Human trajectory tracking using a single omnidirectional camera

Atsushi Kawasaki, Dao Huu Hung and Hideo Saito
Graduate School of Science and Technology
Keio University
3-14-1, Hiyoshi, Kohoku-Ku, Yokohama, Kanagawa, Japan
{kawasaki, hungdaohuu, saito}@hvrl.ics.keio.ac.jp

Abstract

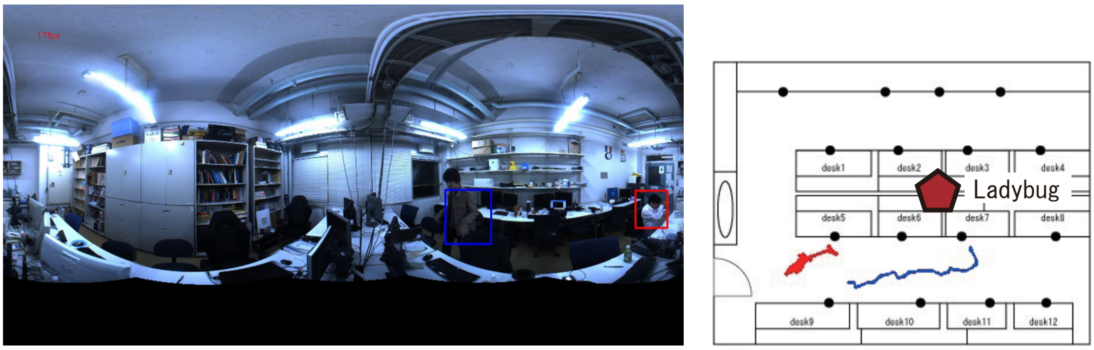
We propose a method to detect and track humans from an omnidirectional camera image, and to visualize human trajectories in the room plan. There are two problems to solve to achieve this issue. One is a robust algorithm of human detection in the case that humans are moving or sedentary in complicated background such as a office room. Dynamic background subtraction is suitable for detection in the complicated background but is not working to detect of a object having a little movement. In this paper, we propose a detection method based on the combination of static background subtraction, dynamic background subtraction, and Histogram of Oriented Gradients(HOG). The second problem is the way to visualize human trajectories in the room plan from distorted panoramic image. Therefore, we propose a method to create a correspondence relation by sandwiching a perspective image between a room plan and a panoramic image. The panoramic image is divided into multiple areas, and each area is converted into a perspective image. It is feasible to calculate the coordinate of human position in the room plan by using a Homography matrix between the perspective image and the room plan. We conducted the accuracy evaluation of human detection and human trajectory in order to ensure the effectiveness of this method. As a result, the proposed method of human detection reduced false positive detection remarkably in comparison with existing method. The experimental results of visualizing human trajectories demonstrated the range of errors of position estimation is about from 16 cm to 55 cm, but it is sufficient for use in data analysis such as head-count and residence time.

Keywords: Human detection, Human tracking, Omnidirectional camera

1 Introduction

A number of surveillance have widely been studied. Among them, tracking of human trajectory is extensively studied for use in the analysis of surveillance videos. The problem of single camera monitoring is limited visual field, while the problem of multi-camera monitoring is geometrical calibration of multiple cameras. An omnidirctional camera can cover 360° direction and enables to monitor broad range without overlapping. However, it is not easy to accurately understand the human position in the distorted panoramic image. This paper proposes a method which performs human detection and tracking from the omnidirectional camera image, and visualizes human trajectories in the room plan.

There are a lot of researches related to a surveillance camera. For example, Oktavianto et al.[1] proposed attendance logging system. Okabe et al.[2] tracked human trajectory using a stereo camera. The key issue in these research is human detection and tracking. Oktavianto et al.[1] proposed a method for human detection by background subtraction, but overlapped humans can not robustly be detected by just using background subtraction. Human detection using Histogram of Oriented Gradients(HOG)[3] is widely used. However, it is difficult to use



(a) (b)

Figure 1: The result of system.(a) shows the result of human detection. (b) shows the trajectories of detected humans.

HOG features in complex backgrounds because a complex background have much gradient information. Vondrick et al.[4] presented the results HOG features cause false positives in some cases. There are many research to approach this problem. Jin et al.[5] proposed a method of tracking in complex background by particle filter and HOG. Bing-bing et al.[6] proposed a detection method that combines HOG and dynamic background subtraction. This approach detects humans with HOG in the target area extracted by background subtraction. However, in this approach humans need to be moving because sedentary humans are included in the background image by using dynamic background subtraction.

