Clothing Classification Using Image Features Derived from Clothing Fabrics, Wrinkles and Cloth Overlaps

Kimitoshi YAMAZAKI[*1,*2], Masayuki INABA[*3]

Abstract— **This paper describes about a method of clothing classification using a single image. The method assumes to be used for building autonomous systems, with the purpose of recognizing day-to-day clothing thrown casually. A set of Gabor filters is applied to an input image, and then several image features that are invariant to translation, rotation and scale are generated. In this paper, we propose the descriptions of the features with focusing on clothing fabrics, wrinkles and cloth overlaps. Experiments of state description and classification using real clothing show the effectiveness of the proposed method.**

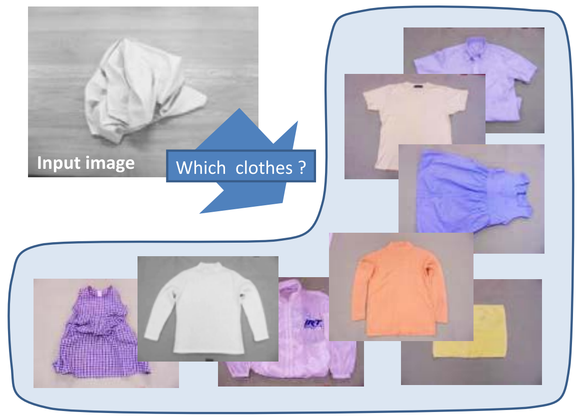

Fig. 1. Concept of our clothing classification. One grayscale image is input, and proper fabric goods is selected from the list of clothing given in advance.

I. Introduction

In daily environments, people use various types of clothing. One of the effective contributions of robots will be to have the ability of doing laundry because people have an excess of tedious housework to accomplish. In this paper, image feature descriptors for classifying clothing are introduced. These methods are useful for building autonomous systems, with the purpose of recognizing day-to-day clothing thrown casually. Fig. 1 shows the problem that we address. The purpose of this research is to correctly guess a class of clothing from one grayscale image that captures a single article of clothing.

Compared to the approaches used for the recognition of solid objects, soft objects such as clothing pose significant challenges because of their variable shape and appearance. Although recently developed image features [8] [11] [13] provide highly reliable results for the purpose of object detection, these features cannot be used for clothing because there is an assumption that the transformations are limited only to solid objects.

However, some robotics researchers who build robots to perform cloth manipulation have proposed the use of various types of image features. Kakikura et al. [7] proposed selecting a target clothing by using color information and achieved success for an isolated task. Willimon et al. [24] also proposed clothing classification. In their isolation phase, graph-based segmentation algorithm was used for deciding target clothing. Ono et al. [16] targeted square-shaped cloth and proposed a description of the bending state based on its contours. Cuén-Rochín et al. [4] proposed action selection for manipulating deformable planar objects. By using physical model, a real robot straightened a square-shaped cloth. Kita et al. [9] utilized a 3D deformable model, and obtained a correspondence between the model and an input pointcloud that was captured by a trinocular stereo camera.

In this previous work, type of the clothing was given in advance. Otherwise, knowledge that was useful for the target identification was simple such as material color. Thus, if there is a recognition module using more generic information, it is useful for developing a general-purpose autonomous system for applications such as daily assistance and industrial laundry. From this viewpoint, other researchers have succeeded in classifying clothing type. Osawa et al. [17] achieved this type of classification while handling clothing. The work by Abbeel et al. [14] demonstrated impressive results and recognized clothing categories by using sensory information while handling the clothing.

If there is a recognition module that enables a robot to identify each clothing without touching, it will be possible to plan an effective handling motion that responds to the clothing. For this reason, we aim to distinguish each clothing without performing any handling. In this paper, we propose image feature descriptors focusing on clothing fabrics, wrinkles and cloth overlaps. Only an image that captures target clothing for recognition is needed. Their distinctiveness is proven by means of experiments that use real images.

The paper is organized as follows: Section II explains how to extract the information of fabrics, wrinkles and cloth overlaps. Section III introduces feature descriptors and their organization. Section IV shows experimental results, and Section V presents the conclusions of this paper.

[*1] Faculty of Engineering, Shinshu University, 4-17-1 Wakasato, Nagano, Nagano, 380-8553, Japan. [*2] PREST, Japan Science and Technology Agency (JST), 4-1-8 Honcho Kawaguchi, Saitama, Japan. [*3] Department of Mechano–Informatics, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan. {yamazaki, inaba}@jsk.t.u-tokyo.ac.jp

