

Person Identification Using Top-View Image with Depth Information

Daichi Kouno, Department of Artificial Intelligence, Kyushu Institute of Technology, Iizuka, Fukuoka, Japan

Kazutaka Shimada, Department of Artificial Intelligence, Kyushu Institute of Technology, Iizuka, Fukuoka, Japan

Tsutomu Endo, Department of Artificial Intelligence, Kyushu Institute of Technology, Iizuka, Fukuoka, Japan

ABSTRACT

In this paper, the authors 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. The authors focus on an image from an overhead camera. The authors utilize depth information for the identification task. By using depth information, the authors can capture the precise person's area and rich information for the identification task as compared with popular RGB cameras. The authors apply four features extracted from images based on depth information to the identification method; (1) estimated body height, (2) estimated body dimensions, (3) estimated body size and (4) depth histogram. In the experiment, the authors evaluated two situations; (a) standing in front of a door and (b) touching a doorknob. The identification accuracy rates are 94.4% and 91.4% on the two situations. The authors obtained the high accuracy by the proposed method.

Keywords: *Computer Vision, Depth Information, Identification Tasks, Person Identification, Top-View Images*

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, such as fingerprint

recognition (Jain et al., 1997), iris recognition (Wildes, 1997) and gait recognition (Kale et al., 2004). The most famous approach of them is to use face information (Zhao et al., 2003). Kanade (1973) has proposed a method using face feature points, such as nose and eyes. The CLAFIC method (Watanabe & Pakvasa, 1973) and the EigenFace method (Turk & Pentland,

DOI: 10.4018/ijsi.2013040106

1991) 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. Figure 2 shows the installation location in our method. We focus on the situation that people stop in front of a door. We handle two situations; (1) standing in front of a door and (2) touching a doorknob. The standing situation occurs in such environments as the front of an automatic door and in an elevator. The touching situation occurs to open a door.

On the other hand, information from overhead camera is not always enough. The lack of information for the identification method leads to the decrease of the accuracy. In this paper, we introduce depth information to the task. The captured image in Figure 2 includes the

depth information. The color values denote the distance from the overhead camera. Our method extracts four features from each image; (1) body height, (2) body dimensions, (3) body size and (4) depth histogram. We apply the four features into the AdaBoost algorithm.

RELATED WORK

Snidaro et al. (2005) have reported a method for tracking and counting people from multiple top-view cameras placed at different locations. Uchida et al. (2009) 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 (2010) 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. Yuan et al. (2011) have proposed a pedestrian detection method using a top-view camera for people counting. They selected the best suitable classifier from several classifiers for the task on the basis of a head candidate location. The purpose of our method is to identify a person in an image.

Iwashita and Stoica (2009) have proposed a gait recognition method using overhead images. They focused on shadow images for the gait recognition. We use four features such as

