Proc. Australia-Japan Advanced Workshop on Computer Vision, Sept. 2003, Adelaide Australia, pp. 32–37

Robust Image Matching Preserving Global Consistency

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

Abstract

We present a new method for detecting point matches between two images. The main issue is how to preserve the global consistency of individual matches. Existing methods propagate local smoothness by relaxation or do combinatorial search for an optimal solution. Our method imposes non-local constraints that should be approximately satisfied across the image. We define the "confidence" of such "soft constraints" to all potential matches. The confidence is progressively updated by "mean-field approximation". Finally, the "hard" epipolar constraint is imposed by RANSAC. Using real images, we demonstrate that our method is robust to camera rotations and zooming changes.

1. Introduction

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

The basic principle is local correlation measurement by template matching. Detecting feature points in the first and second images separately using a corner detector [3, 5, 17, 19, 20, 21], we measure the correlation between the neighborhoods of the two points for each candidate pair and match those that have a high correlation. This works very well if one image is a translated copy of the other. However, if the two images are taken from different positions, the corresponding parts in the images are locally deformed, depending on the 3-D shape of that part of the scene. The correlation significantly diminishes if camera rotations or zooming changes take place during the image capturing process.

It follows that additional constraints are necessary. If the scene is a planar surface or in the distance, the two images are related by an image transformation called *homography* [6]. This strong constraint can be combined with voting techniques such as LMedS [18] and RANSAC [4] to match the images robustly [7, 10, 11].

For a general scene, the only available constraint is the *epipolar equation* [6], and various types of voting schemes based on it have been proposed [1, 6, 22, 23, 25]. However, the epipolar equation is a very weak constraint, admitting many unnatural and inconsistent matches.

To resolve this, some global consistency condition that favors "natural" matches is necessary. A typical approach is to define various attributes to individual points and an affinity measure for each point pair and then search for a global match that maximizes the total affinity between matched pairs. Since the solution is a permutation matrix with at most one 1 in each row and column and all 0 elsewhere, this is a very difficult integer programming task.

In order to avoid this difficulty, many approximation schemes have been proposed including replacing the permutation matrix by a real matrix [13, 14, 24], tensor voting [12], combination of the distant transform with hierarchical search [15], relaxation of graph labels [25], and introduction of the multiresolution approach [2] and graph partition algorithms [16]. Still, a large amount of complicated computations are necessary for iterations. Also, because these methods iteratively propagate local similarity throughout the image, we cannot impose global consistency conditions to spatially apart matches directly.

In this paper, we present a matching algorithm that does not involve such iterative propagation. Instead, global consistency is imposed on all potential matches directly. The basic principle is that we require correct matches to be spatially smooth, assuming that the scene does not have an extraordinary 3-D shape. We also assume that the scene is more or less planar or in the distance so that the image transformation can be roughly approximated by a homography.

The main difficulty is how to deal with requirements that should be satisfied "approximately". For example, if two points have low image correlations, we cannot deny the possibility that they may match. Similarly, non-smooth or seemingly inconsistent matches can be correct. We say such violable constraints are *soft* while inviolable constraints such as the epipolar equation are *hard*.

Our strategy is to define to *all* potential matches *confidence values* that normalize the degree of satisfaction of the soft constraints, whatever they are. Then, we select high confidence matches and estimate from them the global properties, from which the confidence values of all potential matches are updated. This scheme resembles what is known as *mean-field approximation* used in statistical mechanics for describing many-body interactions. The confidence update is done progressively from low-level soft constraint to high-level soft constraint. Finally, the hard epipolar constraint is strictly imposed by RANSAC. Using real images, we demonstrate that our method

is robust to camera rotations and zooming changes.

