A ROBUST ROAD SIGN RECOGNITION METHOD USED BY IN-VEHICLE CAMERA SYSTEMS

Xiaohui XIE[1], Weijie LIU[2], Gang WU[1], Qiang WEI[1]
Panasonic Research and Development Center, China [1]
Panasonic AV Core Technology Center, Japan [2]
Email: xiexiaohui@cmrd.panasonic.com.cn

ABSTRACT

A robust road sign recognition method is proposed in this paper. Templates of road sign images are represented as a hierarchical structure of some component blocks. In the paper approaches for block selection and for hierarchical structure construction are our contributions. Existing block based methods require that template images having sufficient texture, but which is not the situation for road signs. In our method, blocks used as basic matching units (BMU) are selected based on calculations of their penetration numbers. And the hierarchical structure is formed by comparing the cover areas of different subsets of all BMUs. Experiments on actual video clips taken by in-vehicle cameras proved the robustness and efficiency performance of our method.

INTRODUCTION

The main objective of road signs is to regulate traffic. Road signs supply information to help drivers operate their cars in such a way as to enhance traffic safety. However, they may not be noticed sometimes when, for example, a driver gets tired or drunken. A stable road sign recognition system is thus desirable to alert the driver or give a signal to an ACC system at that time. Other applications of road sign recognition include navigation to unmanned driving.

Many techniques have been developed to detect road signs. Piccioli et al [1] retrieved triangular and square contours in an edge image by selecting the edge segments with proper slopes and checking whether their endpoints are close enough and form certain angles. Their approach usually suffers difficulties when retrieving broken edges. Escaleera et al [2] localized triangular signs by seeking the coexistence of three types of corners that form a triangle; and the similar principle were applied to square and circle detection. However, their method strongly relies on corner detection and classification, which are not easy in clutter scenes. Jimenez et al [3] developed a sign shape classification algorithm based on the FFT of the contour signature of a segmented blob. But their classification requires completely-filling blobs; thus an intractable preprocessing of enclosing blob gaps is necessary. Loy and Barnes [4] applied a radial symmetry algorithm to detect regular polygons. They took advantage of the radial symmetry nature of a regular polygon to locate its center. Moreover, they used a rotation-invariant measure, named n-angle gradient, to distinguish polygons with different numbers of side *n*. But their method can only be used for detecting regular polygons.

In our research, we have developed a rapid algorithm to detect both regular and non-regular shape road signs, which is reported in another paper [10].

In the paper, we concentrate on recognition, which verifies detected road sign candidates of a particular bound shape with depicted road sign templates. Compared to detection, it seems few techniques have been proposed for road sign recognition. Normalized cross correlation [11], neural networks [9] and template matching [8] methods are most used in literatures. But all consider seriously the classification of lots of road signs, rather than recognition of one road sign under complex environments. Indeed under a particular bound shape there still exist many road signs, that is, recognition have to deal with more objects than detection, but road signs that should be targeted by an in-vehicle camera system in order to alert a driver are limited. So for our system it is more important to be accurate and robust rather than to be applicable to most road signs.

In fact, many variations occurring to road signs have to be taken into account for an in-vehicle camera system. Such as (some are shown in Fig.9): (1) Colors may fade after long exposure to the sun. Moreover, paint may even flake or peel off, and signs may get damaged. (2) Air pollution and weather conditions may decrease the visibility of road signs. (3) Outdoor lighting conditions vary from day to night and may affect the appearance colors of road signs. (4) Obstacles, such as trees, poles, buildings, and even vehicles and pedestrians, may occlude or partially occlude road signs. (5) Video images of road signs often suffer from blurring in case that the camera is mounted on a moving vehicle. (6) Size of road sign images usually increases when the vehicle moves close to them, as shown in Fig.1.

