MVA2013 IAPR International Conference on Machine Vision Applications, May 20-23, 2013, Kyoto, JAPAN

Comparing Investigation of Images and 3D Mesh Model For Categorizing Electric Appliance Components

Kimitoshi Yamazaki
Shinshu University
4-17-1, Wakasato, Nagano Japan
kyamazaki@shinshu-u.ac.jp

Ryo Hanai
Advanced Industrial Science and Technology
1-1-1, Umezono, Tsukuba, Japan
ryo.hanai@aist.go.jp

Hiroaki Yaguchi
The Univ. of Tokyo
7-3-1, Bunkyo-ku, Hongo, Japan
hyaguchi@jsk.t.u-tokyo.ac.jp

Kotaro Nagahama
The Univ. of Tokyo
7-3-1, Bunkyo-ku, Hongo, Japan
nagahama@jsk.t.u-tokyo.ac.jp

Katsuyoshi Yamagami
Panasonic Corporation
3-4, Seika-cho, Sagara-gun, Kyoto, Japan
yamagami.katsuyoshi@jp.panasonic.com

Masayuki Inaba
The Univ. of Tokyo
7-3-1, Bunkyo-ku, Hongo, Japan
inaba@jsk.t.u-tokyo.ac.jp

1 Introduction

Knowing about the type of objects that exist in real world is important thing for automation of conveyance or classification task. This paper reports our challenges the purpose of which is to categorize objects that are incorporated into home electric appliances. Our target includes small sized objects, flexible objects, and objects that have various appearances by the difference in view-points. Sensing methods we used are two types of measurement data; (1) Mesh model measured by a 3D digitizer, and (2) multi-viewpoint images captured by a high-resolution camera. Using these sensor data, we studied about feature descriptions. These descriptions are designed by regarding the characteristics of target objects and the property of each measurement data.

Figure 1 shows our target objects that are composed of 31 series of electric appliance components. They include small sized objects (e.g. screw and clasp), relatively large objects (e.g. electrical circuit sized hundreds millimeters), shiny and transparent objects (e.g. glass), and flexible objects (e.g. cable and wire). Meanwhile, some of components have a big difference on their appearances or shapes even though they are grouped as a same category. The goal of this research is to design high-precision classifier for these 31 categorized objects.

Figure 1. 31 series of parts

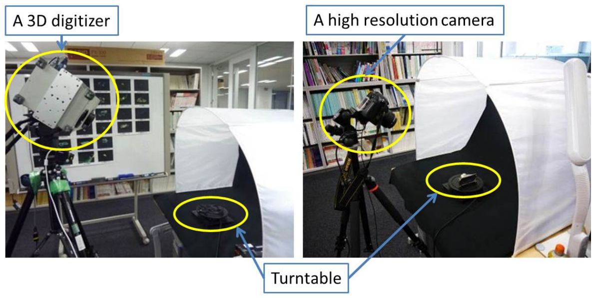

Figure 2. Measurement system

2 Measurement Data

Figure 2 shows two measurement systems that we utilize. They are composed of a sensor (a digital single-lens reflex camera or a 3D digitizer), a turntable, and a computer. During a measurement, an object placed on the turntable is rotated at regular angle intervals. Because all of the measurement procedures are controlled by the computer directly connected to the sensor and the turntable, only we have to do is to just place an object on the turntable.

Figure 3 shows several components we target. The figures numbered (1) show one object which has different shape and appearance changes between two sides. The figures numbered (2) show various appearances caused by their pose, and the figures (3) show some components that can have various shapes despite of the same category. Their colors that are similar to background is also one cause of difficulties. The figures (4) shows two components in the same category. 90 % components in the category is the same color with the left figure, but remaining 10 % components is the same color with right figure.

In general, 3D digitizer is a sensor that projects laser light or regularized patterns, and a measurement will be failed in the part where the projection is insufficient. For instance, the quality of the measurement will become worse or will be lost when object surfaces direct to depth. Other problems are the time of measurement and data merging for getting one complete mesh model. Meanwhile, because the data is represented by 3D shape, getting rid of background data is easier than the case of images.

