A LOCAL DESCRIPTOR FOR HIGH-SPEED AND HIGH-PERFORMANCE PICTOGRAM MATCHING

Terumasa Aoki[1,2] , Kurumi Kaminishi[1]

Tohoku University, Japan
[1]Graduate School of Information Sciences (GSIS)
[2]New Industry Creation Hatchery Center (NICHe)

ABSTRACT

Pictogram, a simple picture-based symbol, is widely used to indicate important facilities (such as "rest room") or important rules (such as "no smoking"). If a practicable pictogram matching technology is available, it must be useful for many applications. Unfortunately, as far as the authors know, there are no practicable methods which satisfy both high-speed (real-time) processing and high performance (high matching ratio). In this paper, we present a novel local descriptor for high-speed and high-performance pictogram matching under a variety of photographing conditions. The proposed method consists of three modules: projection-invariant values called CRN (Cross Ratio Number), shape description scheme for pictograms, and matching acceleration method based on feature vector classification by using the relationship between contour and convex hull. The experiments show our method outperforms the state-of-the-art methods in terms of matching ratio and computation cost.

***Index Terms*—** image matching, pictogram, shape descriptor, local feature

1. INTRODUCTION

Image matching, which is a technique to find the correlations between two images, is one of the fundamental technologies in computer vision [1]. There are wide applications of image matching e.g. image retrieval [2], object recognition [3], image classification [4] and 3D object creation [5], etc. Under this situation, a lot of researches on image matching have been done for the past two decades, and some methods such as SIFT [3] and HOG [7] etc. are widely used in many applications. Most of these methods show good performances for natural images, but they are not suited for simple images such as illustrations, cartoons and pictograms etc. When we use them for these types of simple images, matching performance drastically deteriorates due to fewer characteristics of the images.

Pictogram, which is the target of our research, is defined as "pictorial symbol" and indicates a simple picture-like sign. Most pictograms are designed by monotonous colors and simple figures. The characteristics of pictograms make it easy for humans to understand the meaning of a sign at a glance, but, in contrast, they make it hard for computers to distinguish them. If we obtain a practicable pictogram matching method, many applications can use it to recognize the meanings of traffic signs, computer logos, picture signs on the signboards and others (Fig.1), but there are no practicable methods for pictogram matching as far as we know.

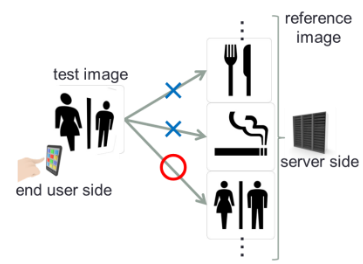

Fig.1. pictogram matching

In recent years, shape descriptor attracts rising attention to match simple images. Shape descriptor is a descriptor to characterize the shape of an object, and it is divided into two categories: region-based [7,8,9,10] and contour-based [11,12,13,14,15]. Contour-based descriptors describe features extracted from contours themselves and/or the surroundings of contours, and they usually ignore inner parts of contours. [14][15] proposed shape descriptors which are robust against articulated transformation, noise and projective transformation. However, they are not suited for pictograms matching because pictograms are often composed of multiple objects (such as "rest room"). Also, some pictograms have the same outer regions although inner regions are different from them (such as "traffic sign").

On the other hand, region-based descriptors use global information extracted from the whole region of an image. Since this type of descriptors meet our requirement (pictogram matching), we focus on them. Representative techniques in this category are Characteristic Number (CN)[9] and Cross Ratio Spectrum (CRS)[10]. These two methods compute one feature vector from a pictogram as a global descriptor. These methods aim to generate robust descriptors against projective transformation, but they are not robust against occlusion. Also, they are time consuming. That's why we cannot use them on real-time applications.

In this paper we present a region-based local shape descriptor, which achieves high-speed and high-performance pictogram matching under various photographing conditions (especially, projective transformation and occlusion as well as scale, luminance, and rotation changes) to establish a practicable pictogram matching technique.

