ELSEVIER

Image and Vision Computing 22 (2004) 93–103

image AND vision COMPUTING

www.elsevier.com/locate/imavis

Image mosaicing by stratified matching

Yasushi Kanazawa[a,*], Kenichi Kanatani[b]

[a]*Department of Knowledge-based Information Engineering, Toyohashi University of Technology, Toyohashi, Aichi 441-8580, Japan*
[b]*Department of Information Technology, Okayama University, Okayama 700-8530, Japan*

Received 23 September 2002; received in revised form 27 June 2003; accepted 2 July 2003

Abstract

We present a robust method for automatically matching points over two images for image mosaicing: after extracting feature points using a feature detector, we progressively estimate the rotation, the scale change, and the projective distortion between the two images by random voting and variable template matching. Using real images, we demonstrate that our method allows accurate image mosaicing even when conventional methods fail.

Keywords: Template matching; Feature point correspondence; Robust estimation; LMedS; Image mosaicing; Homography

1. Introduction

Establishing point correspondences over multiple images is the first step of many video processing applications. Two approaches exist for this purpose: tracking correspondences over successive image frames, and direct matching between separate image frames. This paper focuses on the latter.

The basic principle is local correlation measurement by template matching, but many incorrect matches, or *outliers*, remain. To remove them, we need to apply a robust estimation technique based on a geometric constraint such as the *epipolar equation* [1,6,26]. However, standard techniques such as LMedS [19] and RANSAC [4] do not work unless the initial matches are sufficiently accurate.

In order to resolve this problem, many techniques combining template matching, geometric constraints, multiresolution representation, random sampling and voting, and various types of heuristics have been proposed [2,12–17,22, 24,25]. Still, it is very difficult to match images of a general scene if the relationship between the two images is completely unknown.

In this paper, we specifically focus on image mosaicing applications [10,23,27], for which the two images are related by a transformation called *homography*. Since the entire image undergoes the same transformation (we later allow some parts to deform differently), image matching reduces to estimation of the eight parameters of the homography.

Although the image transformation induced by a homography has globally the same mathematical expression, the actual image distortion can greatly differ from location to location, so it is very difficult to find correspondences by local correlation measurement alone.

In this paper, we introduce a hierarchical scheme: we progressively estimate image distortions by random voting followed by variable template matching compatible with the estimated distortions. In order to distinguish this approach from the traditional multiresolution method [2], we call it *stratified matching*.

The multiresolution method may reduce the computation time for exhaustive search of region matching. For a fixed number of given feature points, however, it is powerless to reduce the combinatorial complexity; lowering the resolution only results in poor matching capability.

In Section 2, we discuss the method of our local correlation measurement. Section 3 describes the computational procedure for our stratified matching. There, we also discuss the merits and limitations of our approach. In Section 4, we show real image examples to demonstrate that our method allows robust image mosaicing even when conventional methods fail. Our automatic thresholding scheme [8] is summarized in Appendix A. The details of the image transformation computations used in Section 3 are

* Corresponding author. Tel.: +81-532-44-6888; fax: +81-532-44-6873.
E-mail addresses: kanazawa@tutkie.tut.ac.jp (Y. Kanazawa), kanatani@suri.it.okayama-u.ac.jp (K. Kanatani).

0262-8856/$ - see front matter © 2003 Published by Elsevier B.V.
doi:10.1016/j.imavis.2003.07.001

given in Appendix B. Our LMedS procedure, which is an extension of the standard LMedS [19] to the general geometric fitting problem, is described in Appendix C.

2. Template matching

Suppose we detect points $P_1, ..., P_N$ in the first image I_1 and points $Q_1, ..., Q_M$ in the second image I_2, using a feature detector [3,5,18,20,21]. We measure the similarity between two points P_α and Q_β by the *residual* (*sum of squares*)

$$J(\alpha, \beta) = \sum_{(i,j)\in\mathcal{N}} |T_{P_\alpha}(i,j) - I_2(i', j')|^2, \tag{1}$$

where $T_{P_\alpha}(i,j)$ is the template obtained by cutting out a square grid $\mathcal{N}$ centered on the point P_α in the first image I_1. We identify the center of $\mathcal{N}$ with the origin (0, 0). The pixel (i', j') to which the template pixel (i, j) is matched in the second image I_2 is given by

$$\begin{pmatrix} i' \\ j' \end{pmatrix} = \begin{pmatrix} x'_\beta \\ y'_\beta \end{pmatrix} + \begin{pmatrix} i \\ j \end{pmatrix}, \tag{2}$$

where (x'_β, y'_β) are the image coordinates of Q_α in I_2, so the template origin (0, 0) is matched to $(i', j') = (x'_\beta, y'_\beta)$. If any pixel (i', j') is outside the image frame of I_2, we regard $J(\alpha, \beta)$ as ∞, which in the actual computation we interpret to be a very large value.

If one image is a translated copy of the other, the neighborhoods of corresponding points should exactly match in the absence of noise. Under a homography, however, the residual $J(\alpha, \beta)$ for corresponding points P_α and Q_β is not 0 due to local image distortions even when no image noise exists.

Such image distortions can be absorbed by distorting the template $\mathcal{N}$ according to the same homography, but we do not know the homography a priori. So, we progressively estimate it as follows.

First, we use the standard template to estimate an approximate image translation. Next, we estimate scale changes and rotations by *similarity template matching*. The image transformation is further refined by *affine template matching*. Finally, we establish the correspondence by *homography template matching*. In each stage, we gradually expand the template and remove outliers using LMedS.

3. Stratified matching

3.1. Initial matching

For the N points $\{P_\alpha\}$ in I_1 and the M points $\{Q_\alpha\}$ in I_2, we compute the residuals $J(\alpha, \beta)$ for all possible pairs using a 9×9 template and apply the automatic thresholding procedure (Appendix A). Then, we enforce the uniqueness of matching. Many algorithms are known for obtaining globally optimal combinations using backtracking. Here, we adopt the following greedy algorithm for the sake of efficiency.

We search the $N \times N$ table of $J(\alpha, \beta)$ for the minimum value $J(\alpha^*, \beta^*)$ and establish the match between points P_{α^*} and Q_{β^*}. Then, we remove from the table the column and row that contain the value $J(\alpha^*, \beta^*)$ and apply the same procedure to the resulting $(N-1) \times (N-1)$ table. Repeating this, we end up with at most $\min(N, M)$ matches, which we use as initial candidates.

Let (x_α, y_α) and (x'_β, y'_β) be the image coordinates of points P_α and Q_β to be matched. We represent them by vectors

$$\mathbf{x}_\alpha = \begin{pmatrix} x_\alpha/f_0 \\ y_\alpha/f_0 \\ 1 \end{pmatrix}, \qquad \mathbf{x}'_\beta = \begin{pmatrix} x'_\beta/f_0 \\ y'_\beta/f_0 \\ 1 \end{pmatrix}, \tag{3}$$

where f_0 is an appropriate scale constant, e.g. the image size, so chosen that x_α/f_0, y_α/f_0, x'_β/f_0, and y'_β/f_0 have the order of 1. We denote the pair of points P_α and Q_β by (α, β).

3.2. Translation matching by one-point voting

We iterate the following computations with $S_m = \infty$ and $\mathbf{t}_m = \mathbf{0}$ as initial values:

1. Randomly sample one pair (a, b) from among the candidate pairs $\{(\alpha, \beta)\}$.
2. Compute the vector
$$\mathbf{t} = \mathbf{x}'_b - \mathbf{x}_a, \tag{4}$$
where $\mathbf{x}_a$ and $\mathbf{x}'_b$ are the vector representations (see Eq. (3)) of P_a and Q'_b, respectively (this convention is understood throughout this paper).
3. Sort the candidate pairs $\{(\alpha, \beta)\}$ with respect to
$$D = \tfrac{1}{2}\|\mathbf{x}'_\beta - \mathbf{x}_\alpha - \mathbf{t}\|^2, \tag{5}$$
and compute its median S.
4. If $S < S_m$, update S_m and $\mathbf{t}_m$: $S_m \leftarrow S$, $\mathbf{t}_m \leftarrow \mathbf{t}$.

