Proc. Statistical Methods in Video Processing Workshop, Copenhagen, Denmark, 1–2 June 2002, pp. 31–36.

Image Mosaicing by Stratified Matching

Yasushi Kanazawa* and Kenichi Kanatani†

*Department of Knowledge-based Information Engineering
Toyohashi University of Technology, Toyohashi, Aichi 441-8580 Japan
†Department of Information Technology, Okayama University, Okayama 700-8530 Japan
kanazawa@tutkie.tut.ac.jp, kanatani@suri.it.okayama-u.ac.jp

Summary

We present a new method for automatically matching feature points over two images: 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. We demonstrate that out method allows robust image mosaicing even when conventional methods fail.

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. In [6], we proposed a method for automatically thresholding the matching residual: we dynamically choose an optimal threshold by analyzing the histogram of the residuals of the potential matches.

Although this method significantly reduces the number of outliers, some outliers still remain. To remove them, we need to apply a robust estimation scheme based on a geometric constraint such as the homography relationship and the epipolar equation [1, 4, 11]. However, standard techniques such as LMedS [9] and RANSAC [2] do not work unless the initial matches are sufficiently accurate.

The reason why correct matches are difficult to obtain by the standard template matching is that image distortions such as rotation and scale change exist over the two images. In this paper, we introduce random voting for hierarchically estimating such image distortions followed by variable template matching compatible with the estimated distortions, which we call *stratified matching*. Using real images, we demonstrate that out method allows robust image mosaicing even when conventional methods fail.

2. Variable Template Matching

We apply a feature detector to two images and detect points P_1, ..., P_N in the first image I_1 and points Q_1, ..., Q_M in the second image I_2 and measure the similarity between two points by the residual (sum of squares)

$$J(P_\alpha, Q_\beta) = \sum_{(i,j)\in\mathcal{N}_\alpha} |I_1(i,j) - I_2(\mathcal{D}_{P_\alpha}^{Q_\beta}(i,j))|^2, \quad (1)$$

where $\mathcal{N}_\alpha$ is a rectangular grid centered on the point P_α in first image I_1, and $\mathcal{D}_{P_\alpha}^{Q_\beta}$ designates a coordinate transformation[1] that maps P_α to Q_β.

The basic transformation $\mathcal{D}_{P_\alpha}^{Q_\beta}$ is the translation that displaces P_α to Q_β. However, the residual $J(P_\alpha, Q_\beta)$ is not 0 in general even if P_α exactly corresponds to Q_β. This is not only due to random fluctuations of the image intensity and illumination changes between the two images but also due to the relative image distortions such as scale change and rotation.

Such image distortions can be absorbed by adjusting the transformation $\mathcal{D}_{P_\alpha}^{Q_\beta}$ accordingly, but we do not know the distortions a priori. Motivated by image mosaicing applications [8, 10, 12], we assume that the entire scene undergoes a homography (we later allow some parts to deform differently) and estimate an image transformation that fits all feature points rather than point-wise adjustment.

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 progressively expand the template and remove outliers using the *least median of squares* (LMedS) method [9]. We call this strategy *stratified matching* to distinguish it from what is known as "hierarchical matching", which upgrades the matching by gradually increasing the resolution from coarse to fine using the standard template.

[1]The coordinates (i, j) are extended to real numbers, for which the image intensity is defined by an appropriate interpolation of the values at surrounding pixels.

3. Stratified Matching

3.1 Initial matching

We detect a fixed number[2] N of feature points $\{P_\alpha\}$ and $\{Q_\beta\}$ in the first and second images separately using the Harris operator [3] and compute the residuals $J(P_\alpha, Q_\beta)$ for all possible pairs. We let $\mathcal{D}_{P_\alpha}^{Q_\beta}$ be the translation that displaces P_α to Q_β with a 9×9 grid $\mathcal{N}_\alpha$. We then apply the thresholding described in [6] and enforce uniqueness as follows.

