Identification of a Specific Person using Color, Height, and Gait Features for a Person Following Robot

K. Koide[a,*], J. Miura[a]

[a]*Department of Computer Science and Engineering, Toyohashi University of Technology, Toyohashi, Aichi, 441-8580, Japan*

Abstract

This paper describes a person identification method for mobile service robots using image and range data. Person identification is a necessary function in order for mobile service robots to locate the target person for those services. Among various sensory features, image-based appearance features have often been used for person identification. They are, however, not effective in severe illumination environments such as a strong backlight. Therefore, we use two illumination-independent features, height and gait, in addition to appearance features for a more robust identification. To this end, we have developed a new method of extracting the gait feature (step length and speed), based on a maximum likelihood estimation of supporting leg positions in accumulated range data. We combine these features and use an online boosting approach to create the specific person classifier. It allows the robot to identify the specific person robustly even in a severe illumination environment. We tested our multi-feature person identification method, combined with a range data-based person tracker, in a specific person following scenario to demonstrate the effectiveness of this method.

Keywords: Multi-feature person identification, gait feature, mobile service robots.

1. Introduction

There is an increasing demand for service robots which can attend and support a person. Such robots are expected to provide services, such as guiding, guarding and elderly care. Person following is a necessary function of personal service robots, in order to provide such services to a specific person. Fig. 1 shows an example of following behavior. To follow a person, the robot has to be able to continuously identify the person, and the identification function has to be robust and usable in both indoor and outdoor environments since persons move across these environments in their daily life.

One way is identification by tracking, that is, to track a person over time and identify him/her based on motion continuity. The person tracking function has been widely studied in many published works. Especially for mobile robots, laser range finders (LRFs) are often used for person detection and tracking thanks to their reliability and wide field of view [1, 2, 3]. These works detect persons from range data and track the persons

based on their positions. However, if a person is occluded by other persons for several seconds, the robot may lose track of the target person or track another person erroneously. It is, therefore, necessary to identify a specific person from only the sensory features obtained for a frame or a short period.

Image-based tracking methods (e.g., [4, 5]) have an advantage in that various information for identification can be obtained while tracking. Possible types of information include color of clothing [6], height [7], face [8],

Figure 1: Person following robot.

---

*Corresponding author
   Email addresses:* `koide@aisl.cs.tut.ac.jp` (K. Koide), `jun.miura@tut.jp` (J. Miura)

(a) outdoor.   (b) indoor.

Figure 2: Extremely severe illumination environments which mobile robots may face.

gait [9, 10], and skeletal information [11].

Many appearance features are used in image-based identification, for example HSV, Lab and XYZ color space histogram [12, 13], Haar-like [14], HOG [15], LBP [15] and SIFT [16] features. Those features are, however, not applicable to severe illumination environments such as a strong backlight or darkness, where colors and edges are not reliably obtained. It is therefore necessary to combine other features, including those from other sensors, for more robust identification.

Person re-identification has been extensively studied in computer vision domain [17, 18, 19, 20, 21]. In these works multiple non-overlapping cameras are put in an environment, and persons are tracked and identified across the cameras for surveillance. Recent person re-identification methods use novel identification procedures, such as automatic discriminative video fragments selection [19]. These methods significantly improve identification performance under observation noise and illumination changes among cameras. However, it is difficult to apply these methods to mobile robots directly for the following two reasons. First, as a robot moves, the background changes and often becomes very complicated. This makes it difficult to obtain an accurate foreground mask for each person and eliminate effects from the background. Secondly, Most of person re-identification methods rely on appearance features. In the case of mobile robots, since cameras are put at a lower height than in surveillance systems, lights or the sun may easily come into sight (see Fig. 2, for example), and mobile robots may face to extremely severe illumination environments where appearance features are significantly degraded.

Person identification using gait analysis has recently become popular [9, 10, 22]. These works extract and use frequency components from silhouette images of a walking person for identification. Since they also assume a static background, these methods cannot be directly applied to mobile robots. Little has been proposed for gait analysis using range data [23, 24, 25].

Cifuentes et al. [23] measured the gait features, such as leg distance and leg orientation, from a mobile robot to realize a smooth human-robot interaction. The relative position between the robot and the person is, however, very limited for avoiding that legs are occluded by the opposite leg. Nakamura et al. [24] and Song et al. [25] put several LRFs on the ground and extracted the gait feature from these data. Since a mobile robot has a single viewpoint, a leg is often occluded by the other leg; the measured gait may be degraded due to this occlusion.

Height features are used for well calibrated and fixed cameras [7]. Since the height of a person is fixed and specific to the person, it is suitable for person identification. In the case of mobile robots, however, it is difficult to measure the height of a person using only one camera because the distance to the person can change largely. In order to use the height feature for mobile robots, another sensor which provides the distance is necessary.

Devices for identification, such as radio frequency identifier (RFID) tags [26, 27] or inertial measurement units (IMUs) [28], are sometimes used by a robot to locate a target person who has the devices. Although using such devices makes target localization easier, it requires users to wear the devices every time when they require services, and this may be inconvenient to the users.

In this paper, we propose a method of robustly identifying a specific person using LRFs and cameras. In order to ensure the redundancy of features in identification, we introduce two illumination-independent features, height and gait, in addition to appearance features. We combine these features to realize a robust person identification even in severe illumination environments. We validate by experiments that these two features greatly increase the identification robustness. The contributions of this paper are threefold. First, we provide a person identification method which adopts gait and height features as well as color features for a robust identification. Secondly, we propose a gait estimation method using LRF. The method extends Nakamura's method [24] so that occlusion between legs are explicitly considered in a maximum likelihood estimation. Thirdly, we propose a new joint feature approach for combining multiple features with different observation cycles.

Fig. 3 shows an overview of the proposed system. The robot is equipped with two-layered laser range finders (Hokuyo UTM-30LX) set at a torso and a leg height and web cameras, and the maximum speed of the robot is about 1.2 [m/s]. The method first uses range data from LRF for tracking persons, and then identifies a
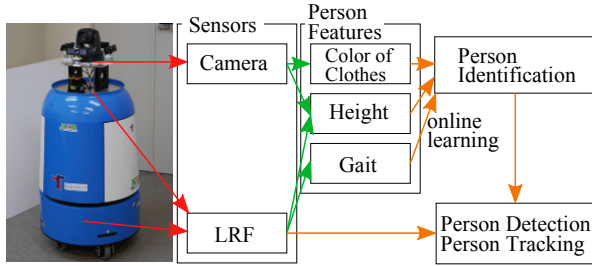
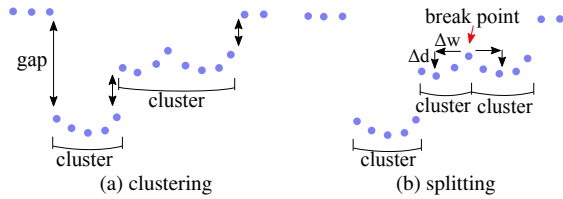Figure 3: Person tracking and identification system.

Figure 4: Torso and leg detection procedure.

Figure 5: Detected torso candidates. Each green circle indicates the position of a torso candidate.

specific person. The color feature is extracted from images while the gait feature is extracted from range data. The height feature is obtained by combining images and range data. The proposed method combines these features in order to identify the person in any environmental conditions.