Repeat the above computations until the median S_m reaches its minimum.[1] Then, do the following computations:

1. Regard those among the candidate pairs $\{(\alpha, \beta)\}$ that satisfy
$$\tfrac{1}{2}\|\mathbf{x}'_\beta - \mathbf{x}_\alpha - \mathbf{t}_m\|^2 < 7S_m \tag{6}$$
as inliers (Appendix C).
2. Compute the vector
$$\mathbf{t} = \frac{1}{N}\sum(\mathbf{x}'_\beta - \mathbf{x}_\alpha) \tag{7}$$

[1] In our experiment, we exhaustively searched all the candidate pairs.

as a representative translation, where $\sum$ sums the selected inliers and N is the number of them.

3. Discard the candidate pairs $\{(\alpha, \beta)\}$, and select from all the points $\{P_\alpha\}$ and $\{Q_\beta\}$ those pairs $\{(\alpha, \beta)\}$ that satisfy

$$\tfrac{1}{2}\|\mathbf{x}'_\beta - \mathbf{x}_\alpha - \mathbf{t}\|^2 < 7S_m. \tag{8}$$

4. From the selected pairs $\{(\alpha, \beta)\}$, remove those that have large residuals, using the automatic thresholding scheme (Appendix A).

The resulting pairs $\{(\alpha, \beta)\}$ are regarded as new candidates.

3.3. Similarity matching by two-point voting

We represent points with image coordinates (x_α, y_α) and (x'_β, y'_β) by complex numbers

$$z_\alpha = x_\alpha + \mathrm{i}y_\alpha, \qquad z'_\beta = x'_\beta + \mathrm{i}y'_\beta, \tag{9}$$

where i is the imaginary unit. We iterate the following computations with $S_m = \infty$, $z_m = 0$, $z'_m = 0$, $Z_m = 1$, and $s_m = 1$ as initial values:

1. Randomly sample two pairs (a_0, b_0) and (a_1, b_1) from among the candidate pairs $\{(\alpha, \beta)\}$.
2. Compute the complex number Z and the real number s

$$Z = \frac{z'_{b_1} - z'_{b_0}}{z_{a_1} - z_{a_0}}, \qquad s = |Z|, \tag{10}$$

   where $|\cdot|$ denotes the absolute value of a complex number.
3. Sort the candidate pairs $\{(\alpha, \beta)\}$ with respect to

$$D = \frac{|z'_\beta - z'_{b_0} - Z(z_\alpha - z_{a_0})|^2}{1 + s^2}, \tag{11}$$

   and compute its median S.
4. If $S < S_m$, update S_m, z_m, z'_m, Z_m, and s_m : $S_m \leftarrow S$, $z_m \leftarrow z_a$, $z'_m \leftarrow z_b$, $Z_m \leftarrow Z$, $s_m \leftarrow s$.

Repeat the above computations until the median S_m reaches its minimum.[2] Then, do the following computations:

1. Regard those among the candidate pairs $\{(\alpha, \beta)\}$ that satisfy

$$\frac{|z'_\beta - z'_m - Z_m(z_\alpha - z_m)|^2}{1 + s_m^2} < 7S_m \tag{12}$$

   as inliers (Appendix C).
2. Optimally fit a similarity transformation to the selected inliers in the form

$$\vec{x}'_\beta = sR\vec{x}_\alpha + \vec{t}, \tag{13}$$

   where $\vec{x}_\alpha$ and $\vec{x}'_\beta$ are two-dimensional vectors consisting of the image coordinates of points P_α and P_β, respectively, and s, R, and $\vec{t}$ are the scale constant, the two-dimensional rotation matrix, and the two-dimensional translation vector, respectively (Appendix B).
3. Discard the candidate pairs $\{(\alpha, \beta)\}$, and select from all the points $\{P_\alpha\}$ and $\{Q_\beta\}$ those pairs $\{(\alpha, \beta)\}$ that satisfy

$$\frac{\|\vec{x}'_\beta - sR\vec{x}_\alpha - \vec{t}\|^2}{1 + s^2} < 7S_m. \tag{14}$$

4. Apply to the selected pairs $\{(\alpha, \beta)\}$ the *similarity template matching*: we enlarge the template $\mathcal{N}$ in Eq. (1) to 17×17 and compute (i', j'), instead of Eq. (2), by[3]

$$\begin{pmatrix} i' \\ j' \end{pmatrix} = \begin{pmatrix} x'_\beta \\ y'_\beta \end{pmatrix} + sR\begin{pmatrix} i \\ j \end{pmatrix}, \tag{15}$$

   and remove those pairs $\{(\alpha, \beta)\}$ that have large residuals, using the automatic thresholding scheme (Appendix A).

The resulting pairs $\{(\alpha, \beta)\}$ are regarded as new candidates.

3.4. Affine matching by three-point voting

We iterate the following computations with $S_m = \infty$, $\mathbf{A}_m = \mathbf{I}$, and $\mathbf{W} = \mathbf{I}$ as initial values ($\mathbf{I}$ denotes the unit matrix):

1. Randomly sample three pairs (a_0, b_0), (a_1, b_1), and (a_2, b_2) from among the candidate pairs $\{(\alpha, \beta)\}$.
2. Compute the matrix

$$\mathbf{A} = (\mathbf{x}'_{b_0}\ \mathbf{x}'_{b_1}\ \mathbf{x}'_{b_2})(\mathbf{x}_{a_0}\ \mathbf{x}_{a_1}\ \mathbf{x}_{a_2})^{-1}. \tag{16}$$

3. Sort the candidate pairs $\{(\alpha, \beta)\}$ with respect to

$$D = (\mathbf{x}'_\beta - \mathbf{A}\mathbf{x}_\alpha, \mathbf{W}(\mathbf{x}'_\beta - \mathbf{A}\mathbf{x}_\alpha)) \tag{17}$$

   and compute its median S. Hereafter, $(\mathbf{a}, \mathbf{b})$ denotes the inner product of vectors $\mathbf{a}$ and $\mathbf{b}$. The matrix $\mathbf{W}$ in Eq. (17) is defined by

$$\mathbf{W} = \begin{pmatrix} & & 0 \\ & \mathbf{W} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad W = (I + AA^{\mathrm{T}})^{-1}, \tag{18}$$

   where I denotes the two-dimensional unit matrix, and A is the top-left 2×2 supmatrix of $\mathbf{A}$.

[2] In our experiment, we stopped the search when no update occurred 100 times consecutively. This criterion was used in the subsequent iterations, too.

[3] The resulting coordinates (i', j') are no longer integers. In all computations involving non-integer image coordinates, we determined the pixel value by bilinear interpolation.

4. If $S < S_m$, update S_m, $\mathbf{A}_m$, and $\mathbf{W}_m$: $S_m \leftarrow S$, $\mathbf{A}_m \leftarrow \mathbf{A}$, $\mathbf{W}_m \leftarrow \mathbf{W}$.

Repeat the above computations until the median S_m reaches its minimum. Then, do the following computations:

1. Regard those among the candidate pairs $\{(\alpha, \beta)\}$ that satisfy

$$(\mathbf{x}'_\beta - \mathbf{A}_m\mathbf{x}_\alpha, \mathbf{W}_m(\mathbf{x}'_\beta - \mathbf{A}_m\mathbf{x}_\alpha)) < 7S_m \tag{19}$$

   as inliers (Appendix C).
2. Optimally fit an affine transformation to the selected inliers in the following form (Appendix B):

$$\mathbf{x}'_\beta = \mathbf{A}\mathbf{x}_\alpha. \tag{20}$$

3. Discard the candidate pairs $\{(\alpha, \beta)\}$, and select from all the points $\{P_\alpha\}$ and $\{Q_\beta\}$ those pairs $\{(\alpha, \beta)\}$ that satisfy

$$(\mathbf{x}'_\beta - \mathbf{A}\mathbf{x}_\alpha, \mathbf{W}(\mathbf{x}'_\beta - \mathbf{A}\mathbf{x}_\alpha)) < 7S_m, \tag{21}$$

   where the matrix $\mathbf{W}$ is computed by Eq. (18) using the matrix $\mathbf{A}$ of the fitted affine transformation.
4. Apply to the selected pairs $\{(\alpha, \beta)\}$ the *affine template* matching: we enlarge the template $\mathcal{N}$ in Eq. (1) to 25×25 and compute (i', j') by

$$\begin{pmatrix} i' \\ j' \end{pmatrix} = \begin{pmatrix} x'_\beta \\ y'_\beta \end{pmatrix} + A\begin{pmatrix} i \\ j \end{pmatrix}, \tag{22}$$

   where A is the top-left 2×2 submatrix of $\mathbf{A}$. Then, remove those pairs $\{(\alpha, \beta)\}$ that have large residuals, using the automatic thresholding scheme (Appendix A).