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

3.2 Translation matching by one-point voting

1. Randomly sample *one* from among the candidate pairs $\{(P_\alpha, Q_\beta)\}$, and compute the translation $\mathcal{T}$ it defines.
2. Compute for each candidate pair (P_α, Q_β) the discrepancy $D(P_\alpha, Q_\beta; \mathcal{T})$ from the estimated translation $\mathcal{T}$ (cf. Appendix), and record the median S of $D(P_\alpha, Q_\beta; \mathcal{T})$ over all the pairs $\{(P_\alpha, Q_\beta)\}$.
3. Repeat this sampling a sufficient number of times[3], and find the translation $\mathcal{T}_m$ that gives the smallest median S_m.
4. Regard those pairs $\{(P_\alpha, Q_\beta)\}$ that satisfy $D(P_\alpha, Q_\beta; \mathcal{T}_m) < 7S_m$ as inliers (cf. Appendix), and optimally fit a translation $\hat{\mathcal{T}}$ to them.
5. *Cancel the existing candidate pairs*, and select from *all the points* $\{P_\alpha\}$ and $\{Q_\beta\}$ those pairs $\{(P_\alpha, Q_\beta)\}$ that satisfy $D(P_\alpha, Q_\beta; \hat{\mathcal{T}}) < 7S_m$.
6. Compute the residuals $J(P_\alpha, Q_\beta)$ for all the combinations of the endpoints $\{P_\alpha\}$ and $\{Q_\beta\}$ of the selected pairs, using the same translation template as the initial one, and apply the thresholding and uniqueness enforcing procedure [6].

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

3.3 Similarity matching by two-point voting

1. Randomly sample *two* from among the candidate pairs $\{(P_\alpha, Q_\beta)\}$, and compute the similarity $\mathcal{S}$ they define.
2. Compute for each candidate pair (P_α, Q_β) the discrepancy $D(P_\alpha, Q_\beta; \mathcal{S})$ from the estimated similarity $\mathcal{S}$ (cf. Appendix), and record the median S of $D(P_\alpha, Q_\beta; \mathcal{S})$ over all the pairs $\{(P_\alpha, Q_\beta)\}$.
3. Repeat this sampling a sufficient number of times[4], and find the similarity $\mathcal{S}_m$ that gives the smallest median S_m.
4. Regard those pairs $\{(P_\alpha, Q_\beta)\}$ that satisfy $D(P_\alpha, Q_\beta; \mathcal{S}_m) < 7S_m$ as inliers (cf. Appendix), and optimally fit a similarity $\hat{\mathcal{S}}$ to them [5].
5. *Cancel the existing candidate pairs*, and select from *all the points* $\{P_\alpha\}$ and $\{Q_\beta\}$ those pairs $\{(P_\alpha, Q_\beta)\}$ that satisfy $D(P_\alpha, Q_\beta; \hat{\mathcal{S}}) < 7S_m$.
6. Compute the residuals $J(P_\alpha, Q_\beta)$ for all the combinations of the endpoints $\{P_\alpha\}$ and $\{Q_\beta\}$ of the selected pairs, using a 17×17 template. We let $\mathcal{D}_{P_\alpha}^{Q_\beta}$ be the translation that displaces P_α to Q_β followed by the rotation and the scale change compatible with the estimated similarity $\hat{\mathcal{S}}$.
7. Apply the thresholding and uniqueness enforcing procedure [6].

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

3.4 Affine matching by three-point voting

1. Randomly sample *three* from among the candidate pairs $\{(P_\alpha, Q_\beta)\}$, and compute the affine transformation $\mathcal{A}$ they define.
2. Compute for each candidate pair (P_α, Q_β) the discrepancy $D(P_\alpha, Q_\beta; \mathcal{A})$ from the estimated affine transformation $\mathcal{A}$ (cf. Appendix), and record the median S of $D(P_\alpha, Q_\beta; \mathcal{A})$ over all the pairs $\{(P_\alpha, Q_\beta)\}$.
3. Repeat this sampling a sufficient number of times and find the affine transformation $\mathcal{A}_m$ that gives the smallest median S_m.
4. Regard those pairs $\{(P_\alpha, Q_\beta)\}$ that satisfy $D(P_\alpha, Q_\beta; \mathcal{A}_m) < 7S_m$ as inliers (cf. Appendix), and optimally fit an affine transformation $\hat{\mathcal{A}}$ to them [5].
5. *Cancel the existing candidate pairs*, and select from *all the points* $\{P_\alpha\}$ and $\{Q_\beta\}$ those pairs $\{(P_\alpha, Q_\beta)\}$ that satisfy $D(P_\alpha, Q_\beta; \hat{\mathcal{A}}) < 7S_m$.
6. Compute the residuals $J(P_\alpha, Q_\beta)$ for all the combinations of the endpoints $\{P_\alpha\}$ and $\{Q_\beta\}$ of the selected pairs, using a 25×25 template. We let $\mathcal{D}_{P_\alpha}^{Q_\beta}$ be the translation that displaces P_α to Q_β followed by an affine transformation centered on Q_β compatible with the estimated affine transformation $\hat{\mathcal{A}}$.
7. Apply the thresholding and uniqueness enforcing procedure [6].