The rest of the paper is organized as follows. Sec. 2 briefly explains a person detection and tracking method using LRF data. Sec. 3 describes an overview of our person identification framework using multiple features. Sec. 4 describes person identification using color and height features. Sec. 5 describes gait extraction and its application to person identification. Sec. 6 describes evaluation of the proposed person identification method. Sec. 7 describes a person following strategy and experimental results. Sec. 8 concludes the paper and discusses future work.

2. Person Tracking

2.1. LRF-based Person Tracking

Multiple layered LRFs are sometimes used for human detection [29]. Typically these sensors are put at the height of torsos and legs, and then both detection results are combined. They assume that the torso of a person is always detected, and if one or two legs are found under a torso, the torso is judged as a true positive. By combining detection results of multiple layered LRFs, we can reduce the number of false positives.

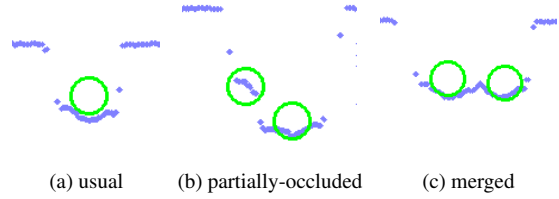Torsos and legs are typically detected as a segment separated from background by finding gaps in range data [1, 3]. However, in populated environments, torsos and legs are not always separated from background or another torso/leg. They are also often partially occluded by another objects. Our method first detects gaps of range data for clustering (see Fig. 4(a)), and then finds break points of merged torsos/legs in range data using two threshold values Δw and Δd (see Fig. 4(b)). For a point in a cluster, if two points separated from the point by Δw on both sides are closer by Δd to the robot than the point, the point is treated as a break point, and the cluster is split at that point. We then apply a size filtering to all the clusters to detect torso/leg candidates. Fig. 5 shows examples of detected candidates for torso.

The detected candidates are classified into torso/leg and other objects using Arras's method [2] and Zainudin's method [3], respectively. Features which represent the shape of the clusters are extracted, and then the classification is performed by machine learning method, such as SVM [30] and Adaboost [31].

We adopt a simple procedure for temporal data association of detected persons, based on a Kalman filter with a constant velocity model and a nearest neighbor (NN) data association. This works well in the majority of tracking cases. If a person is occluded by another person for several seconds, however, it often fails to track the person due to an incorrect data association. We thus take occlusions of persons into account in data association as follows.

We model each person by a circle located at the position predicted by the Kalman filter, and test whether it is occluded or not. We first predict the range data which should be obtained from the circle, and then the predicted range data for the circle are compared with the actual observed range data. If more than a half of the actual range data are closer to the robot than the predicted range data, the person is considered as occluded. The occluded persons are not associated with the detected persons to prevent incorrect data association.
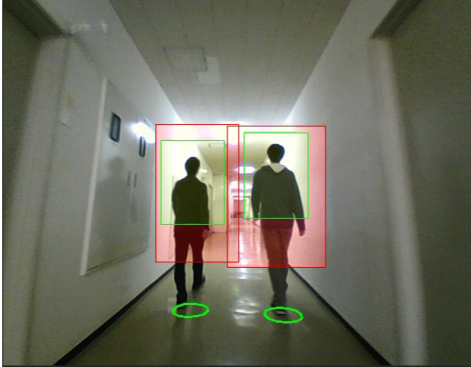
Figure 6: Detecting the person region. The green circles indicates the person position obtained by the LRF-based tracking. The red transparent regions are the ROI calculated from the person position. The green transparent regions are the detected person regions.

2.2. *Detecting Person Region on the image*

A person region on an image is required to extract features for person identification. We first calculate a Region of Interest (ROI) from a person position obtained by the LRF-based tracking, and then detect the upper body of the person from the ROI using the cascaded HOG classifier [32]. To calculate the ROI, we model the person as a cylinder located at the person position and project the cylinder into the image (see Fig. 6). The detected regions are used for extracting the person features.

3. Person Identification Framework

In this section, we briefly describe our person identification framework and a joint feature approach. To identify the person, we employ the color, the height, and the gait feature as it will be explained in detail in Sec. 4 and 5. Those features are merged into a joint feature and learned by online boosting [33].

3.1. *Person identification using online boosting*

Online boosting is one of the online learning methods which constructs an ensemble of weak classifiers and uses it as a strong classifier. This method has been used for people tracking due to its adaptability and real time performance [14, 34]. In our case, each weak classifier uses only one of the three features. Since online boosting selects the best weak classifiers, only the effective features are used for person identification. For example, when they are in a severe illumination environment, the color feature is not effective and only the height and the gait features are used in the classifier. As a result, we can obtain a reliable person classifier even in a severe illumination environment (see experiments in Sec. 6).

While the target person is tracked by the LRF-based tracker, the person classifier is updated with observed features. While the LRF-based tracker is losing the target, updating of the person classifier stops and the robot looks for the target person using the latest person classifier. If a person who is judged as the target person by the classifier is found, the robot sets the person as the target to track, and resumes tracking.

3.2. *Joint feature for online boosting*

The observation cycles of the proposed features are largely different. To apply the features to online boosting, they have to be synchronized and merged into a single joint feature. There are basically two approaches to synchronize features: synchronizing features to the one with the shortest or the longest cycle.

If we take the first approach, while the feature with the shortest cycle is varying, features with long cycles are kept constant or interpolated. Weak classifiers using those with long cycles are updated by one observation until a new observation is obtained. It may cause an overfitting. We thus take the second approach.

In a traditional way for the second approach, every time the feature with the longest cycle is obtained, latest feature values are simply concatenated to construct a feature vector [35]. In our system, however, the observation cycles of the proposed features are very different from each other; those of the color and the height feature are about 30 msec long, while that of the gait feature is about 500 msec long. By using only the latest feature values, a large amount of observations with short cycles are discarded (see Fig. 7(a)), and the identification result may be degraded. Therefore, we also make use of the values of features with shorter cycles obtained during an interval of the feature with the longest cycle by calculating their statistics and concatenating them with the feature with the longest cycle (see Fig. 7(b)).

