*15th Int. Conf. Pattern Recognition (ICPR 2000), September 2000, Barcelona, Spain, Vol. 1, pp. 76–79.*

# Gauge-Based Reliability Analysis of 3-D Reconstruction from Two Uncalibrated Perspective Views

Kenichi Kanatani
Department of Computer Science, Gunma University, Kiryu, Gunma 376-8515 Japan
kanatani@cs.gunma-u.ac.jp

**Abstract**

*We evaluate the reliability of the 3-D (Euclidean) shape reconstructed from two uncalibrated perspective views. Introducing a statistical model of image noise, we optimally compute the fundamental matrix and evaluate its uncertainty in quantitative terms. We then evaluate the covariance matrices of the reconstructed 3-D points by propagating the image noise and the uncertainty in the fundamental matrix using a simple scheme. We show real-image experiments and discuss the effect of the "gauge" on the uncertainty description.*

## 1. Introduction

Many studies have been done for reconstructing 3-D shape from images [1, 12], but quantitative reliability evaluation has not been addressed very much. One reason for this is that 3-D reconstruction has frequently been studied in the *projective* framework because of its mathematical elegance. For error analysis, however, one has to work in the *Euclidean* framework, since noise is a physical quantity that occurs in the image frame.

First order error propagation can be done for calibrated or orthographic views [10, 16], and Cramer-Rao lower bounds can be evaluated for a simple form of bundle adjustment [15]. For general uncalibrated perspective views, however, error propagation or bundle adjustment is very complicated. In this paper, we present an alternative approach: we derive a (non-Cramer-Rao type) *theoretical accuracy bound* and devise an algorithm that attains that bound, which itself can be used as a reliability measure. Our uncertainty analysis is very similar to that of Csurka et al. [2] and in a sense a straightforward extension.

In this paper, we focus on 3-D Euclidean reconstruction from two uncalibrated perspective views and illustrate our technique. Introducing a statistical model of image noise, we optimally compute the fundamental matrix and evaluate its theoretical accuracy bound. Then, we decompose the computed fundamental matrix into the focal lengths and the camera motion parameters and reconstruct the 3-D shape. Finally, we evaluate the covariance matrices of the reconstructed points.

Here arises an important issue, to which little attention has been paid in the past: *3-D reconstruction from images necessarily involves inherent indeterminacy* [15]. These indeterminacies can be removed by imposing normalization constraints; e.g., fixing the coordinate origin to a particular point of the object and normalizing the size of something to unit length. But *such normalization affects the uncertainty description*: the normalized parameters have no uncertainty, and the uncertainty of all parameters is altered by merely changing the normalization conditions. In fact, we cannot speak of uncertainty of a particular parameter in absolute terms unless it is invariant to *gauge transformations* [6, 8, 9, 11]. We discuss the effect of the *gauge* on the uncertainty description by using our real-image examples.

## 2. Fundamental Matrix Computation

Suppose an uncalibrated camera rotates relative to a stationary scene by $\boldsymbol{R}$ (rotation matrix) and translates by $\boldsymbol{t}$. We define arbitrary image coordinate systems in the two images taken before and after the camera motion. Suppose a point $(u, v)$ in the first image corresponds to $(u', v')$ in the second image. We represent them by 3-D vectors

$$\boldsymbol{x} = \begin{pmatrix} u/f_0 \\ v/f_0 \\ 1 \end{pmatrix}, \qquad \boldsymbol{x}' = \begin{pmatrix} u'/f_0 \\ v'/f_0 \\ 1 \end{pmatrix}, \tag{1}$$

where $f_0$ is an appropriate estimate of the focal lengths before and after the camera motion. If no noise exists in the two images, the following *epipolar equation* is satisfied:

$$(\boldsymbol{x}, \boldsymbol{F}\boldsymbol{x}') = 0. \tag{2}$$

Throughout this paper, $(\boldsymbol{a}, \boldsymbol{b})$ denotes the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. The matrix $\boldsymbol{F}$ is singular and is called the *fundamental matrix*.