Some researches on omnidirectional camera have been activity conducted in the past few years. Peri et al.[7] proposed a method to generate perspective and panoramic image from omnidirectional image obtained from a parabolic mirror omnidirectional camera. We used an omnidirectional multi-camera system, Ladybug, Point Grey.

In this paper, we propose a method to detect humans robustly in complex background based on the combination of static background subtraction, dynamic background subtraction and HOG. Furthermore, we propose a method to visualize human trajectories in room plan from the distorted panoramic image. Figure.1 shows an example of results of this system.

2 Overview of the system

This section provides an overview of the major steps in our system. Implementation details are presented in Section 3 through Section 5. The system is composed of the combination of on-line processing and off-line processing . In the off-line processing, We manually select corresponding points between the room plan and panoramic image. After dividing the panoramic image based on the corresponding points, each area is converted into a perspective image. The Homography matrix is then calculated between the perspective image and the room plan.

In the on-line processing, the system obtains a panoramic image as an input from the omnidirectional camera. The target areas are then extracted from a input image by background subtraction. In these areas, human is detected using HOG, and the background image is updated based on the detection result. After human detection, human tracking is performed by data association. The trajectory of each detected human is visualized in the room plan by using Homography between the perspective image and the room plan.

3 Human detection based on background subtraction and HOG

We propose the method to detect moving or sedentary humans in the complicated background. This method is based on combination of static background subtraction, dynamic background subtraction and HOG. Basically, by using Real AdaBoost classifier with HOG, upper bodies are detected in the target areas extracted by background subtraction because upper bodies are certainly visible even when lower body is hidden by other object such as desks. And then, the background image is updated based on the detection result and accumulation of background statistics[8]. In this approach, I is pixel intensity of the background area and is modeled as follow:

$$I = \overline{I} + \sigma \sin\left(2\pi\omega t\right) + k\zeta \tag{1}$$

where $\overline{I}$ is the time average of pixel intensity, σ is the amplitude of intensity, ω is the frequency of intensity, t is time, $k\,(-1 \leq k \leq 1)$ is coefficient, and ζ is the maximum value of the noise depending only on the camera.

Based on the detection result and accumulation of background statistics, one of the following three processes is performed in each pixel. One is in the case that I is not included in detected rectangles and $\overline{I} - \sigma - \zeta \leq I \leq \overline{I} + \sigma + \zeta$ is satisfied. In this case, the pixel is identified to exist in the background area and $\overline{I}$ and σ is updated based on follows:

$$\overline{I}' = \frac{(n-1)\,\overline{I} + I}{n} \tag{2}$$

$$\sigma' = \frac{(n-1)\,\sigma + \sqrt{2\left(I - \overline{I}\right)^2}}{n} \tag{3}$$

where n is the parameter of the update speed.

Another is in the case I is not included in the rectangles and $\overline{I} - \sigma - \zeta \leq I \leq \overline{I} + \sigma + \zeta$ is not satisfied. In this case, the pixel is identified to exist in the area of a moving object except humans, and $\overline{I}$ is not updated but σ is updated as:

$$\sigma' = \frac{(m-1)\,\sigma + \sqrt{2\left(I - \overline{I}\right)^2}}{m} \tag{4}$$

where $m\,(m \geq n)$ is the parameter of the update speed in object areas. By doing this, moving object other than human will be included in the background area gradually. Updating the background image over time prevents undetected error of the detector.

The other is in the case I is included in the rectangles. In this case, the pixel is identified to exist in human areas, and $\overline{I}$ and σ in this areas is not updated. By keeping these values, detected areas are not included in the background image and will always be the subtraction areas. From the empirical results n is $1/50.0$, m is $1/300.0$ and ζ is 10.0 in our implementation.

4 Human tracking by data association

In this system, human tracking is performed by data association that compares the Euclid distances and the distances of the color histograms between the detected rectangle in the current frame and the ones in the last frame. Data association is suitable for omnidirectional multi-camera system whose frame rate is low because data association is simple and does not take processing time. Firstly, the Euclid distance d is calculated. Bhattacharyya distance d_H is then calculated between histograms after making color histogram of each rectangle. Combined distance D is calculated in all combination of rectangles as follow:

$$D = d + \alpha d_H \tag{5}$$

where α is a parameter. When there is a one-to-one correspondence between the rectangle of current frame and the rectangle of the last frame, the combination of the minimal sum of D becomes the tracking result.

However, it is not sufficient to compare the histogram of only the previous frame to make an accurate tracking possible even if human is overlapping or occlusion occurs. In this paper we classify the K clusters from N histograms from N frame before to the current frame by using

K-means. Each cluster's center of gravity is adopted as representative histograms of each human. We use the smallest d_H of K representative histograms when calculating Bhattacharyya distances. From the empirical results N is 100 and K is 3 in our implementation.

