A Hybrid Vision Method for Autonomous Guided Vehicle Navigation

LU Jian IKEDA Hiroyasu HAMAJIMA Kyoko

(*National Institute of Industrial Safety,* 1-4-6, *Umezono, Kiyose, Tokyo* 204 − 0024, *Japan*)

(E-mail: {lujian,ikeda,hamajima}@s.jniosh.go.jp)

Abstract For the autonomous guided vehicle (AGV) used mainly in unfixed work fields, a machine vision method was proposed for the navigation system, in which a series of navigation-signs are placed along the travel route. The navigation system detects and recognizes these signs, and accordingly informs the travel control system. In order for the navigation to have balanced ability of 1) covering a large area and 2) recognizing details of the sign, the proposed vision method was designed to be a hybrid one, using both the stereo vision and the traditional 2D template matching. The former implemented a coarse recognition function for above 1), and the later implemented a fine recognition function for above 2). The results from the coarse recognition were used in the fine recognition for the gaze control to input suitable 2D image of the signs. Experiments on a prototype system show the feasibility of the proposed hybrid method in achieving the objective specifications for a typical AGV.

Key words AGV navigation, machine vision, disparity image, stereo vision, template matching

1 Introduction

The navigation system for AGV (also called mobile robot in this paper) are mainly of two types: tag-based and lane-based. In the tag-based navigation, an AGV is navigated by a series of tags, for example, magnetic tags, placed under the floor along the planned travel route, and the AGV is equipped with the sensor to detect these tags. In the lane-based navigation, the planned route is indicated by the lanes drawn on the floor, and the AGV has the image sensor to simply recognize the lanes. In the case that the planned task and the planned route are relatively fixed, both these two types are effective. However, in many industry applications, the route and the task for AGVs are not fixed, such as in the case for the construction industry. We proposed a navigation system for AGVs that can be planned to travel along some changeable route[1]. In the navigation system, instead of the magnetic tags or lanes mentioned above, a series of image signs, called navigation-signs (Fig. 1), are placed along the route. By change the order and the types of signs in the series, the route can be planned and adjusted flexibly. A vision system is required to recognize navigation-signs met in the front way of the AGV and then to inform the travel control system accordingly. In a practical design for a prototype navigation system, the objective specifications are mainly as follows. 1) Travel speed: 1m/s; 2) Sensing frequency: 5Hz; 3) Size of navigation-sign: 30cm×30cm; 4) Distance of navigation-sign: 1.6m∼5m.

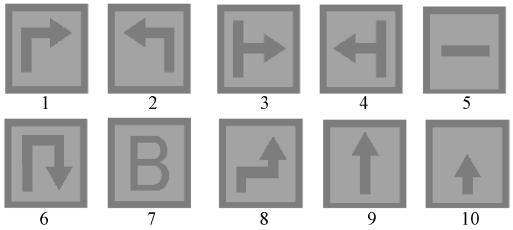


Fig. 1 10 types of navigation-signs

Usually, it is difficult for an individual vision method to achieve the above specifications all in the same time, as it has to monitor a relatively large area, and also to recognize the detail of the a sign image, at a size of 30*cm* level. Our solution here is a hybrid vision method consisted of two functions, *i.e.*, the coarse recognition function, and the fine recognition function. In the following of this paper, after the hybrid method is overviewed in Section 2, the coarse recognition function and the fine recognition function are described in Section 3 and 4, respectively. Finally, the concluding remarks are given in Section 5.

Received December 21, 2005; in revised form April 3, 2006

2 Machine vision and the hybrid method

2.1 Machine vision in industrial applications

Machine vision has been studied and applied for industrial application as an intelligent sensing method for decades. For instance, pattern matching method for 2D image is very effective and widely used in quality inspection on production lines. The reason for this success is mainly because that the sensing objects are within the limited types, usually of expectable and fixed sizes and shapes. However, the sensing objects in case of AGV are of more types and of various shapes, and the range of the sensing distance is also comparatively wide. As the sensing method for such kind of wide space, the advanced and complicated technology, for instance, stereo vision, is desired. On the other hand, although the mathematical theory for stereo vision has been studied for a long period of time[2], its implementation in a realistic application is not so easy, because of the problems from many aspects, such as calibration and synchronization. As the result, development researches on stereo vision remained mainly on developing the special hardware to solve these problems. In addition, the intensive computation also limits the application of stereo vision. As the recent progress in hardware, computation power and stereo vision camera become commercially available. Therefore it is currently possible to pay our main attention to develop algorithm and software for stereo vision applications, and to evaluate them based on commercially available hardware.