Figure 1. The occlusion problem and the solution

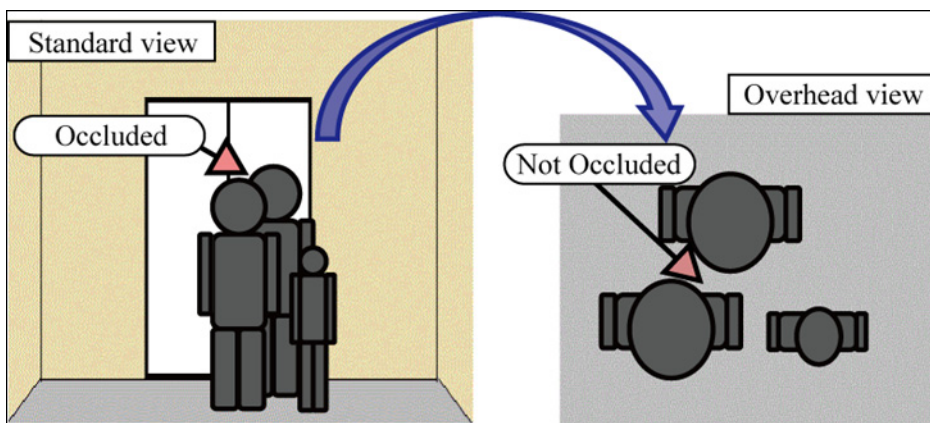
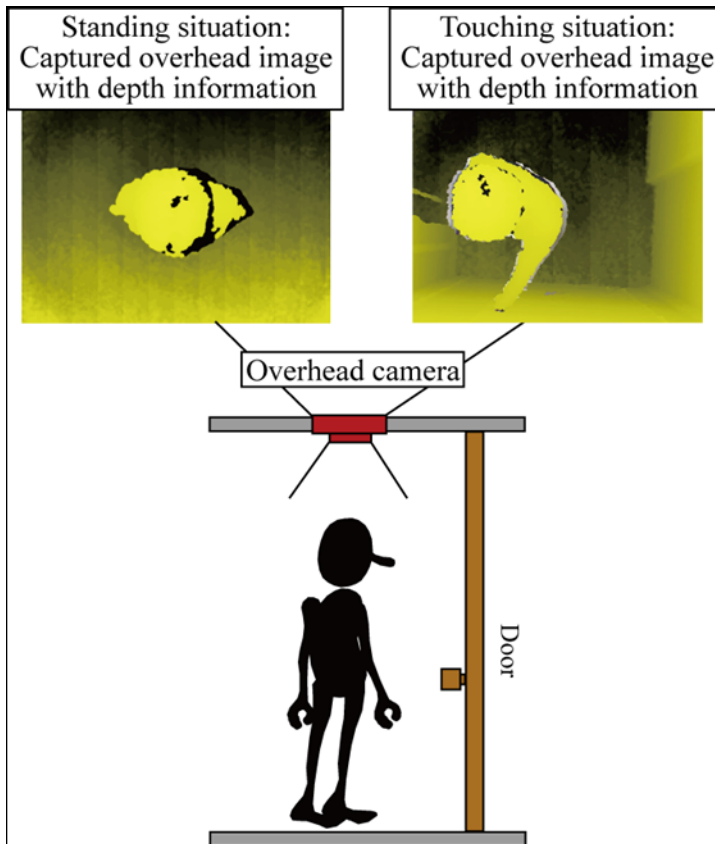


Figure 2. The setting of our task



body height and dimensions for the identification. Fukuzoe et al. (2008) 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. Nakatani et al. (2011) have reported a method using overhead images. However, they used only a simple RGB image. As a result, their method generated occasional mis-detections of the body size in the feature extraction. We use depth information to solve the problem.

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 using depth information, (2) feature extraction and (3) a classifier for person identification. In this paper, we focus on depth information. Although we handle two situations in this paper, the approach to the two situations in the three parts is the same way. Therefore we explain each process with the standing situation only, for the sake of shorthand. We employ Microsoft Kinect (<http://www.xbox.com/en-US/Kinect>) to capture depth information. It is an easy and efficient way to obtain the person's area and depth histograms described below faster and accurately.

Person's Area Extraction

For the person identification, we need to extract the person's area from a captured image. The process is based on background subtraction. The background subtraction approach is generally sensitive to the environment, such as illumination change. To solve this problem, we use depth information from the Kinect camera. In our situation, it captures a person's body and head area and the floor area. There is a significant distance between them. Therefore, we extract the person's area by using a threshold value for the depth values. Figure 4 shows an example of the person's area based on depth information. In the figure, the color strength denotes the depth information.

Feature Extraction

In this section, we describe features for a classifier. In this paper, we handle four features: (1) body height, (2) body dimensions, (3) body size and (4) depth histogram.

Body Height

The body height is one of the most important points to discriminate between persons. We estimate a feature related to the body height of the person in an image by using depth information. The distance between the camera and the top of the head area differs in individuals. We detect the distance of the nearest pixel from the camera in each image. We regard it as the body height feature.

Body Dimensions

The body dimensions are also a intuitive and characteristic feature for the identification. Hence, we introduce the dimension value to our method. In our method, we compute the sum of pixels in the person's area extracted in PERSON'S AREA EXTRACTION.

