A Person Identification Method Using a Top-view Head Image from an Overhead Camera

Ryota Nakatani, Daichi Kouno, Kazutaka Shimada, Tsutomu Endo

Department of Artificial Intelligence, Kyushu Institute of Technology
680-4 Kawazu Iizuka Fukuoka 820-8502 JAPAN
{r_nakatani, d_kouno, shimada, endo}pluto.ai.kyutech.ac.jp

Abstract. In this paper, we describe a novel image-based person identification task. Traditional face-based person identification methods have a low tolerance for occluded situation, such as overlapping of people in an image. We focus on an image from an overhead camera. Using the overhead camera reduces a restriction of the installation location of a camera and solves the problem of occluded images. First, our method identifies the person's area in a captured image by using background subtraction. Then, it extracts four features from the area; (1) body size, (2) hair color, (3) hairstyle and (4) hair whorl. We apply the four features into the AdaBoost algorithm. Experimental result shows the effectiveness of our method.

Keywords: Top-view images, Person identification, Histograms of oriented gradients.

1. Introduction

Person identification is one of the most important tasks in computer vision. One approach to identify a person using a computer is based on analysis of captured images from a camera. Many researchers have studied image based person identification methods. The most famous approach is to use face information. Kanade [6] has proposed a method using face feature points, such as nose and eyes. The CLAFIC method [11] and the Eigen-Face method [9] are also famous approaches using face information. These methods need a face image for the identification process. However, face images are not always captured correctly from a camera. For example, in a traffic-choked situation, the face of a target person might be occluded by other persons .

Here we focus on an overhead camera. By using the overhead camera, the problem of occluded images is solved. Figure 1 shows an example of the occlusion problem and the solution with an overhead camera. In addition, a privacy issue is reduced because the camera does not capture the face image. Furthermore, the restriction of the location of a camera is reduced because the camera does not need to capture the person's face.

In this paper, we describe a novel person identification task using an image from an overhead camera. Figure 2 shows the installation location in our method. We focus on the situation that people stop in front of a door. First, our method extracts the person's area in a captured image by using background subtraction. Then, it extracts four features from the area; (1) body size, (2) hair color, (3) hairstyle and (4) hair whorl. We apply the four features into the AdaBoost algorithm. The main target of our system is person identification in a small-group environment such as offices and labs. Our system can easily obtain location information of each person by furnishing

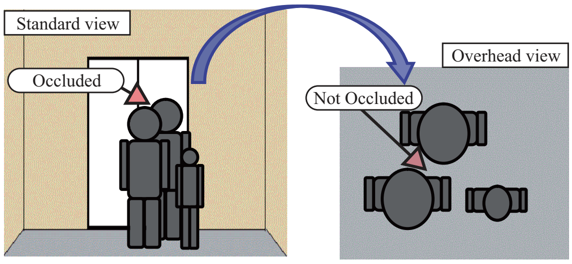

Fig. 1. The occlusion problem and the solution.

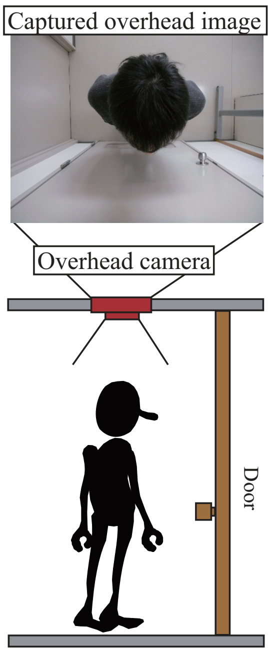

Fig. 2. The setting of our task.