2.2 Overview of the hybrid vision method

As mentioned above, the proposed hybrid vision method consists of two functions, the coarse recognition function and the fine recognition function. For our prototype system, Fig. 2 is the hardware structure and Fig. 3 is the implemented hybrid approach.

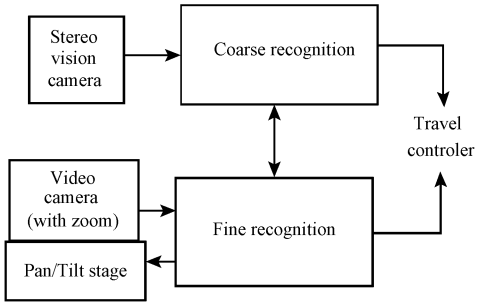


Fig. 2 Hardware Structure for the hybrid vision method

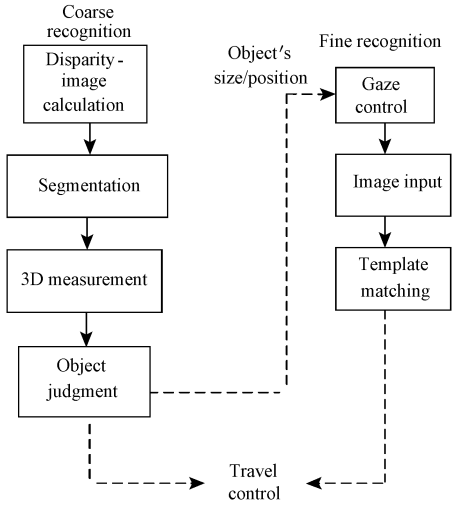


Fig. 3 Approach for the hybrid vision method

The coarse recognition function is performed by using the disparity image obtained from a stereo vision camera. It determines roughly the 3D position and the size of important objects, including the signs. If required, they will be used in fine recognition to control the gaze operations, including pan, tilt and zoom of the video camera.

The fine recognition function recognize the specific type or meaning of a navigation-sign, by using the image obtained from the video camera. As shown in the right side in Fig. 3, the gaze control, *i.e.*, the controls for pan/till and zoom, are performed according to the information obtained by coarse recognition. Then the image input is performed and the template matching is calculated for the input image to recognize the specific type of the sign.

3 Coarse recognition

As shown in Fig. 2 and Fig. 3 the coarse recognition is based on disparity image. Previously, the utilization of disparity data has been studied for transportation robots and automobiles to detect

the front obstacles. However, these applications are limited to the existence detection of the obstacle, and rarely the position measurement is contained. As the set of range data or disparity data, the disparity image is the low resolution one so that its width and height are in the extent of several decade pixels, and the judgment for obstacles could be done directly according to an individual pixel and its transformation to the point in 3D space[3].

In our research, however, not only the existence of objects, including obstacles and signs, but also their 3D position and size are the objective of detection and measurement. Consequently, a stereo vision camera Digiclops, made by Pointgrey Inc., with relatively higher resolution is used. Here the resolution used is 320×240. Since it is not realistic to do the 3D conversion for all pixels in such a resolution, the efficient technique for pixel classification, *i.e.*, the segmentation of disparity image, is required. Although some approaches dedicated to segmenting high-resolution disparity images are proposed[4,5], they are mainly for aerial photograph, unsuitable for the real-time processing task such as obstacle detection. For this purpose, a segmentation method of using region-expanding and the convex decomposition for object extraction was proposed in our previous research[6]. Here for the complete description of the coarse recognition, we will briefly describe the segmentation method in the following Section 3.2, and other steps shown on the left side in Fig. 3 are also described below.

3.1 Calculation of disparity image

A disparity image consists of pixels calculated according to the triangulation, the basic principle for stereo vision. The simplest explanation to stereo vision is for the case that two camera-lens with the same focal length are arranged on one image plane and have their optic axis parallel with each other. As shown in Fig. 4, if a point in the space is expressed as $P(X, Y, Z)$ in world coordinate system, and its projection to the two image are (x_l,y_l) and (x_r,y_r), respectively, the following relational expressions are established.

