魚眼カメラによる職場作業者認識および位置測定

People Recognition and Position Measurement in Workplace by Fisheye Camera

関　海克*　　新西　誠人**
Haike GUAN　　Makoto SHINNISHI

要　旨

オフィスや工場など職場において，作業者の位置測定と可視化は業務改善の重要な課題である．作業者の位置を可視化することで，生産効率向上，作業ミス低減，事故防止効果が得られる．本稿は一枚の360度魚眼カメラ画像による高精度な作業者認識，3次元位置測定，作業者位置可視化方法を提案する．

本提案手法は魚眼画像を単位球面へ投影，透視投影を行うことで，複数方向への歪み補正画像を作成する．職場地面に複数のマーカーを設けることで，歪み補正画像と3次元実空間のキャリブレーションを行い，変換式を求める．複数方向の歪み補正画像から機械学習による作業者認識を行い，求めた変換式を用いて，3次元空間での作業者位置測定を行う．

Ricoh R 360度魚眼カメラにより作業者認識，3次元位置測定実験を行い，本提案手法の有効性を確認した．

ABSTRACT

In a workplace such as a factory or office, it is important to visualize workers' positions to improve efficiency and avoid errors and accidents during work operations.

We propose a method for measuring the 3D positions of people in a workplace using only one fisheye camera. One 360-degree fisheye image is projected onto a unit sphere and several perspective projection images are generated to correct distortions in the image obtained from the fisheye camera. A machine learning method is applied to the corrected images for people recognition, and the results are used to calculate 3D positions in the workplace. Only a few markers need to be placed on the floor of the workplace to calibrate the perspective projection images with the plane of floor.

People recognition and position measurement experiments were conducted in a workplace using Ricoh R 360-degree monocular fisheye cameras. The evaluation results demonstrate the effectiveness of our proposed method.

*　イノベーション本部　AI応用研究センター
Applied AI Research & Development Center, Innovation/R&D Division

**　リコー経済社会研究所
Ricoh Institute of Sustainability and Business

本稿は，IS&T Electronic Imaging学会に帰属の著作権の利用許諾を受け，IS&T Electronic Imaging学会誌*IS&T International Symposium on Electronic Imaging 2019, Intelligent Robotics and Industrial Applications using Computer Vision*, pp. 459-1–459-7 (2019)に掲載した論文を基に作成した．

1. Introduction

In a workplace such as a factory or office, visualizing workers' positions is important for improving efficiency and avoiding mistakes and accidents. The 3D positions can also be used for robot navigation and cooperation between robot and people.

Conventional methods use a monocular fisheye camera to find and track people in a workplace. In these methods, image distortion is first corrected by calibration. Then the people are recognized in the corrected image and their positions in the fisheye image are calculated using the recognition results. However, it is difficult for these methods to correct large amounts of distortion in the image obtained from fisheye camera using the calibration method. In addition, 3D positions cannot be determined by using these methods[1-6]. Stereo fisheye cameras were conventionally used to measure people's 3D positions but calibrating the cameras is difficult and time-consuming[7-9]. Obtaining 3D positions from only one 360-degree fisheye camera is challenging due to the large amounts of distortion in the fisheye camera images and lack of parallax.

Our objective is to recognize people and measure their 3D positions in a workplace using one monocular fisheye camera. Because the fisheye camera has a wide angle of view, fewer cameras are needed to cover a large factory. Setting the fisheye camera is much easier than setting stereo cameras.

Our proposed method consists of one 360-degree equirectangular image from a fisheye camera projected onto a unit sphere. Several perspective projection images are generated to correct distortion. A machine learning method is applied to the corrected images for people recognition, and the recognition results are used to calculate the people's 3D positions in the workplace. A few markers were set on floor of the workplace to calibrate the perspective projection images with the plane of the floor. We conducted experiments in a workplace using Ricoh R 360-degree monocular fisheye cameras and found that our proposed method was effective for recognizing people and determining 3D positions.