The rest of this paper is as follows. In section 2, we present a novel method for pictogram matching. The proposed method computes projection-invariant values from whole region of a pictogram, describes them as local feature vectors and classifies local descriptors by using the relationship between contour and convex hull for fast matching. In section 3, we show some experimental results. The proposed method is compared to some state-of-the art methods (CN and CRS). Finally we conclude this paper in section 4.

2. A LOCAL SHAPE DESCRIPTOR ROBUST TO PHOTOGRAPH CONDITIONS

The proposed method consists of three modules: projection-invariant values called CRN (Cross Ratio Number), local shape description scheme for pictograms, and matching acceleration method based on feature classification by the relationship between contour and convex hull. We show the framework of proposed method Fig.2.

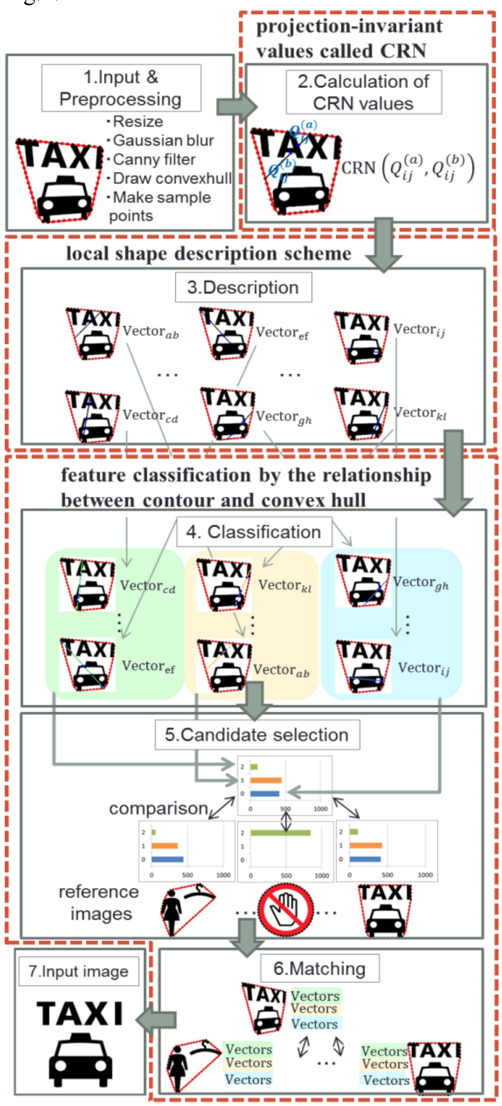

Fig.2 Framework of proposed method

2.1. Projection-invariant values: CRN

We define CRN (Cross Ratio Number), which is based on the cross ratio. CRN is used to compute projection-invariant values at step2 in Fig.2. The CRN algorithm is as follows.

1. Put n sample points P_i on the convex hull of a pictogram equidistantly. Here, n can be given by user.
2. Select two points P_i , P_j from n sample points and draw a line segment.
3. Detect the intersections $Q_{ij}^{(a)}, Q_{ij}^{(b)}$ of the line segment and the inner structure.
4. Compute the CRN value from four points P_i , P_j, $Q_{ij}^{(a)}$, $Q_{ij}^{(b)}$ in by the following equation (1).

$$\mathrm{CRN}\left(Q_{ij}^{(a)}, Q_{ij}^{(b)}\right) = \frac{\mathrm{P}_i Q_{ij}^{(a)}}{Q_{ij}^{(a)} Q_{ij}^{(b)}} \times \frac{Q_{ij}^{(b)} \mathrm{P}_j}{\mathrm{P}_i \mathrm{P}_j} \tag{1}$$

   $\mathrm{CRN}\left(Q_{ij}^{(a)}, Q_{ij}^{(b)}\right)$ denotes the CRN value from four points $P_i, P_j, Q_{ij}^{(a)}, Q_{ij}^{(b)}$ and points $Q_{ij}^{(a)}, Q_{ij}^{(b)}$ denote a-th and b-th $(a \neq b)$ intersections of segments P_iP_j and inner structure respectively (Fig.3)
5. Repeat 3 to 4 for all pairs of intersections and compute ${}_{\alpha}C_2$ CRN values. Here, α is the number of intersections on the line segment.
6. Repeat 2 to 4 for all line segments drawn by selecting all pairs of two points from n sample points.