Body Size

The physical size of the body is one of the most effective features for the person's identification. To be precise, the body size features are the width between shoulders and the width of the head area. 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 marginal distribution of the person's area image. We use the maximum values of the x -axis and y -axis in the person's area. Figure 5 shows an example of the body size feature.

Depth Histogram

The 4th feature is based on depth histograms of each image. It roughly contains height information, such as the distance between the head and shoulder areas. Here the range of the depth information is 0 to 255. We generate the histograms of the depth values from each image. In this feature, we handle each input image including the floor area. Here, note that it is not a person's area image. The reason is that the variation of the background, namely the floor area, often differs greatly in individuals. Figure 6 shows an example of the depth histogram.

Classifier for Identification

Finally we construct a classifier based on four features. In this paper, we employ the AdaBoost algorithm (Freund & Schapier, 1996) 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 (<http://www.cs.waikato.ac.nz/ml/weka/>). We use the C4.5 algorithm (Quinlan, 1993) as the weak classifiers. It is "J48" in Weka. 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.

Figure 3. The outline of our method

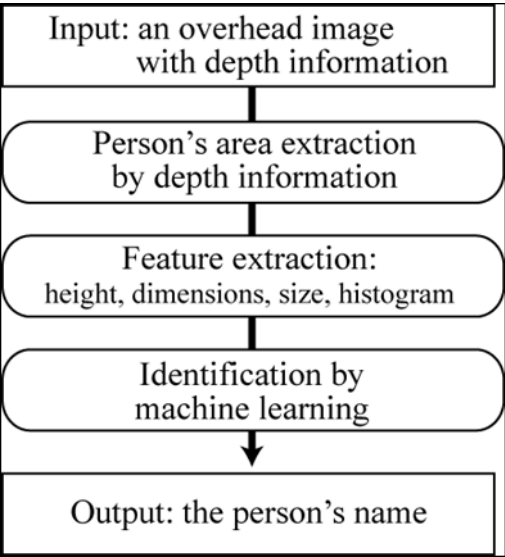


Figure 4. An example of person's area extraction

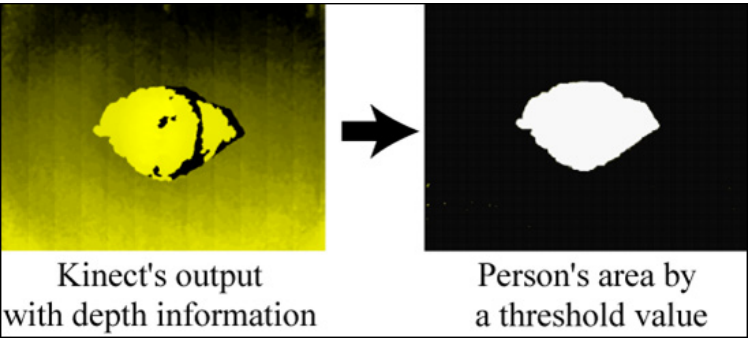


Figure 5. An example of the body size feature

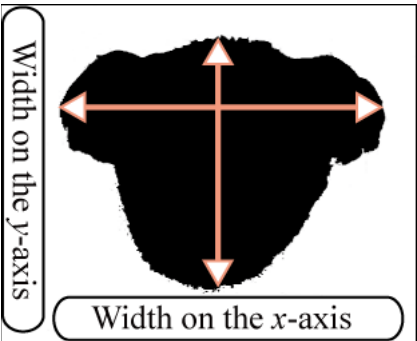
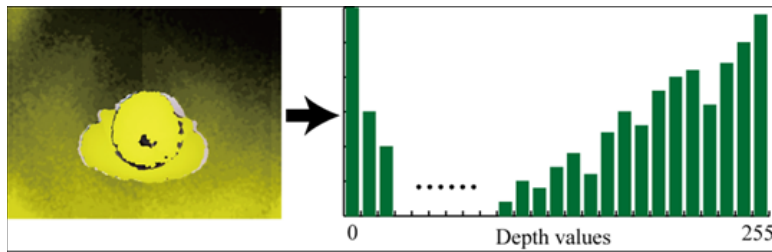


Figure 6. An example of the depth histogram



EXPERIMENT

Settings

In this experiment, we evaluated our method with two situations: standing and touching. The number of test subjects in the standing situation was eight persons. We collected 20 images for each person. As a result, the dataset for the standing situation consisted of 160 images. For the touching situation, we prepared 105 images from seven test subjects, namely 15 images per person. The size of each image in both situations was 640×480 pixels.

