１９９９年情報論的学習理論ワークショップ
1999 Workshop on Information-Based Induction Sciences (IBIS'99)
Izu, Japan, August 26-27, 1999.

Stabilizing Image Mosaicing by the Geometric AIC

Kenichi Kanatani * Yasushi Kanazawa †

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. In this paper, we show that such undesirable distortions can be removed by model selection using the geometric AIC without introducing any empirical thresholds. We also present an optimization scheme based on the Levenberg-Marquardt method and an analytical procedure for computing an initial guess. We demonstrate the effectiveness of our method by real image examples.
Key words: homography, image mosaicing, panoramic image, statistical optimization, model selection, geometric AIC

1. Introduction

Image mosaicing is a technique for integrating multiple images into one continuous image, often referred to as a *panoramic image* [11, 14, 15, 16]. This technique has long been used for creating 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]. Because far-away 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, only a small number of feature points are available to match. In such a case, the selected feature 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 feature points, a large distortion may occur even in the presence of very small noise.

In [7], we have observed this instability using real images and presented an algorithm for computing a homography from point correspondences by a technique called *renormalization*, which not only produces a statistically optimal solution but also evaluates the reliability of the computed solution in quantitative terms. In [8], we have derived a theoretical accuracy bound for the homography computation and confirmed by experiments that our algorithm indeed attains that bound. In this sense, our algorithm is strictly optimal; there is no room for further improvement.

Although our algorithm dramatically reduces the instability of the mapping, as demonstrated in [7, 8], it cannot remove the distortion completely. For further improvement, we need to examine 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 it, 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 this paper, 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 an optimization scheme based on the Levenberg-Marquardt method and an analytical procedure for computing an initial guess. We demonstrate the effectiveness of our method by real image examples.

2. Representation of Homography

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

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

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

$$\boldsymbol{x} = \begin{pmatrix} x/f \\ y/f \\ 1 \end{pmatrix}, \quad \boldsymbol{x}' = \begin{pmatrix} x'/f \\ y'/f \\ 1 \end{pmatrix}, \tag{2}$$

$$\boldsymbol{H} = \begin{pmatrix} A & B & C/f \\ D & E & F/f \\ P/f & Q/f & R/f^2 \end{pmatrix}, \tag{3}$$

eq. (1) can be written as

$$\boldsymbol{x}' = Z[\boldsymbol{H}\boldsymbol{x}]. \tag{4}$$

* Department of Computer Science, Gunma University, Kiryu, Gunma 376-8515 Japan, Tel: (0277)30-1842, Fax: (0277)30-1801, E-mail: kanatani@cs.gunma-u.ac.jp

† Department of Knowledge-Based Information Engineering, Toyohashi University of Technology, Toyohashi, Aichi 441-8580 Japan, Tel: (0532)44-6888, Fax: (0532)44-6873, E-mail: kanazawa@tutkie.tut.ac.jp

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 feature points that correspond to each other. 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 feature points, the uncertainty of matching is characterized by the variation of the matching residual around the detected point [2, 10, 12, 13]. Hence, we assume that the covariance matrices $V[\boldsymbol{x}_\alpha]$ and $V[\boldsymbol{x}'_\alpha]$ are known *up to scale* and write

$$V[\boldsymbol{x}_\alpha] = \epsilon^2 V_0[\boldsymbol{x}_\alpha], \quad V[\boldsymbol{x}'_\alpha] = \epsilon^2 V_0[\boldsymbol{x}'_\alpha]. \tag{5}$$

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 prior 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. (4) can equivalently be written in the form

$$\boldsymbol{x}' \times \boldsymbol{H}\boldsymbol{x} = \boldsymbol{0}. \tag{6}$$

In the presence of Gaussian noise, an optimal estimate of $\boldsymbol{H}$ is obtained by minimizing the squared *Mahalanobis distances*

$$J = \frac{1}{N}\sum_{\alpha=1}^{N} (\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}_\alpha, V_0[\boldsymbol{x}_\alpha]_2^-(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}_\alpha)) + \frac{1}{N}\sum_{\alpha=1}^{N} (\boldsymbol{x}'_\alpha - \bar{\boldsymbol{x}}'_\alpha, V_0[\boldsymbol{x}'_\alpha]_2^-(\boldsymbol{x}'_\alpha - \bar{\boldsymbol{x}}'_\alpha)), \tag{7}$$

subject to the constraint (6). 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 (6) and express eq. (7) in the following form [4]:

$$J = \frac{1}{N}\sum_{\alpha=1}^{N} (\boldsymbol{x}'_\alpha \times \boldsymbol{H}\boldsymbol{x}_\alpha, \boldsymbol{W}_\alpha(\boldsymbol{x}'_\alpha \times \boldsymbol{H}\boldsymbol{x}_\alpha)), \tag{8}$$

$$\boldsymbol{W}_\alpha = \Big(\boldsymbol{x}'_\alpha \times \boldsymbol{H}V_0[\boldsymbol{x}_\alpha]\boldsymbol{H}^\top \times \boldsymbol{x}'_\alpha + (\boldsymbol{H}\boldsymbol{x}_\alpha) \times V_0[\boldsymbol{x}'_\alpha] \times (\boldsymbol{H}\boldsymbol{x}_\alpha)\Big)_2^-. \tag{9}$$