We have already implemented a statistically optimal method for computing $\boldsymbol{F}$ using a technique called *renormalization* [3] as a public domain C++ program[1]

[1] http://www.ail.cs.gunma-u.ac.jp/~kanatani/e/

[4]. We have also confirmed that this algorithm attains the theoretical accuracy bound. This program not only produces an optimal estimate $\hat{\boldsymbol{F}}$ of $\boldsymbol{F}$ but also its *standard deviation pair* $\{\boldsymbol{F}^{(+)}, \boldsymbol{F}^{(-)}\}$, which are the values of $\boldsymbol{F}$ in its parameter space apart from $\hat{\boldsymbol{F}}$ by the standard deviation in both ways in the direction along which errors are the most likely to occur [3]. If $\boldsymbol{F}^{(+)}$ and $\boldsymbol{F}^{(-)}$ coincide up to, say, three significant digits, the solution $\hat{\boldsymbol{F}}$ is guaranteed to have accuracy up to approximately three significant digits.

## 3. Optimal Feature Correction

We assume that each feature point undergoes independent Gaussian noise not necessarily homogeneous or isotropic. We also assume that the covariance matrix of each point is given only *up to scale*. Let $V_0[\boldsymbol{x}]$ and $V_0[\boldsymbol{x}']$ be appropriately normalized covariance matrices of vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$. They can be estimated from the residual of template matching [10, 13, 14]. If the noise has the same isotropic distribution everywhere, we have

$$V_0[\boldsymbol{x}] = V_0[\boldsymbol{x}'] = \mathrm{diag}(1, 1, 0). \tag{3}$$

where $\mathrm{diag}(\cdots)$ is the diagonal matrix with diagonal elements $\cdots$. We use this as the default value when no information is available.

After the fundamental matrix $\boldsymbol{F}$ is computed, we correct $\boldsymbol{x}$ and $\boldsymbol{x}'$ so that they satisfy (2) exactly in a statistically optimal way. This is done as follows [3]:

$$\begin{aligned} \hat{\boldsymbol{x}} &= \boldsymbol{x} - \frac{E(\boldsymbol{x}, \boldsymbol{x}')}{V(\boldsymbol{x}, \boldsymbol{x}')} V_0[\boldsymbol{x}] \boldsymbol{F} \boldsymbol{x}', \\ \hat{\boldsymbol{x}}' &= \boldsymbol{x}' - \frac{E(\boldsymbol{x}, \boldsymbol{x}')}{V(\boldsymbol{x}, \boldsymbol{x}')} V_0[\boldsymbol{x}'] \boldsymbol{F}^\top \boldsymbol{x}. \end{aligned} \tag{4}$$

Here, we define

$$\begin{aligned} E(\boldsymbol{x}, \boldsymbol{x}') &= (\boldsymbol{x}, \boldsymbol{F}\boldsymbol{x}'), \\ V(\boldsymbol{x}, \boldsymbol{x}') &= (\boldsymbol{x}', \boldsymbol{F}^\top V_0[\boldsymbol{x}] \boldsymbol{F} \boldsymbol{x}') + (\boldsymbol{x}, \boldsymbol{F} V_0[\boldsymbol{x}'] \boldsymbol{F}^\top \boldsymbol{x}). \end{aligned} \tag{5}$$

