Human Localization Using a Single Camera Towards Social Distance Monitoring During Sports

Ryosuke Hasegawa[1(✉)], Akira Uchiyama[1], Fumio Okura[1], Daigo Muramatsu[2], Issei Ogasawara[3], Hiromi Takahata[3], Ken Nakata[3], and Teruo Higashino[1]

[1] Graduate School of Information Science and Technology, Osaka University, Suita, Osaka 565-0871, Japan
r-hasegawa@ist.osaka-u.ac.jp
[2] The Faculty of Science and Technology, Seikei University, 3-3-1 Kichijoji-Kitamachi, Musashino, Tokyo 180-8633, Japan
[3] Graduate School of Medicine, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan

Abstract. Coronavirus disease 2019 (COVID-19) is still prevalent in the world. Social distancing is more important during exercise because we may not be able to wear masks to avoid breathing problems, heatstroke, etc. For supporting management of social distancing, we are developing a human localization system using a single camera especially for sports schools and gyms. We rely on a single camera because of the deployment cost. The system recognizes people from a video and estimates the human positions for supporting management of social distancing. The challenge is the error owing to pose variation during sports. In order to solve the problem, we adjust the height of the waist according to the pose of the legs. For evaluation, we collected 80 images with 5 kinds of poses. The results show that we successfully reduce the absolute position error by 23 cm on average.

Keywords: Social distancing · Human detection · Localization

1 Introduction

COVID-19 is still prevalent in the world. Social distancing is more important during sports because we may not be able to wear masks to avoid breathing problems, heatstroke, etc. Because vision-based human detection and tracking has been actively investigated since before the pandemic, vision-based systems have been developed for supporting management of social distancing. These systems detect skeletons [1] or bounding boxes [4] of humans to estimate interpersonal distance. However, position error may become large during sports because human poses change frequently.

T. Hara and H. Yamaguchi (Eds.): MobiQuitous 2021, LNICST 419, pp. 481–486, 2022.
https://doi.org/10.1007/978-3-030-94822-1_27

In order to solve the problem, we propose human localization based on skeletons. Our system uses a single camera for low deployment cost and detects skeletons of people by using *OpenPose-STAF* [3]. We select the waist position estimated by *OpenPose-STAF* to represent the position of the person for its stability in human detection. To improve position error owing to pose variation, we adjust the height of the waist according to the pose of the legs.

For evaluation, we collected 80 images with 5 kinds of pose. The results show that we successfully improve the absolute position error by 23 cm on average. In particular, the error was improved by 60 cm on average when the target was sitting on the ground where the waist height changes significantly.

2 Method

2.1 Overview

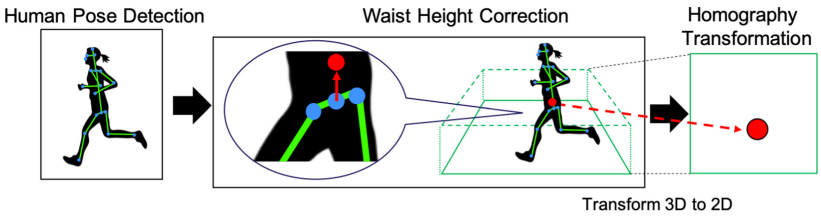

Fig. 1. Method overview

Figure 1 shows the overview of our method. We first detect persons by using the state-of-the-art skeleton detector called *OpenPose-STAF* [3]. We estimate the position of the detected person based on its skeleton and the coordinates of 4 points whose positions are given. We assume that a reference key point of the skeleton used for position estimation is at the same height as these four points. We choose the waist key point as the reference key point since *OpenPose-STAF* tends to detect the waist more reliable than the other body parts and the waist is close to the center of the body. However, the height of the waist changes depending on the pose. Therefore, we correct the height of the waist based on the key points of the legs.

