

# Signboard Recognition by Consistency Checking of Local Features

Jihoon Kim, Taik Heon Rhee, Kee-Eung Kim, and Jin Hyung Kim

Computer Science Department, KAIST  
373-1 Guseong-dong, Yuseong-gu, Daejeon, Republic of Korea  
E-mail: {jihoon, three, kekim, jkim }@ai.kaist.ac.kr

## Abstract

The problem of recognizing signboards in street scenes is defined as matching the input image to pre-stored 2D signboard images. This problem is not as simple as it appears to be due to arbitrary drawings and relative 3D positions. We approached this problem by matching characteristic local features of input image to those of images in the database. Local decisions are verified by the global viewpoint of the homographic consistency and color consistency. The well-known SIFT feature is used as a local feature and the homographic consistency checking is performed using RANSAC, a random sampling method. In order to handle highly perspective-distorted signboards, several perspective-transformed templates are generated offline. In our experiment, with a database of 35 images, our proposed method achieved 95% recognition rate, showing good results despite the highly distorted input images.

**Keywords** Signboard recognition, Homography, Perspective Transform, SIFT, RANSAC

## 1. Introduction

Since signboards are distinct and contain rich information (무엇의?), identification of signboards yields the context of the scene image such as location. We define **the problem of recognizing signboard in street scenes as the matching problem of input image into one of pre-stored 2D signboard images**. Unlike traffic sign or car license plate recognition, model-based top-down approaches are not applicable to this problem because signboards have irregular shapes and colors (Fig.1). In other words, detection of regions of signboard is impossible using prior knowledge on the shape and color of the signboards. This may be the reason why we cannot find cases of general signboard recognition in literatures. However, there was an attempt to recognize signboards of rectangular shapes [7], which is much more restrictive

than our work. This paper focuses on signboard recognition of unrestricted forms.

This paper proposes a bottom-up approach that extracts local features and verifies global consistency. SIFT(Scale-Invariant Feature Transform) [5] is used as the local features. First, SIFT features in query images are matched with pre-extracted features in template images. The matching results of local features contain errors that result in false matches. Such false matches are filtered by consistency checking using the RANSAC(RANDOM SAMple Consensus) [2][3] method. As a result of running RANSAC, we can obtain a homography, which is essentially a transformation matrix between two images. We improved RANSAC by inspecting the validity of the resulting homography. For further verification, we compared the colors of each pixel in the template image to those of corresponding pixels in the query image. Also, for

robust recognition of perspective-distorted signboards, we generated several perspective-distorted templates by adjusting parameters of the perspective transform.



Figure 1. Examples of various signboards

The outline of this paper is as follows. Section 2 provides a system overview of our proposed method. In section 3, SIFT as local feature is explained briefly. Section 4 presents methods for verifying global consistency of local feature matching. In section 5, method for generating perspective-transformed templates is described. Section 6 shows experimental results and example images, followed by the conclusion of the paper in section 7.

## 2. System Overview

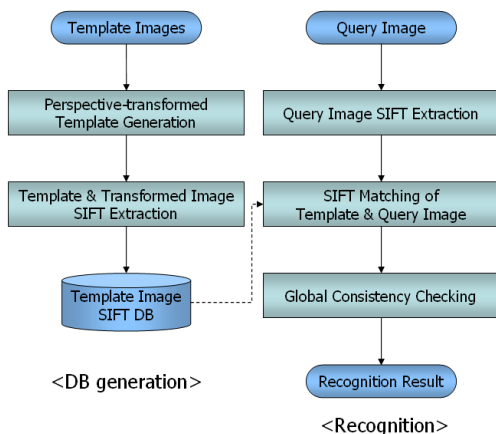


Figure 2. System overview



Figure 3. Template image, query image, transformed templates images, and matching result

Fig.2 shows the overview of the system. Signboard recognition system is composed of two processes: First one is the database (DB) generation process and second one is the recognition process. In the DB generation process, original templates (Fig.3(a)), which are directly front-facing, go through several different transformations (Fig.3(c)) for robust recognition. SIFT features are extracted from the template images and their respective transformed images before being stored.

