Stabilizing Image Mosaicing by Model Selection

Yasushi Kanazawa[1] and Kenichi Kanatani[2]

[1] Department of Knowledge-based Information Engineering
Toyohashi University of Technology
1–1 Hibarigaoka, Tenpaku, Toyohashi, Aichi 441-8580 Japan
kanazawa@tutkie.tut.ac.jp
[2] Department of Computer Science, Gunma University
1–5–1, Tenjin, Kiryu, Gunma 376-8515 Japan
kanatani@cs.gunma-u.ac.jp

Abstract. The computation for image mosaicing using homographies is numerically unstable and causes large image distortions if the matching points are small in number and concentrated in a small region in each image. This instability stems from the fact that actual transformations of images are usually in a small subgroup of the group of homographies. It is shown that such undesirable distortions can be removed by *model selection* using the geometric AIC without introducing any empirical thresholds. It is shown that the accuracy of image mosaicing can be improved *beyond* the theoretical bound imposed on statistical optimization. This is made possible by our *knowledge* about probable subgroups of the group of homographies. We demonstrate the effectiveness of our method by real image examples.

1 Introduction

Image mosaicing is a technique for integrating multiple images into one continuous image, a typical one being a *panoramic image* [10, 13, 15, 16]. This technique has long been used for creating terrain maps from aerial images or analyzing remote sensing satellite images, but recently its applications to virtual reality creation from multiple scene images are attracting much attention. Image mosaicing also plays an important role in automatic surveillance using camera images.

The basic principle underlying image mosaicing is the computation of a *homography*, which is a mapping that typically occurs between two perspective images of a planar surface in the scene [3]. Since faraway scenes can effectively be regarded as planar surfaces, we can register one image to another by computing the homography between them.

If the images have very small overlaps between them, as is often the case for remote sensing images and aerial images, only a small number of matching points are available. In such a case, the selected points in one image may be mapped to the corresponding points in the other image fairly accurately, but if we extrapolate this mapping to portions apart from the matching points, a large distortion may occur even in the presence of very small noise (Fig. 1(a)). Since

(a) (b)

Fig. 1. (a) If images with a very small overlap are used for mosaicing, a large distortion may result even in the presence of small noise. (b) Some part of the image may appear from the other side of the frame.

a homography may map some points to infinity, the part beyond those points can appear from the other side of the image frame (Fig. 1(b)).

In [7] this instability was demonstrated by using real images. An accurate algorithm was also presented for computing a homography from point correspondences using a technique called *renormalization*, which not only produces a statistically optimal solution but also evaluates the reliability of the computed solution in quantitative terms. The algorithm is implemented in C++ and publicly available via the Web[3]. A theoretical accuracy bound was also derived for the homography computation. It was experimentally confirmed that the renormalization algorithm indeed produces estimates in the vicinity of that bound.

Although the renormalization algorithm dramatically reduces the instability of the mapping, as demonstrated in [7], it cannot remove the distortion completely. However, further improvement is theoretically impossible with this technique. In this paper, we will show that this limitation *can* be broken through by incorporating our *knowledge* about the source of the instability. The instability stems from the fact that while homographies constitute an 8-parameter group of transformations, actual transformations are usually in a small subgroup, e.g., the group of translations, the group of rigid motions, the group of similarities, or the group of affine transformations. In the presence of noise, the computed solution moves out of the subgroup to which it should belong, causing a large image distortion.

In the following, we show that such undesirable distortions can be removed by *model selection* using the *geometric AIC* [4, 5] without introducing any empirical thresholds. We also present a Levenberg-Marquardt scheme for optimization and an analytical procedure for computing an initial guess. We demonstrate the effectiveness of our method by real image examples.

[3] `http://www.ail.cs.gunma-u.ac.jp/Labo/programs-e.html`

2 Representation of Homography

A *homography* is an image mapping expressed in the following form:

$$x' = \frac{Ax + By + C}{Px + Qy + R}, \qquad y' = \frac{Dx + Ey + F}{Px + Qy + R}. \tag{1}$$

If we define vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ and matrix $\boldsymbol{H}$ by

$$\boldsymbol{x} = \begin{pmatrix} x/f_0 \\ y/f_0 \\ 1 \end{pmatrix}, \qquad \boldsymbol{x}' = \begin{pmatrix} x'/f_0 \\ y'/f_0 \\ 1 \end{pmatrix}, \qquad \boldsymbol{H} = \begin{pmatrix} A & B & C/f_0 \\ D & E & F/f_0 \\ P/f_0 & Q/f_0 & R/f_0^2 \end{pmatrix}, \tag{2}$$

eq. (1) can be written as

$$\boldsymbol{x}' = Z[\boldsymbol{H}\boldsymbol{x}]. \tag{3}$$

Here, $Z[\,\cdot\,]$ denotes normalization to make the third element 1; f_0 is a scale factor chosen so that x/f_0 and y/f_0 have order 1.

Given two images, we choose matching points from them. Let $\{(x_\alpha, y_\alpha)\}$ and $\{(x'_\alpha, y'_\alpha)\}$ be the coordinates of the points chosen from the first and the second images, respectively. Let $\{\boldsymbol{x}_\alpha\}$ and $\{\boldsymbol{x}'_\alpha\}$ be their vector representations. We regard them as random Gaussian variables with covariance matrices $V[\boldsymbol{x}_\alpha]$ and $V[\boldsymbol{x}'_\alpha]$.