2.2 Homography Transformation

For each frame, we estimate the position of the person whose skeleton is detected by using *OpenPose-STAF*. For localization, we use the homography [2], which is transformation that projects a plane to another plane, given the 4 point correspondences between the two planes. Therefore, a homography transformation matrix can transform pixel coordinates in an image into the actual positions,

given the distance among 4 points in the real world. This means we need to measure the distance between these 4 points in advance.

When a coordinate in an image is (u, v)[pixel], the corresponding coordinate (x, y)[m] in the real world is obtained by the following equation.

$$(x, y) = H(u, v). \tag{1}$$

H is the homography transformation matrix represented by the following equation.

$$H = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & 1 \end{bmatrix}. \tag{2}$$

For each point with a given coordinate, we have two equations. Since H has 8 variables, we can solve H, given the actual positions of the 4 points in the image.

Our method uses the key point of the waist for the reference key point whose position is regarded as the position of the person. This is because the waist key point is stably detected even during movement compared with other key points such as legs. Therefore, the height of the 4 points for the homography transformation matrix is set to 0.9[m], which is the average waist height for adults.

2.3 Waist Height Correction

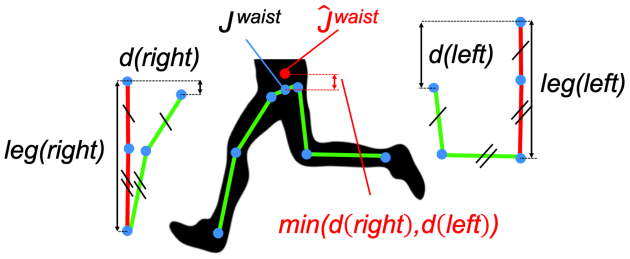

Fig. 2. Waist height correction

While walking and running, the height of the waist does not change largely. However, it can change greatly depending on poses such as sitting on a chair or the ground. Because the height error leads to the position error after the transformation, we mitigate the effect by mapping the position of the waist onto the plane with the height of 0.9[m]. The correction is performed before the homography transformation.

The overview of the correction is shown in Fig. 2. We let a coordinate of key point k be $J^k = [u^k, v^k]$. The length $l(p, q)$ between key points p and q is defined as:

$$l(p, q) = \sqrt{(u^p - u^q)^2 + (v^p - v^q)^2}. \tag{3}$$

For each leg, *OpenPose-STAF* outputs three key points which are the hip, the knee, and the ankle. The length $|leg|$ of the leg is obtained by combining the lengths between these joints as follows.

$$|leg| = l(hip, knee) + l(knee, ankle) \tag{4}$$

We call the difference between the ankle-to-hip height and $|leg|$ the *correction distance* d. The correction distance is defined as below.

$$d = |leg| - (v^{hip} - v^{ankle}). \tag{5}$$

If the leg angle against the ground becomes smaller, d becomes larger. This means the height of the reference key point (i.e. the waist) in the image is less than the assumed average waist height (i.e. 0.9[m]). Therefore, we correct the hip height by adding d to the original hip height. However, there are some cases where a leg is not on the ground because of jumping, balancing, etc. For the waist height correction, we need to use d of the grounded leg because d is calculated assuming the pose of the grounded leg lowers the waist height. If both legs are not on the ground, its duration is usually short. Therefore, we simply ignore such cases. On the other hand, when only one of the left and right legs is not on the ground, the vertical ankle-to-hip distance of the ungrounded leg becomes shorter than the grounded leg. In other words, d of the ungrounded leg is larger than the other since the lengths of the left and right legs should be almost the same. Therefore, we use either of the left or the right leg with the smaller correction distance. The coordinate of the waist $\hat{J}^{waist}$ after correction $\hat{v}^{waist}$ is given by:

$$\hat{v}^{waist} = v^{waist} + \min(d(left), d(right)). \tag{6}$$

We note that, if either of the legs is not detected, we do not perform the correction because we cannot determine detected leg is on the ground.

