PAPER Special Issue on Information Systems for ITS

An Active Vision System for On-Line Traffic Sign Recognition

Jun MIURA†, *Member*, **Tsuyoshi KANDA**†, **Shusaku NAKATANI**†, *Nonmembers*, *and* **Yoshiaki SHIRAI**†, *Member*

SUMMARY
This paper presents an active vision system for on-line traffic sign recognition. The system is composed of two cameras, one is equipped with a wide-angle lens and the other with a telephoto lens, and a PC with an image processing board. The system first detects candidates for traffic signs in the wide-angle image using color, intensity, and shape information. For each candidate, the telephoto-camera is directed to its predicted position to capture the candidate in a larger size in the image. The recognition algorithm is designed by intensively using built-in functions of an off-the-shelf image processing board to realize both easy implementation and fast recognition. The results of on-road experiments show the feasibility of the system.
key words: Traffic sign recognition, Active vision, Real-time vision, ITS

1. Introduction

Traffic signs provide the driver with various information for safe and efficient navigation. Automatic recognition of traffic signs is, therefore, important for automated driving or driver assistance systems [1] [2] [3].

Since these signs are usually painted by distinctive colors, they may be easily detected using color information [4]. However color information is sensitive to the change of weather or lighting condition and, therefore, it is sometime difficult to extract traffic signs reliably only by color. In addition, in urban cluttered scenes, other signboards or buildings with the similar color to traffic signs may make it more difficult to extract an appropriate region for the target signs in the image. To cope with such problems, we additionally use shape information [5] [6]. It is also possible to search the image for the signs using the shape information directly [3].

To recognize traffic signs, it is often necessary to recognize their characters and symbols. For reliable recognition of characters and symbols, it is desirable to capture a sign in a large size in the image. To do this, we use a camera with a telephoto lens and direct it to the target. The camera direction is automatically determined based on the prediction of the target motion in the image.

This paper describes an active vision system that can recognize various traffic signs on-line based on the above-mentioned approach. We design the recognition algorithm by intensively using built-in functions of an off-the-shelf image processing board in order to realize both easy implementation and fast recognition. One of the objectives of this research is to examine the applicability of commercially-available equipments to realization of working traffic recognition systems.

Manuscript received February 28, 2002.
Manuscript revised February 28, 2002.
†The authors are with the Department of Computer-Controlled Mechanical Systems, Osaka University

2. System Overview

Fig. 1 illustrates the system configuration. The system has two cameras, one is equipped with a wide-angle lens (called *wide-camera*) and the other with a telephoto lens (called *tele-camera*), and a PC with an image processing board. The wide-camera is directed to the moving direction of the vehicle. The tele-camera can change the viewing direction to focus the attention to the target sign.

The recognition process is composed of the following steps:

1. Detect candidates of traffic signs using color and shape information.
2. Screen candidates using edge information.
3. Predict the motion of a screened candidate in the image and direct the tele-camera to it.
4. Extract characters and symbols (if necessary).
5. Identify the sign by pattern matching.

The following sections describe these steps. We use as examples two kinds of traffic signs: one is for indicating a speed limit (we call it *speed sign*) and the other is for guiding routes (we call it *guidance sign*).

3. Detecting Sign Candidates

In detecting candidates for a sign, we first extract uniform regions whose color and shape are similar to those of the sign. We then extract edges around each extracted region and verify that the edges form a specific shape. This section describes the sign candidate detection process in detail.

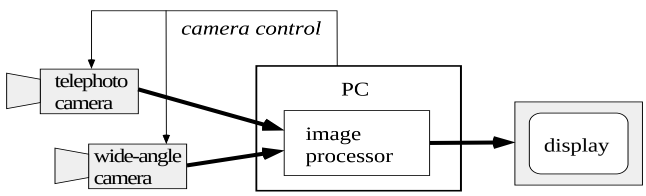


Fig. 1 System overview

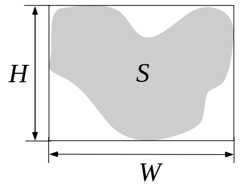
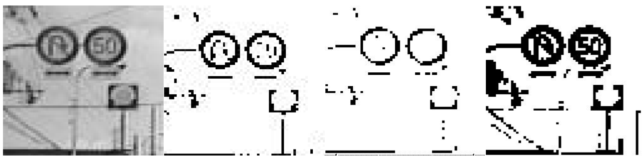


Fig. 2 Shape parameters.