Figure 8 shows examples of top-view images of the eight persons for the standing situation. As you can see Figure 8, it is hard to distinguish between them for human beings. For the two situations, we evaluated our method with leave-one-out cross-validation. It extracts one image from the test data, and regards it as

the current test data and the remaining data as the training data. The validation process is repeated for all test data. In other words, it is the same as k -fold cross-validation in the case that the k is the size of test data. The criterion of the evaluation is the accuracy rate. It was computed by

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

Results for Standing Situation

We evaluated our method in the standing situation first. We compared all combination of four features. Table 1 shows the experimental result of some combinations. In the table, “height”, “dimensions”, “size” and “dHist” denote “body height”, “body dimensions”, “body size” and “depth histogram” features in the feature

Figure 7. Adaboost

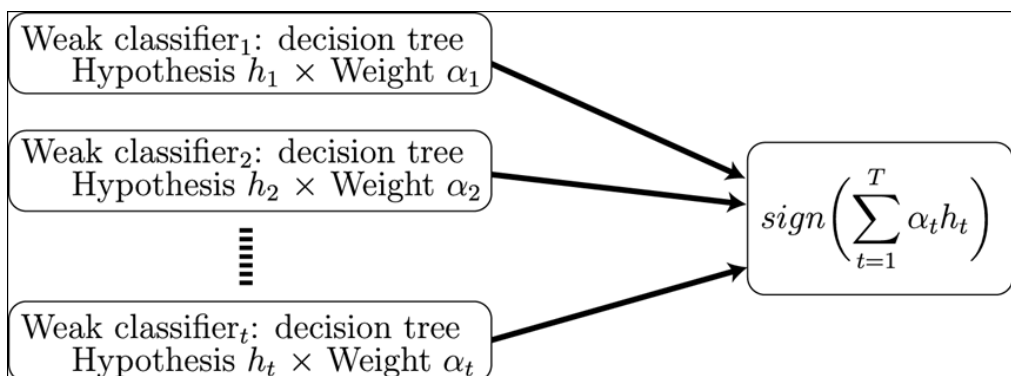
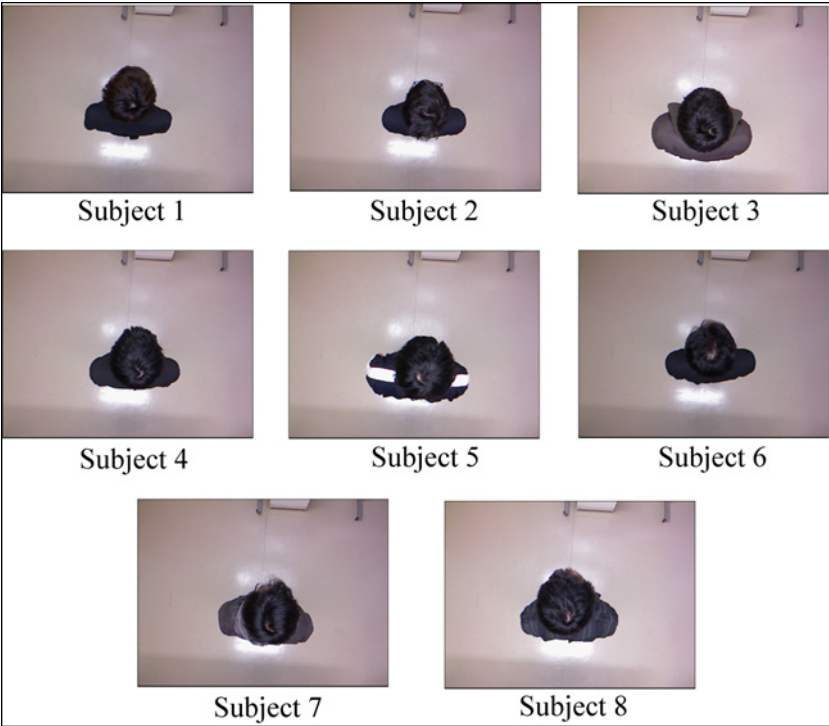


