PAPER Special Section on Machine Vision and its Applications

Precise Top View Image Generation without Global Metric Information

Hiroshi KANO[†a)], ***Member***, **Keisuke ASARI**[†], ***Nonmember***, **Yohei ISHII**[†], ***and*** **Hitoshi HONGO**[†], ***Members***

SUMMARY We describe a practical and precise calibration method for generating a top view image that is transformed so that a planar object such as the road can be observed from a direction perpendicular to its surface. The geometric relation between the input and output images is described by a 3×3 homography matrix. Conventional methods use large planar calibration patterns to achieve precise transformations. The proposed method uses much smaller element patterns that are placed in arbitrary positions within the view of the camera. One of the patterns is used to obtain an initial homography. Then, the information from all of the patterns is used by a non-linear optimization scheme to reach a global optimum homography. The experiment done to evaluate the method showed that the precision of the proposed method is comparable to that of the conventional method where a large calibration pattern is used, making it more practical for automotive applications.

key words: *top view image, homography, calibration*

1. Introduction

The number of rear view backup cameras being installed in automobiles is increasing. The main use of such cameras is to show the blind spots of vehicles, however, it is difficult to correctly perceive distances by simply viewing the direct input image because of the perspective effect of the camera. A top view image transformation, shown in Fig. 1, was proposed to eliminate the perspective effect to help drivers park [1]. It is also used to detect lane markers on the road [2].

To generate a top view image, the camera must be calibrated. This is usually a process of obtaining the camera's parameters such as its focal length, pixel size, mounting location, and mounting angles. When a top view image is generated, a homography matrix [3] is computed from these parameters. The homography matrix is a 3×3 planar projective transformation matrix, transforming the input image into a top view image. In this paper, calibration means the process of computing the 3×3 matrix.

To compute the matrix, there must be at least four sets of corresponding points shared by the input and the transformed images. However, in general, to transform the image precisely, many corresponding points should be laid out evenly across the entire image. This gives rise to a practical problem when conventional calibration methods are applied to an automobile. Figure 2 shows a typical image obtained by a rear view camera. The road surface that the camera covers is large because the field of view is wide and deep. The calibration pattern used in this view must have reference points associated with 2D absolute coordinates. As a result, the calibration pattern becomes larger, for example 10 m × 10 m, and must contain precise reference points.

We propose a practical calibration method to solve the problem that does not require a global 2D coordinate in the calibration area. Instead, it uses only the shape information of several smaller square patterns to compute the homogra-

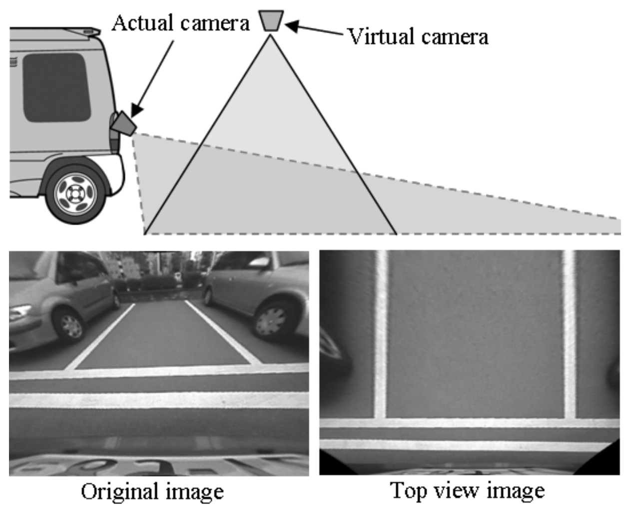


Fig. 1 Top view image generation.