The absolute magnitude of noise is difficult to predict a priori, but its geometric characteristics such as homogeneity/inhomogeneity and isotropy/anisotropy can be relatively easily predicted. For example, if we use template matching for finding corresponding points, the uncertainty of matching is measured by the Hessian of the residual surface around the detected point [2, 9, 11, 12]. Here, we assume that the covariance matrices $V[\boldsymbol{x}_\alpha]$ and $V[\boldsymbol{x}'_\alpha]$ are known *up to scale* and write

$$V[\boldsymbol{x}_\alpha] = \epsilon^2 V_0[\boldsymbol{x}_\alpha], \qquad V[\boldsymbol{x}'_\alpha] = \epsilon^2 V_0[\boldsymbol{x}'_\alpha]. \tag{4}$$

We call the unknown magnitude ϵ the *noise level*. The matrices $V_0[\boldsymbol{x}_\alpha]$ and $V_0[\boldsymbol{x}'_\alpha]$, which we call the *normalized covariance matrices*, specify the relative dependence of noise occurrence on positions and orientations. If no a priori knowledge is available for them, we simply assume isotropy and homogeneity and input the default values $V_0[\boldsymbol{x}_\alpha] = V_0[\boldsymbol{x}'_\alpha] = \mathrm{diag}(1, 1, 0)$ (the diagonal matrix whose diagonal elements are 1, 1, and 0 in that order).

3 Optimal Homography Estimation

Eq. (3) can equivalently be written in the form $\boldsymbol{x}' \times \boldsymbol{H}\boldsymbol{x} = \boldsymbol{0}$. Hence, the task is to estimate $\boldsymbol{H}$ from noisy data $\{\boldsymbol{x}_\alpha\}$ and $\{\boldsymbol{x}'_\alpha\}$ with the knowledge that their true values $\{\bar{\boldsymbol{x}}_\alpha\}$ and $\{\bar{\boldsymbol{x}}'_\alpha\}$ satisfy

$$\bar{\boldsymbol{x}}'_\alpha \times \boldsymbol{H}\bar{\boldsymbol{x}}_\alpha = \boldsymbol{0}. \tag{5}$$

The reliability of an estimate $\hat{\boldsymbol{H}}$ of $\boldsymbol{H}$ can be measured by its *covariance tensor* $\mathcal{V}[\hat{\boldsymbol{H}}]$. A theoretical lower bound on it can be derived in analytical terms [7].

It is well known [4] that an optimal estimate of $\boldsymbol{H}$ which attains the accuracy bound in the first order (i.e., if terms of $O(\epsilon^4)$ are ignored) can be obtained by *maximum likelihood estimation*, minimizing the squared *Mahalanobis distances*

$$J = \frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}_\alpha, V_0[\boldsymbol{x}_\alpha]_2^-(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}_\alpha)) + \frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{x}'_\alpha - \bar{\boldsymbol{x}}'_\alpha, V_0[\boldsymbol{x}'_\alpha]_2^-(\boldsymbol{x}'_\alpha - \bar{\boldsymbol{x}}'_\alpha)), \quad (6)$$

subject to the constraint (5). Here and throughout this paper, $(\boldsymbol{a}, \boldsymbol{b})$ denotes the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. The super script $(\cdot)_r^-$ denotes the (Moore-Penrose) generalized inverse computed after replacing the smallest $n - r$ eigenvalues by zeros.

Using Lagrange multipliers and introducing first order approximation, we can eliminate the constraint (5) and express eq. (6) in the following form [4]:

$$J = \frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{x}'_\alpha \times \boldsymbol{H}\boldsymbol{x}_\alpha, \boldsymbol{W}_\alpha(\boldsymbol{x}'_\alpha \times \boldsymbol{H}\boldsymbol{x}_\alpha)), \quad (7)$$

$$\boldsymbol{W}_\alpha = \Big(\boldsymbol{x}'_\alpha \times \boldsymbol{H}V_0[\boldsymbol{x}_\alpha]\boldsymbol{H}^\top \times \boldsymbol{x}'_\alpha + (\boldsymbol{H}\boldsymbol{x}_\alpha) \times V_0[\boldsymbol{x}'_\alpha] \times (\boldsymbol{H}\boldsymbol{x}_\alpha)\Big)_2^-. \quad (8)$$

Let $\hat{J}$ be the residual, i.e., the minimum of the function J. It can be shown that $\hat{J}/\epsilon^2$ is subject to a χ^2 distribution with $2(N-4)$ degrees of freedom to a first approximation [4]. Hence, an unbiased estimator of ϵ^2 is obtained in the form

$$\hat{\epsilon}^2 = \frac{\hat{J}}{2(1-4/N)}. \quad (9)$$

In [7] a computational technique called *renormalization* was presented. It was experimentally confirmed that the solution practically falls on the theoretical accuracy bound.