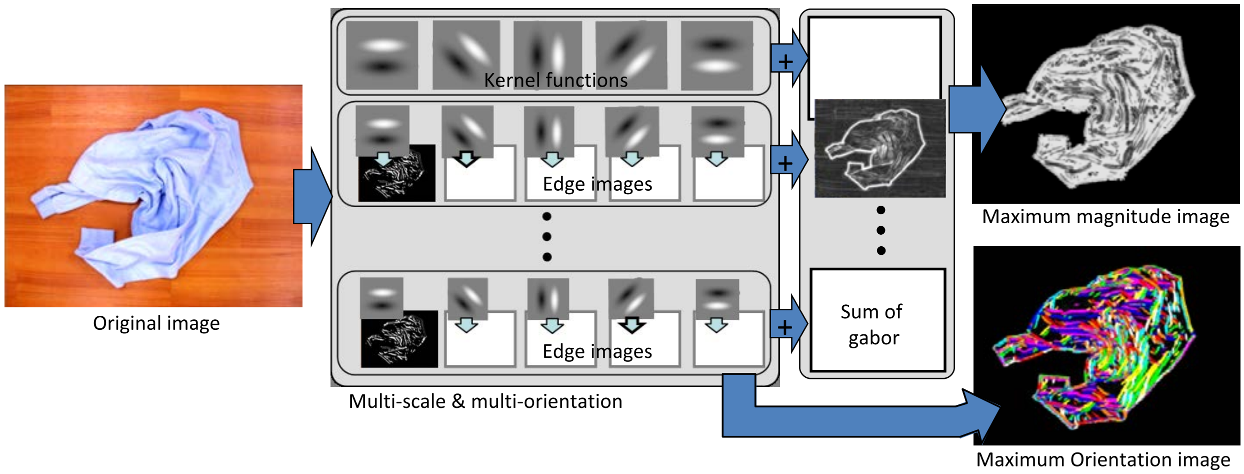

Fig. 2. Multi-scale & multi-orientation filtering. In the maximum magnitude image, the brighter pixel indicates a pixel which reacts in the larger scale parameters. In the maximum orientation image, the difference of pixel color indicates that of the direction of each wrinkle region.

II. Filtering for Describing Clothing Fabrics, Wrinkles and Cloth Overlaps

Our previous paper [21] explains a method for understanding the state of clothing. Multi-scale and multi-orientation filtering are applied to an input image that captures a place of clothing, and these features are extracted and analyzed. In this section, an updated method is described because the filtering process is essential for the generation of feature descriptors, which is the main contribution of this research.

The leftmost figure in Fig.2 shows clothing that is placed on the floor. This image shows a shirt with long sleeves, which is made out of cotton. We can divide this clothing image into several parts: (i) cuffs or other specific parts, (ii) wrinkles and (iii) cloth overlaps. The information in (i) tells us what type of clothing it is, but the result depends on the placement of the clothing. In contrast, (ii) and (iii) are always observable, and could provide us with useful information. Moreover, the type of wrinkle depends on the clothing fabrics. From this reason, we extract and analyze them.

A. Outline

We focus on the fact that the contrast in the image region with respect to the clothing shows gradual changes in the frequency domain. In other words, some parts of the clothing can have stripe-shaped states due to soft material.

To analyze this property, a set of Gabor filters is applied to an input image. This approach is similar to a filter bank [6] [19], the most frequent application of which is texture classification [12]. In our case, because we assume that wrinkles and cloth overlaps in images are derived from a combination of waves that have directionality and gradual frequency, Gabor filters are suitable to describe them. In the filtering, the parameters of the wave profile change, and then, helpful information is extracted from the convolution results. For example, high frequency coefficients often highlight contours and cloth overlaps, whereas low frequency coefficients persistently respond to wrinkles.

Fig.2 shows the concept of the filtering. Various kernel functions of Gabor filters are prepared, and the filtering results are used for generating two proposed images. One intermediate image is called the "maximum magnitude image" for the remainder of this paper. Each pixel in this image has a value that is related to a Gaussian variance, which provides a maximum magnitude during the scale space. Another is called the "maximum orientation image". Each pixel in this image has a value that is related to an angle that indicates, on a scale, the maximum reaction among the Gabor filter.

Especially, maximum magnitude image is an original point of this research. The generation process is similar to the MR8 filter bank [23] halfway, but we produce an image with 8 shades of gray, each picture cell has a frequency coefficient. This image is powerful to describe the difference of clothing fabrics and the type of wrinkles that depend on fabrics and textile form.

B. Gabor Filter

A 2-dimensional Gabor filter [25] is a filter in which the direction and frequency can be arbitrarily changed. The filter has often been applied to scale space analysis. The corresponding equation is as follows:

$$g(\mathbf{x}, \theta, \sigma_x, \sigma_y) = \frac{1}{\sqrt{2\pi}\sigma_x\sigma_y} e^a \cos(2\pi f x_\theta + p), \qquad (1)$$

where

$$\begin{aligned} a &= -\frac{1}{2}\left(\frac{x_\theta^2}{\sigma_x^2} + \frac{y_\theta^2}{\sigma_y^2}\right), \\ x_\theta &= (x - u_x)\cos\theta + (y - u_y)\sin\theta, \\ y_\theta &= -(x - u_x)\sin\theta + (y - u_y)\cos\theta. \end{aligned} \qquad (2)$$

The f is frequency domain, which depends on the variance value σ. The variables x and y are the coordinates of the present pixel, and u_x and u_y are the center coordinates of the Gaussian distribution. The variables σ_x^2 and σ_y^2 are the variances; both of them are represented as σ in the rest of this paper. Thus, $f(\mathbf{x}, \theta, \sigma_x, \sigma_y)$ is represented as $f(\mathbf{x}, \theta, \sigma)$. The variable p is a variable of the phase, and we substitute $\pi/2$ into it because the edge detector should be generated in our case.

C. Multi-scale and multi-orientation filtering

As pre-processing to generate a maximum magnitude image and a maximum orientation image, Gabor filters are applied with a constant σ and a variable θ in eq. (1). Because a Gabor filter has directionality, the resulting images contain various edges that rely on the θ setting.