As is clear in the above algorithm, CRN is projection-invariant and it contains much inner information.

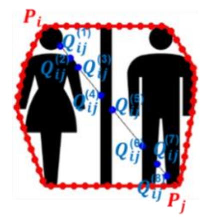

Fig.3. Example how to compute CRN

2.2. A local shape description scheme for pictograms

As described in section 1, existing region-based methods (CN and CRS) generate global descriptors. Namely, one global feature vector is generated by combining all CN (or CRS) values computed from one pictogram. However, global descriptors are not robust against occlusion in nature. Another problem of these methods is difficulty in fixing the position of sample points. When the pictogram rotates from the original position or the pictogram is deformed by projection transformation, the position of each sample point will change. Also, the starting point in sample points, which is used in CN and CRS, will change the position too. CN and CRS need to align the starting points between a test image and a reference image, but this process is time consuming.

To solve these problems, we form a local descriptor by using CRN values instead of a global descriptor. By forming a local descriptor, our method is more robust than the above two methods against occlusion, and alleviates the bad effects of changing the positions of sample points and a starting point. Furthermore, unnecessity to align the starting points and lower calculation cost are the other advantages of local feature vectors.

The proposed description scheme is as follows.

- One feature vector is generated from each line segment.
- A feature vector is M dimensional. Here, M can be given by user.
- Each element of a feature vector is a CRN value, which is computed from two sample points and two intersections of a line segment and inner structure as described in 2.1.
- A feature vector V_{ij} of the segment P_iP_j is defined by the following equation (2).

$$\mathrm{V}_{ij} = \left(\mathrm{CRN}\left(Q_{ij}^{(1)}, Q_{ij}^{(2)}\right), \dots, \mathrm{CRN}\left(Q_{ij}^{(\alpha-1)}, Q_{ij}^{(\alpha)}\right), 0, \dots\right)_M \tag{2}$$

When the number of intersections of a line segment and inner structure is α, the CRN values are lined up as follows: 1st, 2nd, 3rd, ..., α-th intersections, after that, 2nd, 3rd, 4th, ..., α-th intersections, finally $(\alpha$-1)th and α-th intersections. The maximum number of the pairs of intersections is M, and if the number of elements is fewer than M, 0 are put in the remained elements.

As a result, we obtain ${}_nC_2$ vectors described by equation (2) from one pictogram by using all line segments which are drawn by all pair points of all sample points on the convex hull. The set of these vectors is defined as a local shape descriptor in our scheme.

2.3. Matching acceleration method based on feature classification by using the relationship between contour and convex hull

As shown in Fig.4, some sample points are on the contour (black points) and the other sample points are not on the contour (red points). So, for local feature vectors generated by the scheme in 2.2, we can classify them into three groups by the positions of end points (sample points) of each line segment at step4 in Fig.2.

Contour group0: Both two sample points are on the contour.
Contour group1: Only one sample point is on the contour.
Contour group2: None of two sample points are on the contour.

The advantages of this classification are as follows.

- Rates of contour group0, 1 and 2 are the important information for each pictogram. This information can enhance a discriminative power in matching process.
- If this rate of a test image is largely different from that of a reference image, they will not match obviously. We can narrow down the candidates of matching pairs by using this information.
- We do not have to match two feature vectors which belong to different groups because they will not match clearly. We can reduce the calculation cost for matching by using this information.