Figure 8. Top-view images in the test data for the standing situation



extraction section, respectively. “+” denotes a combination of features; e.g., “height+dHist” denotes the method with “body height” and “depth histogram” features. “all” is the system with the four features.

The method with two features, “body height” and “depth histogram”, generated the best performance (94.4%). The combination of “body height”, “body size” and “depth histogram” and all features also produced the same accuracy. The results show that the combination of some features was essentially effective. As the single feature, “depth histogram” produced the higher accuracy as compared with the other features.

First, we investigated the result of the best feature set, namely height+dHist, in detail. Table 2 shows the confusion matrix of the method with the feature set. “Sub1” to “Sub8” denote each test subject. Each test subject in the table was stood in order of height. In other words, the Sub1 was the shortest and the Sub8 was the tallest in

all subjects. Mis-identification was generated between persons whose heights were close. For instance, two case between Sub5 and Sub7. To solve this problem, we need to introduce other features, which are not related to body features.

Next, we discuss each single feature. The body height feature was not significant as a single feature. The accuracy was 52.5%. However, the body height is one of the most effective features in essence because it varies with people. The reason why the feature was not effective in the experiment was that the position of each person depended on the situation at the time. If the position of a person in a captured image was different, the distance between the camera and the top of the head was also different. Here, note that the body height feature in our method was based on the distance from the Kinect camera. It was not the actual body height of a person. Figure 9 shows an example of the problem. In the figure, the values denote the distance output from Kinect. The distance

Table 1. The experimental result for the standing situation

Features	Accuracy
height	52.5
dimensions	40.6
size	73.8
dHist	90.6
height+dHist	94.4
height+size	88.1
dimensions+dHist	92.5
size+dHist	90.6
height+dimensions+size	86.9
height+size+dHist	94.4
all	94.4

value was different even if the person was the same in the case that the standing location was different.

One solution for this problem is to normalize the Kinect outputs by using the person's location. Our method can usually estimate the correct person's area with the depth information from the Kinect. We can compute the correct distance by a normalize function with the correct person's area. By the normalization process, the body height feature might generate higher accuracy if we can obtain the appropriate distance. Figure 10 shows an example of the process. After this experiment, we evaluated the method with this normalization process.

As a result, the accuracy by the body height dramatically increased to 74.4%. This result shows the effectiveness of the normalization process.

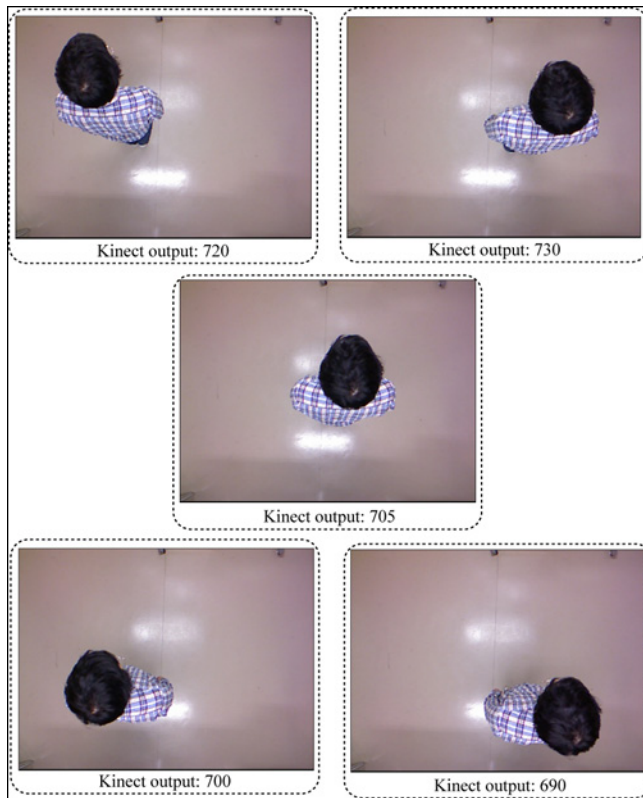
The body dimension feature was also sensitive for the person's location in the image. As a result, the accuracy was the lowest in all single features (40.6%). The normalization process for this feature is of the same importance as the body height feature.