$$X = \frac{b(x_l + x_r)}{2d} \tag{1}$$

$$Y = \frac{b(y_l + y_r)}{2d} \tag{2}$$

$$Z = \frac{bf}{d} \tag{3}$$

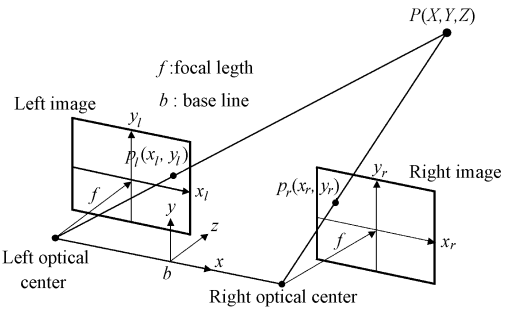


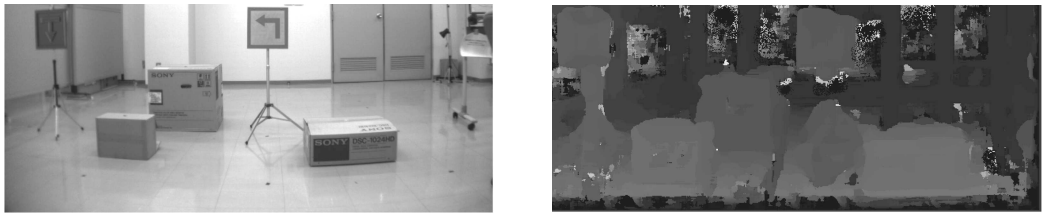
Fig. 4 Principle of stereo vision

Here, f is the focal length, b is the baseline, and d is the disparity:

$$d = x_l - x_r \tag{4}$$

If f and b are already known, the 3D position in (X, Y, Z) can be calculated from their projection positions on the left image and the right image. Especially, the range Z can be calculated from disparity d. Fig. 5 is an example about disparity image. In Fig. 5(b), each pixel takes value d calculated according to (4) and expressed in the brightness proportional to its disparity value. Usually, relatively bright (white) pixels in disparity image corresponding to relatively nearer objects. For example, in the lower half of Fig. 5(b), there are some large blobs containing relatively bright pixels, corresponding to the boxes put on the floor. However, some pixels appear in white spot, *i.e.*, the most brightest pixels,

are due to noise. The problems from this kind of noise are solved by the segmentation processing in later description.



(a) Reference image (b) Disparity image

Fig. 5 Example reference image and disparity image

3.2 Segmentation and 3D measurement

The approaches for segmentation and 3D measurement are described as follows.

1) Segmentation: The pixel classification, or the image segmentation, is carried out for the disparity image in two steps. i) Among the neighbor pixels, the pixels having the near disparity value, *i.e.*, the disparity difference of pixels is less than a threshold, are classified into the same region (region expansion). ii) The regions are further split into convex regions, each of which usually corresponds to only one object(convex decomposition).

2) 3D measurement: For each region obtained at step (1), coordinate (X, Y, Z) corresponding to the pixel at the 2D gravity center of the region is calculated, and is used as the gravity position of corresponding object. Furthermore, $\mathrm{Min}\{X\}$, $\mathrm{Max}\{X\}$, $\mathrm{Min}\{Y\}$ and $\mathrm{Max}\{Y\}$ are calculated for all points (X, Y, Z) corresponding to the boundary pixels of each region, and the size, concretely the width $Xsize$ and the height $Ysize$, of the object corresponding to the region are calculated by $Xsize = \max\{X\} - \min\{X\}$ and $Ysize = \max\{Y\} - \min\{Y\}$.

An example of the resulted segmentation is given in Fig. 6. Note that the segmentation result is super-imposed on the scene instead on the disparity image for easy understanding, although the segmentation is actually performed on disparity image. In addition, though there are more than hundreds of candidate regions actually obtained after the image segmentation step, for simplifying sake here, only the regions larger than 1000 *pixels* are shown.

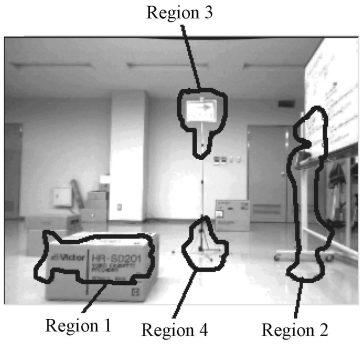


Fig. 6 Example of the extracted objects

3.3 Object judgment based on 3D measurement

The next step is object judgment. This is implemented by evaluating the feature distance between the extracted objects and the expected target objects, *i.e.*, the distance between the measured values and the standard values. Here the standard value is obtained beforehand based on a learning process of statistic pattern recognition. In the prototype system, the target object is the navigation-sign.

