Kindergarten Children Identification Based on Garment Characteristics

Nozomi matsuda ^{a,*}, Qian Fan ^a, Lifeng Zhang ^a

^a Kyushu Institute of Technology, 1-1 Sensui-cho Tobata-ku Kitakyusyu-shi Fukuoka-ken, 8048550, Japan

* p349432n@mail.kyutech.jp

Abstract

Image processing technologies using garment image have been applied for some fields such as similar garment search technology and virtual fitting technology. With regard to similar garment search technology, it is a technique to retrieve same garment images shot from different viewpoints has not been researched sufficiently. However, it may be inferred that by using garment identification technology for person identification, it is possible to propose a detection method that does not depend on the direction of a person, compared to conventional methods based on biometric information like face information. It is thought that this can be effectively used for detecting and identifying a person within a limited space such as a kindergarten or a nursery school.

In this research, we researched a method of discriminating children using features of garment. In this method, feature information is acquired from a pattern of garment, and a garment feature model is created for each child. The information used as a feature is three types. There are shape information by Hough transform, color feature based on color space, and frequency feature by discrete cosine transform.

In the experiments, we created a feature model from the region determined by template matching with the pattern image and compare it with the feature model of the pattern, to measure the discrimination accuracy of the proposed method.

Keywords: garment, garment identification, pattern matching.

1. Background

Image processing techniques utilizing garment images have application fields such as similar garment retrieval technology and person identification technology.

The similar garment retrieval technique is a method of retrieving clothes having features similar to designated garment by using the characteristics of the garment selected by the user. As features of garment, a method of digitizing colors and shapes⁽¹⁾ and a method of utilizing the local positional relationship of color space⁽²⁾ and the like have been proposed. Currently, several applications for similar garment search using these methods have been developed and are being used by people. It is said that it is able to contribute to improvement of user satisfaction and usability, especially by combining it with online shopping site, since clothes matching various tastes can be retrieved from various clothes.

The person identification technology is a technique for identifying the same person from a plurality of images and distinguishing between different people. Currently, this technology is applied to many of photograph management services and software, and research is being carried out to incorporate this technology as one of robot functions⁽³⁾. In human identification technology, face information is often used, but research is being made to supplementary use garment information assuming that a part of the face is chipped due to the influence of shadows or obstacles⁽⁴⁾. The reason why garment information is used only auxiliary is that the clothes worn by the same person are different due to changes in date and time zone. For example, it can be assumed that many people use different clothes in everyday wear when going out during the day and sleeping when in the night at home. Also, even though it is limited to casual wear, it is considered that there are many cases where you wear clothes that are totally different appearance on the day and the following day. Therefore, there is a problem that features that permanently identify individuals can not be acquired from garment. As a solution to this, there is a method of limiting the range of time to be applied to the person identification technology, but the issue of how to apply person identification in a very short period has not

been sufficiently discussed. Therefore, in the current person identification technology, it is mainstream to acquire personal characteristics of the body from the face, the retina, etc., and the garment information is regarded as an auxiliary position.

Several fields of image processing technology utilizing garment images so far, and mentioned that person identification technology can be used as auxiliary information of a method based on physical information among them. This is because garment information is distinctively different in characteristics over time even if the garment information is the same person as physical information. On the other hand, garment information has advantages of less dependence on orientation compared to body information. If the colors and patterns of the entire face, the back face, and the side face are similar, it is also possible to identify the same clothes taken from different orientations by utilizing the features of the garment images taken on one face. This adds the effect that the user does not have to consider the orientation with respect to the camera in the person identification technology. For example, when using physical information for person identification, there are conditions under which the user must face the camera in front of the camera in many ways. However, if you use garment information, you can identify the same person in addition to the case where the user faces the front, and even when turning his back. In this way, garment information has an excellent property of being highly flexible with respect to orientation. In this research, focusing on the flexibility with respect to the orientation of garment information, we have studied person identification technology based on garment information in order to detect a person in a limited space such as a kindergarten or a nursery.

