Human Detection by Fourier descriptors and Fuzzy Color Histograms with Fuzzy *c*-means method

Shohei Akimoto[*1], **Tomokazu Takahashi**[*2], **Masato Suzuki**[*3], **Yasuhiko Arai**[*4] **and Seiji Aoyagi**[*5]

[1] Kansai University, k485204@kansai-u.ac.jp
[*2] Kansai University, t.taka@kansai-u.ac.jp
[*3] Kansai University, m.suzuki@kansai-u.ac.jp
[*4] Kansai University, arai@kansai-u.ac.jp
[*5] Kansai University, aoyagi@kansai-u.ac.jp

It is difficult to use histograms of oriented gradients (HOG) or other gradient-based features to detect persons in outdoor environments given that the background or scale undergoes considerable changes. This study involved the segmentation of depth images. Additionally, P-Type Fourier descriptors were extracted as shape features from two-dimensional coordinates of a contour in the segmentation domains. With respect to the P-type Fourier descriptors, a person detector was created with the fuzzy *c*-means method (for general person detection). Furthermore, a fuzzy color histogram was extracted in terms of color features from the RGB values of the domain surface. With respect to the fuzzy color histogram, a detector of a person wearing specific clothes was created with the fuzzy *c*-means method (specific person detection). The study includes the following characteristics: 1) The general person detection requires less number of images used for learning and is robust against a change in the scale when compared to that in cases in which HOG or other methods are used. 2) The specific person detection gives results close to those obtained by human color vision when compared to the color indices such as RGB or CIEDE. This method was applied for a person search application at the Tsukuba Challenge, and the obtained results confirmed the effectiveness of the proposed method.

Keywords: RGB-D sensor; human detection; Fourier descriptors; color histogram; fuzzy c-means method

1. Introduction

Robot technologies including home robots for housework or nursing care or outdoor robots for security or navigation have been developed in recent years. It was necessary for these robots to have a human detection function in order to avoid collision with persons in an environment with multiple persons and to identify a target person (for example, an person wearing particular clothes in the Tsukuba Challenge).

Typically, studies on human detection use images, and most of the images are taken with fixed cameras. Particularly, it was shown that a method using a feature of HOG (Histograms of Oriented Gradients), which was obtained from intensity gradient data could realize high-speed processing and high detection ability [1]. However, there are considerable changes in the background and scale with time when cameras such as those mounted on a robot move. Since the intensity gradient is significantly affected by the scale change, it is not frequently applied to methods used on a mobile robot.

Previous studies also explored human detection methods using depth data. Satake et al. [2] realized high-precision human detection by comparing depth images obtained from stereo cameras with depth image templates prepared for human detection. However, the method requires prior learning with respect to objects, which could be erroneously detected as persons in the environment. Additionally, the method did not effectively use color information. Munaro et al. [3] proposed a method wherein a three-dimensional point cloud was received from an RGB-D camera to segment the image and SVM detection based on HOG was applied to detect a person. This method realized human detection with both high speed and high precision. However, as mentioned above, the HOG feature was largely affected by a change in the scale, and hence the proposed method could only be used with a mobile robot in limited cases. Kurazume et al. [4] proposed a method for extracting geometric features such as size and curvature from depth data obtained from laser scanners placed at three different heights and creating a learning unit with AdaBoost method to realize robust human detection against environmental changes. However, since the proposed method was unable to acquire color and texture information, it was difficult to identify a specific person by utilizing this method.

Tsukuba Challenge [5] is an activity involving human detection research with mobile robots. Robots are required to detect a person wearing designated clothes while traveling along a specific course. Yamauchi et al. [6] proposed a method in which a robot was provided with the appearance color layout of a specific person in advance, and the color layout was compared with that obtained from cameras to detect the person. However, the method focused on color information, and sometimes wrongly detected a signboard or building as a person. Hosaka et al. [7] discussed a method for extracting domains from depth data obtained from laser scanners and analyzed the domain size to verify if the domain was a person. Nevertheless, it was difficult for the method to detect a specific person since it did not use the color of the person.

Given this background, the present study proposed an alternative human detection method. In this method, depth images were segmented into domains by using an RGB-D sensor, and two features (shapes and colors for a person) were extracted from each domain. The shape feature was used to detect general persons, and both the shape and color features were used to detect a specific person. A P-type Fourier descriptor [8] durable against scale change was used for the shape feature. A fuzzy color histogram [9] durable against lighting change was used for the color feature.

A detector was created based on learning with the fuzzy *c*-means (FCM) method [10] for each feature to enhance the generalization ability of the classifier. The degree of membership obtained from the detector was used to detect persons. The degree of membership assumed a value between 0 and 1, representing the extent to which the object resembled a person. The validity of the detector could be ensured by comparing the results of the detector with that of a human judgment. Additionally, the FCM method had an advantage as it could create a better detector by using lesser number of training data than that used by detectors created with SVM or other methods.

In this study, the effectiveness of the proposed method, which used a detector of FCM method for the features of P-type Fourier descriptors, was verified through the task of specific person detection at the Tsukuba Challenge. The method was then compared with the method that used a detector of SVM with the HOG feature = (hereinafter referred to as HOG+SVM). As detailed in a previous study [3], HOG+SVM was used in a person detection algorithm in Point Cloud Library (an open source library generally used in recent years).

2. Features Used for General Human Detection

In order to extract domains where persons exist, an RGB-D sensor was used to segment a depth image (along the *x* coordinate in the right-left direction, *y* coordinate in the up-down direction, and *z* in the depth direction)[11]. Next, the P-type Fourier descriptors were extracted as features from the domains with a certain size.