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

[2] In our experiment, we detected 100 points in each image.

[3] Actually, we exhaustively sampled all the candidate pairs.

[4] We repeated the sampling until the minimum median was not updated 100 times consecutively. This criterion was applied to the subsequent procedures, too.

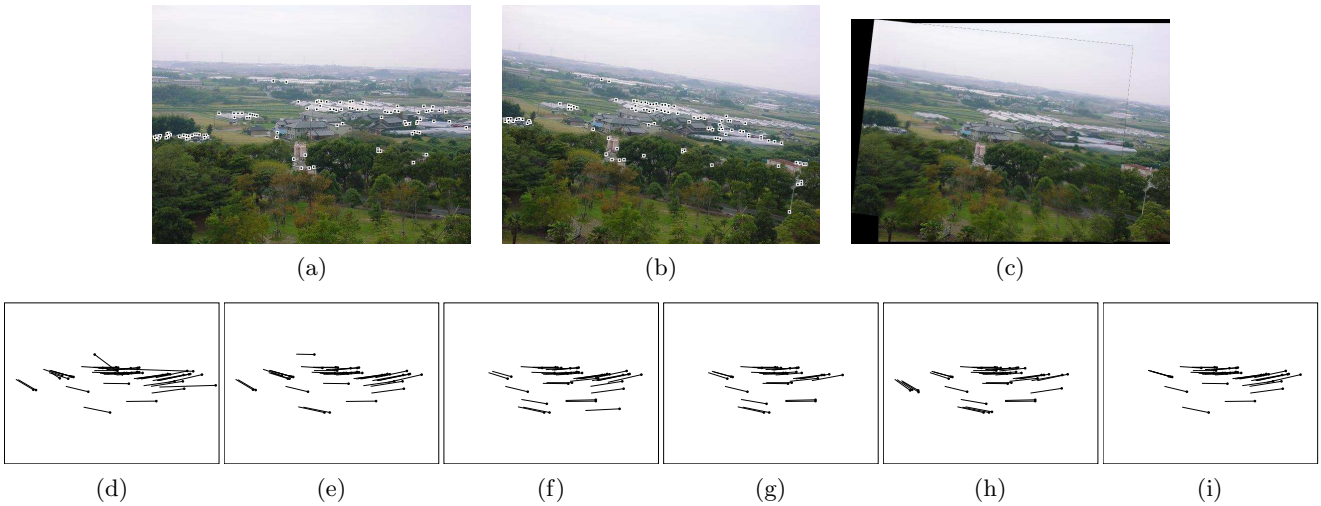


Figure 1: (a), (b) Input images and extracted feature points. (c) Image mosaicing by our method. (d) Initial matches (69.0% correct). (e) Translation matching. (f) Similarity matching (g) Affine matching. (h) Homography matching. (i) Direct estimation by LMedS.

3.5 Homography matching by four-point voting

1. Randomly sample *four* from among the candidate pairs $\{(P_\alpha, Q_\beta)\}$, and compute the homography $\mathcal{H}$ they define.
2. Compute for each candidate pair (P_α, Q_β) the discrepancy $D(P_\alpha, Q_\beta; \mathcal{H})$ from the estimated homography $\mathcal{H}$ (cf. Appendix), and record the median S of $D(P_\alpha, Q_\beta; \mathcal{H})$ over all the pairs $\{(P_\alpha, Q_\beta)\}$.
3. Repeat this sampling a sufficient number of times and find the homography $\mathcal{H}_m$ that gives the smallest median S_m.
4. Regard those pairs $\{(P_\alpha, Q_\beta)\}$ that satisfy $D(P_\alpha, Q_\beta; \mathcal{H}_m) < 7S_m$ as inliers (cf. Appendix), and optimally fit a homography $\hat{\mathcal{H}}$ to them by *renormalization*[5] [7].
5. *Cancel the existing candidate pairs*, and select from *all the points* $\{P_\alpha\}$ and $\{Q_\beta\}$ those pairs $\{(P_\alpha, Q_\beta)\}$ that satisfy $D(P_\alpha, Q_\beta; \hat{\mathcal{H}}) < d^2$, where d is a user-definable admissible discrepancy.
6. Compute the residuals $J(P_\alpha, Q_\beta)$ for all the combinations of the endpoints $\{P_\alpha\}$ and $\{Q_\beta\}$ of the selected pairs using a 33×33 template. We let $\mathcal{D}_{P_\alpha}^{Q_\beta}$ be the translation that displaces P_α to Q_β followed by a homography centered on Q_β compatible with the estimated homography $\hat{\mathcal{H}}$.
7. Apply the thresholding and uniqueness enforcing procedure [6].