4 Models of Image Transformations

Since the elements of $\boldsymbol{H}$ have scale indeterminacy (see eq. (3)), a homography has eight degrees of freedom. However, image transformations that we often encounter have much smaller degrees of freedom. For example, if a moving camera takes images of a faraway scene with varying zooming, the translation of the camera causes no visible changes, so the image transformation is parameterized by the camera rotation $\boldsymbol{R}$ (three degrees of freedom) and the focal lengths f and f' of the two frames (Fig. 2). Such transformations constitute a 5-parameter subgroup of the 8-parameter group of homographies. If the focal length is fixed, we obtain a 4-parameter subgroup.

If the camera translates relative to a nearby scene, we have the group of translations with two degrees of freedom (Fig. 3(b)). If the camera is allowed to rotate around its optical axis, we have the group of 2-D rigid motions with three degrees of freedom (Fig. 3(c)). If the focal length is also allowed to change,

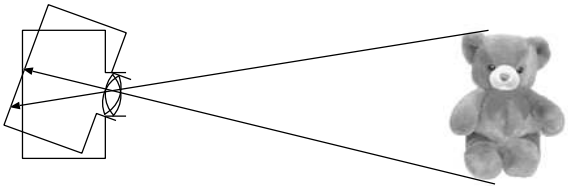


Fig. 2. Image transformation due to camera rotation.

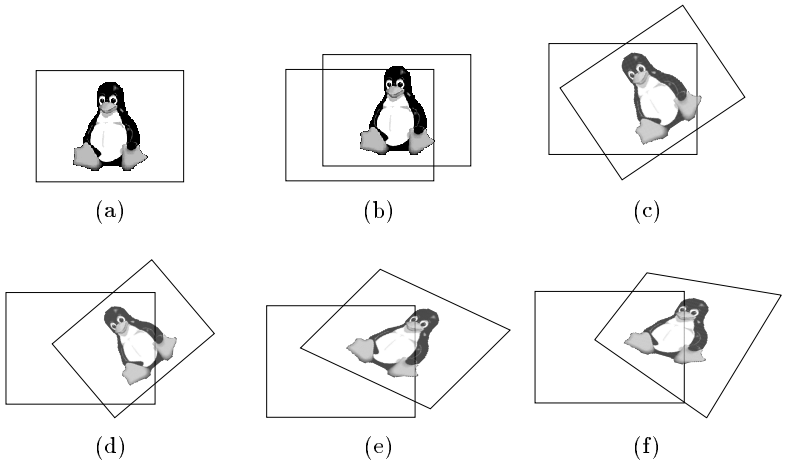


Fig. 3. (a) Original image. (b) Translation. (c) Rigid motion. (d) Similarity. (e) Affine transformation. (f) Homography.

we have the group of similarities with four degrees of freedom (Fig. 3(d)). If the object is a planar surface in the distance, the image transformation can be viewed as an affine transformation with six degrees of freedom (Fig. 3(e)). All these image transformations belong to a subgroup of the group of general homographies (Fig. 3(f)).

Thus, we have a hierarchy of image transformations (Fig. 4). In the presence of noise, however, the computed homography need not necessarily belong to the required subgroup, resulting in a large image distortion that cannot be attributed to any camera motion. Such a distortion can be removed if we find a homography within the required subgroup or *model*. For example, if the image transformation is known to be a 2-D rigid motion, we only need to compute the image rotation and translation optimally. However, we do not know *a priori* to which model the observed transformation should belong.

A naive idea is to choose from among candidate models the one that gives the smallest residual. This does not work, however, because the 8-parameter homography group is always chosen: a model with more degrees of freedom has a smaller residual. For a fair comparison, we need to compensate for the overfit

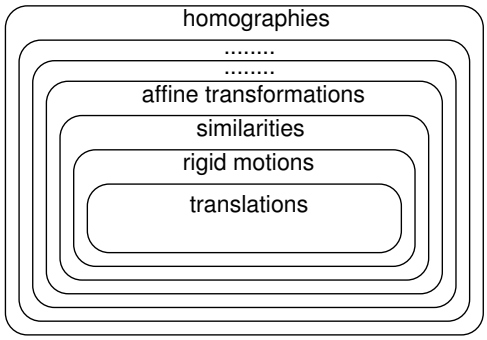


Fig. 4. Hierarchy of image transformations.

caused by excessive degrees of freedom. Here, we measure the goodness of a model by the *geometric AIC* [4, 5, 6], which is a special form of Akaike's *AIC* [1]. The model with the smallest geometric AIC is preferred. See [14] for other model selection criteria.

5 Subgroup Hierarchy

5.1 8-parameter homographies

Let $\hat{J}_{H_8}$ be the resulting residual of eq. (7). The geometric AIC is given by

$$\text{G-AIC}_{H_8} = \hat{J}_{H_8} + \frac{16}{N}\epsilon^2, \tag{10}$$

where the square noise level ϵ^2 is estimated by eq. (9).

