EXTRACTION OF PEDESTRIAN REGIONS USING HISTOGRAM AND LOCALLY ESTIMATED FEATURE DISTRIBUTION

Koji Imaeda, Yasushi Hirano, Shoji Kajita, Kenji Mase[*]

Graduate School of Information Science
Nagoya University
Chikusa, Nagoya 464-8601, Japan

Nobuyuki Shiraki, Akihiro Watanabe

Vehicle Safety Research Center
Toyota Central R&D Labs.
Nagakute, Aichi 480-1192, Japan

ABSTRACT

This paper proposes a novel method for extracting a precise object area from a given rough initial region, where the approach is to extend a conventional kernel density estimator. The proposed method can correctly extract the object area by introducing a brightness histogram as a probability density function with an object distribution map as a priori probability for preprocessing the kernel density estimator. We confirm that the accuracy and computation speed are improved in comparison to those of the original method, even if the background texture is complex and the initial region is not accurate. We also extend the object distribution map to reflect the shape of target objects such as pedestrians.

1. INTRODUCTION

Object extraction is one of the most important and difficult issues in image processing. There are many extraction methods, including semi-automatic ones in which a rough region of an object is initially given to extract a precise object area automatically. A semi-automatic method is useful not only for manually drawn initial regions but also as postprocessing when it is combined with, for instance, stereo vision, inter-frame differentiation, and key object extraction such as face detection. We propose a technique to improve the kernel density estimator, which is one of the semi-automatic object-extraction methods, to obtain more accurate results.

Development of an automatic people locator is one of the highest priorities for automobile driving assistance and autonomous robot navigation. There are many approaches to pedestrian detection from visible and infrared images [1,2], and these have demonstrated good results for restricted imaging conditions, e.g. simple background and straight standing posture. The problem becomes very difficult, however, when the illumination changes, the contrast of the object to the background is low, or the observed people make various postures such as crouching. This is partly because the current approaches rely on edge features of the object region or strong assumptions on the object's contrast. By employing the kernel density function, we are able to exploit general probability density features of the object/background regions as well.

We introduce a Bayesian discriminant function as preprocessing for the kernel density estimator to classify an initial region into object and background regions based on the brightness histogram. Thus, we can supply general brightness-based information, i.e. the contrast between object and background regions, to the density estimator. We confirmed the performance of our proposed method through experimental tests with visible and infrared images in real settings, showing that the accuracy of object extraction and the computational speed are improved over the results of the original kernel density estimator. Moreover, our method has the advantage of controlling a priori probability map of the extracted object according to the expected shape of the target object. The map can be generated as a template from the learning samples to form, for example, an arbitrary posture for extracting pedestrians in various postures. The performance of the template map is also demonstrated by the experimental results.

In the rest of the paper, we refer to related research in Section 2, explain the proposed method in Section 3, and show experimental results, including the template map adaptation, in Section 4. We conclude with discussion and future directions.

2. RELATED RESEARCH

There has been much research on pedestrian extraction for automobile applications. Zhao et. al [3] tried to discriminate pedestrians from other objects by applying a neural network pattern recognition to the extracted region by stereo depth queue. Fang et. al [4] used a far-infrared

[*] Y. Hirano, S. Kajita and K. Mase are also members of Information Technology Center

image to extract the upper-body and lower-body separately. Bertozzi et. al [5] used a multi-resolution infrared image with the features of symmetry and edge density. Soga et. al [2] used the edge features obtained by a Prewitt operator against a stereo range image to achieve good detection performance for straight-standing postures. It is convenient to separate out the front object from the cluttered background image with stereo vision [6]. Since this can only provide a rough region of the object, we still need a refining process to obtain the accurate region of an object to detect it as a pedestrian.

The methods of extracting objects from images can be categorized into two approaches. One approach is to adjust a given contour and/or a region by Active Contour Models [7, 8]. The other is to estimate and determine the category of each pixel, such as Interactive Graph Cuts [9] and Grab Cuts [10]. The former represents an active contour as a parametric curve, such as a Spline curve, and piecewise line segments. By defining some energy functions as a constraint based on the strength of edge, smoothness of curves, etc., the given contour is recursively adjusted to fit the curve to a reasonable contour. However, the resulting contour doesn't closely fit the small, high-curvature or concave curve due to the smoothness constraint.

