chen Wang
Department of Computer Science, Louisiana State Univeristy. Baton Rouge, LA 70803
ashre18@lsu.edu, zdeng6@lsu.edu, chenwang1@lsu.edu

ABSTRACT

We present a demonstration of a human perception-enhanced camera system for web conferencing that protects the user's privacy. Given that people easily forget about their active camera during web conferences, the system advertises the camera's active status via its motions to remind users that they are being watched by others. This prevents inadvertent privacy leakage. The system is developed based on a motorized camera, which moves according to the user's head coordinates just like an eye is looking at the user's face in front of the desk rather than remotely or virtually. The basic idea is to exploit the original human body sense of environmental motions for human-camera interaction, which does not require looking straight at the camera or its LED light to actively check its status. In this demonstration, we showcase our implementation of the human perception-enhanced camera system and invite participants to use the system for web conferences (e.g., Zoom and Google Hangout), which illustrates the system's ability to extend the virtual social interaction to the physical world and the effectiveness of using the camera motion as a non-intrusive awareness indicator.

KEYWORDS

Web Conference, Privacy, Motorized Camera

1 INTRODUCTION

Web conferencing is an online technology that allows users in different locations to hold face-to-face meetings virtually. A user is able to cast live video to other participants using diverse web conferencing Apps such as Zoom, Google Hangouts and Microsoft Teams, which greatly shorten the distance between physically separated people. These Apps have seen a giant surge of subscribers especially during the Covid-19 pandemic [6]. However, during web conferences, people are found to easily neglect that the camera is turned on because it is difficult for such virtual interactions to present real in-person meeting feelings. The user's focus may be dominated by presentation slides or distracted by other incidents, such as a phone call, an email and a visitor. Forgetting an active camera can lead to severe privacy issues and many embarrassing movements (e.g., zoom fails).

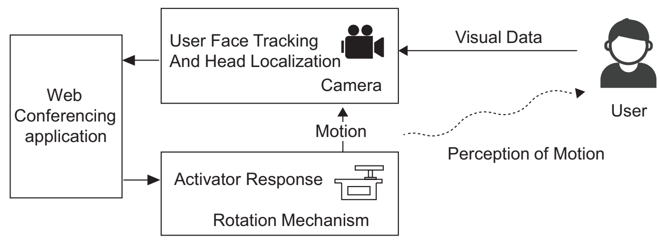

Figure 1: The workflow of the proposed camera system.

The most widely used camera indicator is the LED light on the camera, which signals when the camera is on or off. However, the user must actively look at the LED light to check the camera, which requires high user effort and it is hard to remind the user in time for many scenarios. Moreover, such a constant light signal is not so effective to arouse the user's attention, as humans are less sensitive to the unchanged settings [8]. Besides light indicators, current web conference software also uses notifications, indicator icons and the user's own video displayed on the screen to reminder the user of a working camera. But these methods all need the user to take the initiative for web camera privacy protection.

To solve the web camera awareness challenge, we propose to exploit the inherent human body sense of environmental motions, which originates from the prehistoric man's vigilance of potential large predators in the surrounding [10]. Specifically, we use a motorized camera to imitate a live eye and convey the "in-person" meeting feeling to the user. The camera broadcasts its active status, and the user passively receive the signal through the peripheral vision, hearing and the skin hair. Prior works have studied the effectiveness of varying and static stimuli and the advantage of using motions over luminance and colors for capturing human attention [4, 5, 11]. This work will, for the first time, demonstrate the use of such concepts in the human-camera interaction for protecting the user's web camera privacy.

The workflow of our system is shown in Figure 1, which works with current web conferencing Apps on laptops or

ACM MobiCom '21, March 27-April 1, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-8342-4/22/01.
https://doi.org/10.1145/3447993.3510591

desktops. The camera system creates two loops: 1) The regular camera data input to the Apps are used for tracking the user's face and estimating the head coordinate, which controls the actuators to move the camera like a person is looking at the user's face in the talk. Data adjustment is performed in the loop to coordinate the face tracking and actuator response. 2) The motions of the camera is perceived by the user passively, who consistently realizes the camera status and behaves in the camera like in a real face-to-face meeting. In this demonstration, we showcase the proposed camera system in a web-conference session and invite participants to experience this web camera awareness technique and compare it with four types of existing camera indicators.

2 SYSTEM DESIGN

The proposed system is designed to take regular video inputs for processing and feedback camera motions to let users consistently realize that they are being watched in the meeting. Moreover, we track the user's face during web conferences to allow the user to stay centered in the camera. The system design mainly consists of two components.

Face Tracking module: This module processes each video frame to detect faces in real time and track the face of the nearest user, if there are more than one users in the camera. After testing three face detection methods regarding the frame rate, memory consumption, latency, accuracy and hardware limitations, we use the Histograms of Gradients(HOG) over the other two methods, Haar cascade filters and Convolutional Neural Networks(CNN)[3, 7, 9]. Moreover, we use OpenCV libraries with our pre-trained model to generate a rectangle that bounds the user's face. The rectangle with the highest area is selected, and its center is calculated as (x_{face}, y_{face}). This is taken as the current coordinate of the face in the frame at time t and compared with the frame center (x_{frame}, y_{frame}). Based on the deviation of the face center from the frame center, we calculate the motor angles for the activator response.