(a) input (b) good (c) too low (d) too high

Fig. 3 Binarization results using various thresholds.

3.1 Extracting Candidates by Color and Shape

3.1.1 Speed Sign

A speed sign is characterized by its circular shape and a white region with a red boundary (see Fig. 3(a)). One may think that the red boundary region is used as the cue for extraction. However, since the red region has low intensity values, color information is often unreliable. Therefore, we use the white circular region as the first cue for detection. Candidate white regions are extracted by binarization with *area filtering*, which only keeps white regions whose areas are within a predetermined range.

The shape of each extracted regions are then checked by using simple shape features. Using the height H, the width W, and the area S (see Fig. 2), we define the following two shape features:

$$R = \frac{W}{H}, \quad (1)$$

$$E = \frac{S}{WH}, \quad (2)$$

where R is the aspect ratio and E indicates how the shape is close to a rectangle. We here keep speed sign candidates which satisfy the following conditions:

$$0.7 < R < 1.1,$$
$$0.5 < E.$$

The appropriate threshold for binarization is sensitive to the change of lighting condition, as shown in Fig. 3, and is difficult to determine uniquely in advance. Therefore, we perform binarization multiple times using different thresholds, and check the above condition for every white regions in the binarized images to collect candidate regions.

The same detection method is applied to zoomed-up images. Fig. 4 shows the detection results in zoomed-up images under various lighting conditions. In these examples,the intensity of the white circle ranges from 165 to 215 in 8bits; nevertheless the speed sign candidates are properly detected thanks to the multiple-thresholds approach.

(a) front light. (b) backlight. (c) cloudy weather.

Fig. 4 Detection of speed sign candidates under various lighting conditions.

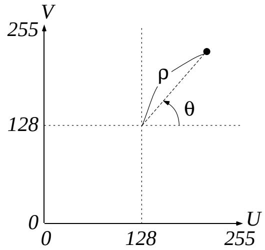


Fig. 5 Relationship between U-V and ρ-θ.

Although this multiple-thresholds approach is effective in not missing any candidates, at the same time it may extract many candidate regions that do not come from the target signs. So an appropriate screening process must follow.

3.1.2 Guidance Sign

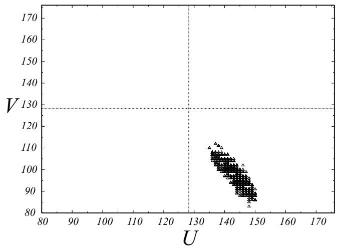
A guidance sign is characterized by its rectangular shape and blue color with white characters and symbols (see Fig. 16(a)). As in the case of speed signs, we perform binarization with area filtering in the color image space multiple times using different thresholds. The thresholds are determined by analyzing the actual distribution of data in the YUV color space taken in various weather and lighting conditions.

The YUV color space, which is used in many commercial image processing boards including ours, is defined by intensity Y and color differentials U and V. If U and V values are digitized in 8 bits, hue θ and saturation ρ can be calculated by (see Fig. 4):

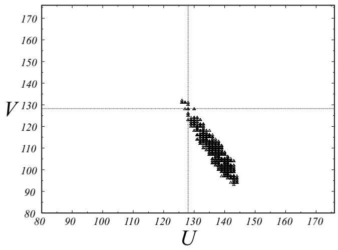
$$\theta = \arctan\left(\frac{V-128}{U-128}\right),$$
$$\rho = \sqrt{(U-128)^2 + (V-128)^2}.$$

Fig. 6 shows some example color distributions of guidance signs. The figures show that while the saturation value is affected by the change of lighting condition, the hue value is not. Therefore, we use three thresholds for the hue value with constant thresholds for the hue and the intensity values.

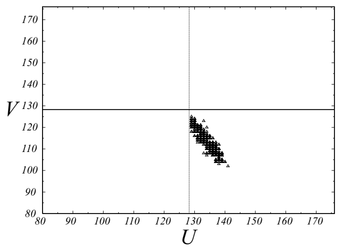
Concerning the shape information, we use the following two conditions on the shape features (see eqs. (1) and



(a) sunny daytime.