In this paper, the statistics of two features, the height and the gait, are calculated. Since the height of a person is fixed, the distribution of the height feature can be expected to be unimodal. On the other hand, if we observe a person for a while, the distribution of the color feature may become multimodal due to illumination changes. However, in our case, the duration for summarizing the features is about 0.5 [sec] (i.e., the observation cycle of the gait feature). We assume that the duration is small enough to model the distribution of the color features as unimodal. We thus employ mean and standard deviation to summarize the features.
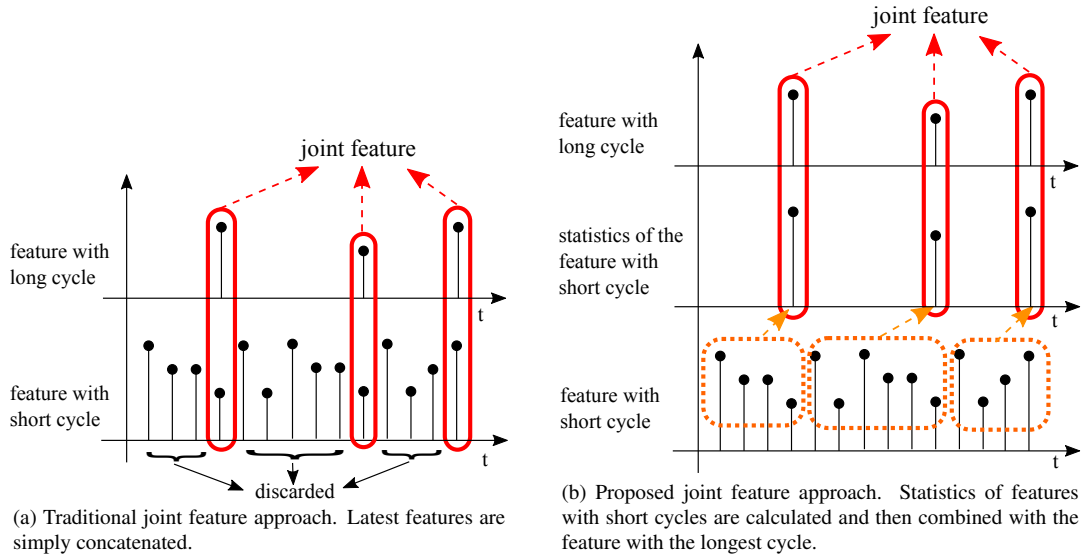
(a) Traditional joint feature approach. Latest features are simply concatenated.

(b) Proposed joint feature approach. Statistics of features with short cycles are calculated and then combined with the feature with the longest cycle.

Figure 7: Comparing joint feature approaches.

4. Image-based Person Identification

4.1. Color feature

Color features can easily be extracted from an image and are effective for identifying a person by their clothing color [12, 14, 34]. Texture and shape features, such as HOG [15] and SIFT [16], are also used for a more robust person identification. However, all such appearance features are weak under severe illumination environments [17]. We thus use only a color feature as an appearance feature for simplicity.

Color histogram is one of the most popular representations for color modeling. We use a hue-saturation histogram (HS-histogram) to reduce the effect of light intensity changes. To obtain a histogram, we follow Luber's histogram extraction approach [14]. A HS-histogram is constructed from pixels in a rectangular region with randomized positions and sizes in the person region. By online boosting [33], histogram extraction regions are sampled randomly, and regions with better identification rate are used for constructing an ensemble of classifiers. An example of histogram extraction regions generated by online boosting is shown in Fig. 8.

4.2. Height feature

The height of a person can be used as another feature for person identification. Even if there are multiple persons with similar heights, the height is useful for reducing the number of candidates for the target person. To calculate the height of a person, we first determine the topmost position (i.e., *sinciput* of the head region) in

Figure 8: An example of regions for color histogram extraction selected by online boosting. The blue rectangles are the regions for hue histogram extraction, and the green rectangles are the regions for saturation histogram extraction.

the image, and then estimate the height using the camera geometry.

A saturation-intensity histogram of a hair region is computed from the hair images in advance, and then a Gaussian mixture model (GMM) is fitted to the histogram. Hair images are collected from about fifty people in various environments, and the total number of hair images are about two hundred. Since we collected hair images from Asian people, most of pixels will be the ones with zero saturation (black or gray pixels). We thus fit a separate univariate GMM to the intensity distribution of zero saturation pixels. The resultant GMM is used as the hair color model (see Fig. 9 and Fig. 10). Currently, the hair color model is specialized for people with black or gray hair. However, the model can be
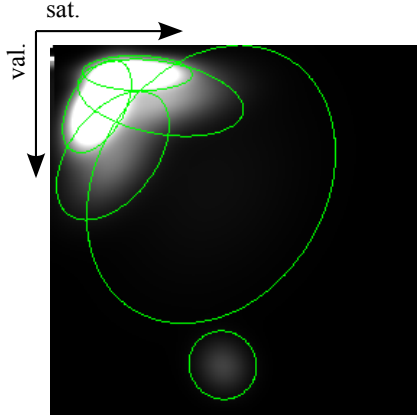
Figure 9: Hair color model. The green circles indicate gaussian distributions. The gaussian in the high value region corresponds to bright pixels caused by direct reflections of light.
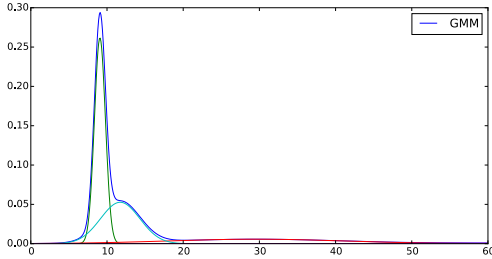

Figure 10: Hair color model for zero saturation pixels. The blue line indicates the hair color model and the rest lines indicate the gaussian distributions which compose the model.

extended for other people by adding their hair images.

We make two images from an input image, one representing the similarity of hair color and the other representing the magnitude of the gradient, and calculate the pixel-wise product of the images. The pixel which has the highest product value is considered as the sinciput of the person (see Fig. 11).

The sinciput position in the image is combined with the person position obtained by the LRF-based tracking to calculate the person height. The relationship between the 3D coordinate relative to the camera (X, Y, Z) and the projected screen coordinate (u, v) in the pinhole camera model is given by

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, \tag{1}$$

where (f_x, f_y) is the focal length, and (c_x, c_y) is the center
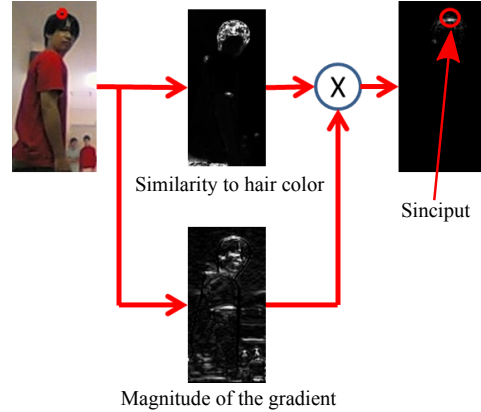

Figure 11: Sinciput detection procedure.

point of the image. From this equation, we obtain:

$$Y = \frac{Z(v - c_y)}{f_y} \tag{2}$$

The depth Z between the camera and the person is obtained by the LRF-based tracking, and v is the sinciput height in the image. By putting these values into eq. (2), we obtain the persons height.

To reduce the effects of a failure of the sinciput detection, we apply a robust estimation to the person height calculation. We adopt the M estimation with Tukey's biweight function [36] to estimate the person's height.

5. LRF-based Person Identification

5.1. Gait feature

In computer vision, gait recognition has been studied widely [9, 10, 37]. It is, however, difficult to apply their methods to mobile robots since they assume a static background to extract silhouette images of a walking person. By using an RGB-D camera, such as Kinect, we can separate the person region from the background region, and then extract gait features [38, 39]. However, Stone et al. reported that the gait analysis using depth images shows a lower accuracy than those using RGB images [38]. Furthermore, infrared depth cameras, like Kinect, are not usable in outdoor scenes.