On the other hand, the latter pixel-based extraction method uses the probability density functions (p.d.f.) of an object and background regions for each to represent the brightness distribution of the image. The p.d.f. is often estimated from a brightness histogram [9] or a mixture Gaussian model [10]. By labeling each pixel with the category that gives a higher value of p.d.f., the object region is extracted. This approach outperforms the active contour model in terms of extraction performance for small and high-curvature objects. However, since it uses the same p.d.f.s independently of the pixel's location in the image, the category estimation sometimes fails when the brightness distribution of the pixel neighborhood is closer to the opposite category. It also fails when the initial region is given to the extent of providing accurate p.d.f.s.

Takeshima et. al [11] proposed an object-extraction method based on a kernel density estimator [12, 14] in order to solve these problems. It uses a joint probability function of multiple random variables for position, color and category of each pixel. The use of random variables of pixel position works effectively to reduce the estimation error for similar color regions, and it has robustness against the ambiguity of initial region specification. However, it still has problems when the initial object region contains a background edge. We propose introducing a brightness histogram of the initial region as a p.d.f. with

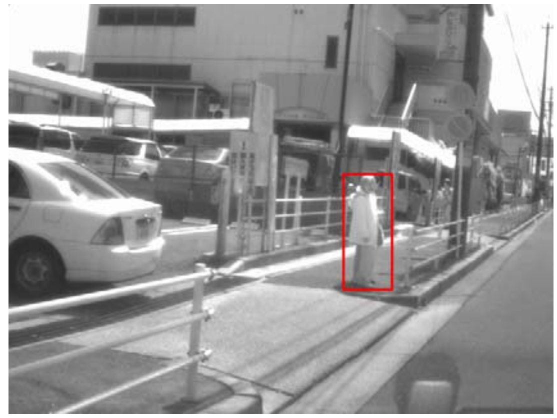


Fig. 1. Input image and bounding box as a given initial region (rectangle)

Fig. 2. Bounding box (rectangle) and enlarged ROI

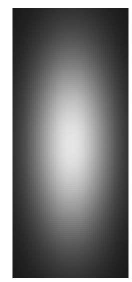


Fig. 3. Probability distribution function for the ROI

an object distribution map as the a priori probability. This is applied to the preprocessing of object extraction of a kernel density estimator.

3. KERNEL DENSITY ESTIMATOR WITH BAYSIAN DISCRIMINANT FUNCTION

We attempted to improve the kernel density estimator by introducing a method based on Bayesian discrimination as a preprocessing of the initial estimation of an object's region. We assume in this paper that the initial region is given as a rectangular bounding box, which is manually given in the experiment in the next section. Such a bounding box can be obtained easily from the stereo vision range image in real applications.

3.1. Region of Interest for Processing

The Region of Interest (ROI) is set according to the size and location of the bounding box. In our experiment, we assume that the outskirts of the bounding box represents the background. The ROI is set by enlarging by 1.5 times the width and height of the original bounding box to include enough pixels of the background (see Fig.1, Fig.2).

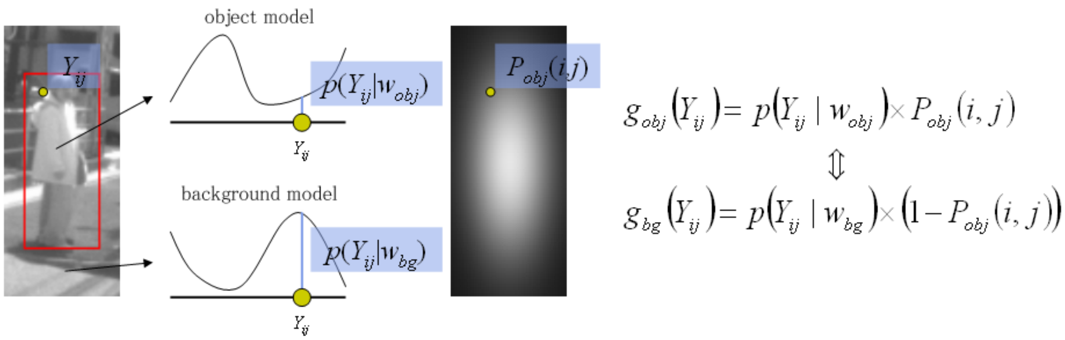


Fig. 4. Category estimation flow by brightness histogram

3.2. Modifying Probability Distribution Function

We introduce a p.d.f. as an existing-object probability map in order to give a priori knowledge of object position. This is considered useful for reducing the estimation error when there is an area of similar brightness between the object and the background.