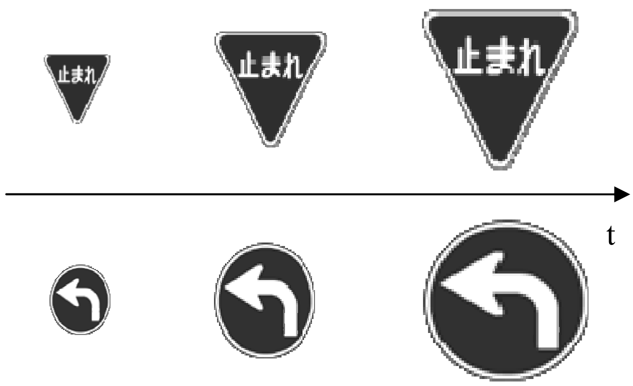


Fig.1 Road sign samples at different distances

To be efficient and robust under difficult conditions, object recognition methods based on local features have been proposed. One approach uses feature points. Under the approach, the SIFT method [6] is well known for its invariance property of the generated features to image scaling, translation, rotation, and partially to illumination changes and affine projection. But the first step in SIFT, that is, to determinate the points of interest by constructing a Gaussian pyramid, limits greatly its adoption to road sign recognition, because images captured at present usually are of small size and are blurred. Another approach uses feature blocks. A recent proposal under the approach suggested using feature hierarchies [7], where image blocks of locally maximal mutual information are extracted hierarchically to represent objects. But the method is not effective to pick up suitable blocks from road sign templates because the templates have not sufficient texture, with their only two or three colors, i.e. gray-levels, and definite patterns. Simple measures for block selection and a flexible structure for grouping blocks are necessary.

A novel recognition method based on a block hierarchy is presented in this paper. The blocks are selected based on a simple measure suitable for road signs. Rather than as usual tree hierarchies, the selected blocks are organized as a different multi-layer structure in our method. The block hierarchy generated takes into consideration that road signs are of few grey levels and simple patterns. One block hierarchy is generated by searching one template of a kind of road signs and is scaled and overlaid onto candidates detected. Matching between a template and a candidate is calculated based both block features and their hierarchy relations. Experiment results show that the proposed method has good robustness and calculation efficiency performance.

The rest of the paper is organized as follows. Firstly, the approaches to block selection and hierarchy construction are described respectively in the following sections. Then the matching process is shown. Finally, experimental results and conclusions are given.

BLOCK SELECTION

As a prerequisite, candidate road signs of a particular bound shape are assumed to be detected from the input images. As every candidate may have different size and varying color i.e. grey level, appearance, it is necessary to use some primary features of a candidate in order to ensure robust and real time recognition. It is noticed that road signs are usually painted with few colors, that is, they should show few gray levels with respect to their grayscale templates. So we can normalize and binarize the detected candidates, and to implement matching using binary images.

We definite a concept named penetration number for each image block to measure its information. According to the penetration number of blocks we search basic matching units (BMU) of an image, which are defined as the blocks with local maximum information. Details are described as below.

Penetration Number

Penetration number of an image block is generally defined as the intensity change frequency between foreground and background. For grey and color images, it could be obtained by counting edge points. As binary images are sufficient in our case, we give only a detailed description to binary images.

Suppose A is an l_m * l_n rectangle block area in a binary image I, its penetration number is calculated by

$$F = \frac{1}{l_m}\sum_i p(x_i) + \frac{1}{l_n}\sum_j p(y_j) \qquad (1)$$

where, $p(x_i)$ means value change times from 0 to1 and 1 to 0 at line $x = x_i$, while $p(y_j)$ value change times from 0 to1 and 1 to 0 at line $y = y_j$. As shown in Fig. 2, line $x = x_0$ has a penetration number of two while line $y = y_0$ four. From the definition we could say that a block with a bigger penetration number contains more information. So block A with penetration number F_A is said to be an informative block with F_A.

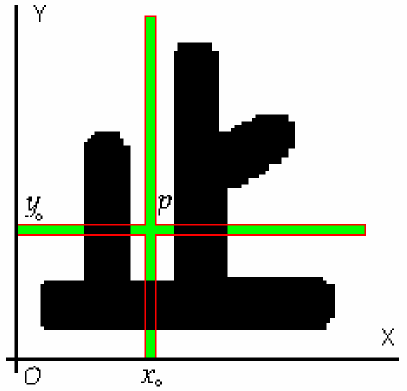


Fig.2 Penetration number

Basic Matching Unit (BMU)

A basic matching unit (BMU) is defined as such an image block which has a bigger penetration number than that of all its surrounding image blocks. Distribution of all BMUs with different scales in an image depends on contents of the image. Based on the penetration number, two BMU areas may be separate, overlapping and containing each other.