When a person is walking, the legs of the person swing and stop alternately. The interval when a leg is stopping is referred to as *stance phase*, and the interval when a leg is swinging is referred to as *swing phase* [40]. During the stance phase, the leg which stops and supports the body of the person is referred to as a *supporting leg*. If we can obtain the supporting leg positions (where the leg touches the ground), we can cal-
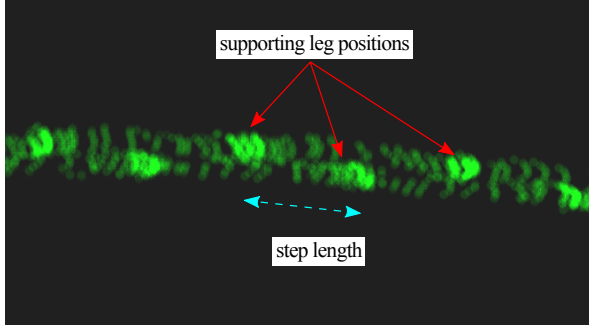
6

Figure 12: Accumulated range data of the legs of a walking person. High-density regions are considered as the over around the supporting legs.

culate gait features, such as a step length and a stance width, from these positions.

Nakamura et al. [24] proposed a method of detecting the supporting leg positions from LRF data. They observed the legs of walking persons by several LRFs from different directions at a railway station and accumulated range data over time. Since the supporting leg positions have high accumulated values, they are extracted by Mean shift method [41]. Fig. 12 shows an example accumulation of range data; supporting leg positions can be found at high-density positions. We basically use their approach but a difference is that a mobile robot has a single viewpoint for LRF. This causes occlusion of supporting legs by the other ones, which may degrade the spotting of supporting leg positions in the accumulated range data.

We thus develop a method of reliably spotting support leg positions based on maximum likelihood estimation which takes such occlusions into account.

Let $X = [x_1, y_1, \cdots, x_n, y_n]$ be positions of supporting legs, $Y = [x'_1, y'_1, \cdots, x'_n, y'_n]$ be their observed positions, and $\Sigma = [\sigma_1^2, \cdots, \sigma_n^2]$ be the observation variances. The Likelihood function L is defined as:

$$L = \prod_{i=1}^{n} \frac{1}{2\pi\sigma_i^2} \exp\left(-\frac{(x'_i - x_i)^2 + (y'_i - y_i)^2}{2\sigma_i^2}\right) \quad (3)$$

We minimize the following objective function J.

$$J = -\log L$$
$$= \sum_{i=1}^{n} \log 2\pi\sigma_i^2 + \sum_{i=1}^{n} \frac{1}{2\sigma_i^2}\left\{(x'_i - x_i)^2 + (y'_i - y_i)^2\right\} \quad (4)$$

Since the step length of a person at a stationary walk is constant [37], we assume that and obtain:

$$(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 = \text{const.}$$
$$i = (1, 2, \cdots, n-1) \quad (5)$$

From this equation, we obtain the following constraint function g_i:

$$g_i = (x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 -$$
$$(x_i - x_{i-1})^2 - (y_i - y_{i-1})^2 = 0 \quad (6)$$
$$i = (2, 3, \cdots, n-2, n-1)$$

According to the method of Lagrange multiplier, we define the following function.

$$F = J - \sum_{i=2}^{n-1} \lambda_i g_i \quad (7)$$

Then we find a set of leg positions which satisfies the following equations:

$$\frac{\partial F}{\partial x_i} = 0, \quad \frac{\partial F}{\partial y_i} = 0, \quad \frac{\partial F}{\partial \lambda_i} = 0 \quad (8)$$

The partial differentiations of F are introduced as following equations. Note that $\lambda_i = 0$ for $i \leq 0$.

$$\frac{\partial F}{\partial x_i} = -\frac{1}{\pi\sigma_i^2}(x'_i - x_i) - 2\lambda_i(x_{i+1} - x_i)$$
$$+ 2\lambda_{i-1}(x_{i+1} - x_{i-1}) - 2\lambda_{i-2}(x_i - x_{i-1}) \quad (9)$$
$$\frac{\partial F}{\partial y_i} = -\frac{1}{\pi\sigma_i^2}(y'_i - y_i) - 2\lambda_i(y_{i+1} - y_i)$$
$$+ 2\lambda_{i-1}(y_{i+1} - y_{i-1}) - 2\lambda_{i-2}(y_i - y_{i-1}) \quad (10)$$
$$\frac{\partial F}{\partial \lambda_i} = x_{i+2}^2 - 2x_{i+1}(x_{i+2} - x_i)$$
$$- x_i^2 + y_{i+2}^2 - 2y_{i+1}(y_{i+2} - y_i) - y_i^2 \quad (11)$$

We use five walking steps for estimation of supporting leg positions and the duration of the observation is about 2.5 [sec]. We assume that the walking speed is constant for this duration.

When the robot observes a walk from a side position, a leg on the robot side is always visible while the other is sometimes occluded. We thus give the observation of the supporting leg on the robot side a small variance (i.e., high reliability) and that of the other leg a large variance (low reliability).

Fig. 13 shows how to determine the side of a supporting leg. We draw a line every two positions and see if the point between them is on the same side as the robot. In the case of the figure, p_{t-1} is given a small variance while p_t a large variance.

We use the pair of step length and walking speed as the gait feature since those are determined by physical characteristics of the person (e.g. weight, height, and lengths of limbs) and specific to an individual [37].
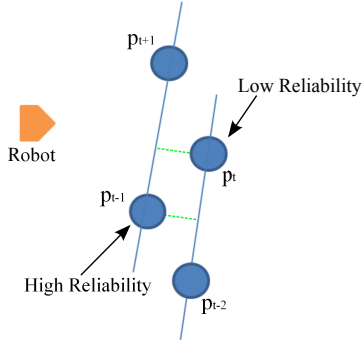
7

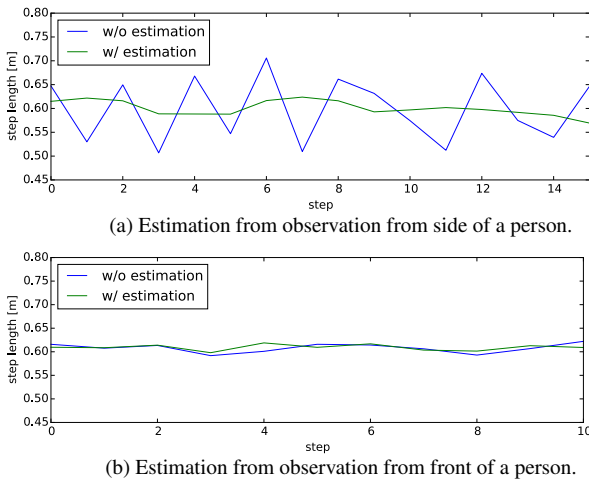Figure 13: Assigning reliabilities to the measurement of supporting legs.

(a) Estimation from observation from side of a person.

(b) Estimation from observation from front of a person.

Figure 14: Estimated step length.

5.2. Gait estimation evaluation

We describe an evaluation of our gait estimation method. We placed markers evenly on the ground every $0.6\,m$ and a person walked by stepping at every marker so that we could obtain a constant step length. The robot observed the walk both from the front and the side of the person for comparison.

When the robot observes the person from the side (see Fig. 14(a)), the measured step lengths fluctuate due to the occlusion of the supporting leg. The effect of occlusion is then largely reduced by the proposed estimation method. On the other hand, when the robot observes from the front (see Fig. 14(b)), the measured step length is much more stable since no occlusions occur.