(b) cloudy daytime.



(c) evening.

Fig. 6 Change of color distributions of guidance signs. U, V values are quantized to $0 \sim 255$. $(U, V) = (128, 128)$ indicates that the saturation component is zero.

(2)):

$$0.5 < R < 1.8,$$
$$0.7 < E.$$

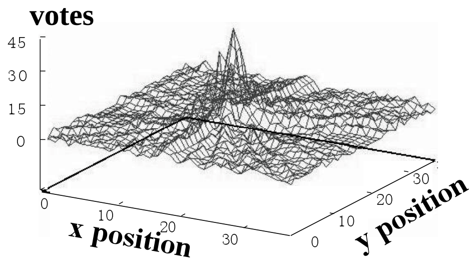
3.2 Screening Candidates by Edges

For screening candidates using edge information, we set a search area around each candidate region detected at the previous step and extract edges in the area to see if the edges form a specific shape.

3.2.1 Speed Sign

We apply the Hough transform to detecting the circular shape of speed signs. If an edge is a part of a circle, the center of the circle should exist on the line which passes the edge and has the same direction as the gradient of the edge. Using this fact, we calculate such a line for each edge and vote the line in the search area. If there is a prominent peak in the area, we determine that a circle exists. Fig. 7 shows an example of circle detection. Fig. 8 shows a detection result of speed sign candidates screened using edge information.

(a) search area (b) edge image



(c) voting result

Fig. 7 Detecting a circle.

(a) original image (b) before screening (c) after screening

Fig. 8 Detection of speed sign candidates.

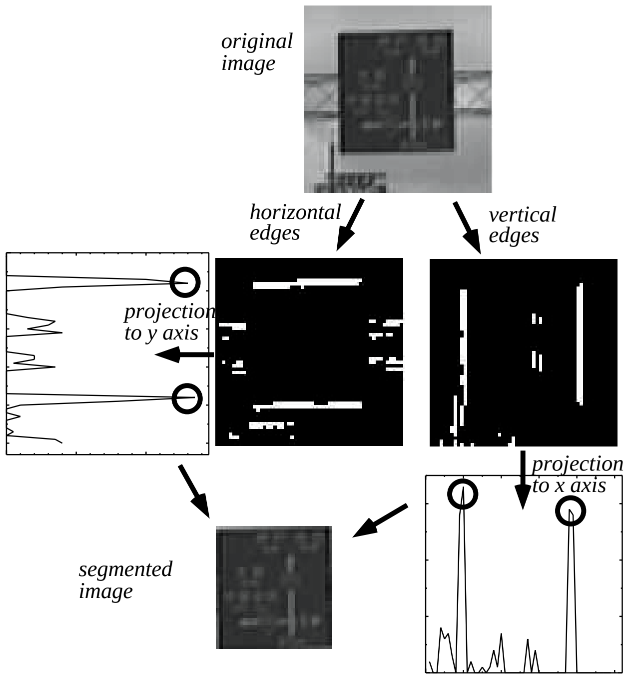


Fig. 9 Detecting four boundary line segments of guidance sign.

3.2.2 Guidance Sign

For guidance signs which have rectangular shapes, we detect four boundary line segments. Boundary detection is done by a simple vertical and horizontal projection of edges, as shown in Fig. 9. Sometimes four surrounding line segments are not fully detected due to, for example, strong edges of a near object in the image. In such a case, if we have at least one strong line segment, we predict the position of other segments and try to find them with a low threshold for edge detection. Fig. 10 shows detection results of guidance sign candidates.

Fig. 10 Detection of guidance sign candidates.

4. Controlling Camera Direction

4.1 Prediction of Sign Position in the Image

After detecting a candidate by the wide-camera for several times (currently, 3), the tele-camera is directed to it for capturing a zoomed-up image. Since it takes some time to switch the camera and direct the tele-camera, we need to predict the position *in the image* of the candidate at the time of the image input by the tele-camera. Assuming that the vehicle moves straight at a constant speed from the time of candidate detection to the end of identification, prediction is performed as follows.