5.2 5-parameter homographies

If the camera rotates by $\boldsymbol{R}$ around the center of the lens and changes the focal length from f to f', the resulting homography has the form

$$\boldsymbol{H} = \boldsymbol{F}'^{-1}\boldsymbol{R}^\top\boldsymbol{F}, \tag{11}$$

where

$$\boldsymbol{F} = \text{diag}(1, 1, \frac{f}{f_0}), \qquad \boldsymbol{F}' = \text{diag}(1, 1, \frac{f'}{f_0}). \tag{12}$$

We use the Levenberg-Marquardt method (LM method) for minimizing eq. (7). First, we define the following non-dimensional variables:

$$\phi = \frac{f}{f_0}, \qquad \phi' = \frac{f'}{f_0}. \tag{13}$$

The minimization procedure goes as follows:

1. Let c=0.001. Analytically compute initial guesses of ϕ, ϕ', and $\boldsymbol{R}$ (see Appendix A), and evaluate the residual $J = J(\phi, \phi', \boldsymbol{R})$.
2. Compute the gradient ∇J and Hessian $\nabla^2 J$ (see Appendix B for the detailed expressions):

$$\nabla J = \begin{pmatrix} \partial J/\partial \phi \\ \partial J/\partial \phi' \\ \nabla_{\Omega} J \end{pmatrix}, \tag{14}$$

$$\nabla^2 J = \begin{pmatrix} \partial^2 J/\partial \phi^2 & \partial^2 J/\partial \phi \partial \phi' & (\nabla_{\Omega} \partial J/\partial \phi)^\top \\ \partial^2 J/\partial \phi' \partial \phi & \partial^2 J/\partial \phi'^2 & (\nabla_{\Omega} \partial J/\partial \phi')^\top \\ \nabla_{\Omega} \partial J/\partial \phi & \nabla_{\Omega} \partial J/\partial \phi' & \nabla^2_{\Omega} J \end{pmatrix}. \tag{15}$$

3. Let $\boldsymbol{D}$ be the diagonal matrix consisting of the diagonal elements of $\nabla^2 J$, and solve the following simultaneous linear equation:

$$(\nabla^2 J + c\boldsymbol{D}) \begin{pmatrix} \Delta\phi \\ \Delta\phi' \\ \Delta\boldsymbol{\Omega} \end{pmatrix} = -\nabla J. \tag{16}$$

4. Compute the residual $J' = J(\phi + \Delta\phi, \phi' + \Delta\phi', \mathcal{R}(\Delta\boldsymbol{\Omega})\boldsymbol{R})$.
   - If $J > J'$, let $c \leftarrow 10c$ and go back to Step 3.
   - If $J < J'$ and $|J - J'|/J < \epsilon_J$, return ϕ, ϕ', and $\boldsymbol{R}$ and stop.
   - Else, let $c \leftarrow c/10$, update ϕ, ϕ', and $\boldsymbol{R}$ in the form

$$\phi \leftarrow \phi + \Delta\phi, \qquad \phi' \leftarrow \phi' + \Delta\phi', \qquad \boldsymbol{R} \leftarrow \mathcal{R}(\Delta\boldsymbol{\Omega})\boldsymbol{R}, \tag{17}$$

     and go back to Step 2.

Here, ϵ_J is a threshold for convergence, and $\mathcal{R}(\Delta\boldsymbol{\Omega})$ denotes the rotation around $\Delta\boldsymbol{\Omega}$ by an angle $\|\Delta\boldsymbol{\Omega}\|$. Let $\hat{J}_{H_5}$ be the resulting residual. The geometric AIC is given by

$$\text{G-AIC}_{H_5} = \hat{J}_{H_5} + \frac{10}{N}\epsilon^2, \tag{18}$$

where the square noise level ϵ^2 is estimated by eq. (9).

5.3 4-parameter homographies

If we let $f = f'$ in eq. (11), we obtain the 4-parameter group of homographies, for which optimal values of ϕ and $\boldsymbol{R}$ are obtained by slightly modifying the LM method described above. Let $\hat{J}_{H_4}$ be the resulting residual. The geometric AIC is given by

$$\text{G-AIC}_{H_4} = \hat{J}_{H_4} + \frac{8}{N}\epsilon^2. \tag{19}$$

5.4 Similarities

A similarity is a special homography that has the following form:

$$\boldsymbol{H} = \begin{pmatrix} s\cos\theta & -s\sin\theta & t_1/f_0 \\ s\sin\theta & s\cos\theta & t_2/f_0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{20}$$

By this transformation, the image is rotated by angle θ around the origin, scaled by s, and translated by (t_1, t_2). If we define

$$\vec{x}_\alpha = \begin{pmatrix} x_\alpha/f_0 \\ y_\alpha/f_0 \end{pmatrix}, \qquad \vec{x}'_\alpha = \begin{pmatrix} x'_\alpha/f_0 \\ y'_\alpha/f_0 \end{pmatrix}, \tag{21}$$

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \qquad \vec{\tau} = \begin{pmatrix} t_1/f_0 \\ t_2/f_0 \end{pmatrix}, \tag{22}$$

eq. (7) is rewritten in the following form:

$$J = \frac{1}{N}\sum_{\alpha=1}^{N} (\vec{x}'_\alpha - sR\,\vec{x}_\alpha - \vec{\tau}, W_\alpha(\vec{x}'_\alpha - sR\,\vec{x}_\alpha - \vec{\tau})), \tag{23}$$

$$W_\alpha = \left(s^2 R V_0[\vec{x}_\alpha] R^\top + V_0[\vec{x}'_\alpha]\right)^{-1}. \tag{24}$$