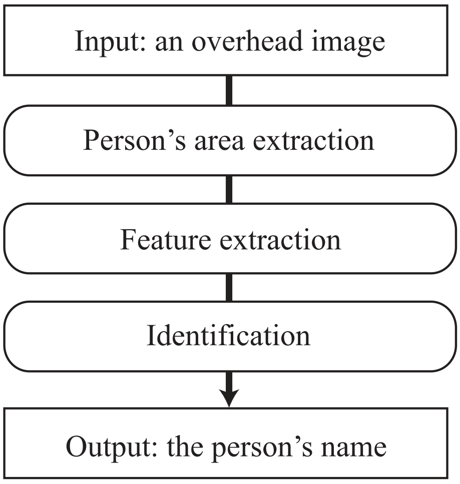

Fig. 3. The outline of our method.

with an overhead camera in front of a door of each room in the office or lab. It is essentially valuable for many systems, such as attendance management, person’s location retrieval and behavior recognition.

2. Related work

Uchida et al. [10] have reported a method for recognizing people at elevator lobby, using an overhead camera. The purpose of their study is to count people and recognize wheelchairs. Onishi and Yoda [7] have proposed a visualization system of customer flows using top-view images. The purposes of these studies were counting and tracking of people although they handled top-view images. The purpose of our method is to identify a person in an image.

Iwashita and Stoica [5] have proposed a gait recognition method using overhead images. They focused on shadow images for the gait recognition. We use four features such as body size and hair color for the identification. Fukuzoe et al. [3] have reported a person identification method using body shape. However, they did not use an overhead camera. Therefore, the system essentially included the problem of occluded images. By using top-view images, our method can resolve the problem of occluded images.

3. Proposed method

In this section, we explain the proposed method. Figure 3 shows the outline of our method. It consists of three parts: (1) person’s area extraction, (2) feature extraction and (3) a classifier for person identification.

3.1. Person’s area extraction

For the person identification, we need to extract the person’s area from a captured image. An example of

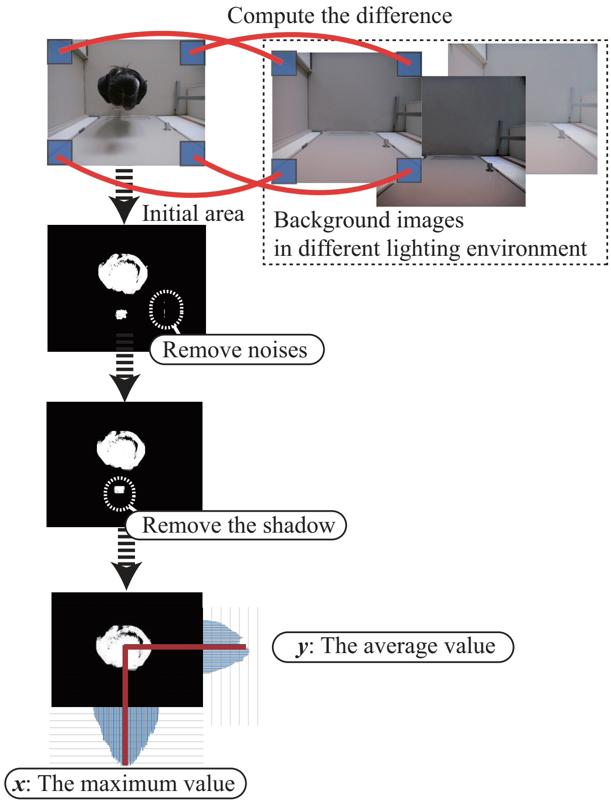

Fig. 4. An example of person’s area extraction process.

the person’s area extraction process is shown in Figure 4. The process is based on background subtraction and noise reduction. It estimates the person’s area by using histograms of the target image. The process is as follows:

1 Background subtraction
First, we apply a background subtraction method to the person’s area extraction process. The background subtraction approach is generally sensitive to the environment, such as illumination change. To solve this problem, we prepare several background images for the process. Then we compute the difference of brightness between four corners of the input image and each background image. By using the background image which possesses the smallest difference, we delete the background area in the image.