Since a sign moves on a line passing the F.O.E. (focus of expansion), we only need to predict the position of the sign on the line. We here explain how to predict the horizontal position (on the x axis) of the target (see Fig. 11). In the figure, f is the focal length, m_i is the camera position at the time of image input, x_i is the x position of the target at that time. The origin of the x axis is at the center of image, which is the intersection of the image plane and the optical axis originating from the vehicle position. p_i is the distance between camera positions. d and w are the distance to the target from m_1 in the longitudinal and the lateral direction, respectively. We here show how to predict the position x_3 in the third image from two observations x_1 and x_2.

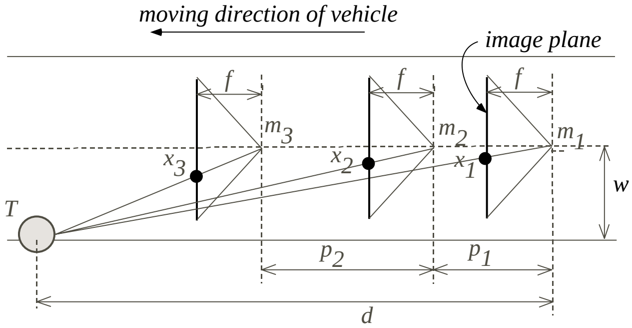


Fig. 11 Predicting target motion.

Let t_i be the time to move by distance p_i. t_1 is calculated by measuring the time of input of the first and the second images. t_2 depends on the time needed for camera control, and can be estimated in advance. Let v be the vehicle speed; we can calculate p_i as $p_i = vt_i$. From the similarity of triangles in the figure, we obtain:

$$\frac{f}{x_1} = \frac{d}{w},$$
$$\frac{f}{x_2} = \frac{d-p_1}{w},$$
$$\frac{f}{x_3} = \frac{d-(p_1+p_2)}{w}.$$

By solving these equations, we obtain:

$$x_3 = \frac{p_1 x_1 x_2}{(p_1+p_2)x_2 - p_2 x_1} = \frac{x_1 x_2}{(1+\alpha)x_2 - \alpha x_1},$$

where $\alpha = p_2/p_1$. Since v is constant, $\alpha = t_2/t_1$. This result shows that the target position can be predicted without vehicle speed v and focal length f. From this predicted x position and the equation of the line passing the F.O.E., the direction of the camera (pan and tilt angles) is calculated.

4.2 Tracking Sign in the Image

If the tele-camera does not capture the target in a sufficiently large size in the image, the camera continuously tracks the target to obtain a larger target image. We use a simple tracking strategy, in which the camera direction is moved by a predetermined angle only when the sign region is not fully included in the image. After finishing the tracking (i.e., after the sign has gone out of the image), the image which captures the sign at the largest size is used for recognition. Fig. 12 shows an example sequence of tracking a guidance sign by the tele-camera.

The current tracking strategy is not reliable enough; fig. 13 shows an example of tracking failure. In this case, although the prediction of the first sign position in the zoomed-up image is correct, the tele-camera failed to catch up the movement of the sign in the image owing to a fast vehicle movement and the constant angular movement of the camera per frame. We have also tested a simple proportional control for tracking; however, since the control cycle time is long due to a slow communication between the camera and the PC, we could not realize a stable tracking. More robust tracking method should be considered in the future.

5. Identification of Signs

5.1 Identification by Pattern Matching

Identification of signs is carried out by a normalized correlation-based pattern matching with a traffic sign image

Fig. 12 Tracking a guidance sign by tele-camera.

Fig. 13 A tracking failure.