A maximum orientation image $I_{ori}(\mathbf{x})$ is generated from these results. The pixel information is written as follows:

$$I_{ori}(\mathbf{x}) = argmax F_1(\mathbf{x}, \theta), \tag{3}$$

where $\mathbf{x}$ denotes the pixel coordinates and θ denotes the inclination angle of a kernel function in equation (2). The $F_1(\cdot)$ is a continuous function concerning θ and the neighboring pixels.

$$F_1(\mathbf{x}, \theta) = \int_w f(\mathbf{x}) g(\mathbf{x} + \mathbf{x}_0, \theta) d\mathbf{x}_0, \tag{4}$$

where $f(\mathbf{x})$ indicates an input image and w denotes the window size of the convolution.

In practice, $I_{ori}(\cdot)$ is calculated from the discrete values $(\theta_1, \theta_2, \cdots, \theta_K)$. After pre-processing with varying θ, an image is generated by collecting the highest radiance value at the same pixel coordinates in the set of filtered images. In the case of clothing, ellipse-like regions are extracted along wrinkle directions.

In contrast, for generating a maximum magnitude image, the following procedure is needed. First, the multi-orientation filtering mentioned above is performed, and then, we obtain an image whose pixels are the sum of the result of the Gabor filtering with varying θ. (In Fig.2, 'Sum of Gabor' indicates it.) We call the image 'temporal image' in the rest of this section.

By varying σ, a set of temporal images is generated. The maximum magnitude image consists of pixels that indicate the maximum radiance in the temporal images. We define a pixel of the maximum magnitude image $I_{mag}(\mathbf{x})$, which can be written as the following:

$$I_{mag}(\mathbf{x}) = argmax F_2(\mathbf{x}, \sigma), \tag{5}$$

where

$$F_2(\mathbf{x}, \sigma) = \int_\theta \int_w f(\mathbf{x}) g(\mathbf{x} + \mathbf{x}_0, \theta, \sigma) d\mathbf{x}_0 \mathbf{d}\theta. \tag{6}$$

The window width w and the frequency are decided from the σ value automatically.

To provide the same emphasis to different frequency octaves, some techniques have been proposed. For example, Rubner and Tomasi [18] applied log-Gabor filters [1] because the centers of these filters in the frequency domain are equally spaced in a log polar representation of the spectrum of an image. We do not use this representation; instead, our frequency intervals are determined by reference to Rubner's results.

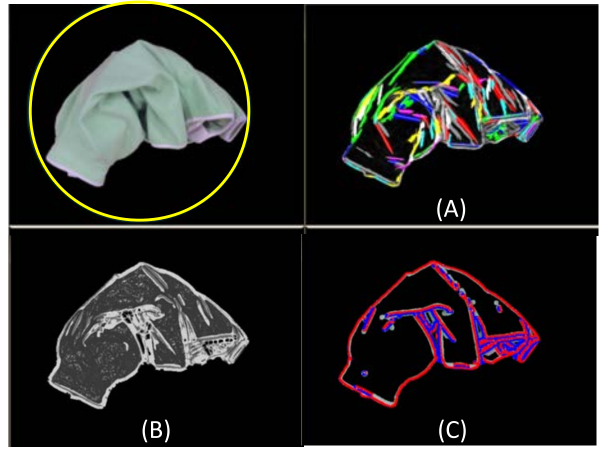

Fig. 3. Examples of the pre-processing. Image (A) shows position and orientation distribution. Wrinkles are divided into a group of ellipses, and are colored depending on the orientation. Image (B) shows density of wrinkles and reaction to fabrics. Pixel radiance shows the difference of maximum frequency in the pixel. Image (C) shows extracted cloth-overlaps. Red pixels is belonging to upside, and blue pixels are of downside.

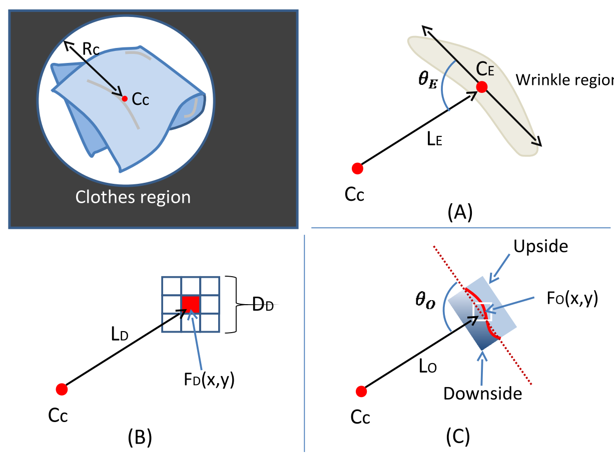

Fig. 4. Feature description

III. Feature Description

In this section, we propose three types of features. All of them are invariant to translation, rotation and scale.

A. Pre-processing

First, the clothing region is extracted from the input image. In the present form, we apply the mean shift-based image segmentation [2], and remove the pixels that belong to the background. After that, as shown in the upper left figure in Fig.3 and 4, a circle that sufficiently includes the target clothing is defined. The radius R_C and its center coordinates C_C are used in the process for making features, as described next.