2 Noise reduction
The background subtraction usually generates many small nosies in the output image because of mismatching of the background image. In this step, we eliminate non-person’s areas by using surrounding area’s information of each pixel. In other words, we delete areas with a small quantity of pixels.

3 Shadow deletion
The image often contains the shadow of the target person. We delete the shadow area from the image. The shadow appears in the door area because of the location of room lighting. Here we estimate the per-

son's area candidate. We suppose that the biggest area is the person's area in the image. On the basis of suppositions of the shadow location and the person's candidate, we remove the shadow area.

4 Center detection
Finally, we estimate the center of the person's area from the image. For the estimation process, we use the histograms of the person's area. We use the maximum value for the x-axis and the average value for y-axis as the center (x,y), respectively. The reason why we use the maximum value for the x-axis is that the background subtraction occasionally remove the actual person's area extremely. This result leads to the estimation error of the center detection. In our preliminary experiment, this problem was evident in the x-axis. Shoulder areas were occasionally removed by the background subtraction method. It depended on the color of clothing. The estimation error occurs in the center detection if one shoulder is removed as the background and the method uses the average value of the histogram[1]. Therefore, we apply the maximum value to the x of the center (x,y).

3.2. Feature extraction

In this section, we describe features for a classifier. In this paper, we handle four features. The features are classified into one body feature, namely body size, and three hair features, namely hair color, hairstyle and hair whorl. For the hair features, we need to detect the head area. For the detection, we use the center information of the person's area detected in Section 3.1. In this paper, the size of the head area is 50×50 pixels.

- body size
  The physical size of the body[2] is one of the most effective features for the person's identification. In our method, we use the width size of x-axis and y-axis as the body size features. These values are extracted from the histograms of the person's area image.

- hair color
  The hair color is often different between persons. In our method, we use the sum of brightness values of each pixel in the head area.

- hairstyle
  The hairstyle is a characteristic element of each person. We compute the feature from edge information on the head area.

- hair whorl
  The hair color and hairstyle are often changed. On the other hand, the location and shape of hair whorl is a robust feature in time change. Hence, we use the hair whorl as the feature for the identification process. Our system manipulates the head area with a smoothing filter. As a result, the edges about hairstyle are deleted. We use the remaining edges as the hair whorl features.

1. On the other hand, the y-axis is less affected by this mistake.
2. To be precise, the body size features are the width between shoulders and the width of the head area.

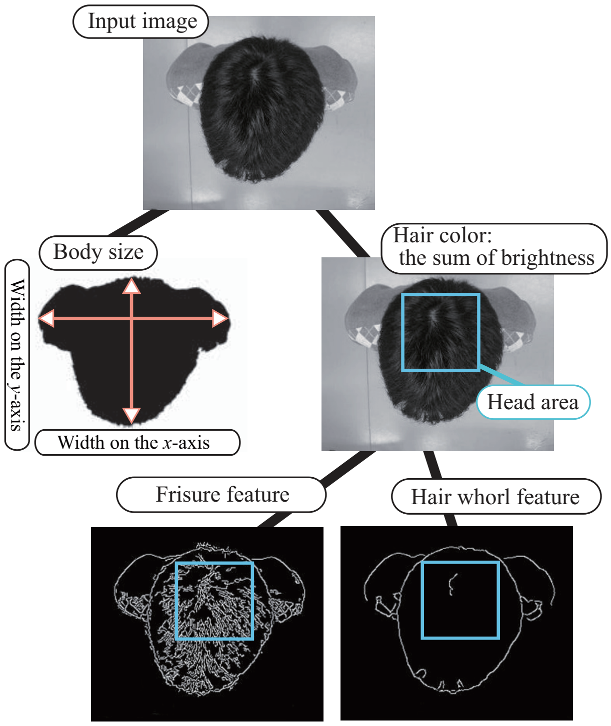

Fig. 5. An example of 4 features.

Figure 5 shows an example of the extracted features.

