Improved Matching Accuracy in Traffic Sign Recognition by Using Different Feature Subspaces

Arihito Ihara, Hironobu Fujiyoshi
Dept. of Computer Science
Chubu University
Aichi, Japan
ai@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp

Masanari Takagi, Hiroaki Kumon, Yukimasa Tamatsu
DENSO CORPORATION
Aichi, Japan
{MASANARI_TAKAGI, yukimasa_tamatsu }@denso.co.jp
kumon@rd.denso.co.jp

Abstract

This paper presents a traffic sign recognition method based on keypoint classification by AdaBoost using PCA-SIFT (principal component analysis scale invariant feature transform) features in different feature subspaces. A technique for recognizing traffic signs from an image taken with an in-vehicle camera has already been proposed to assist drivers. SIFT features are used for traffic sign recognition because they are robust to changes in scaling and rotation of traffic signs, but real-time processing is difficult because the computation cost of SIFT feature extraction and matching is high. In our method, two different feature subspaces are constructed from gradients in traffic sign images and those in general images. Detected keypoints are projected into both subspaces, and AdaBoost is used to classify whether they are on the traffic sign or not. Experimental results show that the computation cost for keypoint matching can be reduced to about half that of the conventional method.

1 Introduction

Traffic accidents are a serious problem, so traffic signs are very important to safe driving. Drivers can avoid having many traffic accidents if they drive in accordance with the signs, but some traffic signs cannot be recognized by drivers because of degradation or occlusion or because the drivers overlook them. Previous work in this area has included a method of constructing a classifier or color features from traffic sign images[1]. Template matching needs various template images covering scaling, rotation, and changes in illumination[2]. Therefore, the computation cost of matching is high. Traffic sign recognition using color is difficult because a given traffic sign can appear to have a range of color variation under different environmental conditions. On the other hand, Takagi et al.[3] presented traffic sign recognition using scale invariant feature transform (SIFT) features, which make this traffic sign recognition robust to changes in scaling and rotation of traffic signs[4]. However, real-time processing is difficult because the computation cost of SIFT feature extraction and matching is high.

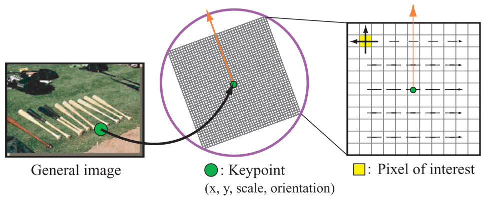

Figure 1: Gradient extraction in a local patch.

In this paper, we propose a technique for improving the matching speed and accuracy by using PCA-SIFT features projected into different feature subspaces and describe the results of an evaluation experiment performed to determine this method's effectiveness.

2 PCA-SIFT

PCA-SIFT was presented by Ke et al.[5]. It applies principal component analysis (PCA) to gradient features for a 41 x 41 patch centered at the keypoint detected by SIFT. The feature space is reduced by projecting the gradient features into a subspace constructed using PCA. This approach has the advantage of faster matching than SIFT because its feature space is composed of 36-element vectors compared with 128-element vectors in SIFT. The SIFT keypoint detection is shown below.

1. Scale-space extrema detection
2. Keypoint localization
3. Orientation assignment

PCA-SIFT is more accurate and much faster than SIFT.

2.1 PCA-SIFT descriptor

The PCA-SIFT descriptor describes the horizontal and vertical gradients of a 41 x 41 patch centered on the keypoint detected by SIFT. The patch is a region with a given scale and rotated to align its dominant orientation to a canonical direction (Figure 1). Hence, the PCA-SIFT descriptor has a feature vector $\mathbf{X}$ with $2 \times 39 \times 39 = 3042$-elements. The subspace into which the feature vector is projected is constructed by

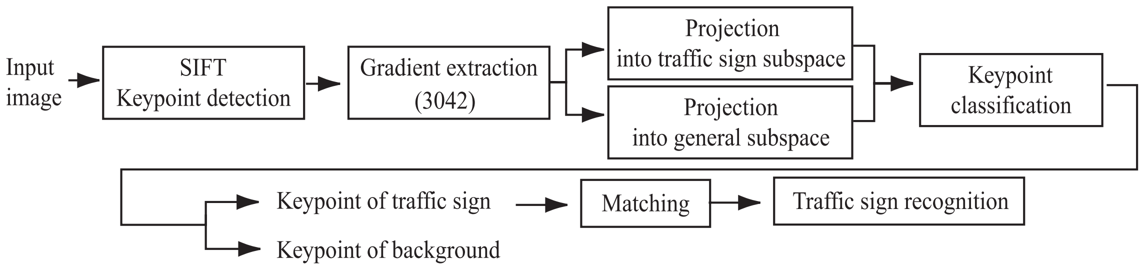

Figure 3: Flow of our method.