B. Position and orientation distribution of wrinkles (DST-W)

One of the most important and specific piece of information that describes the state of clothing is the wrinkles. Because this research mainly targets crumpled clothing, as

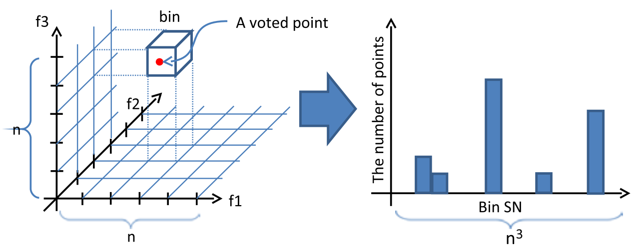

Fig. 5. Histogram calculation from parameter distribution

shown in Fig.2 'original image', we obtain the distribution of wrinkles in a maximum orientation image, which we are able to use to describe the wrinkle distribution.

First, all of the wrinkle regions are divided and approximated into a group of ellipses. Then, to define a wrinkle-based descriptor, the following three criteria are used (see also Fig.4(A)):

1) Proportion of R_C to the long axis of a wrinkly ellipse.
2) Proportion of R_C to L_E, where L_E is a line segment connecting C_C with the center of the ellipse C_E.
3) Relative angle θ_E between line segment L_E and the long axis of the ellipse.

All of the wrinkles that were extracted from an image are parameterized according to the above three criteria, and then the results are projected into the 3D parameter space. After that, as shown in Fig.5, the space is regarded as divided into small voxels, and number of points in each voxel are counted. As a result, we can obtain a frequency histogram to describe the clothing in the image.

C. Clothing fabric and wrinkle density (CF-WD)

Clothing used in daily living is made of various material such as cotton, polyester, acrylic. In addition, there are many types of clothing fabric, such as pile and shirring. Extracting these characteristics from the visible surface of clothing is a great help in classifying them. Moreover, when analyzing the organization of wrinkles, clothing fabric is dominant to the organization. A maximum magnitude image makes it possible to describe these elements distinctively.

A feature descriptor derived from clothing fabric and wrinkle density is generated from all of the pixels that form the clothing region in a maximum magnitude image. There are three criteria as follows (see Fig.4(B)):

1) The proportion of R_C to L_D, where L_D is the distance between C_C and an arbitrary pixel D that is in the clothing region with coordinates (x, y).
2) Pixel brightness $F_D(x, y)$.
3) The sum of the difference between D and its surrounding 8 neighbors, in other word, $D_D = \sum_{i,j \in W}(F_D(x, y) - F_D(x+i, y+j))$.

These results are translated into a frequency histogram by the same procedure as described in the above subsection.

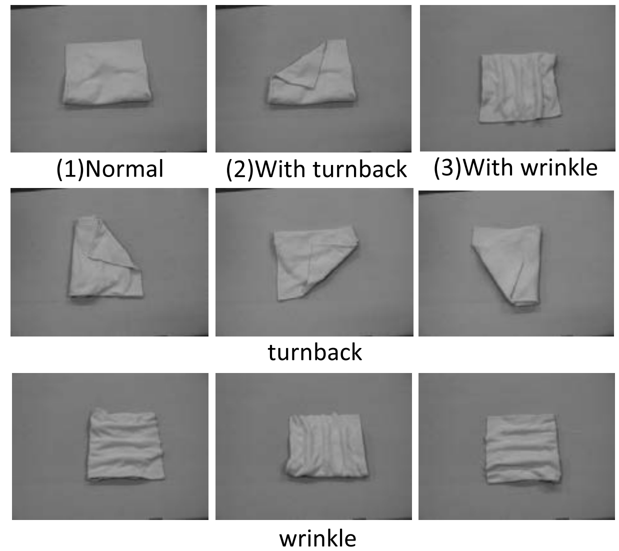

Fig. 6. A set of cloth images with simple variation in the form

D. Existence of cloth-overlaps (OVLP)

For making a feature out of the state of cloth overlaps, a pre-processing that estimates the up and down relationship of clothing edges and boundaries is performed.

The procedure is as follows: Pick up a certain temporary image (Sum of Gabor in Fig.2) which is filtered by a relatively high frequency σ, and an edge line is detected from the image. The edge lines become candidates for the boundaries of the cloth overlaps. After that, a narrow band is set around the edge lines, and then the brightness difference between the regions divided by the line is calculated at an original image. In our implementation, which is described in the next section, we selected an image that is filtered by a $\sigma = 4.0$ Gaussian, and we set the band-width to 5 pixels. An example of the extraction result is shown in Fig.3(C).

From these results, we obtain the following three types of features (see Fig.4(C)):

1) The proportion of R_C to L_O, where L_O is the distance between C_C and an arbitrary pixel O, which is in the clothing region with coordinates (x, y).
2) The value E_O, which is the level of overlapping and its side (up or down).
3) The relative angle θ_O between the boundary edge and L_O.

These results are also translated into frequency histograms by the same procedure as described in the above subsection.