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

$$\hat{\epsilon}^2 = \frac{\hat{J}}{2(1 - 4/N)}. \tag{10}$$

4. Models of Image Transformations

Since the elements of $\boldsymbol{H}$ has scale indeterminacy (see eq. (4)), 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 far-away 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. 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 of it.

If the camera translates relative to a nearby scene, we have the group of translations with two degrees of freedom. 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. If the focal length is also allowed to change, we have the group of similarities with four degrees of freedom. It can be generalized to the group of affine transformations with six degrees of freedom.

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. But this does not work, because the 8-parameter homography group is always chosen: in general a model with more degrees of freedom has a smaller residual. For a fair comparison, we need to compensate for the overfit effect 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 AIC is preferred.

5. Subgroup Hierarchy

5.1 5-parameter homographies

(a) (b) (c)

Figure 1: (a) Real images of an outdoor scene and the selected feature points. (b) Mapping computed by an optimally computed homography. (c) Mapping computed by model selection.

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

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

We use the Levenberg-Marquardt method (LM method) for minimizing eq. (8). First, we define the following non-dimensional variables:

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

The minimization procedure goes as follows:

1. Let c=0.001. Analytically compute initial guesses of ϕ, ϕ', and $\boldsymbol{R}$ [9] (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):

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

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

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

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

Table 1: Geometric AIC for Fig. 1

8-parameter homography	$9.92E-06$
5-parameter homography	$4.80E-02$
4-parameter homography	$1.57E-02$
similarity	$7.32E-06$
rigid motion	$1.57E-02$
translation	$1.57E-02$
affine transformation	$8.92E-06$

4. Compute the residual $J' = J(\phi + \Delta\phi, \phi' + \Delta\phi', \mathcal{R}(\Delta\boldsymbol{\Omega}))$. 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

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

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

$$AIC_{H_5} = \hat{J}_{H_5} + \frac{10}{N}\epsilon^2, \tag{18}$$

where ϵ^2 is estimated by eq. (10).

5.2 4-parameter homographies

If we let $f = f'$ in eq. (11), we obtain a 4-parameter group of homographies model, 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

$$AIC_{H_4} = \hat{J}_{H_4} + \frac{8}{N}\epsilon^2. \tag{19}$$

5.3 Similarities

A similarity is a special homography that has the form

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

Figure 2: Two images of an outdoor scene and the selected feature points.

Table 2: Geometric AIC for Fig. 2.

8-parameter homography	$1.25E-05$
5-parameter homography	$3.65E-03$
4-parameter homography	$4.39E-02$
similarity	$8.54E-06$
rigid motion	$3.55E-04$
translation	$3.53E-04$
affine transformation	$1.08E-05$

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

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

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

eq. (8) is rewritten in the following form:

$$J = \frac{1}{N}\sum_{\alpha=1}^{N} (\vec{x}'_\alpha - sR\vec{x}_\alpha - \vec{t}, W_\alpha(\vec{x}'_\alpha - sR\vec{x}_\alpha - \vec{t})), \tag{23}$$

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

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

$$AIC_S = \hat{J}_S + \frac{8}{N}\epsilon^2. \tag{25}$$

5.4 Rigid motions

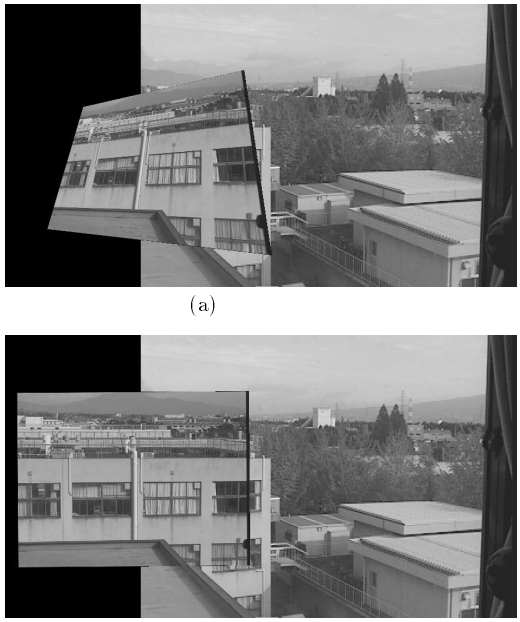
The image transformation reduces to a 2-D rigid motion if we let $s = 1$ in eq. (20). We can apply the same LM method for minimizing J and the procedure for computing an initial guess after an appropriate modification. Let $\hat{J}_M$ be the resulting residual. The geometric AIC is given by

$$AIC_M = \hat{J}_M + \frac{6}{N}\epsilon^2. \tag{26}$$

5.5 Translations

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

$$AIC_T = \hat{J}_T + \frac{4}{N}\epsilon^2. \tag{27}$$



(a)

(b)

Figure 3: (a) Mosaicing by an optimally computed homography. (b) Mosaicing by model selection.