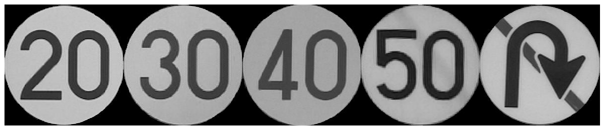
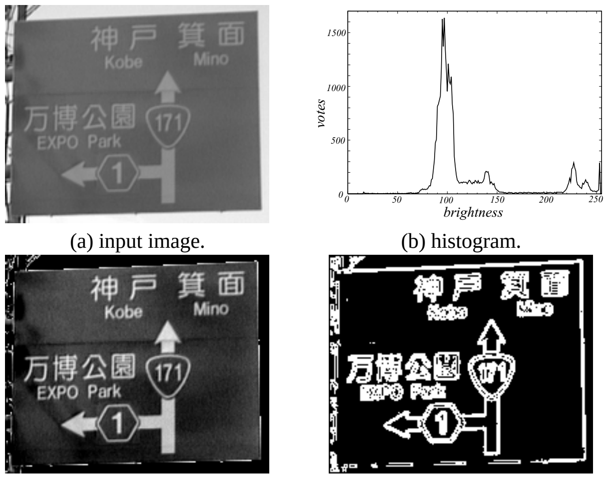


Fig. 14 Template images for speed signs.

(a) detected candidates. (b) recognition result.

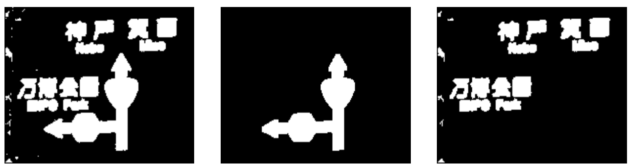
Fig. 15 Recognition of speed signs.



(a) input image. (b) histogram.

(c) transformed image. (d) extracted edges.

Fig. 16 Intensity transformation and edge extraction.



(a) filled. (b) symbol. (c) characters.

Fig. 17 Extracting characters and symbols.

database. Normalized correlation is robust to the change of lighting conditions and is suitable for pattern matching in outdoor scenes. The image processing board we are using (IP-5005 by Hitachi) has the built-in function of calculating normalized correlation between a test image and a template image. For pattern matching, a set of candidate positions is first generated around each screened sign candidate obtained in the previous step. Then the normalized correlation values are calculated and the highest among them is used for evaluation.

Salgian and Ballard [7] proposed to use responses to be obtained by convolving the input image with a set of *steerable filters* in order to localize the position of a traffic sign. This method is faster than the normalized correlation-based method, but it may be weak for the change of lighting conditions.

Fig. 14 shows the template images of speed signs in the database. We use two thresholds for the correlation value to evaluate the appropriateness and the uniqueness of the recognition; that is, if the correlation value of the best candidate is above the first threshold and the ratio of correlation values for the best and the second best candidates is above the second threshold, the best candidate is considered to be identified.

Fig. 15 shows a recognition result of speed signs. Fig. 15(a) is the result of candidate detection in the zoomed-up image. The correct templates (recognition result) are displayed at the bottom-left of the Fig. 15(b).

5.2 Extracting Characters and Symbols

Information on a guidance sign is classified into characters (Japanese Kanji and alphabets) and symbols (indicating road structure). Since information is different from sign to sign, we need to segment such characters and symbols for recognition. This segmentation is performed as follows.

If we properly locate a guidance sign region, characters and symbols may be extracted by a simple thresholding using intensity. This extraction method, however, sometimes fails due to a bad lighting condition such as a large intensity gradation over the sign. So we use edges for segmentation. However, from our experience, an appropriate threshold for extracting edges is still sensitive to the change of lighting condition; thus we first perform an intensity transformation based on the intensity histogram of the image.

Fig. 16(a) shows an input image from which charac-

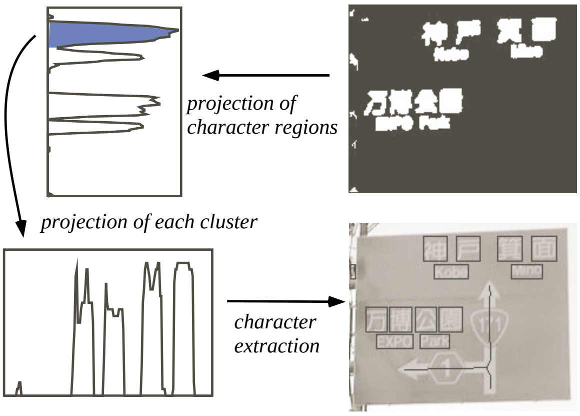


Fig. 18 Result of character extraction.