On the other hand, camera is a completely passive sensor. Even if objects with transparent or shiny surface are targeted, an image captures their influence and property. Sensor information processing with feasible

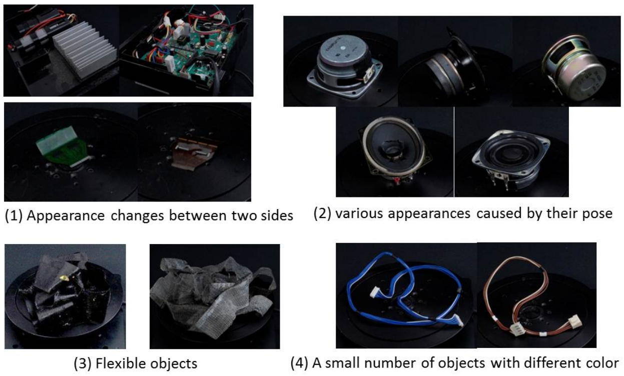

Figure 3. Examples of electric parts

compensation can provide us to get the information of object surfaces even if the cases. Meanwhile, one of the problems is to know the existence of a target object when the object looks like background because of the surface color.

3 Composition of Classifiers

3.1 Basic policy for feature description

One of the keypoints to solve categorization problem with high accuracy is 'feature description'. Feature description that represents large distance for different categories and small distance for the same category enables to implement high accuracy categorization, even if simple linear discriminant function is provided. For this reason, we focus on the design of feature description. In our investigation, using both information about whole shape and local shape is important. Following subsections explain about our feature description calculated from 3D shape model and multiview images.

3.2 An approach for 3D shape-based categorization

As related work, 3D shape retrieval have been studied for classifying an input object into known object categories. Most of these researches, an input model composed of a group of 3d meshes is converted to feature description, and it is evaluated the similarity with pre-defined categories.

Several feature descriptions have been proposed; Extended Gaussian Image (EGI) [1], Concrete Radialized Spherical Projection (CRSP) [2], Spherical Wavelet Descriptor (SWD) [3], and so on. Density Based Framework (DBF) [4] is a framework based on EGI representation. Methods described above need a preprocessing that aligns the direction of shape models. For this purpose, statistical methods, such as CPCA [5] and NPCA [2], are applied.

On the other hand, feature descriptions without pose alignment is needed in our targets. This is because the alignment process do nothing for flexible objects (e.g. cables). For this reason we focus on SPRH (Surflet Pair Relation Histogram) feature [7]. SPRH feature is based on one frequency histogram that captures rough shape information, which is as same as other descriptors described above. In addition, if an object includes many of locally-resembled shape, SPRH feature enables to represent the situation. We implement a discriminant function that includes feature description considering both rough shape information and local shape patterns. Several techniques such as multi-scale description and bag of features are included in this function.

3.3 Feature description by multi-scale SPRH

In our assumption, a shape model is composed of a group of surfaces. Wahl et al. [7] calls a pair of a surface and its normal 'surflet', and proposed SPRH feature by describing the relation between many pairs of surflets sampled from a model.

For a given surflet pair $(\mathbf{p}_1, \mathbf{n}_1), (\mathbf{p}_2, \mathbf{n}_2)$, first a coordinate system is defined. If the origin is chosen to be $\mathbf{p}_1$, the following base vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ are defined:

$$\mathbf{u} = \mathbf{n_1}, \mathbf{v} = \frac{(\mathbf{p_2} - \mathbf{p_1}) \times \mathbf{u}}{\|(\mathbf{p_2} - \mathbf{p_1}) \times \mathbf{u}\|}, \mathbf{w} = \mathbf{u} \times \mathbf{v}. \tag{1}$$

Using these base vectors, the relation between two surflets are described by four parameters:

$$\begin{aligned} \alpha &= \arctan(\mathbf{w} \circ \mathbf{n_2}, \mathbf{u} \circ \mathbf{n_2}), && (2)\\ \beta &= \mathbf{v} \circ \mathbf{n_2}, && (3)\\ \gamma &= \mathbf{u} \circ \frac{\mathbf{p_2} - \mathbf{p_1}}{\|\mathbf{p_2} - \mathbf{p_1}\|}, && (4)\\ \delta &= \|\mathbf{p_2} - \mathbf{p_1}\|. && (5) \end{aligned}$$

SPRH features are given by quantizing these parameters and binning into a histogram. Since the histogram is computed over surflet pairs on the whole object, it includes information on global geometry as well as local geometry.