Next, we explain the feature extraction method for hairstyle and hair whorl in detail. The features are based on edge information. Hence, we need to convert the edge information into feature vectors. Here we apply the Histograms of Oriented Gradients (HOG) descriptors, which are reported by Dalal and Triggs [1], into the hairstyle and hair whorl features. The HOG is one of the most effective features for human detection and vehicle detection tasks in computer vision. The HOG descriptors are based on counting occurrences of gradient orientation in localized portions of an image. First, the method computes gradient magnitude $m(x,y)$ and orientation $\theta(x,y)$ in each pixel (x,y).

$$m(x,y) = \sqrt{f_x(x,y)^2 + f_y(x,y)^2}$$

$$\theta(x,y) = \tan^{-1}\frac{f_y(x,y)}{f_x(x,y)}$$

where

$$f_x(x,y) = I(x+1,y) - I(x-1,y)$$

$$f_y(x,y) = I(x,y+1) - I(x,y-1)$$

Then, it generates the cell histogram consisting of 5×5 pixels. Finally, it normalizes the block consisting of 3×3 cells.

$$L2 - norm = \frac{v}{\sqrt{||\mathbf{V}||^2 + \varepsilon^2}}$$

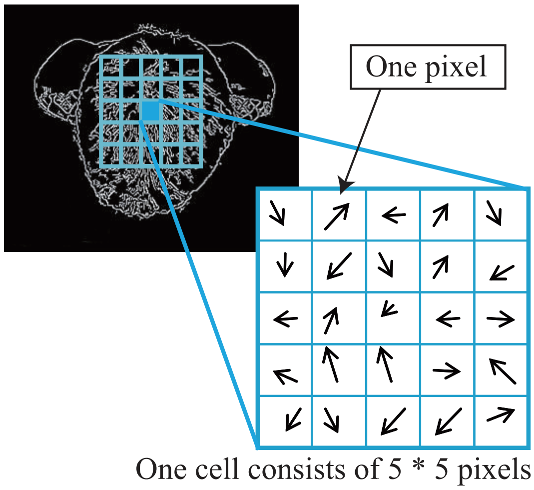

Fig. 6. Histogram of oriented gradients.

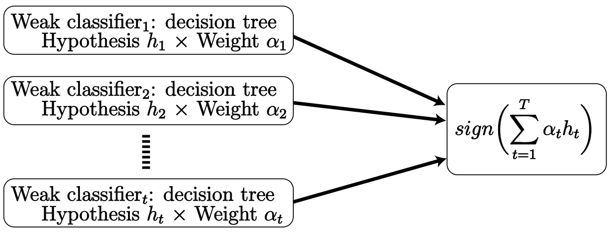

Fig. 7. Adaboost.

where **V** is the feature vector of the block. ε is the factor for the block normalization and $\varepsilon = 1$. Figure 6 shows the outline of the process.

3.3. Classifier for identification

Finally we construct a classifier based on four features. In this paper, we employ the AdaBoost algorithm [2] as the classifier. The AdaBoost algorithm is one of the most famous machine learning techniques. It generates a strong classifier by combining some weak classifiers. In this paper, we implement the AdaBoost with the open source software Weka[3]. We use the C4.5 algorithm [8] as the weak classifiers[4]. C4.5 is also one of the famous machine learning techniques, which generates a decision tree. Figure 7 shows the outline of the AdaBoost algorithm. We extend the AdaBoost into multi-label classification with the One-versus-Rest method.

4. Experiment

In this experiment, the number of test subjects was eight persons (six males and two females). We collected 30 images for each person. As a result, the dataset consisted of 240 images. The size of each image was 400 × 500 pixels. We evaluated our method with 10-fold cross-validation.

3. http://www.cs.waikato.ac.nz/ml/weka/
4. Actually, it is "J48" in Weka.

Table 1. The experimental result.