The resulting pairs $\{(\alpha, \beta)\}$ are regarded as new candidates.

3.5. Homography matching by four-point voting

We iterate the following computations with $S_m = \infty$ and $\mathbf{H}_m = \mathbf{I}$ as initial values:

1. Randomly sample four pairs (a_0, b_0), (a_1, b_1), (a_2, b_2), and (a_3, b_3) from among the candidate pairs $\{(\alpha, \beta)\}$.
2. Compute the homography $\mathbf{H}$ determined by these four pairs (Appendix B).
3. Sort the candidate pairs $\{(\alpha, \beta)\}$ with respect to

$$D = (\mathbf{x}'_\beta \times \mathbf{H}\mathbf{x}_\alpha, \mathbf{W}(\mathbf{x}'_\beta \times \mathbf{H}\mathbf{x}_\alpha)) \tag{23}$$

   and compute its median S, where $\mathbf{P}_\mathbf{k} = \mathrm{diag}(1, 1, 0)$. The matrix $\mathbf{W}$ is defined by

$$\mathbf{W} = (\mathbf{x}'_\beta \times \mathbf{H}\mathbf{P}_\mathbf{k}\mathbf{H}^{\mathrm{T}} \times \mathbf{x}'_\beta + (\mathbf{H}\mathbf{x}_\alpha) \times \mathbf{P}_\mathbf{k} \times (\mathbf{H}\mathbf{x}_\alpha))_2^-, \tag{24}$$

   where $(\cdot)_2^-$ denotes the Moore-Penrose generalized inverse with rank 2 (i.e. the smallest eigenvalue is replaced by 0) [7]. For a vector $\mathbf{a}$ and a matrix $\mathbf{A}$, the product $\mathbf{a} \times \mathbf{A}$ denotes the matrix whose columns are the vector products of $\mathbf{a}$ and the columns of $\mathbf{A}$; the product $\mathbf{A} \times \mathbf{a}$ denotes the matrix whose rows are the vector products of $\mathbf{a}$ and the rows of $\mathbf{A}$.
4. If $S < S_m$, update S_m and $\mathbf{H}_m : S_m \leftarrow S$, $\mathbf{H}_m \leftarrow \mathbf{H}$.

Repeat the above computations until the median S_m reaches its minimum. Then, do the following computations:

1. Regard those among the candidate pairs $\{(\alpha, \beta)\}$ that satisfy

$$(\mathbf{x}'_\beta \times \mathbf{H}_m\mathbf{x}_\alpha, \mathbf{W}_{m\lambda}(\mathbf{x}'_\beta \times \mathbf{H}_m\mathbf{x}_\alpha)) < 7S_m \tag{25}$$

   as inliers (Appendix C), where the matrix $\mathbf{W}_m$ is defined by Eq. (24) after replacing $\mathbf{H}$ by $\mathbf{H}_m$.
2. Optimally fit a homography $\mathbf{H}$ to the inliers (Appendix B).
3. Discard the candidate pairs $\{(\alpha, \beta)\}$, and select from all the points $\{P_\alpha\}$ and $\{Q_\beta\}$ those pairs $\{(\alpha, \beta)\}$ that satisfy

$$(\mathbf{x}'_\beta \times \mathbf{H}\mathbf{x}_\alpha, \mathbf{W}(\mathbf{x}'_\beta \times \mathbf{H}\mathbf{x}_\alpha)) < \frac{d^2}{2f^2}, \tag{26}$$

   where d is a user-definable constant and the matrix $\mathbf{W}$ is defined by Eq. (24) from the optimally fitted homography $\mathbf{H}$.
4. Apply to the selected pairs $\{(\alpha, \beta)\}$ the *homography template matching:* we enlarge the template $\mathcal{N}$ to 33×33 in Eq. (1) and compute (i', j') by

$$\begin{pmatrix} i'/f_0 \\ j'/f_0 \\ 1 \end{pmatrix} = Z\left[\mathbf{T}\mathbf{H}\begin{pmatrix} (x_\alpha + i)/f_0 \\ (y_\alpha + j)/f_0 \\ 1 \end{pmatrix}\right], \tag{27}$$

   where $Z[\cdot]$ denotes scale normalization to make the Z component 1. The matrix $\mathbf{H}$ deforms the template according to the estimated homography, and the matrix

$$\mathbf{T} = (\mathbf{i}\ \mathbf{j}\ \mathbf{k} + \mathbf{x}'_\beta - Z[\mathbf{H}\mathbf{x}_\alpha]) \tag{28}$$

   adjusts the position of the template so that $\mathbf{x}_\alpha$ exactly matches $\mathbf{x}'_\beta$, where $\mathbf{i} = (1, 0, 0)^{\mathrm{T}}$, $\mathbf{j} = (0, 1, 0)^{\mathrm{T}}$, and $\mathbf{k} = (0, 0, 1)^{\mathrm{T}}$. Then, we remove those pairs $\{(\alpha, \beta)\}$ that have large residuals, using the automatic thresholding scheme (Appendix A).

The resulting pairs $\{(\alpha, \beta)\}$ are regarded as the final matches.

3.6. Summary of the procedure

In the above process, we progressively estimate the image transformation using LMedS (Appendix C), starting from translation to similarity, affine transformation, and homography. At each stage, we discard the previous candidate pairs and match the points all over again using a new template compatible with the estimated transformation; the template size is increased to 9, 17, 25, and 33 to

upgrade the discriminative power. As a result, those pairs initially rejected can be accepted in the later stage.

At each stage, we apply the automatic thresholding procedure (Appendix A). To do this, we need an estimate of the ratio of the percentage p of the correct matches to its maximum value $p_{\max}$ (Appendix A). We gradually increase it to $p/p_{\max} = 0.6$, 0.7, 0.8, and 0.9.

If we know that the two images should undergo a particular transformation to a specified degree, the best choice may be RANSAC [4]. Here, however, the intermediate transformations are all approximate, so we do not know to what extent they should be satisfied. LMedS best suits such a case. On the other hand, we know that the true transformation is a homography. So, in the final stage we introduced a user-definable admissible discrepancy d. In our experiment, we set $d = 3$ (pixels).

Since our method is initialized by the standard template matching, the image rotation and the zooming change between the two images should not be extremely large. Otherwise, the use of template matching would be meaningless. If large zooming changes and rotations are known to exist, we need to estimate them by other means. The multiresolution approach [2] may be very effective for such estimation.

We are also assuming that the entire scene undergoes the same homography. This is a strong limitation as compared with other general matching schemes. However, since our method is based on majority voting, minority parts may undergo different transformation. This property can be exploited for intruder detection applications as shown in Section 4.

4. Real image examples

We extracted 100 feature points from the images in Fig. 1(a) and (b) using the Harris operator [5], as marked in the images. Fig. 1(d) is the 'optical flow' (line segments connecting the matching positions) of the initial matches. Fig. 1(e)–(h) shows the upgraded matches obtained by the translation matching, the similarity matching, the affine matching, and the final homography matching, respectively. We can see that the accuracy progressively increases in each stage. Fig. 1(c) is the resulting mosaiced image.