In this case, we first introduce a simple two-dimensional Gaussian map with the peak in the middle of the ROI and the standard deviation computed from the width and height of the ROI (Fig.3). That is, the p.d.f. $P_{obj}(i, j)$ of an object at a pixel (i, j) in ROI I is given by the following equations.

$$P_{obj}(i,j) = \frac{1}{\sqrt{2\pi}\sigma_{width}} \exp\left(-\frac{(i - I_{width}/2)^2}{2\sigma_{width}}\right) \times \frac{1}{\sqrt{2\pi}\sigma_{height}} \exp\left(-\frac{(i - I_{height}/2)^2}{2\sigma_{height}}\right)$$

where I_{width} and I_{height} are width and height of the ROI, respectively, and σ_{width} and σ_{height} are set as $\sigma_{width} = I_{width}/4$ and $\sigma_{height} = I_{height}/4$ so that the region of 2σ will fit within the ROI.

The background probability is given by P_{bg} $(= 1 - P_{obj})$ for each pixel. In order to maintain the total equality of object probability over the ROI, we need to stretch the density function to obtain a normalized existing-object probability map. The stretching factor is determined so as to obtain a value P_{obj} between 0.1 and 0.9.

3.3. Category Estimation by Histograms

We first estimate the object and the background regions roughly by using the bounding box and the ROI. We compute a brightness histogram within the bounding box as an object-region brightness model (object model) and another histogram within the ROI, excluding the bounding box region, as a background-region brightness model (background model). The object model is represented by $p(Y_{ij} \mid w_{obj})$ with brightness $Y(i, j) = Y_{ij}$ as the random variable.

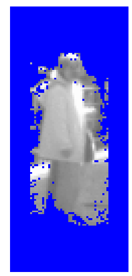


Fig. 5. Preprocessing result by histogram matching

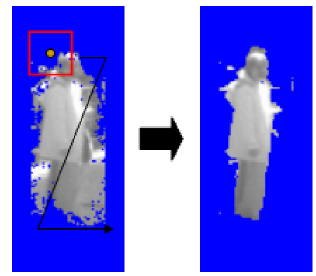


Fig. 6. Object region estimated by kernel density estimator with proposed preprocessing

Then, the discriminant functions $g_{obj}(Y_{ij})$ and $g_{bg}(Y_{ij})$ for object and background regions, respectively, can be represented by the following equations.

$$g_{obj}(Y_{ij}) = p(Y_{ij} \mid w_{obj}) \times P_{obj}(i,j)$$
$$g_{bg}(Y_{ij}) = p(Y_{ij} \mid w_{bg}) \times P_{bg}(i,j)$$

where $P_{bg} = 1 - P_{obj}$. Then, we can compute the values of the discriminant functions $g_{obj}(Y_{ij})$ and $g_{bg}(Y_{ij})$ in order to estimate the category of the pixel (i, j). This category is given by the function that gives the larger value (Fig.4, Fig.5).

3.4. Kernel density estimation with proposed preprocessing

The category estimation is performed by the kernel density estimator [11] with the estimated preprocessing result of the initial category in the previous section (Fig.6). The estimated category is used as initial values of a random variable. We also use the position and the brightness of each pixel and the brightness variance over a neighborhood of 5 by 5 pixels to represent a textual feature.

The pixel position's bandwidth for the kernel density estimation is 1/10 of width and height of the ROI with a uniform kernel, while the brightness bandwidth is 1/2 of the standard deviation of the brightness of pixels within the ROI with an Epanechnikov kernel, and the brightness variance bandwidth is 1/2 of the brightness variance across the pixels within the ROI with a Gaussian kernel. These parameters and kernels are chosen heuristically based on a pre-experiment with learning samples. The iteration of estimation stops when the category estimation is fixed at all pixels.