Histogram generated through sampling and binning process is susceptible to the effects of small difference of surfaces, measurement error, and defect of surface data. To absorb them, kernel density estimation is applied with binning the relation of surflets.

Another improvement for reducing the effect of random sampling, poisson sampling is applied to surflet extraction. Important parameters are R and r ($R < r$) ; the former indicates a radius of spherical range for evaluating local shape, and the latter is a distance between two samples. In our implementation, the r is gradually decreased until the number of samples reaches a predefined number.

3.4 An approach for image-based categorization

As same attitude as using 3D shape data, both rough overall shape and local shape are considered. First, an input image is divided into aligned rectangular regions. With each region, three features, that is shape feature, appearance feature, and color feature, are calculated. The feature vectors with normalized by means of L1 norm are combined, and are used for discrimination by SVM (Support Vector Machine).

The point of the method we applied is a preprocessing for extracting shape information. The processing composes of three steps; (i) edge extraction, (ii)

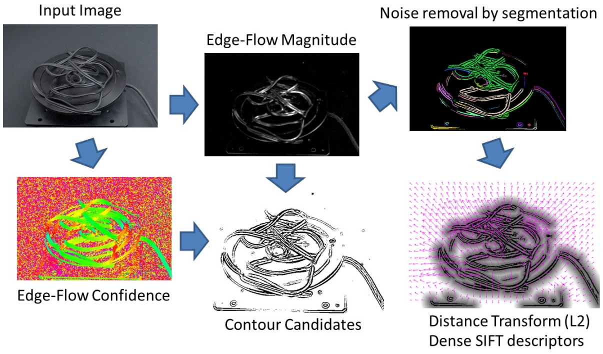

Figure 4. Image processing for feature description

noise removal, and (iii) calculation of Distance Transform [8]. It has an effect in description of roughness of object shape and texture on the object.

The discrimination process is differ from 3D shape-based categorization. Images that capture an object are used for similarity calculation individually, whereas shape-based approach uses a shape model that combines several measurement data.

3.5 Feature description from multi-view images

Figure 4 shows the processing flow of the pre-processing (i) to (iii) described above. After that, feature description is performed based on SIFT feature that is calculated about ordinary-divided lattice points in an image.

The first step of the pre-processing is a process that calculates edge flow. Eigenvalues are calculated from structural tensor in the same manner as Harris corner detector. The tensor with gaussian weight function is represented as follows:

$$M_\sigma(x,y) = G_\sigma(x,y) * (\nabla I)(\nabla I)^T, \tag{6}$$

where I denotes image, G_σ denotes gaussian function, and σ denotes variance. Using eigenvalues $\lambda_1, \lambda_2 (\lambda_1 \geq \lambda_2)$ and corresponding eigen vector $\mathbf{u}_1, \mathbf{u}_2$ calculated from M_σ, we can get edge direction as $\mathbf{u}_1$ and edge magnitude as λ_1. Reliability of the edge is calculated by $E = (\lambda_1 - \lambda_2)/(\lambda_1 + \lambda_2)$.

Contour information of an object is extracted by edge magnitude with relative ease. However, it is not the case with objects the color of which is similar to that of background. On the other hand, edge flow provides weak edges though it is sensitive to noise. From these properties, we extract edges on ridge line by selecting regions where both consistency of edge flow and edge magnitude become high.

The consistency of edge flow is evaluated by averaging of edge direction $\{\mathbf{u}_i\}_{i=1}^N$ in local region. That is,

$$val = \left\| \frac{1}{N} \sum_i \mathbf{u}_i \right\|. \tag{7}$$

The next step is to extract a group of points that are on edges with high continuity. For that purpose, segmentation process is performed based on three criteria; the distance between pixels, the consistency of edge direction, and magnitude. Minimum Spanning Tree (MST) is applied to the segmentation process. Distance Transform is calculated from the segmentation results, and we can get the information of object contour and shape patterns.

