PAPER

Reliability of 3-D Reconstruction by Stereo Vision

Yasushi KANAZAWA† *and* Kenichi KANATANI††, *Members*

SUMMARY Theoretically, corresponding pairs of feature points between two stereo images can determine their 3-D locations uniquely by triangulation. In the presence of noise, however, corresponding feature points may not satisfy the epipolar equation exactly, so we must first correct the corresponding pairs so as to satisfy the epipolar equation. In this paper, we present an optimal correction method based on a statistical model of image noise. Our method allows us to evaluate the magnitude of image noise *a posteriori* and compute the covariance matrix of each of the reconstructed 3-D points. We demonstrate the effectiveness of our method by doing numerical simulation and real-image experiments.

key words: *stereo vision, reliability of 3-D reconstruction, model of image noise, epipolar equation, statistical optimization, noise estimation*

1. Introduction

Stereo vision is one of the most fundamental means of 3-D sensing from images and is widely used as a visual sensor for autonomous navigation of robots [1], [8]. In the past, the study of stereo vision has mainly focused on the *correspondence detection* between the two images. In fact, detecting correspondences is a very difficult task to automate efficiently, and many practical techniques have been proposed for it [1]. However various other issues arise when we reconstruct 3-D from detected correspondences. First of all, the 3-D reconstruction should be accurate. Hence, we must maximize the accuracy by using an optimization technique based on the statistical characteristic of image noise. At the same time, *the reliability of the reconstructed 3-D must be evaluated* [6], [7]. If the errors involved in the reconstructed 3-D cannot be estimated, robots cannot take appropriate actions to archive given tasks effectively. This paper presents a new method for reconstructing 3-D by stereo vision in a statistically optimal way and evaluating the reliability of the reconstruction in quantitative terms.

Theoretically, we can easily determine 3-D from corresponding pairs of feature points between the two stereo images by triangulation. In the presence of noise, however, the lines of sight determined by corresponding feature points do not necessarily intersect in the 3-D scene. A simple method often used is computing the nearest point from the corresponding lines of sight by least squares. By this method, however, it is very difficult to estimate the reliability of the reconstructed position theoretically. In this paper, we first correct each corresponding pair of feature points so that their lines of sight intersect in the scene. We present an optimal scheme for this correction based on a statistical model of image noise. Our method allows us to evaluate the magnitude of image noise *a posteriori* and compute the covariance matrix of each of the reconstructed 3-D points. Finally, we demonstrate the effectiveness of our method by doing numerical simulation and real-image experiments.

Manuscript received February 2, 1995.

†The author is with the Department of Information and Computer Engineering, Gunma College of Technology, Maebashi-shi, 371 Japan.

††The author is with the Department of Computer Science, Gunma University, Kiryu-shi, 376 Japan.

2. Epipolar Geometry of Stereo Vision

We take the first camera as a reference coordinate system and place the second camera in a position obtained by translating the first camera by vector $\boldsymbol{h}$ and rotating it around the center of the lens by matrix $\boldsymbol{R}$. We call $\{\boldsymbol{h}, \boldsymbol{R}\}$ the *motion* (or *stereo*) *parameters.* The two cameras may have different focal lengths f and f'.

Let (x, y) be the image coordinates of a feature point projected onto the image plane of the first camera, and (x', y') those for the second camera. We use the following three-dimensional vectors to represent them:

$$\boldsymbol{x} = \begin{pmatrix} x/f \\ y/f \\ 1 \end{pmatrix}, \quad \boldsymbol{x}' = \begin{pmatrix} x'/f' \\ y'/f' \\ 1 \end{pmatrix}. \tag{1}$$

As shown in Fig. 1, vectors $\boldsymbol{x}$, $\boldsymbol{R}\boldsymbol{x}'$, and $\boldsymbol{h}$ are coplanar, so they satisfy the following *epipolar equation* [1], [2], [8]:

$$|\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{R}\boldsymbol{x}'| = 0. \tag{2}$$

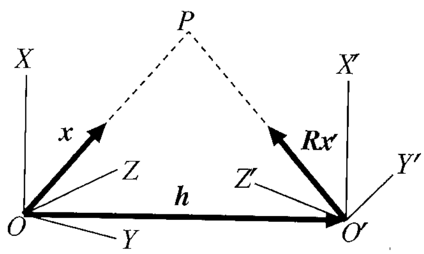


Fig. 1 The camera model and the coordinates systems.