The body size feature was relatively robust as compared with the body height and dimension features. Nakatani's work (Nakatani et al., 2011) also reported the same tendency. The reason was that the variation of the values of the body

Table 2. The confusion matrix of the best features set (standing): height+dHist

Subject	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7		Sub8
Sub1	20	0	0	0	0	0	0	0	
Sub2	1	19	0	0	0	0	0	0	
Sub3	0	0	19	0	0	1	0	0	
Sub4	0	1	0	18	0	1	0	0	
Sub5	0	0	0	1	18	1	0	0	
Sub6	0	0	0	0	0	20	0	0	
Sub7	0	0	1	0	2	0	17	0	
Sub8	0	0	0	0	0	0	0	20	

Figure 9. An example of the difference of the Kinect outputs



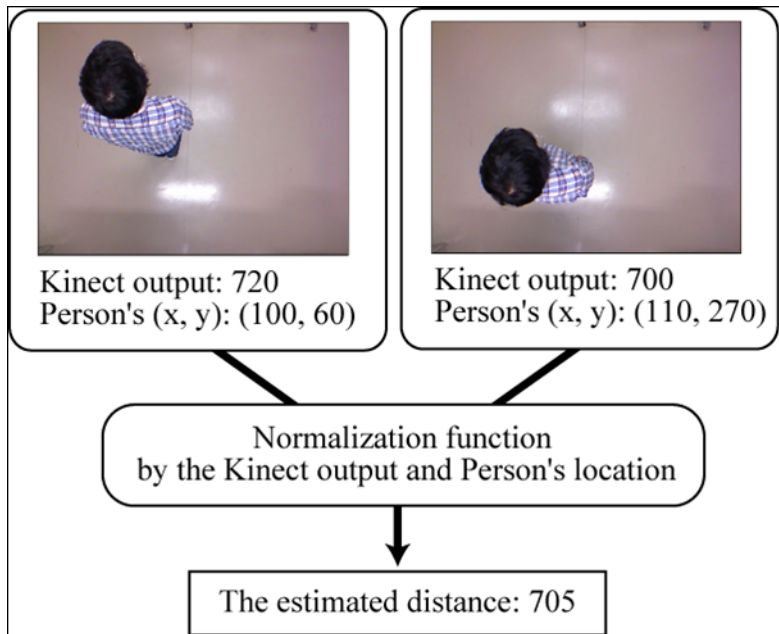
size, namely x and y sizes of the person's area, was usually small in the various positions as compared with the body height and dimension features. We think that the normalization for the body size feature also leads to the improvement of the accuracy.

Results for Touching Situation

Next, we evaluated our method in the touching situation. We also compared all combination of four features. Table 3 shows the experimental result of some combinations for the touching situation. The method with two features, "body height" and "depth histogram", generated the best performance (91.4%). As the single feature, "depth histogram" produced the higher accuracy as compared with the other features. These results show the same tendency of the standing situation.

As a single feature, the "body height" feature was dramatically improved in this situation (52.5% in Table 1 vs. 88.6%3). The problem of "body height" was the position of each person. It was often different. In the standing situation, we solve this problem by using a normalization function. On the other hand, the positions of each person were relatively stable in the touching situation. In this situation, test subjects touch a doorknob to open the door. Therefore, our method could capture similar positions from each image as compared with the standing situation. In this situation, location information of each person might be a new feature for the identification because it is different for each person. Although the height feature was effective, the accuracy of the size feature decreased dramatically. The reason was that our method could not capture the body size,

Figure 10. An example of the normalization process



namely x and y , correctly because the direction of each person became diagonally forward (See the upper right image in Figure 2). Therefore, our method did not estimate the correct body size. To achieve higher accuracy, we need to solve the problem.