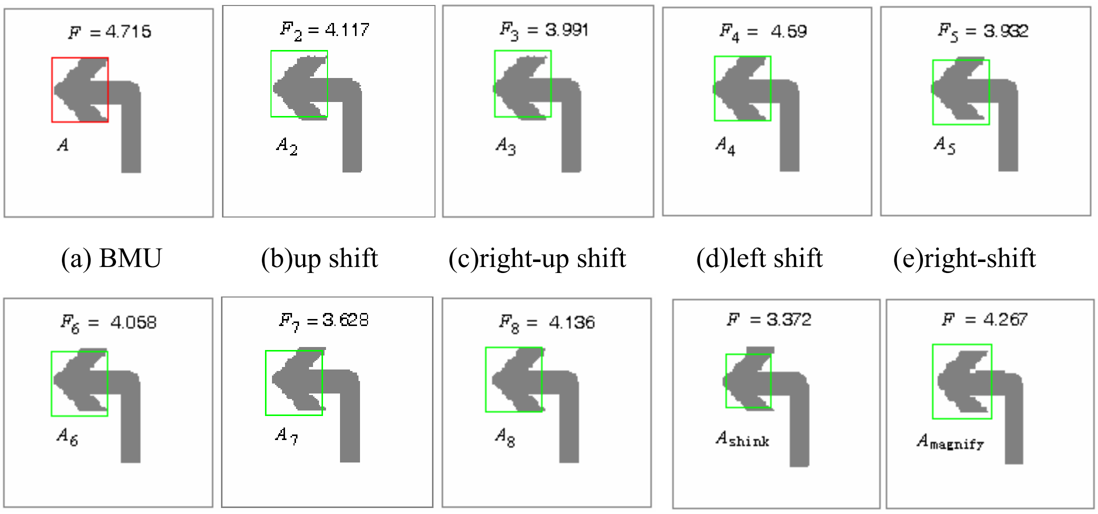
BMUs are searched only from templates of road signs. For candidates to be detected, the corresponding BMUs are extracted by scaling and overlaying one template onto it. In another words, BMUs are found at training phase. So we can use an exhaustive search.

For a template road sign, blocks of different sizes, say from 10% up to 100% of image size, are considered and all their penetration numbers are calculated. Then with a particular block, we investigate its surrounding blocks to determine if it is a BMU. In detail, suppose A is a block with a penetration number F_A, we can translating A by [Δx, Δy] or scaling A by [1-k, 1+k] to obtain a block A', where Δx and Δy are small constants and $0<k<1$. Assume A' with a penetration number F'_A, if condition

$$F_A > F_A^{'}(\Delta x, \Delta y, k) \qquad (2)$$

is satisfied, A is determined to be a BMU.

As penetration number is an information measure, from the definition of formula (2) we know that a BMU points to an area of a road sign where a local maximum with respect to the information measure occurs. If compared to [7], a BMU corresponds to a variant of the informative blocks which is specifically applicable to images of simple texture. Fig. 3 gives an example for extracting a BMU for a turn-left road sign. When moving a block up, down, to left and to right, the corresponding penetration numbers are calculated. In the figure, (i) and (j) show the results for shrunk and magnified blocks respectively.



(a) BMU (b)up shift (c)right-up shift (d)left shift (e)right-shift

(f) left-down shift (g) down shift (h)right-down shift (i) shrink (j) magnify

Fig.3 A sample of BMU

HIERARCHY CONSTRUCTION

After all BMUs of an image are searched, a flexible hierarchy structure that organizes those BMUs is necessary. Two reasons exist for needing such a hierarchy. One is that road sign candidates detected from an input image are usually noise, so not all details can be extracted always. A coarse-to-fine multi-layer feature structure is then effective. Another one is for speed-up. As road signs are observed from small size to big size (see Fig. 1), a coarse-to-fine matching process may give a decision at an early stage.