Fig.3(b) shows a query image that has a perspective distortion. In the recognition process, SIFT features are extracted from the query image and matched to those of the templates. Fig.3(d) shows an example of a matching result Depicted by solid lines that connect the matching points. Once the template matching is done, the class having maximum number of matching points is chosen as the recognition result. Also, the signboard region is detected through estimated transform from the template image to the query image. If the maximum number of matching points is too small, the case is rejected. In such cases where reliability is low, rejection is preferred because the risk of rejection is much smaller than misrecognition.

## 3. Local Feature Extraction

SIFT(Scale-Invariant Feature Transform) [5] is used as local features. SIFT extracts stable key-points from the

images and calculates the geometrical descriptor for each key-point. SIFT is invariant to translation, scaling and rotation and partially invariant to illumination and 3-D viewpoint change. It is proven to be the most robust local feature [6]. Fig.3(d) and Fig.4(b) show examples of SIFT feature matching. Even though slight 3D viewpoint change exists between the two images, images are matched properly due to the distinctiveness and invariance of SIFT feature.

Unfortunately, the matching result of local features contains false matches. Thus, the next stage performs consistency checking of the matching points to reject false matches.

## 4. Global consistency verification

### 4.1 Finding adequate homography

In the situation where samples contain inliers and outliers, RANSAC [2][3] can reject invalid samples and estimate parameters of the model. Even though RANSAC cannot find an optimal solution, it can find an approximate solution efficiently. In RANSAC, samples are chosen randomly and a model consisting of such samples is constructed. After that, a set of samples explained by the model is calculated. These steps are repeated. Finally, the set of data having the maximum number of elements is decided and the final model parameters are estimated with the elements.

RANSAC can also be utilized for global consistency checking of local feature matching result. During the process, estimated model is homography.(?) When planar objects are imaged, the relation between the original object and the image can be expressed as a linear perspective transform. The transform matrix is called a homography [3]. A homography can be expressed as follows.

$$\mathbf{X}' = \mathbf{H}\mathbf{X}$$

$$\begin{bmatrix} tx_1' \\ tx_2' \\ t \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}$$

$\mathbf{H}$  is a homography, where  $\mathbf{X}$  and  $\mathbf{X}'$  are  $3 * 1$  vectors and correspond to the same point. We need 8 parameters because  $h_{33}$  equals to 1. Thus, 4 pairs of correspondence point are required for solving  $\mathbf{H}$ . In the same context, the 4-point correspondences are randomly chosen by RANSAC.

When random correspondence points are chosen and a homography is calculated by RANSAC, the calculated homography may not be valid. Rejecting invalid homographies improves RANSAC efficiently. We introduce two methods (for what?). First is homography decomposition and the other is heuristic rules. To provide meanings to the parameters of a homography, the homography is decomposed as such[3]:

$$\mathbf{H} = \mathbf{H}_s \mathbf{H}_A \mathbf{H}_p = \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{K} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{v}^T & v \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^T & v \end{bmatrix}$$

If a homography decomposition is impossible, random samples composing the homography are rejected and new random samples are chosen.

Even though homography decomposition is successful, it is probable that the values of the parameters are impossible in practical situations. Thus, we can set valid ranges of each parameter.

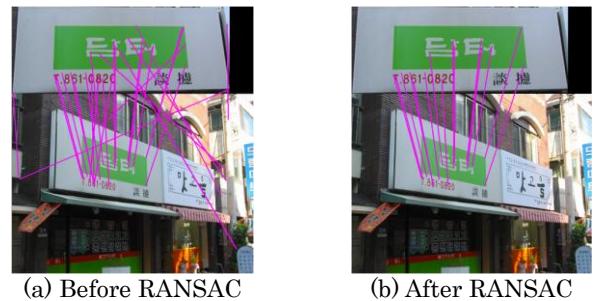


Figure 4. Effect of RANSAC

Fig.4(a) shows a matching result of SIFT features between a template image and a query image. Fig.4(b) shows the result after the false matches were removed by applying RANSAC.