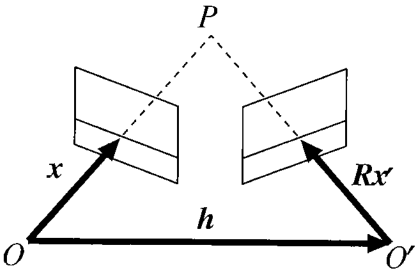


Fig. 2 Epipolar constraint.

In this paper, $|\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}|$ denotes the scalar triple product of vectors $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$. The *essential matrix* [1], [2], [8] is defined by

$$\boldsymbol{G} = \boldsymbol{h} \times \boldsymbol{R}, \tag{3}$$

where the right-hand side is the matrix defined by the vector product of vector $\boldsymbol{h}$ and each column of $\boldsymbol{R}$. The epipolar equation (2) can be written as

$$(\boldsymbol{x}, \boldsymbol{G}\boldsymbol{x}') = 0. \tag{4}$$

Throughout this paper, $(\boldsymbol{a}, \boldsymbol{b})$ denotes the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$.

For a fixed value of $\boldsymbol{x}'$, Eq. (4) is an equation of a line on the image plane of the first camera. This line is a projection of the line of sight that starts from the origin O' and passes through $\boldsymbol{x}'$; it is called the *epipolar* of $\boldsymbol{x}'$. Similarly, Eq. (4) is an equation of a line on the image plane of the second camera for a fixed value of $\boldsymbol{x}$; it is called the *epipolar* of $\boldsymbol{x}$. Put differently, the epipolars are the intersections of the image planes with the plane defined by the two camera origins O and O' and the feature point P that we are observing (Fig. 2). Hence, the projection of a feature point on one image plane must be on the epipolar of the corresponding feature point on the other image plane. This relationship is called the *epipolar constraint* [1], [2], [8].

3. Statistical Model of Image Noise

The epipolar equation implies that for a given point in one image, we need to search only along the epipolar in the other image for the corresponding point. In the presence of noise, however, the epipolar constraint is not exactly satisfied, so we must also search the neighborhood of the epipolar. Correspondence pairs found in this way do not necessarily satisfy the epipolar constraint. In this paper, we assume such a circumstance.

Let $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{x}}'$ be corresponding points in the absence of image noise. In the presence of noise, we observe

$$\boldsymbol{x} = \bar{\boldsymbol{x}} + \Delta\boldsymbol{x}, \quad \boldsymbol{x}' = \bar{\boldsymbol{x}}' + \Delta\boldsymbol{x}'. \tag{5}$$

We regard the noise terms $\Delta\boldsymbol{x}$ and $\Delta\boldsymbol{x}'$ as random variables of means $\boldsymbol{0}$ and covariance matrices

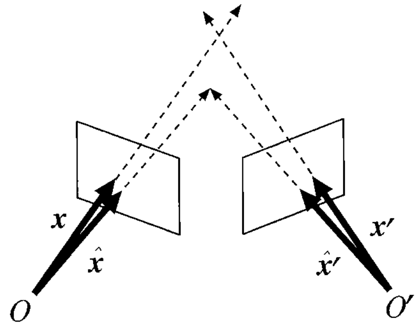


Fig. 3 Correction of feature points.

$$V[\boldsymbol{x}] = E[\Delta\boldsymbol{x}\Delta\boldsymbol{x}^\top], \quad V[\boldsymbol{x}'] = E[\Delta\boldsymbol{x}'\Delta\boldsymbol{x}'^\top], \tag{6}$$

where the symbol $E[\,\cdot\,]$ denotes expectation, and the superscript $\top$ denotes transpose.