1-1 Related Work

A monocular fisheye surveillance camera has been used for recognizing and tracking people. The image distortion from the fisheye camera is first corrected by calibration[1-5], and then people are recognized from the images[10-15]. However, it is difficult for these methods to correct a 360-degree equirectangular image captured by fisheye camera using the calibration method, as shown in Fig. 1. Though recognizing pedestrians directly from a distorted fisheye image was proposed[6], collecting the learning image data was too time-consuming. Because the image distortion from a fisheye camera differs depending on the position in the image, many recognition models are made for one fisheye image. The people recognition models are made for a specific camera. To apply the model to a different fisheye camera, we would need to collect learning image data and train models from scratch. Furthermore, 3D positions cannot be obtained by these methods[1-6].

Stereo cameras have been used to recognize people and measure distance to obtain 3D positions[7,8]. However, calibrating the cameras is difficult and time-consuming, especially for stereo fisheye cameras[9].

It is difficult for conventional methods to recognize people and obtain their 3D position from one 360-degree fisheye camera.

1-2 Proposed Method

We propose a method for measuring people's 3D positions in a workplace using only one monocular fisheye camera. Fig. 3 shows a flow chart of the proposed method. One 360-degree equirectangular image from a fisheye camera is used to recognize people and obtain their 3D position. The equirectangular image is projected onto a unit sphere and several perspective projection images are generated to correct image distortion.

Fig. 1 360-degree equirectangular image from fisheye camera.

Fig. 2 Fisheye image with corrected distortion.

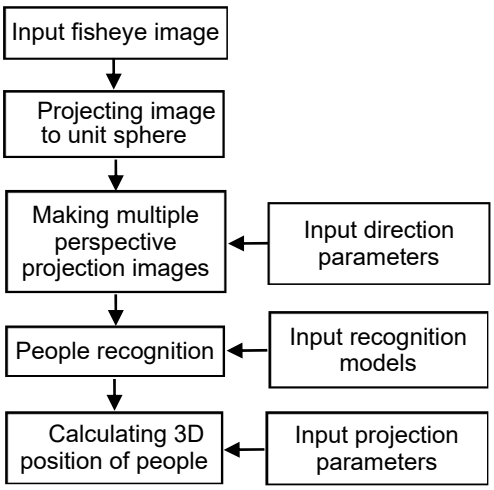


Fig. 3 Flow chart of proposed method.

A machine learning method is applied to the corrected images for people recognition and the results are used to calculate the people's 3D positions in the workplace. A few markers were placed on the floor of the workplace to calibrate the perspective projection images and the plane of the floor. The proposed method is described in detail in the following sections.

2. Fisheye Distortion Correction

In our proposed method, several projection transforms are used to correct image distortion from the fisheye camera. Fig. 1 shows an example of an equirectangular projection image with large distortion taken by a 360-degree fisheye camera. Fig. 2 shows the same image with the distortion corrected using our method.

2-1 Equirectangular projection

Equirectangular projection is a type of projection used to map a portion of the surface of a sphere to a flat image. The horizontal coordinate is longitude, and the vertical coordinate is latitude. Equirectangular images cover 360° horizontally and 180° vertically. Equirectangular projection is the most commonly used format for 360-degree fisheye cameras. To correct distortion, the equirectangular image is projected onto a unit sphere.

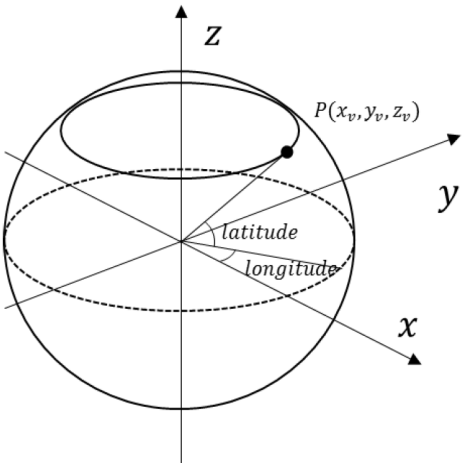


Fig. 4 Unit sphere used for image projection transform.

2-2 Unit Sphere Projection

To correct the image distortion from the fisheye camera, the equirectangular projection image is projected onto a unit sphere, as shown in Fig. 4. The radius of the sphere is 1. The light input to the fisheye lens passes point P (x_v, y_v, z_v) on the unit sphere and the center of the sphere O. The angle between the input light ray and the equatorial plane in the vertical direction is *latitude* and the horizontal direction angle is *longitude*.