4. GENERAL EXPERIMENT AND DISCUSSION

We conducted a general experiment with the original kernel density estimator and the proposed method. The same test image and bounding box are given for each method. The test images are snow (93×87 pixels), cup (96×84), drink (90×90), pedestrian1 (59×138), pedestrian2 (59×136), and pedestrian3 (58×139) (see Fig.7(a)). The results are listed in Table 1 and shown in Fig.7. The number of error pixels in the table 1 represents the sum of true-negative and false-positive pixels. The details of the error distribution are given in Table 2. The ground-truth is given by manual drawing over images. The number of false-positive pixels is reduced significantly in most of the examples except for the snow image. Furthermore, the total number of error pixels shows a decreasing trend.

For simple objects such as the snow image, the extraction quality of the proposed method is as good as the original. The proposed method can help to reduce the iteration time, i.e. the computational cost. For complex shapes such as the cup, drink and pedestrian1 images, the extraction results with the proposed method are better than the original. However, for difficult examples such as the pedestrian2 and pedestrian3 images, the proposed method fails to extract some parts of the objects. This is because the given object's (a priori) probability map from the p.d.f assumes a general two-dimensional Gaussian distribution. Consequently, the true object region close to the bounding box tends to be classified as the background.

5. APPLICATION TO PEDESTRIAN EXTRACTION

We introduced an a priori probability map that reflects the target object shape, i.e. a pedestrian, to obtain a better result. The a priori probability map is generated from the example images of pedestrians from a car-mounted camera. The map is generated by taking the average of each position's pixel values, and then it is blurred by a Gaussian filter (5×5 pixels) after normalizing the width and height of the manually extracted pedestrian shapes.

A preliminary experiment with a pedestrian-shaped map was conducted. Figure 8 shows some good results obtained after adjusting the parameters of the bandwidths. A better result was achieved by reflecting the target shape in the a priori probability map for these examples. However, if the shapes of the map and the target object don't match, the extraction may fail easily. Employing the edge information in the map positioning is a candidate technique to achieve more accurate extraction.

6. CONCLUSION

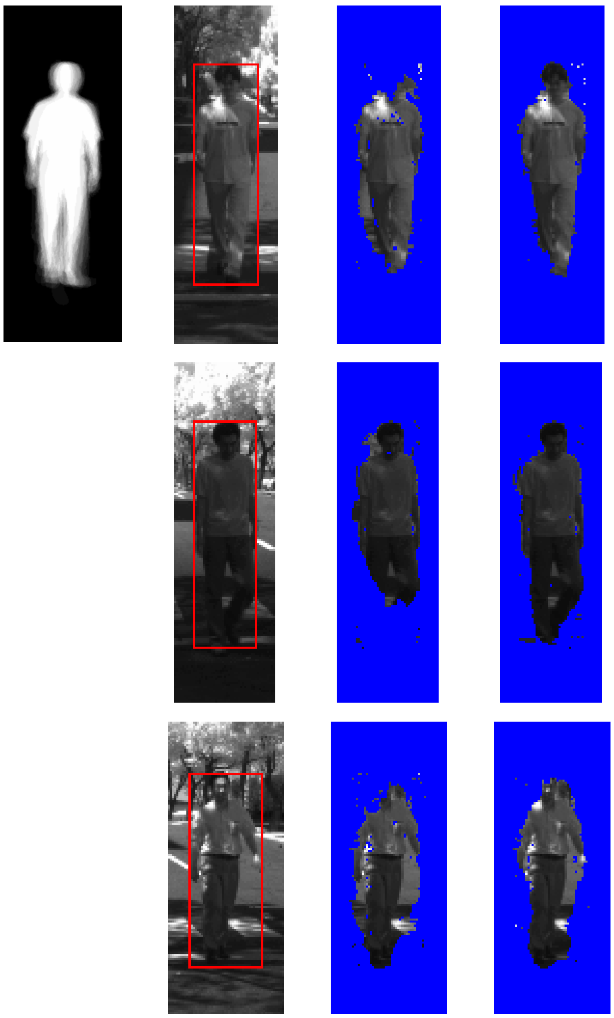
We proposed a method that introduces the preprocessing of category estimation to a kernel density estimator in order to achieve better performance in extracting an object from a given bounding box region. The discriminant function derived from a histogram-based probability density function can give a better initial categorization for the kernel estimator. Using the proposed method, the computational cost can be reduced for simple objects, and accuracy can be improved for objects having complicated shapes. Since an object's probability map is adjusted in advance by an example object, object extraction enough becomes robust and accurate for pedestrians in various