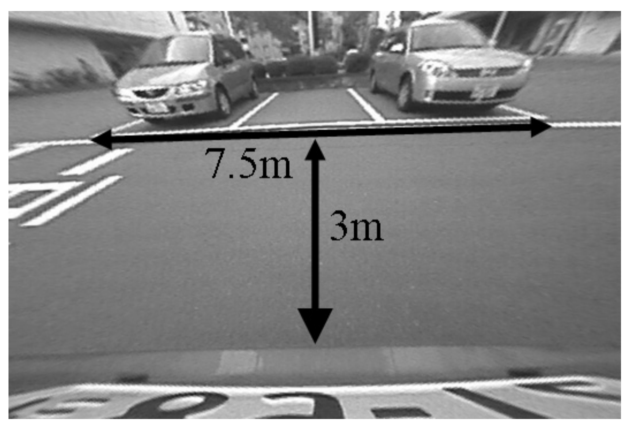


Fig. 2 Image of rear view camera.

Manuscript received October 26, 2007.
Manuscript revised January 25, 2008.

[†]The authors are with the Digital Technology Research Center, R&D H.Q., SANYO Electoric Co., Ltd., Daito-shi, 574–8534 Japan.
a) E-mail: Hiroshi.Kano@sanyo.com

DOI: 10.1093/ietisy/e91–d.7.1893

Copyright © 2008 The Institute of Electronics, Information and Communication Engineers

phy matrix. It is enough to place them roughly so that they cover a large area of a camera view as a whole. The proposed method is useful especially when a rear view camera is installed in a location such as an automotive body shop where it is difficult to keep a wide space for camera calibration.

The rest of the paper is organized as follows: in Sect. 2, we describe conventional calibration methods, Sect. 3 describes the basic idea of the proposed method, Sect. 4 presents results of both a simulation and an experiment using actual images, and in Sect. 5, we conclude the paper.

2. Camera Model, Calibration, and Homography

2.1 Camera Model and Homography

From the point of view of geometrical characteristics, a pinhole camera is a device that transforms a 3D world coordinate into a 2D image coordinate [3]. This relation is described by Eq. (1). Here, matrix $\mathbf{P}_{3\times4}$ is a 3×4 matrix called a projective transformation matrix. $[X, Y, Z, 1]^t$ is a homogeneous 3D world coordinate point in the scene, and $[u, v, 1]^t$ is the corresponding homogeneous 2D image coordinate point. There is equality in the homogeneous coordinate in (1).

$$[u, v, 1]^t \simeq \mathbf{P}_{3\times4} \cdot [X, Y, Z, 1]^t \tag{1}$$

Matrix $\mathbf{P}_{3\times4}$ is decomposed into Eq. (2). Here, $\mathbf{K}_{3\times3}$ is a 3×3 matrix, the elements of which are computed from internal camera parameters such as the focal length and pixel size. $\mathbf{R}_{3\times3}$ is a 3×3 rotation matrix, $\mathbf{t}_{3\times1}$ is a 3×1 translation vector, and $[\mathbf{R}_{3\times3}|\mathbf{t}_{3\times1}]$ is a 3×4 matrix concatenating $\mathbf{R}_{3\times3}$ and $\mathbf{t}_{3\times1}$; The camera position and angles relative to the 3D world coordinate frame. These are the external camera parameters.

$$\mathbf{P}_{3\times4} = \mathbf{K}_{3\times3} \cdot [\mathbf{R}_{3\times3}|\mathbf{t}_{3\times1}] \tag{2}$$

When there is a plane in the scene, a 3D coordinate of a point on the plane is described by $x \cdot \mathbf{i} + y \cdot \mathbf{j} + \mathbf{d}$, where (x, y) is the 2D coordinate on the plane, $\mathbf{i}$ and $\mathbf{j}$ are the directional unit vectors along the axes, and $\mathbf{d}$ is the point of origin. Therefore, the homogeneous 3D coordinate of the point on the plane is described by Eq. (3), where $\mathbf{Q}$ is a 4×3 matrix. By substituting (3) for (1), we obtain Eq. (4). Equation (4) indicates that the relation between the 2D coordinates of a plane in the scene and the image plane is described by a 3×3 matrix $\mathbf{H}$ in the homogeneous sense. The 3×3 matrix is called the homography matrix.

