Enhancing Gesture Perception in Learning for Prosthetic Vision with Emerging Event Cameras

Anran Meng
Beijing Technology and Business University
Beijing, China
2230702051@st.btbu.edu.cn

Fuwei Dong
Beijing Technology and Business University
Beijing, China
2230702065@st.btbu.edu.cn

Ming Li
Beijing Technology and Business University
Beijing, China
2230702018@st.btbu.edu.cn

Xiaoming Chen (Corresponding Author)
Beijing Technology and Business University
Beijing, China
xiaoming.chen@btbu.edu.cn

***Abstract*—Gestures are powerful tools for enhancing student learning and facilitating human-computer interaction in digital learning environments. However, students with blindness or severe vision impairment are unable to benefit from gestures, resulting in inequitable learning conditions where they rely solely on passive listening. To address this issue, we propose the utilization of event cameras, which are novel image sensors capable of capturing the "essence" of motion, including gestures. By harnessing the power of event cameras, we can focus on the motion dynamics of gestures that are often inadequately represented in traditional videos. This potentially enables visually impaired students to have a better understanding and improved interaction with gesture-based content. In our approach, we leverage prosthetic vision to enable gesture perception for visually impaired students in learning. Our approach consists of several steps, including gesture event capture, event denoising, and prosthetic vision simulation, to facilitate accurate gesture recognition for effective learning. Through evaluation, we demonstrate the effectiveness of utilizing event information to assist gesture perception for visually impaired students. This research opens up new possibilities for inclusive education by utilizing emerging event cameras to provide equitable learning experiences for all students, regardless of their visual impairment.**

***Index Terms*—Gesture, prosthetic vision, event camera**

I. Introduction

Physical movements of instructors, such as gestures or facial expressions, can effectively enhance student learning. Among them, hand gestures play a vital role in human communication and human-computer interaction in digital learning environment. As such, gestures are often intertwined with spoken language, and the combination of gestures and speech has beneficial effects on comprehension. However, students with blindness or severely impaired vision are unable to access such benefits - they often can only listen passively to lectures without visually acquiring supplementary information.

Fortunately, prosthetic vision [1] provides an opportunity to mitigate these deficits. Visual prosthetics involve implanting multi-electrode arrays into the retina or visual pathway to apply artificial electrical stimulation, exciting nerve cells to simulate light perception. An illustrative example of this process can be seen in Fig. 1. With proper coordination, stimulus pulses can facilitate the attainment of partial vision for individuals with blindness or severe visual impairments. However contemporary electrode matrices have relatively low resolution. For example, modern implants boast approximately 1000 channels, equivalent to 32×32 pixels. Such low-resolution displays necessitate effective key information concentration, e.g., concentrating on the motion dynamics in the case of gestures. Notably, traditional video cameras retain immense information that could overwhelm low-acuity prosthetics. Event cameras [2], in contrast, are a novel type of image sensors that capture sparse and asynchronous visual data, with a primary emphasis on capturing motion dynamics. Compared to traditional video cameras, event cameras have the ability to detect and record pixel-level light changes independently, generating events with precise timestamps. These events encapsulate valuable information about alterations in light intensity, thereby offering detailed insights into the direction and intensity of motion. Fig. 2 showcases some captured images by event cameras as examples. The distinctive imaging principle of event cameras enables them to effectively capture the essence and intricacies of motion, making them highly promising for capturing and visualizing gesture motion specifically for visually impaired students [3]. Leveraging the accurate timestamped events, we can effectively capture the dynamic elements inherent in gestures, often insufficiently captured by traditional video cameras. Consequently, visually impaired students can gain a better understanding and enhanced perception of gesture-based content, facilitating their learning process.

Specifically, we propose in this paper a framework that utilizes emerging event cameras to capture motion dynamics and enhance motion perception, specifically focusing on hand gesture perception and learning for visually impaired students with prosthetic vision. We conducted an evaluation to provide evidence of the effectiveness of our approach. To facilitate

This work was supported in part by National Natural Science Foundation of China under Grant 62177001 and Beijing Natural Science Foundation under Grant 4222003.

a more accessible evaluation process, we utilized a prosthetic simulator to simulate the visual perception of visually impaired users and demonstrated the positive impact of the events captured by an event camera in improving gesture perception.

(a) The captured praising (thumbs up) gesture (b) Simulated prosthetic vision based on the captured regular video (c) Simulated prosthetic vision based on the captured event video

Fig. 1. Prosthetic vision simulated with pulse2percept for the praising (thumbs up) gesture

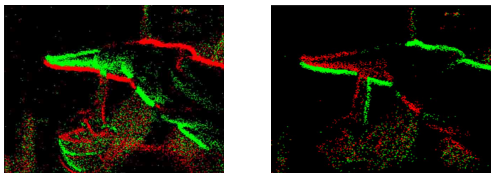
II. Related Work

A. The Role of Gestures in Education

Gestures are commonly used to express emotions and intentions and are a common nonverbal communication method closely related to speech. Gestures are considered complex hand movements and are equally important as other language patterns. They can not only convey experiential concepts but also express abstract concepts. This versatility makes gestures a rich and flexible nonverbal communication tool that can encompass a wide range of concepts and emotional expressions [4]. The combination of gestures and speech can convey meanings that are difficult to capture in a language alone, as they share a common relationship at the neural level. Therefore, the gestures expressed by instructors in classrooms play a very important role in promoting efficient and in-depth learning for students [5] [6].