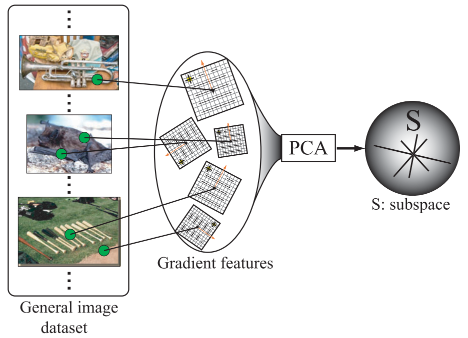

Figure 2: Construction of subspace by PCA.

a large number of such vectors $\mathbf{X}$ from a general image dataset (Figure 2). PCA-SIFT features are extracted by projecting the horizontal and vertical gradients of the patch into the subspace.

3 Proposed method

Our method constructs different feature subspaces with a dataset of traffic sign images and a dataset of general images. In addition, we construct a Real AdaBoost classifier to discriminate keypoints. The method's flow is illustrated in Figure 3 and described below.

1. Detect keypoint in input image by SIFT.
2. Horizontal and vertical gradient features centered on the keypoint are extracted.
3. Project the horizontal and vertical gradients into different feature subspaces.
4. Combine the projected features.
5. Discriminate the keypoints by using the Real AdaBoost classifier that has learned the combined feature.
6. Match only keypoints that were discriminated as points over the traffic sign.

3.1 Construction of different feature subspaces

We construct different feature subspaces with the traffic sign image dataset and general image dataset (Figure 4). The feature subspace of the traffic sign images is better expressed in local space than the feature subspace of the general images is. Therefore, the subspace of traffic sign images can express features that are difficult to express in the subspace of general images. The dimensionality of the feature space is 72; the cumulative proportion is over 60%.

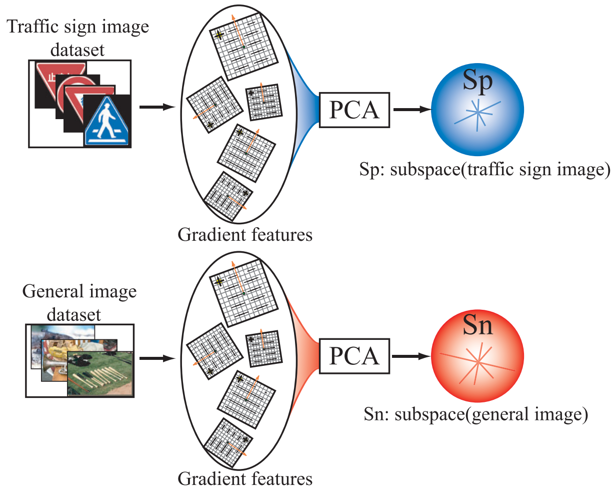

Figure 4: Different feature subspaces.

3.2 Extraction of combined feature

Gradient features extracted from input images are projected into different feature subspaces and combined by equation (1).

$$\mathbf{V} = [\mathbf{x}^T S_n, \mathbf{x}^T S_p] \tag{1}$$

where $\mathbf{V}$ is the combined feature, $\mathbf{x}$ is the gradient feature of horizontal and vertical, S is the projection matrix, S_n is the subspace constructed with the general image dataset, and S_p is the subspace constructed with the traffic sign image dataset. The keypoint involves features of both the general image class and the traffic sign image class (i.e., the keypoint has 144 elements) and is discriminated by the Real AdaBoost classifier trained with the combined feature.

3.3 Construction of Real AdaBoost classifier

The Real AdaBoost[7] classifier is used to discriminate the keypoint. Real AdaBoost calculates the degree of similarity from the positive and negative probability density distributions and trains the most different dimensionality of the feature space. A real number is used for training and discrimination because the degree of similarity becomes the evaluated value. Training data consists of images taken with an in-vehicle camera, the traffic sign class has 22,449 keypoints detected from 566 images, and the background class has 35,827 keypoints detected from 14 images. The number of the weak classifier is 100 in discrimination cost and accuracy.