Activator Module: We use the pan-tilt servo motors to enable the camera to rotate both horizontally and vertically according to the user's face movements. We apply a Proportional Integral Derivative (PID) control loop for generating the panning and tilting rotations based on the face tracking results [2]. The aim is to adjust the camera's focus back to the center of the user's face. The PID controller calculates the difference between the video frame center and the current face center in the frame as an error term and consistently compensates for the error. 2 independent processes with their own PID controller calculate pan and tilt angles for the servo motors to rotate to. Additionally, we adjust the proportional gain, the integral gain and the derivative gain of the PID controller based on five users' experience to smooth the camera motions and reduce the intrusiveness.

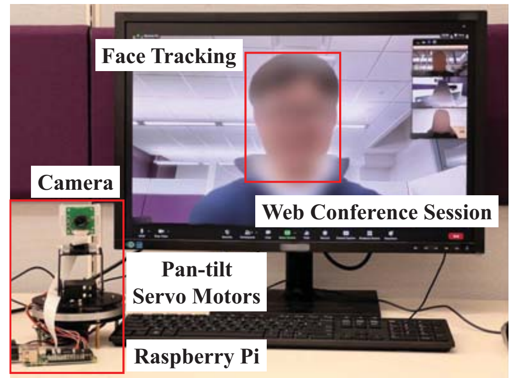

Figure 2: The implemented system in a web conference.

3 DEMONSTRATION

We implement the proposed motorized camera system on a Raspberry Pi with a regular RGB camera and pan-tilt servo motors as shown in Figure 2. Specifically, we use the Arducam for the Raspberry Pi to capture the user's face with a resolution of 5mp [1]. We use the Raspberry Pi 3 model with 64 Gb of memory to process video frames in real time. A separate PID process runs simultaneously with the video processing in the Raspberry Pi. Two servo motors are interfaced with the Raspberry Pi to rotate the camera for motion-based response. Furthermore, to reduce the security concern that giving more freedom to the camera may potentially expose unwanted areas of the user's environment, we constrained each servo's rotation within a narrow angle (e.g., between -25°to 25°). If the user moves out of this range, the camera system will be temporarily shut down.

For the demonstration, we use the above prototype to present the experimental validation of the proposed system that exploits human's peripheral nerves to improve the user's awareness of a working web camera. We show that motion can be a better way of making the user perceive the active status of the camera and that a motorized camera can achieve the goal when it is integrated with the face tracking technique for unobtrusive human-camera interaction. Participants can join our mimic Google Hangout meeting session and observe this novel motion-based camera indicator with their peripheral vision, hearing and even the body skin hair. They can experience passively receiving the indicator signals while focusing on a primary task on the screen without directly looking at the camera, which extends the virtual social interaction to the physical world. For comparison, four current camera indicators will also be presented, including the LED light indicator, system notification messages, software indicator icons and the user's own video displayed on the screen. Participants are able to experience the differences of these methods and choose which they prefer.

Acknowledgment. This work was partially supported by LEQSF(2020-23)-RD-A-11.

REFERENCES

[1] Arducam for raspberry pi camera module. https://www.amazon.com/Arducam-Camera-Raspberry-Interchangeable-M12x0-5/dp/B0867C9SZP?ref_=ast_sto_dp&th=1, 2021.

[2] Kiam Heong Ang, G. Chong, and Yun Li. Pid control system analysis, design, and technology. *IEEE Transactions on Control Systems Technology*, 13(4):559–576, 2005.

[3] O. Déniz, G. Bueno, J. Salido, and F. De la Torre. Face recognition using histograms of oriented gradients. *Pattern Recognition Letters*, 32(12):1598–1603, 2011.

[4] Steven L. Franconeri and Daniel J. Simons. Moving and looming stimuli capture attention. *Perception & Psychophysics*, 65(7):999–1010, October 2003.

[5] Steven L. Franconeri and Daniel J. Simons. The dynamic events that capture visual attention: A reply to Abrams and Christ (2005). *Perception & Psychophysics*, 67(6):962–966, August 2005.

[6] Nir Kshetri. Covid-19 meets big tech. *Computer*, 53(8):10–13, Aug 2020.

[7] Bogdan Kwolek. Face detection using convolutional neural networks and gabor filters. In Włodzisław Duch, Janusz Kacprzyk, Erkki Oja, and Sławomir Zadrożny, editors, *Artificial Neural Networks: Biological Inspirations – ICANN 2005*, pages 551–556, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[8] Rebecca S. Portnoff, Linda N. Lee, Serge Egelman, Pratyush Mishra, Derek Leung, and David Wagner. Somebody's watching me? assessing the effectiveness of webcam indicator lights. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 1649–1658, New York, NY, USA, 2015. Association for Computing Machinery.

[9] Anirudha B Shetty, Bhoomika, Deeksha, Jeevan Rebeiro, and Ramyashree. Facial recognition using haar cascade and lbp classifiers. *Global Transitions Proceedings*, 2(2):330–335, 2021. International Conference on Computing System and its Applications (ICCSA- 2021).

[10] Ian M Thornton, Ronald A Rensink, and Maggie Shiffrar. Active versus Passive Processing of Biological Motion. *Perception*, 31(7):837–853, July 2002.

[11] A. von Muhlenen, M. I. Rempel, and J. T. Enns. Unique Temporal Change Is the Key to Attentional Capture. *Psychological Science*, 16(12):979–986, December 2005.