In addition to the edge-based description, appearance and color feature are used. The appearance feature is based on SIFT descriptor. SIFT is calculated by grid sampling, and bag of features are applied to generate frequency histogram. Meanwhile, Opponent Color [9] is used to represent color information. To overcome illumination changes and shooting angle difference, kernel density estimation(KDE) is applied. A feature vector is a frequency histogram that is of color space divided into ordinary cubes.

3.6 Integration of multi-view information

In the case of image-based categorization, a final discrimination result is obtained by integrating discrimination results at each view.

Platt [10] proposed a method for deciding parameters that maximize the likelihood of a posterior probability with sigmoid function. In our case, we use the result of SVM on behalf of the posterior probability about each view. That is,

$$p(C|V) = \frac{p(C)}{p(V)} p(V|C) = \frac{p(C)}{p(V)} p(v_1, \ldots, v_n|C), \tag{8}$$

where $V = \{v_1, \ldots, v_n\}$ is a set of views. Assuming that each of views is independently generated from one model, right side of the equation (8) can be written as follows:

$$\frac{p(C)}{p(V)} \prod_i p(v_i|C) = \frac{\prod_i p(v_i)}{p(C)^{n-1} p(V)} \prod_i p(C|v_i), \tag{9}$$

where $p(C)$ is a prior probability. Assuming that the $p(C)$ is constant, maximizing $p(C|V)$ is equal to maximizing the following equation:

$$\sum_i ln\{p(C|v_i)\}, \tag{10}$$

where $p(C|v_i)$ is a probability calculated at ith view. However, independency assumed above is not typically true. As other concern, an input view is far from any view included in training data. From these reasons, we use following equation that replace ln to identity function.

$$\sum_i p(C|v_i) \tag{11}$$

4 Experiments

First, we prepared a training dataset. About 10 objects in each categories were selected, their 3D shapes and multi-view images were captured. Feature vectors were calculated from them in advance.

N-fold cross validation is applied to evaluate categorization accuracy. Figure 5 – 7 shows categorization results. The horizontal axis indicates the number of categories (No.1, 11 and 32 were unused number). 1.0 in the vertical axis means that the categorization succeeded with 100 % accuracy.