2.1. Preprocessing of features extraction

The adjacent pixels located within a Euclidean distance of 50 mm were labeled with the same label to form a domain. The floor part which constituted a point group existing at a certain height was eliminated based on the assumption that the height at which the RGB-D sensor was installed was known. Instead of RGB images, depth images were used and segmented by labeling based on the robustness against lighting change [12].

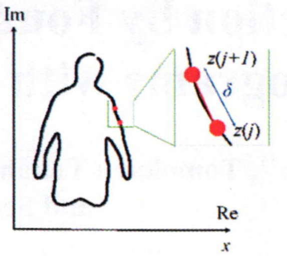


Fig. 1. Schematic of points on contour for obtaining P-type Fourier descriptor.

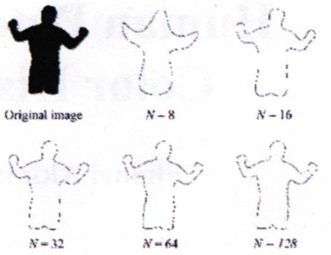


Fig. 2. Contour reconstruction with various terms of P-type Fourier descriptor.

On an empirical basis, the domains within an area of height 0.5 m - 1.9 m (in the *y* direction) and width 0.3 m - 1.5 m (in the *x* direction) were to be detected as persons.

2.2. P-type Fourier descriptors

The P-type Fourier descriptors include the descriptors obtained by the Fourier expansion of a complex function with a cumulative angular function of a curve (contour of a person in our case) in the exponent part [8]. The P-type Fourier descriptors did not change in a scale change or parallel translation. The P-type Fourier descriptors were also suitable since they use the contour of a person and require less learning data when compared to features such as HOG or other intensity gradients.

The P-type Fourier descriptors were extracted as shape features by using the *xy* coordinates of the contour of an individual person detected above. For the task of specific person detection at Tsukuba Challenge, the features of a seated person were extracted from the contours on the upper half of the extracted domains.

A method of calculating the features is explained in detail below. As shown in **Fig. 1,** a two-dimensional curve was regarded as a group of points on a complex plane and approximated to a polygonal figure with *n* line segments of the length *δ*. Each point was written as the following equation:

$$z(j) = x(j) + iy(j) \tag{1}$$

and each line segment is normalized by the length *δ* to define *w*(*j*) according to the following equation:

$$w(j) = \frac{(z(j+1) - z(j))}{\delta} \tag{2}$$

where $w(j)$ is the discrete Fourier transform to obtain the coefficient $c(k)$.

$$c(k) = \frac{1}{n}\sum_{j=0}^{n-1} w(j)\exp\left(-2\pi i\frac{jk}{n}\right) \qquad (3)$$

Here, $k = -N \cdots 0 \cdots N$. A collection of the coefficients given by Eq. (3) corresponds to N th-order P-type Fourier descriptors.

Figure 2 shows the contours reproduced in different orders. The low order components of the P-type Fourier descriptions could be used as features since they contained a considerable amount of information about the shape of the contours. Specifically, $4N$ coefficients including real term coefficients and imaginary term coefficients were used as features. According to a previous study [13], the expansion order was set to N = 20 and δ was set to 0.1 pixels.

3. CREATION OF DETECTOR WITH FUZZY *C*-MEANS METHOD

3.1. Algorithm of fuzzy *c*-means method

Fuzzy *c*-means method (hereinafter referred to as FCM method) is a method of stochastic clustering that incorporates fuzzy theory in the *k*-means method of deterministic input data clustering to *k* groups[10]. Input data can belong to multiple clusters with different degrees of membership. A data set X with n data and c cluster centers V are defined according to Eqs. (4) and (5), respectively.

$$\boldsymbol{X} = (x_1, \cdots x_k, \cdots x_n) \qquad (4)$$

$$\boldsymbol{V} = (v_1, \cdots v_i, \cdots v_c) \qquad (5)$$

The FCM method solves Eq. (6) according to the following expression:

$$\mathrm{V} = \arg\min\left(\sum_{k=1}^{n}\sum_{i=1}^{c}(\mu_{ik})^m \parallel x_k - v_i \parallel^2\right) \; for \; m \qquad (6)$$

where m is a parameter representing the degree of fuzziness and assumes a real value equal to or larger than 1. Additionally, μ is the degree of membership for each cluster and assumes a real value between 0 and 1.

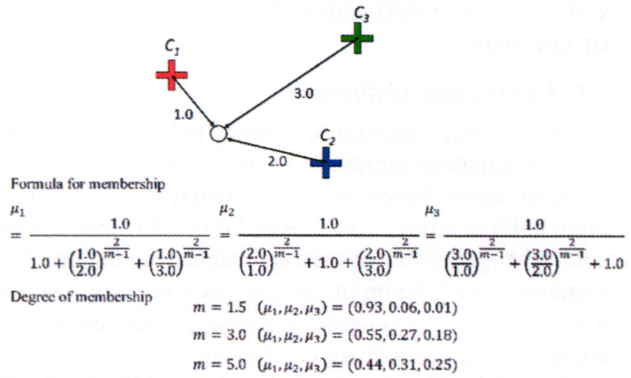
In the clustering algorithm, the cluster center is calculated by Eq. (7) with an initial value given to the degree of membership for each x.

$$v_i = \sum_{k=1}^{n}(\mu_{ik})^m x_k \Big/ \sum_{k=1}^{n}(\mu_{ik})^m \qquad (7)$$

Furthermore, μ is updated by Eq. (8) based on the obtained cluster center as detailed below:

$$\mu_{ik} = \left(\sum_{j=1}^{c}\left(\frac{\parallel x_k - v_i \parallel^2}{\parallel x_k - v_j \parallel^2}\right)^{\frac{1}{m-1}}\right)^{-1} \quad (i = 1 \cdots c) \qquad (8)$$

