Upper-Body Posture Estimation Employing both Palm and Back of the User's Hand-Mounted Cameras

Satoshi Honda*
Graduate School of Science and Technology
Kwansei Gakuin University
Sanda, Japan

Yasuyuki Kono
kono@kwansei.ac.jp
Kwansei Gakuin University
Sanda, Japan

ABSTRACT

We propose a posture-estimation method that employs wearable devices. By attaching fisheye cameras to the palms and backs of both hands, the system captures the surrounding environment in all directions. The implementation of these features enables the simple and comfortable tracking of fingertip and upper-body motions that can be reflected in a virtual space. This paper describes the concept and implementation of the proposed system.

CCS CONCEPTS

• **Human-centered computing → Virtual reality**.

KEYWORDS

Fisheye/omnidirectional camera, finger posture estimation, head posture and position estimation

ACM Reference Format:
Satoshi Honda and Yasuyuki Kono. 2024. Upper-Body Posture Estimation Employing both Palm and Back of the User's Hand-Mounted Cameras. In *International Conference on Advanced Visual Interfaces 2024 (AVI 2024), June 3–7, 2024, Arenzano, Genoa, Italy.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3656650.3656712

Figure 1: Wearing fisheye cameras on the user's hand

1 INTRODUCTION

This paper proposes a system that estimates hand and upper-body postures in real time by attaching fisheye cameras to the palms and backs of both hands of a user. Communication and experience in virtual spaces, such as online conferencing and teleworking, have increased significantly in recent years. However, manipulating objects in a virtual space requires the user to use a controller that lacks immersion because it does not directly reflect real-world actions. The purpose of this research is to improve immersive experiences. The upper body is an important medium for a wide range of interpersonal communication, such as eye contact and gestures, and many studies on upper-body posture estimation have been conducted in the field of human-computer interaction. Ahuja et al. have enabled full-body posture estimation by attaching a wearable camera and a sphere-shaped mirror to the user's head[1]. However, wearing a device on the head limits their activity.

This research proposes a simple and user-friendly system that allows users to estimate posture using fisheye cameras attached to the palms and backs of both hands. Kuno et al. proposed a wearable device that measures skin deformation on the back of the hand using a reflective optical sensor [2]. They estimated hand posture using a CNN for interaction methods in virtual spaces and conducted experiments to evaluate the accuracy of the device. This method can only estimate hand postures and cannot interpret gestures made with the human body. Maruyama and Kono developed an AR pottery wheel-throwing system employing an HMD and omnidirectional cameras, each of which was attached to the center of the user's palm [3]. The omnidirectional cameras enable the estimation of the user's finger postures and the three-dimensional relative position and orientation between the user's hands and the virtual clay model on the wheel. Their system only estimated the relative position of the head and hand and could not estimate the posture of the upper body, including the arms. Our proposed system employs fisheye cameras attached to the user's hands to capture 360° videos of each hand and estimate the position and posture of the head and fingers without restricting the posture of the upper half of the body. Figure 1 shows the appearance of the proposed system. We believe that this research will enhance communication among multiple individuals through virtual objects without compromising immersion.

2 RESEARCH OUTLINE

Fisheye cameras wear attached to the palms and backs of both hands to enable posture estimation without restricting the activities of the upper body and fingers, and image processing technology was employed to enable upper-body posture estimation, including finger joint postures. An AR checkerboard was placed on

*Currently, Mitsubishi Electric Inc.

AVI 2024, June 3–7, 2024, Arenzano, Genoa, Italy

ACM ISBN 979-8-4007-1764-2/24/06
https://doi.org/10.1145/3656650.3656712

a desk to establish the user's position in the virtual space, and the user's position relative to the checkerboard in the physical space was then mirrored in the virtual space. The user's head in the fisheye image is normalized by rotating the image based on frame-by-frame rotation differences. The hand posture was estimated using the method described by Maruyama and Kono [3]. The facial region, which is a part of the fisheye image, has limited pixels and low resolution, which causes the feature points of the facial parts to deviate from the actual coordinates. This research employed both the histogram gradient of the face area and the distance between the human eyes to estimate the position and posture of the head. The head posture was determined by analyzing the facial features in the histogram of the face region. The distance from the camera to the head is calculated by consulting the average interocular distance (PD), which is 64mm plus or minus 5 mm for an adult Japanese, as reported by Taya [4]. By implementing these methods and reflecting the estimated hand and head postures in the virtual space, users can simply wear small wearable devices on their hands to replicate both their fingertip and body movements in the virtual space. The upper-body posture estimation method in this research requires users to wear camera devices attached to both the palms and backs of their hands, and the blind spots of the camera on the back of each hand can be compensated for by the camera on the palm.

