Drone Control for Monitoring a Walking Person from Constant Distance

Hiroto Yamashita, Takashi Morimoto, Ikuhisa Mitsugami

Abstract We propose a drone system that automatically follows a walking person keeping capturing him/her from a certain distance. In this system, the person is detected from images captured by the drone camera using OpenPose, the position and orientation of the drone are controlled so as to make his/her size in the images constant. We experimentally controlled the drone to follow a person constantly from behind him and confirmed it runs with reasonable response time and positional stability.

1 Introduction

Human video analysis in the sports field has been one of the active research interests in the computer vision field. It can be applied not only to the analysis of player's behaviors but also to the production of visual contents that makes it more comprehensible and attractive for the audience and TV viewers. Most of such analysis, however, usually cope with just some kinds of sports such as gymnastics, figure skating, and football, where their players move within such a limited area. It is, however, difficult to capture and analyze marathon runners or race-walkers, who usually run or walk in too long ways to capture by stationary (even pan-tilt-zoom) cameras.

This limitation of those existing systems motivates us. For overcoming the limitation, we propose a new drone system that can follow a person walking in a wide area from a constant distance. The proposed system consists of a drone and a laptop PC. From images captured by the drone camera, the system detects a person and

Hiroto Yamashita
Hiroshima City University, e-mail: yamashita@sys.info.hiroshima-cu.ac.jp

Takashi Morimoto
Hiroshima City University, e-mail: morimoto@sys.info.hiroshima-cu.ac.jp

Ikuhisa Mitsugami
Hiroshima City University, e-mail: mitsugami@hiroshima-cu.ac.jp

estimates his/her 2D pose using OpenPose. PID control is then performed so that the neck and middle-hip points of the estimated human pose should be at their reference positions defined in the drone image coordinate. We conducted experiments to investigate the accuracy on measured values of pose information used when following a person, and the control performance when the person walks.

2 Related Works

Commercial drones equipped with an automatic tracking function [1, 2] are already becoming popular. There are also several drone systems that automatically follow one or more pedestrians [3, 4]. Those systems, however, follow a person with quite a large distance, and so they are not suitable for monitoring details of human behaviors.

Pestana et al. [5] have proposed a method that uses a drone that detects a dynamic object in a captured image and follows it while keeping a certain distance. Mel et al. [6] also have developed a system in which drones automatically follow the walking path of a person. It also follows the person within a certain range so that the person can always be observed in order to predict the walking path. In those studies, however, the drones detect a person as a bounding box and thus do not recognize the posture of the person. On the other hand, our system extracts a skeletal pose to track the person more smoothly. Huang et al. [7] also have proposed a drone system that can autonomously capture cinematic shots of action scenes. They present a method to extract human 3D pose estimation based on stereo camera and design a real-time camera planning strategy that fulfills the aesthetic objectives for filming. The system, however, utilizes the depth information of the stereo camera for pose estimation. On the other hand, our system performs visual-based following using only a monocular camera.

3 Proposed System

3.1 System Configuration

Fig. 1 shows the system configuration of the proposed system. This system consists of a drone Tello [8] and a PC. The drone and PC are connected via WiFi, and images captured by the drone camera are transmitted to the PC. The PC performs human detection and pose estimation on the captured images. Parameters for the drone control is then calculated according to the position the person detected in the image, and are transmitted to the drone.

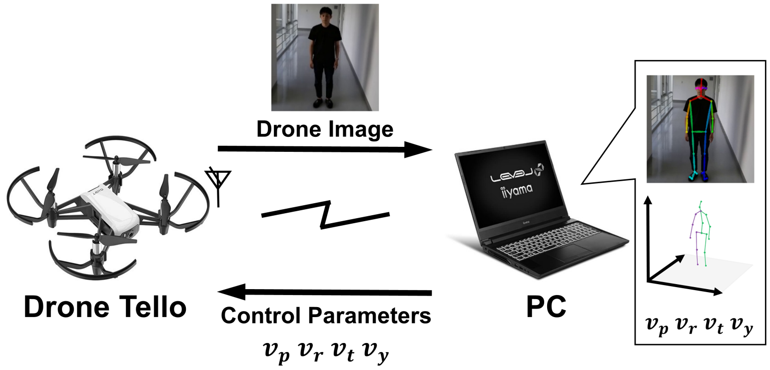

Fig. 1 Overview of the proposed system.