3 Evaluation

3.1 Evaluation Setting

For evaluation of human localization performance, we collected images from one subject. The subject was located at one of the lattice points in Fig. 3, and took 5 types of poses as shown in Fig. 4. The poses are standing, sitting (ground), sitting (chair), half-sitting, and crouching to evaluate the effect of the waist height correction. For each pose and position, we obtained images in which all key points of the lower body were detected.

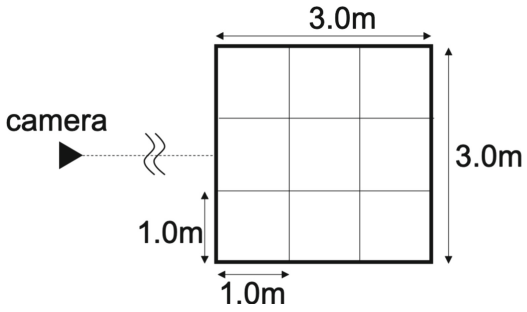

Fig. 3. Evaluation layout

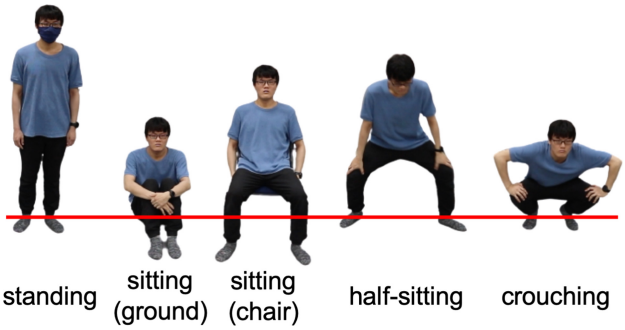

Fig. 4. Types of poses

3.2 Result

Table 1 shows the mean absolute error distance for each pose. From this result, we see that we can estimate the position of the standing person with low error. However, the error becomes larger as the waist height becomes closer to the ground. In addition, we succeeded to decrease the error by 23 cm on average by the waist height correction. However, we could not observe significant improvement for the pose of sitting on a chair. This is because the elevation angle of the camera and the angle of the leg are almost equal, which means the appearance of the leg length in the image is shorter than the actual length. In order to deal with this problem, we may need to obtain more accurate leg length by using a technique of estimating a 3D pose from a skeleton, for example.

Table 1. Mean absolute error for each pose [m]

Pose	Corrected	Original	Corrected - Original
Standing	0.056	0.064	−0.008
Sitting (ground)	0.729	1.338	−0.609
Sitting (chair)	0.716	0.779	−0.063
Half-sitting	0.370	0.641	−0.271
Crouching	0.712	0.915	−0.203
Average	0.517	0.747	−0.231

4 Conclusion

In this paper, we proposed human localization using a single camera during sports towards social distance monitoring. For evaluation, images with 5 poses were collected. As a result, we successfully reduced the absolute position error by 23 cm on average.

As our future work, we investigate a method using the upper body skeleton and/or the lower body skeleton in the previous frame for the waist height correction even when the lower body skeleton is not detected. Moreover, we are planning to use the proposed localization method for close-contact detection and tracking to quantify the risk of infection.

References

1. Aghaei, M., Bustreo, M., Wang, Y., Bailo, G., Morerio, P., Del Bue, A.: Single image human proxemics estimation for visual social distancing. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 2785–2795, January 2021
2. Capel, D., Zisserman, A.: Computer vision applied to super resolution. IEEE Signal Process. Mag. **20**(3), 75–86 (2003). https://doi.org/10.1109/MSP.2003.1203211
3. Raaj, Y., Idrees, H., Hidalgo, G., Sheikh, Y.: Efficient online multi-person 2D pose tracking with recurrent spatio-temporal affinity fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4620–4628 (2019)
4. Rezaei, M., Azarmi, M.: Deepsocial: social distancing monitoring and infection risk assessment in COVID-19 pandemic. Appl. Sci. **10**(21), 7514 (2020)