For comparison, we did the standard LMedS procedure [19], directly computing the least-median homography by random 4-point voting followed by outlier removal; Fig. 1(i) shows the resulting matches. In this example, the distortion between the two images is relatively small, so the automatic thresholding (Appendix A) alone produces sufficiently correct matches (69.0% correct). As a result, the direct LMedS can also gives a satisfactory result. However, our procedure produces denser matches, because our method can generate new matches by adjusting the template.

Fig. 2 is another example similarly arranged. In this case, the image distortion is large. In addition, periodic patterns exist in the scene. As a result, the inlier ratio is very low (28.3% correct), so the direct LMedS fails, as shown in Fig. 2(i). However, our method produces correct matches, as shown in Fig. 2(h). The reason is as follows.

Although only 28.3% of the matches in Fig. 2(d) are compatible with a correct homography, most of them are compatible with an approximate translation or an approximate similarity with a large error allowance given by LMedS (Appendix C). Hence, we can successively narrow down the correct matches.

The road images in Fig. 3(a) and (b) are taken from different positions, and an intruding object (a car) appears in one image. Also, non-planar parts (poles and bushes) exist in the scene, so we cannot map one image to the other entirely by a single homography.

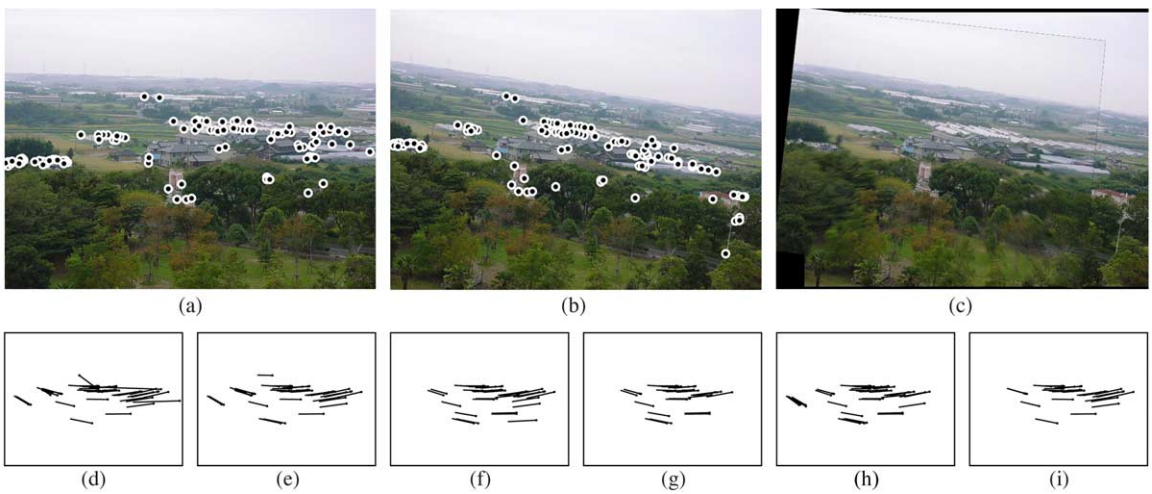

Fig. 1. (a) and (b) Input images and extracted feature points. (c) Image mosaicing by our method. (d) Initial matches (69.0% correct). (e) Translation matching. (f) Similarity matching. (g) Affine matching. (h) Homography matching. (i) Direct estimation by LMedS.

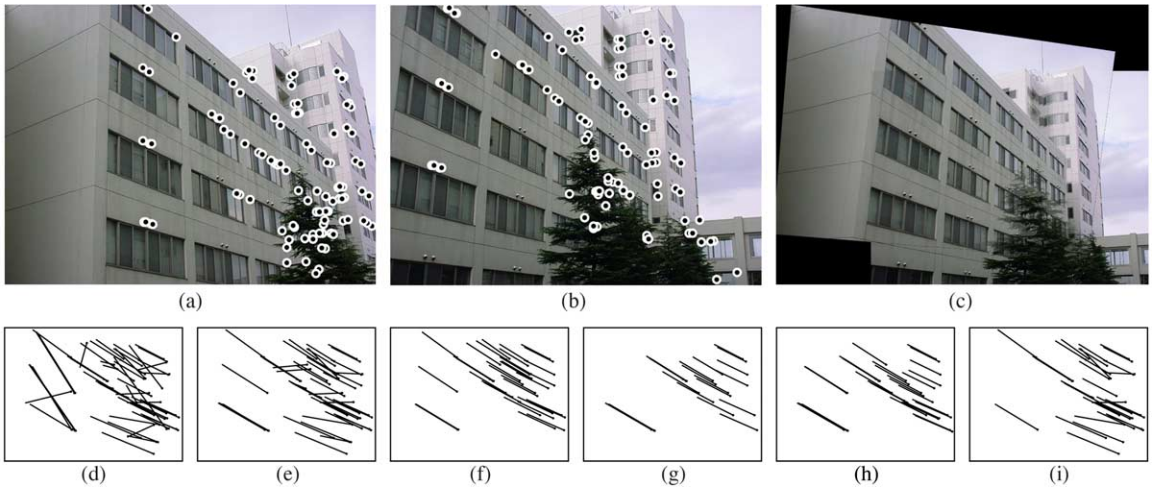

Fig. 2. (a) and (b) Input images and extracted feature points. (c) Image mosaicing by our method. (d) Initial matches (28.3% correct). (e) Translation matching. (f) Similarity matching. (g) Affine matching. (h) Homography matching. (i) Direct estimation by LMedS.

The initial matches are shown in Fig. 3(d); they are only 44.8% correct. In the end, however, correct matches are generated in the planar part of the scene, as shown in Fig. 3(e). Also, some of the initially incorrect matches are correctly recombined.

The small square (33×33 pixels) in Fig. 3(a) indicates the template region around a feature point in the final stage; the corresponding deformed template is superimposed around the matched point in Fig. 3(b). The evolution of its shape (to scale) through the translation matching, the similarity matching, the affine matching, and the homography matching is shown in Fig. 3(c).

Fig. 3(f) is the superposition of the mapped images, and Fig. 3(g) displays the absolute value of their difference. The non-overlapping parts indicate non-planar parts of the scene.

Fig. 4 shows another example similarly arranged. In this case, the initial matches in Fig. 4(c) are *almost entirely incorrect* (16.3% correct). Yet, our procedure can successfully recombine them into correct matches.

Fig. 5 shows an example of image mosaicing using our method. The camera was somewhat rotated in the course of shooting, but the panoramic image below was automatically generated from the seven images above.

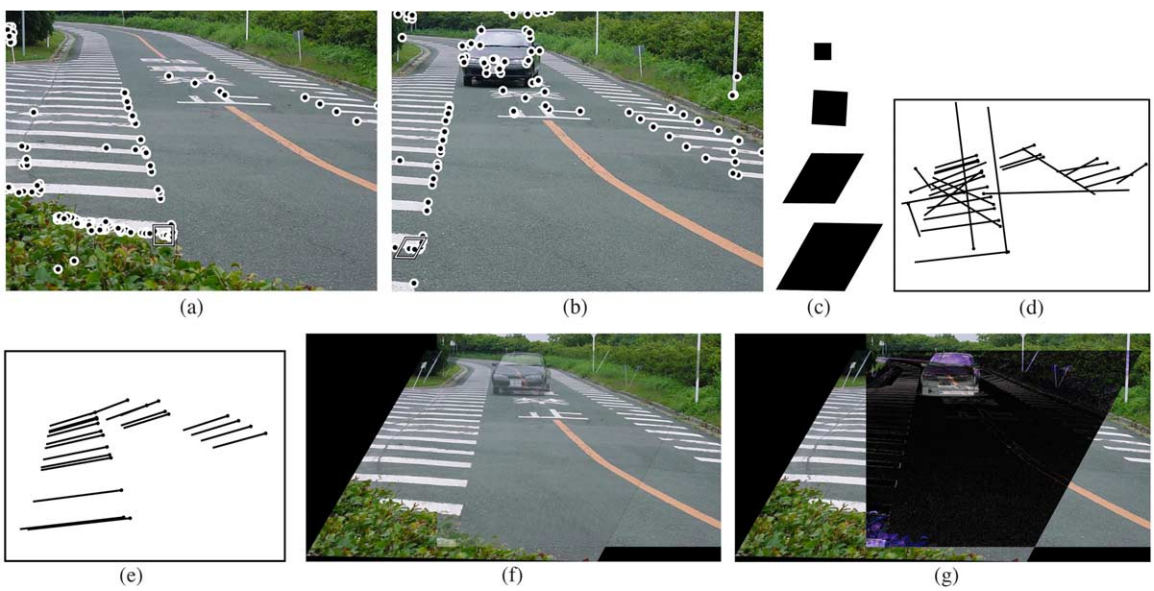

Fig. 3. (a) and (b) Input images and detected feature points. The template regions in the final stage are superimposed around a pair of matched points. (c) The evolution of the template shape (to scale) for the match indicated in (a) and (b). (d) Initial matches (44.8% correct). (e) Final matches. (f) Image mosaicing. (g) Difference image.