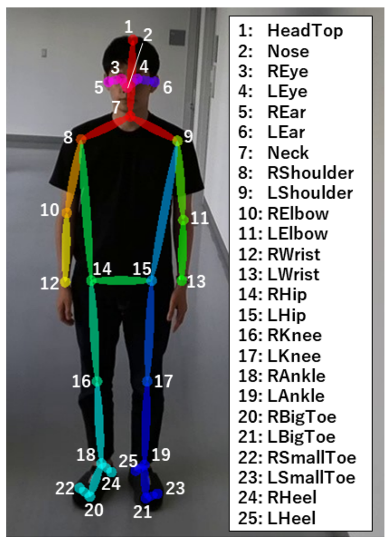

Fig. 2 25 joint points extracted by OpenPose.

3.2 Human Detection

For human detection, we apply CNN-based 2D pose estimation OpenPose [9] to images captured by the drone camera. It estimates 2D positions of 25 joint points (hands, feet, eyes, etc.) of a person in the image, as shown in Fig. 2.

3.3 Drone Control

The proposed system refers to the 2D joint points obtained by OpenPose for continuously locating the drone at a relatively constant distance from a person. In order for a drone to follow a person while keeping a certain distance, it is necessary to obtain the distance information between the drone and the person, but it is practically impossible to obtain metric distance information only from two-dimensional images. Though it is a possible way to give the metric information manually (e.g., the height of the person) to estimate the metric distance, it is troublesome that we need to ask every time a target person to input his/her height to the system when we consider the practical use. The proposed system thus does not use metric distance information but controls the distance constant by keeping the size of a person in the image constant.

The choice of the joint points for the references is important for stable drone control. The simplest idea is to use the head and ankle points for the references for keeping capturing the whole body. When capturing a pedestrian from his/her front or back, however, even while the pedestrian walks straight at a constant speed, the position of limbs moves up and down greatly in the image, so that their positions the images change largely. It is thus inappropriate to refer to those joint points. It is also an important condition for stable control that the joint points for the references have to be always detected stably. For example, the nose or hand points cannot be detected by self-shielding when observed from a certain direction. In consideration of the above requirements, our system adopts the Neck and MiddleHip (Midpoint between RightHip and LeftHip) as the reference points for the drone control.

The system controls the position and orientation of the drone so that those two joint points estimated at each frame should get close to the corresponding reference points. The reference points for Neck and MiddleHip are defined at $(480, 210)$ and $(480, 330)$, respectively, in the image coordinate (the image size is 960×720), as described in Fig. 3. We apply PID controls: input values are the longitudinal movement speed, vertical movement speed, and rotational angular speed of the drone.

4 Experimental Evaluation

We evaluated the performance of the proposed system from two viewpoints: estimation stability of the two joint points that are used as the references for the drone control, and performance of the PID control. In these experiments, the drone was intended to follow a walking person from behind him.

4.1 Stability of Estimated Joint Points

The drone control was performed using the Neck and MiddleHip points in the image coordinate for following a person maintaining a certain distance. We investigated