The absolute magnitude of the image noise is very difficult to estimate *a priori*, but its geometric characteristics, such as uniformity and isotropy, can be easily predicted. Hence, we assume that the covariance matrices are known *only up to scale* and write

$$V[\boldsymbol{x}] = \epsilon^2 V_0[\boldsymbol{x}], \quad V[\boldsymbol{x}'] = \epsilon^2 V_0[\boldsymbol{x}']. \tag{7}$$

We call the constant ϵ the *noise level*, which indicates the average magnitude of the image noise and is assumed unknown. The matrices $V_0[\boldsymbol{x}]$ and $V_0[\boldsymbol{x}']$ are called the *normalized covariance matrices* and assumed known.

4. Optimal Correction of Feature Points

Since detected corresponding points $\boldsymbol{x}$ and $\boldsymbol{x}'$ do not necessarily satisfy the epipolar equation (4), we correct $\boldsymbol{x}$ and $\boldsymbol{x}'$ in the form

$$\hat{\boldsymbol{x}} = \boldsymbol{x} - \Delta\boldsymbol{x}, \quad \hat{\boldsymbol{x}}' = \boldsymbol{x}' - \Delta\boldsymbol{x}', \tag{8}$$

so as to satisfy Eq. (4) (Fig. 3):

$$(\boldsymbol{x} - \Delta\boldsymbol{x}, \boldsymbol{G}(\boldsymbol{x}' - \Delta\boldsymbol{x}')) = 0. \tag{9}$$

Taking a linear approximation, we obtain

$$(\boldsymbol{G}\boldsymbol{x}', \Delta\boldsymbol{x}) + (\boldsymbol{G}^\top\boldsymbol{x}, \Delta\boldsymbol{x}') = (\boldsymbol{x}, \boldsymbol{G}\boldsymbol{x}'). \tag{10}$$

This equation describes a hyperplane in the six-dimensional space of $\Delta\boldsymbol{x}$ and $\Delta\boldsymbol{x}'$; we are to find the most likely values of $\Delta\boldsymbol{x}$ and $\Delta\boldsymbol{x}'$ on this hyperplane. If the distribution of the image noise is Gaussian, the most likely values of $\Delta\boldsymbol{x}$ and $\Delta\boldsymbol{x}'$ are given by the point on this hyperplane nearest from origin O measured in the *Mahalanobis distance*. We can write this condition as the following optimization:

$$J = (\Delta\boldsymbol{x}, V_0[\boldsymbol{x}]^-\Delta\boldsymbol{x}) + (\Delta\boldsymbol{x}', V_0[\boldsymbol{x}']^-\Delta\boldsymbol{x}') \to \min. \tag{11}$$

Here, $V_0[\boldsymbol{x}]^-$ and $V_0[\boldsymbol{x}']^-$ are the (Moore-Penrose) generalized inverses of $V_0[\boldsymbol{x}]$ and $V_0[\boldsymbol{x}']$, respectively. In statistical terms, this minimization is equivalent to *maximum likelihood estimation*[4].

The solution of the minimization (11) under the constraint (10) is given as follows (we omit the derivation [4]):

$$\Delta \boldsymbol{x} = \frac{(\boldsymbol{x}, \boldsymbol{G}\boldsymbol{x}')V_0[\boldsymbol{x}]\boldsymbol{G}\boldsymbol{x}'}{(\boldsymbol{x}', \boldsymbol{G}^\top V_0[\boldsymbol{x}]\boldsymbol{G}\boldsymbol{x}') + (\boldsymbol{x}, \boldsymbol{G}V_0[\boldsymbol{x}']\boldsymbol{G}^\top \boldsymbol{x})},$$
$$\Delta \boldsymbol{x}' = \frac{(\boldsymbol{x}, \boldsymbol{G}\boldsymbol{x}')V_0[\boldsymbol{x}']\boldsymbol{G}^\top \boldsymbol{x}}{(\boldsymbol{x}', \boldsymbol{G}^\top V_0[\boldsymbol{x}]\boldsymbol{G}\boldsymbol{x}') + (\boldsymbol{x}, \boldsymbol{G}V_0[\boldsymbol{x}']\boldsymbol{G}^\top \boldsymbol{x})}. \tag{12}$$

Since Eq. (10) is obtained by a linear approximation, the corrected values $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}'$ may not strictly satisfy the epipolar equation (4). Hence, we repeat this correction by letting $\boldsymbol{x} \leftarrow \hat{\boldsymbol{x}}$ and $\boldsymbol{x}' \leftarrow \hat{\boldsymbol{x}}'$ until Eq. (4) is sufficiently satisfied.