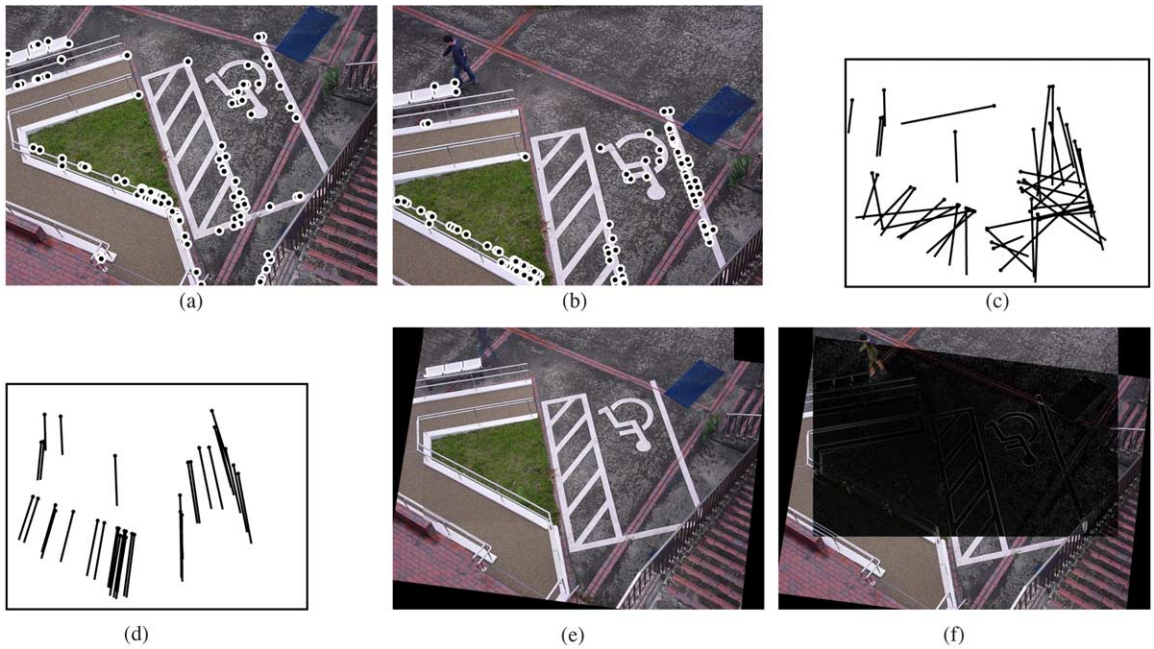

Fig. 4. (a) and (b) Input images and extracted feature points. (c) Initial matches. (d) Final matches (16.3% correct). (e) Image mosaicing. (f) Difference image.

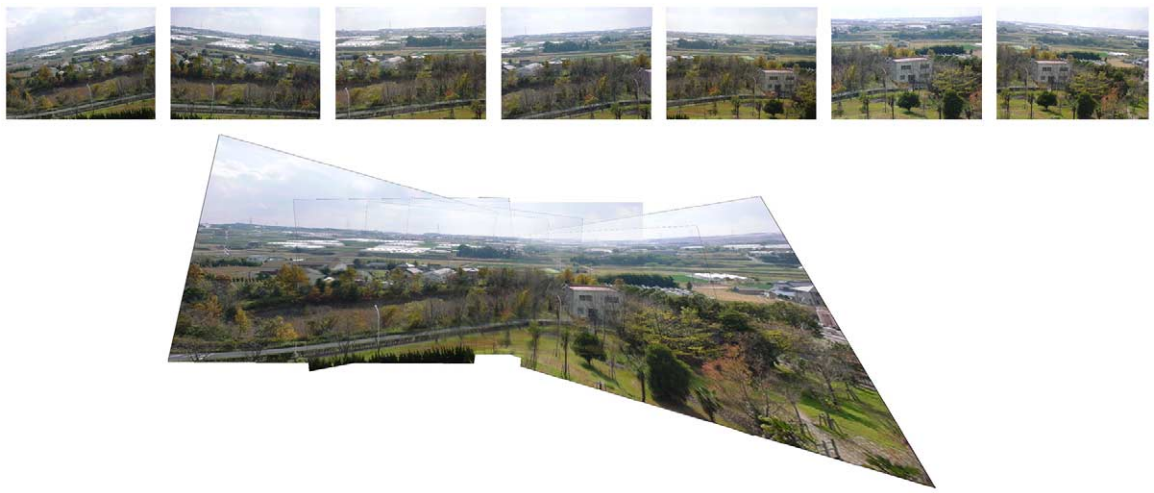

Fig. 5. An example of image mosaicing using our method. Above: seven input images. Below: the resulting panoramic image.

5. Conclusions

We have presented a robust method for automatically matching feature points over two images for image mosaicing. After extracting feature points using a feature detector, we progressively estimate the rotation, scale change, and the projective distortion between the two images by random voting and variable template matching.

Traditional approaches for image mosaicing based on local correlations and optical flow [23,27] cannot deal with a large amount of camera rotation, zooming, and perspective distortion, such as the images in Fig. 3. Our method works even in the presence of a large percentage of initial outliers.

Acknowledgements

This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, under the Grant for the 21st Century COE Program 'Intelligent Human Sensing' and the Grant in Aid for Scientific Research C(2) (No. 15500113), the Support

Center for Advanced Telecommunications Technology Research, and Kayamori Foundation of Informational Science Advancement.

Appendix A. Automatic thresholding for template matching

Here, we summarize the procedure for our automatic thresholding scheme [8].

Given N points in the first image and M points in the second, we compute the template matching residual J defined by Eq. (1) for *all* the NM combinations of the points (even though some of them apparently do not satisfy required geometric constraints). Let $J_1 \leq J_2 \leq \cdots \leq J_{NM}$ be the resulting NM residuals sorted in ascending order.

Intuitively, we may imagine that $J_1 \leq \cdots \leq J_c$ for some J_c are correct matches and the remaining $J_{c+1} \leq \cdots \leq J_{NM}$ incorrect matches. In reality, however, the distribution of correct matches is in large part included in the distribution of incorrect matches with a long tail, as depicted in Fig. A1, indicating that no obvious threshold exists [8]. Hence, if we want to pick out a large number of correct matches, we need to set a high threshold, which inevitably accepts many incorrect matches.

It follows that in order to set an optimal threshold, we need to model the residual distribution and determine the value that best balances the conflicting goals of collecting as many correct matches as possible and rejecting as many incorrect ones as possible. To this end, we fit a χ^2 density function to the residual histogram.

Our starting point is the observation that the residual J of a correct match should be due to small distortions between the two images as well as random fluctuations of the image intensity. If we model the intensity difference between the matching pixels as a Gaussian distribution of mean 0 and standard deviation σ_0, the ratio J/σ_0^2 should be subject to a χ^2 distribution with n^2 degrees of freedom, where n is the template size, provided the pixel value difference is pixel-wise independent.