## 4.2 Color verification

There can still be false results after RANSAC. Thus, a verification step using color is added to overcome SIFT features that are extracted from grayscale image. From the homography, corresponding points between the template and the query image are calculated. Color values of all pixels in the template and corresponding pixels in the query are compared. To reduce the illumination effect, hue values composed of several bins are used. If the difference between template and query image is more than some threshold, the matching is rejected. The difference metric is given as :

$$\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N |Hue(\mathbf{H}\mathbf{x}_{ij}) - Hue(\mathbf{x}_{ij})|$$

where the size of the template image is  $M \times N$  and  $\mathbf{x}_{ij}$  means the  $(i, j)$  pixel in template image.  $\mathbf{H}$  is a homography matrix from the template image to the query image. The  $Hue()$  function converts an RGB value of a pixel to a hue value.

Color verification cannot improve recognition rate but it has an effect of reducing false recognition by rejecting the cases where color information between the template and the query is drastically different.





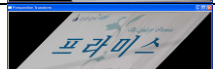






## 5. Perspectively-transformed template generation

For robust recognition in the case of excessive perspective distortion, several perspective-transformed shapes of templates are generated and used as additional templates. Even though the template image and the query image cannot be matched, the query image can be matched with one of the transformed templates and thus recognized. To parameterize perspective transform, we start from a homography because it means perspective transform(?). The expansion of homography decomposition equation is as follows:

$$\mathbf{H} = \begin{bmatrix} s \cos \theta & -s \sin \theta & t_1 \\ s \sin \theta & s \cos \theta & t_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} k_1 & k_2 & 0 \\ 0 & 1/k_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ v_1 & v_2 & 1 \end{bmatrix}$$

To express a perspective transform, 8 parameters -  $s$ ,  $\theta$ ,  $t_1$ ,  $t_2$ ,  $k_1$ ,  $k_2$ ,  $v_1$ , and  $v_2$  - are required. Fortunately, SIFT is perfectly invariant to translation, rotation and scaling. Thus, the scaling parameter  $s$ , rotation parameter  $\theta$ , and translation parameter  $t_1$ ,  $t_2$  can be ruled out. The characteristics of the remaining 4 parameters are summarized in Table 1.

**Table 1. Characteristics of parameters**

	Value	Shape	Meaning
Original	-		Original
$k_1$	0.5		Aspect ratio
	2		
$k_2$	1		Shear
	-1		
$v_1$	0.001		Horizontal perspective
	0.005		
	-0.001 1		
$v_2$	0.001		Vertical perspective
	0.005		
	-0.001 1		

$k_1$ ,  $k_2$ ,  $v_1$ , and  $v_2$  are the aspect ratio change parameter, the shear transform parameter, the horizontal perspective parameter, and the vertical perspective parameters,

respectively. Values of these four parameters are chosen and the perspective transforms of the template are generated by combining the four parameters. Using such transformed images as additional templates, it is possible to recognize images that are excessively perspective-distorted.

## 6. Experiments

### 6.1 Experiment setup

Signboard images are composed of template images and query images. Template images are gathered without distortion, whereas query images include distortions such as scaling, rotation, perspective, and illumination change.

We have two sets of databases. The first set has high quality images and the second set contains images that are captured frame-by-frame from a digital camcorder recording. Therefore, the second set of images contain noises and motion blurs. Table 2 shows the description of the 2 databases.

**Table 2. Databases**

	DB1	DB2
Template #	35	67
Query #	41	242
Device	CANON G2 digital camera (4 mega pixel)	SONY DCR-TRV 940 digital camcorder (1 mega pixel)
Location	Outdoor	Indoor
Quality	High	Motion blur & noise

Unless recognizing a highly-located signboard, vertical perspective transform rarely happens. Slight perspective can be covered by SIFT. Thus, the  $v_2$  parameter is ignored and the remaining three parameters -  $k_1$ ,  $k_2$  and  $v_1$  - were used to generate additional templates.