This is minimized by the LM method (we omit the details). See Appendix C for the procedure for computing an initial guess. Let $\hat{J}_S$ be the resulting residual. The geometric AIC is given by

$$\text{G-AIC}_S = \hat{J}_S + \frac{8}{N}\epsilon^2. \tag{25}$$

5.5 Rigid motions

The image transformation reduces to a 2-D rigid motion if we let $s = 1$ in eq. (20). We can apply the same LM method for minimizing J and the procedure for computing an initial guess after an appropriate modification. Let $\hat{J}_M$ be the resulting residual. The geometric AIC is given by

$$\text{G-AIC}_M = \hat{J}_M + \frac{6}{N}\epsilon^2. \tag{26}$$

5.6 Translations

A 2-D rigid motion reduces to a translation if we let $\theta = 0$. Let $\hat{J}_T$ be the resulting residual. The geometric AIC is given by

$$\text{G-AIC}_T = \hat{J}_T + \frac{4}{N}\epsilon^2. \tag{27}$$

(a) (b)

Fig. 5. Real images of an outdoor scene and the selected points.

(a) (b)

Fig. 6. (a) The image mapping computed by an optimal homography. (b) The image mapping by model selection.

5.7 Affine transformations

An affine transformation is a special homography that has the form

$$\boldsymbol{H} = \begin{pmatrix} a_{11} & a_{12} & t_1/f_0 \\ a_{21} & a_{22} & t_2/f_0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{28}$$

Optimal values of $\{a_{ij}\}$ and $\{t_i\}$ are obtained by the LM method, and an initial guess can be computed analytically (we omit the details). Let $\hat{J}_A$ be the resulting residual. The geometric AIC is given by

$$\text{G-AIC}_A = \hat{J}_A + \frac{12}{N}\epsilon^2. \tag{29}$$

5.8 Principle of model selection

The geometric AIC consists of the residual and the penalty term that is proportional to the degree of freedom of the model. The penalty term is determined by

Fig. 7. Two images of an outdoor scene and the selected points.

(a) (b)

Fig. 8. (a) Mosaicing by an optimally computed homography. (b) Mosaicing by model selection.

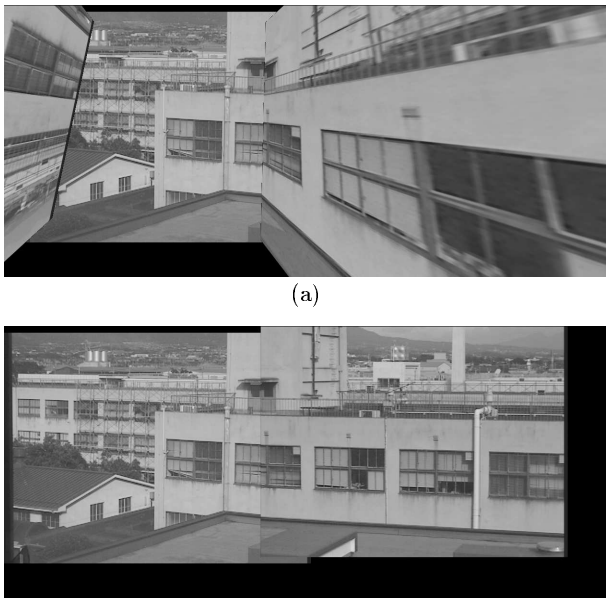
analyzing the decrease of the residual caused by overfitting the model parameters to noisy data [1, 4, 5]. Adopting the model with the smallest geometric AIC is equivalent to checking how much the residual will increase if the degree of the freedom of the model is reduced and adopting the simpler model *if the resulting increase of the residual is comparable to the decrease of the degree of freedom*, which can be interpreted as a symptom of overfitting.

6 Real Image Experiments

Fig. 5(a) is an image of an outdoor scene. Fig. 5(b) is a zoomed image of the same scene corresponding to the white frame in Fig. 5(a). We manually selected the seven points marked in the images and computed the homography for each of the candidate models described in the preceding section, using the default noise model. The computed geometric AICs of the candidate models are listed in Table 1. As we can see, the similarity model is preferred. Fig. 6(a) shows the resulting superimposed image using the homography computed by the optimal algorithm given in [7]. Fig. 6(b) is the result using the selected similarity.

Fig. 7 is a pair of images of an outdoor scene with a small overlap. Using the six points marked in the images, we computed the geometric AICs of the candidate models as shown in Table 1. Again, the similarity model is preferred.

Fig. 9. Two images of an outdoor scene and the selected points.



(a)

(b)

Fig. 10. (a) Mosaicing by an optimally computed homography. (b) Mosaicing by model selection.

Fig. 8(a) is the mosaiced image using the homography computed by the optimal algorithm. Fig. 8(b) is the result using the selected similarity.

Fig. 9 shows a different pair of images. Using the five points marked there, we computed the geometric AICs shown in Table 1, which indicate that the translation model is preferred. Fig. 10(a) is the mosaiced image using the optimal homography; Fig. 10(b) is the result using the selected translation.

Fig. 11 shows the same scene as Fig. 7. This time, we used twenty two points distributed over a large region. The resulting geometric AICs are listed in Table 1. The best model is the 8-parameter homography; the second best model

Fig. 11. Matching many points.

(a) (b)

Fig. 12. (a) Mosaicing using the best model. (b) Mosaicing using the second best model.

is the 5-parameter homography. The difference between their geometric AICs is very small, indicating that the image transformation can be viewed almost as the 5-parameter homography. Fig. 12(a) is the mosaiced image using the best model; Fig. 12(b) is the result using the second best model.