The residual of an incorrect match, on the other hand, is due to the inhomogeneity of that particular scene. If we model the intensity difference between the matching pixels as a Gaussian distribution of mean 0 and standard deviation σ_1, the ratio J/σ_1^2 should be subject to a χ^2 distribution with n^2 degrees of freedom, provided the pixel value difference is pixel-wise independent.

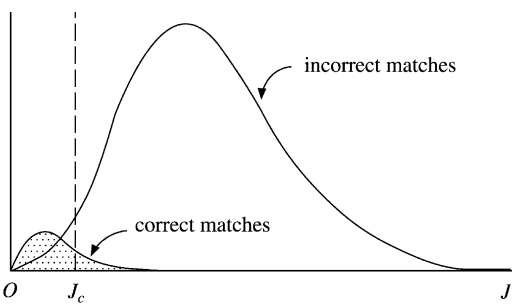

Fig. A1. The residual distribution of correct matches is in large part included in the residual distribution of incorrect matches.

In reality, the pixel value difference should hardly be independent. However, the interpixel correlations are difficult to analyze, so we introduce the following approximation. If there are N points in the first image and M points in the second, the number of correct matches is at most $\min(N, M)$, which is much smaller than the total number NM of all the pairs. If most of the matches are incorrect, the average μ_J and the variance σ_J^2 of the residual J over all the matches should be approximately $n^2\sigma_1^2$ and $2n^2\sigma_1^4$, respectively, in the absence of correlations. Eliminating σ_1, we obtain $n^2 \approx 2\mu_J^2/\sigma_J^2$. However, n^2 should be much smaller than this value due to correlations. So, we define the *effective template size* as follows [8]:

$$n = \frac{\sqrt{2}\mu_J}{\sigma_J}. \tag{A1}$$

In other words, we regard each pixel value as if independent within the template of that size, which need not be an integer.

Let p and q $(= 1 - p)$ be the ratios of the correct and incorrect matches, respectively, among the total NM matches. According to our model, the probability density of the residual J for all the matches is

$$f(J) = \frac{p}{\sigma_0^2}\phi_{n^2}\left(\frac{J}{\sigma_0^2}\right) + \frac{q}{\sigma_1^2}\phi_{n^2}\left(\frac{J}{\sigma_1^2}\right), \tag{A2}$$

where $\phi_d(x)$ denotes the probability density of the χ^2 distribution with d degrees of freedom. We determine the model parameters σ_0 and σ_1 by maximum likelihood estimation from the NM residuals $J_1 \leq J_2 \leq \cdots \leq J_{NM}$. Differentiating $\log \prod_{i=1}^{NM} f(J_i)$ with respect to σ_0^2 and σ_1^2 and letting the results be zero, we obtain

$$\sigma_0^2 = \frac{\sum_{i=1}^{NM} A_i J_i}{n^2 \sum_{i=1}^{NM} A_i}, \qquad \sigma_1^2 = \frac{\sum_{i=1}^{NM} B_i J_i}{n^2 \sum_{i=1}^{NM} B_i}, \tag{A3}$$

where we define

$$A_i = \frac{1}{1 + \frac{q}{p}\left(\frac{\sigma_0}{\sigma_1}\right)^{n^2} \exp\left(\frac{J_i}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)\right)},$$
$$B_i = \frac{1}{1 + \frac{p}{q}\left(\frac{\sigma_1}{\sigma_0}\right)^{n^2} \exp\left(\frac{J_i}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\right)}. \tag{A4}$$

The values of σ_0 and σ_1 are obtained by iterations [8]. Experiments using real images have shown that this model

with σ_0, σ_1, and n estimated as described above fits very well to the actual residual histogram [8].

The ratios p and q $(= 1 - p)$ are given empirically. Since the number of correct matches between N points and M points is at most $\min(N, M)$, we let $p_{\max} = \min(N, M)/NM$ and set, for example, $p = 0.6p_{\max}$ if no knowledge is available about the correctness of the matches. Experiments have shown that the final results are only slightly affected by the choice of $p/p_{\max}$ [8].

Suppose we set a threshold J_c for the residual J and accept those matches with $J \leq J_c$ as correct. Let α be the ratio of the accepted correct matches among all the correct ones; we call it the *detection ratio*. A correct match with residual J is accepted with the probability

$$\alpha = P_0[J < J_c] = P_0\left[\frac{J}{\sigma_0^2} < \frac{J_c}{\sigma_0^2}\right], \tag{A5}$$

where $P_0[\cdot]$ denotes the probability for correct matches. Let $\chi^2_{n^2}(\alpha)$ be the αth percentile of the χ^2 distribution with n^2 degrees of freedom. Since J/σ_0^2 for a correct match is subject to a χ^2 distribution with n^2 degrees of freedom, Eq. (A5) implies that J_c/σ_0^2 equals $\chi^2_{n^2}(\alpha)$. Hence, the threshold J_c is given by

$$J_c = \sigma_0^2 \chi^2_{n^2}(\alpha). \tag{A6}$$

Some incorrect matches are necessarily accepted by this thresholding. An incorrect match with residual J is accepted with the probability

$$\gamma = P_1[J \leq J_c] = P_1\left[\frac{J}{\sigma_1^2} \leq \left(\frac{\sigma_0}{\sigma_1}\right)^2 \chi^2_{n^2}(\alpha)\right], \tag{A7}$$

where $P_1[\cdot]$ denotes the probability for incorrect matches. Let $\Phi_{n^2}(X)$ $(= \int_0^X \phi_{n^2}(x)\mathrm{d}x)$ be the accumulated probability function of the χ^2 distribution with n^2 degrees of freedom. Since J/σ_1^2 for an incorrect match is subject to a χ^2 distribution with n^2 degrees of freedom, Eq. (A7) implies

$$\gamma = \Phi_{n^2}\left(\left(\frac{\sigma_0}{\sigma_1}\right)^2 \chi^2_{n^2}(\alpha)\right). \tag{A8}$$

Among the NM possible matches, the numbers of correct and incorrect matches are pNM and qNM, respectively. After the thresholding, we obtain αpNM correct matches and γqMN incorrect ones on average. Hence, the *inlier ratio*, i.e. the ratio of correct matches among the accepted matches, is approximately

$$\beta = \frac{\alpha pNM}{\alpha pNM + \gamma qMN} = \frac{\alpha p}{\alpha p + \gamma q}. \tag{A9}$$

The detection ratio α should be large if we want to collect many correct matches, but the number of incorrect matches also increases, lowering the inlier ratio β as a result. So, we determine the threshold J_c so that the detection ratio α *equals* the inlier ratio β. This balances the ratio $1 - \alpha$ of rejecting correct matches and the ratio $1 - \beta$ of accepting incorrect ones. Letting $\beta = \alpha$ in Eq. (A9), we obtain

$$\alpha = 1 - \frac{q}{p}\Phi_{n^2}\left(\left(\frac{\sigma_0}{\sigma_1}\right)^2 \chi^2_{n^2}(\alpha)\right), \tag{A10}$$

from which α is obtained by iterations [8]. Then, threshold J_c is given by Eq. (A6).

Appendix B. Optimal fitting of image transformations

A *homography* from a point $\mathbf{x}_\alpha$ to a point $\mathbf{x}'_\alpha$ is written in the form

$$\mathbf{x}'_\alpha = Z[\mathbf{H}\mathbf{x}_\alpha], \qquad \mathbf{H} = \begin{pmatrix} A & B & C \\ D & E & F \\ P & Q & R \end{pmatrix}. \tag{B1}$$

(Recall the notation of Eq. (3) and the normalization operation $Z[\cdot]$.) Since the matrix $\mathbf{H}$ has nine elements up to scale, it can be uniquely determined from four matches in general position.