The resulting pairs $\{(P_\alpha, Q_\beta)\}$ are our final matches.

3.6 Summary of the procedure

In the above process, we progressively estimate the image transformation by LMedS (see Appendix for the details), *cancel the existing candidate pairs*, and match the points all over again using a new template compatible with the estimated transformation, increasing the template size to 9, 17, 25, and 33. In doing so, we also computed the residuals of the pairs that are not compatible with the estimated transformation. This artificial introduction of incorrect matches is for applying our automatic thresholding scheme[6] [6]. After setting the threshold, these incorrect matches are removed.

Unlike the standard LMedS [9, 11], where we *choose* correct matches from an initial set, we *generate* correct matches by redoing template matching each time a new deformation is estimated, using a new template and changing the threshold. As a result, those pairs initially rejected can be accepted in the later stage.

If we know that the two images should undergo a particular transformation to a specified degree, the best choice may be RANSAC [2]. Here, however, the intermediate transformations are all approximate, and we do know 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 last stage we introduced a user-definable admissible discrepancy p. In our experiment, we set $d = 3$ (pixels).

4. Real Image Examples

We extracted 100 feature points from the images in Figs. 1(a), (b) using Harris operator [3], as marked there. Fig. 1(d) is the "optical flow" (line segments connecting the matching positions) of the initial matches obtained by the procedure described in Sec. 3.1. Figs. 1(e)–(h) show the upgraded matches

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

[6]As is described in [6], the ratio of the percentage p of the correct matches to its maximum value $p_{\max}$ needs to be specified. We gradually increase it to $p/p_{\max} = 0.6$, 0.7, 0.8, and 0.9.