Eqs. (7) and (8) were used repeatedly until the cluster center converged. The obtained μ always satisfies the following equation:



C_1
C_3
C_2
3.0
1.0
2.0
Formula for membership
$\mu_1 = \frac{1.0}{1.0 + \left(\frac{1.0}{2.0}\right)^{\frac{2}{m-1}} + \left(\frac{1.0}{3.0}\right)^{\frac{2}{m-1}}}$
$\mu_2 = \frac{1.0}{\left(\frac{2.0}{1.0}\right)^{\frac{2}{m-1}} + 1.0 + \left(\frac{2.0}{3.0}\right)^{\frac{2}{m-1}}}$
$\mu_3 = \frac{1.0}{\left(\frac{3.0}{1.0}\right)^{\frac{2}{m-1}} + \left(\frac{3.0}{2.0}\right)^{\frac{2}{m-1}} + 1.0}$
Degree of membership
$m = 1.5 \;\; (\mu_1, \mu_2, \mu_3) = (0.93, 0.06, 0.01)$
$m = 3.0 \;\; (\mu_1, \mu_2, \mu_3) = (0.55, 0.27, 0.18)$
$m = 5.0 \;\; (\mu_1, \mu_2, \mu_3) = (0.44, 0.31, 0.25)$

Fig. 3. A schematic example showing the calculation of the degree of membership (μ) in fuzzy *c*-means method.

$$\sum_{i=1}^{c}\mu_{ik} = 1 \quad for \; \forall k \qquad (9)$$

3.2. Characteristics of fuzzy *c*-means method

Figure 3 shows an example of the calculation of the degree of membership. With respect to large m, the degree of membership of the input data for each cluster becomes uniform and the clustering becomes more vague. This is because the exponential term with the exponent $1/(m\text{ -}1)$ in Eq. (8) becomes closer to 1 with larger m and the degree of membership, and the degree of membership to each of the clusters 1 to c takes the same value. Therefore, the FCM method allows for vague clustering.

3.3. Calculation of degree of membership with fuzzy *c*-means method

The center of the cluster could be obtained from Eq. (7) by clustering the feature vectors of the learning data with the FCM method. This completes the learning of the detector. FCM detectors require a lower number of training data to achieve detection accuracy similar to or higher than that of SVM detectors [14].

Next, this study considered the identification of unknown data. Equation 8 was used to calculate the degree of membership of the unknown data to each cluster. Since the degree of membership to a cluster is determined by the identification with the FCM method, it is possible to quantitatively determine the extent to which the unknown data are close to the cluster (for example, in the present case, the extent to which the unknown data resemble the upper human body can be determined). Additionally, the degree of membership can be compared with human judgment to ensure the effectiveness of the detector. For example, given a degree of membership that is less than 0.5, a binary classifier such as SVM could result in detection errors. However, the present method allows for improvements in person detection.. For example, in case that the object resembles a person but the degree of membership is approximate 0.4, adjusting the threshold from 0.5 to 0.4 would be possible and effective in this method.

4. Features Used for Specific Human Detection

4.1. Fuzzy color histogram

A fuzzy color histogram proposed by Han et al.[9] was used as a feature representing the color of a person. An ordinary color histogram was created by assigning an input color to a bin of the closest color. However, a fuzzy color histogram was created by calculating the degree of membership of the input color to each bin and assigning it to each bin. This enables obtaining color features that are robust against a change in lighting.

Specifically, 13 categorical colors [15] (red, pink, orange, yellow, green, blue, purple, brown, white, gray, black, cyan, and beige) proposed by Uchikawa et al. were used for the bins of the histogram. According to experimental studies with Japanese subjects, persons tend to classify a color into one of these 13 colors. **Figure 4** shows 13 categorical colors and their RGB values.

In order to calculate the similarity (color difference) between each pixel color and the categorical colors, the RGB color of a pixel was converted to an $L^*a^*b^*$ value, which was then subjected to a CIEDE2000 color difference formula [16]. The CIEDE2000 color difference formula was experimentally proposed such that the difference between two colors could be matched with human visual sensation.

The process of creating a fuzzy color histogram includes the following steps:

STEP 1. 13 clusters centering on $L^*a^*b^*$ values of the 13 categorical colors are prepared.

STEP 2. All pixels of the obtained domain are used as inputs. The RGB values of each pixel are converted to $L^*a^*b^*$ values.

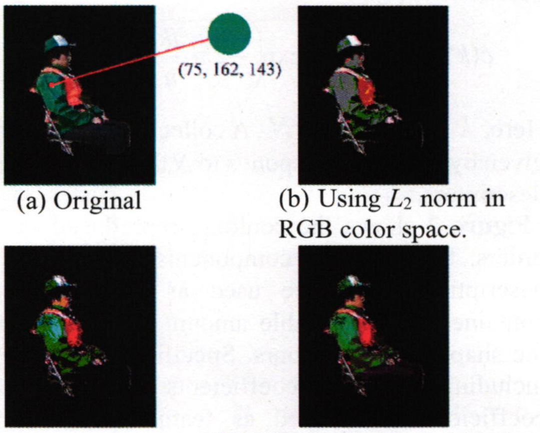
STEP 3. The CIEDE2000 color difference formula is used to calculated the color difference between the $L^*a^*b^*$ values of the input and those of the 13 categorical colors.

STEP 4. The degree of membership of a given color into the 13 categorical colors is calculated by using Eq. (8). The color difference obtained in STEP 3 is used to calculate the distance of Eq. (8) between the input and each cluster. The degree of fuzziness m is empirically set to 2.0.

STEP 5. The obtained degree of membership is assigned to each bin of the histogram.

STEP 6. Each bin of the histogram is divided by the number of input pixels for normalization purposes.