This procedure is essentially equivalent to the *triangulation* of Hartley and Sturm [7], but the above form is computationally much more efficient. Also, we can easily derive statistical properties of the resulting variables. In fact, the corrected positions $\hat{\boldsymbol{x}}$ are $\hat{\boldsymbol{x}}'$ have the following (normalized) covariance matrices [3]:

$$\begin{aligned} V_0[\hat{\boldsymbol{x}}] &= V_0[\boldsymbol{x}] - \frac{(V_0[\boldsymbol{x}]\boldsymbol{F}\boldsymbol{x}')(V_0[\boldsymbol{x}]\boldsymbol{F}\boldsymbol{x}')^\top}{V(\boldsymbol{x}, \boldsymbol{x}')}, \\ V_0[\hat{\boldsymbol{x}}'] &= V_0[\boldsymbol{x}'] - \frac{(V_0[\boldsymbol{x}']\boldsymbol{F}^\top\boldsymbol{x})(V_0[\boldsymbol{x}']\boldsymbol{F}^\top\boldsymbol{x})^\top}{V(\boldsymbol{x}, \boldsymbol{x}')}. \end{aligned} \tag{6}$$

We are assuming that errors in $\boldsymbol{x}$ and $\boldsymbol{x}'$ are independent, but the corrected values $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}'$ have the following (normalized) correlation matrix [3]:

$$V_0[\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}'] = -\frac{(V_0[\boldsymbol{x}]\boldsymbol{F}\boldsymbol{x}')(V_0[\boldsymbol{x}']\boldsymbol{F}^\top\boldsymbol{x})^\top}{V(\boldsymbol{x}, \boldsymbol{x}')}. \tag{7}$$

## 4. Self-Calibration Procedure

If the principal point is at the center of the frame and no image skew exists, we can self-calibrate the focal lengths $f$ and $f'$ of the two images and the translation $\boldsymbol{t}$ and the rotation $\boldsymbol{R}$ of the camera by an analytical method [1, 5]

If the estimate $f_0$ is replaced by the true values $f$ and $f'$ in the representation (1), the depths of the point we are viewing are estimated with respect to the first and the second camera coordinate systems in the following form [3]:

$$\hat{Z} = (\boldsymbol{t} \times \boldsymbol{R}\hat{\boldsymbol{x}}', \boldsymbol{n}), \qquad \hat{Z}' = (\boldsymbol{t} \times \hat{\boldsymbol{x}}, \boldsymbol{n}) \tag{8}$$

Here, we have defined

$$\boldsymbol{n} = \frac{\hat{\boldsymbol{x}} \times \boldsymbol{R}\hat{\boldsymbol{x}}'}{\|\hat{\boldsymbol{x}} \times \boldsymbol{R}\hat{\boldsymbol{x}}'\|^2}. \tag{9}$$

The 3-D position $\hat{\boldsymbol{r}}$ with respect to the first camera coordinate system is given by

$$\hat{\boldsymbol{r}} = \hat{Z}\hat{\boldsymbol{x}}. \tag{10}$$

Its normalized covariance matrix is given by

$$V_0[\hat{\boldsymbol{r}}] = \hat{Z}^2 V_0[\hat{\boldsymbol{x}}] + 2\hat{Z} S[V_0[\hat{Z}, \hat{\boldsymbol{x}}]\hat{\boldsymbol{x}}^\top] + V_0[\hat{Z}]\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^\top, \tag{11}$$

where $S[\,\cdot\,]$ denotes the symmetrization operation ($S[\boldsymbol{A}] = (\boldsymbol{A} + \boldsymbol{A}^\top)/2$). The normalized covariance matrix $V_0[\hat{\boldsymbol{x}}]$ is given by the first of eqs. (6) (the estimate $f_0$ is replaced by its true values $f$ and $f'$ appropriately in its representation). From eq. (8), we obtain the normalized variance $V_0[\hat{Z}]$ and the normalized correlation vector $V_0[\hat{Z}, \hat{\boldsymbol{x}}]$ in the form

$$\begin{aligned} V_0[\hat{Z}] = \frac{1}{\|\hat{\boldsymbol{x}} \times \boldsymbol{R}\hat{\boldsymbol{x}}'\|^2} \Big( & \hat{Z}^2 (\boldsymbol{m}, V_0[\hat{\boldsymbol{x}}]\boldsymbol{m}) \\ & - 2\hat{Z}\hat{Z}'(\boldsymbol{m}, V_0[\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}']\boldsymbol{R}^\top \boldsymbol{m}) \\ & + \hat{Z}'^2 (\boldsymbol{m}, \boldsymbol{R} V_0[\hat{\boldsymbol{x}}']\boldsymbol{R}^\top \boldsymbol{m}) \Big), \end{aligned} \tag{12}$$