The corrected values $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}'$ are random variables, because they are computed from data. So let us write $\hat{\boldsymbol{x}} = \bar{\boldsymbol{x}} + \Delta\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}' = \bar{\boldsymbol{x}}' + \Delta\hat{\boldsymbol{x}}'$. The normalized *a posteriori* covariance matrices $V_0[\hat{\boldsymbol{x}}] = E[\Delta\hat{\boldsymbol{x}}\Delta\hat{\boldsymbol{x}}^\top]/\epsilon^2$ and $V_0[\hat{\boldsymbol{x}}'] = E[\Delta\hat{\boldsymbol{x}}'\Delta\hat{\boldsymbol{x}}'^\top]/\epsilon^2$ are obtained in the following form (we omit the derivation [4]):

$$V_0[\hat{\boldsymbol{x}}] = V_0[\boldsymbol{x}] - \frac{(V_0[\boldsymbol{x}]\boldsymbol{G}\hat{\boldsymbol{x}}')(V_0[\boldsymbol{x}]\boldsymbol{G}\hat{\boldsymbol{x}}')^\top}{(\hat{\boldsymbol{x}}', \boldsymbol{G}^\top V_0[\boldsymbol{x}]\boldsymbol{G}\hat{\boldsymbol{x}}') + (\hat{\boldsymbol{x}}, \boldsymbol{G}V_0[\boldsymbol{x}']\boldsymbol{G}^\top \hat{\boldsymbol{x}})},$$
$$V_0[\hat{\boldsymbol{x}}'] = V_0[\boldsymbol{x}'] - \frac{(V_0[\boldsymbol{x}']\boldsymbol{G}^\top\hat{\boldsymbol{x}})(V_0[\boldsymbol{x}']\boldsymbol{G}^\top\hat{\boldsymbol{x}})^\top}{(\hat{\boldsymbol{x}}', \boldsymbol{G}^\top V_0[\boldsymbol{x}]\boldsymbol{G}\hat{\boldsymbol{x}}') + (\hat{\boldsymbol{x}}, \boldsymbol{G}V_0[\boldsymbol{x}']\boldsymbol{G}^\top \hat{\boldsymbol{x}})}. \tag{13}$$

We also obtain the normalized *a posteriori* correlation matrix $V_0[\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}'] = E[\Delta\hat{\boldsymbol{x}}\Delta\hat{\boldsymbol{x}}'^\top]/\epsilon^2$ in the following form (we omit the derivation [4]):

$$V_0[\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}'] = -\frac{(V_0[\boldsymbol{x}]\boldsymbol{G}\hat{\boldsymbol{x}}')(V_0[\boldsymbol{x}']\boldsymbol{G}^\top\hat{\boldsymbol{x}})^\top}{(\hat{\boldsymbol{x}}', \boldsymbol{G}^\top V_0[\boldsymbol{x}]\boldsymbol{G}\hat{\boldsymbol{x}}') + (\hat{\boldsymbol{x}}, \boldsymbol{G}V_0[\boldsymbol{x}']\boldsymbol{G}^\top \hat{\boldsymbol{x}})}. \tag{14}$$

Equation (14) implies that although the errors $\Delta\boldsymbol{x}$ and $\Delta\boldsymbol{x}'$ are statistically independent, their corrected values $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}'$ are correlated.

