EXController: Enhancing Interaction Capability for VR Handheld Controllers using Real-Time Vision Sensing

Junjian Zhang
Digital Nature Group
University of Tsukuba
tyookk@digitalnature.slis.tsukuba.ac.jp

Yaohao Chen
Digital Nature Group
University of Tsukuba
yaohao.chen@digitalnature.slis.tsukuba.ac.jp

Satoshi Hashizume
Digital Nature Group
University of Tsukuba
hashizume@digitalnature.slis.tsukuba.ac.jp

Naoya Muramatsu
Digital Nature Group
University of Tsukuba
naoya.muramatsu@digitalnature.slis.tsukuba.ac.jp

Kotaro Omomo
Digital Nature Group
University of Tsukuba
okkotaro@digitalnature.slis.tsukuba.ac.jp

Riku Iwasaki
Digital Nature Group
University of Tsukuba
riku@digitalnature.slis.tsukuba.ac.jp

Kaji Wataru
Digital Nature Group
University of Tsukuba
wkj@digitalnature.slis.tsukuba.ac.jp

Yoichi Ochiai
Digital Nature Group
University of Tsukuba
PixieDustTechnologies,Inc.
wizard@slis.tsukuba.ac.jp

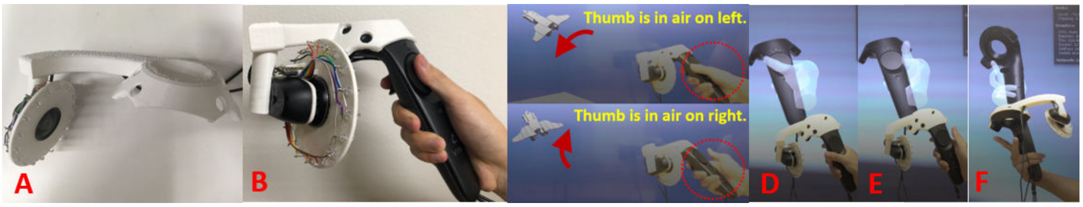

Figure 1: (A) Our prototype includes a NIR camera sensor. (B) While holding the Vive controller, the system is used to predict finger postures in real-time. EXController provides extra input from finger postures that can interact with VR content. (C) is an application that provides thumb postures in air to control an object's movements in the air. We developed a real-time Unity demo that maps the postures to matched finger animations (D, E, F), which is able to enhance "hand presence" while holding the controller.

ABSTRACT

This paper presents EXController, a new controller-mounted finger posture recognition device specially designed for VR handheld controllers. We seek to provide additional input through real-time vision sensing by attaching a near infrared (NIR) camera onto the controller. We designed and implemented an exploratory prototype with a HTC Vive controller. The NIR camera is modified from a traditional webcam and applied with a data-driven Convolutional Neural Network (CNN) classifier. We designed 12 different finger gestures and trained the CNN classifier with a dataset from 20 subjects, achieving an average accuracy of 86.17% $across - subjects$, and, approximately more than 92% on three of the finger postures, and more than 89% on the top-4 accuracy postures. We also developed a Unity demo that shows matched finger animations, running at approximately 27 fps in real-time.

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**; **Gestural input**;

KEYWORDS

Virtual reality, Gesture recognition, Handheld controller

ACM Reference format:
Junjian Zhang, Yaohao Chen, Satoshi Hashizume, Naoya Muramatsu, Kotaro Omomo, Riku Iwasaki, Kaji Wataru, and Yoichi Ochiai. 2018. EXController: Enhancing Interaction Capability for VR Handheld Controllers using Real-Time Vision Sensing. In *Proceedings of VRST 2018: 24th ACM Symposium on Virtual Reality Software and Technology, Tokyo, Japan, November 28-December 1, 2018 (VRST '18),* 2 pages.
https://doi.org/10.1145/3281505.3283385

1 INTRODUCTION

With recent rapid advances of head-mounted display (HMD) technology, virtual reality (VR) and augmented reality (AR) equipment systems have entered the consumer market. VR consumer HMD systems, such as HTC Vive, Oculus Rift, and PlayStation (PS) VR headsets, have handheld Vive, Touch, or Move controllers as primary interactional input devices. With these controllers, hand interaction is mainly based on sensing components such as buttons, joysticks and touch capacitive sensors, which can detect input events with

VRST '18, November 28-December 1, 2018, Tokyo, Japan
© 2018 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-6086-9/18/11.
https://doi.org/10.1145/3281505.3283385