$$\begin{aligned}[X, Y, Z, 1]^t &= \begin{pmatrix} x\cdot\mathbf{i} + y\cdot\mathbf{j} + \mathbf{d} \\ 1 \end{pmatrix} \\ &= \mathbf{Q}_{4\times3} \cdot [x, y, 1]^t \end{aligned} \tag{3}$$

$$\begin{aligned}[u, v, 1]^t &\simeq \mathbf{P}_{3\times4} \cdot \mathbf{Q}_{4\times3} \cdot [x, y, 1]^t \\ &= \mathbf{H}_{3\times3} \cdot [x, y, 1]^t \end{aligned} \tag{4}$$

2.2 Calibration of Homography Matrix

One elaborate method of obtaining the homography matrix is to follow the steps, described in 2.1, one by one. For example, by using Zhang's multiple-plane calibration [4], it is possible to obtain Eqs. (1) and (2), and then by using the equation of a plane corresponding to the road, we can obtain homography that transforms the image plane into the road plane. However, this method is impractical for automotive applications because the calibration pattern is large and has to be observed from several different directions.

Another similar method is to obtain the internal camera parameters beforehand and then use the design values of the external camera parameters to determine Eqs. (1), (3) and (4). However, this method does not work well, because the design values of the external parameters are not sufficiently accurate. As a result, the top view images generated by this method frequently become distorted, as illustrated in Fig. 3.

The two methods described above are indirect in the sense that the camera is calibrated as a 3D-to-2D projective device first, and then the 2D-to-2D projective homography is computed afterwards. As a matter fact, no 3D-to-2D projective characteristics are needed to determine the homography.

In a direct method, a large calibration pattern is placed on the road. On the surface, there are some reference points with absolute 2D coordinate values. Four sets of corresponding points between the reference points of the input image and the calibration pattern are sufficient to compute the matrix. However, in general, to obtain a precise matrix, many corresponding points should be laid out evenly across the entire image. The precision of this method is illustrated in Fig. 4. Though this method is stable, it requires a large calibration pattern. This direct method is hereafter referred to as the conventional method.

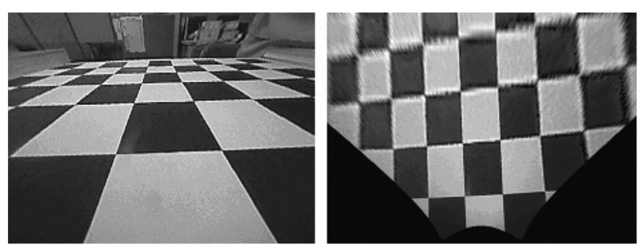


Fig. 3 Left: input image, right: top view image using design values of external parameters.

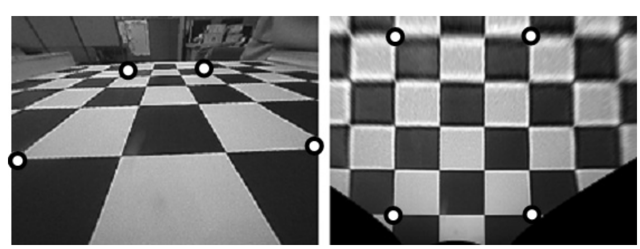


Fig. 4 Left: input image, right: top view image by direct method using a large pattern (circles are reference points used for calibration).

3. Proposed Method without Global Metric Information

3.1 Basic Idea

It is thought that the calibration pattern must cover a large area of the input image for the calibration to be precise. The conventional method uses a large planar pattern, such as a checkerboard. The proposed method uses smaller elemental patterns and places them at arbitrary positions on the road's surface, as illustrated in Fig. 5. It is the same as the conventional method in that the patterns cover a major part of the input image as a whole. However, it is different in that the patterns are not precisely aligned with each other.

Not all element patterns are necessarily the same, but the dimensions of all of them must be known in advance. One of them, which is selected in order to estimate an initial homography, must have at least four reference points. In this paper, all of the element patterns are the same and are square in shape. They are placed evenly, but not necessarily aligned precisely. Starting from an initial homography computed using one of the element patterns, the tentative homography matrix is optimized by iteratively minimizing the error function, taking into account all patterns.