2. Proposed Method

In this section, garment identification method based on matching of color features and shape features of patterns is described. This method consists of two stages: a process of creating a feature model of a garment as a comparison source, and a process of preparing a feature model from the image to be compared and calculating the degree of matching with the comparison source modeling. First, the creation processing of the comparison source model will be described, then the creation of the comparison object model and the calculation processing of the coincidence rate will be described. Evaluation for the proposed method is described in the next section.

Fig. 1. Garment pattern image cropping from a child image.

Fig. 2. Flow chart of child's base model generating.

2.1 Process of Creating Comparison Source Model

In this process, when an arbitrary clothing pattern image is given, plural types of features are extracted from the image and stored as a clothing feature model of the child. The clothing pattern image is an image that can be judged that it can be clipped out from the clothing area and has high frequent appearance. In this research, the image shown in Figure 1c was treated as a clothing pattern image. Fig. 1a shows an image of a child from which a clothing pattern image was cut out. When clothes area is visually cut out from Fig. 1a, it is considered to be approximately Fig. 1b. In Fig. 1b, since patterns like Fig. 1c appear frequently, this is treated as a clothing pattern image.

Introduction Figure 2 shows a flow chart of the comparison source model creation process. In Fig. 2, "input a base image" processing corresponds to giving any clothing

pattern image previously described. In the next "extract color feature" processing, features based on the color information of the image are acquired. In the "extract hough feature" processing, shape features of the image based on the Hough transform^(5, 6) are acquired. In the "extract DCT feature" process, the frequency characteristic of the image based on the discrete cosine transform (DCT) is acquired. In the last "save data" processing, the acquired features are recorded as clothing feature model data of the comparison source.

2.2 Shape Features of Images Based on Hough Transform

Hough transform⁽⁵⁾ is a method used as one of feature extraction methods in image processing. By using the Hough transformation, it is possible to extract shapes constituting straight lines, circles, etc. in the image. This method is called standard Hough transformation, and among them, the theory to detect a straight line is the basis. By applying this, shape such as a circle can be detected.

In the proposed method, three types of circles, horizontal lines, and vertical lines are defined as the features possessed by the image, using the shape that can be acquired by Hough transform. Further, in the case of having each feature, detailed features of each shape such as the position of the center point coordinate with respect to the length of the image and the difference between the start point and the end point of the line segment are calculated.

Introduction Figure 3 shows a flowchart for detecting circles. In Fig. 3, from "Smoothing" processing to "Detect contour processing", it is preprocessing for Hough transformation. The preprocessing converts the inputted pattern image to a line image which can detect the shape by Hough transformation. First of all, in the smoothing process, the input image is smoothed to reduce the noise in the image. A Gaussian filter is used for smoothing.

In the "Binarization" process, the smoothed image is converted into a binary image. A binary image is an image in which each pixel takes a value of 0 or 255. First, since elements of pixels of the smoothed image are composed of BGR, they are converted to grayscale. Next, a binary image is obtained by applying a binarization process to the grayscale image. For the binarization threshold setting, we use Otsu's method⁽⁷⁾ which is generally used.

In the "Reduce small area" process, a small area included in the binary image is painted out and deleted. The value used for filling is set to 255 when the value of the

Fig. 3. Flow chart of circle feature generating.

region is 0, and 0 if the value of the region is composed of 255. First, in order to determine the area to be filled, a labeling process is performed on the binary image. In the labeling process, the same number is assigned to adjacent pixels having the same value except 0, thereby separating the image into several regions. There are two standards to be regarded as adjacent, there are four connection standards targeting one pixel above, below, right and left of the pixel of interest and eight connection standards targeting one pixel around the pixel of interest. In this case, labeling processing based on eight consolidation standards is performed. As a result, numbers are allocated to each area in which the value in the binary image is composed of 255. Also, the total number of numbers for each area corresponds to the size of the area. Therefore, those with an area size of n% or less with respect to the largest area size are subject to filling. This allows you to delete small areas in the image. Also, in the labeling process, since the same labels are allocated to all areas whose value is 0, filling can not be done. Hence, after applying a filling process to the binary image, the value of the binary image is inverted and the filling process is performed again.