Figure 5 shows images where the color of each pixel is replaced with the closest categorical color (out of 13 colors) by various color difference calculation methods. When the L_2 norm of RGB (the square root of the sum of the squares of the component differences) or the CIEDE2000 color difference formula was used, then each pixel color was replaced with the color of the smallest color difference. When the fuzzy color histogram was used, then each pixel color was replaced with a synthesized color wherein the RGB values of the 13 categorical colors were weighted by the degree of membership. As a result, it could be observed that several colors were replaced with gray in the replacement based on the RGB values. However, replacement accuracy was improved in the replacement based on the CIEDE2000 color difference formula and the replacement based on the fuzzy color histogram. In this example, a person wearing green clothes served as a target person in the Tsukuba Challenge. However, the example of the pixels extracted from the clothes in **Fig. 5(a)** indicated that the RGB values were 75, 162, and 143. Furthermore, the comparison between the 13 categorical colors and the L_2 norm of RGB revealed that the closest color was gray (128, 128, 128), and not green (0, 128, 0) as shown in **Fig. 4**. Conversely, the color of the clothes was replaced with green or cyan when the CIEDE2000 color difference formula was used. This is closer to the human visual sensation. Both the results of the CIEDE2000 color difference formula and the weighting of the degree of membership were used for the color replacement when the fuzzy color histogram was used. Hence, the color of the clothes and the color of the trousers were much closer to those of the original image. For example, the arm color is cyan in Fig. (c) but green in Fig. (d), which coincides with the original color.



(a) Original

(b) Using L_2 norm in RGB color space.

(c) Using CIEDE2000 color difference.

(d) Proposed method (using FCM method and CIEDE2000 color difference).

Fig. 5. Results of transforming RGB value of each pixel to 13 categorical colors via various methods for calculating color difference.

5. Proposed Method for General and Specific Human Detection

5.1. Learning

Two classes of training data, human (for search targets in the Tsukuba Challenge) and non-human were used. The shape feature (P-type Fourier descriptors) and color feature (fuzzy color histogram) were extracted by using an RGB-D sensor from a seated person who wore the same clothes as the persons in the Tsukuba Challenge. It was necessary to prepare the training images since the data set of the target persons in the Tsukuba Challenge was not available. However, an effective human detector can be created with a small number of training data if the P-type Fourier descriptors and the FCM method are used. (See the detection result shown below). The RGB-D Object Dataset [17] was used with respect to non-human training data.

With respect to each of the shape features and color features, the cluster center of each class (i.e. human and non-human) was learned by using the FCM method. The initial degree of membership of each training data to the human cluster was randomly set to a value between 0.8 and 1.0, and that to the non-human cluster was set to a value between 0 and 0.2, in order to realize an easy conversion of the cluster.

The learning of the detector was conducted with different degrees of fuzziness *m* and the accuracy of training data classification with the created detector is shown in **Fig. 6**. With respect to the shape feature, the accuracy had a peak as a function of *m*. With respect to the color feature, the accuracy initially retained the same level and then monotonically decreased as *m* increased. Particularly, the detection becomes more vague, and the detection accuracy decreases for large values of *m*. A detector with $m = 1.6$ was used based on this result. Thus, the accuracy was highest for the shape feature. A certain level of accuracy was retained for $m = 2.0$, however, it had considerable versatility (degree of vagueness) with respect to the color feature.

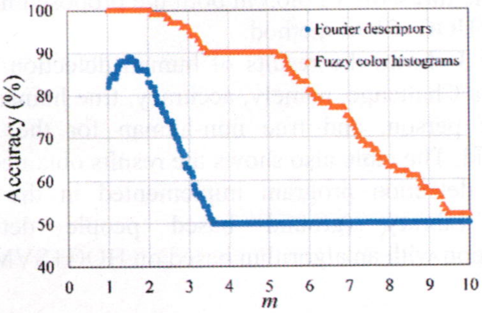


Fig. 6. Accuracy by changing degree of fuzziness (*m*).

5.2. Identification

Figure 7 shows a flow of human detection processing. The depth image data from the RGB-D sensor were segmented into domains, and the features were extracted from the upper half of each domain. For the P-type Fourier descriptors, the domains were expanded by morphology processing to prevent domain loss.

With respect to each feature, the detector that learned by using the FCM method was used to calculate the degree of membership to the human or non-human cluster. If the degree of membership of the P-type Fourier descriptor to the human cluster exceeded a threshold, then it was regarded as person. Furthermore, if the degree of membership of the fuzzy color histogram exceeded a threshold, it was regarded as a person wearing the Tsukuba Challenge clothes. Section 6.2 describes the process of determining the thresholds.

6. Person Detection Experiment at Tsukuba Challenge

6.1. Experiment method

The proposed method was used in conjunction with the RGB-D data of search target persons acquired at the Tsukuba Challenge 2015 to conduct an experiment involving the detection of a specific person in an outdoor environment. The data were acquired by Microsoft KINECT for Windows v2 (Kinect v2) mounted on a developed mobile robot with wheels. The Kinect v2 was set horizontally at a height of 0.74 m, and the robot was manually driven around the search area. Based on the ToF method, Kinect v2 acquired three-dimensional data of a depth ranging from 0.5 m to 8.0 m. The data could contain noises if an RGB-D sensor such as Kinect v2 was used in outdoor environment. This study examined the performance of Kinect v2 in an outdoor environment and found that it could acquire a lesser number of points under sunlight due to the interference of infrared rays but could still obtain the same level of information as the LRF (Laser Range Finder) because Kinect v2 could acquire sufficiently dense points [18].