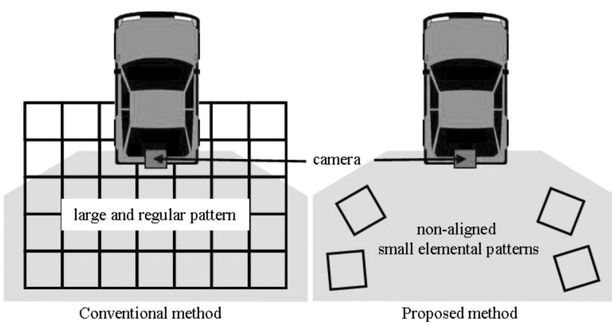


Fig. 5 Conventional method and proposed method.

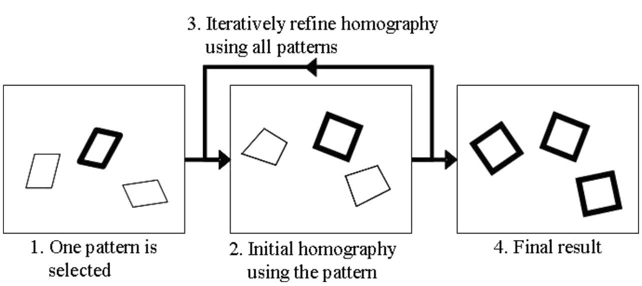


Fig. 6 Homography optimization procedure.

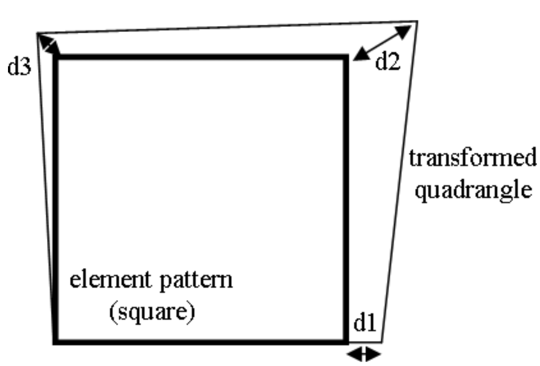


Fig. 7 Error between element pattern and transformed quadrangle.

3.2 Optimization Procedure

The procedure for optimizing the homography is shown in Fig. 6. In the first step, one of the element patterns is selected. In the second step, a homography matrix that transforms the selected pattern into a fixed size square is computed. Then, this matrix is used as the initial value of an iterative process in the third step, where the matrix is modified iteratively using an evaluation function obtained from all patterns.

The evaluation function is defined by Eq. (5). Here dn is the distance between two corresponding corners of an element pattern and a pattern transformed by a tentative homography, as illustrated in Fig. 7. The two quadrangles are laid out as shown. There are three dn's for each pattern and all dn's over all patterns are summed up to obtain one evaluation value. Powell's quadratically convergent method is used as the algorithm to minimize the evaluation function.

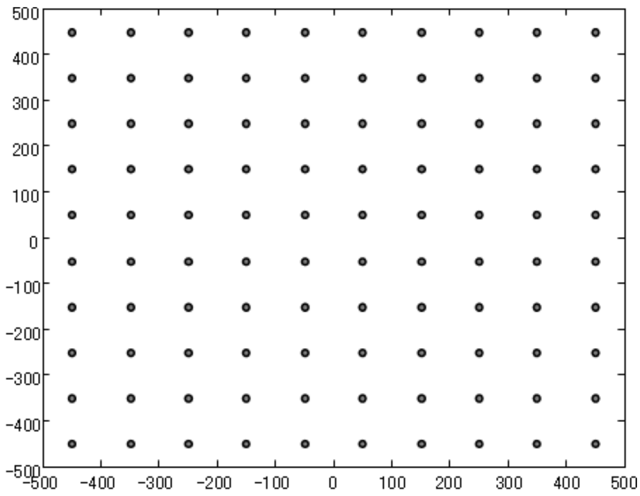
$$D = \sum_{all\ patterns} \left(\sum_{each\ pattern} |dn|^2 \right) \tag{5}$$