Table 1 summarizes the evaluation. The fluctuation of the observation from the side is larger than the observation from the front obviously due to the occlusion. The proposed estimation could reduce the fluctuation in the both cases.

5.3. Gait identification experiment

A gait identification experiment was conducted. We recorded gait data of about 30 steps long (about 20 seconds long) for eight persons at a normal walking speed as a training set and constructed a classifier for each person using online boosting [33]. In the experiments, the number of weak classifier selectors is five and each selector contains ten weak classifiers. Fig. 15 shows the gait data for training; we can see some persons (e.g., persons C, E, and H) have distinctive gaits.

We recorded another set of data in the same settings for evaluating the identification performance. Table 2 shows the result of the experiment. The first row indicates the constructed models, and the first column indicates the test data. Each cell indicates the acceptance rate of the test data by the constructed model. For model C and D, the correct person shows a higher identification rate than the others. For model A, B, E, F, and H, the correct person's are the second highest. These results show that the gait feature is mostly effective to identify a person or to reduce the number of possible identities of a person.

The model B, however, shows the lower identification rate for the correct person, since the gait data of person B is in the most dense area. It is difficult to identify the person using only the gait feature in some cases, such as person B. This will be dealt with by combining with the other features.

6. Person Identification Experiment

6.1. Person identification experiment

In order to compare the effectiveness of the features, we conducted person identification experiments. In the experiments, two people walk side by side while the robot is controlled manually and follows both persons and measures their person features. To evaluate the effectiveness of the each feature, five person classifiers are constructed. These classifiers use the following features, respectively.

1. Height feature
2. Gait feature
3. Color feature
4. All proposed features with the traditional joint feature approach
5. All proposed features with the proposed joint feature approach

For the classifiers with all the proposed features, we tested two methods: one with a traditional joint feature

Table 1: Step length estimation result.

		w/ estimation		w/o estimation	
	No. of data	mean [m]	SD [m]	mean [m]	SD [m]
observation from side	42	0.6006	0.01387	0.5988	0.06469
observation from front	32	0.6078	0.00862	0.6049	0.0114
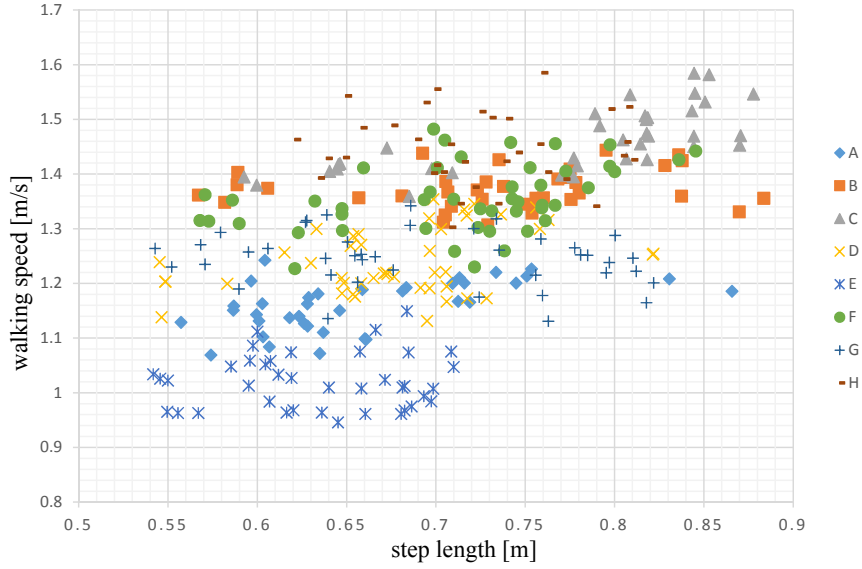
Figure 15: Observed gait data.

Table 2: Gait-based Identification Results.

model / test data	A	B	C	D	E	F	G	H
A	0.946	0.486	0.000	0.676	1.000	0.054	1.000	0.270
B	0.711	0.658	0.553	0.789	0.184	0.947	0.500	0.474
C	0.361	0.667	0.944	0.694	0.056	0.722	0.167	0.972
D	0.978	0.609	0.022	0.848	0.761	0.500	0.783	0.174
E	0.564	0.231	0.000	0.103	1.000	0.000	1.000	0.282
F	0.854	0.563	0.375	0.833	0.313	0.917	0.500	0.542
G	0.878	0.366	0.000	0.805	0.780	0.512	0.854	0.268
H	0.529	0.412	0.882	0.500	0.118	0.824	0.206	0.882

Table 3: Precision of person identification.

	height	gait	color	all features with traditional joint feature approach	all features with proposed joint feature approach
experiment 1	0.542	0.712	0.949	0.949	0.966
experiment 2	0.924	0.532	0.684	0.937	0.937
experiment 3	0.810	0.726	0.903	0.921	0.948

(a) Experiment 1. Persons with the similar heights and the different color clothes.

(b) Experiment 2. Persons with the different heights and the similar color clothes.

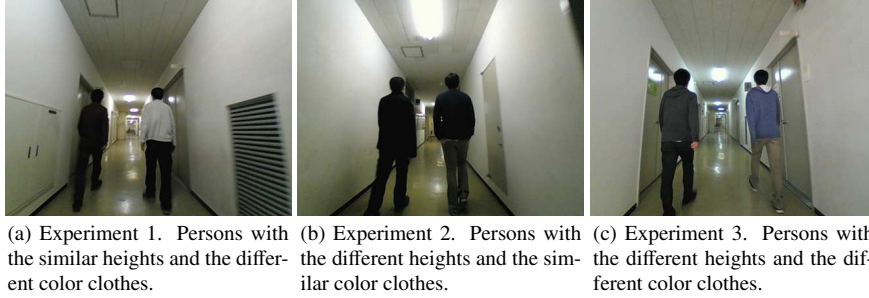(c) Experiment 3. Persons with the different heights and the different color clothes.

Figure 16: The environments of the person identification experiment.

approach and the other with the proposed one. In the all experiments, online boosting contains 10 weak classifier selectors, and each selector contains 10 weak classifiers.

The experiments are conducted in three different cases (see Fig. 16); two persons are with similar colors and different heights in case (a); those with different colors and similar heights in case (b); those with different colors and heights in case (c). The learning process of the classifier with all the proposed features takes about 20 msec long for one person. We tested the proposed system in this experiment, and calculate the precision of the identification. Table 3 shows the result of the experiments.

The classifiers using a single feature show a good precision in specific cases but not in the others. The classifiers with all the proposed features show superior performances in all cases. In addition, the classifier with the proposed joint feature shows equal or greater precision than the one with traditional one. This shows the effectiveness of the proposed joint feature.

6.2. Person identification experiment in severe illumination environments

We conducted person identification experiments for two target persons and for two different illumination environments. Fig. 17 shows snapshots of the experiments. In experiments 1 and 3, most of the persons were wearing similarly colored clothes and sometimes entered shadowed areas. In experiments 2 and 4, color information is almost lost due to a strong backlight. In all cases, it is very difficult to identify the target person using color information only.