2. Template Matching

We measure the local correlations between the neighborhoods of point p in the first image and point q in the second by the *residual* (*sum of squares*)

$$J(p,q) = \sum_{(i,j)\in\mathcal{N}} |T_p(i,j) - T_q(i,j)|^2, \qquad (1)$$

where $T_p(i,j)$ and $T_q(i,j)$ are the intensity values of the templates defined by cutting out an $w \times w$ pixel region $\mathcal{N}$ centered on p and q, respectively[1]. If we normalize them to $\sum_{(i,j)\in\mathcal{N}} T_p(i,j)^2 = 1$ and $\sum_{(i,j)\in\mathcal{N}} T_q(i,j)^2 = 1$, eq. (1) is equivalent to the use of the *normalized correlation*.

The basic procedure for point matching is as follows. We extract N points p_1, ..., p_N in the first image and M points q_1, ..., q_M in the second, using a feature detector such as the Harris operator [5] and SUSAN [21]. Then, we compute the residuals $\{J(p_\alpha, q_\beta)\}$, $\alpha = 1, ..., N$, $\beta = 1, ..., M$, for all NM combinations of the extracted points. We search the $N \times M$ table of $\{J(p_\alpha, q_\beta)\}$ for the minimum value $J(p_{\alpha^*}, q_{\beta^*})$ 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 $J(p_{\alpha^*}, q_{\beta^*})$ and do the same procedure to the resulting $(N-1)\times(N-1)$ table. Repeating this, we end up with $L = \min(N, M)$ matches. We call this procedure *uniqueness enforcement* with respect to the residual J.

However, this procedure cannot be done directly, since the selected pairs may not be all correct while some of the discarded pairs may be correct. In order to take all potential matches into consideration, we introduce confidence values to all pairs.

3. Confidence of Local Correlations

We define the confidence of local correlations for the pair (p, q) via the *Gibbs distribution* in the form

$$P = e^{-sJ(p,q)}, \qquad (2)$$

so that high confidence is given for a smaller residual $J(p, q)$. Physicists usually put $s = 1/kT$ and call T *temperature*, where k is the Boltzmann constant. If $s = 0$ (or $T = \infty$), we uniformly have $P = 1$ irrespective of the residual $J(p, q)$. As s increases (or T decreases), the confidence of those with large residuals quickly decreases, and ultimately the confidence concentrates only on the smallest residual.

Here, we determine the attenuation constant s (or temperature T) as follows. Among all the NM pairs $\{(p_\alpha, q_\beta)\}$, at most L $(= \min(N, M))$ of them can be correct. We require that the average of the L smallest residuals equal the overall weighted average with respect to the confidence (2). If the NM potential matches (p_α, q_β) are sorted in ascending order of $J(p_\alpha, q_\beta)$ and the λth residual is abbreviated as J_λ, this condition is written in the form

$$\frac{1}{Z}\sum_{\lambda=1}^{NM} J_\lambda e^{-sJ_\lambda} = \bar{J}, \qquad (3)$$

where

$$Z = \sum_{\lambda=1}^{NM} e^{-sJ_\lambda}, \qquad \bar{J} = \frac{1}{L}\sum_{\lambda=1}^{L} J_\lambda. \qquad (4)$$

The solution of eq. (3) is easily computed by Newton iterations to search for the zero of $\Phi(s) = 0$, starting from $s = 0$, where we define

$$\Phi(s) = \sum_{\lambda=1}^{NM} (J_\lambda - \bar{J})e^{-sJ_\lambda}. \qquad (5)$$

Let $P_\lambda^{(0)}$ be the thus defined confidence of local correlations for the λth pair.

4. Confidence of Spatial Consistency

Next, we introduce the confidence of spatial consistency, assuming that the scene does not have an extraordinary 3-D shape. For this, we choose tentative candidates for correct matches by enforcing uniqueness with respect to $P_\lambda^{(0)}$ to those pairs that satisfy[2]

$$P_\lambda^{(0)} > e^{-k^2/2}. \qquad (6)$$

We enumerate the resulting matches by the index $\mu = 1, ..., n_0$ in an arbitrary order. Let $\vec{r}_\mu$ be the 2-dimensional vector that connects the two points of the μth match, starting from the one in the first image and ending at the other in the second. We call it the "flow vector" of the μth match.