3.3.1 The standard features

In the prototype system, the navigation-sign is a board painted with sign image for AGV navigation. The size of the board is $0.3m \times 0.3m$, and the board is mounted on a stand such that its center has the height of $0.85m$ from the floor, the same height with the optical center of the stereo vision camera.

Fig. 7 is an example navigation-sign. If we define feature space consisting of three feature vectors Y, Xsize, and Ysize, the feature of navigation-sign can be expressed as the followings.

$$Y = 0 \tag{5}$$

$$Xsize = 0.3 \tag{6}$$

$$Ysize = 0.3 \tag{7}$$

In addition, the following evaluation function is used for judging whether the object is a sign.

$$W_t = Y^2 + (Xsize - 0.3)^2 + (Ysize - 0.3)^2 \tag{8}$$

From (8), we know that W_t is the square value of distance between extracted object and the sign within the feature space. Originally, the right-hand of (8) should be the square root. For less computation cost, however, the square value is used instead of the square root.

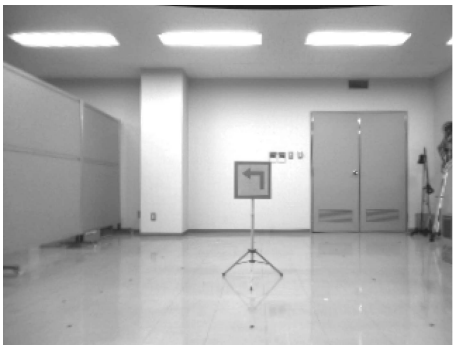
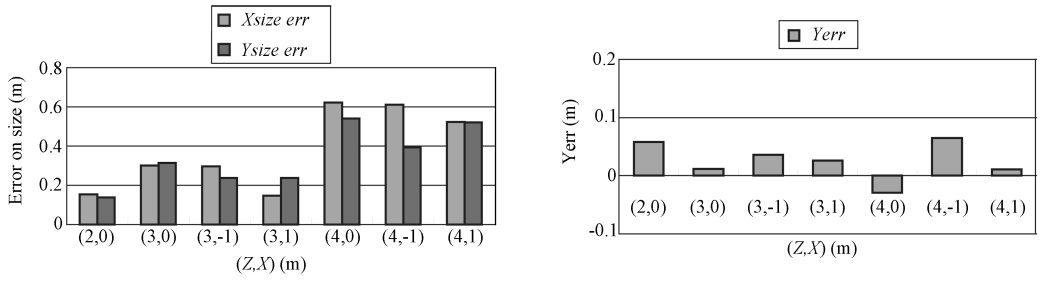


Fig. 7 Example of navigation-sign

3.3.2 The calibration to the standard feature

When the evaluation value for W_t in (8) is 0, the extracted object is sure to be judged as the sign. However, it is only the ideal expression for the features of the sign. For the real application environment, because of the erroneous in 3D measurement X, Y, Z and on $Xsize$ and $Ysize$, it is necessary to calibrate the values related to the standard feature of sign.

Therefore, the error between the measurement and the actual value are measured by performing 3D measurement experiment on the signs located in some typical positions, with the depths range from $2m$ to $4m$. The results of these experiments are shown in Fig. 8. Here, the horizontal axis expresses the actual locations of signs, and the vertical axis are errors on the size (Fig. 8(a)) and on the position (Fig. 8(b)). As shown in the same figure, for all experiments with the depth in the range $2m \cdot 4m$, the errors on $Xsize$ and on $Ysize$ are in the range (0.14, 0.62), and the central value of the error is 0.38. As the calibration plan, instead of 0.3, 0.3+0.38 may be used as the standard value for $Xsize$ and $Ysize$ of sign. However, we believe that the measure and control of the robot travel should be performed in



(a) Reference image (b) Disparity image

Fig. 8 Measurment error to navigation-sign in our experiment

higher accuracy for short distance. For the depth ranging from $2m$ to $3m$, the errors on $Xsize$ and on $Ysize$ are in the range (0.14, 0.31), with the central value of 0.23. Accordingly, 0.3+0.23 is used as the calibration to 0.3 in (8) as the standard value of $Xsize/Ysize$ related to sign. On the other hand, errors on Y are relatively smaller than $Xsize$ and $Ysize$, no calibration is considered for Y. Accordingly, after the calibration, the evaluation function for sign judgment is as follows.

$$W_t = Y^2 + (Xsize - 0.53)^2 + (Ysize - 0.53)^2 \tag{9}$$

3.3.3 Judgment by using threshold

In practice, for the judgment by (9), rather than using $W_t = 0$ directly, a threshold is used, otherwise, the tolerance to the judgment error will be too small. In order to obtain this threshold, in the experiments similar with in Section 3.3.2, Wt calculated by equation (9) is shown in Fig. 9. As shown in Fig. 9, W_t is at most 0.3. As the result, if $W_t < 0.3$ for an extracted object, it is judged as sign, otherwise, it is judged as an obstacle.