Table 4 shows the confusion matrix of the method with the feature set. Each test subject in the table was stood in order of height in the

same manner as Table 2. Most of mistakes were also generated in the case that the heights of persons were close, e.g., *Sub3* and *Sub4*. To solve this problem is our important future work.

Discussion

In this section, we discuss the experimental results for both the situations and the

Table 3. The experimental result for the touching situation

Features	Accuracy
height	88.6
dimensions	48.6
size	47.6
dHist	89.5
height+dHist	91.4
height+dimensions+size	89.5
dimensions+size+dHist	91.4
height+size+dHist	91.4
all	91.4

Table 4. The confusion matrix of the best features set (touching): height+dHist

Subject	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7
Sub1	15	0	0	0	0	0	0
Sub2	1	14	0	0	0	0	0
Sub3	0	1	12	2	0	0	0
Sub4	0	0	2	12	0	1	0
Sub5	0	0	0	0	15	0	0
Sub6	0	0	0	1	0	14	0
Sub7	0	0	0	0	1	0	14

effectiveness of our system. The body dimension and size features are sensitive for clothing change. Handling the clothing information is one of the most important issues for identification tasks (Hossain et al., 2010). If clothes that target person were changed, it might have a negative impact on the accuracy because the body dimension and size estimated 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 (2009) have reported the effectiveness of clothing information to recognize people in images. Yamaguchi et al. (2011) 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 depth histogram feature was the most effective feature as a single feature. The feature contains the several distance information such as the distance between the camera and head, the distance between head and shoulder areas, and the relative distance between the camera and the floor area. In other words, the feature expressed the relative range difference computed from several areas. Therefore, it was effective by combining the body height, which expressed the absolute distance from the camera. On the other hand, the histogram feature was also a sensitive feature for the person's location. If the location of the person changed, the distance

between each part also changed. It was the main reason for the misidentification by the histogram feature. Moreover, the normalization process might become more complex as compared with the height feature because the histogram feature contains many values. The normalization method for the depth histogram feature is one of the important issues in our method.

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.

In human surveillance, tracking of people is one of the most important tasks. Person re-identification is the task of recognizing an individual in images from non-overlapping cameras (Farenzena et al., 2010), (Gray & Tao, 2008). Applying our method focusing on top-view head images to the re-identification task is interesting future work.

CONCLUSION

In this paper, we proposed a person identification method with top-view images and depth information. By using the overhead camera, the problems of occluded images and privacy were solved. In addition, depth information was useful for the top-view image based person identification task. We handled two situations; standing situation and touching situation.

Our system classified images with four features; body height, body dimensions, body size and depth histogram. We employed the AdaBoost algorithm with C4.5 as the classifier. For standing situation, we obtained 94.4% with the method with “body height” and “depth histogram”. The experimental result shows the effectiveness of our task and method. The critical problem was the variation of the distance value from Kinect. It was caused by the person’s location in captured images. The difference of the location generated the difference of the distance value. As a result, our method estimated the distance values for features incorrectly. Applying a normalization process to the height feature led to the improvement of the accuracy (52.5% to 74.4% as a single feature). This result shows the significance of the normalization process. It is important future work, especially the depth histogram feature. On the other hand, for the touching situation, the problem of the location was least disruptive to the body height feature.

The body dimension and size features are sensitive for clothing change. On the other hand, clothing information has an important role in person identification as context information. Applying the context features to our method is also our future work.

We evaluated our method with 160 images for eight persons in the standing situation and 105 images for seven persons in the touching situation. However, the test data set was too small. We need to verify the validity of our method by using a large dataset, namely many persons and many images.