5.6 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

$$AIC_A = \hat{J}_A + \frac{12}{N}\epsilon^2. \tag{29}$$

6. Real Image Experiments

Fig. 1(a) is an image of an outdoor scene. We manually selected feature points marked in the image and the corresponding ones in a zoomed image of the same scene. We computed the homography from these for each of the candidate models described in the preceding section by using the default noise model. Table 1 lists the geometric AICs of the candidate models. As we can see, the similarity model is preferred. Fig. 1(b) shows the resulting superimposed image via by the homography computed by the optimal algorithm of Kanatani [7, 8]. Fig. 1(c) is the result obtained by the selected similarity.

Fig. 2 is a pair of images of an outdoor scene with a small overlap. Table 2 lists the geometric AICs of the candidate models for the feature points marked in the images. Again, the similarity model is preferred. Fig. 3(a) is the mosaiced images obtained by applying the homography computed by the optimal algorithm of Kanatani [7, 8]. Fig. 3(b) is the result obtained by the selected similarity.

Figure 4: Two images of an outdoor scene and the selected feature points.

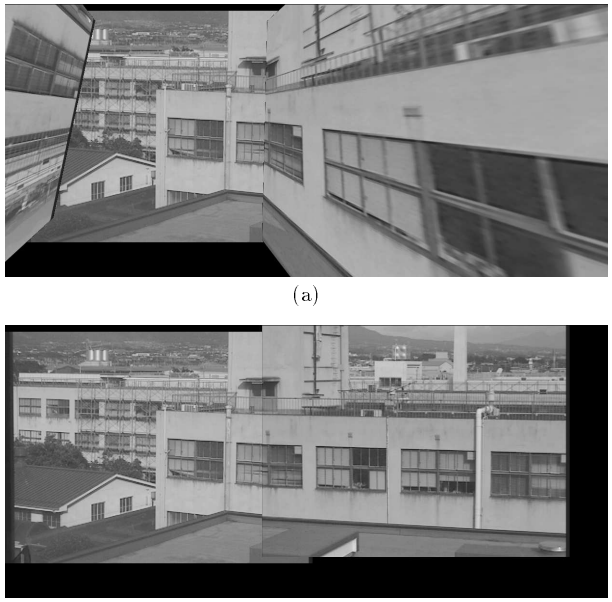
Table 3: Geometric AIC for Fig. 4.

8-parameter homography	$4.01E-05$
5-parameter homography	$4.69E-05$
4-parameter homography	$4.45E-05$
similarity	$4.38E-05$
rigid motion	$4.00E-05$
translation	$3.65E-05$
affine translation	$4.10E-05$

Fig. 4 shows a different pair of images, and Table 3 is the list of the geometric AICs. This time, the translation model is preferred. Fig. 5(a) is the mosaiced image obtained by the optimal homography; Fig. 5(b) is the result obtained by the selected translation.

7. Concluding Remarks

As we can see from Figs. 3(a) and 5(a), the image mapping given 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. This instability stems from the fact that actual transformations of images are usually in a small subgroup of the group of homographies. Such undesirable distortions can be removed by model selection using the geometric AIC, as demonstrated in Figs. 3(b) and 5(b). The selection is completely automated without introducing any empirical thresholds to adjust.



(a)

(b)

Figure 5: (a) Mosaicing by an optimally computed homography. (b) Mosaicing by model selection.

References

[1] H. Akaike, "A new look at the statistical model identification," IEEE Trans. Automation Control, vol.19, no.6, pp.176–723, 1974.

[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, Optimal homography computation with a reliability measure, *Proc. IAPR Workshop on Machine Vision Applications*, November 1998, Makuhari, Japan, pp. 426–429.

[8] K. Kanatani and N. Ohta, Accuracy bounds and optimal computation of homography for image mosaicing applications, *Proc. Int. Conf. Comput. Vision*, September 1999, Kerkyra, Greece, to appear.

[9] C. Matsunaga and K. Kanatani, Calibration of a moving camera using a planar pattern, *5th Symposium on Sensing via Image Information (SII'99)*, June 1999, Yokohama, Japan.

[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] 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.

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

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

[14] 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.

[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. IEEE Conf. Comput. Vision Patt. Recog.*, June 1997, Puerto Rico, pp. 420–425.