### 6.2 Experiment result

We examined the recognition rate change against the number of perspective templates. Parameter setups of each experiment is summarized in Table 3. In experiment 1, query images are matched only with templates without

transform. In experiment 2, 3, and 4, the number of perspective-transformed templates is increased as the number of possible values for the parameters is increased.

**Table 3. Experiment setup of changing additional templates**

	Exp. 1	Exp. 2	Exp. 3	Exp. 4
$k_1$	-	1	1	0.7 1
$k_2$	-	-0.5 0.5	-0.5 0.5	-0.5 0.5
$v_1$	-	0	-0.0005 0 0.001	-0.0005 0 0.001
Additional templates	0	1*2*1=2	1*2*3=6	2*2*3=12

Table 4 shows the result. As the number of perspective transform templates increased, recognition rates increased as well. From this result, we can infer that recognition rate and speed have a tradeoff relationship. Table 5 shows 12 perspective templates used in experiment 4. These shapes cover most practical perspectives. Even though perspective distortions in images are slightly different from these generated shapes, characteristics of SIFT, partially invariant to perspective transform, can overcome the differences.

**Table 4. Experiment result of changing additional templates**

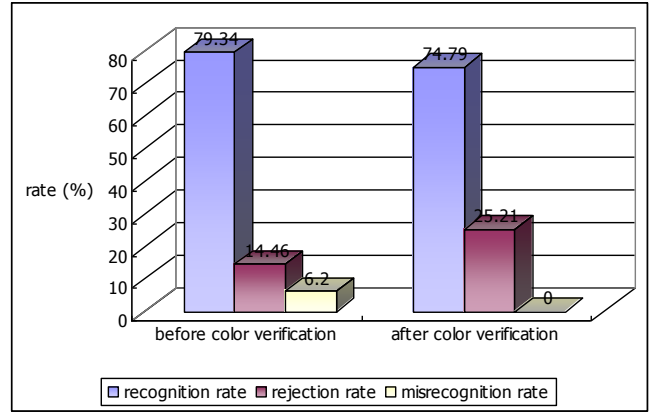
	Rate	Exp. 1	Exp. 2	Exp. 3	Exp. 4
DB1	Recognition rate(%)	39.02	73.17	90.24	95.12
	Rejection rate(%)	60.98	21.95	7.32	0.00
	Misrecognition rate(%)	0.00	4.88	2.44	4.88
	Recog. speed (query/min.)	13.3	9.1	6.1	5.0
DB2	Recognition rate(%)	34.30	66.94	72.73	79.34
	Rejection rate(%)	65.29	30.99	22.73	14.46
	Misrecognition rate(%)	0.41	2.07	4.55	6.20
	Recog. speed (query/min.)	20.7	10.9	6.3	4.4

**Table 5. Perspectively-transformed templates**

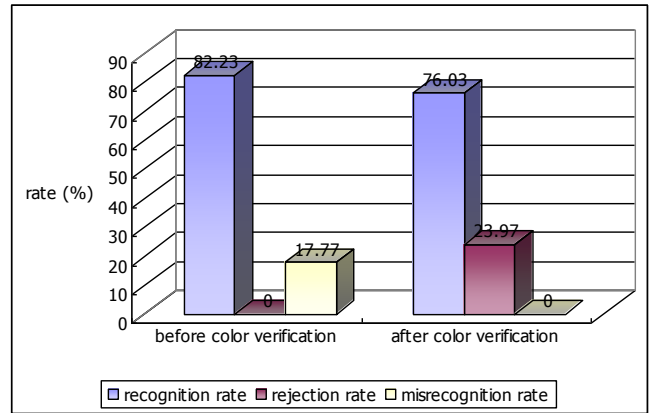
		$v_1$		
		-0.0005	0	0.001
$k_1 = 0.7$	$k_2 = -0.5$			
	$k_2 = 0.5$			
$k_1 = 1.0$	$k_2 = -0.5$			
	$k_2 = 0.5$			

Also, we conducted experiments on recognition rate changes against matching point threshold and the effect of color verification. A minimum of four matching points are required to construct a homography. If the number of matching points is less than four, the results are rejected. If the number is almost equal to four, the cases are also rejected because of the low confidence level. We chose 4, 5, and 6 as the value of matching point thresholds. Moreover we checked color verification effect.