Features	Accuracy
size	88.8
color	28.3
style	63.3
whorl	64.6
size+color	89.2
size+style	91.7
size+whorl	88.8
style+color	62.9
style+whorl	66.7
size+style+color	**92.5**
size+style+whorl	88.3
size+color+whorl	88.8
style+color+whorl	66.3
all	88.8

4.1. Results

We evaluated all combination of four features. Table 1 shows the experimental result. In the table, "size", "color", "style" and " whorl" denote "body size", "hair color", "hairstyle" and "hair whorl" features respectively. "+" denotes a combination of features; e.g., "size+color" denotes the method with "body size" and "hair color" features. "all" is the system with the four features. The accuracy was computed by

$$\frac{\#\text{ of images identified correctly}}{\#\text{ of images}} \times 100.$$

The method with three features, "body size", "hair color" and "hairstyle", generated the best perfomance (92.5%). As the single feature, "body size" produced higher accuracy as compared with the other features. The combination of all features was not always effective in this experiment.

4.2. Discussion

In this section, we discuss the experimental results and the effectiveness of our system. In this experiment, 8 images of 240 images failed in the preprocessing of the person identification process, namely person's area extraction (Section 3.1). In the extraction process, the background subtraction was a major bottleneck. The mistake of the preprocessing leads to the decrease of the person identification accuracy. We need to improve the background subtraction process.

Next, we investigated the result of the best feature set, namely size+style+color, in detail. Table 2 shows the confusion matrix of the method with the feature set. "Sub_1" to "Sub_8" denote each test subject. "(M)" and "(F)" denote "male" and "female", respectively. "Height" is the body

Table 2. The confusion matrix of the best features set: size+style+color.

Subject	Sub_1	Sub_2	Sub_3	Sub_4	Sub_5	Sub_6	Sub_7	Sub_8	Height
Sub_1 (M)	28	0	0	0	2	0	0	0	175
Sub_2 (F)	0	28	0	0	0	0	2	0	155
Sub_3 (M)	0	0	27	0	1	1	1	0	171
Sub_4 (M)	0	0	0	30	0	0	0	0	181
Sub_5 (M)	0	0	2	0	28	0	0	0	176
Sub_6 (M)	0	0	0	0	0	25	0	5	168
Sub_7 (F)	0	0	0	0	0	0	30	0	148
Sub_8 (M)	0	0	0	1	0	3	0	26	175

height (centimeter) of each test subject. Mis-identification between the sexes was slight: one case (Sub_3 vs. Sub_7) in this experiment. For "Sub_6" and "Sub_8", there were 8 mistakes although their heights, which were inherently related to the body size feature, were different. It was caused by clothes. The test subject "Sub_6" often wore thick clothes such as a down jacket. As a result, our system overestimated the true size of his body in the person's area extraction process.

The body size was the most effective feature as a single feature for the identification (88.8 in Table 1), as well as a sensitive feature. As mentioned above, if clothes that he/she wore were changed, it might have a negative impact on the accuracy because the body size estimated by our method was occasionally incorrect. On the other hand, clothing information is one of the most effective features for person identification tasks as context information. Gallagher and Chen [4] have reported the effectiveness of clothing information to recognize people in images. Yamaguchi et al. [12] have applied clothing features as context information to a person identification task for partially occluded images. If our system handles the clothing information as context, it leads to the improvement of the accuracy. This is future work of our research.

The hair color was the weakest single feature for the task (28.3 in Table 1). The reason was that many people in Asia have black hair. On the other hand, it was effective for particular persons with color-treated hair.

The hair whorl was not always effective for the identification although it was essentially a discriminating feature. It was caused by the detection process of the hair whorl. The method deleted hairstyle information by using a smoothing filter for the hair whorl detection. However, the smoothed images often contained noise, namely parts of the hairstyle edge. As a result, the feature did not contribute to improve the accuracy. To achieve higher accuracy, we need to consider the method for the hair whorl detection process.