5 Visualizing human trajectory

We consider distortion of the panoramic image when visualizing human trajectories. In this paper this problem is solved by converting perspective images from divided panoramic images.

In the off-line processing, we first select corresponding points between the room plan and the panoramic image. The area between a desk and the other desk is utilized because the floor is hidden by desks in the environment such as a office. (In the case that the aisle is broad like figure.4(b), we select corresponding points on the floor.) We create some perspective images which respectively contain an area constituted by every four input points. Assuming there is the sphere with a focus central on omnidirectional camera, Figure.2 shows a relationship between the panoramic image and the sphere. This relationship enables creating a perspective image by projecting the pixel intensity of the spherical surface on a plane including four input points. The Homography is calculated for each area utilizing the corresponding points between the room plan and the perspective image.

In the on-line processing, human trajectories are visualized using the Homography and the coordinate value of the detected rectangle. Assuming the lower side of the rectangle exists on a plane between a desk and the other desk, we can estimate the coordinate value of the human position in the room plan by multiplying the center of the lower side on the perspective image by the Homography. Figure.3 shows the flow of visualizing.

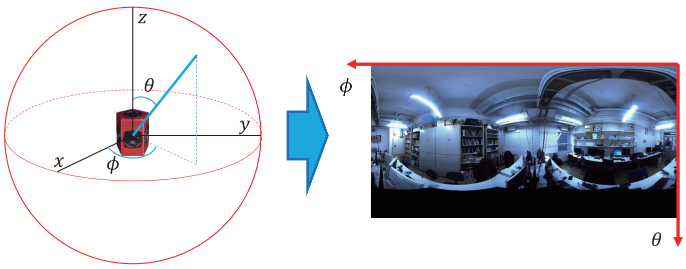


Figure 2: Assuming the sphere of the constant radius exists with a focus central on omnidirectional camera, pixel intensities are projected to the spherical surface. When the latitude is ϕ and the longitude is θ, we can make a panoramic image by using polar coordinate.

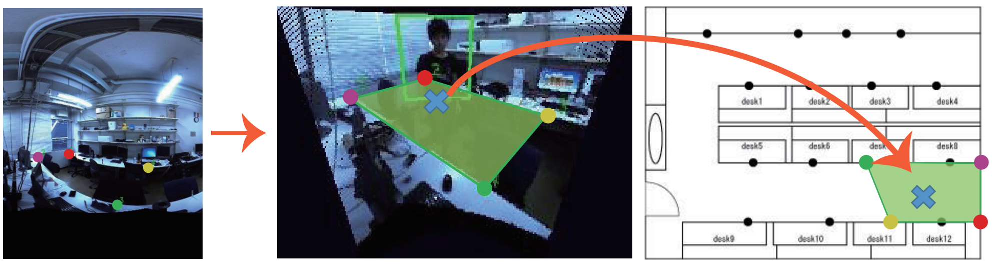


Figure 3: In off-line process, After inputing the corresponding points between the panoramic image(left image) and the room plan(right image), the perspective image(center image) is created. In on-line process, human position in the room plan is calculated by multiplying the center of the lower side of the rectangle on the perspective image by the Homography between the perspective image and the room plan.

6 Experiment

We conducted the experiments for evaluating the human detection and trajectory tracking. Figure.4 shows two environments of the experiments. Figure.4(a) is the environment where the aisle is narrow, and Figure.4(b) is one where the aisle is broad. Training images were cropped from panoramic image to minimize the loss of accuracy of detection by distortion.

(a) The room interior

(b) The room interior

Figure 4: The environment of the experiment

6.1 Evaluation of human detection

We compared our method with the combination method of static background subtraction and HOG. In this experiment, a video of 8800 frames was prepared. The ROC curve is given in Figure.5[a], in which the x-axis is False Positives Per Image(FPPI), and the y-axis is the detection rate. Figure.5[a] shows FPPI of our method is markedly lower than the comparative method. Since our method takes advantage of dynamic background subtraction, the false positive detection of our method is lower in subtraction area generated by movement of objects and illumination changes. However, the maximum value of the detection rate is also lower. The cause is possibly human areas included in the background image when human is not detected.