7 Concluding Remarks

As we can see from Figs. 8(a) and 10(a), the image mapping defined by the optimally computed homography is very unstable and can cause a large unnatural distortion if the matching points are small in number and concentrated in a small region in each image. Theoretically, the accuracy cannot be improved any further. We have shown that the accuracy *can* be improved nonetheless if we incorporate our *knowledge* about source of the instability.

The instability stems from the fact that actual transformations of images are usually in a small subgroup of the group of homographies. It follows that undesirable distortions can be removed by selecting an appropriate model by using the geometric AIC. The improvement is dramatic as demonstrated in Figs. 8(b) and 10(b). As Fig. 12 shows, model selection is not necessary if a large number of matching points are distributed over a large region, and the general 8-parameter

Table 1. The geometric AICs and the selected models.

Model	Fig. 5	Fig. 7	Fig. 9	Fig. 11
8-parameter homography	$9.92E-06$	$1.25E-05$	$4.01E-05$	◯ $2.946E-06$
5-parameter homography	$4.80E-02$	$3.65E-03$	$4.69E-05$	$2.954E-06$
4-parameter homography	$1.57E-02$	$4.39E-02$	$4.45E-05$	$2.976E-03$
affine transformation	$8.92E-06$	$1.08E-05$	$4.10E-05$	$3.507E-06$
similarity	◯ $7.32E-06$	◯ $8.54E-06$	$4.38E-05$	$4.887E-06$
rigid motion	$1.57E-02$	$3.55E-04$	$4.00E-05$	$2.976E-03$
translation	$1.57E-02$	$3.53E-04$	◯ $3.65E-05$	$2.990E-03$

homography is chosen if model selection is applied. Thus, an appropriate mapping is always selected whether a sufficient number of matching points are available or not. This selection process does not require any empirical thresholds to adjust. Our technique is very general and can be applied to a wide range of vision applications for increasing accuracy and preventing computational instability (e.g., [8]).

Acknowledgments. This work was in part supported by the Ministry of Education, Science, Sports and Culture, Japan under a Grant in Aid for Scientific Research C(2) (No. 11680377).

References

1. H. Akaike, A new look at the statistical model identification, *IEEE Trans. Automation Control*, **19**-6 (1974), 716–723.
2. W. Förstner, Reliability analysis of parameter estimation in linear models with applications to mensuration problems in computer vision, *Comput. Vision Graphics Image Process.*, **40** (1987), 273–310.
3. K. Kanatani, *Geometric Computation for Machine Vision*, Oxford University Press, Oxford, 1993.
4. K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier Science, Amsterdam, 1996.
5. K. Kanatani, Geometric information criterion for model selection, *Int. J. Comput. Vision*, **26**-3 (1998), 171–189.
6. K. Kanatani, Statistical optimization and geometric inference in computer vision, *Phil. Trans. Roy. Soc. Lond.*, A-**356** (1998), 1303–1320.
7. K. Kanatani and N. Ohta, Accuracy bounds and optimal computation of homography for image mosaicing applications, *Proc. 7th Int. Conf. Comput. Vision*, September, 1999, Kerkya, Greece, pp. 73–78.
8. C. Matsunaga and K. Kanatani, Calibration of a moving camera using a planar pattern: Optimal computation, reliability evaluation and stabilization by model selection, *Proc. 6th Euro. Conf. Comput. Vision*, June–July 2000, Dublin, Ireland, Vol.2, pp. 595–609.
9. 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.

10. H. S. Sawhney, S. Hsu and R. Kumar, Robust video mosaicing through topology inference and local to global alignment, *Proc. 5th Euro. Conf. Comput. Vision*, June 1998, Freiburg, Germany, Vol. 2, pp. 103–119.
11. J. Shi and C. Tomasi, Good features to track, *Proc. Conf. Comput. Vision Patt. Recogn.*, June 1994, Seattle, WA, pp. 593–600.
12. A. Singh, An estimation-theoretic framework for image-flow computation, *Proc. 3rd Int. Conf. Comput. Vision*, December, 1990, Osaka, Japan, pp. 168–177.
13. R. Szeliski and H.-U. Shum, Creating full view panoramic image mosaics and environment maps, *Proc. SIGGRAPH'97*, August 1997, Los Angeles, CA, U.S.A., pp. 251–258.
14. P. H. S. Torr, Model selection for two view geometry: A review, in, D. A. Forsyth, J. L. Mundy, V. D. Gesú, R. Cipolla (Eds.): *Shape, Contour and Grouping in Computer Vision*, LNCS 1681, Springer-Verlag, Berlin, 1999, pp. 277–301.
15. T. Werner, T. Pajdla and V. Hlaváč, Efficient 3-D scene visualization by image extrapolation, *Proc. 5th Euro. Conf. Comput. Vision*, June 1998, Freiburg, Germany, Vol. 2, pp. 382–396.
16. I. Zoghlami, O. Faugeras and R. Deriche, Using geometric corners to build a 2D mosaic from a set of images, *Proc. Conf. Comput. Vision Patt. Recogn.*, June 1997, Puerto Rico, pp. 420–425.