IV. Proof with experiments on clothing classification

A. Basic evaluation of the proposed feature descriptions

Fundamental experiment was performed to investigate the function of feature descriptions against to clothing image. For each clothing shown in Fig. 10, three patterns of configurations were used: (1) an article of clothing that was folded into a square, (2) one corner of the clothing was turned down and

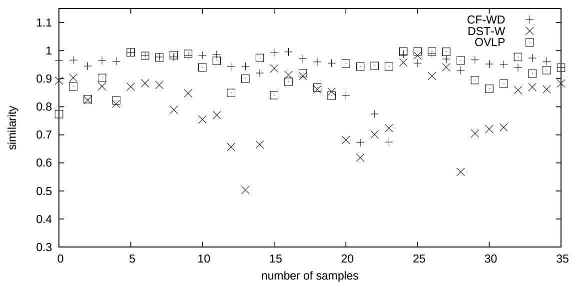

Fig. 7. Similarity values of clothes with different shape

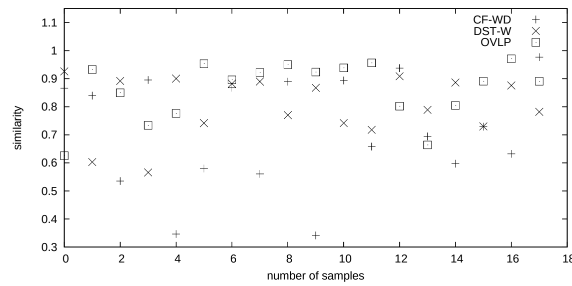

Fig. 8. Similarity values of the different type of clothes

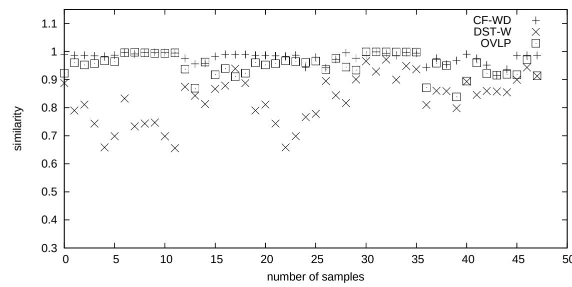

Fig. 9. Similarity values of clothes with different placement

(3) three rectilinear wrinkles were added. For each patterns, four series of placements that rotated the clothing at intervals of 90 degrees were captured. Fig. 6 shows an example that is of an article of clothing as shown in Fig. 10, (11). Images in the middle row shows clothing images with turning down one corner, and images in the lowest row shows wrinkled clothing with rotational changes.

Image pairs were generated according to the following three rules: (A) Same type of clothing but one is folded into a square and another has a turnback, (B) Different type of clothing but they have turnback in the same position and (C) Same type of clothing and they have turnback. However, their placements (rotation angle) are different. Feature descriptions were generated from between two images, and cosine similarity between them were calculated. Fig. 7, 8 and 9 show a part of results. Three descriptions **CF-WD**, **DST-W** and **OVLP** were indicated in these graphs. Horizontal axis presents serial number of the pairs, and vertical axis is similarity value. Fig. 7 shows the result of (A), which shows that highly similar values were presented on **CF-WD** regardless of whether turnback exists. Fig. 8 shows the result of (B), which compared different type of clothing. It could expect to show low similarity. Although there were inter-individual variability, **CF-WD** and **DST-W** indicated strong tendency of it comparing with **OVLP**. Fig.9 shows the result of (C). **CF-WD** had high similar values and **OVLP** was second. Meanwhile, **DST-W** indicated basically low similarity value because it was influenced the difference of placement.

These results conducted the following discussion. **CF-WD** is suitable for clothing classification under the condition that an article of clothing was thrown casually. This is because all of similarity values were over 0.95 in Fig. 7 while values were lower than 0.9 in Fig. 9 that compared different type of clothing. **CF-WD** focuses on the difference of fabric patterns and wrinkles instead of a shape of clothing. In the meanwhile, because **DST-W** involves the information of the shape of clothing, it may suit better for judgement of the shape classification. This is our future work.

B. Settings for the experiment of clothing classification

An image database that included 21 types of clothing was prepared, so that experiments could be performed on clothing classification. Fig.10 shows the clothing. In regard to the top rank of the figures, the list of clothing is as follows: (1) turtleneck (100% cotton), (2) parka (100% polyester), (3) sweater (100% cotton), (4) cardigan (100% acrylic fablic), (5) one-piece suit (100% cotton), (6) T-shirt (100% cotton), (7) hand-towel (100% cotton). Both (7) and (8) are pile fabric but their spreading size was different.

About 200 to 300 images the size of which was VGA (640×480 pixels) were captured from each piece of clothing by throwing it randomly, so that the total number of images in the database was over 6200. The distance between a camera and a cloth was about 900 [mm]. Fig.11 shows some images in the dataset. There were various state of clothing that were scrumpled or smoothed.

The color information was not used in the following experiments.

C. Additional feature descriptor

Besides the feature descriptors that were presented in section III, we attempted to implement some other representations.

Scale space extrema (SSEX): The well-known image feature descriptor SIFT [13] uses a scale space extrema for generating features with scale invariance. Based on the success of previous research that copes with object recognition, we also uses scale space extrema observed at clothing region. The procedure is as follows: First, pyramid images are generated from an input image. Next, a set of DoG (Difference of Gaussian) filters are applied to the images, and then a minimum or maximum value at 3×3 pixels from 3 consecutive pyramid images is extracted.