We regard the points $\{\mathbf{x}_\alpha\}$ and $\{\mathbf{x}'_\alpha\}$ as Gaussian random variables whose means are their true values $\{\bar{\mathbf{x}}_\alpha\}$ and $\{\bar{\mathbf{x}}'_\alpha\}$. Let $\{V[\mathbf{x}_\alpha]\}$ and $\{V[\mathbf{x}'_\alpha]\}$ be their covariance matrices. We write

$$V[\mathbf{x}_\alpha] = \epsilon^2 V_0[\mathbf{x}_\alpha], \qquad V[\mathbf{x}'_\alpha] = \epsilon^2 V_0[\mathbf{x}'_\alpha], \tag{B2}$$

and call ϵ the *noise level*. The matrices $V_0[\mathbf{x}_\alpha]$ and $V_0[\mathbf{x}'_\alpha]$ given up to scale are called the *normalized covariance matrices*. We can assume

$$V_0[\mathbf{x}_\alpha] = V_0[\mathbf{x}'_\alpha] = \mathrm{diag}(1, 1, 0) \tag{B3}$$

in the usual situation [11].

Since Eq. (B1) can equivalently be written in the form $\mathbf{x}'_\alpha \times \mathbf{H}\mathbf{x}_\alpha = \mathbf{0}$, an optimal estimate of $\mathbf{H}$ for $N(\geq 4)$ matches $\{\mathbf{x}_\alpha\}$, $\{\mathbf{x}'_\alpha, \alpha = 1, \ldots, N$, is obtained by minimizing

$$K = \sum_{\alpha=1}^{N} (\mathbf{x}_\alpha - \bar{\mathbf{x}}_\alpha, V_0[\mathbf{x}_\alpha]_2^-(\mathbf{x}_\alpha - \bar{\mathbf{x}}_\alpha)) + \sum_{\alpha=1}^{N} (\mathbf{x}'_\alpha - \bar{\mathbf{x}}'_\alpha, V_0[\mathbf{x}'_\alpha]_2^-(\mathbf{x}'_\alpha - \bar{\mathbf{x}}'_\alpha)), \tag{B4}$$

subject to the constraint

$$\bar{\mathbf{x}}'_\alpha \times \mathbf{H}\bar{\mathbf{x}}_\alpha = \mathbf{0}. \tag{B5}$$

Using Lagrange multipliers and introducing first order approximation, we can eliminate the constraint (B5) to express Eq. (B4) in the following form [7]:

$$K = \sum_{\alpha=1}^{N} (\mathbf{x}'_\alpha \times \mathbf{H}\mathbf{x}_\alpha, \mathbf{W}_\alpha(\mathbf{x}'_\alpha \times \mathbf{H}\mathbf{x}_\alpha)), \tag{B6}$$

$$\mathbf{W}_\alpha = (\mathbf{x}'_\alpha \times \mathbf{H}V_0[\mathbf{x}_\alpha]\mathbf{H}^{\mathrm{T}} \times \mathbf{x}'_\alpha + (\mathbf{H}\mathbf{x}_\alpha) \times V_0[\mathbf{x}'_\alpha] \times (\mathbf{H}\mathbf{x}_\alpha))_2^-. \tag{B7}$$

In our experiment, we computed the solution by a technique called *renormalization*[4] [9].

An *affine transformation* is a special homography that takes the form

$$\mathbf{x}'_\alpha = \mathbf{A}\mathbf{x}_\alpha, \qquad \mathbf{A} = \begin{pmatrix} A & t_1/f_0 \\ & t_2/f_0 \\ 0\,0 & 1 \end{pmatrix}, \tag{B8}$$

where A is a 2×2 non-singular matrix. The matrix $\mathbf{A}$ is determined from three matches $\{\mathbf{x}_\alpha\}$, $\mathbf{x}'_\alpha$, $\alpha = 1, 2, 3$, in general position:

$$\mathbf{A} = (\mathbf{x}'_1\ \mathbf{x}'_2\ \mathbf{x}'_3)(\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3)^{-1}. \tag{B9}$$

For N (≥ 3) matches $\{\mathbf{x}_\alpha\}$, $\{\mathbf{x}'_\alpha\}$, $\alpha = 1, \ldots, N$, substitution of Eqs. (B3) and (B8) into Eq. (B4) yields

$$K = \sum_{\alpha=1}^{N} (\mathbf{x}'_\alpha - \mathbf{A}\mathbf{x}_\alpha, \mathbf{W}(\mathbf{x}'_\alpha - \mathbf{A}\mathbf{x}_\alpha)), \tag{B10}$$

where $\mathbf{W}$ is given by Eq. (18). Optimal values of A and $\{t_i\}$ are obtained using the Levenberg–Marquart method.

A *similarity* is a special affine transformation obtained by replacing the matrix $\mathbf{A}$ in Eq. (B8) by

$$\mathbf{S} = \begin{pmatrix} s\cos\theta & -s\sin\theta & t_1/f_0 \\ s\sin\theta & s\cos\theta & t_2/f_0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{B11}$$

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

In terms of the two-dimensional vectors $\vec{x}_\alpha$ and $\vec{x}_\beta$ consisting of the x and y image coordinates, the transformation is written in the form

$$\vec{x}'_\alpha = sR\vec{x}_\alpha + \vec{t}, \tag{B12}$$

where

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \qquad \vec{t} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}. \tag{B13}$$

The mapping is determined uniquely if two distinct matches are given.

For $N(\geq 2)$ matches $\{\vec{x}_\alpha\}$, $\{\vec{x}'_\alpha\}$, $\alpha = 1, \ldots, N$, Eq. (B10) reduces to

$$K = \frac{1}{1+s^2} \sum_{\alpha=1}^{N} \|\vec{x}'_\alpha - sR\vec{x}_\alpha - \vec{t}\|^2. \tag{B14}$$

The minimizing solution is easily obtained using the Levenberg–Marquart method.

Finally, *translation* is a special similarity with $s = 1$ and $R = I$. It is uniquely determined by a single pair of points. For $N(\geq 1)$ matches $\{\vec{x}_\alpha\}$, $\{\vec{x}'_\alpha\}$, $\alpha = 1, \ldots, N$, an optimal translation $\vec{t}$ is given by the displacement of the centroid:

$$\vec{t} = \frac{1}{N} \sum_{\alpha=1}^{N} (\vec{x}'_\alpha - \vec{x}_\alpha). \tag{B15}$$

[4] The program code is publicly available from http://www.ail.cs.gunma-u.ac.jp/Labo/programs-e.html.

Appendix C. LMedS for geometric fitting

Here, we describe our LMedS procedure, which is an extension of the standard LMedS [19] to the general geometric fitting problem.

Geometric fitting [7] is to fit a d-dimensional manifold $\mathscr{M}$ defined by a constraint equation $\mathbf{F}(\xi; \mathbf{u}) = \mathbf{0}$ parameterized by a p-dimensional vector $\mathbf{u}$ to N data points $\{\xi_\alpha\} \in \mathscr{R}^n$. Each data point ξ_α is assumed to be disturbed from its true position $\bar{\xi}_\alpha$ by independent Gaussian noise of mean 0 and standard deviation σ in each coordinate. The true positions $\{\bar{\xi}_\alpha\}$ are assumed to be in the manifold $\mathscr{M}$.

The constraint that a point (x, y) in one image should be mapped to a point (x', y') in another by a translation $\mathscr{T}$, a similarity $\mathscr{S}$, an affine transformation $\mathscr{A}$, or a homography $\mathscr{H}$ defines a two-dimensional manifold $\mathscr{M}$ with 2, 4, 6, or 8 parameters, respectively, in the four-dimensional joint space of (x, y, x', y'). Each data point is specified by the four-dimensional vector $\xi_\alpha = (x_\alpha, y_\alpha, x'_\alpha, y'_\alpha)^{\mathrm{T}}$ consisting of the coordinates of the corresponding points (x_α, y_α) and (x'_α, y'_α).