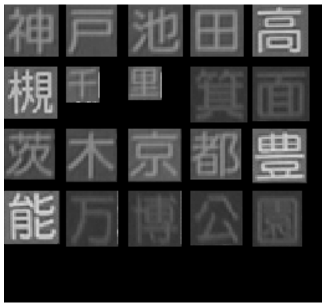


Fig. 19 Template images of characters on guidance signs. These templates are extracted from a videotaped image sequence.

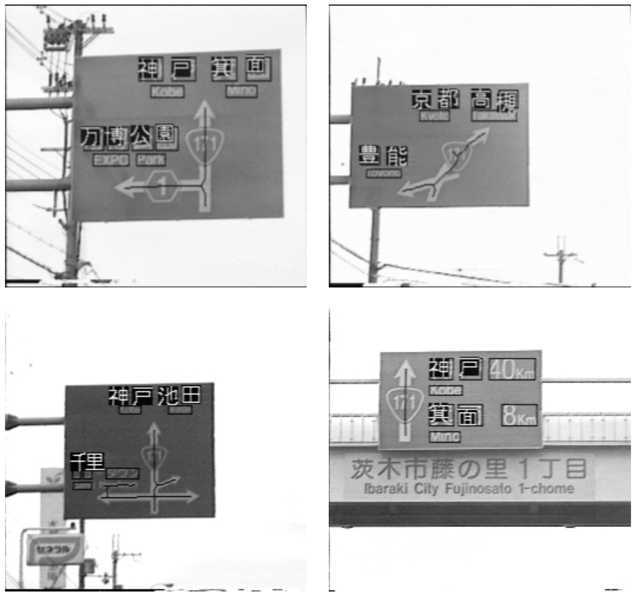


Fig. 20 Four examples of on-line character recognition.

ters and symbols are segmented. Fig. 16(b) is the intensity histogram. An input image usually contains a sky region, to which the high-intensity cluster in the histogram corresponds. The biggest cluster corresponds to the guidance sign. We extract this cluster by first suppressing data in the histogram which are less than a certain threshold (currently, 20) to generate clusters, and then selecting the largest one among them. Let I_{min} and I_{max} be the minimum and the maximum intensity of the cluster. We perform an intensity transformation, from I_{org} to I_{new}, using the following equation:

$$I_{new} = \begin{cases} 255 \cdot \dfrac{I_{org} - I_{min}}{I_{max} - I_{min}} & \text{if } I_{min} \leqq I_{org} \leqq I_{max}, \\ 0 & \text{otherwise.} \end{cases}$$

Fig. 16(c) is the image after transformation and Fig. 16(d) is the extracted edges. Thanks to this intensity transformation, we can extract characters and symbols stably using one threshold in various lighting conditions.

After eliminating edges around the sign and filling small holes, we perform a labeling on the remaining regions. Since a symbol usually forms a large region, we separate symbols from characters using size information (see Fig. 17).

Segmentation of characters is done by *profile projection cuts*, which is a popular method in document image analysis [8]. Since characters are aligned almost horizontally in the image, we first project the character regions onto the vertical axis to calculate the histogram. Next we search the histogram for prominent peaks that indicate vertical position ranges where characters exist. Then for each range, we project the regions onto the horizontal axis to calculate the histogram, and to determine character positions. Fig. 18 shows the result of character extraction.

Identification of each character is performed by the normalized correlation-based matching, as in the case of speed sign recognition. Fig. 19 shows a set of template images of characters. In this identification, we can also use the knowledge of possible names of destination; that is, a pair (or set) of neighboring characters should make sense as the name of an actual destination. For example, “Kyoto” is represented by two Kanji characters “京” (kyo) and “都” (to); thus, if we have found “京”, its right-side character is most likely to be “都”. We have made a database of such destination names. The score of a name is calculated as the multiplication of the correlation values of all characters. The highest-scored name is selected as the final identification result.

Fig. 20 shows four examples of on-line character identification. The identified characters are superimposed on the position of the corresponding character region. Concerning symbols indicating road structure, the skeletons of symbol regions are extracted. In these examples, the intensity (digitized in 8bits) of the blue region of the guidance sign ranges from 72 to 165 and the difference between the maximum and the minimum intensity of a region reaches 35 at most; nevertheless, the character regions are properly extracted thanks to the above-mentioned intensity transformation.