Fig. 18 shows snapshots of the experiment 2; The experiment was conducted in the most severe illumination environment. Green rectangles in the images indicate detected persons and the red triangles above them indicate the target person. At the beginning of the experiment, the robot learned the features of the target per-

son and created a person classifier (Fig. 18 (a)), and then some persons occluded the target person (Fig. 18 (b)(c)). The LRF-based tracker failed to track the target person several times due to the occlusion of the person (Fig. 18 (d)(e)). The robot however, found the correct target person using the person classifier, and resumed correct tracking (Fig. 18(f)). In this experiment, the robot successfully continued to track a specific person in spite of temporarily-lost situations thanks to the height and the gait feature.

Table 4 shows the result of the four experiments. The total time for the experiments was about 765 [sec] and the target person was occluded by others 43 times through all of the experiments. The robot lost track of the target person 16 times due to occlusions. The person classifier, however, found the correct target person and resumed the tracking every time. The robot tracked a wrong person as the target for 6 [sec] (0.8% of the experiments) due to a wrong data association. However, that person was then judged not to be the target and the robot then found the correct person. Among the rest of the time, the robot correctly tracked the target person for 620 [sec] (81.0%) and looked for him while calculating the gait feature values for 139 [sec] (18.2%)

The person classifier with only the color feature was also tested in the experiments. The robot with the classifier successfully tracked the target person in experiment 1. In the other experiments, however, the robot tracked wrong persons in many frames (221 [sec] (28.9%)) due to severe illumination environments.

7. Person Following Framework

7.1. Tracking strategy

The LRF-based tracking method described above may sometimes fail to track the target person. The robot has to be able to recover from such a failure situation. We therefore define three states which switch in the operation as follows (see Fig. 19).

(a) Experiment 1. Direct light con-
dition. Target person is subject A.

(b) Experiment 2. Backlight con-
dition. Target person is subject A.

(c) Experiment 3. Direct light con-
dition. Target person is subject B.

(d) Experiment 4. Backlight con-
dition. Target person is subject B.

Figure 17: The environments of the person identification experiment.
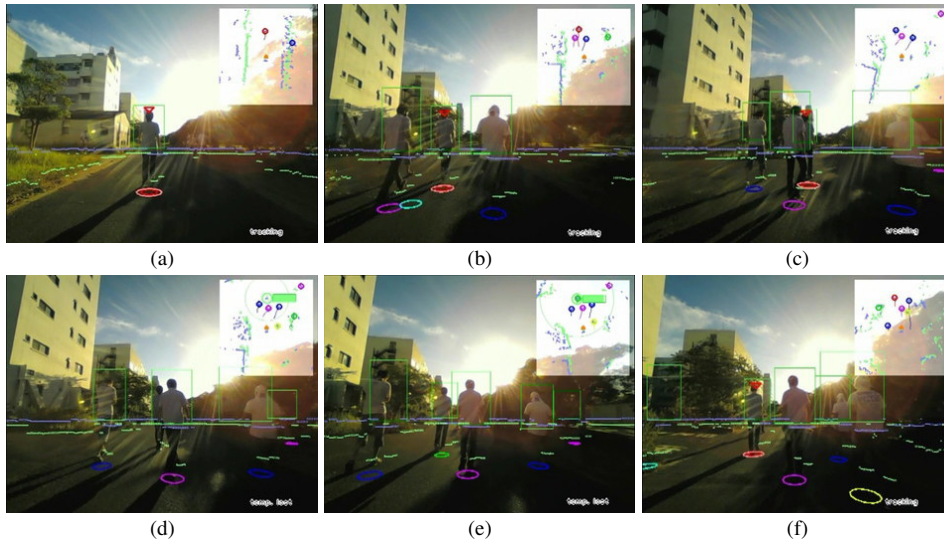

Figure 18: Person identification experiment in a severe illumination environment: Red triangles above green rectangles indicate the identified target person.

Table 4: Result of the experiments in the severe environments.

		exp. 1	exp. 2	exp. 3	exp. 4	total	
	time [sec]	157	213	195	200	765	
	occlusion of the target [times]	11	11	8	13	43	
all features	successfully tracked [sec]	128	168	167	157	620	(81.0%)
	lost track of the target [sec]	29	45	22	43	139	(18.2%)
	tracked wrong person [sec]	0	0	6	0	6	(0.8%)
	lost track of the target [times]	3	6	3	4	16	
	wrong association [times]	0	0	1	0	1	
color feature	successfully tracked [sec]	128	102	74	105	409	(53.5%)
	lost track of the target [sec]	29	61	24	21	135	(17.6%)
	tracked wrong person [sec]	0	50	97	74	221	(28.9%)
	lost track of the target [times]	3	1	1	2	7	
	wrong association [times]	0	0	1	0	1	

the person features are measured enough   the target person is lost

learning   tracking   temporary lost
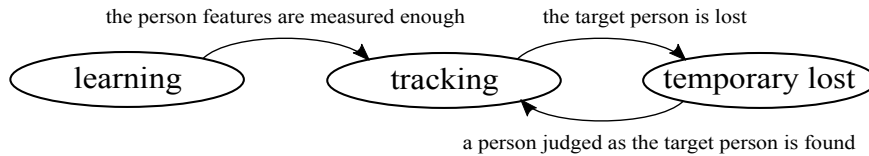
a person judged as the target person is found

Figure 19: State machine for person following behavior.

In the *initial* state, the robot measures the person features while following the target person. If a sufficient number of person features are measured, the robot constructs a person classifier from the features and transits to the *tracking state*. In the *tracking* state, the robot performs the usual tracking and identification. When the LRF-based tracking loses the target, the robot transits to the *temporary lost* state. In this state, while the robot is looking for the target person using the person classifier, the position of the target person is predicted from the most recent person movement, and the robot moves toward the position. If a person is judged as the target, that is, the target person is re-identified, the robot transits to the *tracking* state.

7.2. Person following experiment

We applied the proposed system to person following experiment. The experiment was conducted in both indoor and outdoor environments. The experimental environment is a public space in Toyohashi university of technology, and there were many ordinary persons. Fig. 20 shows snapshots of the experiment. The left images show experimental scenes. The right images show the images captured by the robot. The rectangular region in the upper right corner of the right images indicates range data and the conditions of the LRF-based tracker. The circles in the region indicate the tracked persons by the LRF-based tracker. The circles under the persons in the images also indicate the position of the tracked persons. Green rectangles in the images indicate detected person regions and the red triangles above them indicate the target person.

The experiment started in a populated outdoor environment. The robot followed a target person while measuring his features (Fig. 20(a)). Then, the robot constructed the person classifier and continued the following behavior. Several persons walked with the target person, and often occluded the target person (Fig. 20(b)). The LRF-based tracker lost the target person due to the occlusion (Fig. 20(c)). The green circle in the upper right rectangular region in the right image of Fig. 20(c) indicates the predicted target person position to witch the robot was moving. Once the target person appeared and walked for several steps (Fig. 20(d)), the

robot realized that the person was the correct target to track (Fig. 20(e)). After the robot followed the person for a while, the target person moved to the indoor environment (Fig. 20(f)). While the person and the robot were moving into the indoor environment, a strong illumination change occurred (Fig. 20(g)) and the target person was also occluded by another person (Fig. 20(h)). However, the robot successfully found the target person (Fig. 20(i)(j)). After that, the person returned to the outdoor environment and continued the following behavior (Fig. 20(k)(l)).