The LMedS [19] for fitting the manifold $\mathscr{M}$ to $\{\xi_\alpha\}$ is to minimize

$$S = \mathrm{med}_{\alpha=1}^{N} D(\xi_\alpha; \mathscr{M}), \tag{C1}$$

where $D(\xi_\alpha; \mathscr{M})$ measures the square distance of point ξ_α from the manifold $\mathscr{M}$: its actual form is given by Eqs. (5), (11), (17) and (23) for the translation $\mathscr{T}$, the similarity $\mathscr{S}$, the affine transformation $\mathscr{A}$, and the homography, respectively.

Minimization of Eq. (C1) is done by repeating random sampling of the minimum number $\lceil p/r \rceil$ of points that can define the manifold $\mathscr{M}$ a sufficient number of times, evaluating the median S each time, and choosing the manifold $\hat{\mathscr{M}}$ that gives the least median S_m [19].

If the noise is small, $D(\xi_\alpha; \mathscr{M})/\sigma^2$ should be subject to a χ^2 distribution with r degrees of freedom [7], where $r = n - d$ is the *codimension* of $\mathscr{M}$, i.e. the number of independent equations of the constraint $\mathbf{F}(\xi; \mathbf{u}) = \mathbf{0}$. Hence,

$$\mu = \mathop{\mathrm{med}}_{\alpha=1}^{N} \frac{D(\xi_\alpha; \mathscr{M})}{\sigma^2}, \tag{C2}$$

equals in expectation the 50th percentile $\chi^2_{r,50}$ of the χ^2 distribution with r degrees of freedom. It follows that the variance σ^2 can be estimated by

$$\hat{\sigma}^2 = \frac{S}{\chi^2_{r,50}}. \tag{C3}$$

Here, the median S is defined by Eq. (C1) using the true manifold $\mathscr{M}$. In actual computation, $\mathscr{M}$ is approximated by

the manifold $\hat{\mathscr{M}}$ fitted to $\lceil p/r \rceil$ sample points. In other word, we use instead of S

$$S_m = \operatorname*{med}_{\alpha=1}^{N} D(\xi_\alpha; \hat{\mathscr{M}}). \tag{C4}$$

Because we repeat sampling so as to minimize S_m, it is in general smaller than S. So, we apply the correction

$$\hat{\sigma}^2 = \left(1 + \frac{10}{rN - p}\right)\frac{S_m}{\chi^2_{r,50}}. \tag{C5}$$

The term $rN - p$ in the denominator is introduced to account the fact that (i) if $N = p/r$, the fitted manifold $\hat{\mathscr{M}}$ exactly passes through the data points with 0 median, so the variance σ^2 cannot be estimated and (ii) Eq. (C5) gives a good approximation for large N. The number 10 in the numerator is determined so as to make Eq. (C5) agree with the formula in Ref. [19] for $r = 1$:

$$\begin{aligned}\hat{\sigma} &= \sqrt{\left(1 + \frac{10}{N - p}\right)}\sqrt{\frac{S_m}{\chi^2_{1,50}}} \\ &\approx 1.4826\left(1 + \frac{5}{N - p}\right)\sqrt{S_m}.\end{aligned} \tag{C6}$$

Using thus estimated variance $\hat{\sigma}^2$, we can remove outliers with confidence level $\alpha\%$ by rejecting those data ξ_α that satisfy

$$\frac{D(\xi_\alpha; \hat{\mathscr{S}})}{\hat{\sigma}^2} \geq \chi^2_{r,\alpha}, \tag{C7}$$

where $\chi^2_{r,\alpha}$ is the αth percentile of the χ^2 distribution with r degrees of freedom.

For example, we have $r = 2$ and $p = 8$ for fitting a homography, so for $\alpha = 99$ Eq. (C7) reduces to

$$D(\xi_\alpha; \hat{\mathscr{M}}) \geq 6.44\left(1 + \frac{5}{N - 4}\right)S_m. \tag{C8}$$

References

[1] P. Beardsley, P. Torr, A. Zisserman, 3D model acquisition from extended image sequences, Proceedings of the Fourth European Conference on Computer Vision, Cambridge, UK, April 2 (1996) 683–695.

[2] J.R. Bergen, P. Anandan, K.J. Hanna, R. Higorani, Hierarchical model-based motion estimation, Proceedings of the Second European Conference on Computer Vision, Santa Margerita, Italy May (1992) 237–252.

[3] F. Chabat, G.Z. Yang, D.M. Hansell, A corner orientation detector, Image and Vision Computing 17 (10) (1999) 761–769.

[4] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of ACM 24 (6) (1981) 381–395.

[5] C. Harris, M. Stephens, A combined corner and edge detector, Proceedings of the Fourth Alvey Vision Conference, Manchester, UK August (1988) 147–151.

[6] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, Cambridge, 2000.

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

[8] K. Kanatani, Y. Kanazawa, Automatic thresholding for correspondence detection, International Journal of Image and Graphics 3 (2003) in press.

[9] K. Kanatani, N. Ohta, Accuracy bounds and optimal computation of homography for image mosaicing applications, Proceedings of the Seventh International Conference on Computer Vision, Kerkyra, Greece, September 1 (1999) 73–78.

[10] Y. Kanazawa, K. Kanatani, Stabilizing image mosaicing by model selection, in: M. Pollefeys, L. Van Gool, A. Zisserman, A. Fitzgibbon (Eds.), 3D Structure from Images—SMILE 2000, 2001, Springer, Berlin, 2001, pp. 35–51.

[11] Y. Kanazawa, K. Kanatani, Do we really have to consider covariance matrices for image features?, Proceedings of the Eighth International Conference on Computer Vision, Vancouver, Canada, July 2 (2001) 586–591.

[12] M.-S. Lee, G. Medioni, P. Mordohai, Inference of segmented overlapping surfaces from binocular stereo, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (6) (2002) 824–837.

[13] M.I.A. Lourakis, S.V. Tzurbakis, A.A. Argyros, S.C. Orphanoudakis, Feature transfer and matching in disparate stereo views through the use of plane homography, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2) (2003) 271–276.

[14] J. Maciel, J. Costeira, Robust point correspondence by concave minimization, Image and Vision Computing 20 (9/10) (2002) 683–690.

[15] J. Maciel, J. Costeira, A global solution to sparse correspondence problems, IEEE Transactions Pattern Analysis and Machine Intelligence 25 (2) (2003) 187–199.

[16] C.F. Olson, Maximum-likelihood image matching, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (6) (2002) 853–857.

[17] M. Pilu, A direct method for stereo correspondence based on singular value decomposition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico June (1997) 261–266.

[18] D. Reisfeld, H. Wolfson, Y. Yeshurun, Context-free attentional operators: the generalized symmetry transform, International Journal of Computer Vision 14 (2) (1995) 119–130.

[19] P.J. Rousseeuw, A.M. Leroy, Robust Regression and Outlier Detection, Wiley, New York, 1987.

[20] C. Schmid, R. Mohr, C. Bauckhage, Evaluation of interest point detections, International Journal of Computer Vision 37 (2) (2000) 151–172.

[21] S.M. Smith, J.M. Brady, SUSAN—a new approach to low level image processing, International Journal of Computer Vision 23 (1) (1997) 45–78.

[22] F. Schaffalitzky, A. Zisserman, Multi-view matching for unordered image sets, or 'How do I organize my holiday snaps?', Proceedings of the Seventh European Conference on Computer Vision, Copenhagen, Denmark, May 1 (2002) 414–431.

[23] R. Szeliski, H.-Y. Shum, Creating full view panoramic image mosaics and environment maps, Proceedings of the SIGGRAPH'97, Los Angeles, CA August (1997) 251–258.

[24] P.H.S. Torr, C. Davidson, IMPSAC: synthesis of importance sampling and random sample consensus, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (3) (2003) 354–364.

[25] M.A. van Wyk, T.S. Durrani, B.J. van Wyk, A RKHS interpolator-based graph matching algorithm, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (7) (2002) 988–995.

[26] Z. Zhang, R. Deriche, O. Faugeras, Q.-T. Luong, A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry, Artificial Intelligence 78 (1995) 87–119.

[27] I. Zoghlami, O. Faugeras, R. Deriche, Using geometric corners to build a 2D mosaic from a set of images, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico June (1997) 420–425.