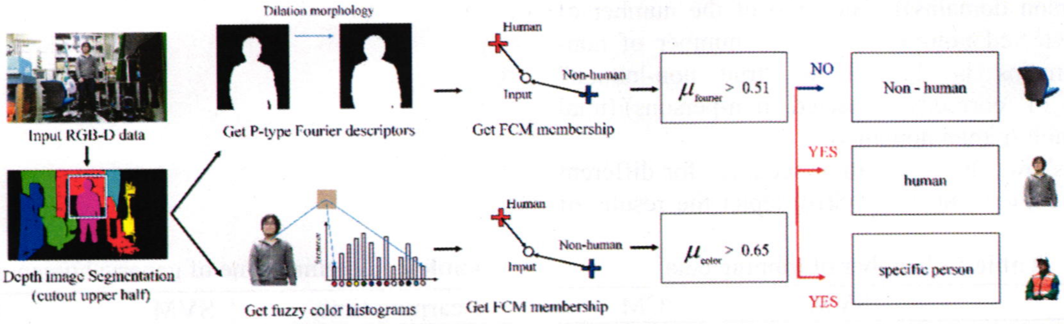


Fig. 7. Flow of human detection process.

The person detection accuracy of the proposed method (P-type Fourier descriptors and FCM method) and that of the conventional method of HOG+SVM were compared with each other.

The HOG feature was extracted from the upper half of the domains in the data set of NICTA Pedestrian Dataset [19] and learning was conducted by using the LibSVM [20]. As shown in **Table 1**, the number of training images (including humans and non-humans) for the P-type Fourier descriptors was 100, and that for the HOG+SVM was 10,000. The following parameters were included in the HOG feature. The image size was 32x40 pixels, the cell size was 4x4 pixels, the block size was 2x2 cells, and the angle ranging between 0° and 180° was divided into 9 directions with an interval of 20°. Specifically, the number of dimensions of the HOG feature was (32/4-1)x(40/4-1)x2x2x9=2,268.

Both methods applied a detector with a fuzzy color histogram to a domain that was judged as a person, in order to verify whether or not if the person was the specific target person (i.e., the person dressed in the Tsukuba Challenge clothes).

The calculation was performed on Windows 8.1 with CPU Intel Core i5-4310U (2.00GHz) and 4 GB memory.

6.2. Experiment results of all extracted domains

Table 2 shows the time required for learning. Therefore, the proposed method required only a short learning time (approximately 1/400 per frame) as shown in **Table 1**.

There were 93 domains extracted from the depth image data of the Tsukuba Challenge. Of these domains, 42 were human domains (in which 28 were domains of target persons) and 51 (=93-42) were non-human domains.

The ratio of the sum of the number of correctly detected persons and that of correctly detected non-persons to the total number of extracted domains is defined as "accuracy" (=((Number of correctly detected non-persons)+(Number of correctly detected persons))/(Total number of domains)). The ratio of the number of correctly detected persons to the number of human domains is defined as "true human" (=(number of correctly detected persons)/(total number of human domains)). The ratio of the number of correctly detected specific persons to the number of specific-person domains is defined as "true specific person" (=(number of correctly detected specific persons)/(total number of specific-person domains)). The ratio of the number of correctly detected non-persons to the number of non-human domains is defined as "true non-human" (=(number of correctly detected non-persons)/(total number of non-human domains)).

Figure 8 shows the recognition accuracy for different thresholds. **Figures 8(a)** and **8(b)** depict the results of human detection and the results of specific human detection, respectively. With respect to human detection, the proposed method used the feature of P-type Fourier descriptors and a detector made with the FCM method, yielding an output ranging from 0 to 1. The domain was regarded as person when the output value exceeded a threshold. From **Fig. 8(a),** it is observed that the accuracy was the highest when the threshold was 0.51. Conversely, this type of curve could not be obtained in the HOG+SVM method, since the LibSVM automatically determines the optimal threshold. Since both the proposed method and the conventional HOG+SVM method use the fuzzy color histogram and create a detector with the FCM method, curves presenting changes of the "accuracy of detecting target" (=((number of correctly detected non-targets)+(number of correctly detected targets)/(total number of domains)) against the threshold can be obtained for both methods as shown in **Fig. 8(b)**. This indicated that the accuracy was highest when the threshold of 0.65 in both the proposed method and the HOG+SVM method.

Table 3 shows the results of human detection at the Tsukuba Challenge, namely, accuracy, true human, true specific person, and true non-human for the above threshold. The table also shows the results obtained by a human detection program implemented in the Point Cloud Library (ground based people detection application with an algorithm based on HOG+SVM with

Table 1. Number of training data

	SVM	FCM
Positive (frame)	5,000	50
Negative (frame)	5,000	50

Table 2. Learning time of generic human detector

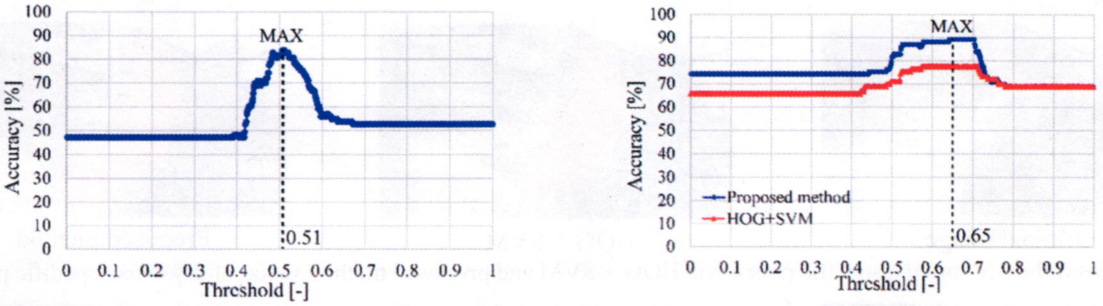
Learning time	SVM	FCM
Total (ms)	66,000	1.5
per frame (ms)	6.6	1.5×10^{-2}