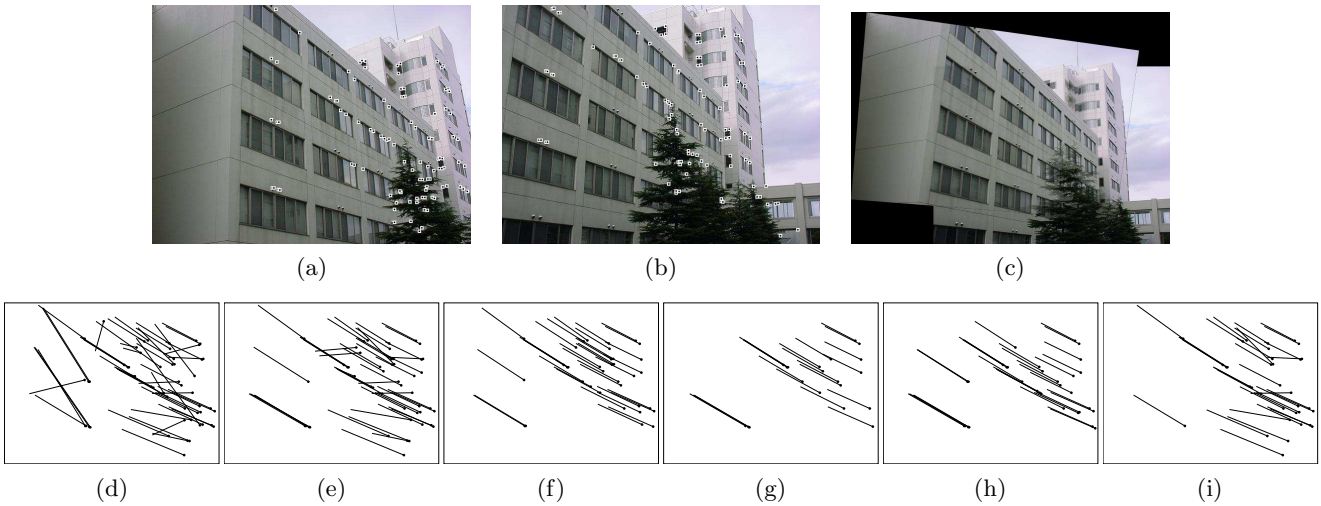


Figure 2: (a), (b) Input images and extracted feature points. (c) Image mosaicing by our method. (d) Initial matches (28.3% correct). (e) Translation matching. (f) Similarity matching (g) Affine matching. (h) Homography matching. (i) Direct estimation by LMedS.

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 [9], directly computing the minimum-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 template matching procedure described in [6] 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. This is because we do not "choose" correct matches but "generate" correct matches by adjusting the template.

Fig. 2 is another example similarly arranged. In this case, the deformation between the images in Figs. 2(a), (b) 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 homography, most of them are compatible with an approximate translation or an approximate similarity with a large error estimate predicted by LMedS (see Appendix). Hence, we can gradually narrow down correct matches by applying a compatible template of a larger size, thereby increasing the number of correct matches and removing outliers.

As is well known, two images undergo a homography when the camera is rotated without translating, the scene is a planar surface, or the scene is sufficiently far away. This is the case for the images in Figs. 1 and 2. In Fig. 3 in contrast, the two images 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 another entirely by a single homography. In this case, the initial matches in Fig. 3(c) are 44.8% correct. Yet, in the end correct matches are generated only in the planar part of the scene, as shown in Fig. 3(d). This is because our method is based on voting and hence minor portions may undergo different transformations.

Fig. 4(e) is the superposition of the mapped images, and Fig. 4(f) displays the absolute value of their difference. The overlapping parts indicate non-planar parts of the scene. This is the well known principle for detecting intruding objects, such as pedestrians and vehicles, on a planar surface, such as the floor and the ground.

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

5. Conclusion

We have presented a new method for automatically matching feature points over two images. 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.

Many image mosaicing techniques have been presented in the past [10, 12], but mostly the image transformations are estimated by directly comparing the gray/color levels of consecutive frames of an image sequence. Such techniques cannot be applied to separate

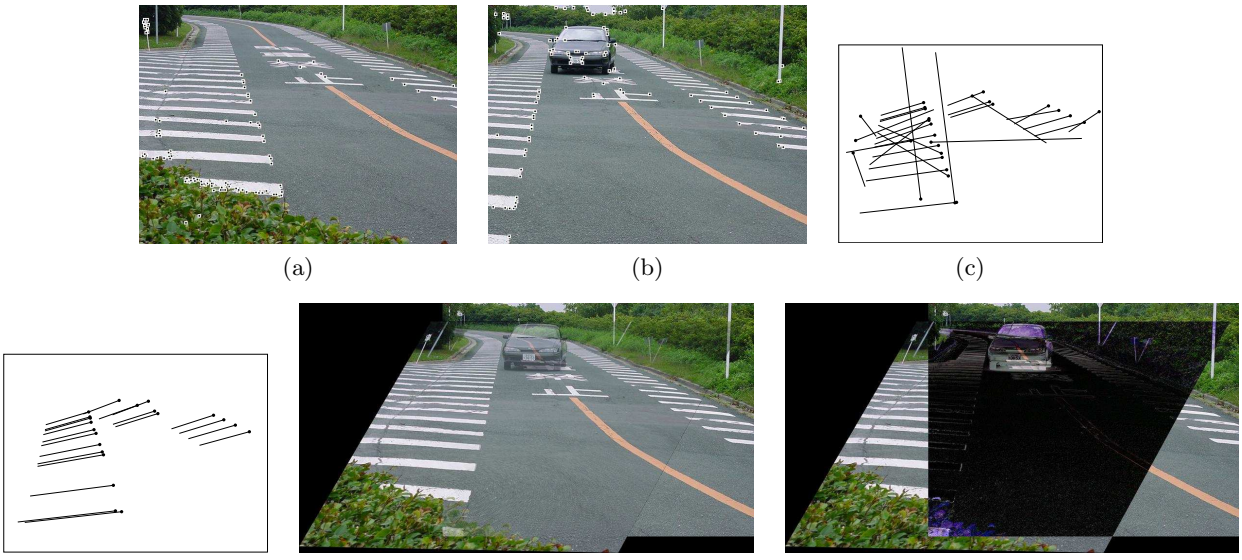


Figure 3: (a), (b) Input images and detected feature points. (c) Initial matches (44.8% correct). (d) Final matches. (e) Image mosaicing. (f) Difference image.

images involving 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 outliers. As a byproduct, intruding objects can be automatically detected.