The relation between coordinates (x_v, y_v, z_v) and the input angles of the light ray (*latitude, longitude*) is shown in Equations (1)–(3). One pixel on the equirectangular projection image has the coordinates (*latitude, longitude*). This image pixel is projected to the point on the unit sphere which has coordinates (x_v, y_v, z_v).

$$x_v = \cos(latitude)\cos(longitude) \tag{1}$$

$$y_v = \cos(latitude)\sin(longitude) \tag{2}$$

$$z_v = \sin(latitude) \tag{3}$$

2-3 Unit Sphere Rotation

As shown in Fig. 1, the 360-degree equirectangular image of fisheye camera covers a large area of the workplace. To cover the area using images with corrected fisheye distortion , multiple images are made by using the unit sphere shown in Fig. 4. Each corrected image is made in a specific direction. The unit sphere is rotated in several directions to make the distortion corrected images. The angles (α, β, γ) are the rotation angles by which the unit sphere is rotated. α, β and γ are the rotation angles around the x, y and z axis. Coordinates (x, y, z) of a point on the sphere are transferred to (x', y', z') by Equation (7). Coordinate transforms are made by Equations (4)–(6) separately. The combined rotation by angles (α, β, γ) are calculated by Equation (7).

As shown in Fig. 5, the whole surface of the unit sphere can cover the entire 3D space. Because the equirectangular image in Fig. 1 is projected onto the entire unit sphere, part of the unit sphere surface can cover a large area of the workplace. Thus, the workplace can be covered by very few fisheye cameras which can be set up easily. The proposed method can be applied to future surveillance camera products.

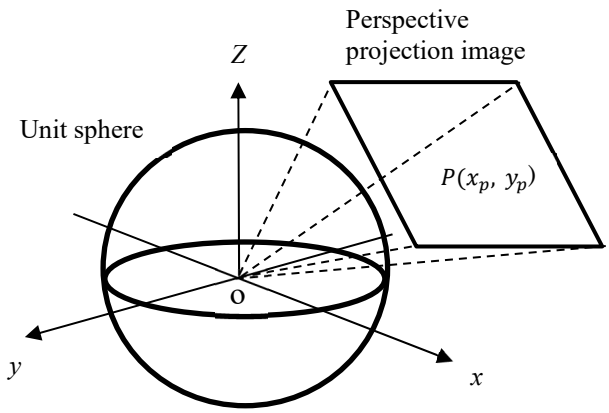


Fig. 5 Perspective projection image created from unit sphere.

The corrected image can be created by rotating the unit sphere in a specific direction by angles (α, β, γ). Multiple corrected images can be created to cover a large area by rotating the unit sphere several times.

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos\alpha & -sin\alpha \\ 0 & sin\alpha & cos\alpha \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = R_x(\alpha) \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{4}$$

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} cos\beta & 0 & sin\beta \\ 0 & 1 & 0 \\ -sin\beta & 0 & cos\beta \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = R_y(\beta) \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{5}$$

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} cos\gamma & -sin\gamma & 0 \\ sin\gamma & cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = R_z(\gamma) \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{6}$$

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = R_x(\alpha) R_y(\beta) R_z(\gamma) \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{7}$$

2-4 Perspective Projection

As shown in Fig. 5, the perspective image is made by projecting the image from the unit sphere onto the perspective image plane. Point $P(x_p, y_p)$ on the perspective image plane is related to the input light ray angles (*latitude, longitude*) as expressed by Equations (8) and (9). The perspective image plane is the tangent plane to the unit sphere.