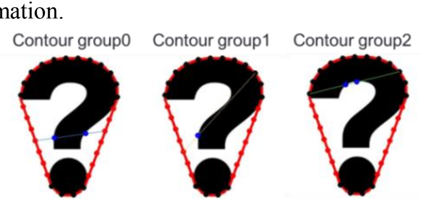

Fig.4 Contour groups
(Color of sample points,
black: on the convex hull, red: not on the convex hull)

In fact, the number of feature vectors in each group largely varies from the content of pictograms (Fig.5). So, we select the matching candidates in preprocessing of feature matching at step5 in Fig.2. We first compute the similarity between the contour group histograms of an input image and reference images by histogram intersection. After that, the reference images, the similarity of which is higher than the threshold, are selected as the matching candidates.

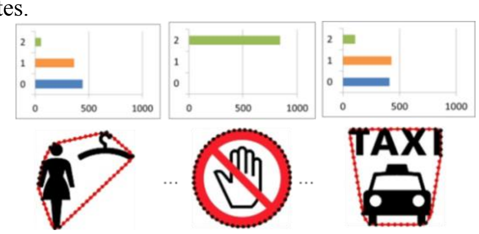

Fig. 5 contour group histograms

After classification and candidate selection, we match local feature vectors in an input image and matching candidates in each contour group respectively at setp6 in Fig.2. Finally, the result of matching between an input image and a reference candidate image is obtained by calculating the sum of the cross-matched feature vectors in all groups (Fig.6).

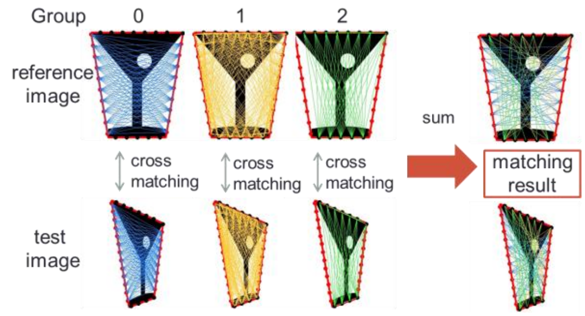

Fig.6 Contour group based matching

3. EXPERIMENTAL RESULTS

We show some experimental results for the proposed method, CN and CRS in this section.

3.1. Experimental conditions

For experiments, we used the pictogram data set published by Public Information Symbols [16]. We used all pictograms (125 pictograms) of this data set. To evaluate the proposed method on various photographing conditions, we generated the following two types of test images (Fig.7).

- 1125 (125×9) pictograms with projective transformations: We generated 9 test images from each original pictogram. Test images are deformed by projective transformation (θ, φ). Here, θ, φ are rotation about the vertical axis and about the horizontal axis, and values of them are 20, 40 and 60 degrees respectively.
- 375 (125×3) pictograms with occlusions: We generated 3 test images from each original pictogram. Three sizes of obstacles (1/8, 1/16 and 1/32) are put on each original pictogram randomly.

All of images used in the experiments are JPEG images, and resolution of them is 300×300 pixel. And all experiments were done by the following computer.

- CPU : Intel(R) Core(TM) i3-3240 CPU 3.40GHz
- Memory (RAM) : 12.0GB

Also, the following parameters are used.

- the number of sample points on the convex hull is n=60
- the dimension is M=28
- threshold of histogram similarity is 0.80.