In the "Detect contour" processing, the outline of the binary image is detected. There are edge detection method⁽⁸⁾ by Canny filter and contour detection⁽⁹⁾ by joining line segments in contour detection method, and the latter method is used in this research. As a result, for the pixels in the binary image, the boundary between 0 and 255 is detected as a contour.

In the "Hough transform" processing, a circle is detected from a binary image by the Hough transform of circle detection. In the proposed method, it is assumed that the voting is 35 or more and the radius r of the circle is detected when $0.2 \leq r \leq l \cdot 0.5$ with respect to the length l of the image side. Also, assuming that a circle is detected in the order of voting, we assumed that the distance between the circle and the center point of the circle is separated by $l \cdot 0.5$ pixels or more as a condition for detecting multiple circles. By this limitation, it is possible to limit the circle detection to about the two most similar shapes, and then the circle-like shape. In the next branching process, if the entire circle is not contained in the image, the parameter of the circle is removed. Thereafter, if a circle is held, it is assumed that there is a circle in the image and it is considered that there is a circular feature. In the case of giving a circular feature, as a detailed feature, a value based on the image size is calculated for the center coordinates of the circle.

Using equation (1) for this calculation. l_x and l_y correspond to the vertical length and the horizontal length of the image. Also, c_x and c_y correspond to the x and y coordinates of the center point of the circle. The above information is recorded as a circular feature.

$$\begin{pmatrix} cf_x \\ cf_y \end{pmatrix} = \begin{pmatrix} 1/l_x & 0 \\ 0 & 1/l_y \end{pmatrix} \begin{pmatrix} c_x \\ c_y \end{pmatrix} \quad (1)$$

Fig. 4 shows a flowchart for detecting a straight line. In Fig. 4, the processing from "Smoothing" processing to "Extract contour" processing is the same preprocessing as in FIG. 3. In the straight line detection, "Linear approximation" processing is added to preprocessing. In the "Linear approximation" processing, a straight line approximation is made to the contour line detected by the "Extract contour" processing. By converting the contour into a straight line to some extent, it makes it detectable by stochastic Hough transformation. The contour line after "Extract contour" processing produces a contour at the edge of the image. Therefore, unnecessary contours are deleted by painting an area of 5 pixels from the edge of the image with 0. For the linear approximation, we use the Ramer-Douglas-Peucker algorithm⁽¹⁰⁾. In this method, straight line approximation is realized by thinning point clouds constituting a contour line depending on an arbitrary approximation precision with respect to a contour line having a start point and an end point.

Fig. 4. Flow chart of line feature generating.

In the "probabilistic Hough transform" process, a straight line is detected by performing stochastic Hough transformation on the linearly approximated contour image. In order to detect a remarkable straight line in the pattern, it is assumed that only a straight line having a length of $0.85 \cdot l$ or more with respect to the length l of one side of the image is detected. Also, in the probabilistic Hough transform, if the end points of a plurality of line segments are within a certain distance, there is a mechanism of considering it as one line segment. For this reason, straight lines consisting of two or more line segments detected after stochastic Hough transform are rejected from the detection result. This process is executed at the first branch. In the second branch, it is judged which of the vertical line and the horizontal line the detected straight line corresponds to. Therefore, the difference between the end points in the x direction and the y direction of the straight line is calculated by Expression (2).

$$\begin{aligned} difference_x &= |line_{sx} - line_{ex}| \\ difference_y &= |line_{sy} - line_{ey}| \end{aligned} \quad (2)$$

The $line_s$ indicates the start point of the straight line and $line_e$ indicates the end point of the straight line, but the subscripts x and y represent the coordinates of that point. If $difference_x \leq difference_y$, it is judged that the straight line corresponds to the vertical line in particular, otherwise, judged that the straight line corresponds to the characteristic of the horizontal line. After judging the nature of the straight line, it is assumed that there is a feature of a straight line shape with properties of the direction judged by the formula (2) on the image. In the case of the vertical line

feature, in the case of the horizontal line feature, the difference is treated as a detailed feature. The above information is recorded as a feature of a linear shape.