Our method in this paper did not deal with the rejection task. In other words, it handled only the registered persons. The main purpose of our system is to obtain location information of each person belonging to particular groups, such as members of a laboratory, for attendance management. To detect people who are not registered, namely outsider and suspicious individuals, we need to apply another approach.

Moreover, our system handled only the situation that people stop in front of a door. Although this precondition is natural, we need to expand the target into walking or moving persons, for constructing a more flexible system. This is also future work for our system.

5. Conclusions

In this paper, we proposed a novel person identification task and method. The main target of the person identification was a small-group environment such as offices and labs. The method is valuable for many systems, such as attendance management, person retrieval and behavior recognition. We focused on images from an overhead camera. By using the overhead camera, the problems of occluded images and privacy were solved.

Our system classified images with four features extracted from the person's area detected by the background subtraction. We employed the AdaBoost algorithm with C4.5 as the classifier. We obtained 92.5% with the method using "body size", "hair color" and "hairstyle". The experimental result shows the effectiveness of our task and method. On the other hand, the "hair whorl" feature did not lead to the improvement of the accuracy although it was expected to increase the accuracy as a robust feature. The reason was that the correctness of the hair whorl detection was insufficient. We need to improve the hair whorl detection method for a higher accuracy.

Future work includes (1) improvement of the preprocessing especially the hair whorl detection method and (2) evaluation of the method with a large-scale dataset and other datasets, (3) applying context features of each person, such as accessories (e.g., bags), type of clothes (e.g., T-shirt and down jacket) and clothing information (colors and patterns) into the method, and (4) handling information of moving persons.

References:

[1] N. Dalal and B. Triggs. "Histograms of Oriented Gradients for Human Detection," Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 886-893, 2005.

[2] Y. Freund and R. E. Schapier. "Experiments with a new boosting algorithm, " Proc of ICML pp. 148–156, 1996.

[3] T. Fukuzoe, M. Itou, D. Mito and M. Watanabe. "A peron identifica-

tion method unifying image features of body shape and daily habit based on probabilistic recognition framework," IEICE TRANSACTIONS on Information and Systems, D Vol.91-D, No. 5, pp. 1369-1379, 2008.

[4] A. C. Gallagher and T. Chen. "Using Context to Recognize People in Consumer Images," IPSJ Transactions on Computer Vision and Applications, Vol. 1, pp. 115-126, 2009.

[5] Y. Iwashita and A. Stoica. "Gait Recognition using Shadow Analysis," Proc. of Symposium on Bio-inspired, Learning, and Intelligent Systems for Security 2009, pp. 26-31, 2009.

[6] T. Kanade. "Picture processing by computer complex and recognition of human face,' Technical report, Kyoto University, Dept. of Information Science, 1973.

[7] M. Onishi and I. Yoda. "Visualization of Customer Flow in an Office Complex over a Long Period," Proceedings of International Conference on Pattern Recognition (ICPR), pp. 1747-1750, 2010.

[8] J. R. Quinlan. "C4.5 Programs for Machine Learning," Morgan Kaufmann Publishers, 1993.

[9] M. Turk and A. P. Pentland. "Eigenfaces for recognition," Journal of Cognitive Neuroscience, Vol. 3, No. 1, pp. 71-86, 1991.

[10] H. Uchida, R. Inoda, T. Tsuji and S. Abe. "Counting people and recognizing wheelchairs at elevator lobby by real-time image processing," IEEJ Transactions on Industrial Applications, Vol. 129, No. 6, pp. 578-584, 2009.

[11] S. Watanabe and N. Pakvasa. "Subspace method in pattern recognition," Proc. of 1st Int. J. Conf on Pattern Recognition, pp. 2-32, 1973.

[12] J. Yamaguchi, K. Shimada, S. Enokida, T. Ejima and T. Endo. "Personal Identification Using Facial Feature and Context," Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol. 23, No. 2, pp. 13-21, 2011.