But unlike the tree structure of [7], where sub-blocks are further extracted from a parent layer block, we develop a different structure here. In our case, a tree structure is helpless because our BMUs with different scales usually overlap to each other. Thus a tree structure can not organize all BMUs necessary.

We give a brief description about a relationship between BMUs and the image. Then propose our method for hierarchy construction.

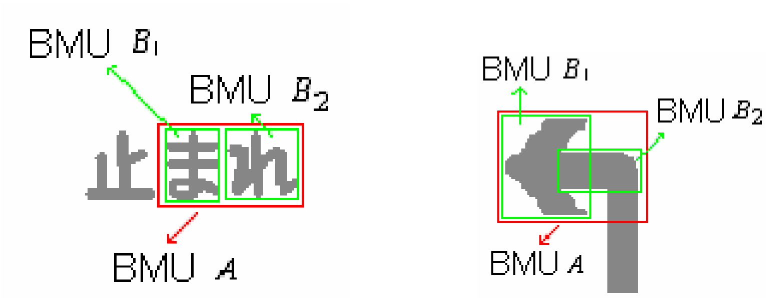
Maximal Cover of Images

A subset of BMUs give a cover to the foreground of an image. Let Z_A is the foreground of image *I,* and A_i *(i=1,…,N)* are a subset of BMUs, if A_i satisfies the condition that $\bigcup_i A_i$ covers the largest area of Z_A than any other subsets, and $A_{i1} \not\subset A_{i2}$ $(\forall i1, i2,$ *i1≠i2, i1=1,…,N, i2=1,…,N*), A*i* *(i=1,…,N)* is defined as the maximal cover of the image *I*. Similarly, the maximal cover can be defined for blocks too.

Construct Hierarchical Structure

Firstly, we find all maximal covers from all subsets of BMUs. Then we decide the one with the smallest number of BMUs as the top-layer of the hierarchical structure. And its BMUs are called top-layer BMUs. In our experiment, the top-layer is formed with only one BMU, which corresponds to the rectangular circumscribing the foreground of an image. Next, for each top-layer BMU we find all its maximal covers, but the covers are required not containing any top-layer BMUs. Likely, we find the cover with the smallest number of BMUs for each top-layer BMU and form the second-layer of the hierarchy. The procedure can be repeated until no more cover exists, and thus a complete multi-layer hierarchical structure is built. It should be noted that BMUs at a child layer may cover a larger area than its parent.

Two examples are shown in Fig. 4, where (a) depicts a stop, and (b) a turn-left road sign in Japan BMU *A* is a parent-layer BMU, and *B1, B2* are two child-layer BMUs.



(a) Stop road sign (b) Turn-left road sign

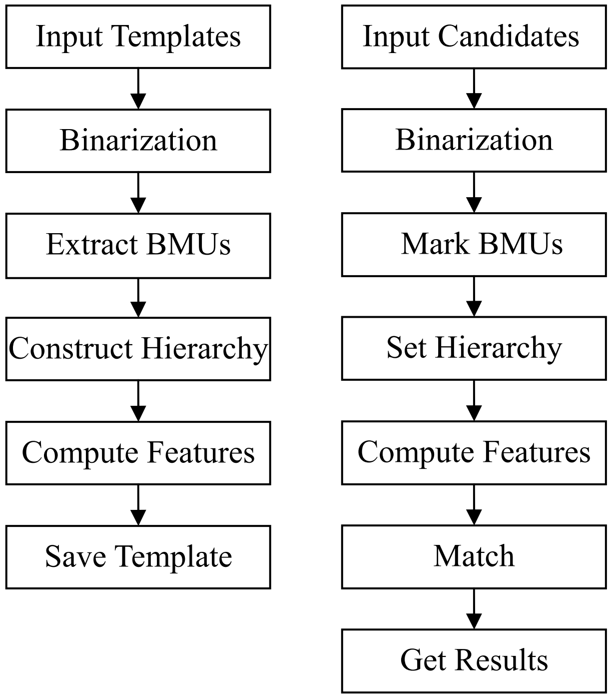
Fig.4 BMU hierarchy samples

HIERARCHICAL STRUCTURE MATCHING

Recognition Flowchart

BMUs and their hierarchical structure are built offline at a training step, where templates of target road signs are used as inputs. Otherwise, an online matching process is enforced between detected road sign candidates and depicted templates. BMUs of the candidates are marked by scaling and overlaying corresponding templates onto them.