2.3 Image Features Based on Color Space

Color information is used as one of the basic features in image processing. It is treated as an important factor especially when measuring similarity of images independently of shape. Various kinds of color space systems are proposed for expressing colors. Among them, RGB space and HSV space can be mentioned as representative color space systems. The RGB space is a color space system used when handling image data on a computer. Red, green, blue), the color can be expressed. While it is considered to be the most general color space system, it is said that human color perception makes it difficult to understand how colors change by changing intensity of RGB. For the RGB space, the HSV space is said to be a color space that closely resembles human color perception. It is composed of three elements of H (hue) S (saturation) V (brightness), and it is possible to handle colors in terms of how to adjust the vividness and brightness of a color for a certain color. The similarity comparison method based on color information is practiced in various color spaces including the above two kinds of color spaces and it is important to use a color space suitable for the purpose of implementation. Particularly, when using color space, it is pointed out that the accuracy depends on the brightness of the image.

In this research, RGB space was used as the color space for acquiring color information. It aims to compare similarity by using rough hue in the pattern image. Also, in order to suppress the influence of brightness change of the image, color reduction processing and lightness normalization are performed as preprocessing. This processing procedure is shown in Fig. 5.

In this method, as a "Smoothing" process, smoothen a clothing pattern image. In "Lightness normalization" processing, the brightness of the image is adjusted by normalization. As a result, it is possible to correct the color elements of the image in which the overall color tone is too bright or too dark. Specifically, the original image is separated into an R element image, a G element image, and a B element image, and the dynamic range is corrected respectively. Color processing "reduces each gradation of an image from 256 gradations to n gradations lower than that.

In the "make Histogram" process, the frequency of the value of each color element of the image is converted into a histogram. The creation of the histogram in this method is

Fig. 5. Flow chart of color feature generating.

performed according to the following procedure.

- (a) Create three matrices h with 256 voting areas for RGB.
- (b) The image I that is the object of creating the histogram is divided into images for each component.
- (c) Scan the value v for each image pixel by pixel and increment by 1 the element with the number corresponding to the value v for the matrix.
- (d) Suppose that after scanning all the pixels, the matrix being held is handled as a histogram corresponding to the color element.

However, the sum of the elements of the histogram is proportional to the size of the image. Therefore, by normalizing each element of the histogram with the image size, dependence on the image size is solved. Moreover, it is pointed out that when a feature is created for the entire image, it is not possible to utilize the positional relationship of colors⁽¹¹⁾. Therefore, the image is divided into 2×2 , 4 regions to acquire histograms of the image regions. As a result, twelve histograms are created by 4 regions $\times$ 3 components per clothing pattern image.

In the "Calculation centroid of histogram" process, a weighted average is calculated for each histogram with elements as weights. By using the center of gravity, it is possible to emphasize element numbers with large values within the elements of the histogram. Also, the influence of element numbers with small values can be almost ignored. Can be regarded as aggregating low pixels. In the proposed method, twelve centroids are treated as color features of the

image.

Also, when using the center of gravity of the histogram, the same center of gravity position may be taken even if the shape of the histogram is different. This is the case where the center of gravity is placed in the vicinity of the center of the region where the values concentrate in a plurality of regions far away from the center of gravity when the histogram is concentrated in the limited region. This problem is solved by determining the shape of the histogram according to the distance between each center of gravity position and the maximum value element number of the histogram. Specifically, if the distance between each center of gravity position and the maximum value element number of the histogram is within the threshold value, it is unimodal, otherwise it is determined that it has a mountain of more than two peaks. By adding this information to the position of the center of gravity and using it as the feature information, it is possible to handle information as to what kind of shape histogram the center of gravity position is placed.

2.4 Frequency Characteristics of Images Based on Discrete Cosine Transform

Discrete cosine transform (DCT) is a technique often used for transforming images into frequency space. A major application example is adaptation to JPEG. By using the components in the frequency domain, there is an advantage that the compression ratio can be adjusted at the time of image compression. DCT transformation for images is called two-dimensional DCT transformation.