Fig.5 shows the result of the experiment. The smaller the matching point threshold is, the higher the recognition rate and misrecognition rate become. Therefore, to increase recognition rate, matching point threshold should be as small as possible. The increase of misrecognition rate can be overcome by color verification step. After color verification, all misrecognition cases are converted to rejection and some recognition cases are also converted to rejection. However, the number of converted recognition cases is less than that of the converted misrecognition cases. Therefore, color verification method is effective in reducing misrecognition rates. The increase of the rejection rate seems concerning in this case. However, considering practical applications, users do not get false result but are rather requested to send a higher quality query image in case of a rejection. From this, risk of rejection is much less than that of misrecognition.



(a) Matching point threshold: 5



(b) Matching point threshold: 4

**Figure 5. Color verification effect**

**6.3 Result examples**

**Table 6. Recognition examples**

	1	2
Query		
Template		
Matching		
Parameter	$k_1=0.7$ $k_2=-0.5$ $v_1=-0.0005$	$k_1=0.7$ $k_2=0.5$ $v_1=0.001$

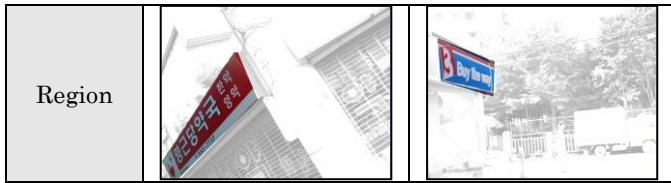


Table 6 shows examples of signboards that are recognized correctly. The template images are perspective-transformed by the parameters. The transformed image is matched against the query image. Then, the global consistency is checked before finally detecting the correct signboard region. Even though signboards in both query images are perspective-distorted, the proposed method can recognize signboards perfectly.

In case 4 (Table 4), two query images cannot be recognized correctly because both of them contain non-planar signboards. Also, in the experiment using DB2, low quality images containing motion blurs and noises are rejected. The proposed method cannot deal with non-planar signboards and low quality images.

## 7. Conclusion

This paper proposed a signboard recognition method which is robust to perspective transform. We took an approach of conducting consistency checking by creating multiple perspective-transformed templates. By using SIFT local feature, the number of types of parameters can be reduced.

The problem of recognizing objects having rich prior knowledge mainly takes a top-down approach. At first, the region at which the object exists is detected. However, we took a top-down approach for signboard recognition because signboards rarely have prior knowledge. Local features of images are extracted and matched before checking the global consistency.

In the experiment using high quality images, the recognition rate was above 90%. The more perspective templates are used, the better recognition rate we can get. However, this method cannot be applied to non-planar, noisy and blurred signboard.

## Acknowledgement

This research is supported by Ubiquitous Computing and Network (UCN) Project, the Ministry of Information and Communication (MIC) 21st Century Frontier R&D Program in Korea.

## References

- [1] A. de la Escalera, J. Armingol, and M. Mata. *Traffic sign recognition and analysis for intelligent vehicles*. Image and Vision Computing, 2003.
- [2] M. Fischler and R. Bolles. *Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography*. Comm ACM, 24(6):381–395, 1981.
- [3] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.
- [4] E. Lee, P. Kim, and H. Kim. *Automatic recognition of a car license plate using color image processing*. International Conference on Image Processing, 2:301–305, 1994.
- [5] D. G. Lowe. *Distinctive image features from scaleinvariant keypoints*. International Journal of Computer Vision, 60(2):91–110, 2004.
- [6] K. Mikolajczyk and C. Schmid. *A performance evaluation of local descriptors*. IEEE Transactions on Pattern Analysis & Machine Intelligence, 27(10):1615–1630, 2005.
- [7] A. Tam, H. Shen, J. Liu, and X. Tang. *Quadrilateral signboard detection and text extraction*. International Conference on Imaging Science, Systems and Technology, pages 708–713, 2003.