$$longitude = \arctan\left(\frac{y_p}{x_p}\right) \tag{8}$$

$$latitude = \arctan\left(\frac{z_p}{\sqrt{x_p{}^2 + y_p{}^2}}\right) \tag{9}$$

2-5 Correcting Distorted Image

The fisheye distortion is corrected by using the projections described above. The intensity of the image pixel on the perspective image plane at position $P(x_p, y_p)$ is calculated from the equirectangular projection image pixel with the coordinates (*latitude, longitude*). First, coordinates (x_p, y_p) are mapped to point (x', y', z') on the unit sphere using Equations (1)–(3), (8), and (9). Then the coordinates (x, y, z) on the sphere prior to rotation is calculated by Equations (4)–(7). The rotation angles (α, β, γ) are parameters set before calculation. Finally, the coordinates (*latitude, longitude*) on the equirectangular projection image are calculated using point (x, y, z) on the unit sphere, using Equations (1)–(3). The coordinates (*latitude, longitude)* may not be integers. Bi-linear interpolation is used to calculate the pixel intensity of the equirectangular projection images. Several perspective images are made to cover the entire work area.

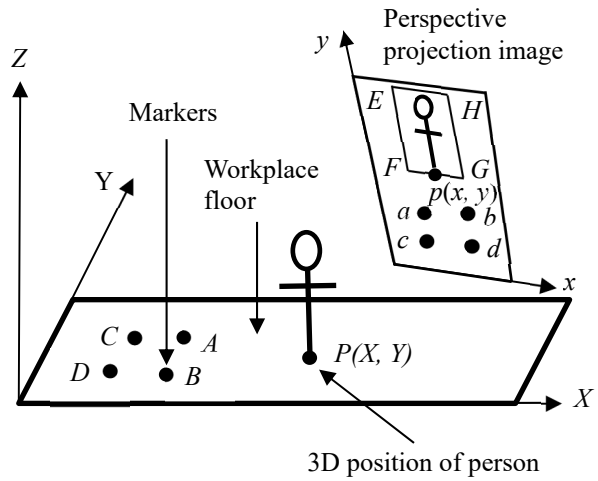


Fig. 6 Measurement of worker's 3D position.

3. 3 Dimension Position Measurement

Our proposed method determines the 3D position of a worker on the floor plane of the workplace, as shown in Fig. 6. The position at point *P(X, Y)*, the point on the floor plane, is used as the position of a worker in the workplace. It is assumed that a worker is standing. The position of their feet is used as their 3D position in the workplace. In the perspective projection image, the worker's 3D position is represented by point *p(x, y)*. By calibrating the perspective plane with the 3D plane of the workplace floor, we can calculate the 3D position *P(X, Y)* using point *p(x, y)*. The floor plane and the perspective projection image are planes, which can be expressed by perspective projection in Equation (10). The parameters of m_{ij} can be calculated by calibration. By setting several markers on the floor, the parameters m_{ij} can be calibrated. Four points can be used to calculate the transform parameters between one perspective image and the floor plane. *A, B, C, D* represent the markers on the floor plane. Points *a, b, c* and *d* on the perspective image plane are the images of points *A, B, C,* and *D*. Because positions *A, B, C, D, a, b, c, d* are known values, eight equations can be formulated to obtain eight unknown values of m_{ij}. We project several perspective images to cover a large area of the

workplace. The perspective images are calibrated using the markers on the floor plane. Then, we can calculate the 3D position of a worker from his or her position on the perspective image.

$$\begin{pmatrix} X \\ Y \\ u \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (10)$$

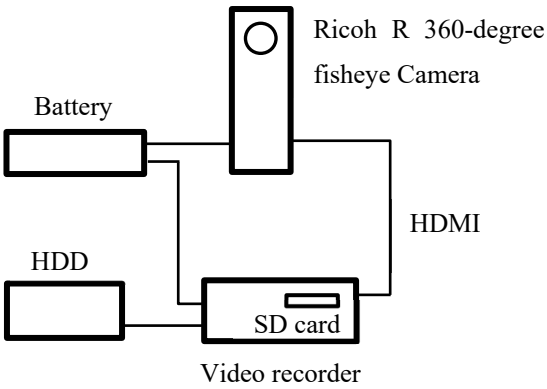


Fig. 7 Experimental setup.