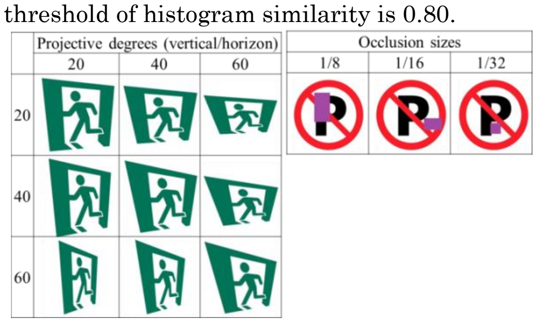

Fig.7 Examples of test images

In the experiments, we compared the following 6 methods. We can evaluate the performances of the proposed method for CN and CRS by comparing method 1 to 3. Also, we can evaluate each module (CRN values, local description scheme, and feature vector classification) by comparing method 1 and method 4 to 6.

method1) the proposed method
method2) CN
method3) CRS
method4) method 1, but CN values are used instead of CRN values.
method5) method 1, but global descriptors (like CN and CRS) are used instead of local descriptors. (Note that feature vector classification described in 2.3 can not be used inevitably in this method.)
method6) method 1, but feature vector classification is not used.

Performances of these 6 methods are compared in terms of matching ratio and computation cost.

For matching ratio, the system computes the similarities between a test image and all reference images, and selects the most similar one as the matching result. In local descriptor based methods (method 1, 4, and 6), similarity is computed by the number of cross-matched feature vectors between a test image and a reference image. Cross-matched feature vectors are a pair of points which are the closest point for each other on the multidimensional vector space. In global descriptor based method (method 2 and 5), similarity is computed by histogram intersection. And DTW (Dynamic Time Warping) is used in method 3.

On the other hand, Computation cost is divided into two: generation cost of descriptors per one image (description time) and matching cost between one test image and one reference image (matching time). We evaluate both of them because they are both important for real-time applications.

3.2. Evaluation of matching ratio

Table 1 and Table 2 show the experimental results of matching ratio for projective transformation and occlusion respectively.

Table 1 Matching ratio of the pictograms with projective transformation [%]

method	vertical/ horizon	20°	40°	60°	all images
method1 (proposed method)	20°	**100**	**97.6**	**94.4**	**96.6**
	40°	**100**	**98.4**	**96.8**	
	60°	**91.2**	**94.4**	**96.8**	
method2 (CN)	20°	79.2	51.2	30.4	43.8
	40°	54.4	44.8	30.4	
	60°	42.4	36.0	25.6	
method3 (CRS)	20°	76.0	47.2	28.0	46.8
	40°	62.4	58.4	37.6	
	60°	38.4	31.2	41.6	
method4 (not CRN)	20°	96.0	75.2	59.2	70.9
	40°	80.8	72.8	62.4	
	60°	66.4	64.0	61.6	
method5 (not local)	20°	93.6	70.4	40.0	50.0
	40°	68.0	48.0	37.6	
	60°	33.6	31.2	27.2	
method6 (not classification)	20°	96.8	88.8	81.6	87.6
	40°	92.0	89.6	84.0	
	60°	78.4	86.4	90.4	

Table 2 Matching ratio of the pictograms with occlusion [%]

methods	occlusion size			all images
	1/8	1/16	1/32	
method1 (proposed method)	**97.6**	**100**	**100**	**99.2**
method2 (CN)	91.2	94.4	96.0	93.9
method3 (CRS)	77.6	86.4	86.4	83.5
method5 (not local)	97.6	97.6	99.2	98.1
method6 (not classification)	90.4	96.8	98.4	95.2

As is clear in Table 1 and 2, we can see that matching ratio of method 1 (the proposed method) is much higher than the existing methods (method 2 and 3). These results prove robustness of the proposed method for projective transformation and occlusion. Furthermore, we can also see that 3 modules described in 2.1 to 2.3 are all useful for performance improvement respectively because matching ratio of method 1 is higher than method 4 to 6.

3.3. Evaluation of computation cost

Table 3 shows the experimental result of computational cost.

Table 3 Average description time and comparison time

methods	description time [ms]	matching time[ms]
method1 (proposed method)	**299.3**	**28.5**
method2 (CN)	14902	278.7
method3 (CRS)	173.6	1031

From this result, we find that description time in method 2 (CN) and matching time in method 3 (CRS) are very huge. In fact, CN inherently includes redundant calculations for CN values, and DTW, which is used in CRS, is a well-known time consuming process. As a result, we can see that the proposed method is the best method regarding computation cost.