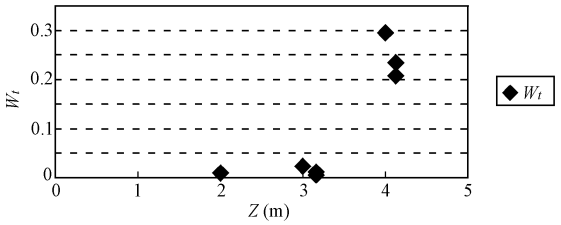


Fig. 9 Distribution of evaluation function W_t for standard navigation-signs

3.4 The Experiment and the result

The evaluation experiment on the coarse recognition is performed in the environment like in Fig. 7. The results are shown in Table 1. Each line in the table corresponds to the measurement to a sign placed at some typical positions expressed by (Z, X). At all positions, the sign is set at a fixed height of $0.85m$ (from the floor), the same height with the optic center of the stereo vision camera, such that $Y = 0$. It is noticed that all value of W_t in (9) for the candidate regions corresponding to signs are smaller than 0.3, meaning that a sign can be successfully recognized according to discussion in Section 3.3.3. As for the speed, the coarse recognition program was run at $4Hz$ on a PC with a $700MHz$ Pentium III CPU. An old and slow PC was used just due to license problems involved in some software tools we used. The specification of $5Hz$ is possible to be reached if update PCs are used.

Table 1 Example results from 3D measurement for detected signs

(Z, X)	W_t	X	Y	Z	$Xsize$	$Ysize$
(2,0)	0.008	0.073	−0.02	2.003	0.459	0.422
(2,-1)	0.019	−0.91	0.009	1.997	0.37	0.455
(3,0)	0.009	0.096	0.025	3.12	0.575	0.549
(3,-1)	0.059	−0.95	0.009	3.172	0.743	0.488
(3,1)	0.023	1.089	0.013	2.959	0.583	0.627
(4,0)	0.137	0.152	0.047	4.308	0.821	0.678
(4,-1)	0.229	−1	0.1	4.448	0.74	0.902
(4,1)	0.204	1.213	0.064	4.33	0.831	0.801

4 Fine recognition

This section describes the main steps of the approach for the fine recognition. Among the three steps shown in right side in Fig. 3, the gaze control and the template matching are described mainly, since the image input is almost a hardware operation.

4.1 Gaze control

The first step of the fine recognition is the gaze control. Fig. 10 shows the flowchart for the gaze control .As shown in Fig. 10, gaze control includes controls for pan/tilt and for zoom. In this section, the description about the principles for pan/tilt control and for zoom control is given, respectively.

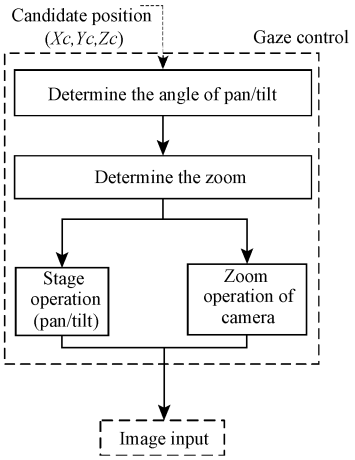


Fig. 10 Flowchart for the gaze control

4.1.1 Determination for the pan/tilt angle

As shown in Fig. 2, the video camera used in the fine recognition is installed on a camera stage with pan/tilt functions. In order for the visual field of the camera to catch the image center of the extracted object, as the first step in the gaze control, the pan and the tilt angles are to be determined according to the position information of the candidate object. If the position information is (xc, yc, zc) for the coordinate system shown in Fig. 11, the pan angle should be

$$\theta = -\tan(xc/zc) \tag{10}$$

and the tilt angle should be

$$\phi = \tan^{-1}(yc/\sqrt{xc^2 + zc^2}) \tag{11}$$

4.1.2 Determination for the zoom value

After the stage is adjusted to the suitable angle to catch the image of the object, in the next step, in order for the camera to input the image with enough large size of the navigation-sign image, the zoom of the camera should be set to make the camera have a suitable image angle. As shown in Fig. 12, in addition to the image angle B corresponding to object′s size S, a margin angle Ma is required. As the result, the objective image angle A is determined as follows.