B. Prosthetic Vision

For students with blindness or severe visual impairment, prosthetic vision can offer partial vision capabilities. Many ongoing research efforts focus on developing retinal prostheses that deliver electrical pulses to the retina [7]. Visual prostheses can be integrated at various positions along the visual pathways, including the retina, optic nerve, lateral geniculate body, or visual cortex [8]. These prosthesis aim to provide electrical stimulation to the appropriate neural structures to elicit visual perceptions. As technology continues to advance, we anticipate



(a) Captured event image (b) Denoised event image

Fig. 2. Sample images of captured event videos (before and after event denoising with EventZoom [11]): Positive events (signifying brightness increment) are shown in green while negative events (signifying brightness decrement) are shown in red.

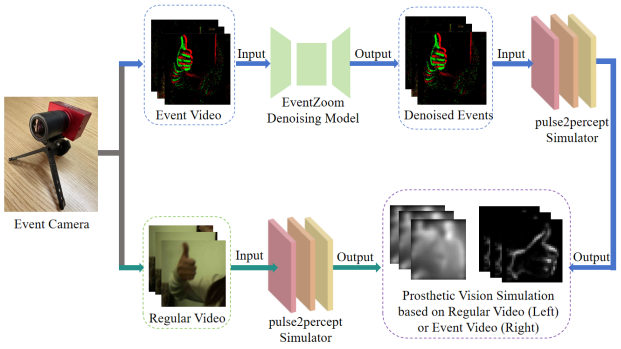


Fig. 3. Flowchart for our methodology and evaluation

that prosthetic vision will become increasingly technologically mature and widely adopted. This progress holds promise for enhancing the visual experiences and educational opportunities of individuals with visual impairments.

C. Event Video and Denoising

Due to its unique working principle, an event camera generates streams of time and pixel position-related events instead of traditional images. The captured event streams allow for detailed analysis of dynamic scenes, making event cameras ideal for high-speed environments and real-time applications. Previous research [9] has shown that event cameras excel at capturing motion, making them suitable for motion recognition. In this paper, we leverage the capabilities of event cameras to capture dynamic hand gestures and simulate prosthetic vision for visually impaired students, aiding them in recognizing gestures. However, event data captured by event cameras may be affected by sensor noise and environmental factors. Several methods have been proposed to reduce noise in event camera data, including filtering techniques [10]. These techniques involve adjusting triggering thresholds, applying spatial filters based on neighboring information, and utilizing deep learning methods, particularly convolutional neural networks (CNNs), to automatically remove noise.

III. Methodology

A. Pipeline

The flowchart of our methodology and evaluation process is illustrated in Fig. 3. In our design, we capture five distinct hand gestures including applauding, waving, counting, affirming, and praising (thumbs up). These gestures were recorded using a Davis346 event camera capable of outputting both regular videos and event videos. To enhance the quality of the captured event video, we applied EventZoom filtering [11], resulting in denoised event images as demonstrated in Fig. 2. The denoised event images of the event video, along with the frames of the regular video, are then used as inputs in a prosthetic vision simulator to simulate the visual experience of visually impaired students for comparison purpose.

B. Prosthetic Vision Simulation

For prosthetic vision simulation, Pulse2Percept [12] serves as an open-source simulator designed to predict the per-

ceptual experience of retinal prosthesis users across various implantation configurations. The simulator offers a modular and scalable user interface, enabling seamless simulation of new implants, stimuli, and retinal models. By utilizing Pulse2Percept, we can flexibly adjust implant configurations and simulate the perceptual experiences of visually impaired students in different contexts. In this work, we utilized the Pulse2Percept simulator to replicate and compare the impaired visual perception based on the captured regular video and the captured event video. For example, Fig. 1 showcases Pulse2Percept-simulated vision of the "praise" (thumbs up) gesture as perceived by visually impaired students, based on the captured regular video or event video, respectively. Notably, the event video, capturing the essential motion of the gesture, renders a clear depiction, while the regular video results in a blurry representation of the gesture. By employing Pulse2Percept, we were also able to recruit participants with normal vision as "virtual patients" [13] following established research protocols.

IV. Evaluation

In our evaluation, a user study was conducted with a total of 35 participants, comprising 30 students, 2 faculty members, and 3 professionals. The gender distribution was 15 males and 20 females, with an average age of 27 and a standard deviation of 3.07. Participants were randomly assigned to two groups: the regular video group and the event video group. Using the Pulse2Percept simulator, each group was able to perceive the prosthetic vision of the five captured gestures. These prosthetic vision simulations were based on either regular video or event video inputs. After observing the simulated prosthetic vision, participants were then asked to select one of 5 choices to determine the meaning conveyed by each gesture. This task allowed for the evaluation and comparison of the effectiveness of regular video and event video in conveying gesture meaning through prosthetic vision.