The duration of the experiment is 920 [sec], and the LRF-based tracker lost the target person 11 times due to occlusions. However, the robot re-identified the target every time and successfully continued to follow the target throughout the experiment. The average re-identification time after the person appeared was 5.6 [sec]. Since during that time, the robot kept moving towards the predicted position of the target, it was able to find the target when he appeared again.

8. Conclusions and Discussion

This paper has described a method of identifying a specific person using color, height and gait features. An HS histogram is extracted from the randomized rectangular region in the person region and used as a color feature. To measure the height of a person, we detect the sinciput position in the image using the hair color model, gradient image, and the person position obtained by the LRF-based tracking. We also develop a new method of estimating the gait feature from accumulated range data by spotting supporting leg positions using mean shift and a maximum likelihood estimation with a constant step length constraint. These features are combined into the joint features and learned using online boosting. We tested the proposed multi-feature identification method in a scenario where occlusions frequently occurred under a severe illumination environment to show the effectiveness of the method.

We currently suppose that the target person stays close enough to the robot so that the robot can recognize them when they are not occluded. If the person happens to be far from the robot, it then has to move around and

search a wide area for the person. To reduce the possibility of encountering such a situation, the robot has to be able to reliably predict the movement of the person in various cases. By using the predicted position, we can narrow down areas where the person may be there, and reduce the number of the candidates for the person. We are now working on modeling person movements and incorporating the predicted position information into the current identification method.

References

[1] D. Schulz, W. Burgard, D. Fox, A. B. Cremers, People tracking with mobile robots using sample-based joint probabilistic data association filters, The International Journal of Robotics Research 22 (2) (2003) 99–116. doi:10.1177/0278364903022002002.

[2] K. O. Arras, O. M. Mozos, W. Burgard, Using boosted features for the detection of people in 2D range data, in: IEEE International Conference on Robotics and Automation, IEEE, 2007, pp. 3402–3407. doi:10.1109/ROBOT.2007.363998.

[3] Z. Zainudin, S. Kodagoda, G. Dissanayake, Torso detection and tracking using a 2d laser range finder, in: Australasian Conference on Robotics and Automation, ARAA, 2010.

[4] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, IEEE, 2005, pp. 886–893. doi:10.1109/CVPR.2005.177.

[5] B. Han, S.-W. Joo, L. S. Davis, Probabilistic fusion tracking using mixture kernel-based Bayesian filtering, in: IEEE International Conference on Computer Vision, IEEE, 2007, pp. 1–8. doi:10.1109/iccv.2007.4408938.

[6] N. Noceti, A. Destrero, A. Lovato, F. Odone, Combined motion and appearance models for robust object tracking in real-time, in: IEEE International Conference on Advanced Video and Signal Based Surveillance, IEEE, 2009, pp. 412–417. doi:10.1109/avss.2009.40.

[7] G. Berdugo, O. Soceanu, Y. Moshe, D. Rudoy, I. Dvir, Object reidentification in real world scenarios across multiple non-overlapping cameras, in: European Signal Processing Conference, 2010, pp. 1806–1810.

[8] C. Cruz, L. E. Sucar, E. F. Morales, Real-time face recognition for human-robot interaction, in: IEEE International Conference on Automatic Face & Gesture Recognition, IEEE, 2008, pp. 1–6. doi:10.1109/AFGR.2008.4813386.

[9] S. D. Mowbray, M. S. Nixon, Automatic gait recognition via Fourier descriptors of deformable objects, in: Audio-and Video-Based Biometric Person Authentication, Springer, 2003, pp. 566–573. doi:10.1007/3-540-44887-x_67.

[10] Y. Makihara, H. Mannami, Y. Yagi, Gait analysis of gender and age using a large-scale multi-view gait database, in: Computer Vision–ACCV 2010, Springer, 2011, pp. 440–451. doi:10.1007/978-3-642-19309-5_34.

[11] M. Munaro, S. Ghidoni, D. T. Dizmen, E. Menegatti, A feature-based approach to people re-identification using skeleton keypoints, in: IEEE International Conference on Robotics and Automation, IEEE, 2014, pp. 5644–5651. doi:10.1109/icra.2014.6907689.

[12] N. Bellotto, H. Hu, Multisensor data fusion for joint people tracking and identification with a service robot, in: IEEE International Conference on Robotics and Biomimetics, IEEE, 2007, pp. 1494–1499. doi:10.1109/ROBIO.2007.4522385.

[13] L. Nanni, M. Munaro, S. Ghidoni, E. Menegatti, S. Brahnam, Ensemble of different approaches for a reliable person re-identification system, Applied Computing and Informatics-doi:10.1016/j.aci.2015.02.002.

[14] M. Luber, L. Spinello, K. O. Arras, People tracking in RGB-d data with on-line boosted target models, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2011, pp. 3844–3849. doi:10.1109/iros.2011.6095075.

[15] V. Alvarez-Santos, X. M. Pardo, R. Iglesias, A. Canedo-Rodriguez, C. V. Regueiro, Feature analysis for human recognition and discrimination: Application to a person-following behaviour in a mobile robot, Robotics and autonomous systems 60 (8) (2012) 1021–1036.

[16] J. Satake, M. Chiba, J. Miura, A SIFT-based person identification using a distance-dependent appearance model for a person following robot, in: IEEE International Conference on Robotics and Biomimetics, IEEE, 2012, pp. 962–967. doi:10.1109/robio.2012.6491093.

[17] A. Bedagkar-Gala, S. K. Shah, A survey of approaches and trends in person re-identification, Image and Vision Computing 32 (4) (2014) 270–286. doi:10.1016/j.imavis.2014.02.001.

[18] R. Mazzon, S. F. Tahir, A. Cavallaro, Person re-identification in crowd, Pattern Recognition Letters 33 (14) (2012) 1828–1837. doi:10.1016/j.patrec.2012.02.014.

[19] T. Wang, S. Gong, X. Zhu, S. Wang, Person Re-identification by Video Ranking, Springer, 2014, pp. 688–703. doi:10.1007/978-3-319-10593-2_45.

[20] W. LI, Y. WU, M. MUKUNOKI, M. MINOH, Bi-level relative information analysis for multiple-shot person re-identification, IEICE Transactions on Information and Systems E96.D (11) (2013) 2450–2461. doi:10.1587/transinf.e96.d.2450.

[21] O. Javed, K. Shafique, M. Shah, Appearance modeling for tracking in multiple non-overlapping cameras, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, IEEE, 2005, pp. 26–33. doi:10.1109/cvpr.2005.71.

[22] A. Hayder, J. Dargham, A. Chekima, G. M. Ervin, Person identification using gait, International Journal of Computer and Electrical Engineering 3 (4) (2011) 477–482. doi:10.7763/ijcee.2011.v3.364.

[23] C. A. Cifuentes, A. Frizera, R. Carelli, T. Bastos, Human robot interaction based on wearable IMU sensor and laser range finder, Robotics and Autonomous Systems 62 (10) (2014) 1425–1439. doi:10.1016/j.robot.2014.06.001.

[24] K. Nakamura, H. Zhao, X. Shao, R. Shibasaki, Human Sensing in Crowd Using Laser Scanners, InTech, 2012. doi:10.5772/33276.