$$A = B + 2Ma \tag{12}$$

$$B = 2\tan^{-1}(S/2L) \tag{13}$$

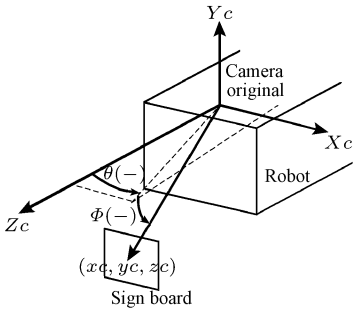


Fig. 11 Camera coordinate system

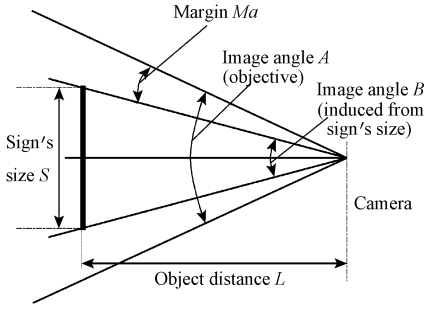


Fig. 12 Sign′s size and the objective zoom value for the image angle

At present, the video camera we used is FCB-IX47 made by SONY. In order to set the image angle to A for the camera of this model, the zoom value V, *i.e.*, the instruction for zoom operation in

percentage, has the following relationship[2)] with A:

$$V = 33.02622 - 0.49466 \times A + 0.00814 \times A^2 \tag{14}$$

4.2 Template matching

In addition to the core processing of template matching , this step includes also the pre-processing, *i.e.*, colour conversion and binarization. The colour conversion is to convert the colour input image in RGB to the colour image in HSL. A colour image in HSL is easier to extract colour and to perform the colour analysis. An image after binarization is easier to perform template matching. The colour conversion, binarization and template matching are well known image processing operation, and the details about them can be found in reference[2]. Although the detail is ignored, Table 2 is the correlation coefficient between the signs, with the numbers in the first row and the first column representing the signs in Fig. 1. We explain the followings.

1) Correlation coefficients in Table 2 are values in percentage.

2) The correlation coefficient less than 10% is not listed.

3) The correlation coefficient between different sign is at most 42%. By setting a threshold larger than 42%, if the correlation coefficient is well larger than the threshold when matching a candidate sign image with a template sign image, the candidate sign will be judged as of the same type with the template image.

Table 2 Correlation coefficient (%) between signs

	1	2	3	4	5	6	7	8	9	10
1		14	15			15				
2										
3				12	37	18		15		
4					42					
5										
6							27	18		
7										
8										
9										33
10										

4.3 The time required for the fine recognition

In this section, by using measured data, we examine some problems involved in required time for the fine recognition. Since the size of sign is fixed to $30cm \times 30cm$, and the distant sign (within $5m$) can be zoomed large enough to be used in template matching, the main purpose here is to examine whether the time(speed) specifications are practical to be achieved.

As shown in the right side in Fig. 3, the time to perform fine recognition is the total time for the following three steps, *i.e.*, gaze control, image input, and image processing (including pre-process and template matching). According to the measured data on our prototype system, Fig. 13 shows the time required for the fine recognition.

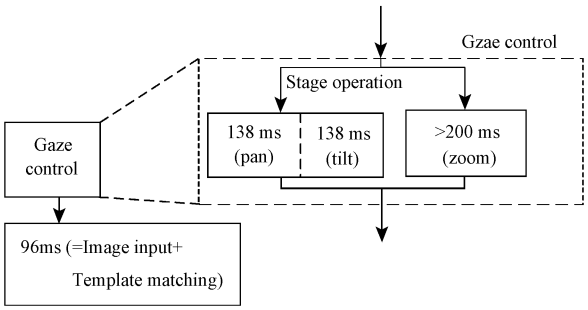


Fig. 13 The time required for fine recognition

[2)] The relationship is obtained by calibration and computation according to the specification in "Instruction Manual" for SONY FCB-IX47.

It is noticed in Fig. 13 that the time for fine recognition will be $96ms$ (about $100ms$) plus the time for gaze control. Because calculation time for determining the value for gaze control can be ignored, compared with the mechanical operation time for zoom and pan/tilt, and because the operations for zoom and for pan/tilt can performed in parallel, the time for gaze control will be the larger one among the time for zoom and for pan/tilt.