3 ESTIMATION OF HEAD POSITION AND POSTURE

3.1 Head Tracking in Roll Direction

The aberration of the fisheye camera hinders the detection of face rotation in the roll direction by simply applying Haar-like feature detection functions included in the Dlib image-processing library. The rotational difference between two successive frames, in which a face is detected, is rotated and normalized in advance in the next frame to ensure that feature points can be consistently detected in consecutive frames.

3.2 Head Posture and Position Estimation

The histogram gradient and interocular distance were used to estimate head posture. Takeoka et al. conducted experiments on automatic face detection and face orientation recognition using images that include skin color information [5]. We employed histogram gradients to estimate the facial posture based on their research. The proposed method first determines the facial area using the tracking method described in Section 3.1. The cropped image is sliced into 15-px segments along the x- and y-axes. The positions of the pupil centers and eye contours are estimated by analyzing the gradient of the HSV color space histogram of the skin color pixels in each segmented image. The histograms near the nose show that the gradients of saturation (s) and brightness (v) of the skin color are small, and the histograms are clustered around the same value because the skin color is similar to that of the nose. The s and v values of the skin color near the chin were almost the same as those near the nose, and only a limited number of pixels in other areas had similar s and v values. These pixels can be removed by eliminating small skin-colored regions. The translation vector of the head from the camera is estimated by consulting the pupil positions, eye contours, and PD described in Section 2.

4 CONCLUDING REMARKS

In this research, a VR communication tool that is more versatile, immersive, and intuitive than existing research was developed. Fisheye cameras were attached to the back and palm of both hands, and image processing techniques were employed for upper-body posture estimation, as well as finger joint posture estimation. For the upper-body posture estimation process, an AR checkerboard was used to determine the user's position in the virtual space, head tracking was performed in the roll direction using rotation differences, and a face position and posture estimation method was proposed using histogram gradients and interocular distance. To demonstrate the superiority of the proposed method over existing methods that that rely solely on Haar-like features, we conducted an evaluation experiment involving seven subjects. The experimental results showed that the proposed method is more effective than the existing methods under certain conditions and that the proposed method is particularly effective when there are significant changes in head direction.

The appearance of the system, as shown in Figure 1, is somewhat unrealistic. This is because the prototype system employs "commercial" camera devices, each of which is RICOH THETA. We believe that the appearance is significantly realistic when the devices are embedded in the skin of the hand. Creatures or monsters that have an eye on each parm, such as the Japanese ancient mythical nymph or monster "Tenome" or Pale Man in the movie "Pan's Labyrinth," directed by Guillermo del Toro, have frequently appeared in legends or literature. Incorporating elements reminiscent of such creatures into our system enhances its functions and enriches our lives.

REFERENCES

[1] Karan Ahuja, Chris Harrison, Mayank Goel, and Robert Xiao. 2019. MeCap: Whole-body digitization for low-cost VR/AR headsets. *Proc. UIST'19 (ACM Symposium on User Interface Software and Technology)*, pp.453-462. DOI: https://doi.org/10.1145/3332165.3347889

[2] Wakaba Kuno, Yuta Sugiura, Nao Asano, Wataru Kawan and Maki Sugimoto. 2017. Estimation of 3D finger postures with wearable device measuring skin deformation on back of hand. *Proc. ICAT-EGVE'17 (International Conference on Artificial Reality and Telexistence & Eurographics Symposium on Virtual Environments)*, pp.41-42. DOI: https://doi.org/10.2312/egve.20171384

[3] Yusuke Maruyama and Yasuyuki Kono. 2019. AR pottery wheel-throwing by attaching omnidirectional camera to the center of the user's palms, *Proc. AH2019 (10th Augmented Human International Conference)*, Article No.47, DOI: https://doi.org/10.1145/3311823.3311856

[4] Shuichiro Taya. 2017. Individual differences in visual perception and visual cognition produced by differences in binocular distance, *Proc. Grant-in-Aid for Scientific Research in Japan* (in Japanese).

[5] Saori Takeoka, Masahiro Ozaki, Hiromi Kawada, and Yoshinori Adachi. 2003. A basic experiment of face image detection and face orientation recognition for learner recognition, *Bulletin of Nagoya Women's University, Faculty of Humanities and Society*, 49, pp.129-136 (in Japanse).