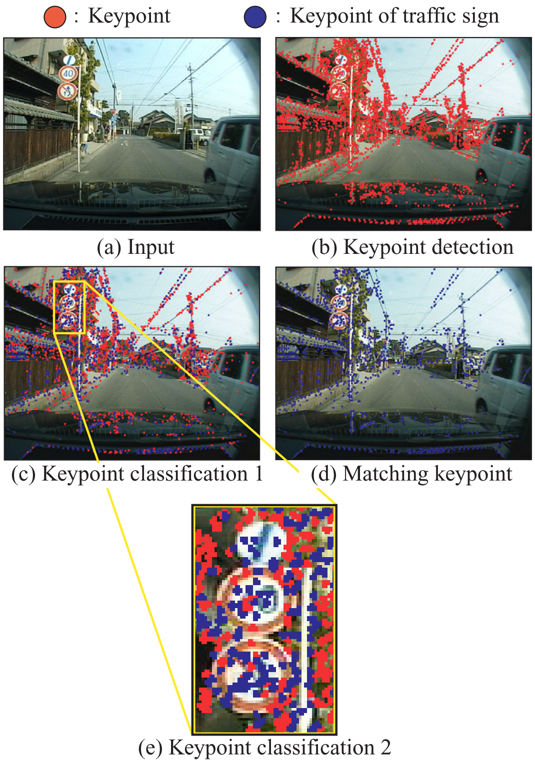

Figure 5: Example of keypoint classification.

3.4 Keypoint discrimination

Keypoints in the input image are detected by SIFT. Gradient features around a detected keypoint are projected to the general and traffic sign image classes, and they are combined. Keypoints are discriminated by the Real AdaBoost classifier with combined features. An example of keypoint discrimination is shown in Figure 5. The red points are keypoints detected using SIFT, and the blue points are keypoints detected over the traffic sign. Figure 5(a) is the input image and Figure 5(b) shows the detected keypoints (3542 keypoints). Figure 5(c) shows keypoints removed as general-class ones by the Real AdaBoost classifier (2150 keypoints). The region with traffic signs in Figure 5(c) and (e) shows keypoints discriminated as being in the traffic sign class. Figure 5(d) shows matching keypoints (1392 keypoints). Thus, the proposed method can lower the matching cost compared with the conventional method.

3.5 Matching

The template and discriminated keypoints are matched by Euclidean distance, which is calculated by equation (2). The Euclidean distance calculation uses the features of traffic sign image class (72-elements).

$$d(\mathbf{V}_{I1}, \mathbf{V}_{I2}) = ||\mathbf{x}_{I1}^T S_p - \mathbf{x}_{I2}^T S_p||^2 \qquad (2)$$

where $\mathbf{V}_{I1}$ and $\mathbf{V}_{I2}$ express keypoints of the traffic sign image $I1$ and input image $I2$, respectively. If $d(\mathbf{V}_{I1}, \mathbf{V}_{I2})$ is small, the degree of similarity is high, and thresholding is used for the matching.

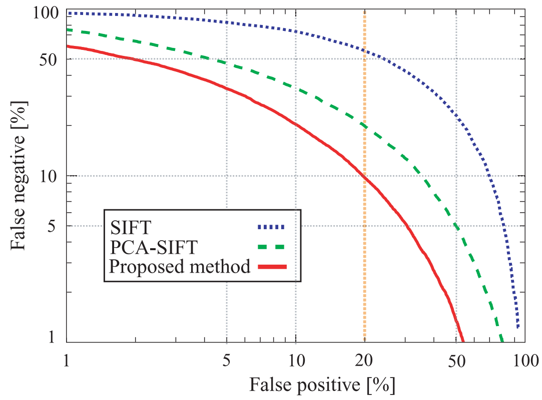

Figure 6: Results of keypoint classification.

4 Estimation of discrimination and matching

4.1 Experiment 1: keypoint discrimination

The estimation of keypoint discrimination shows the effectiveness of different feature subspaces of the general and traffic sign image classes. The keypoint classifier was constructed from SIFT features, PCA-SIFT features, and our method's features.

4.1.1 Overview of experiment 1

The training dataset consisted of template images, which were illustrations of traffic signs from the Caltech256 dataset[6]. Keypoints detected from the template images were trained using Real AdaBoost. Training keypoints were the traffic sign image class of 41,919 keypoints and the general image class of 575,598 keypoints. The test dataset used the training dataset to compare the abilities to express the features of subspaces. The results were evaluated using the detection error tradeoff (DET) curve expressed on a double logarithmic chart.