4. People Recognition and Position Measurement

As shown in Fig. 6, people recognition is performed on the perspective projection image plane. A machine learning method is used in the proposed method. Histograms of Oriented Gradients (HOG) features and Support Vector Machine (SVM) machine learning are used to recognize people in our proposed method[10]. A pre-trained model was used for people recognition and the results are expressed as rectangle *EFGH*. The position at the middle of line *FG*, the bottom line of the resultant rectangle, is used as the position of the worker on the perspective image. By using the calibrated parameters of m_{ij} and Equation (10), the 3D position *P(X, Y)* can be calculated from position *p(x, y)*.

5. Experiment Results

Ricoh R 360-degree fisheye cameras, such as the one shown in Fig. 8, were used for people recognition and 3D position measurements. The experimental setup is illustrated in Fig. 7. A video stream is outputted from the camera to a video recorder via an HDMI cable. The video data is recorded to an SD card or hard disk (HDD) in the recorder. The power supply to the camera and video recorder is provided by a mobile battery. The Ricoh R fisheye cameras are attached to walls or pillars in the workplace.

The output fisheye image of Ricoh R is a 1920×1080-pixel equirectangular projection image. An example of the fisheye camera is shown in Fig. 1. 3D perspective projection images were made to correct fisheye distortion. The views from 30, 90, and 30 degrees were used to obtain the left, front, and right perspective projection images, respectively. The size of each perspective image was 640×480 pixels.

In the experiment, we did not train the machine learning model by collecting images of people. Instead, a pre-trained model was used for people recognition. Because the distortion of the equirectangular image was well corrected, the pre-trained model was able to attain accurate recognition results.

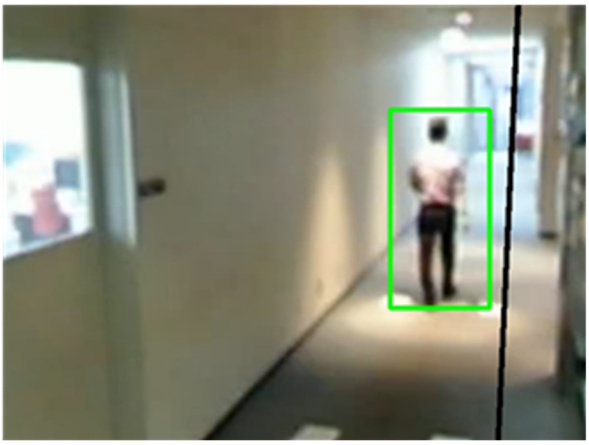
Fig. 9 shows the results of people recognition. Fig. 9 (a) is the image from the left perspective and the recognition results which are indicated by the green rectangles. Fig. 9 (b) is the image from the right perspective and the respective recognition results. In Fig. 10, the red points indicate the feet positioning of the people in the perspective image, which are used for 3D position measurement.

The recognition rate and false positive rate were evaluated, the results of which are shown in Table 1. 33,298 images were used for evaluation. The measurement results of the people's 3D positions are shown in Fig. 11. The red points indicate the 3D positions

which are the result of 290 frames. X and Y are distance in the horizontal and vertical direction in millimeters. The accuracy of distance measurement was 6.6%.

Fig. 8 Ricoh R 360-degree fisheye camera.

(a)



(b)

Fig. 9 People recognition results on perspective projection images.
(a) Left direction people recognition results.
(b) Right direction people recognition results.

Fig. 10 Example of people recognition results. Red points indicate positions in perspective projection image.

Table 1 People recognition evaluation results.

	People recognition results
Recognition rate (%)	92.5
False positive rate (%)	0.2

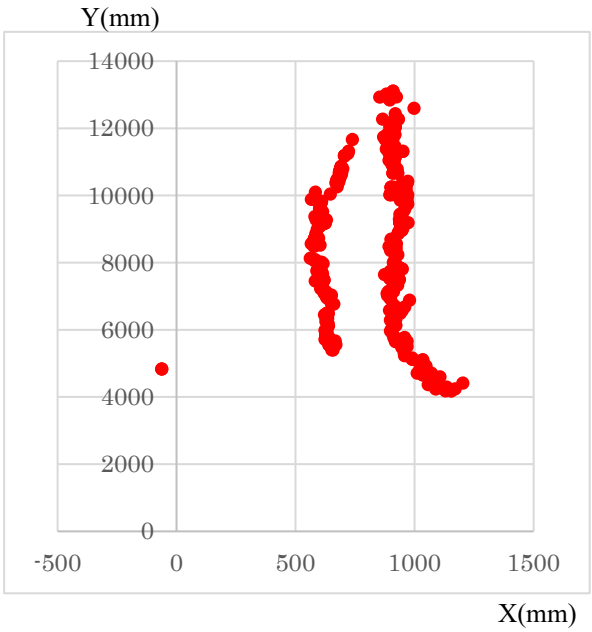


Fig. 11 3D position of people measurement results. Peoples' 3D position calculated from 290 fisheye image frames are printed at the same graph by red points.