TABLE I
Accuracy of gesture recognition for prosthetic vision simulated based on regular video and event video

	Gesture Recognition Accuracy [%]	
Gestures	Prosthetic vision based on regular video	Prosthetic vision based on event video
Applauding	91.43	88.57
Waving	94.29	100
Counting	88.57	97.14
Affirming	82.86	94.29
Praising	88.57	94.29
Average	**89.14**	**94.86**

The results of our study, as presented in Table IV, indicate that the prosthetic vision generated from the event video showed higher accuracy in gesture recognition compared to the regular video. For four out of the five gestures, the prosthetic vision generated from the event video exhibited improved gesture recognition accuracy. On average, there was a 5.72% increase in gesture recognition accuracy. The only exception was observed for the applauding gesture, where the prosthetic vision generated from the regular video demonstrated a slightly higher accuracy in gesture recognition compared to the event video. We believe this discrepancy is due to the repetitive applauding motion, which can be easily recognized in traditional videos. However, for more challenging gestures, the prosthetic vision generated from event videos consistently showed improved recognition accuracy. These findings suggest that utilizing event videos can enhance gesture recognition accuracy in prosthetic vision, particularly for more complicated gestures.

V. Discussion and conclusion

In this study, we aimed to capture hand gestures using emerging event cameras. Denoising techniques were specifically applied to the event video to enhance the quality of captured events. To assess gesture recognition during learning for visually impaired students, we employed a prosthetic vision simulator that mimicked the visual experience of them. Then our user study involved 30 participants using the prosthetic vision simulator. Our results demonstrated that event video, which better captured the essence of gesture motion, significantly improved gesture recognition for visually impaired students compared to traditional regular video. This represents a notable departure from their previous reliance solely on auditory cues. The findings underscore the potential of utilizing emerging event cameras to assist visually impaired students in various learning tasks, exemplified by the success in gesture learning demonstrated in this study.

Moving forward, our future research will explore the broader application of event cameras in other learning tasks. By expanding our understanding of the capabilities and benefits of event cameras, we aim to further enhance the overall learning experience and opportunities for visually impaired students.

References

[1] Allen, Penelope J., and Lauren N. Ayton. "Development and Experimental Basis for the Future of Prosthetic Vision", *Macular Surgery: Current Practice and Trends* (2020): 449-462.

[2] G. Gallego et al., "Event-Based Vision: A Survey," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 154-180, 1 Jan. 2022.

[3] Rueckauer, Bodo, and Marcel van Gerven. "Experiencing prosthetic vision with event-based sensors", *Proceedings of the International Conference on Neuromorphic Systems*, 2022.

[4] Anu, S., R. Babitha , and N. Muthukumaravel . "Role of speech associated gestures of the teacher in Medical education - a comparative study", *National Journal of Clinical Anatomy* 2.3(2013): 150.

[5] Abels, and Simone. "The role of gestures in a teacher–student-discourse about atoms", *Chemistry Education Research & Practice* (2016):10.1039.C6RP00026F.

[6] Lewis, Tasha Noel. "The role of motion event gestures in L2 development in a study abroad context", *Dissertations & Theses - Gradworks* (2009).

[7] Xia, Xuan, et al. "Semantic translation of face image with limited pixels for simulated prosthetic vision", *Information Sciences (2022)*.

[8] Memon, Muhammad, and J. F. Rizzo. "The Development of Visual Prosthetic Devices to Restore Vision to the Blind", *Neuromodulation* 2.1(2009):723-742.

[9] Pan, Jing Samantha, and G. P. Bingham. "With an Eye to Low Vision: Optic Flow Enables Perception Despite Image Blur", *Optometry and vision science*, 90.10(2013): 1119-1127.

[10] Mohamed, Sherif AS, et al. "DBA-Filter: A dynamic background activity noise filtering algorithm for event cameras", *Intelligent Computing: Proceedings of the 2021 Computing Conference*, Volume 1. Springer International Publishing, 2022.

[11] P. Duan, Z. W. Wang, X. Zhou, Y. Ma and B. Shi, "EventZoom: Learning to Denoise and Super Resolve Neuromorphic Events", *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2021, pp. 12819-12828.

[12] Michael, Beyeler, et al. "pulse2percept: A Python-based simulation framework for bionic vision", *Python in Science* 2017.

[13] Nicole Han, Sudhanshu Srivastava, Aiwen Xu, Devi Klein, and Michael Beyeler. "Deep Learning–Based Scene Simplification for Bionic Vision", In *Proceedings of the Augmented Humans International Conference 2021 (AHs '21)*, ACM, New York, NY, USA, 45–54.