4. Experimental Results

4.1 Accuracy Evaluation by Simulation

We conducted a simulation experiment as follows. It was assumed that there were 100 points on a plane surface. They were aligned on an equally-spaced 10×10, 100 mm square grid, as shown in Fig. 8 (a). These reference points were observed using a camera. It was assumed that a VGA image such as the one shown in Fig. 8 (b) was obtained. Random noise of average 0 and standard deviation σ was added to the points in the camera image to obtain the observed points. The small dots in Fig. 8 (b) are the points corresponding to the reference points, and the small crosses are the observed points that fluctuate around the dots. (The small crosses in the figure do not represent the actual computed positions, the actual positions are actually much closer to the dots. They are shown only for ease of understanding).

In the experimental evaluation of the proposed method's accuracy, the vertices of several squares were used as reference points to compute the homography matrix. Point correspondences were used for the conventional method. Figure 9 shows the reference points and the squares used for these evaluations. Sixteen evenly distributed points on the plane were used for the conventional method (A). We evaluated four cases for the proposed method: two squares



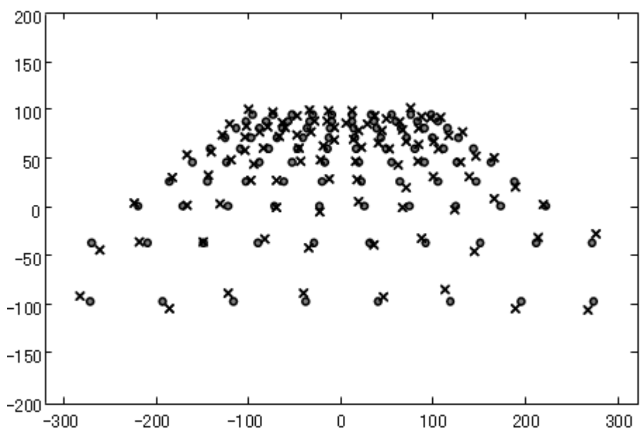


Fig. 8 Simulation data.

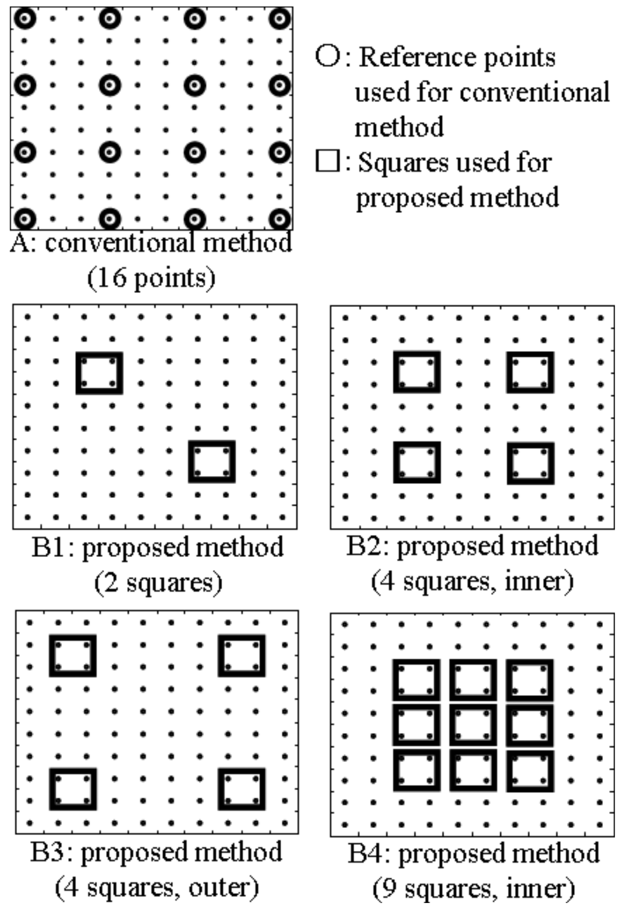


Fig. 9 Accuracy evaluation by simulation: five cases.