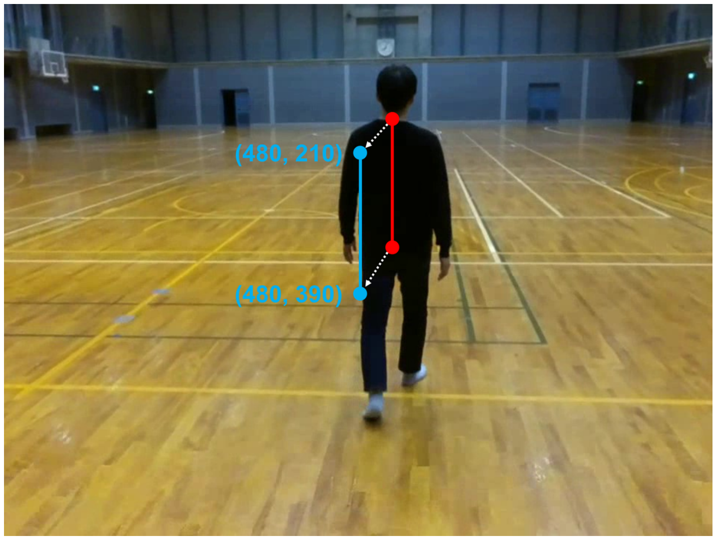

Fig. 3 The image captured by the drone when following a person. Red points show the estimated joint points of Neck and MiddleHip, and blue points show the corresponding reference points.

the stability of locations of these joint points estimated from images captured by the drone camera under walking conditions. We asked a participant to walk on a treadmill and located the drone at a certain distance behind him, so that the videos captured under the settings would be similar to those when the drone captures following a walking person. The distance between the drone and the subject, then, was set to 4 m. Fig. 4 shows an example of the captured images under the settings.

Fig. 5 shows changes in the length between the Neck and MiddleHip points in the image coordinate when the participant walked. Though there were small fluctuations, it can be regarded to be almost constant. For comparison, the figure also shows a curve the length between the head and right ankle points, which is much more fluctuated, which should be undesirable for the stable drone control.

4.2 Performance of Drone Control

To evaluate the control performance of the proposed system, we controlled the drone so as to follow a participant who walked and stopped in a straight line. In this experiment, he was standing for a while, started walking, and after a while, he stopped again. Fig. 7 shows the experimental settings.

Since it is difficult to measure the physical distance between the drone and the participant while walking, we evaluated the length of the Neck and MiddleHip points, instead. Fig. 6 shows its changes during the scenario above described. From this figure, it is confirmed that the length was kept almost constant except for moments

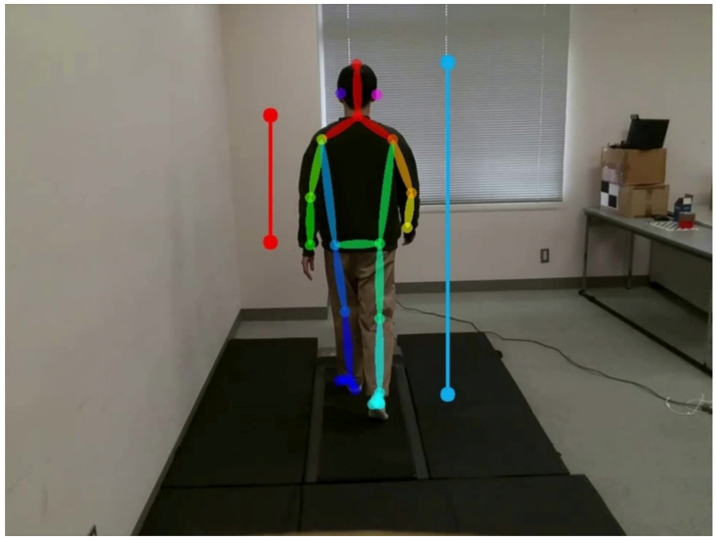

Fig. 4 An example of images captured by the drone camera. The red line denotes the distance between Neck and MiddleHip, and the blue line denotes that between Head and RightAnkle.