A primary flowchart of our recognition process is shown in Fig. 5.



(a) Offline training (b) Online matching

Fig. 5 Flowchart of recognition

Examples of Hierarchical Structure

Fig. 6 shows two examples of hierarchical structure with two Japanese road signs. Both are constructed with three layers. The left one gives a direct result in coincidence with our intuition, where the text area is suitably divided at different levels.

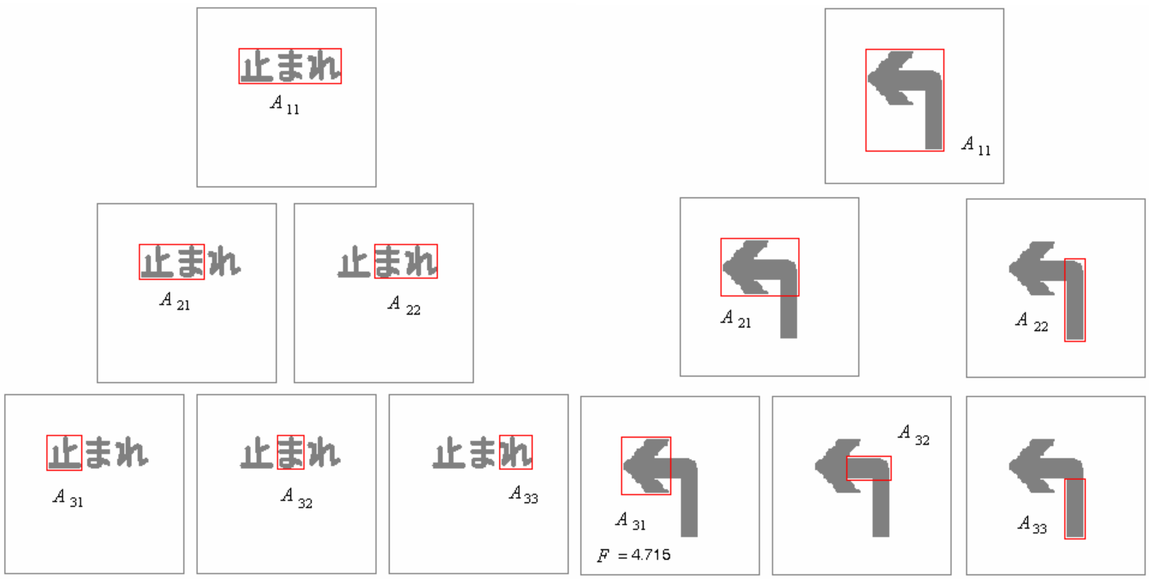


Fig.6 Examples of hierarchical structure

Features and Similarity Measure

Matching between two hierarchical structures of BMUs are implemented based on features of BMUs and their similarity measures. At present, several primary features are used in our development. Among them, one is the gravity center positions of BMUs within an image, which represents the space distribution of all BMUs. Other features include vertical and horizontal histograms of BMUs, which are thought suitable block features especially for low resolution images. Examples are shown in Fig. 7.

Similarity measure is implemented both with features and with hierarchical layers. For example, gravity center positions are compared with their relative coordinates within the bound shape of road signs, and histograms are compared through calculating the correlation of their outline curves. Other features, e.g. errors occurring when overlay templates onto detected candidates, can also be considered, but we indicate that simple features work well for our systems.