Our strategy is to view those matches which are consistent with the resulting "optical flow" $\{\vec{r}_\mu\}$ as more likely to be correct. Specifically, we compute the confidence weighted mean $\vec{r}_m$ and the confidence weighted covariance matrix V of the optical flow by

$$\vec{r}_m = \frac{1}{Z}\sum_{\mu=1}^{n_0} P_\mu^{(0)}\vec{r}_\mu, \qquad Z = \sum_{\mu=1}^{n_0} P_\mu^{(0)},$$
$$V = \frac{1}{Z}\sum_{\mu=1}^{n_0} P_\mu^{(0)}(\vec{r}_\mu - \vec{r}_m)(\vec{r}_\mu - \vec{r}_m)^\top. \qquad (7)$$

Then, we go back to the original NM potential matches. We define their confidence of spatial consistency via the Gaussian distribution in the form

$$P_\lambda^{(1)} = e^{-(\vec{r}_\lambda - \vec{r}_m, V^{-1}(\vec{r}_\lambda - \vec{r}_m))}, \qquad (8)$$

where $(\vec{a}, \vec{b})$ designates the inner product of vectors $\vec{a}$ and $\vec{b}$. Thus, a flow vector $\vec{r}_\lambda$ has low confidence if it largely deviates from the optical flow $\{\vec{r}_\mu\}$.

5. Confidence of Global Smoothness

We then introduce the confidence of global smoothness, assuming that the scene is more or less planar or in the distance so that the image transformation can be roughly approximated by a homography.

[1] We let $w = 9$ in our experiments.

[2] we let $k = 3$ in our experiment.

First, we choose tentative candidates for correct matches. This time, we enforce uniqueness with respect to $P_\lambda^{(0)} P_\lambda^{(1)}$ to those pairs that satisfy

$$P_\lambda^{(0)} P_\lambda^{(1)} > e^{-2k^2/2}. \tag{9}$$

We enumerate the resulting matches by the index μ = 1, ..., n_1 in an arbitrary order.

Let (x_μ, y_μ) and (x'_μ, y'_μ) make the μth pair. We represent these two points by 3-D vectors

$$\boldsymbol{x}_\mu = \begin{pmatrix} x_\mu/f_0 \\ y_\mu/f_0 \\ 1 \end{pmatrix}, \quad \boldsymbol{x}'_\mu = \begin{pmatrix} x'_\mu/f_0 \\ y'_\mu/f_0 \\ 1 \end{pmatrix}, \tag{10}$$

where f_0 is an appropriate scale factor, e.g., the image size. Then, a *homography* is written as an image mapping in the form

$$\boldsymbol{x}' = Z[\boldsymbol{H}\boldsymbol{x}], \tag{11}$$

where $Z[\,\cdot\,]$ means normalization to make the third component 1.

We optimally fit a homography to the n_1 candidate matches. Let the true positions (in the absence of noise) of $\{\boldsymbol{x}_\mu\}$ and $\{\boldsymbol{x}'_\mu\}$ be, respectively, $\{\bar{\boldsymbol{x}}_\mu\}$ and $\{\bar{\boldsymbol{x}}'_\mu\}$. Taking account of their confidence, we compute the homography matrix $\boldsymbol{H}$ by minimizing

$$J = \sum_{\mu=1}^{n_1} P_\mu^{(0)} P_\mu^{(1)} (\|\boldsymbol{x}_\mu - \bar{\boldsymbol{x}}_\mu\|^2 + \|\boldsymbol{x}'_\mu - \bar{\boldsymbol{x}}'_\mu\|^2), \tag{12}$$

subject to the constraint $\bar{\boldsymbol{x}}'_\mu = Z[\boldsymbol{H}\bar{\boldsymbol{x}}_\mu]$, μ = 1, ..., n_1, with respect to $\{\bar{\boldsymbol{x}}_\mu\}$, $\{\bar{\boldsymbol{x}}'_\mu\}$, and $\boldsymbol{H}$. The solution is easily obtained by modifying existing optimization techniques. We used the method of Kanatani and Ohta[3] [8].