Although we have not tried to recognize alphabets on

Table 1 Processing time

contents of processing		time(msec)
speed signs	binarization & labeling	7
	screening by edges (hough transform)	8
	identification (pattern matching)	20
guidance signs	binarization & labeling	7
	screening by edges (edge segment detection)	5
	character and symbol extraction	650
	character identification (pattern matching)	30
processing time for candidate detection per frame		90-170
control of tele-camera + switch cameras		350

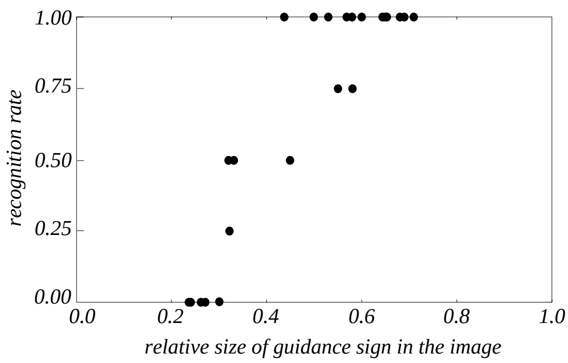


Fig. 21 Relationship between sign size and recognition rate.

Table 2 Result of on-line guidance sign recognition.

# of signs	# of detection	# of successful tracking
17	17 (100%)	4 (23.5%)
# of characters in the successfully-tracked signs		# of identified characters
26		26 (100.0%)

Table 3 Result of on-line speed sign recognition.

# of signs	# of detection	# of identification
71	69 (97.2%)	33 (46.5%)

guidance signs, the same method can basically be applied to alphabet recognition.

6. Implementation and Experiments

The experimental system is composed of a PC (Pentium II (400 MHz), Linux 2.2.13), an image processing board (IP-5005 by Hitachi), and two cameras (EVI-G20 by SONY). This camera has a zoom lens; one camera uses the wide-angle position (the horizontal viewing angle is 45 degrees) and the other uses the telephoto position (the horizontal viewing angle is 15 degrees).

The recognition algorithm mentioned above has been implemented using over 50 built-in functions of IP-5005. Only circular shape detection using edges (see Fig. 7) is performed on the PC. Table 1 summarizes the processing time of each operation.

We first tested the recognition algorithm using the images captured from a manually-taken video sequence. We picked up 24 images, each of which includes one guidance sign in a different size. Fig. 21 illustrates the relationship between the relative size of a sign with respect to the image size (512 x 480) and the recognition rate, which is the ratio of the number of correctly recognized characters to that of characters on the sign. The figure shows that in order to correctly identify all characters on a guidance sign, its relative size in the zoomed-up image should be above around $0.55 \sim 0.6$.

We then conducted on-road experiments using the system. The results shown below were obtained during one round-cruising between two points on Route 171, about 6 kilometers apart from each other; this means the lighting conditions were not constant during the cruising. In addition, the weather often changed between sunny and cloudy, even in cruising in one direction.

Table 2 shows the result of recognizing guidance signs. The signs which were not visible at the time of experiment due to, for example, occlusion by other vehicles are left out of count. In addition, in the table, the tracking is considered successful if the relative size of the sign is larger than 0.55. The success rate is low mainly because the current tracking method by the tele-camera is simple and is not reliable enough, as described above. However, once a good image was taken by the tele-camera after tracking, the system succeeded in recognizing all characters; this results shows the effectiveness of the recognition algorithm.

Table 3 shows the result of recognizing speed signs. Again, the signs which were not visible are not counted. Since, at the time of experiment, we used only one threshold for binarization in the tele-image, if there are several signs in an image (this is a usual case) and if the optimal threshold is different for each sign, the system detects only one sign and fails to detect the others; this is a cause of failure.

Another cause of failure for both speed signs and guidance signs is a long switching time from the wide-camera to the tele-camera. Although we assume the constant velocity during switching and controlling the cameras (see Sec. 4.1), if the velocity perturbation is large in speed and/or direction, this assumption may not be satisfied due to a long switching time and the signs are lost in the first input of the tele-camera. Use of an image processing board which has dual image input ports would be a solution.