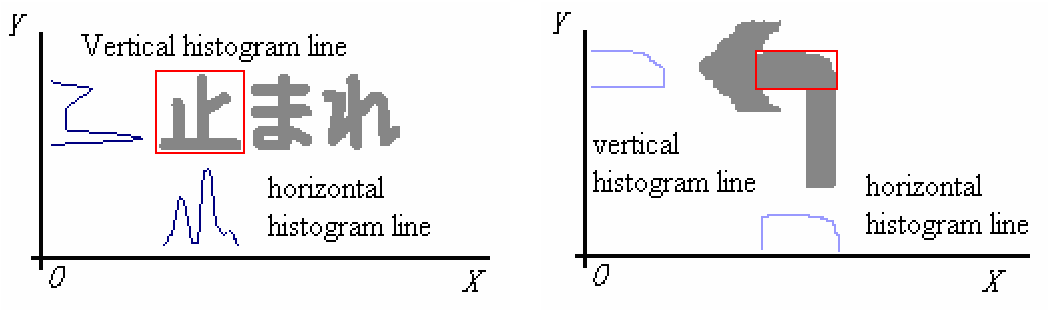
As BMUs are organized as a multi-layer structure, similarity measure for hierarchies is a fusion of that for individual layers. And the similarity measure for a layer is also a fusion of that for its BMUs. Simply, suppose the similarity of a BMU pair j at level i is S_{ij}, and the similarity of layer i is S_i, $1 \leqslant i \leqslant M$, $1 \leqslant j \leqslant N$, the fusions can be implemented by weighted sums as follows

$$S_i = \sum_j \beta_j \cdot S_{ij} \tag{3}$$

$$S = \sum_i \alpha_i \cdot S_i = \sum_i \alpha_i \cdot (\sum_j \beta_j \cdot S_{ij}) . \tag{4}$$

where, αi and βj are suitable weights.

One important note is that we calculate the similarity from the top-layer BMUs. This is the situation for an in-vehicle camera system. And we can know if the similarity is low enough or high enough. If too low we can decide that the candidates are not real road signs at an early time when the in-vehicle camera is still far from the road sign and can not supply more details. If high enough we also decide a success for the recognition, according to how many BMUs have been considered. Either gives a speed up for the matching.



(a) Stop road sign (b) Turn-left road sign

Fig. 7 Feature selection for BMUs

EXPERIMENTAL RESULTS

Experiments on real videos captured with in-vehicle cameras are conducted to evaluate the proposed method. All our test data includes 12 video clips taken under different weather conditions, including in the morning, daylight and nightfall. Road signs include the Stop and the Turn-left road signs in Japan. Some videos also include blurred and occluded road signs.

Table 1 gives the total experimental results and Fig.8 shows the corresponding ROC curves.

Video Clip Number	Frame Number	Candidates Number	False Positive Rate	False Negative Rate	Correct Rate
12	1429	1615	0.74%	0.43%	98.83%

Table 1 Experiment results on 12 video clips

Table 1 shows that the proposed method achieves a recognition rate of 98.83% with 1615 detected candidates. It is better than that of using other methods, such as template matching with about 70%, and neural network with about 80%. At present our method recognizes candidates stably if the candidates have a size larger than 28x28, which corresponds to about 50 meters of distance with respect to our camera parameters.

Efficiency is also tested at the experiment. By using a PC with a Pentium CPU 1.8GHz and 256M memory, our method costs 4.5ms to match one candidate, that is, 222 matches can be done in one second. Figure 9 gives some variations of the Stop road sign. It is obvious that the proposed method gives a satisfied robustness performance besides its better accuracy and efficient real-time performance.