6.2 Evaluation of trajectory tracking

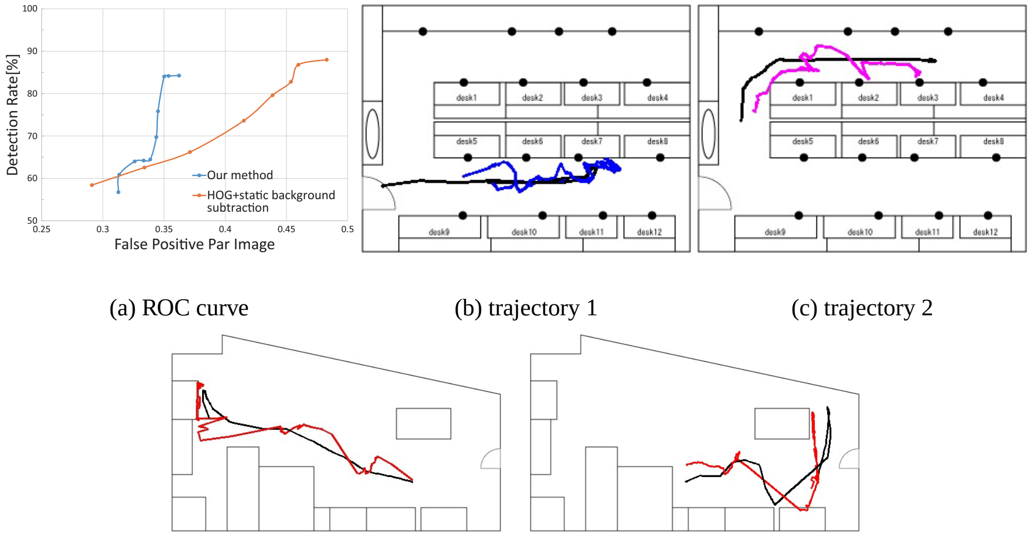
We measured the error of the Euclid distance between the visualizing result and the ground truth in all frame, and calculated the average, the standard deviation, and the maximum of the error. In this experiment, four videos of 200 frames were prepared. The ground truth is the coordinate on the room plan obtained by manually plotting human position on the room plan. The results of the trajectories are given in Figure.5 (b)-(e). The trajectories of black lines indicate the ground truth. Table.1 shows the evaluation index of the error of the trajectories. The range of distance error is about from 16 cm to 55 cm. The assumed cause of the distance error is as follows. Since the classifier can detect only upper bodies, the amount of information is not enough to decide human areas. When the rectangle can not surround a human accurately, the coordinate multiplied by the Homography is misaligned.

Table 1: The error of trajectory

	(b)	(c)	(d)	(e)
Average[cm]	16.50	36.75	50.91	55.31
Standard deviation[cm]	10.98	22.91	28.88	26.23
Maximum[cm]	47.41	110.29	141.61	135.79

7 Conclusion

In this paper, we presented the method to detect and track humans using the omnidirectional camera, and to visualize trajectories in the room plan. Our method which is based on combination of static background subtraction, dynamic background subtraction, and HOG, can detect moving or sedentary human in the complicated background. This method overcomes the issue that dynamic background subtraction is not working to detect a object having a little movement.



(a) ROC curve (b) trajectory 1 (c) trajectory 2

(d) trajectory 3 (e) trajectory 4

Figure 5: The result of the experiments

Human tracking is performed by data association which is suitable for this system because the frame rate of omnidirectional multi-camera system is low. On the other hand, in order to visualize trajectory from distorted panoramic image, we proposed the method that the panorama image is divided into multiple areas and each area is converted into a perspective image. As an experimental result, the proposed method of human detection reduced false positive detection remarkably. Furthermore, the result of visualizing human trajectories showed trajectory error of position estimation is about from 16 cm to 55 cm.

References

[1] H. Oktavianto, H. Gee-Sern, and C. Sheng-Luen. Image-based intelligent attendance logging system. *System Science and Engineering (ICSSE), 2012 International Conference on*, pages 1–6, 2012.

[2] A. Okabe, M. Ambai, and S. Ozawa. In-store human detection based on abs method for flow line analysis. *IEEJ Transactions on Electronics, Information and Systems*, 127:506–512, 2007.

[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 1:886–893, 2005.

[4] C. Vondrick, A. Khosla, T. Malisiewicz, and A. Torralba. Hog-gles: Visualizing object detection features. 2013.

[5] L. Jin, J. Cheng, and H. Huang. Human tracking in the complicated background by particle filter using color-histogram and hog. pages 1–4, 2010.

[6] W. Bing-bing, C. Zhi-xin, W. Jia, and Z. Liquan. Pedestrian detection based on the combination of hog and background subtraction method. pages 527–531, 2011.

[7] V. Peri and Shree K N. Generation of perspective and panoramic video from omnidirectional video. 1:243–245, 1997.

[8] S. Morita, K. Yamazawa, and N. Yokoya. Networked video surveillance using multiple omnidirectional cameras. 3:1245–1250, 2003.