A Analytical decomposition

We first compute the homography $\boldsymbol{H} = (H_{ij})$ (up to scale) that maps $\{\boldsymbol{x}_\alpha\}$ to $\{\boldsymbol{x}'_\alpha\}$, say, by the optimal algorithm given in [7] or simply by least squares. The non-dimensional focal lengths ϕ' and ϕ and the rotation matrix $\boldsymbol{R}$ that satisfy eq. (11) are computed analytically by the following procedure. First, ϕ' and ϕ are given by

$$\phi' = \sqrt{-\frac{A}{B}}, \qquad \phi = \sqrt{\frac{H_{13}^2 + H_{23}^2 + H_{33}^2\phi'^2}{K}}, \tag{30}$$

where

$$A = (H_{11}H_{12} + H_{21}H_{22})H_{31}H_{32} + (H_{12}H_{13} + H_{22}H_{23})H_{32}H_{33} + (H_{13}H_{11} + H_{23}H_{21})H_{33}H_{31}, \tag{31}$$

$$B = H_{31}^2H_{32}^2 + H_{32}^2H_{33}^2 + H_{33}^2H_{31}^2, \tag{32}$$

$$K = \frac{H_{11}^2 + H_{21}^2 + H_{12}^2 + H_{22}^2 + (H_{31}^2 + H_{32}^2)\phi'^2}{2}. \tag{33}$$

Then, compute the following singular value decomposition:

$$\boldsymbol{F}^{-1}\boldsymbol{H}^\top\boldsymbol{F}' = \boldsymbol{V}\begin{pmatrix}\sigma_1 & & \\ & \sigma_2 & \\ & & \sigma_3\end{pmatrix}\boldsymbol{U}^\top. \tag{34}$$

Here, $\sigma_1 \geq \sigma_2 \geq \sigma_3$ (≥ 0) are the singular values, and $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices. The rotation matrix $\boldsymbol{R}$ is given by

$$\boldsymbol{R} = \boldsymbol{V}\begin{pmatrix}1 & & \\ & 1 & \\ & & \det(\boldsymbol{V}\boldsymbol{U}^\top)\end{pmatrix}\boldsymbol{U}^\top. \tag{35}$$

This procedure produces an exact solution if noise does not exist. In the presence of noise, the solution is optimal in the least squares sense.

B Gradient and Hessian

We put

$$\boldsymbol{e}_\alpha = \boldsymbol{x}'_\alpha \times \boldsymbol{H}\boldsymbol{x}_\alpha. \tag{36}$$

In computing the gradient ∇J of eq. (7), we ignore terms of $O(\boldsymbol{e}_\alpha)^2$ ($O(\cdots)^n$ denotes terms of order n or higher in $\cdots$). This is justified because J has the form $\sum_{\alpha=1}^N (\boldsymbol{e}_\alpha, \boldsymbol{W}_\alpha \boldsymbol{e}_\alpha)/N$ and hence ∇J is $O(\boldsymbol{e}_\alpha)$. In particular, $\boldsymbol{W}_\alpha$ can be regarded as a constant matrix, since the terms involving derivatives of $\boldsymbol{W}$ in ∇J are $O(\boldsymbol{e}_\alpha)^2$. This approximation causes only higher order errors in the solution of $\nabla J = 0$.

Under this approximation, the gradient of J with respect to ϕ, ϕ', and $\boldsymbol{R}$ is given by

$$\frac{\partial J}{\partial \phi} = -\frac{2}{\phi N} \sum_{\alpha=1}^N (\boldsymbol{H}\boldsymbol{k}, \boldsymbol{x}'_\alpha \times \boldsymbol{W}_\alpha \boldsymbol{e}_\alpha), \tag{37}$$

$$\frac{\partial J}{\partial \phi'} = \frac{2}{\phi' N} \sum_{\alpha=1}^N (\boldsymbol{k}, \boldsymbol{H}\boldsymbol{x}_\alpha)(\boldsymbol{k}, \boldsymbol{x}'_\alpha \times \boldsymbol{W}_\alpha \boldsymbol{e}_\alpha), \tag{38}$$

$$\nabla_\Omega J = \frac{2}{N} \sum_{\alpha=1}^N (\boldsymbol{F}\boldsymbol{x}_\alpha) \times \boldsymbol{F}^{-1}\boldsymbol{H}^\top (\boldsymbol{x}'_\alpha \times \boldsymbol{W}_\alpha \boldsymbol{e}_\alpha), \tag{39}$$

where $\boldsymbol{k} = (0, 0, 1)^\top$.

In computing the Hessian of $\nabla^2 J$ of eq. (7), we ignore terms of $O(\boldsymbol{e}_\alpha)$. This is justified because J has the form $\sum_{\alpha=1}^N (\boldsymbol{e}_\alpha, \boldsymbol{W}_\alpha \boldsymbol{e}_\alpha)/N$ and hence $\nabla^2 J$ is $O(1)$. In particular, $\boldsymbol{W}_\alpha$ can be regarded as a constant matrix, since the terms involving derivatives of $\boldsymbol{W}$ in $\nabla^2 J$ are $O(\boldsymbol{e}_\alpha)$. This approximation does not affect the accuracy of Newton iterations, since the Hessian $\nabla^2 J$ controls merely the speed of convergence, not the accuracy of the solution. Newton iterations with this approximation, called *Gauss-Newton iterations*, are known to be almost as efficient as Newton iterations.