5. Reliability of 3-D Reconstruction

If the corrected values $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}'$ satisfy the epipolar equation (4), the position $\boldsymbol{r}$ of the reconstructed point is computed in the form

$$\boldsymbol{r} = Z\hat{\boldsymbol{x}}, \tag{15}$$

where Z is the *depth* from the XY plane of the first camera coordinate system. From the geometry shown in Fig. 1, the depth Z is obtained in the following form:

$$Z = \frac{(\boldsymbol{h} \times \boldsymbol{R}\hat{\boldsymbol{x}}', \hat{\boldsymbol{x}} \times \boldsymbol{R}\hat{\boldsymbol{x}}')}{\|\hat{\boldsymbol{x}} \times \boldsymbol{R}\hat{\boldsymbol{x}}'\|^2}. \tag{16}$$

From Eq. (15), the covariance matrix of the reconstructed 3-D position $\boldsymbol{r}$ is given in the following form:

$$V[\boldsymbol{r}] = Z^2 V[\hat{\boldsymbol{x}}] + Z(V[\hat{\boldsymbol{x}}, Z]\hat{\boldsymbol{x}}^\top + \hat{\boldsymbol{x}}V[\hat{\boldsymbol{x}}, Z]^\top) + V[Z]\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^\top. \tag{17}$$

In this equation, $V[\hat{\boldsymbol{x}}]$, $V[\hat{\boldsymbol{x}}']$, and $V[\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}']$ are, respectively, the a posteriori covariance matrices and the a posteriori correlation matrix given in the form

$$V[\hat{\boldsymbol{x}}] = \epsilon^2 V_0[\hat{\boldsymbol{x}}], \quad V[\hat{\boldsymbol{x}}'] = \epsilon^2 V_0[\hat{\boldsymbol{x}}'],$$
$$V[\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}'] = \epsilon^2 V_0[\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}']. \tag{18}$$

The correlation vector $V[\hat{\boldsymbol{x}}, Z]$ and the variance $V[Z]$ in Eq. (17) are given in the following form (the derivation is straightforward but lengthy [4]):

$$V[Z] = \frac{1}{\|\hat{\boldsymbol{x}} \times \boldsymbol{R}\hat{\boldsymbol{x}}'\|^2}\Big((Z^2(\hat{\boldsymbol{a}}', V[\hat{\boldsymbol{x}}]\hat{\boldsymbol{a}}') - 2ZZ'(\hat{\boldsymbol{a}}', V[\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}']\boldsymbol{R}^\top\hat{\boldsymbol{a}}') + Z'^2(\hat{\boldsymbol{a}}', \boldsymbol{R}V[\hat{\boldsymbol{x}}']\boldsymbol{R}^\top\hat{\boldsymbol{a}}')\Big), \tag{19}$$

$$V[\hat{\boldsymbol{x}}, Z] = -\frac{ZV[\hat{\boldsymbol{x}}] - Z'V[\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}']\hat{\boldsymbol{a}}'}{(\hat{\boldsymbol{a}}', \hat{\boldsymbol{x}})}. \tag{20}$$

Here, vectors $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{a}}'$ are defined by

$$\hat{\boldsymbol{a}} = N[\boldsymbol{h} \times \hat{\boldsymbol{x}}], \quad \hat{\boldsymbol{a}}' = \hat{\boldsymbol{a}} \times \boldsymbol{R}\hat{\boldsymbol{x}}', \tag{21}$$

and the symbol $N[\cdot]$ denotes normalization into a unit vector. The depth Z' with respect to the second camera is obtained from the geometry shown in Fig. 1 in the following form:

$$Z' = -\frac{(\hat{\boldsymbol{x}} \times \boldsymbol{h}, \hat{\boldsymbol{x}} \times \boldsymbol{R}\hat{\boldsymbol{x}}')}{\|\hat{\boldsymbol{x}} \times \boldsymbol{R}\hat{\boldsymbol{x}}'\|^2}. \tag{22}$$

An unbiased estimator of the squared noise level $\hat{\epsilon}^2$ is given by

$$\hat{\epsilon}^2 = \frac{(\boldsymbol{x}, \boldsymbol{G}\boldsymbol{x}')^2}{\|V_0[\hat{\boldsymbol{x}}']\boldsymbol{G}^\top\hat{\boldsymbol{x}}\|^2 + \|V_0[\hat{\boldsymbol{x}}]\boldsymbol{G}\hat{\boldsymbol{x}}'\|^2}. \tag{23}$$

This equation is a consequence of the fact that $1/\epsilon^2$ times the residual of the optimization (11) is a χ^2 variable of one degree of freedom (we omit the details [4]).