[25] X. Song, J. Cui, H. Zhao, H. Zha, R. Shibasaki, Laser-based tracking of multiple interacting pedestrians via on-line learning, Neurocomputing 115 (2013) 92–105. doi:10.1016/j.neucom.2013.02.001.

[26] T. Germa, F. Lerasle, N. Ouadah, V. Cadenat, M. Devy, Vision and RFID-based person tracking in crowds from a mobile robot, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2009, pp. 5591–5596. doi:10.1109/iros.2009.5354475.

[27] R. Liu, G. Huskic, A. Zell, Dynamic objects tracking with a mobile robot using passive UHF RFID tags, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2014, pp. 4247–4252. doi:10.1109/iros.2014.6943161.

[28] L. Wu, Z. An, Y. Xu, L. Cui, Poster abstract: Human tracking based on lrf and wearable IMU data fusion, in: International conference on Information processing in sensor networks, ACM, 2013, pp. 349–350. doi:10.1145/2461381.2461442.

[29] A. Carballo, A. Ohya, S. Yuta, Fusion of double layered multiple laser range finders for people detection from a mobile robot,

in: IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, IEEE, 2008, pp. 677–682. doi:10.1109/MFI.2008.4648023.

[30] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297. doi:10.1007/BF00994018.

[31] R. E. Schapire, Y. Singer, Improved boosting algorithms using confidence-rated predictions, in: Annual conference on Computational learning theory, Vol. 37, ACM, 1998, pp. 297–336. doi:10.1145/279943.279960.

[32] Q. Zhu, M.-C. Yeh, K.-T. Cheng, S. Avidan, Fast human detection using a cascade of histograms of oriented gradients, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, IEEE, 2006, pp. 1491–1498. doi:10.1109/cvpr.2006.119.

[33] H. Grabner, H. Bischof, On-line boosting and vision, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, IEEE, 2006, pp. 260–267. doi:10.1109/cvpr.2006.215.

[34] M. Munaro, E. Menegatti, Fast RGB-d people tracking for service robots, Autonomous Robots 37 (3) (2014) 227–242. doi:10.1007/s10514-014-9385-0.

[35] D. A. Klein, D. Schulz, S. Frintrop, Boosting with a joint feature pool from different sensors, in: Computer Vision Systems, Springer, 2009, pp. 63–72. doi:10.1007/978-3-642-04667-4_7.

[36] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, Numerical Recipes in FORTRAN: The Art of Scientific Computing, Cambridge University Press, 1992.

[37] C. BenAbdelkader, R. Cutler, L. Davis, Stride and cadence as a biometric in automatic person identification and verification, in: IEEE International Conference on Automatic Face Gesture Recognition, IEEE, 2002, pp. 372–377. doi:10.1109/afgr.2002.1004182.

[38] E. E. Stone, M. Skubic, Passive in-home measurement of stride-to-stride gait variability comparing vision and kinect sensing, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2011, pp. 6491–6494. doi:10.1109/iembs.2011.6091602.

[39] E. E. Stone, M. Skubic, Unobtrusive, continuous, in-home gait measurement using the microsoft kinect, IEEE Transactions on Biomedical Engineering 60 (10) (2013) 2925–2932. doi:10.1109/tbme.2013.2266341.

[40] B. R. Umberger, Stance and swing phase costs in human walking, Journal of The Royal Society Interface 7 (50) (2010) 1329–1340. doi:10.1098/rsif.2010.0084.

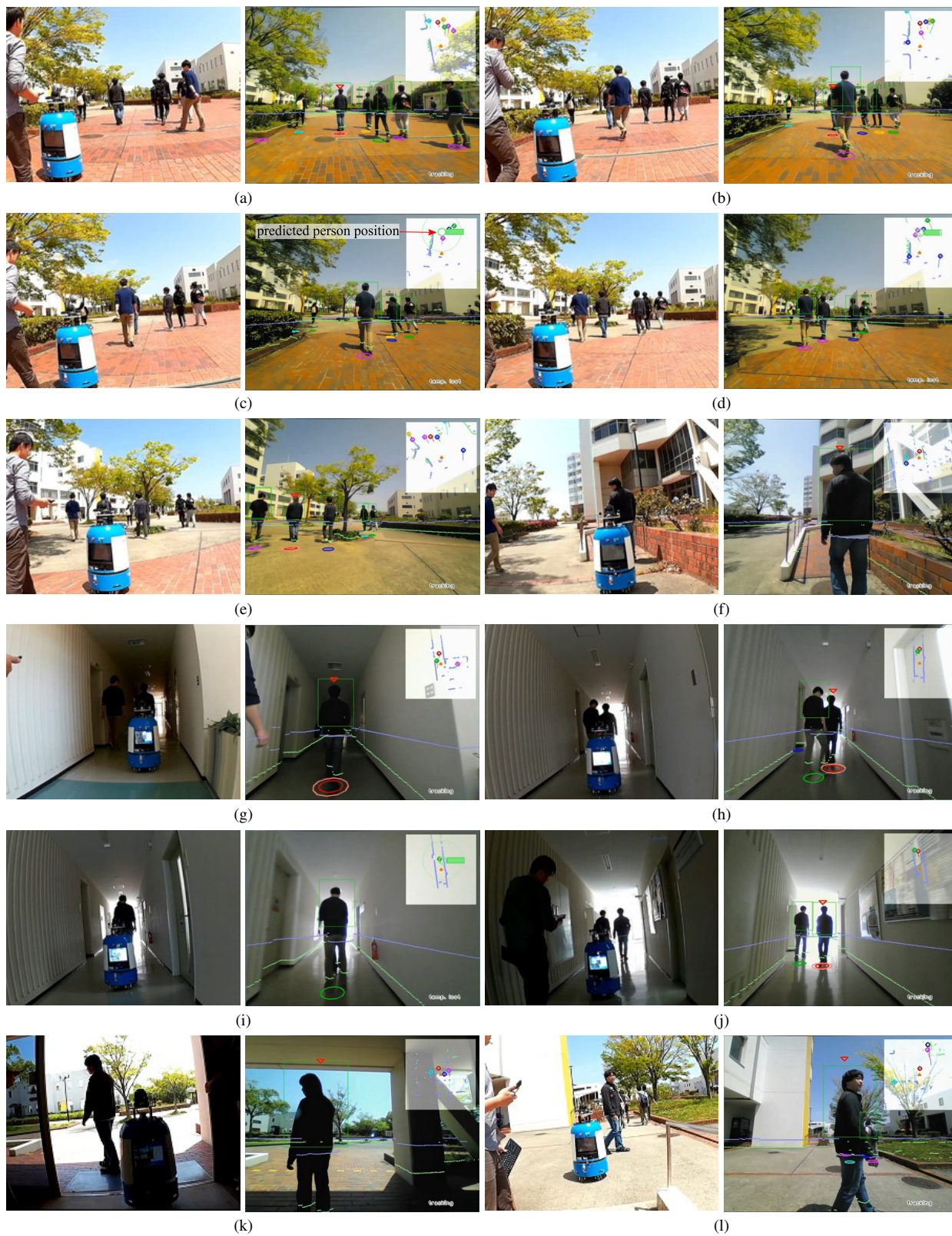[41] Y. Cheng, Mean shift, mode seeking, and clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence 17 (8) (1995) 790–799. doi:10.1109/34.400568.

Figure 20: The person following experiment.