Table 1 Results of accuracy evaluation by simulation.

		A	B1	B2	B3	B4
$\sigma =$ 0.1 pixel	Ave.	0.44	2.79	0.70	0.36	0.43
	SD	0.37	2.90	0.71	0.31	0.41
$\sigma =$ 0.5 pixel	Ave.	2.01	10.01	2.69	2.27	2.39
	SD	1.34	10.40	2.31	2.20	1.87
$\sigma =$ 1.5 pixel	Ave.	5.16	30.19	10.94	5.66	7.10
	SD	4.04	28.93	9.64	4.42	6.75

Unit:mm

(B1), four inner squares (B2), four outer squares (B3) and nine inner squares (B4). We considered that the precision of homography obtained by A was the standard that should be achieved by the proposed method.

Table 1 summarizes the results. In this table, σ is the standard deviation of random noise added to the observed points, Ave. is the average of the distances between the reference points mapped onto the image coordinate using the computed homography and the observed points, and SD is their standard deviation. Table 1 shows that B1 (two squares) is the worst and B2 (four inner squares) is not as good as A. B3 is comparable to A and B4 is a little bit worse than A. We also found that the more squares we used for the proposed method, the more precise the obtained homography became. From this experiment, we found that placement of four square patterns in the outer area is comparable to the conventional method using 16 reference points evenly spaced within the calibration area.

4.2 Accuracy Evaluation Using Actual Images

We conducted an experiment to test the effectiveness of the proposed method. We used the CCA-BC200 manufactured by SANYO Electric Co., Ltd., as the rear view camera. The horizontal view angle of the camera is 135 degrees. The lens distortion was corrected in advance as shown in Fig. 10. We prepared two patterns for the accuracy evaluation. One

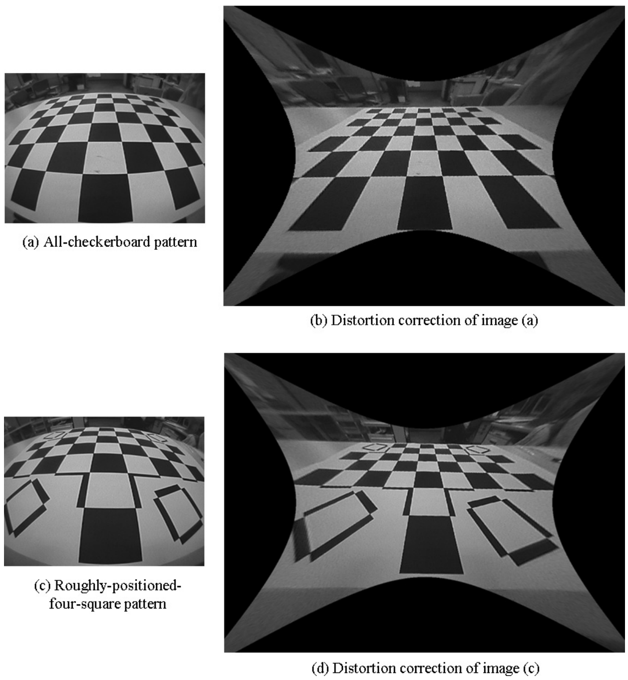


Fig. 10 Distortion correction of wide angle camera.

was an all-checkerboard pattern. The other was a roughly-placed-four-square pattern. The four-square pattern also had a checkerboad on it for the accuracy evaluation.

As shown in (AA) of Fig. 11, for the accuracy evaluation of the conventional method, ten intersecting points were selected from the all-checkerboard pattern. The ten pairs of input and transformed image coordinates were carefully extracted. For each evaluation of the proposed method using the all-checkerboard pattern, a different number of squares on the checkerboard was used, as shown in (BB1) to (BB3) in Fig. 11. The number used varied between one, two and four. We also conducted the accuracy evaluation of the proposed method using the roughly-placed-four-square pattern. For this pattern, the four squares surrounding the checkerboard were used for the calibration and the checkerboard was used only for the accuracy evaluation.