The geometric interpretation of the above process is schematically illustrated in Fig. 4. The covariance matrices $V[\boldsymbol{x}]$ and $V[\boldsymbol{x}']$ define the *Mahalanobis metric* in each image. An image is a two-dimensional space. A correspondence pair of points determines a point in the direct product space of the two images, and a natural metric is introduced into this four-dimensional space in the form of the direct product of the Mahalanobis metrics in the two images. The set of points that satisfy the epipolar equation (4) defines a three-dimensional *manifold* (hypersurface) in this four-dimensional space. Since the direct product of corresponding points detected in the presence of image noise is not necessarily

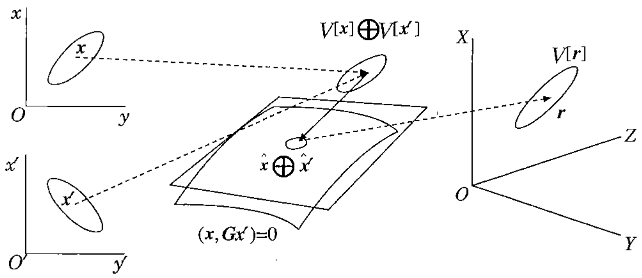


Fig. 4 The principle of 3-D reconstruction by stereo vision.

in this manifold, the direct product point is *projected* onto the nearest point in the manifold measured in the above defined metric. At the same time, the covariance matrix is projected onto the *tangent space* to this manifold. The estimator of the squared noise level given by Eq. (23) is simply the squared distance of this projection. The reconstructed position $\boldsymbol{r}$ and its covariance matrix $V[\boldsymbol{r}]$ are obtained by the mapping defined by Eqs. (15) and (16): the three-dimensional manifold and its tangent space are mapped into the three-dimensional space defined by the first camera coordinate system.

6. Visualization of Reliability

Let λ_1, λ_2, and λ_3 be the eigenvalues of the covariance matrix $V[\boldsymbol{r}]$, and $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ the corresponding orthonormal system of eigenvectors. We can define an ellipsoid with axes $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, and $\boldsymbol{u}_3$ and the corresponding radii $\sqrt{\lambda_1}$, $\sqrt{\lambda_2}$, and $\sqrt{\lambda_3}$. This ellipsoid indicates the reliability of the reconstructed 3-D point $\boldsymbol{r}$. We call the inside of this ellipsoid the *standard region.*

Define

$$\boldsymbol{r}^+ = \boldsymbol{r} + \sqrt{\lambda_{\max}}\boldsymbol{u}_{\max},$$
$$\boldsymbol{r}^- = \boldsymbol{r} - \sqrt{\lambda_{\max}}\boldsymbol{u}_{\max}, \tag{24}$$

where $\lambda_{\max}$ is the largest eigenvalue of $V[\boldsymbol{r}]$ and $\boldsymbol{u}_{\max}$ is the corresponding unit eigenvector. The vector $\boldsymbol{u}_{\max}$ indicates the orientation of the most likely deviation, and $\sqrt{\lambda_{\max}}$ is the standard deviation in that orientation. By displaying the vectors $\boldsymbol{r}^+$ and $\boldsymbol{r}^-$, we can visualize the reliability of the reconstructed point $\boldsymbol{r}$. We call these two vectors the *primary deviation pair.*

7. Numerical Simulation

We illustrate the effectiveness of our method by doing numerical simulation. We define a grid pattern on a cylindrical surface placed in space and regard the grid points as feature points (Fig. 5 (a)). The two cameras are assumed to have the same focal length $f = 600$ (pixels). After projecting the feature points onto the image planes, we add Gaussian random noise with standard deviation 2 (pixels) to each coordinate independently. Hence, the noise level ϵ is equal to $1/300$. The normalized covariance matrices are

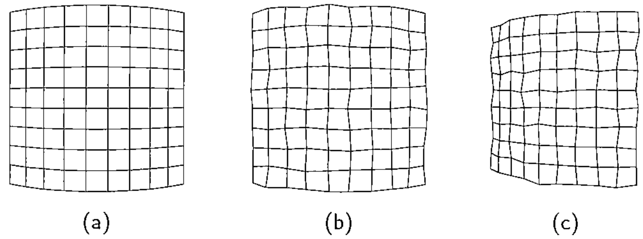


Fig. 5 (a) Grid pattern of a cylindrical surface. (b) Simulated stereo image with noise (left). (c) Simulated stereo image with noise (right).