7. Recognition of Other Signboards

At the roadside, there are a lot of signboards of shops, supermarkets, restaurants, and so on. These signboards can also be useful landmarks for navigation. People often use such information when they teach a route to others. Since all of such shops are not necessarily on electronic maps and the map information may not be updated timely, a map may not be able to provide necessary landmark information. In such a case, automatic recognition of such signboards could be useful.

Since signboards are, as in the case of traffic signs,

(a) (b)

Fig. 22 Recognition of signboards.

characterized by distinctive colors and shapes with clear characters and symbols, the above-mentioned strategy, which combines candidate detection with the wide-camera and character recognition with the tele-camera, can be applied in a similar way. Fig. 22 shows preliminary experiments of signboard detection in the wide-camera. The signboard in Fig. 22(a) is represented by three rectangles, green, white, and blue ones, aligned vertically. The signboard in Fig. 22(b) is represented by an orange rectangle. We are now investigating character identification of such signboards for more reliable recognition.

8. Conclusions and Discussion

This paper has presented an active vision system that can detect and recognize traffic signs on-line. The multiple-thresholds binarization in conjunction with screening by edges has been effective for robust detection of traffic signs in various lighting conditions. To read the characters on guidance signs, the system captures the large-sized image of a sign using the tele-camera, which is automatically directed to the sign. We succeeded in recognizing characters on guidance signs on-line. Although the recognition algorithm itself seems effective, the current tracking method to capture a guidance sign in a large size by the tele-camera is not reliable enough; this degrades the overall performance of the system. Adopting a more robust tracking method is a future work.

A feature of the system is that the whole system is composed of only off-the-shelf equipments. Especially, the recognition algorithm is implemented by intensively using the built-in functions of the general-purpose image processing board. This shows that widely available hardwares are suitable for building blocks of really working systems.

Various traffic information systems based on communication technologies [10] can be used for providing the driver with useful information for navigation. The deployment of such systems, however, entails the development of infrastructures. On the other hand, vision-based information gathering systems like the one presented in this paper are basically stand-alone and seem less costly. Since both types of system could provide different kinds of information with different reliability, a cooperative use of them might be a future direction.

References

[1] S. Azami, S. Katahara, and M. Aoki. Route Guidance Sign Identification Using 2-D Structural Description. In *Proceedings of the 1996 IEEE Symp. on Intelligent Vehicles*, pp. 153–158, 1996.

[2] I. Masaki. Machine-Vision Systems for Intelligent Transportation Systems. *IEEE Intelligent Systems*, Vol. 13, No. 6, pp. 24–31, 1998.

[3] U. Franke et al. Autonomous Driving Goes Downtown. *IEEE Intelligent Systems & Their Applications*, Vol. 13, No. 6, pp. 40–48, 1998.

[4] L. Priese et al. Traffic Sign Recognition Based on Color Image Evaluation. In *Proceedings of the 1993 IEEE Symp. on Intelligent Vehicles*, pp. 95–100, 1993.

[5] H. Kawakami, T. Ikeda, and J. Tajima. Road Sign Detection Using Information about Color and Shape. Technical Report PRU92-29, IEICE, 1992. (in Japanese).

[6] G. Piccioli et al. Robust Method for Road Sign Detection and Recognition. *Image and Vision Computing*, Vol. 14, pp. 209–223, 1996.

[7] G. Salgian and D.H. Ballard. Visual Routines for Autonomous Driving. In *Proceedings of the 6th Int. Conf. on Computer Vision*, pp. 876–882, 1998.

[8] Y.Y. Tang, S.-W. Lee, and C.Y. Suen. Automatic Document Processing: A Survey. *Pattern Recognition*, Vol. 29, No. 12, pp. 1931–1952, 1996.

[9] E. Ichihara, H. Takao, and Y. Ohta. Visual Assistance for Drivers Using Roadside Cameras. In *Proceedings of the 1999 IEEE Int. Conf. on Intelligent Transportation Systems*, pp. 170–175, Tokyo, Japan, October 1999.

[10] Yilin Zhao. *Vehicle Location and Navigation Systems*. Artech House Inc., Boston, 1997.