Table 1. Experimental results: original and proposed methods(or: original, pr: proposed)

	# of error pixels (ratio%)		# of iteration	
	or.	pr.	or.	pr.
snow	21(2)	21(2)	15	8
cup	727(37)	233(12)	31	66
drink	706(30)	333(14)	31	66
pedestrian1	1781(101)	311(14)	22	63
pedestrian2	1137(60)	821(43)	85	55
pedestrian3	510(25)	690(33)	29	16

Table 2. Error comparison between original and proposed method

	True-negative (pixels)		False-positive (pixels)	
	original	proposed	original	proposed
snow	18	18	3	3
cup	153	172	574	61
drink	38	229	1772	91
pedestrian1	7	222	1772	91
pedestrian2	674	776	463	45
pedestrian3	117	375	393	315



(a) (b) (c) (d)

Fig. 8. Pedestrian Extraction Results: (a) Priori probability map (b) Input image and bounding box, (c) Unmodified method, (d) Modified method.

postures. Future issues include a map-generation method, criteria of template selections and feature selection for more robustness against high-contrast backgrounds.

7. REFERENCES

[1] Ismail Haritaoglu, David Harwood, Larry S. Davis,"W4: Real-Time Surveillance of People and Their Activities," IEEE Trans. on PAMI, Vol. 22, No. 8, pp. 809-830 (2000).

[2] Mineki Soga, Takeo Kato, Mitsuhiko Ohta, Yoshiki Ninomiya, "Pedestrian Detection Using Stereo Vision and Tracking," IEEE ICDE Workshop on Realworld Multimedia Corpora in Mobile Environment, pp. 1-8 (2005).

[3] Liang Zhao, Charles E. Thorpe, "Stereo- and Network-Based Pedestrian Detection," IEEE Trans. On ITS, Vol. 1, No. 3, pp. 148-154 (2000).

[4] Yajin Fang, Keiichi Yamada, Yoshiki Ninomiya, Berthold K.P. Horn, Ichiro Masaki, "A Shape-Independent Method for Pedestrian Detection with Far-Infrared Images," IEEE Trans. on Vehicular Technology, Vol. 53, No. 6, pp. 1679-1697 (2004).

[5] Massimo Bertozzi, Alberto Broggi, Alessandra Fascioli, Thorsten Graf, Marc-Michael Meinecke, "Pedestrian Detection for Driver Assistance Using Multiresolution Infrared Vision", IEEE Transaction on Vehicular Technology Vol. 53 No. 6, pp. 1666-1678 (2004).

[6] M. Bertozzi, E. Binelli, A. Broggi, M. Del Rose, "Stereo Vision-based approaches for Pedestrian Detection", Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPRW'05) Workshops p. 16 (2005).

[7] M. Kass, A. Witkin, D. Terzopoulos, "Snakes:Active contour models", International Journal of Computer Vision Vol. 1 No. 4, pp. 321-331 (1988).

[8] N. Paragios, R. Deriche, "Geodesic active contours and level sets for the detection and tracking of moving objects", IEEE Transactions on Pattern Analysis and Machine Intelligence Vol. PAMI-22 No. 3, pp. 266-280 (2000).

[9] Y.Y. Boykov, M. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images", In Proceedings of International Conference on Computer Vision, pp. 105-112 (2001).

[10] A. Blake, C. Rother, M. Brown, P. Perez, P. Torr, "Interactive image segmentation using an adaptive GMMRF model", In Proceedings of the 8th European Conference on Computer Vision, pp. 428-441 (2004).

[11] Hidenori Takeshima, Toshimitsu Kaneko, Takashi Ida, "Extracting Object Regions Using Locally Estimated Color Distributions", Meeting on Image Recognition and Understanding (MIRU2006), IS1-4 (2006).

[12] M. Singh, N. Ahuja, "Regression based bandwidth selection or segmentation using parzen windows", In Proceedings 338 of IEEE International Conference on Computer Vision, pp. 2-9 (2003).

[13] D. Comaniciu, V. Ramesh, P. Meer, "The variable bandwidth mean shift and data-driven scale selection", In Proceedings of IEEE International Conference on Computer Vision, pp. 438-445 (2001).

[14] D. Comaniciu, "An algorithm for data-driven bandwidth selection", IEEE Transactions on Pattern Analysis and Machine Intelligence Vol. PAMI-25 No. 2, pp. 281-288 (2003).