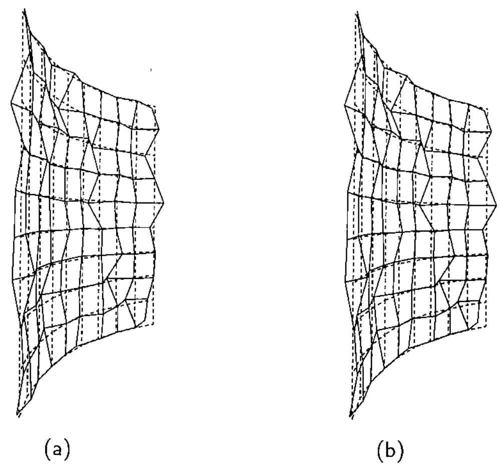


Fig. 6 (a) Reconstructed positions by our method (a side view). (b) Reconstructed positions by least squares (a side view).

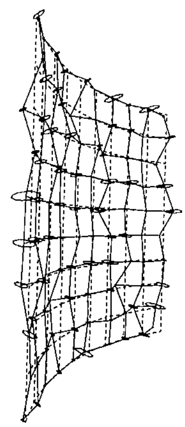


Fig. 7 Standard regions of the grid points.

$$V_0[\boldsymbol{x}] = V_0[\boldsymbol{x}'] = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 0 \end{pmatrix}. \tag{25}$$

However, the value of ϵ is regarded as unknown in the simulation. Figures 5(b) and 5(c) show simulated stereo images. The shape reconstructed by our method is shown in Fig. 6 (a). For the sake of comparison, we show in Fig. 6 (b) the reconstruction by the usual least-squares method (as mentioned in Sect. 1). Although the two results are almost identical, our method has the

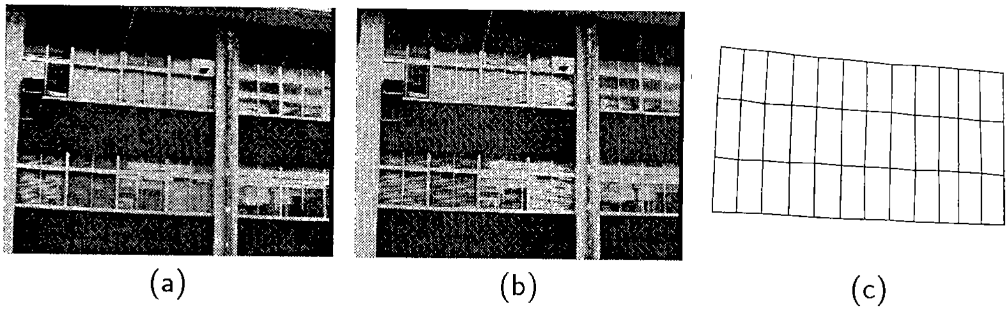


Fig. 8 (a) Left image. (b) Right image. (c) Grid pattern extracted from the left image.

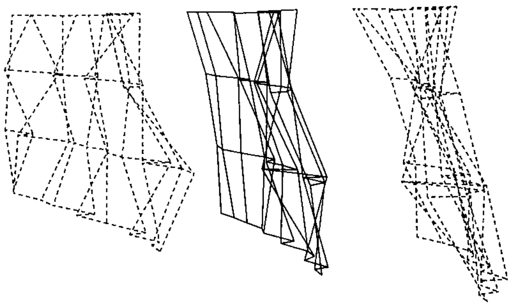


Fig. 9 Reconstructed position and its reliability (a side view).

advantage that the reliability of the reconstructed 3-D points is quantitatively evaluated. Figure 7 shows the standard regions of the reconstructed grid points. The true shape is drawn in dashed lines. We can see that most of the true positions are inside of the ellipsoids.