Under this approximation, the individual elements of the Hessian $\nabla^2 J$ are given as follows:

$$\frac{\partial^2 J}{\partial \phi^2} = \frac{2}{\phi^2 N} \sum_{\alpha=1}^N (\boldsymbol{H}\boldsymbol{k}, (\boldsymbol{x}'_\alpha \times \boldsymbol{W}_\alpha \times \boldsymbol{x}'_\alpha)\boldsymbol{H}\boldsymbol{k}), \tag{40}$$

$$\frac{\partial^2 J}{\partial \phi'^2} = \frac{2}{\phi'^2 N} \sum_{\alpha=1}^N (\boldsymbol{k}, \boldsymbol{H}\boldsymbol{x}_\alpha)^2 (\boldsymbol{k}, (\boldsymbol{x}'_\alpha \times \boldsymbol{W}_\alpha \times \boldsymbol{x}'_\alpha)\boldsymbol{k}), \tag{41}$$

$$\frac{\partial^2 J}{\partial \phi \partial \phi'} = -\frac{2}{\phi \phi' N} \sum_{\alpha=1}^N (\boldsymbol{k}, \boldsymbol{H}\boldsymbol{x}_\alpha)(\boldsymbol{H}\boldsymbol{k}, (\boldsymbol{x}'_\alpha \times \boldsymbol{W}_\alpha \times \boldsymbol{x}'_\alpha)\boldsymbol{k}), \tag{42}$$

$$\nabla_\Omega \frac{\partial J}{\partial \phi} = -\frac{2}{\phi N} \sum_{\alpha=1}^{N} (\boldsymbol{F}\boldsymbol{x}_\alpha) \times \boldsymbol{F}^{-1}\boldsymbol{H}^\top (\boldsymbol{x}'_\alpha \times \boldsymbol{W}_\alpha \times \boldsymbol{x}'_\alpha)\boldsymbol{H}\boldsymbol{k}, \tag{43}$$

$$\nabla_\Omega \frac{\partial J}{\partial \phi'} = \frac{2}{\phi N} \sum_{\alpha=1}^{N} (\boldsymbol{k}, \boldsymbol{H}\boldsymbol{k})(\boldsymbol{F}\boldsymbol{x}_\alpha) \times \boldsymbol{F}^{-1}\boldsymbol{H}^\top (\boldsymbol{x}'_\alpha \times \boldsymbol{W}_\alpha \times \boldsymbol{x}'_\alpha)\boldsymbol{k}, \tag{44}$$

$$\nabla^2_\Omega J = \frac{2}{N} \sum_{\alpha=1}^{N} (\boldsymbol{F}\boldsymbol{x}_\alpha) \times \boldsymbol{F}^{-1}\boldsymbol{H}^\top (\boldsymbol{x}'_\alpha \times \boldsymbol{W}_\alpha \times \boldsymbol{x}'_\alpha)\boldsymbol{H}\boldsymbol{F}^{-1} \times (\boldsymbol{F}\boldsymbol{x}_\alpha). \tag{45}$$

Here, the product $\boldsymbol{a} \times \boldsymbol{T} \times \boldsymbol{a}$ of a vector $\boldsymbol{a} = (a_i)$ and matrix $\boldsymbol{T} = (T_{ij})$ is a symmetric matrix whose (ij) element is $\sum \varepsilon_{ikl}\varepsilon_{jmn}a_k a_m T_{ln}$, where ε_{ijk} is the Eddington epsilon, taking 1, -1, and 0 when (ijk) is an even permutation of (123), an odd permutation of it, and otherwise, respectively.

C Analytical similarity solution

We represent the coordinates (x_α, y_α) and (x'_α, y'_α) and the translation $\vec{\tau} = (\tau_1, \tau_2)^\top$ by the following complex numbers:

$$z_\alpha = \frac{x_\alpha}{f_0} + i\frac{y_\alpha}{f_0}, \qquad z'_\alpha = \frac{x'_\alpha}{f_0} + i\frac{y'_\alpha}{f_0}, \qquad \tau = \tau_1 + i\tau_2. \tag{46}$$

Let z_C and z'_C be the centroids of the feature points:

$$z_C = \frac{1}{N}\sum_{\alpha=1}^{N} z_\alpha, \qquad z'_C = \frac{1}{N}\sum_{\alpha=1}^{N} z'_\alpha. \tag{47}$$

Compute the deviations of the individual feature points from the centroids:

$$\tilde{z}_\alpha = z_\alpha - z_C, \qquad \tilde{z}'_\alpha = z'_\alpha - z'_C. \tag{48}$$

The scale s and the angle θ of rotation are given by

$$s = \frac{1}{N}\sum_{\alpha=1}^{N} \mathrm{abs}[\frac{\tilde{z}'_\alpha}{\tilde{z}_\alpha}], \qquad \theta = \arg[\frac{1}{N}\sum_{\alpha=1}^{N} S[\frac{\tilde{z}'_\alpha}{\tilde{z}_\alpha}]], \tag{49}$$

where we define

$$S[Z] = \frac{Z}{\mathrm{abs}[Z]}. \tag{50}$$

The translation τ is given by

$$\tau = z'_C - se^{i\theta} z_C. \tag{51}$$

This article was processed using the LaTeX macro package with LLNCS style