REFERENCES

- Farenzena, M., Bazzani, L., Perina, A., Murino, V., & Cristani, M. (2010). Person reidentification by symmetry-driven accumulation of local features. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2360-2367).
- Freund, Y., & Schapier, R. E. (1996). Experiments with a new boosting algorithm. In *Proceedings of ICML* (pp. 148-156).
- Fukuzoe, T., Itou, M., Mito, D., & Watanabe, M. (2008). A person identification method unifying image features of body shape and daily habit based on probabilistic recognition framework. *IEICE Transactions on Information and Systems*, D 91(5), 1369-1379.
- Gallagher, A. C., & Chen, T. (2009). Using context to recognize people in consumer images. *IPSI Transactions on Computer Vision and Applications*, 1, 115-126. doi:10.2197/ipsjtva.1.115.
- Gray, D., & Tao, H. (2008). Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Proceedings of European Conference on Computer Vision (ECCV)* (pp. 262-275).
- Hossain, M. A., Makihara, Y., Wang, J., & Yagi, Y. (2010). Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition*, 43(6), 2281-2291. doi:10.1016/j.patcog.2009.12.020.
- Iwashita, Y., & Stoica, A. (2009). Gait recognition using shadow analysis. In *Proceedings of Symposium on Bio-inspired, Learning, and Intelligent Systems for Security 2009* (pp. 26-31).
- Jain, A., Lin, H., & Bolle, R. (1997). On-line fingerprint verification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(4), 302-314. doi:10.1109/34.587996.
- Kale, A., Sundaresan, A., Rajagopalan, A. N., Cuntoor, N. P., Roy-Chowdhury, A. K., Kruger, V., & Chellappa, R. (2004). Identification of humans using gait. *IEEE Transactions on Image Processing*, 13(9), 1163-1173. doi:10.1109/TIP.2004.832865 PMID:15449579.
- Kanade, T. (1973). *Picture processing by computer complex and recognition of human face. Technical report, Kyoto University.* Dept. of Information Science..
- Nakatani, R., Kouno, D., Shimada, K., & Endo, T. (2011). A person identification method using a top-view head image from an overhead camera. In *Proceedings of 2nd International Workshop on Advanced Computational Intelligence and Intelligent Informatics (IWACIII2011)* (pp. SS4-1).
- Onishi, M., & Yoda, I. (2010). Visualization of customer flow in an office complex over a long period. In *Proceedings of International Conference on Pattern Recognition (ICPR)* (pp. 1747-1750).
- Quinlan, J. R. (1993). *C4.5 programs for machine learning.* Morgan Kaufmann Publishers..

- Snidaro, L., Micheloni, C., & Chiavedale, C. (2005). Video security for ambient intelligence. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans*, 35(1), 133–144. doi:10.1109/TSMCA.2004.838478.
- Turk, M., & Pentland, A. P. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 71–86. doi:10.1162/jocn.1991.3.1.71 PMID:23964806.
- Uchida, H., Inoda, R., Tsuji, T., & Abe, S. (2009). Counting people and recognizing wheelchairs at elevator lobby by real-time image processing. *IEEE Transactions on Industrial Applications*, 129(6), 578–584. doi:10.1541/ieejias.129.578.
- Watanabe, S., & Pakvasa, N. (1973). Subspace method in pattern recognition. In *Proceedings of 1st International Journal Conference on Pattern Recognition* (pp. 2-32).
- Wildes, R. P. (1997). Iris recognition: An emerging biometric technology. *Proceedings of the IEEE*, 85, 1348–1363. doi:10.1109/5.628669.
- Yamaguchi, J., Shimada, K., Enokida, S., Ejima, T., & Endo, T. (2011). Personal identification using facial feature and context. *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, 23(2), 13–21. doi:10.3156/jsoft.23.137.
- Yuan, X., Wei, X.-Y., & Song, Y.-D. (2011). Pedestrian detection for counting applications using a top-view camera. *IEICE Transactions on Information and Systems*, 94(6), 1269–1277. doi:10.1587/transinf.E94.D.1269.
- Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys*, 35(4), 399–458. doi:10.1145/954339.954342.