$$V_0[\hat{Z}, \hat{\boldsymbol{x}}] = -\frac{(\hat{Z} V_0[\hat{\boldsymbol{x}}] - \hat{Z}' V_0[\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}']\boldsymbol{R}^\top)\boldsymbol{m}}{(\boldsymbol{m}, \hat{\boldsymbol{x}})}, \tag{13}$$

where we have defined

$$\boldsymbol{m} = N[\boldsymbol{t} \times \hat{\boldsymbol{x}}] \times \boldsymbol{R}\hat{\boldsymbol{x}}'. \tag{14}$$

The operator $N[\,\cdot\,]$ denotes normalization into a unit vector ($N[\boldsymbol{a}] = \boldsymbol{a}/\|\boldsymbol{a}\|$).

**Figure 1. Real images of an indoor scene.**

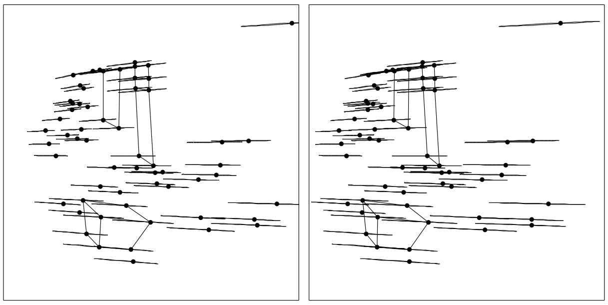


**Figure 2. 3-D reconstruction and uncertainty ellipsoids (stereogram).**

The term "normalized" means that the absolute noise magnitude $\epsilon$, which we call the *noise level*, is normalized to 1. An estimate $\hat{\epsilon}$ of the noise level $\epsilon$ is automatically evaluated in the course of computing the fundamental matrix $\boldsymbol{F}$ [4]. Hence, the absolute covariance matrix $V[\hat{\boldsymbol{r}}]$ equals $\hat{\epsilon}^2 V_0[\hat{\boldsymbol{r}}]$.

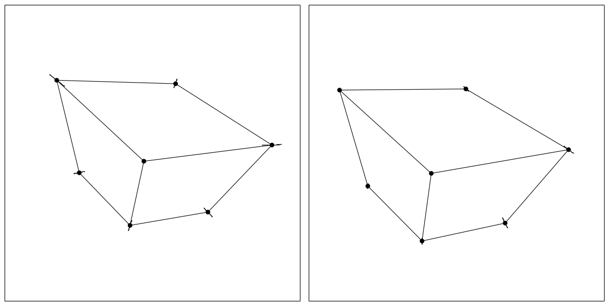
## 5 Reliability of 3-D Reconstruction

The above covariance analysis is based on the assumption that the fundamental matrix $\boldsymbol{F}$ is correct. However, $\boldsymbol{F}$ itself has uncertainty because it is computed form the data. In fact, we can evaluate its "covariance tensor" in the process of computing $\boldsymbol{F}$ [4]. The covariance of the focal lengths and the motion parameters could be computed from it, but the correlations among themselves and with all other quantities are intractably complicated. Here, we introduce the following approximation.

Since our renormalization-based algorithm for computing the fundamental matrix $\boldsymbol{F}$ can compute the primary deviation pair $\{\boldsymbol{F}^{(+)}, \boldsymbol{F}^{(-)}\}$, we can obtain the corresponding 3-D positions $\boldsymbol{r}^{(+)}$ and $\boldsymbol{r}^{(-)}$ from this. Here, we identify the line segment connecting them with the uncertainty of the reconstructed position $\hat{\boldsymbol{r}}$, i.e., we regard $(\boldsymbol{r}^{(+)} - \hat{\boldsymbol{r}})(\boldsymbol{r}^{(+)} - \hat{\boldsymbol{r}})^\top$ as an approximate covariance matrix of $\hat{\boldsymbol{r}}$, although this is not a purely geometric property, since analysis based on $\boldsymbol{F}^{(\pm)}$ depends on this particular parameterization of $\boldsymbol{F}$.