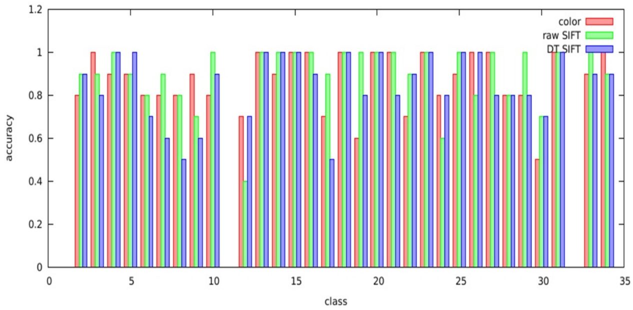

Figure 6. Category classification by means of image-based feature

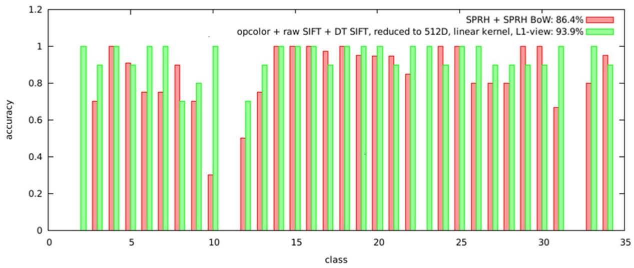

Figure 7. Category classification by means of two proposed methods

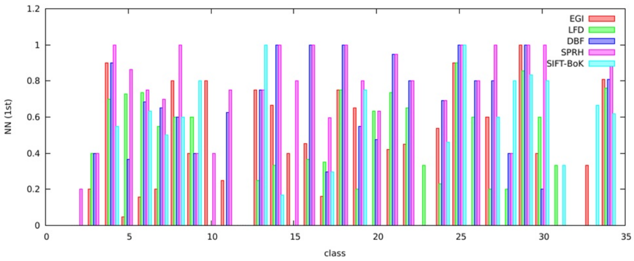

Figure 5. Category classification by means of 3D shape-based feature

Figure. 5 is a result about 3D shape model. Several existing feature descriptions such as EGI, LFD[11], DBF, SPRH, and SIFT-BoK [12] were also examined. SPRH shows high accuracy against No.17 (seal material), No.19 (power strip), No.24 (wiring cable) and No.27 (hose) that were flexible objects.

Figure 6 shows results of the image-based categorization. One image dataset per one object includes 12 images captured at different views. This graph shows the results of three features; Color feature (color), bag of features generated from SIFT descriptors with grid sampling (raw SIFT), and the proposed method (DT SIFT). Compared with shape-based categorization, they indicated high accuracy result.

Figure 7 is a graph that shows two methods proposed in this paper. 3D shape-based feature that combines common SPRH and improved SPRH by using bag of features representation shows 86% success rate. On the other hand, image-based categorization was performed by combining three image features that were individually indicated at Figure 6. The success rate was especially went up for small objects; No.2 (screws with various shapes), No.21 (metal frame), No.22 (metal component). Note that most of No.12(cables) and No.19 (power strip) objects whose color was difficult to distinguish from background could be classified accurately. The conclusive success rate was 94%.

5 Conclusions

This paper described the categorization of objects that were incorporated into home electric appliances. We used two types of measurement data; 3D shape model and multi-view images. Feature descriptions considering both rough overall shapes and local shape were proposed, and experiments showed high accuracy rate, especially for using multi-view images.

For future work, we extend our feature description for the case that only a part of object can be measured and several objects are measured with one sensing.

References

[1] Berthold K. P. Horn. Extended gaussian images. *Proc. of the IEEE*, Vol. 72, No. 2, pp. 1671–1686, 1984.

[2] Panagiotis Papadakis, Ioannis Pratikakis, Stavros Perantonis, and Theoharis Theoharis. Efficient 3d shape matching and retrieval using a concrete radialized spherical projection representation. *Pattern Recogn.*, Vol. 40, pp. 2437–2452, September 2007.

[3] Hamid Laga, Hiroki Takahashi, and Masayuki Nakajima. Spherical wavelet descriptors for content-based 3d model retrieval. *Int'l Conf. on Shape Modeling and Applications*, pp. 15 – 25, 2006.

[4] Ceyhun Burak Akgul, Bulent Sankur, Yucel Yemez, and Francis Schmitt. 3d model retrieval using probability density-based shape descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, pp. 1117–1133, 2009.

[5] D. V. Vranic et al. Tools for 3d-object retrieval: Karhunen-loeve transform and spherical harmonics. In *IEEE MMSP 2001*, pp. 293–298, 2001.

[6] Philip Shilane, Patrick Min, Michael Kazhdan, and Thomas Funkhouser. The princeton shape benchmark. In *Shape Modeling International*, June 2004.

[7] Eric Wahl, Ulrich Hillenbrand, and Gerd Hirzinger. Surflet-pair-relation histograms: A statistical 3d-shape representation for rapid classification. *3D Digital Imaging and Modeling, Int'l Conf. on*, Vol. 0, p. 474, 2003.

[8] P. F. Felzenszwalb and D. P. Huttenlocher. Distance transforms of sampled functions. In *Technical report, Cornell Computing and Information Science*, 9 2004.

[9] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Trans. on PAMI*, Vol. 32, No. 9, pp. 1582–1596, 2010.

[10] J C Platt. Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. B Sch 旦 lkopf C J C Burges A J Smola editors Advances in Kernel Methods Support Vector Learning, pp. 61–74, 2000.

[11] D. Y. CHEN. On visual similarity based 3d model retrieval. *Proc. Eurographics, 2003*, 2003.

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