In case of comparing image types using frequency components of DCT, by creating a model based on low frequency components, it is possible to detect images that approximate rough shapes⁽¹²⁾. However, in this research, in order to identify the same shape, we created a model focusing on the input components of DCT. Figure 6 shows a flow chart that creates a feature model of a pattern by DCT by discarding low components and adopting components as a special feature of how the edge in the image are distributed indicate.

In this method, as a "Smoothing" process, smoothen a clothing pattern image. Here, smoothing is performed to reduce image noise. In addition, "Convert color space RGB to Gray" processing converts the color space of the image to gray scale in order to adapt DCT to the image. In the "DCT" process, DCT is performed on the gray image to acquire the frequency component of the image. In the "remove spectrum" processing, frequency components in an arbitrary band are retained and the remaining frequency components

Fig. 6. Flow chart of DCT feature generating.

are discarded. By this processing, only high frequency components can be stored. In "Coding", replace the frequency component with two kinds of codes of 0 or 1. If the DCT coefficient is larger than 0, it becomes 0 if it is 0 or less. Suppose that this code is treated as a frequency feature model of the image.

2.5 Process of Creating Comparative Model and Calculation Similarity

In this process, in order to calculate similarity with the comparison source model, a feature model of the input image is created, and similarity between models is calculated. The features of the model to be compared are similar to those of the comparison source model, and three types of shape features based on Hough transform, features based on color space, and frequency features based on DCT are used. However, while creating a comparison source model gives a pattern image of clothing, creation of a comparison object model does not limit the input image. Therefore, it is necessary to determine which region in the image to create the feature model. Therefore, in this method, template matching is performed between the clothing pattern image as the comparison source and the input image to be compared, and a comparison target model is created from the detected region. Use zero mean normalized mutual function for template matching.

In this method, template matching is performed on the target image while changing the size of the template image, thereby determining the image region for creating the comparison target feature model. Thereafter, as a similarity

calculation of the feature model, when each feature matches $n\%$ or more, it is judged that it is the same pattern image.

3. Experiment

In order to measure the effectiveness of the proposed method, a child identification experiment was conducted. The image to be identified is shown in Fig. 7. The (a) and (c) in Fig. 7 are images of the child to be identified, and the (b) and (d) are garment pattern images manually cut out. In this chapter, children of these subjects are called A, B from the left. The size of the clothing pattern image is 128×128 .

We use 155 images taken by children as a test image in the facility. First, for these images, an algorithm⁽¹³⁾ for detecting the human region is applied and a region of the child is cut out from the image. As a result, 416 person region images were acquired from 155 test images. The proposed method was applied to these images, and when a model matching the comparison source garment pattern model was detected, it was judged that the child of the comparison source model was shown. The areas containing each child are shown in Figures 8 and 9. For each figure, the area image enclosed by the black frame shows that it was detected from the same image. The evaluation criterion is expressed by the formula (3).

$$f1 - measure = \frac{2 * precision * recall}{precision + recall}$$

$$precision = \frac{true\ positive}{ture\ positive + false\ positive} \quad (3)$$

$$recall = \frac{true\ positive}{ture\ positive + false\ negative}$$

$f1-measure$ is the evaluation criterion and takes a range from 0 to 1. The closer this value is to 1, the higher the effect of the method is. There are two elements in the calculation of $f1-measure$, which represent $precision$ and $recall$,

Table 1. Title.

Text	Text	Text
0.0	1.0	2.0
3.0	4.0	5.0
6.0	7.0	8.0

respectively. The $precision$ indicates how much of the correct image is included in the detected image. $Recall$ represents how far a correct answer image that should be detected should be detected. In equation (3), $true\ positive$ represents the correct image detected by the method. Also, $false\ positives$ represent incorrect images detected by the method,

Fig. 7. Target images of garment pattern.

Fig. 8. Child images. (A).

Fig. 9. Child images. (B).

and $false\ negative$ is not detected correct images.