Since the fundamental matrix $\boldsymbol{F}$ is computed from all feature points by optimally fitting the epipolar constraint, the correlation of $\boldsymbol{F}$ with a particular feature point is expected to be very small. Hence, we approximate the covariance matrix $V[\hat{\boldsymbol{r}}]$ as the sum of the two error sources:

$$V[\hat{\boldsymbol{r}}] = \hat{\epsilon}^2 V_0[\hat{\boldsymbol{r}}] + (\boldsymbol{r}^{(+)} - \hat{\boldsymbol{r}})(\boldsymbol{r}^{(+)} - \hat{\boldsymbol{r}})^\top. \tag{15}$$



(a) (b)

**Figure 3. Gauge dependence of uncertainty description. (a) Normalization based on the centroid. (b) Normalization based on three vertices.**

| | ratio | angle (deg) |
|---|---|---|
| computed value | 1.014 | 96.5 |
| true value | 1.000 | 90.0 |
| predicted standard deviation | 0.003 | 2.4 |

**Table 1. Reliability of gauge invariants.**

## 6. Real Image Experiment

We reconstructed the 3-D shape from the two images shown in Fig. 1, using the feature points marked in the images. Fig. 2 is a side view of the reconstructed points (stereogram); wireframes consisting of some of them are shown. On each reconstructed point is centered the uncertainty ellipsoid defined by the covariance matrix given by eq. (15). All ellipsoids look like thin needles, indicating that the uncertainty is very large along the depth orientation.

*This description is deceptive, however.* This uncertainty description is based on a particular *gauge*, i.e., a choice of normalization, such that the world coordinate system coincides with the first camera frame and the camera translation is normalized to unit length [6, 11]. This gauge hides the fact that the uncertainty is mostly due to the camera translation. In fact, what is uncertain is the depth of the object *as a whole*, not the object shape.

For example, if we take the centroid of the polyhedral object as the coordinate origin and normalize the root-mean-square distance to the vertices from the centroid to unit length, we obtain the description shown in Fig. 3(a). By construction, the uncertainty is almost symmetric with respect to the centroid, and the object shape has very little uncertainty.

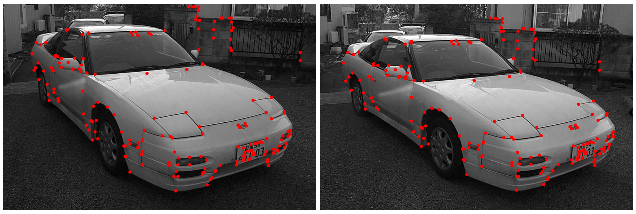


**Figure 4. Two images of a car.**

**Figure 5. Reconstructed and texture-mapped 3-D shape.**

Fig. 3(b) is the uncertainty description for yet another gauge: one of the object vertex is taken to be the coordinate origin, another is taken to be (1,1,0), and a third one is on the $XY$ plane. By definition, the first two points have no uncertainty.

It follows that *uncertainty of individual quantities has no absolute meaning*. In other words, the discrepancy of the reconstructed quantities from their true values is not a meaningful measure of accuracy if artificial normanizations are involved.

Let us call the description changes due to choosing different gauges (i.e., normanizations) *gauge transformations*. Absolute meaning can be given only to *gauge invariants* [6], i.e., quantities invariant to gauge transformations. Typical gauge invariants for Euclidean reconstruction are ratios of lengths and angles of lines. Table 1 lists the ratio of two sides of the polyhedral objects and the angle they make along with their true values and their standard deviations derived by the covariance matrices of the vertices.