Acknowledgements: This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, under a Grant in Aid for Scientific Research C(2) (No. 13680432), the Support Center for Advanced Telecommunications Technology Research, and Kayamori Foundation of Informational Science Advancement.

References

[1] P. Beardsley, P. Torr and A. Zisserman, 3D model acquisition from extended image sequences, *Proc. 4th Euro. Conf. Comput. Vision*, April 1996, Cambridge, U.K., Vol. 2, pp. 683–695.

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

[3] C. Harris and M. Stephens, A combined corner and edge detector, *Proc. 4th Alvey Vision Conf.*, August 1988, Manchester, U.K., pp. 147–151.

[4] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, U.K., 2000.

[5] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier Science, Amsterdam, the Netherlands, 1996.

[6] K. Kanatani and Y. Kanazawa, Automatic thresholding for correspondence detection, *Workshop on Statistical Methods in Video Processing*, June 2002, Denmark, Copenhagen.

[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, Vol. 1, pp. 73–78.

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

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

[10] R. Szeliski and H.-Y. Shum, Creating full view panoramic image mosaics and environment maps, *Proc. SIGGRAPH'97*, August 1997, Los Angeles, CA, U.S.A., pp. 251–258.

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

[12] 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. Recog.*, June 1997, Puerto Rico, pp. 420–425.

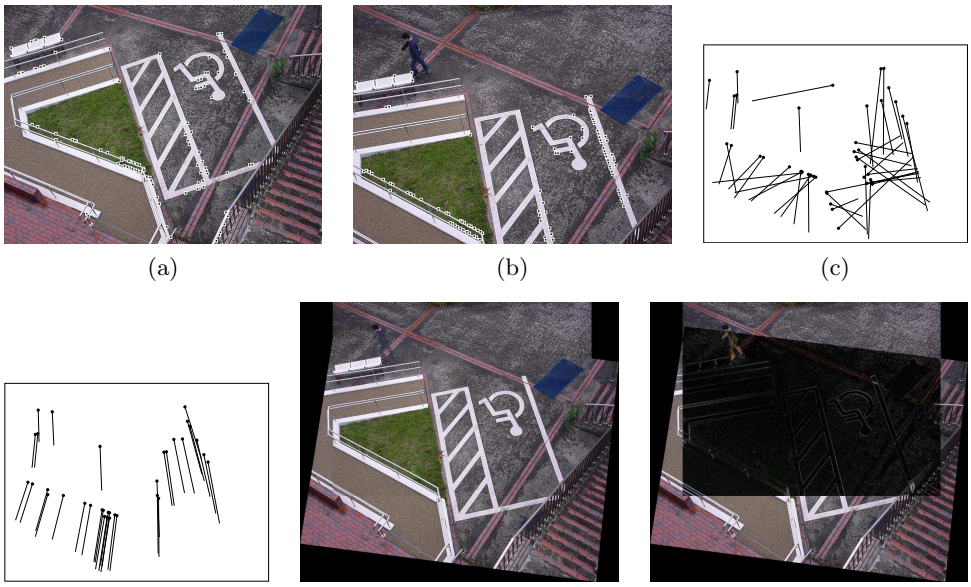
Appendix: LMedS for Geometric Fitting

Geometric fitting [5] is to fit a d-dimensional manifold $\mathcal{M}$ defined by a constraint equation $\boldsymbol{F}(\boldsymbol{x};\boldsymbol{u}) = \boldsymbol{0}$ parameterized by a p-dimensional vector $\boldsymbol{u}$ to N data points $\{\boldsymbol{x}_\alpha\} \in \mathcal{R}^n$. Each data point $\boldsymbol{x}_\alpha$ is assumed to be disturbed from its true position $\bar{\boldsymbol{x}}_\alpha$ by independent Gaussian noise of mean 0 and standard deviation σ in each coordinate. The true positions $\{\bar{\boldsymbol{x}}_\alpha\}$ are assumed to be in the manifold $\mathcal{M}$.