4. CONCLUSIONS

In this paper, we presented a novel local descriptor for high-speed and high-performance pictogram matching. The proposed method consists of three modules: projection-invariant values called CRN, shape description scheme for pictograms, and matching acceleration method based on feature vector classification by using the relationship between contour and convex hull. Experiment results show the effectiveness of proposed three modules. And the proposed method shows much higher matching ratio than the existing methods (96.6% for projective transformations and 99.2% for occlusion). In addition, the computation time of the proposed method is much faster than existing methods. These facts prove that the proposed method outperforms the existing methods such as CN and CRS.

5. REFERENCES

[1]YinanYu, Kaiqi Huang and Wei Chen," A Novel Algorithm for View and Illumination Invariant Image Matching",IEEE Transaction on Image Processing, vol.21, issue.1, pp.229-240, 2012.

[2] J. Sivic, A. Zisserman "Video Google: a text retrieval approach to object matching in videos", Ninth IEEE International Conference on Computer Vision, vol.2, pp.1470-1477, 2003.

[3] D. G. Lowe, "Object recognition from local scale invariant features", The Proceedings of the Seventh IEEE International Conference on Computer Vision, Sep, 1999.

[4] Donghua Xu, Zhiyi Qu "An Image Classification Method Based on Matching Similarity and TF-IDF Value of Region", 2013 Sixth International Symposium on Computational Intelligence and Design (ISCID), vol.2, pp.112-115, 2013.

[5] J. W. H. Tangelder, R. C. Veltkamp "A survey of content based 3D shape retrieval methods", Proceeings of Shape Modeling Applications, pp.145-156, 2004.

[6] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection", 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) , vol.1, pp.886-893, 2005.

[7] R.B. Yadav, N.K. Nishchal, A.K. Gupta, and V.K. Rastogi, "Retrieval and classification of objects using generic fourier, legendre moment, and wavelet zernike moment descriptors and recognition using joint transform correlator," Optics &Laser Technology, vol. 40, no. 3, pp. 517-527, 2008.

[8] F. Ghorbel, S. Derrode, R. Mezhoud, T. Bannour, S. Dhahbi, "Image reconstruction from a complete set of similarity invariants extracted from complex moments", Pattern recognition letters, vol. 27, no. 12, pp. 1361-1369, 2006.

[9] Zhongxuan Luo, Daiyun Luo, Xin Fan, Xinchen Zhou, and Qi Jia, "A shape descriptor based on new projective invariants", IEEE International Conference on Image Processing (ICIP), 2013

[10] Linlin Li and Chew Lim Tan," Recognizing Planar Symbols with Severe Perspective Deformation",IEEE Transaction on Pattern Analysis And Machine Intelligence ,vol.32, issue.4, 2010 ,pp755-762,2010

[11] Miroslaw Bober, "MPEG-7 visual shape descriptors", IEEE Transactions on Circuits and Systems for Video Technology, pp.716-719, 2001.

[12] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pp. 509-522, 2002.

[13] Xiang Bai, Cong Rao, Xinggang Wang, "Shape Vocabulary: A Robust and Efficient Shape Representation for Shape Matching", IEEE Transactions on Image Processing, val.23, issue.9, pp3935-3949, 2014.

[14] Haoran Xu, Jianyu Yang, Junsong Yuan, "Invariant multi-scale shape descriptor for object matching and recognition", 2016 IEEE International Conference on Image Processing (ICIP), pp644-648, 2016

[15] Qi Jia, Xin Fan, Yu Liu, Haojie Li, Zhongxuan Luo, He Guo, "Hierarchical projective invariant contexts for shape recognition", Pattern Recognition, vol.52, pp358-374, 2016

[16] "Public Information Symbols", Foundation for Promoting Personal Mobility and Ecological Transportation, http://www.ecomo.or.jp/english/picto_top.html