Table 3 shows the experiment result to estimate the time for zoom operation. Both the Table 3(a) and 3(b) are measured from a series of zoom operations, performed for the initial object distance of $5000mm$ to the last object distance of $1500mm$, with the difference in that the object distance change between the consequent step is $500mm$ for Table 3(a) and $1000mm$ for Table 3(b) respectively. Table 3(c) is measured for one single zoom operation of changing the object distance from $1600mm$ to $5000mm$. Table 4 shows the experiment result to estimate the time for pan/tilt operations. From the Table 3 and 4, it is noticed that

1) The time for a zoom operation is comparatively long: the time for any single zoom operation is longer than $200ms$.

2) The time for a pan or tilt operation is estimated less than $100ms$ in average, considering the case that the angle for a pan operation or for a tilt operation is usually less than 10 *degree*.

Therefore, as the bottleneck for the fine recognition, the zoom operation should be avoid if possible.

Table 3 Measured time for zoom operation

(a) 500 mm steps

Step	Objective distance (mm)	Zoom value(%)	Time(ms)
0	5000	73.3	
1	4500	72.3	260
2	4000	71	261
3	3500	67.5	310
4	3000	62.3	351
5	2500	55.7	361
6	2000	46.6	380
7	1500	33.4	470
Average		342(ms)	

(b) 1000mm steps

Step	Objective distance (mm)	Zoom value(%)	Time(ms)
0	5000	73.3	
1	4000	71	281
2	3000	62.3	381
3	2000	46.6	521
Average		394(ms)	

(c) One step

Distance (mm)	time(ms)
1600 → 5000	852

Table 4 Measured speed for pan/tilt operation

Operation	Parameter	Maximum(deg./sec)	Minimum(deg./sec)
Pan	Speed	147.14	2.93
	Start speed	51.38	
Tilt	Speed	145.7	2.98
	Start speed	57.14	

4.4 Reducing the time for gaze control

To avoid frequent zoom operation, the zoom value should be fixed for some certain period of time, *i.e.*, one operation of zoom setting should cover a wide range of object distance. As the result, the image taken on the sign board may not be in a suitable size for template matching. In addition, the image taken at different distance with a fixed zoom value may be defocused, such that the focus operation may be required, although it is not preferred that the additional time be taken for the focus operation. In the following, we examine whether the fine recognition is possible for fixed zoom and for the defocusing condition.

4.4.1 Fine recognition under large image angle

The large image angle (large visual field) of camera, corresponds to the low magnification of zoom. In this case, the sign board takes small area in the whole image taken by the camera. In order to investigate the results of template matching operations when the sign image become small, for a sign at $5000mm$ distance, experiments are performed for different zoom values, corresponding to different sizes of sign image. The experiment results are shown in Table 5. It is noticed that

1) The largest size is the one in the first line, where the height of sign is 100 *pixels*, and the smallest size is in the last line where the height of sign is 26 *pixels*.

2) While the height of the sign image changed from 100 to 26, about 4:1, the correlation coefficient changes from 87% down to 66%. Though there is about 20% down in correlation coefficient, the template matching is still possible (Fig. 14).

Fig. 14 Example for small sign image (zoom: 30%, distance: 5000mm)

3) Considering the distance range in the specification is 1600mm to 5000mm, a ratio of about 3:1 in distance, compared with the ratio of 4:1 in size change mentioned above, it is expected that for different distance, fine recognition is possible even the zoom value is fixed, say to 20% as shown in the last line in Table 5.

Table 5 Correlation coefficients for different image magnifications

Zoom value(%)	Height of sign image(pixels)	Coefficiency (%)
73	100	87
60	66	84
50	50	87
40	39	77
30	32	66
20	26	66

distance 5000mm
zoom value=73% is used as the reference condition

Table 6 Correlation coefficients under defocusing condition

Focus value (%)	Focused distance (mm)	Correlation coefficient (%)
12.7	5000	84
32	2000	85
48	1000	66

distance 5000mm

4.4.2 Fine recognition under defocusing condition

In order to investigate the results of template matching operations under defocusing conditions, for a sign at 5000mm distance, experiments are performed for different focus distances of 5000mm, 2000mm, and 1000mm. The experiment results are shown in Table 6. It is noticed that even the focus is set to 1000mm, the correlation coefficient is 66%. The correlation coefficient 66%, still significant enough for fine recognition (Fig. 15).