6. Conclusion

We proposed a method for recognizing people and measuring their 3D positions in a workplace by using only one fisheye camera. One 360-degree equirectangular image captured by a fisheye camera is projected onto a unit sphere and several perspective projection images are generated to correct distortion. A machine learning method is applied to the corrected images for people recognition, and the recognition results are used to calculate the 3D positions of the people in the workplace. Only a few markers were needed to be placed on the floor of the workplace to calibrate the perspective projection images and the plane of the workplace floor.

People recognition and position measurement experiments were conducted in a workplace using Ricoh R 360-degree monocular fisheye cameras. The rate of people recognition was 92.5% with a false positive rate of 0.2%, and the measurement accuracy of the people's positions was 6.6%, thus demonstrating the effectiveness of our proposed method.

References

1) K. V. Beeck, T. Goedeme: Real-time accurate pedestrian detection and tracking in challenging surveillance videos, *Proceedings of the 10th International Conference on Computer Vision Theory and Applications*, Vol. 3, pp. 325–334 (2015).
2) K. V. Beeck, T. Goedeme, T. Tuytelaars: A warping window approach to real-time vision-based pedestrian detection in a truck's blind spot zone, *Proceedings of the ninth international conference on informatics in control, automation and robotics*, Vol. 2, pp. 561–568 (2012).
3) S. Silberstein et al.: Vision-Based Pedestrian Detection for Rear-View Cameras, *2014 IEEE Intelligent Vehicles Symposium*, pp. 853–860 (2014).
4) M. T. Bui et al.: People Detection in Heavy Machines Applications, *International Conference on Cybernetics and Intelligent System & Robotics, Automation and Mechatronics*, pp. 18–23 (2013).
5) Z. Zhang, P. L. Venetianer, A. J. Lipton: A Robust Human Detection and Tracking System Using a Human-Model-Based Camera Calibration, *The Eighth International Workshop on Visual Surveillance*, pp. 1643–1653 (2008).
6) M. Demirkus et al.: People Detection in Fish-eye Top-views, *In Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, pp. 141–148 (2017).
7) C. G. Keller, M. Enzweiler, D. M. Gavrila: A new benchmark for stereo-based pedestrian detection, *2011 IEEE Intelligent Vehicles Symposium (IV)*, pp. 691–696 (2011).
8) S. Nedevschi, S. Bota, C. Tomiuc: Stereo-Based Pedestrian Detection for Collision-Avoidance Applications, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 10, No. 3, pp. 380–391 (2009).
9) J. Ahn: Calibration of stereo fish-eye camera for Driver Assistance Systems, *25th International Conference of Image and Vision Computing New Zealand*, pp. 1–5 (2010).
10) N. Dalal, B. Triggs: Histograms of Oriented Gradients for Human Detection, *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 886–893 (2005).
11) P. Dollar et al.: Pedestrian Detection: A Benchmark, *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 304–311 (2009).
12) A. Angelova et al.: Real-Time Pedestrian Detection with Deep Network Cascades, *Proceedings of the British Machine Vision Conference (BMVC)*, pp. 32.1–32.12, BMVA Press (2015).

13) D. Ribeiro et al.: A Real-Time Deep Learning Pedestrian Detector for Robot Navigation, *2017 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, pp. 165–171 (2017).
14) Y. Tian et al.: Deep Learning Strong Parts for Pedestrian Detection, *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1904–1912 (2015).
15) D. Xu et al.: Learning Cross-Modal Deep Representations for Robust Pedestrian Detection, *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5363–5371 (2017).