8. Real Image Experiment

Figures 8 (a) and 8 (b) show real stereo images. We regard the corners of the windows as feature points (Fig. 8 (c)). Figure 9 shows the reconstructed 3-D points. The 3-D shapes that envelope the primary deviation pairs of the reconstructed grid points are drawn in dashed lines. In this experiment, the distance between the two cameras is very short as compared with the distance to this building (approximately 1/16). Also, the noise level ϵ is estimated from the degree to which the epipolar equation (4) is not satisfied, so the error in the motion parameters $\{\boldsymbol{h}, \boldsymbol{R}\}$ is treated as "image noise". As a result, the reliability of this reconstruction is computed to be low. We can also see that the reliability of the right side of the surface is lower than the left side because the right side is farther away from the cameras than the left side.

9. Conclusion

We have presented an optimal 3-D reconstruction scheme for stereo vision by modeling statistical properties of image noise. It has turned out that the usual least-squares method is almost optimal, so our method does not improve the accuracy as far as the reconstructed 3-D shape is concerned. However, our method has the advantage that not only an optimal 3-D reconstruction but also its reliability can be computed in quantitative terms. This has a great significance in robot operations in real environments. We have also presented a scheme for visualizing the reliability by means the "primary deviation pairs".

In this paper, we did not assume any knowledge about the shape of the object. If the object is known to have a particular shape, say a planar surface, this knowledge can be used to increase the accuracy of 3-D reconstruction [7].

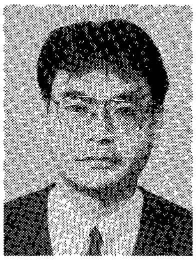
Acknowledgment

The authors thank Mr. Toru Kawashima of NTT for developing the graphics system for displaying 3-D for this research and helping our real image experiments. This work was in part supported by the Ministry of Education, Science, Sports and Culture, Japan under a Grant in Aid for Scientific Research B (No. 07458067) and the Okawa Institute of Information and Telecommunication, Japan.

References

[1] O. Faugeras, "Three-Dimensional Computer Vision: A Geometric Viewpoint," MIT Press, Cambridge, MA, U.S.A., 1993.
[2] K. Kanatani, "Geometric Computation for Machine Vision," Oxford University Press, Oxford, U.K., 1993.
[3] K. Kanatani, "Renormalization for unbiased estimation," Proc. 4th Int. Conf. Comput. Vision (ICCV'93), Berlin, Germany, pp.599–606, May 1993.
[4] K. Kanatani, "Statistical Optimization for Geometric Computation: Theory and Practice," Lecture Note, Second Edition, Department of Computer Science, Gunma University, June 1995.
[5] K. Kanatani and T. Maruyama, "Optimal focal length calibration system," Proc. IEEE/RSJ Int. Conf. Intell. Robots Sys. (IROS'93), Yokohama, Japan, pp.1816–1821, July 1993.
[6] K. Kanatani, "Statistical analysis of geometric computation," CVGIP: Image Understanding, vol.59, no.3, pp.286–306, 1994.
[7] Y. Kanazawa and K. Kanatani, "Direct reconstruction of

planar surfaces by stereo vision," IEICE Trans. Inf. & Syst., vol.E78-D, no.7, pp.917–922, 1995.
[8] J. Weng, T.S. Huang, and N. Ahuja, "Motion and Structure from Image Sequences," Springer, Berlin, Germany, 1993.



Yasushi Kanazawa received his B.E. and M.E. degrees in information and computer science from Toyohashi University of Technology, Aichi, Japan, in 1985 and 1987, respectively. He is currently Assistant Professor of information and computer engineering at Gunma College of Technology. His research interests include image processing and computer vision.

Kenichi Kanatani received his Ph.D. in applied mathematics from the University of Tokyo in 1979. He is currently Professor of computer science at Gunma University. He is the author of *Group-Theoretical Methods in Image Understanding* (Springer, 1990) and *Geometric Computation for Machine Vision* (Oxford Univ. Press, 1993).