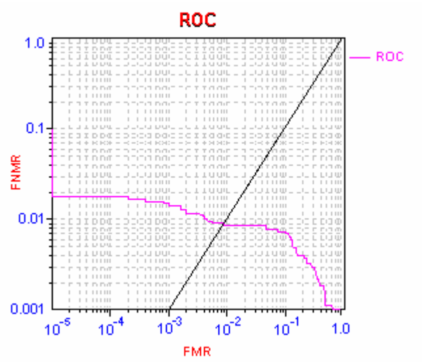


Fig.8 ROC curves of the experiment results

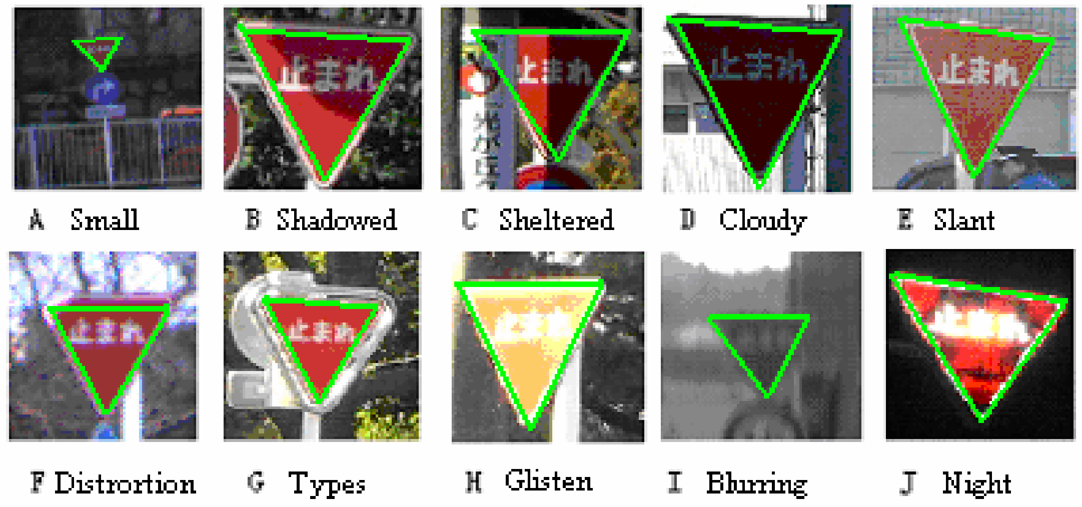


Fig.9 Samples of recognition objects

CONCLUSION

In this paper, we proposed a novel road sign recognition method. The method represents a road sign templates using a hierarchical structure of image blocks, which we have named as basic matching units (BMU). The blocks are selected based on penetration numbers and their hierarchical structure is constructed based on calculating maximal covers of BMU subsets. Rather than existing approaches, the hierarchical BMU structure is especially suitable for recognizing road signs. Experiment results show that our method has a satisfied robustness and efficiency performance.

REFERENCES

[1]. G. Piccioli, E. De Micheli, P. Parodi and M. Campani, “Robust Method for Road Sign Detection and Recognition”, Image and Vision Computing, Vol. 14, 1996.

[2]. Arturo de la Escalera, Luis E. Moreno, Miguel Angel Salichs and Jose Maria Armingol, “Road Traffic Sign Detection and Classification”, IEEE Transactions on Industrial Electronics, Vol. 44, No. 6, 1997.

[3]. P. Gil Jimenez, S. Lafuente Arroyo, H. Gomez Moreno, F. Lopez Ferreras and S. Maldonado Bascon, “Traffic Sign Shape Classification Evaluation II: FFT Applied to the Signature of Blobs”, IEEE Intelligent Vehicles Symposium, 2005.

[4]. Gareth Loy and Nick Barnes, “Fast Shape-based Road Sign Detection for a Driver Assistance System”, IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004.

[5]. Milan Sonka, Valclav Hlavac and Roger Boyle, “Image Processing, Analysis, and Machine Vision”, Second edition, Brooks/Cole press, 2001.

[6]. D.G.Lowe. Distinctive Image Features from Scale-Invariant Keypoints. IJCV, 2004.
[7]. B.Epshtein, S.Ullman, Feature Hierarchies for object Classification, ICCV 2005, Beijing, China
[8]. Y.B.Lauziere, D.Gingras, F.P.Ferrie. A model-based road sign identification system. IEEE International Conference on Computer Vision and Pattern Recognition, 2001.
[9]. H.Ohara, I.N.S. Miki, N.Yabuki. Detection and recognition of road signs using simple layered neural networks. International Conference on Neural Information, 2002.
[10]. G. Wu, W. J. Liu, X. H. Xie, and Q. Wei, A shape-based method for traffic sign detection, submitted to ITS World Congress 2007, Beijing.
[11]. N. Barnes, G. Loy, D. Shaw, and A. Robles-Kelly, Regular polygon detection, ICCV 2005.