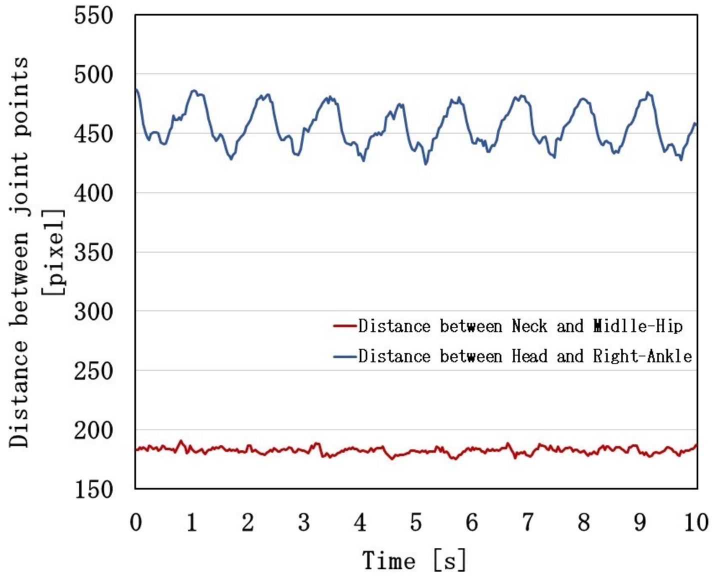

Fig. 5 Time-series of the distances.

when the person started walking or stopped it. But, even at those moments, it changed slightly. Through the whole scenario, the length was kept within a certain range.

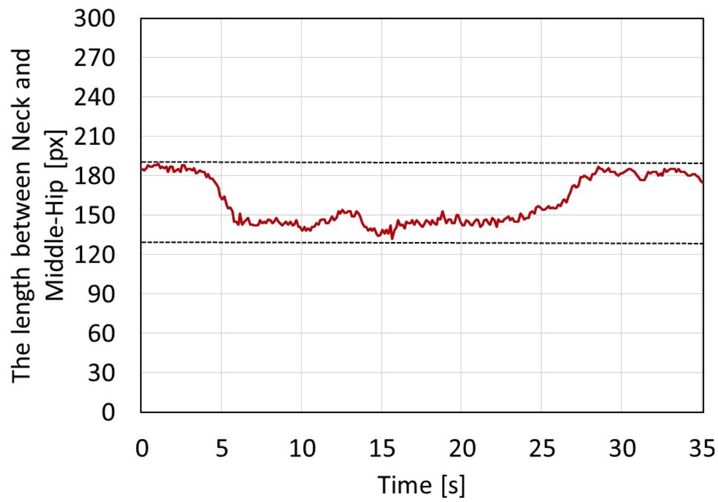

Fig. 6 Time-series of the distance between Neck and MiddleHip during the walking.

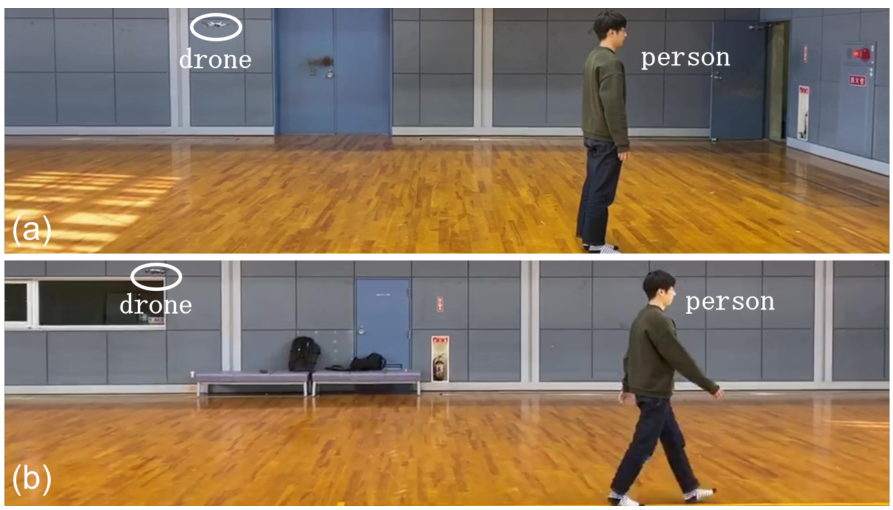

Fig. 7 Experimental settings of control performance evaluation: (a) A person is stopped, (b) A person is walking