The calibration results are shown in Table 2. An error is the distance between a coordinate value computed by the estimated homograpy and the ideal coordinate value of each intersecting point, as shown by the grid lines in Fig. 11. Max. is the maximum value of the errors of the all intersecting points, Min. is the minimum value of the errors, Ave. is the average, and SD is the standard deviation. Area in Table 2 is the area covered by the calibration pattern for each case. The unit value of Area is the area of one square in the checkerboard.

Table 2 shows that, in terms of SD, AA is 2.92, BB3 is 3.10 and CC is 2.89, and in terms of Ave., AA is 3.94, BB3 is 4.73 and CC is 3.18. Therefore, it can be seen that the accuracy of BB3 and CC, both of them using the proposed method where four squares placed at similar positions, is comparable to AA. In terms of Area, AA is 54. BB3 and

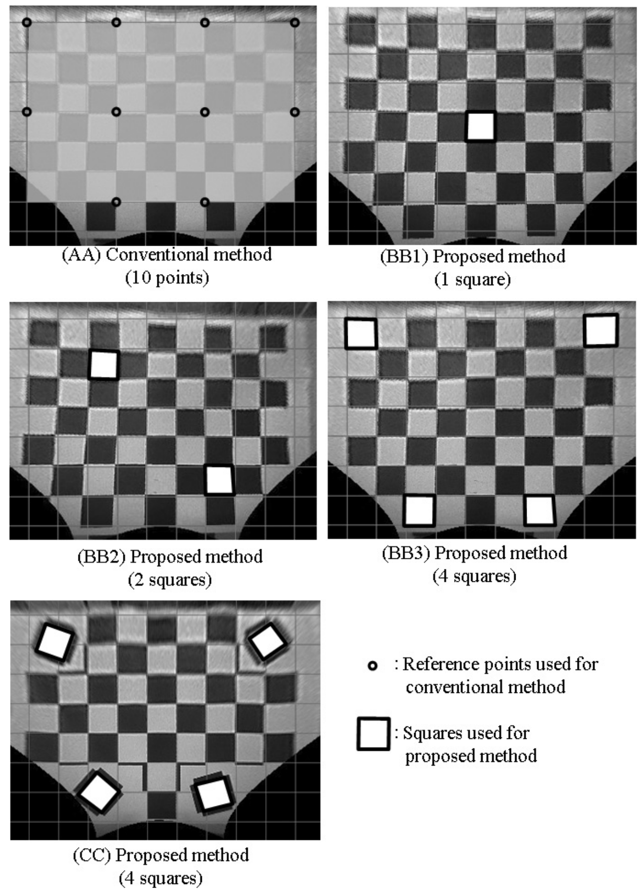


Fig. 11 Accuracy evaluation in real environment: four cases.

Table 2 Experimental accuracy results.

Error	AA	BB1	BB2	BB3	CC
Max.	13.98	39.48	33.81	14.09	12.56
Min.	0.45	0.70	1.10	0.37	0.16
Ave.	3.94	14.41	13.23	4.73	3.18
SD	2.92	8.11	8.79	3.10	2.89
Area	54	1	2	4	4

Error: pixel, Area: unit square

CC are only 4. This means that when these methods are used for the rear view camera of an automobile, the cost of installing the calibration pattern on the road surface is much smaller for BB3 and CC than for AA.

4.3 Application to Actual Rear View Camera

The proposed method using two square patterns was applied to a rear view camera installed onto a vehicle. The pattern is $1.5\,\mathrm{m} \times 1.5\,\mathrm{m}$. The generated image is shown in Fig. 12. In the top view image, the lower portion around the two square patterns are precise but a little distortion is found in the upper portion of the image. This is because we used only two