After that, a feature descriptor is generated from all of the extrema with three criteria listed next:

Fig. 10. The list of clothing used for generating image database

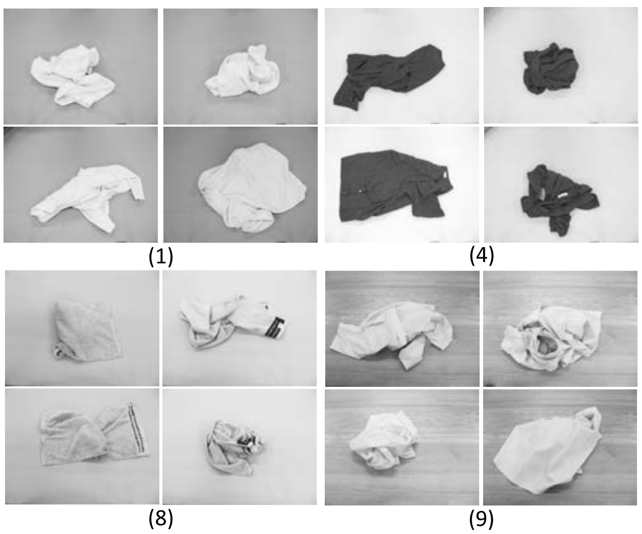

Fig. 11. Examples of image in database

1) Scale value of the extrema.
2) Proportion of R_C to a distance between C_C and the position of the extrema.
3) Proportion of R_C to a distance from the present extrema position to the nearest extrema position.

These results are translated into a frequency histogram by the same procedure as other descriptors described in Section III.

Contour (CNTR): Using the pre-processing of feature descriptions that are mentioned above, we finish a regional extraction of the clothing. Thus, contour information can be used as one feature.

First, a two-valued image that divides a clothing region and a background region is prepared. Next, a log polar transformation is applied to the image for generating the $\theta - r$ graph with centering at C_C. A distance histogram is calculated by discretizing the θ axis, and a feature description is achieved by registering the values of the histogram bins.

TABLE I
CLUSTERING RESULT (CF-WD DESCRIPTOR, 10 CLUSTERS)

Cluster No	1	2	3	4	5	6	7	8	9	10
(6) T shirt	0	64	5	0	0	0	0	47	112	1
(7) Towel	58	0	2	3	82	85	12	47	1	9
(8) Towel	2	0	0	0	28	24	95	1	11	72
(9) Shirt	0	1	119	124	0	0	0	2	0	0

TABLE II
CLUSTERING RESULT (SSEX DESCRIPTOR, 10 CLUSTERS)

Cluster No	1	2	3	4	5	6	7	8	9	10
(6) T shirt	40	68	1	71	5	28	1	30	0	2
(7) Towel	3	10	34	2	23	11	102	28	53	33
(8) Towel	10	3	12	2	23	6	14	52	14	64
(9) Shirt	3	30	41	15	10	92	16	5	16	1

D. *Feature analysis based on unsupervised clustering*

Five types of features (**DST-W**, **CF-WD**, **OVLP**, **SSEX**, and **CNTR**) were generated from respective images in the database. That is, one clothing image was translated into five different features. For investigating the capability of clothing classification, clustering by means of repeated bisection [22] was applied to these features. If the features are suitable for clothing classification, then this clustering will make well-divided clusters.

In this trial, the dimensions of the feature descriptor were set at 64, 512 and 4096. Each dimension was derived from a resolution of voxels in 3D feature space; in other words, if each axis is divided by n, then the feature dimension becomes $n \times n \times n$. We attempted an application with $n = 4$, $n = 8$ and $n = 16$, respectively.

Table I and Table II shows certain clustering results by using about 1000 features from 4 classes of clothing. The clustering results of the **CF-WD** and **SSEX** descriptor, when setting the number of clusters to 10, are represented. Obviously, the **CF-WD** descriptor shows specific clusters

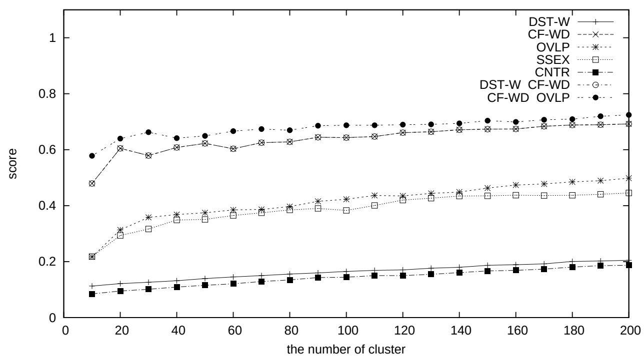

Fig. 12. Clustering results (n = 4)

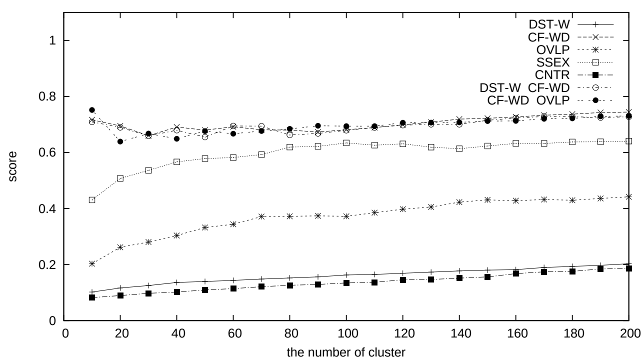

Fig. 13. Clustering results (n = 8)

that rely on the pre-defined clothing class, which is notable because the **CF-WD** descriptor tends to make a similar feature when it is applied to the same type of fabric. Clusters 5 and 6 both contained towels that are made of pile fabric.

The clustering result is evaluated by using the following equation:

$$score = \frac{1}{m}\sum_{i=1}^{m}\frac{N_{max}}{N_c}, \tag{7}$$

where m is the number of clusters. N_c is the number of the features that are contained in each cluster, and N_{max} is the number of features that make up the greatest potion of the cluster. If most of features in one cluster are derived from one type of clothing, N_{max} approaches N_c, and $score$ becomes high. As another case, the score becomes the highest when the number of clusters is set to that of the input features.