5 Conclusion

In this paper, we proposed a new drone system that automatically follows a walking person keeping the constant distance from him/her, which should be useful for video analysis of a person who moves in a wide area. The proposed system detects a person and estimates his/her 2D pose from images captured by the drone camera. PID control is then applied so that the Neck and MiddleHip points should coincide with the reference points defined in the image coordinate to keep the drone position at a constant distance from the person. In the experiments, we investigated the stability of the estimated joint points using the treadmill and the fixed drone camera, which simulated the case where the drone followed while capturing a walking person. We also evaluated the performance of the drone control by actually controlling the drone to follow a walking person. Through those experiments, we confirmed that the drone had the ability to keep following a person even he/she repeatedly start or stop walking.

As future work, we need to improve the drone control performance. We have not yet tried to find the optimal parameters for the PID control. We expect more smooth and stable control should be achieved by finding the best parameters. It is also an important task to improve the system to recognize and reflect the rotations of the person. To do this, we are planning to apply a 3D pose estimation method that can recover 3D skeletal poses from the 2D poses estimated by OpenPose, like those proposed by Martinez et al. [10] or Zhou et al. [11]. Another difficult task is to keep locating at the front of the walking person. It is much more difficult than locating behind him/her because it requires an accurate and quick forecast of his/her future trajectory. For this problem, we plan to apply a gaze estimation method proposed by Murakami et al. [12, 13] that estimates the gaze direction from the head motion because the gaze direction should well indicate the direction he/she would go next.

Acknowledgements This work was supported by JSPS KAKENHI Grant Number JP18H03312.

References

1. MAVIC 2 PRO, https://www.dji.com/jp/mavic-2/
2. Skydio 2, https://www.skydio.com/
3. C. Bian, Z. Yang, T. Zhang, H. Xiong, “Pedestrian Tracking from an Unmanned Aerial Vehicle,” IEEE International Conference on Signal Processing, 2016.
4. B. Risse, M. Mangan, L. D. Pero, B. Webb, “Visual Tracking of Small Animals in Cluttered Natural Environments Using a Freely Moving Camera,” IEEE International Conference on Computer Vision, pp.2840-2849, 2017.
5. J. Pestana, J. L. Sanchez-Lopez, S. Saripalli, P. Campoy, “Computer Vision Based General Object Following for GPS-denied Multirotor Unmanned Vehicles,” American Control Conference, 2014.
6. D. D. Mel, K. A. Stol, J. A. D. Mills, B. R. Eastwood, “Vision-Based Object Path Following on a Quadcopter forGPS-Denied Environments,” International Conference on Unmanned Aircraft Systems, 2017.
7. C. Huang, F. Gao, J. Pan, Z. Yang, W. Qiu, P. Chen, X. Yang, “ACT: An Autonomous Drone Cinematography System for Action Scenes,” IEEE International Conference on Robotics and Automation, pp.7039-7046, 2018.
8. Tello, https://www.ryzerobotics.com/jp/tello
9. Z. Cao, T. Simon, S.-E. Wei, Y. Sheikh, “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” IEEE Conference on Computer Vision and Pattern Recognition, 2017.
10. J. Martinez, R. Hossain, J. Romero, and J. J. Little, “A simple yet effective baseline for 3d human pose estimation,” IEEE Conference on Computer Vision, pp.2659-2668, 2017.

11. X. Zhou, S. Liu, G. Pavlakos, V. Kumar, K. Daniilidis, "Human Motion Capture Using a Drone," IEEE International Conference on Robotics and Automation, pp.2027-2033, 2018.
12. J. Murakami, I. Mitsugami, "Gaze from Head: Gaze Estimation without Observing Eye," The 5th Asian Conference on Pattern Recognition, 2019.
13. J. Murakami, I. Mitsugami, "VR-based Eye and Head Motion Collection for Modeling Their Coordination," IEEE 8th Global Conference on Consumer Electronics, 2019.