4.1.2 Results of keypoint discrimination

The results of keypoint discrimination are shown in Figure 6. The blue line is the SIFT features, the green line is the PCA-SIFT features, and the red line is our method's features. With a false positive rate of 20% (orange line), our method had about a 10% lower false negative rate than PCA-SIFT. This means that different feature subspaces can express features that PCA-SIFT cannot express. The dimensions of features selected by Real AdaBoost are shown in Figure 7. Because of the use of PCA, low-dimensional features are expressed better than high-dimensional features. And the Boosting algorithm of AdaBoost selects low-dimensional features. Our method involves more effective features for discrimination than the general image class because the features of the traffic sign class are selected by Boosting. Therefore, separating the general and traffic sign image classes is effective.

4.2 Experiment 2: Matching accuracy

Next, we compared the matching accuracies of the conventional method and our method. The conventional method was traffic sign recognition using SIFT features.

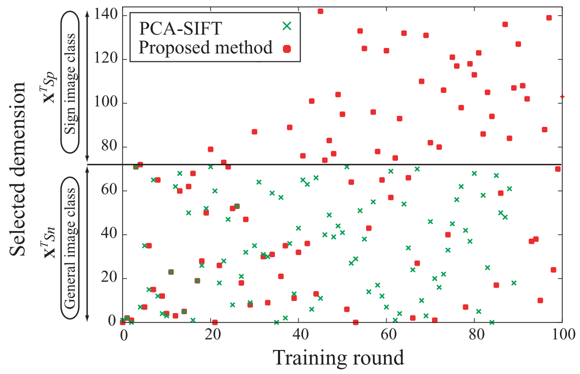

Figure 7: Features selected by Boosting.

Table 1: Matching accuracy [%]

	True matching rate	False matching rate
SIFT[3]	8.0 (163/2034)	91.9 (1871/2034)
Proposed method	14.4 (39/269)	85.5 (230/269)

4.2.1 Overview of experimental 2

The test images were 500 images taken with an in-vehicle camera. The conventional and proposed methods were compared in terms of false matching rate and true matching rate, which were computed by:

$$False\ matching\ rate = \frac{Num\ of\ false\ matches}{total\ matches} \quad (3)$$

$$True\ matching\ rate = \frac{Num\ of\ true\ matches}{total\ matches}. \quad (4)$$

4.2.2 Results for matching accuracy

The results for matching accuracy are given in Table 1. Our method had a 6.4% lower false matching rate than the conventional method. Although the number of true matches was lower than with the conventional method, the true matching rate was improved by our method because keypoints in the background were rejected by the Real AdaBoost classifier.

4.3 Experiment 3: Matching cost

Finally, we compared the matching cost of the conventional and proposed methods.

4.3.1 Overview of experiment 3

We iterated matching ten times per image and estimated the average. The two methods ran on machines with the same configuration: Xeon 3.00-GHz CPU with 8 cores. Our method involves not only a matching cost but also a discrimination cost. The number of matching keypoint was 1234.

4.3.2 Results for matching cost

The results for matching cost are given in Table 2. The matching cost for our method was reduced to about half that of the conventional method because background keypoints were rejected and the feature dimensions were reduced.

Table 2: Matching cost[ms]

	Classification	Matching	Total
SIFT[3]	—	425.4	425.4
Proposed method	20.3	182.5	202.8

Figure 8: Examples of traffic sign recognition.

4.4 Examples of traffic sign recognition

For traffic sign recognition, we used the voting method proposed by Takagi et al. [3]. Some examples of traffic sign recognition are shown in Figure 8. Our method could recognize traffic signs that the conventional method could not because background keypoints were rejected.

5 Conclusion

This paper presented a traffic sign recognition method using PCA-SIFT in different feature subspaces. The effectiveness of using different feature subspaces was confirmed by experiments.

References

[1] K. Myochin, Y. Kageyama, M. Nishida, "A Study on Recognition of Circular Road Signs in Color Scene Image," Technical report of IEICE PRMU, Vol. 104, pp. 181–185, 2005.

[2] K. Doman, T. Takahashi, Y. Mekada, I. Ide, H. Murase, "Generative Learning for Traffic Sign Detection Using Cascaded Classifiers," Meeting on Image Recognition and Understanding 2007, IS-5-04, pp. 1396–14011, 2007.

[3] Masanari Takagi, Hironobu Fujiyoshi, "Road Sign Recognition using SIFT feature," SSII07, LD2-06, 2007.

[4] D. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision (IJCV), 60(2), pp. 91–110, (2004).

[5] Y. Ke, R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 511–517, 2004.

[6] http://www.vision.caltech.edu/Image_Datasets/Caltech256/

[7] R. E. Schapire, Y. Singer, "Improved Boosting Algorithms Using Confidence-rated Predictions," Machine Learning, No. 37, pp. 297–336, 1999.