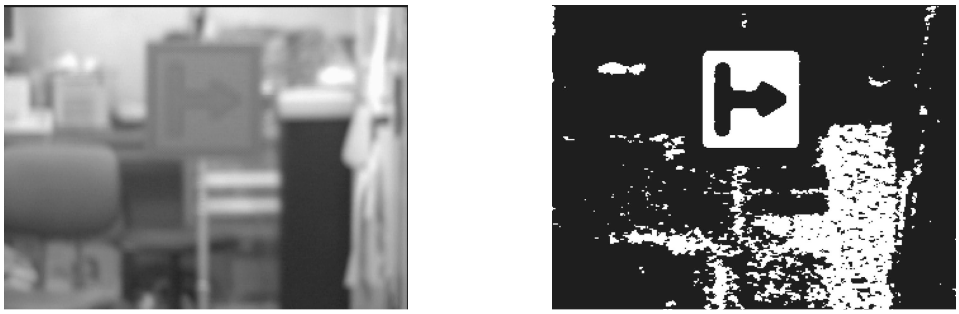


Fig. 15 Example of defocusing condition (focus value: 48%)

4.5 Conclusions about the fine recognition

The discussions in Section 4 are concluded as follows.

1) Since any single zoom operation takes time longer than $200ms$, the zoom operation should be avoid as possible.

2) Even in the condition that the zoom is set to have a wide image angle, such that the sign image becomes small, the fine recognition(template matching) is still possible.

3) Even under the defocusing condition, the fine recognition is still possible within the distance range of $1600mm$ to $5000mm$.

4) By fixing the zoom value for a wide range of distance to avoid frequent zoom operation, it is possible to sustain the specification of sensing frequency of $5Hz$.

5 Concluding remarks

For AGVs used mainly in unfixed work fields, a hybrid vision method that combines the stereo vision and the traditional 2D vision was described. It provided the balanced ability of covering a large area and recognizing details of the navigation sign. Experiments on a prototype system show the feasibility of the proposed hybrid method in achieving the objective specifications for a typical AGV. We also notice that the traditional 2D machine vision method would be still applicable to a wide area case like in AGV navigation, as far as the 2D image used is suitably selected, such as through the gaze control assisted by the stereo vision. Finally, we believe that the approach here for object extraction and judgment based on the disparity image will be applicable to objects other than the navigation-sign used as the example in this paper, as far as the features of the target object are suitably established through a learning process.

References

1 Lu J, Yasuda K, Nakagawa S, Ikeda H. A navigation vision system for industrial mobile robots. In: Proceedings of the 33rd International Symposium on Robotics. Stockholm, Sweden: International Federation of Robotics, 2002. 7∼11

2 Chen C H, Pau L F, Wang P S P. Handbook of Pattern Recognition and Computer Vision. 2nd Edition, Singapore: World Scientific Press, 1999

3 Willamson T, Thorpe C. A trinocular stereo system for highway obstacle detection. In: Proceedings of IEEE International Conference on Robot and Automation. Detroit, Michigan, USA: IEEE Press, 1999. 2267∼2273

4 Koster K, Spann M. MIR: An approach to robust clustering-application to range image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(5): 430∼444

5 Bellon O R P, Direne A I, Silva L. Edge detection to guide range image segmentation by clustering techniques. In: Proceedings of IEEE International Conference on Image Processing, Kobe, Japan: IEEE Press, 1999. 725∼729

6 Lu J, Hamajima K, Ishihara K. A hybrid machine vision method for autonomous guided vehicles. In: Proceedings of SPIE-IS and T Electronic Imaging. Santaa, Clara, Californin, USA: International Society for Optical Engineering, 2003. 118∼127

LU Jian Reveived his Ph. D. degree from Toyohashi University of Technology, Japan, in 1997. He is currently a researcher in National Institute of Industrial Safety, Japan. His research interests include image processing, pattern recognition, and computer application in industrial control. He is a member of The Institute of Electronics, Information and Communication Engineer, and a member of Japan Society of Ergonomics.

IKEDA Hiroyasu Received his master in electrical engineering from Chuo University , Japan, in 1985. He is currently a chief researcher in National Institute of Industrial Safety, Japan. His research interests include robotics, and safety control. He is a member of the Japan Society of Mechanical Engineers, and a member of the Robotics Society of Japan.

HAMAJIAMA Kyoko Received her Ph. D. degree from Tokyo Denki University, Japan, in 1999. She is currently a researcher in National Institute of Industrial Safety, Japan. Her research interests include image processing, robotics, and computer application in industrial control. She is a member of The Japan Society of Mechanical Engineers.