Table 3. Evaluation results in Tsukuba challenge (among 93 domains of raw data)

	Accuracy	True human	True specific person	True non-human
Proposed method	83.9 %	75.0 %	69.0 %	91.8 %
HOG + SVM	58.1 %	26.2 %	28.6 %	84.3 %
Point Cloud Library	72.1 %	38.9 %	48.1 %	88.0 %

Table 4. Evaluation results in Tsukuba challenge (among 78 domains of good condition data)

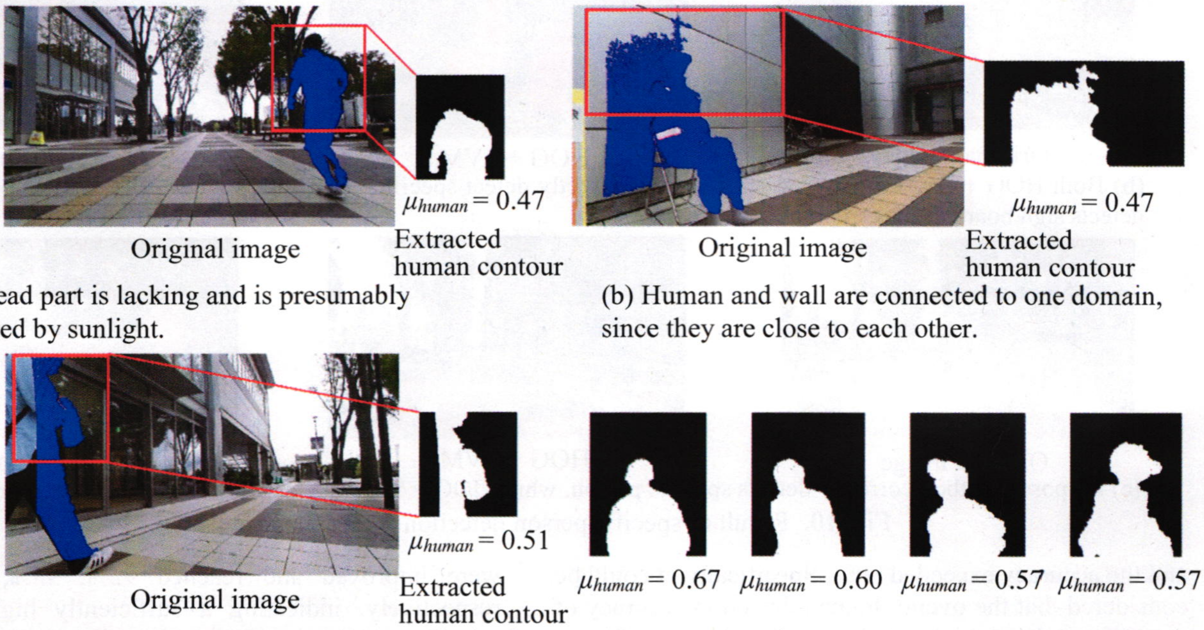
	Accuracy	True human	True specific person	True non-human
Proposed method	92.3 %	93.1 %	90.5 %	91.8 %
HOG + SVM	69.2 %	40.7 %	42.1 %	84.3 %
Point Cloud Library	78.9 %	5.0 %	8.3 %	98.7 %



(a) accuracy-threshold curve in human detection.

(b) accuracy-threshold curve in specific person detection.

Fig. 8. Accuracy by changing threshold in proposed method.



(a) Head part is lacking and is presumably affected by sunlight.

(b) Human and wall are connected to one domain, since they are close to each other.

(c) Only half part of human is detected due to limit of detectable range of Kinect v2. Border of success/failure.

(d) Examples of extracted contours of successfully detected humans.

Fig. 9. Failure and success examples of extracted human contours. μ_{human} denotes the degree of human membership, of which threshold is over 0.51 for humans.