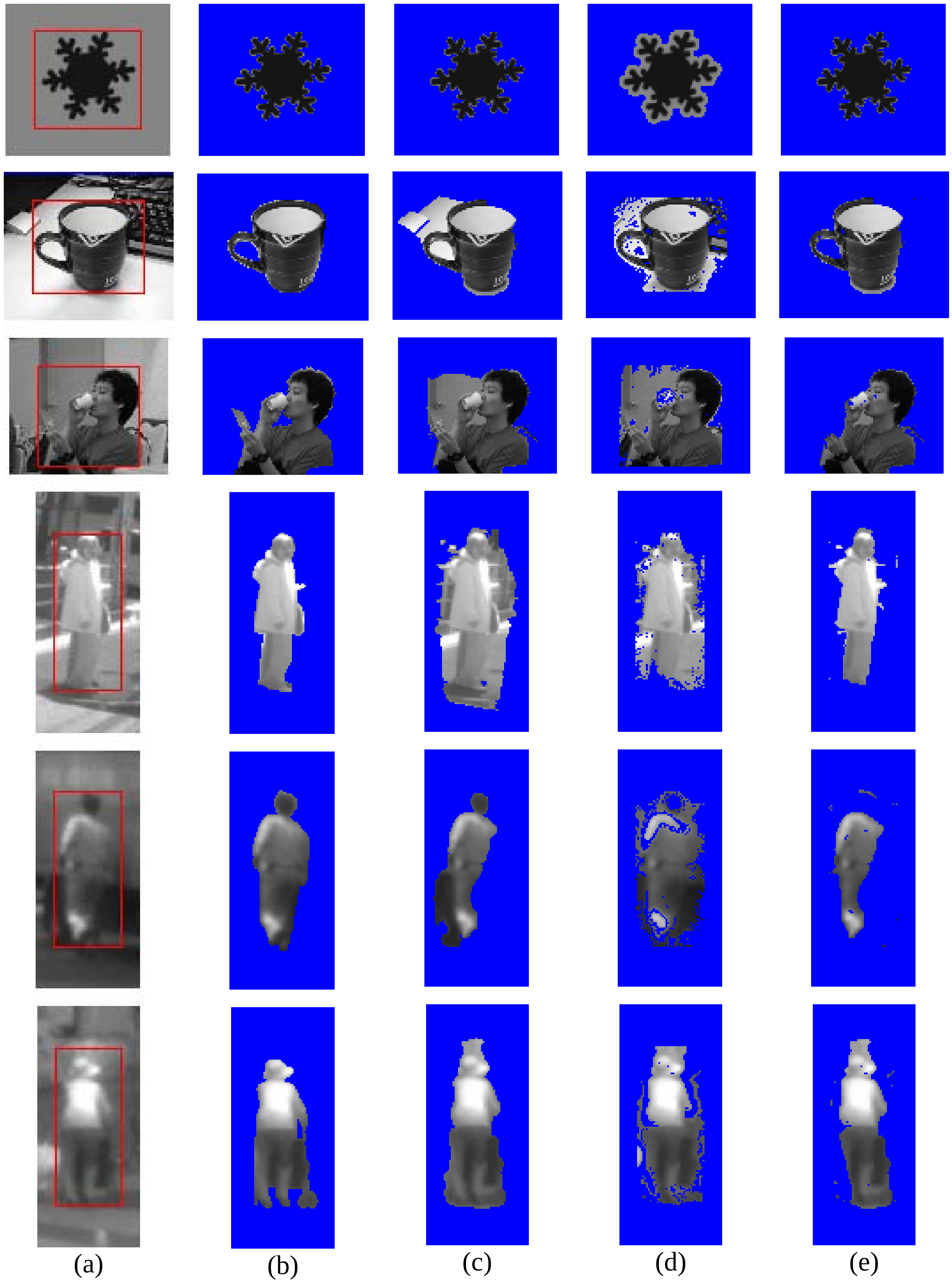


Fig. 7. Results of general experiment: (a) Input image and bounding box, (b) Ground-truth (manual), (c) Original method, (d) Preprocessing result by histogram, (e) Proposed method. From top to bottom: snow, cup, drink, pedestrian1, pedestrian2, and pedestrian3.