Fig.12 and 13 show the score transition depending on the number of clusters. Not only single features, but also some combinations of features were applied by simply concatenating them. The highest score was found in the case of **CF-WD** at $n = 8$. We also tested an $n = 16$ division, but the result was almost the same as 8.

Fig.14 shows some images, the maximum orientation images and their enlarged views. According to the difference in the materials and the types of fabric, various tendencies not only about the pixel radiance but also about the composition of wrinkles were found.

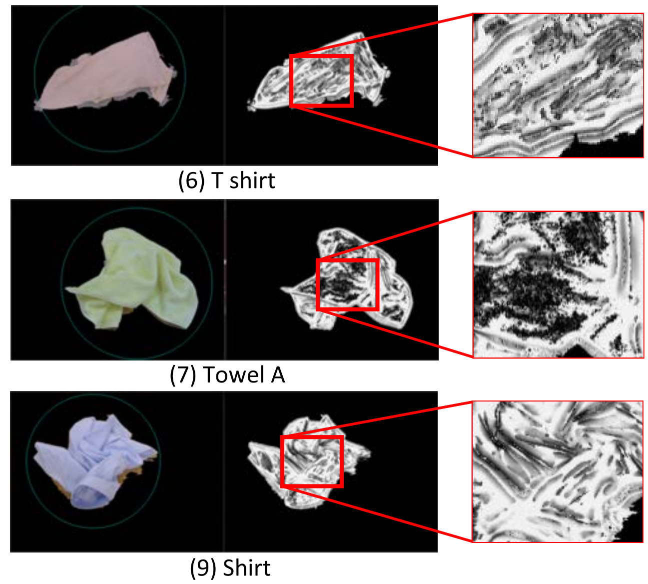

Fig. 14. Examples of maximum magnitude image

E. Classification using supervised learning

Using all of the 21 classes shown in Fig.10, the discriminative function for clothing classification was generated by means of a multi-class SVM (Support Vector Machine). We used LIBSVM [26], and the type of classifier and kernel function were C-SVC and RBF, respectively. Because every description method represented in this paper makes one feature from one image, classifying features is synonymous with the classifying clothing images in dataset.

For evaluating the learning result, N-fold cross validation was applied. Fig.15 shows the accuracy rate of validation when changing N from 2 to 10. As mentioned in section IV.B, about 200 to 300 images were prepared for one class. After they were divided into N groups, one group was used as test data, and the rest of them were training data. **CF-WD** was the best feature, and the next in line was **OVLP**. A combination of these descriptors was also tried, and the best accuracy was achieved when **CF-WD**, **OVLP** and **SSEX** were used simultaneously. The best accuracy rate was 99.07%.

As another comparison, we also implemented a popular method of generic object recognition proposed by Csurka et al. [3]. It is based on “Bag of Keypoints” approach whose keypoint is provided by SIFT descriptor. In our experiments, the number of visual words was determined by means of grid search, and the best number was 1000 that resulted 79% accuracy rate. This was by no means a bad rate, but there was about 20% difference from the proposed method.

F. Robustness to illumination changes

The original dataset introduced in the top of this section was captured in the illumination level of $760[lx]$. To investigate the robustness to illumination changes, additional datasets were prepared. Two datasets that included about 100

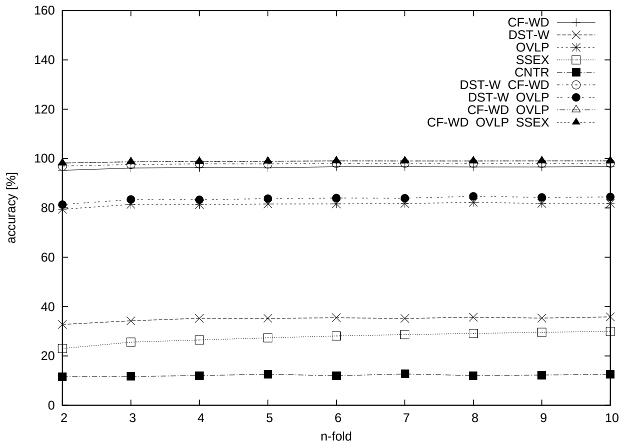

Fig. 15. The result of cross validation

images in each of 21 types of clothing were captured under the illumination conditions of $1060[lx]$ and $1400[lx]$.

First, the two datasets were examined by using a distinctive function trained by the original dataset. SVM was used for the training with the same setting described in the previous subsection. The accuracy rate according to 10-fold cross validation was 85% in $1060[lx]$ case, and 84% in $1400[lx]$ case. From these results, illumination change gives a certain level of influence to the result of feature calculation.

On the other hand, when all of three datasets were used for generating a distinctive function, the accuracy rate was 97.2%. This result tells us that we can perform high-accuracy classification by using a distinctive function trained from images that are captured with various degrees of illuminance.

V. CONCLUSIONS

In this paper, we proposed a state description method for clothing that are randomly placed in a daily environment. A set of Gabor filters are applied to an input image with a range of frequencies and directions, and useful information such as wrinkles and cloth overlaps are detected based on the maximum magnitudes and orientations. Using these results, several feature descriptors are generated. In the experiments on clothing classification that used real images, we achieved more than a 99% success rate.