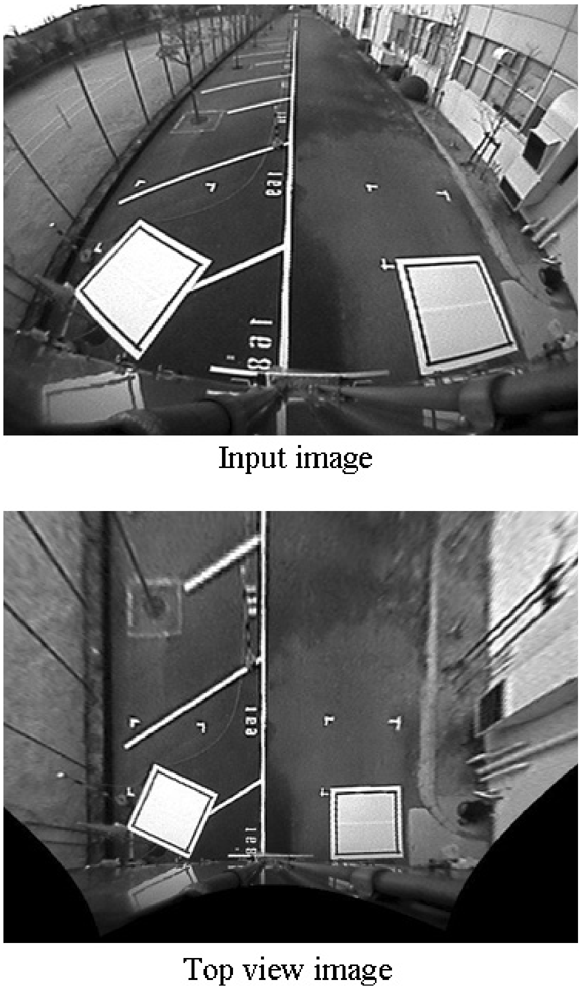


Fig. 12 Top view generation using camera mounted on a vehicle.

square patterns placed close to the camera. However, the calibration process itself was easy and the result shows that a practical top view image was obtained.

5. Conclusion

A practical method for generating a top view image based on homograpy was proposed. In this method, several small patterns are spaced evenly so that they cover the entire view of a camera. However, they are not necessarily aligned precisely with each other. Therefore global metric information is not necessary for this proposed method. The experimental results showed that the procedure was more practical than the conventional method but had the same precison.

In future work, we will attempt to reduce the computation cost of the proposed method for the embedded implementation into an automobile.

References

[1] H. Shimizu, "Overhead view parking support systems," Proc. 11th World Congress on ITE, 2004.
[2] A. Takahashi, Y. Ninomiya, M. Ohta, M. Nishida, and N. Yoshikawa, "Image processing technology for rear view camera (1): Development of lane detection system," R&D Review of Toyota CRDL, vol.38, no.2, pp.31–36, 2003.
[3] R. Hartley and A. Zisserman, Multiple view geometry in computer vision, Second ed., p.87–93, Cambridge Univ. Press, 2003.
[4] Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. Pattern Anal. Mach. Intell., vol.22, no.11, pp.1330–1334, 2000.
[5] W. Press, S. Teukolsky, W. Vettering, and B. Flannery, Numerical Recipes in C, Second ed., Cambridge University Press, 1992.

Hiroshi Kano received his B.E., M.E. and D.E. degrees in information engineering from Kyoto University in 1982, 1984 and 1998 respectiviley. He is currently a senior manager at the Digital Technology Research Center in SANYO Electric Co., Ltd. His research interest includes image processing, computer vision, and their applications to camera equipment such as digital cameras and on-vehicle cameras.

Keisuke Asari received his B.E. and M.E. degrees in information engineering from Okayama University in 2004 and 2006 respectiviley. He is currently a researcher at the Digital Technology Research Center in SANYO Electric Co., Ltd. His research interests include image processing and its application to automobile use.

Yohei Ishii received his B.Ag. and M.Ag. degrees in agricultural engineering from Kyoto University in 1996. He is currently a senior researcher at the Digital Technology Research Center in SANYO Electric Co., Ltd. His research interests include image processing and its application to automobile use.

Hitoshi Hongo received his D.E. degree in information science from Gifu University in 1999. He is currently a manager at the Digital Technology Research Center in SANYO Electric Co., Ltd. His research interests include image processing, pattern recognition and their applications to automotive use.