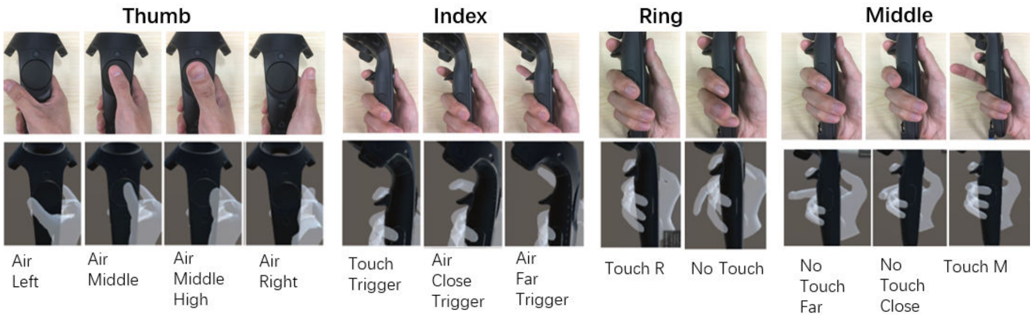

Figure 2: Finger postures recognized by our classifier applied to the NIR camera sensor

Table 1: EXController results compared to MobileNetV2.

	Thumb				Index			Ring		Middle			
	Air Left	Air Middle	Air Middle High	Air Right	Touch Trigger	Air Close Trigger	Air Far Trigger	Touch R	No Touch	No Touch Far	No Touch Close	Touch M	Mean
EXController no DA	91.64	77.96	87.29	92.24	93.52	69.96	86.78	81.12	79.40	87.27	71.22	88.20	83.88
EXController DA*	**94.08**	84.37	87.95	**92.24**	**96.33**	79.06	86.55	78.96	82.87	88.32	74.15	**89.22**	**86.17**
MobileNet2 no DA	98.64	89.34	89.09	95.38	97.07	83.61	92.46	81.68	91.36	89.37	70.95	93.66	89.38

* "DA" means the training is implemented with Data Augmentation.

100% accuracy and fast recognition speed. Oculus Touch controllers take advantage of an ingenious ergonomic design combined with touch capacitive sensing, providing users with an enhanced sense of "hand presence". It uses capacitive sensors to determine the discrete position of fingers, rendering a quasi-reconstructed hand model to display the hand posture. Recently Valve redesigned their controllers, now called Knuckles EV3, improving the performance of finger tracking using a combination of high-fidelity force and capacitive sensors. However, none of these controllers support any gesture interaction when fingers are in the air or not touching the sensor. Besides, current VR handheld controllers still hardly show the motion posture of individual fingers when not in contact with the sensor. Our aim is to extend and enhance the input capabilities for VR handheld controllers using vision recognition technology.

2 IMPLEMENTATION AND RESULTS

The modified traditional webcam is informed by Fanello et al.'s work [1]. The number of training images for each gesture of each subject is 1000, while there are 300 test images. Hence, our initial dataset includes a training set of 240,000 images and a test set of 72,000 images. Twelve different finger postures are shown in Figure 2. For a real-time gesture recognition system, we need to select a model that does not require high computational effort. In the Unity demo, we used the TensorFlowSharp library that currently does not support a graphics processing unit (GPU), consequently we implemented a LeNet-5-based CNN model. The finger segmentation is currently implemented by cropping the input image according to the position of pixels. After segmentation of the initial input image, each finger image is resized and passed to the classifier, which generates the predicted posture. The accuracy results in Table 1 show that with data augmentation, accuracy is improved by approximately 2.29%. MobileNetV2 classified the postures with higher accuracy and the top-4 accuracy results also improved noticeably. We found that subjects with straight thumbs (thumb profiles of a few subjects are not straight) resulted in better recognition of thumb postures. Subjects that have strong grip strength or a big hand size seem to achieve higher recognition rates.

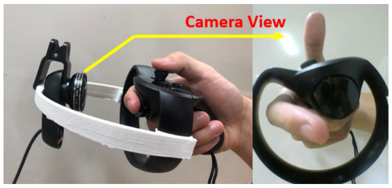

Figure 3: A tentative design for Touch controller.

3 LIMITATIONS AND FUTURE WORK

The main limitation of our work is the added weight of the controller. However, our approach is considered feasible for other VR controllers. As example, for other VR controllers such as Oculus Touch, even though the grip profile is completely different from the Vive controller, our approach could be applied as shown in Figure 3.

REFERENCES

[1] Sean Ryan Fanello, Cem Keskin, Shahram Izadi, Pushmeet Kohli, David Kim, David Sweeney, Antonio Criminisi, Jamie Shotton, Sing Bing Kang, and Tim Paek. 2014. Learning to Be a Depth Camera for Close-range Human Capture and Interaction. *ACM Trans. Graph.* 33, 4, Article 86 (July 2014), 11 pages. https://doi.org/10.1145/2601097.2601223