Future work would include more feasible information about the state of the clothing, which would be added to the classification method. After that addition, we will attempt to develop a method that enables a daily assistive robot to handle daily laundry. Collaboration this research with clothing detection method [20] would have a important role for such application.

ACKNOWLEGEMENT

This work was supported by JST PRESTO program.

REFERENCES

[1] J. Bigü and J. M. du buf: “N-folded symmetries by complex moments in gabor space and their application to unsupervised texture segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-16(1):80–87, 1994.

[2] C. Christoudias, B. Geogescu and P. Meer: “Synergism on Low Level Vision,” Int’l Conf. on Pattern and Recognition, pp. 150 – 155, 2002.

[3] G. Csurka, C. Bray, C. Dance, and L. Fan: “Visual categorization with bags of keypoints,” in Proc. of ECCV Workshop on Statistical Learning in Computer Vision, pp. 59 – 74, 2004.

[4] S. Cuén-Rochín, J. Andrade-Cetto and c. Torras: “Action Selection for Robotic Manipulation of Deformable Planar Objects,” in Proc. of Frontier Science Conference Series for Young Researchers: Experimental Cognitive Robotics, pp. 1-6, 2008.

[5] M. Galun, E. Sharon, R. Basri and A. Brandt: “Texture Segmentation by Multiscale Aggregation of Filter Responses and Shape Elements,” Proc. of IEEE Int’l. Conf. on Computer Vision, pp. 716-723, 2003.

[6] J. Geusebroek, A. Smeulders and J. Weijer: “Fast Anisotropic Gauss Filtering,” IEEE Trans. on Image Processing, 12(8):938-943, 2003.

[7] K. Hamajima and M. Kakikura: “Planning Strategy for Unfolding Task of Clothes – Isolation of clothes from washed mass –,” in Proc. of Int’l. Conf. on Robots and Systems, pp. 1237 – 1242, 2000

[8] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool: “SURF: Speeded Up Robust Features, “ Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346–359, 2008

[9] Y. Kita, F. Saito and N. Kita: “A deformable model driven method for handling clothes,” Proc. of Int. Conf. on Pattern Recognition, 2004.

[10] H. Kobori, Y. Kakiuchi, K. Okada and M. Inaba: “Recognition and Motion Primitives for Autonomous Clothes Unfolding for Humanoid Robot,” in Intelligent Autonomous Systems 11, pp.57–66, 2010.

[11] K. Mikolajczyk and C. Schmid: “Scale and Affine Invariant Interest Point Detectors,” Int’l Journal of Computer Vision vol. 60, No. 1, pp.63 – 86, 2004.

[12] T. Leung and J. Malik: “Representing and recognizing the visual appearance of materials using three-dimensional textons,” International Journal of Computer Vision, 43(1):29-44, June 2001.

[13] D. G. Lowe: “Distinctive image features from scale-invariant keypoints,” Int’l Journal of Computer Vision, vol. 60, No. 2, pp. 91-110, 2004.

[14] J. Maitin-Sphepard, M. Cusumano-Towner, J. Lei and P. Abbeel: “Cloth Grasp Point Detection based on Multiple-View Geometric Cues with Application to Robotic Towel Folding,” Int’l. Conf. on Robotics and Automation, pp.2308 – 2315, 2010

[15] S. Miller, M. Fritz, T. Darrell and P. Abbeel: “Parametrized Shape Models for Clothing,” In the proceedings of the International Conference on Robotics and Automation (ICRA), 2011.

[16] E. Ono, H. Okabe, H. Ichijo and N. Aisaka: “Robot Hand with Sensor for Cloth Handling,” In Proc. 1990, Japan, U.S.A. Symp. on Flexible Automation, pp. 1363–1366, 1990.

[17] F. Osawa, H. Seki, and Y. Kamiya: “Unfolding of Massive Laundry and Classification Types by Dual Manipulator,” Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.11 No.5, pp. 457– , 2007.

[18] Y. Rubner and C. Tomasi: “Coalescing Texture Descriptors,” ARPA Image Understanding Workshop, pp. 927 – 935, 1996.

[19] C. Schmid: “Constructing models for content-based image retrieval,” In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 39–45, 2001.

[20] K. Yamazaki and M. Inaba: “A Cloth Detection Method Based on Image Wrinkle Feature for a Daily Assistive Robots,” IAPR Conf. on Machine Vision Applications, pp.366–369, 2009.

[21] K. Yamazaki, K. Nagahama and M. Inaba: “Daily Clothes Observation from Visible Surfaces Based on Wrinkle and Cloth-Overlap Detection,” IAPR Conf. on Machine Vision Applications, pp., 2011.

[22] Y. Zhao and G. Karypis: “Comparison of agglomerative and partitional document clustering algorithms,” University of Minnesota - Computer Science and Engineering Technical Report, No. 02-014, 2002.

[23] M. Varma and A. Zisserman: “A Statistical Approach to Texture Classification from Single Images,” International Journal of Computer Vision Volume 62, Numbers 1-2, 61-81, 2005.

[24] B. Willimon, S. Birchfleld, I. Walker: “Classification of clothing using interactive perception,” in Proc. of IEEE Int’l Conf. on Robotics and Automation, pp. 1862 - 1868, 2011.

[25] “Tutorial on Gabor Filters,” http://mplab.ucsd.edu/tutorials/tutorials.html

[26] C. Chang and C. Lin: “LIVSVM – A Livrary for Support Vector Machines,” http://www.csie.ntu.edu.tw/ cjlin/libsvm/.