Fig. 4 shows two real images of a car. Fig. 5 shows its 3-D shape computed from the feature points marked in the images. We defined a wireframe with triangular meshes from the reconstructed points and mapped the texture onto it. A fairly accurate 3-D shape is created even though only two views are used.

## 7. Concluding Remarks

We have shown how to evaluate the reliability of the 3-D (Euclidean) shape reconstructed from two uncalibrated perspective views. Introducing a statistical model of image noise, we optimally computed the fundamental matrix and evaluated its uncertainty in quantitative terms. We then evaluated the covariance matrices of the reconstructed 3-D points by propagating the image noise and the uncertainty in the fundamental matrix using a simple scheme. We showed real-image experiments and discussed the effect of the gauge on the uncertainty description.

## References


[1] S. Bougnoux, From projective to Euclidean space under any practical situation, a criticism of self calibration, *Proc. 6th Int. Conf. Comput. Vision.*, January 1998, Bombay, India, pp. 790–796.

[2] G. Csurka, C. Zeller, Z. Zhang and O. D. Faugeras, Characterizing the uncertainty of the fundamental matrix, *Comput Vision Image Understand.*, **68**-1 (1997), 18–36.

[3] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier, Amsterdam, 1996.

[4] K. Kanatani, Optimal fundamental matrix computation: Algorithm and reliability analysis, *Proc. 6th Symp. Sensing via Image Information*, June 2000, Yokohama, Japan.

[5] K. Kanatani and C. Matsunaga, Closed-form expression for focal lengths from the fundamental matrix, *Proc. 4th Asian Conf. Computer Vision*, January, 2000, Taipei, Taiwan, pp. 128–133.

[6] K. Kanatani and D. D. Morris, Gauges and gauge transformations in 3D reconstruction from a sequence of images, *Proc. 4th Asian Conf. Computer Vision*, January 2000, Taipei, Taiwan, pp. 1046–1051.

[7] R. I. Hartley and P. Sturm, Triangulation, *Comput. Vision. Image Understand.*, **68**-2 (1997), 146–157.

[8] P. F. McLauchlan, Gauge invariance in projective 3D reconstruction, *IEEE Workshop on Multi-View Modeling and Analysis of Visual Scenes*, June 1999, Fort Collins, CO, U.S.A..

[9] P. F. McLauchlan, Coordinate-frame independence in optimization algorithms for 3D vision, *IEEE Workshop on Vision Algorithms: Theory and Practice*, September 1999, Corfu, Greece.

[10] D. D. Morris and T. Kanade, A unified factorization algorithm for points, line segments and planes with uncertainty models, *Proc. Int. Conf. Comput. Vision*, January 1998, Bombay, India, pp. 696–702.

[11] D. D. Morris, K. Kanatani and T. Kanade, Uncertainty modeling for optimal structure from motion, *IEEE Workshop on Vision Algorithm: Theory and Practice*, September 1999, Corfu, Greece.

[12] M. Pollefeys, R. Koch and L. Van Gool, Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters, *Int. J. Comput. Vision*, **32**-1 (1999), 7–26.

[13] J. Shi and C. Tomasi, Good features to track, *Proc. IEEE Conf. Comput. Vision Patt. Recogn.*, June 1994, Seattle, WA, pp. 593–600.

[14] A. Singh, An estimation-theoretic framework for image-flow computation, *Proc. 3rd Int. Conf. Comput. Vision*, December, 1990, Osaka, Japan, pp. 168–177.

[15] R. Szeliski and S. B. Kang, Shape ambiguities in structure from motion, *IEEE Trans. Patt. Anal. Mach. Intell.*, **19**-5 (1997), 506–512.

[16] J. Weng, N. Ahuja and T. S. Huang, Optimal motion and structure estimation, *IEEE Trans. Patt. Anal. Mach. Intell.*, **15**-9 (1993), 864–884.