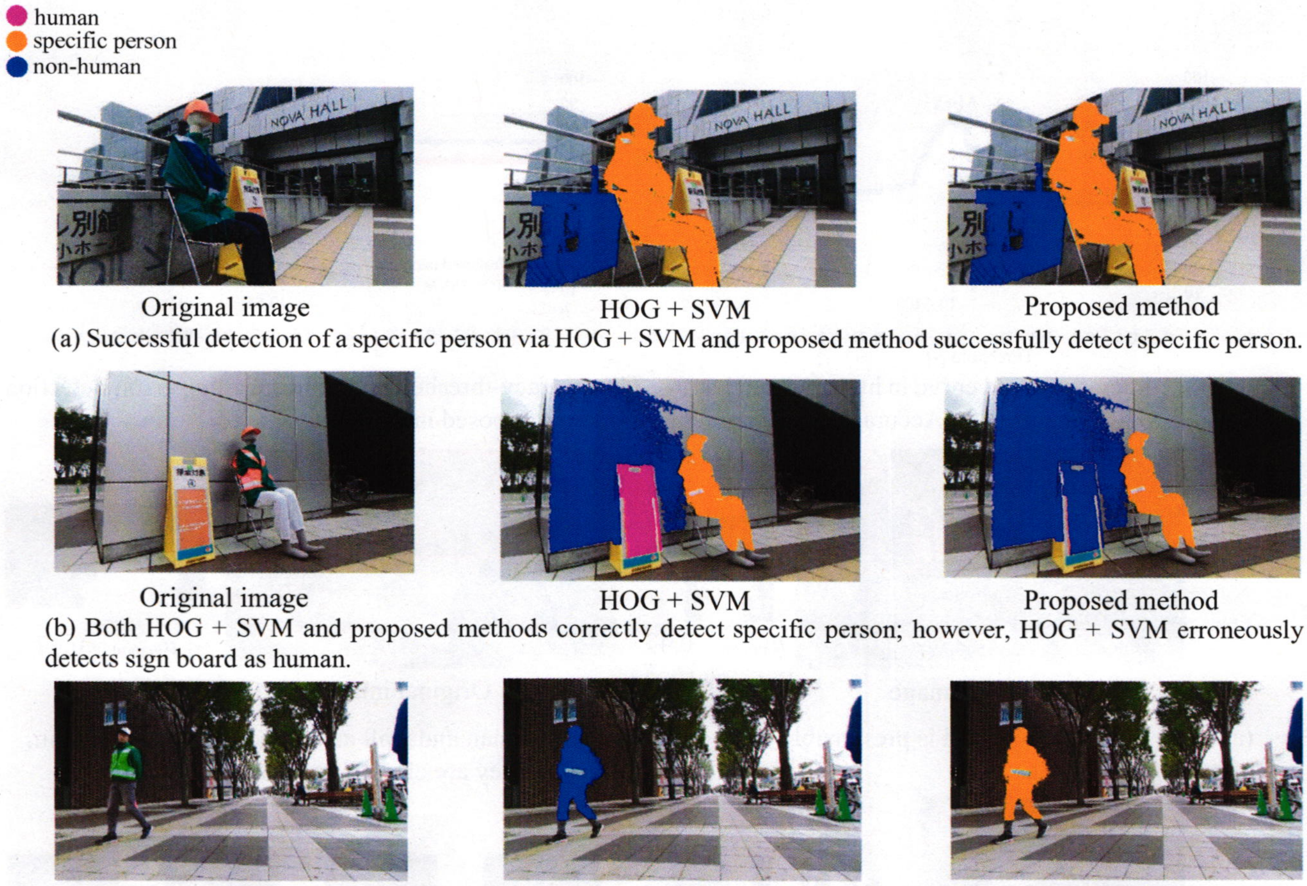
a classifier in which learning was already completed). Although the Tsukuba Challenge required the detection of a seated person, this detector was trained for a standing person and had lower accuracy than the proposed method.

6.3. Discussion on error factors

Table 3 indicates that the proposed method realized a higher detection accuracy in all the evaluation indices when compared to the other methods. This was because the HOG feature that depended on the intensity gradient did not correctly respond to the changes since the lighting and background changes occurred largely in an outdoor environment. Furthermore, a vertically long object tended to be wrongly detected as a person in an outdoor environment as long as the HOG feature was used. This was another cause for the low accuracy of the conventional method.

Conversely, even with the proposed method including the P-type Fourier descriptors and the FCM method, the index of true human is 75%. This indicated that the method could detect a person as non-person. This wrong detection occurred almost when the domain extraction had a problem. **Figures 9 (a)-(c)** give a few examples of the wrong detection, showing original images and extracted contour images. In (a), the point cloud of the head disappeared probably because the infrared ray of the Kinect v2 sensor interfered with the outdoor light. In (b), the point cloud of the person was connected to that of another area, i.e., a wall in this case (the point cloud of the wall was located at a height of 1.9 m or lower due to the sensor view field and satisfied the person domain condition). In (c), a person existed at the edge of the view field of the Kinect v2 sensor and only half of his/her point cloud could be acquired.

Figure 9(d) shows the successfully detected contour images. It could be observed that the proposed method determined a judgment with respect to a person or a non-person by using the contour of the upper body of a person. Particularly, with respect to data on the round head, thin neck, and wide shoulder length. In the erroneous detection cases shown in **Fig. 9**, this type of contour was not obtained and the P-type Fourier descriptors did not work appropriately. However, **Table 3** indicated that the true non-human of the proposed method exceeded 90%



(a) Successful detection of a specific person via HOG + SVM and proposed method successfully detect specific person.

(b) Both HOG + SVM and proposed methods correctly detect specific person; however, HOG + SVM erroneously detects sign board as human.

(c) Proposed method correctly detects specific person, while HOG + SVM erroneously detects human as non-human.

Fig. 10. Result of specific person detection in Tsukuba challenge.

and the accuracy exceeded 80%. Specifically, it could be considered that the overall human detection accuracy of the proposed method was high even in an outdoor environment.

6.4. Discussion on number of training data

The number of training data is shown in **Table 1**. HOG+SVM required a total of 10,000 images. For the detection targets at the Tsukuba Challenge, the detection accuracy shown in **Table 3** still remained at the same level even when the number of the images was reduced to 100 (50 human images and 50 non-human images). In general, the HOG+SVM method required many training images but the FCM did not. However, even the former did not need many training images for this detection target. This could be because the total number of the detection targets was as low as 93, and the accuracy was as low as 58%, but the mechanism was not clear.

6.5. Experiment results with successful domain extraction

In order to solve the problem in **Fig. 9**, the missing head area, half-body area, and the area merging with another object were manually identified, and these 15 domains were eliminated from the studied domains. Therefore, the number of domains was reduced from 93 to 78. The proposed method and the conventional method were both again applied to these 78 domains to enable human detection. **Table 4** shows the results. The accuracy, true human, and true specific person of the proposed method were improved and reached 92%, 93%, and 90%, respectively, indicating a sufficiently high detection precision.

Therefore, high recognition precision could be achieved with the proposed method if the domain extraction from the depth images was successful. In other words, the development of a high precision domain extraction method is an important task in the future.

6.6. Result of target human detection

Figure 10 shows an example of detecting target persons at the Tsukuba Challenge. The domains considered as persons are shown in gray, those considered as the Tsukuba Challenge's target person are shown in light gray, and those considered as non-persons are shown in dark gray.