Appendices

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 of Kanatani [7, 8] 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

$$\phi' = \sqrt{-\frac{A}{B}}, \tag{30}$$

$$\phi = \sqrt{\frac{H_{13}^2 + H_{23}^2 + H_{33}^2 \phi'^2}{K}}, \tag{31}$$

where

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

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

$$K = \frac{H_{11}^2 + H_{21}^2 + H_{12}^2 + H_{22}^2 + (H_{31}^2 + H_{32}^2)\phi'^2}{2}. \tag{34}$$

Then, compute the following singular value decomposition:

$$\boldsymbol{F}^{-1}\boldsymbol{H}^\top\boldsymbol{F}' = \boldsymbol{V}\begin{pmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \sigma_3 \end{pmatrix}\boldsymbol{U}^\top. \tag{35}$$

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

$$\boldsymbol{R} = \boldsymbol{V}\begin{pmatrix} 1 & & \\ & 1 & \\ & & \det(\boldsymbol{V}\boldsymbol{U}^\top) \end{pmatrix}\boldsymbol{U}^\top. \tag{36}$$

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

B. Gradient and Hessian

We put

$$\boldsymbol{e}_\alpha = \boldsymbol{x}'_\alpha \times \boldsymbol{H}\boldsymbol{x}_\alpha. \tag{37}$$

In computing the gradient ∇J of eq. (8), 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

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

$$\frac{\partial J}{\partial \phi'} = \frac{2}{\phi' N}\sum_{\alpha=1}^N (\boldsymbol{k}, \boldsymbol{H}\boldsymbol{x}_\alpha)(\boldsymbol{k}, \boldsymbol{x}'_\alpha \times \boldsymbol{W}_\alpha \boldsymbol{e}_\alpha), \tag{39}$$

$$\nabla_{\boldsymbol{\Omega}} J = \frac{2}{N}\sum_{\alpha=1}^N (\boldsymbol{F}\boldsymbol{x}_\alpha) \times \boldsymbol{F}^{-1}\boldsymbol{H}^\top (\boldsymbol{x}'_\alpha \times \boldsymbol{W}_\alpha \boldsymbol{e}_\alpha), \tag{40}$$

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

In computing the Hessian of $\nabla^2 J$ of eq. (8), 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:

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

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

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

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

$$\begin{aligned} \nabla_{\boldsymbol{\Omega}} \frac{\partial J}{\partial \phi'} = & \frac{2}{\phi N}\sum_{\alpha=1}^N (\boldsymbol{k}, \boldsymbol{H}\boldsymbol{k})(\boldsymbol{F}\boldsymbol{x}_\alpha) \times \boldsymbol{F}^{-1}\boldsymbol{H}^\top \\ & (\boldsymbol{x}'_\alpha \times \boldsymbol{W}_\alpha \times \boldsymbol{x}'_\alpha)\boldsymbol{k}, \end{aligned} \tag{45}$$

$$\begin{aligned} \nabla^2_{\boldsymbol{\Omega}} J = & \frac{2}{N}\sum_{\alpha=1}^N (\boldsymbol{F}\boldsymbol{x}_\alpha) \times \boldsymbol{F}^{-1}\boldsymbol{H}^\top (\boldsymbol{x}'_\alpha \times \boldsymbol{W}_\alpha \times \boldsymbol{x}'_\alpha) \\ & \boldsymbol{H}\boldsymbol{F}^{-1} \times (\boldsymbol{F}\boldsymbol{x}_\alpha). \end{aligned} \tag{46}$$

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:

$$z_\alpha = \frac{x_\alpha}{f_0} + i\frac{y_\alpha}{f_0}, \quad z'_\alpha = \frac{x'_\alpha}{f_0} + i\frac{y'_\alpha}{f_0}, \quad \tau = \tau_1 + i\tau_2. \tag{47}$$

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

$$z_C = \frac{1}{N}\sum_{\alpha=1}^N z_\alpha, \quad z'_C = \frac{1}{N}\sum_{\alpha=1}^N z'_\alpha. \tag{48}$$

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

$$\tilde{z}_\alpha = z_\alpha - z_C, \quad \tilde{z}'_\alpha = z'_\alpha - z'_C. \tag{49}$$

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

$$s = \frac{1}{N}\sum_{\alpha=1}^N \mathrm{abs}[\frac{\tilde{z}'_\alpha}{\tilde{z}_\alpha}], \quad \theta = \arg[\frac{1}{N}\sum_{\alpha=1}^N S[\frac{\tilde{z}'_\alpha}{\tilde{z}_\alpha}]], \tag{50}$$

where we define

$$S[Z] = \frac{Z}{\mathrm{abs}[Z]}. \tag{51}$$

The translation τ is given by

$$\tau = z'_C - s e^{i\theta} z_C. \tag{52}$$