The constraint that a point (x, y) in one image should be mapped to a point (x', y') in another by a translation $\mathcal{T}$, a similarity $\mathcal{S}$, an affine transformation $\mathcal{A}$, or a homography $\mathcal{H}$ defines a 2-dimensional manifold $\mathcal{M}$ with 2, 4, 6, or 8 parameters, respectively, in the 4-dimensional joint space of (x, y, x', y'). Each $\boldsymbol{x}_\alpha$ is a 4-dimensional vector consisting of the coordinates of the corresponding points.

The LMedS [9] for fitting the manifold $\mathcal{M}$ to $\{\boldsymbol{x}_\alpha\}$ is to minimize

$$S = \mathrm{med}_{\alpha=1}^{N} D(\boldsymbol{x}_\alpha; \mathcal{M}), \tag{2}$$



(a) (b) (c) (d) (e) (f)

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

where $D(\boldsymbol{x};\mathcal{M})$ designates the square of the distance of point $\boldsymbol{x}$ from the manifold $\mathcal{M}$—the actual forms of $D(\boldsymbol{x};\mathcal{M})$ for a translation $\mathcal{T}$, a similarity $\mathcal{S}$, an affine transformation $\mathcal{A}$, and a homography $\mathcal{H}$ together with their optimal fitting procedures are found in [8].

Minimization of eq. (2) is done by repeating random sampling of the minimum number $\lceil p/r \rceil$ of points that can define the manifold $\mathcal{M}$ a sufficient number of times, evaluating the median S each time, and choosing the manifold $\hat{\mathcal{M}}$ that gives the minimum median S_m [9].

If the noise is small, $D(\boldsymbol{x}_\alpha;\mathcal{M})/\sigma^2$ is subject to a χ^2 distribution with r degrees of freedom [5], where $r = n - d$ is the *codimension* of $\mathcal{M}$, i.e., the number of independent equations of the constraint $\boldsymbol{F}(\boldsymbol{x};\boldsymbol{u}) = \boldsymbol{0}$. If we put

$$\mu = \mathrm{med}_{\alpha=1}^{N} \frac{D(\boldsymbol{x}_\alpha;\mathcal{M})}{\sigma^2}, \tag{3}$$

the ratio $D(\boldsymbol{x}_\alpha;\mathcal{M})/\sigma^2$ is larger than μ for half the data points and smaller than μ for the remaining half. Hence, μ 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

$$\hat{\sigma}^2 = \frac{S}{\chi^2_{r,50}} \tag{4}$$

Here, the median S is defined by eq. (2) using the true manifold $\mathcal{M}$. In actual computation, it is approximated by the manifold $\hat{\mathcal{M}}$ fitted to $\lceil p/r \rceil$ sample points, so the median S is approximated by

$$S_m = \mathrm{med}_{\alpha=1}^{N} D(\boldsymbol{x}_\alpha;\hat{\mathcal{M}}). \tag{5}$$

Since we repeat sampling so as to minimize this, the sample median S_m is in general smaller than the true median S. So, we apply the correction

$$\hat{\sigma}^2 = \Big(1 + \frac{10}{rN - p}\Big)\frac{S_m}{\chi^2_{r,50}}. \tag{6}$$

The term $rN - p$ in the denominator is introduced to account the fact that (i) if $N = p/r$, the fitted manifold $\hat{\mathcal{M}}$ exactly passes through the data points with 0 median, so the variance σ^2 cannot be estimated and (ii) eq. (4) gives a good approximation for large N. The number 10 in the numerator is determined so as to make eq. (4) agree with the formula in [9] when $r = 1$:

$$\hat{\sigma} = \sqrt{\Big(1+\frac{10}{N-p}\Big)}\sqrt{\frac{S_m}{\chi^2_{1,50}}} \approx 1.4826\Big(1+\frac{5}{N-p}\Big)\sqrt{S_m}. \tag{7}$$

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

$$\frac{D(\boldsymbol{x}_\alpha;\hat{\mathcal{S}})}{\hat{\sigma}^2} \geq \chi^2_{r,\alpha}, \tag{8}$$

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 with $\alpha = 99$, we have

$$D(\boldsymbol{x}_\alpha;\hat{\mathcal{M}}) \geq 6.44\Big(1 + \frac{5}{N-4}\Big)S_m. \tag{9}$$