Figure (a) shows an example where both the proposed method and the reference method (HOG+SVM) successfully detected a person. The domain detected as a person was further detected as a target person (specific person) by the fuzzy color histogram. The detector was created by color training in an indoor environment. However, it successfully detected a person even in an outdoor environment. This indicated that the present method was robust with respect to lighting changes. In (b), the proposed method was successful in the detection but the conventional method wrongly detected a signboard as a person. In (c), the conventional method wrongly detected a person as a non-person. This could be because the background image and thereby the brightness gradient were considerably different from that in the training. Conversely, the proposed method detected a standing person who wore the clothes of the Tsukuba Challenge. However, since the Tsukuba Challenge required robots to detect seated persons, detection of a standing person was considered as erroneous detection. Countermeasures against this error will be explored in a future study.

CONCLUSIONS AND FUTURE PROBLEMS

In this study, a method of creating a human detector with the FCM method by employing the P-type Fourier descriptors as shape features and the fuzzy color histogram as color features was proposed. This method allowed a shorter training period of the detector when compared to those of previous methods. Experiments using actual images with target persons taken at the Tsukuba Challenge indicated that 1) human detection using the P-type Fourier descriptors was more robust against scale changes than that using the HOG feature, 2) a color feature robust against lighting changes could be obtained by using the fuzzy color histogram, and 3) a combination of these features and the FCM method could be used to detect a person wearing specific clothes in an outdoor environment, thereby indicating the effectiveness of the proposed method.

Future research includes the development of more accurate domain extraction methods and evaluating the robustness of the proposed method against occlusion.

Acknowledgement

A part of the study was financially supported by Promotion Grant for Higher Education and Research 2014 at Kansai University under the title ”Tsukuba Challenge and RoboCup@Home.”

REFERENCES

[1] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection." In Proc. CVPR, CA, USA, pp. 886-893, 2005.

[2] Satake Junji, and Jun Miura, "Robust stereo-based human detection and tracking for a person following robot." In Workshop on People Detection and Tracking IEEE ICRA, 2009.

[3] Matteo Munaro, et al., "Tracking people within groups with RGB-D data." IEEE/RSJ IEEE International Conference on Intelligent Robots and Systems, pp.2101-2107, 2012.

[4] Mozos, Oscar Martinez, Ryo Kurazume, and Tsutomu Hasegawa, "Multi-part people detection using 2d range data." International Journal of Social Robotics, Vol. 2, No. 1, pp. 31-40, 2010.

[5] Yuta Shin'ichi, "Open Experiment of Autonomous Navigation of Mobile Robots in the City: Tsukuba Challenge 2014 and the Results (Special Issue on Real World Robot Challenge in Tsukuba: Autonomous Technology for Useful Mobile Robot)." Journal of robotics and mechatronics, Vol. 27, No. 4, pp. 318-326, 2015.

[6] Yamauchi Kenji, et al, "Person Detection Method Based on Color Layout in Real World Robot Challenge 2013." Journal ref: Journal of Robotics and Mechatronics, Vol. 26, No. 2, pp. 151-157, 2014.

[7] Hosaka Kento, et al., "A Person Detection Method Using 3D Laser Scanner: Proposal of Efficient Grouping Method of Point Cloud Data (Special Issue on Real World Robot Challenge in Tsukuba: Autonomous Technology for Useful Mobile Robot)." Journal of robotics and mechatronics, Vol. 27, No. 4, pp. 374-381, 2014.

[8] Uesaka Yoshinori, "A new Fourier descriptor applicable to open curves." Trans. of IEICE, Vol. 67-A (Vol. 3), pp. 166-173, 1984 (in Japanese).

[9] Han Ju, and Kai-Kuang Ma, "Fuzzy color histogram and its use in color image retrieval." IEEE Trans. Image Process, Vol. 11, No. 8, pp. 994-952, 2002.

[10] Bezdek, et al., and William Full, "FCM: The fuzzy c-means clustering algorithm." Computers & Geosciences 10.2, pp. 191-203, 1984.

[11] Y. Inaura, T. Takahashi, M. Suzuki, and S. Aoyagi, "To propose a office furniture recognizing method based on general shape concept and recoginition example using a depth map." The 30 Th Annual Conference Of The Robotics Society of Japan, 2J1-6, 2012. (in Japanese).

[12] O. Hori, 2004, Labeling. In: Digital Image Processing. Tokyo: Computer Graphic Arts Society. pp. 181-182.

[13] De Leon, et al., "Human silhouette recognition with Fourier descriptors." In Proc. of 15th International Conference on Pattern Recognition, pp. 709-712, Sep. 2000.

[14] Ichihashi Hidetomo, et al., "Benchmarking parameterized fuzzy c-Means classifier." Proc. IEEE International Conference on Fuzzy System, pp.1137-1144, Aug. 2009.

[15] Uchikawa K., "Categorical perception of surface color.", Kogaku, 17. 12, pp.47-55, 1988. (in Japanese).

[16] Sharma, et al., "The CIEDE2000 color - difference formula: Implementation notes, supplementary test data, and mathematical observations." Color Research & Application, Vol. 30, pp. 21-30, 2004.

[17] Kevin Lai, et al., "A Large-Scale Hierarchical Multi-View RGB-D Object Dataset." IEEE, International Conference on Robotics and Automation (ICRA), pp. 1817-1824, 2011.

[18] K. Takahashi, T. Takahashi, M. Suzuki, and S. Aoyagi, "Application of Kinect v2 Sensor to a Mobile Robot and Its Characterization in the Outdoor Environment (Comparison with LRF)." The Robotics and Mechatronics Conference, 2015. (in Japanese).

[19] Overett, Gary, et al., "A new pedestrian dataset for supervised learning." IEEE Intelligent Vehicles Symposium, 2008.

[20] Chang, et al., "LIBSVM: a library for support vector machines." 2001.