Then, we go back to the original NM potential matches. The discrepancy of each potential match from the estimated homography is measured by

$$D_\lambda^H = \|\boldsymbol{x}'_\lambda - Z[\boldsymbol{H}\boldsymbol{x}_\lambda]\|^2, \tag{13}$$

where $\boldsymbol{x}_\lambda$ and $\boldsymbol{x}'_\lambda$ represent the two points of the λth pair, λ = 1, ..., NM. We define the confidence of global smoothness via the Gibbs distribution in the same way as the confidence of local correlations. Namely, we let

$$P_\lambda^{(2)} = e^{-tD_\lambda^H}. \tag{14}$$

The constant t is determined by solving

$$\frac{1}{Z} \sum_{\lambda=1}^{NM} D_\lambda^H e^{-tD_\lambda^H} = \bar{D}^H, \tag{15}$$

where

$$Z = \sum_{\lambda=1}^{NM} e^{-tD_\lambda^H}, \qquad \bar{D}^H = \frac{1}{L} \sum_{\lambda=1}^{L} D_\lambda^H. \tag{16}$$

The solution is easily obtained by doing Newton iterations to eq. (5) after J_λ is replaced by D_λ^H.

[3]We used the program code placed at
`http://www.ail.cs.gunma-u.ac.jp/Labo/programs-e.html`

6. Voting the Epipolar Constraint

Finally, we strictly enforce the *epipolar constraint*. For a matching pair $\{\boldsymbol{x}, \boldsymbol{x}'\}$, the *epipolar equation*

$$(\boldsymbol{x}, \boldsymbol{F}\boldsymbol{x}') = 0. \tag{17}$$

should hold. The matrix $\boldsymbol{F}$ is called the *fundamental matrix* [6].

First, we choose tentative candidates for correct matches by enforcing uniqueness with respect to $P_\lambda^{(0)} P_\lambda^{(1)} P_\lambda^{(2)}$ to those pairs that satisfy

$$P_\lambda^{(0)} P_\lambda^{(1)} P_\lambda^{(2)} > e^{-3k^2/2}. \tag{18}$$

We enumerate the resulting matches by the index μ = 1, ..., n_2 in an arbitrary order. From these candidate matches, we robustly fit the epipolar equation (17) using RANSAC [4, 6]. Letting $S_m = 0$ and $\boldsymbol{F}_m = \boldsymbol{O}$ as initial values, we do the following computation:

1. Randomly choose eight among the n_2 pairs.
2. From them, compute the fundamental matrix $\boldsymbol{F}$.
3. For each of the n_2 pairs, compute

$$D_\mu^F = \frac{(\boldsymbol{x}_\mu, \boldsymbol{F}\boldsymbol{x}'_\mu)^2}{\|\boldsymbol{P}_\mathbf{k}\boldsymbol{F}^\top\boldsymbol{x}_\mu\|^2 + \|\boldsymbol{P}_\mathbf{k}\boldsymbol{F}\boldsymbol{x}'_\mu\|^2}, \tag{19}$$

   where $\boldsymbol{P}_\mathbf{k} = \mathrm{diag}(1, 1, 0)$ (the diagonal matrix with diagonal elements 1, 1, and 0 in that order).

4. Let S the sum of the confidence $P_\mu^{(0)} P_\mu^{(1)} P_\mu^{(2)}$ of those pairs that satisfy

$$D_\mu^F \leq \frac{2d^2}{f_0^2}, \tag{20}$$

   where d (pixel) is a user definable threshold[4].
5. If $S > S_m$, update $S_m \leftarrow S$ and $\boldsymbol{F}_m \leftarrow \boldsymbol{F}$.