In the proposed method, there are multiple parameters to be set. Therefore, measurement experiments of identification accuracy were conducted for a plurality of combinations for each parameter. As the result of experiment using each pattern image, the combination with the highest accuracy is shown. In the case of A, the number of detections of correct images was 8 out of 11, and that of erroneous detection of incorrect images was 3 out of 405. As a result, the $precision$ was 0.72, the $recall$ was 0.72, and the $f1-measure$ at this time was 0.72. In the case of B, 4 out of 11 detections of correct images were detected, and 0 out of 405 erroneous detections of incorrect images were detected. As a

result, the *precision* was 1 and the *recall* was 0.36, and the *f1-measure* at this time was 0.53.

4. Conclusions

In this research, we examined garment identification method to apply to person identification. This is to utilize the fact that it is possible to identify even an image that does not depend on the orientation of a person by using clothing information and can not be detected by a method using biometric information. Therefore, we examined the method assuming to use it for kindergartens · nursery school. As a concrete method, we proposed a garment identification method based on matching of color features and shape features of patterns. Three types of feature information, shape feature, color feature, and frequency feature are used, and these feature models are created from pattern images of children's clothes that are arbitrarily specified. It is judged whether or not a similar pattern exists in the image depending on the created model and the similarity of the feature model created from the image to be compared. When it is judged that there is a pattern, it can be regarded that there is a child wearing the clothing having the pattern in the image.

In the experiment, a feature model was created from two types of pattern images, and an attempt was made to identify an image in which a similar pattern was displayed from the test image group, and its accuracy was measured. We confirmed that correct answer images can be identified with high accuracy by setting appropriate parameters from each experiment. As a future task, we propose a decision formula for setting appropriate parameters for arbitrary pattern images. Therefore, it is necessary to analyze the relationship between each feature and the parameters for comparing them and clarify the components of the decision formula. Further, in order to reduce the dependence on the change of the parameters, it is also a problem to devise more a method of creating each feature model.

References

- (1) T. Urai, D. Okunaka and M. Tokumaru, "Garment image retrieval based on a similarity evaluation method for Kansei retrieval system", Fuzzy System Symposium, pp.876-881, 2012
- (2) M. Mizouti, A. Kanezaki and T. Harada, "Garment Retrieval Based on Local Similarity with Multiple Images", Proceedings of the 22nd ACM international conference on Multimedia, pp.1165-1168, 2014
- (3) M. Chiba, J. Satake and J. Miura, "Improvement of a SIFT-based people identification for a people following robot Use of a distance-dependent appearance model", 54th Automatic Control Union Lecture Meeting, pp.76-81, 2011
- (4) J. Yamaguchi, K. Shimada, S. Enokida and T. Ejima, "Personal Identification Using Facial Feature and Context for Partially Occluded Images", Technical research report of the Institute of Electronics, Information and Communication Engineers, vol.109, No.470, pp.25-30, 2010
- (5) R. Fisher, S. Perkins, A. Walker and E. Wolfart, "Hough Transform", <http://homepages.inf.ed.ac.uk/rbf/HIPR2/hough.htm>
- (6) J. Matas, C. Galambos and J.V. Kittler, "Robust Detection of Lines Using the Progressive Probabilistic Hough Transform", Computer Vision and Image Understanding, vol.78, No.1, pp.119-137, 2000
- (7) N. Otsu, "A threshold selection method from gray-level histogram", IEEE Trans. Systems, Man and Cybernetics, vol.SMC-9, pp.64-69, 1979
- (8) J. Canny, "A Computational Approach to Edge Detection", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.8, No.6, pp.679-698, 1986
- (9) S. Suzuki and K. Abe, "Topological Structural Analysis of Digitized Binary Images by Border Following", CVGIP, vol.30, No.1, pp.32-46, 1985
- (10) S. H. Paul and M. Garland, "Survey of Polygonal Surface Simplification Algorithms", <http://www.cs.cmu.edu/afs/cs/user/garland/www/Papers/simp.pdf>, 1997
- (11) T. Shibai and Y. Ueno, "Image Retrieval using Color-Position Features in Color Space", IFT 2002, pp.47-48, 2002
- (12) Q. Chen, "Similar Image Retrieval Using Histogram Features", TELECOM FRONTIER, pp.1-7, No.86, 2015
- (13) J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", CVPR 2016, <https://arxiv.org/abs/1506.02640>, pp.1-10, 2015