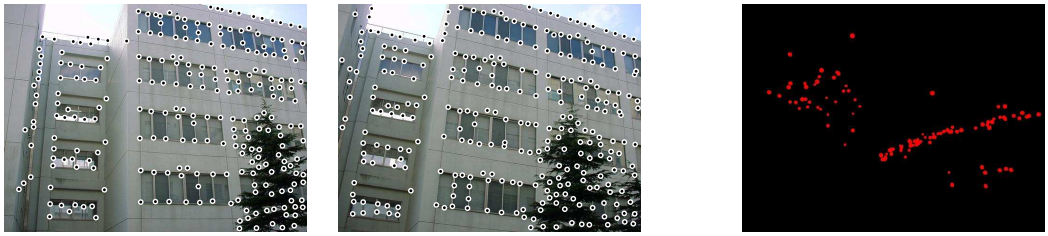
We repeat this computation a sufficient number of times[5] to find the matrix $\boldsymbol{F}_m$ that gives the largest total confidence S_m.

Then, we go back to the original NM potential matches. We measure the degree of fit to the epipolar equation by D_λ^F in eq. (19) after replacing $\boldsymbol{x}_\mu$ and $\boldsymbol{x}'_\mu$, respectively, by $\boldsymbol{x}_\lambda$ and $\boldsymbol{x}'_\lambda$ that represent the λth pair, λ = 1, ..., NM. We choose from among the NM pairs those that satisfy eq. (20). The resulting pairs are thresholded by the criterion (18). Finally, we enforce uniqueness with respect to $P_\lambda^{(0)} P_\lambda^{(1)} P_\lambda^{(2)}$ to obtain the final matches.

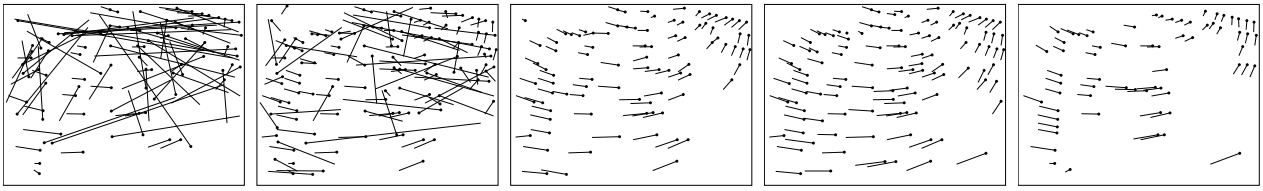
Note that the confidence for different types of constraint can be compared or multiplied on an equal footing, because it is normalized into the interval [0,1] in such a way that the L most favorable matches have approximately the same level of confidence. This is the reason why we used the Gibbs distribution in the form of eqs. (2) and (14) and determined the attenuation constants s and t from the conditions (3) and (15).

[4]We let d = 3 in our experiment.

[5]We stopped the search when no update occurred 100 times consecutively.



Left image. Right image. 3-D reconstruction.



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

Figure 1: Upper row: Input images and 3-D reconstruction. Bottom row: (a) Initial matches based on local correlations. (b) Matches with spatial consistency incorporated. (c) Matches with global smoothness added. (d) Final matches with the epipolar constraint imposed. (e) The method of Zhang et al. [25].

7. Real Image Examples

Using the two images in the upper left of Fig. 1, we detected 300 feature points separately using the Harris operator [5], as marked there. Fig. 1(a) is the "optical flow" of the initial candidate matches based on local correlations (we used the normalized correlation for this example). This scene has many periodic patterns, so the template matching based only on local correlations produces many mismatches.

Fig. 1(b) shows the matches after spatial consistency is imposed; Fig. 1(c) shows the matchers after global smoothness is added. As we can see, the accuracy increases as we impose more constraints. Doing RANSAC to the matches in Fig. 1(c), we obtained the final matches in Fig. 1(d).

For comparison, we used the method of Zhang et al.[6] [25] and obtained the flow shown in Fig. 1(e). As can be seen, our method produces denser matches than their method. This is because the confidence values of mismatches, if they exit in some stage, are likely to decrease in the next stage while the latent correct matches can gain more confidence. As a result, the order of preference changes from one stage to the next, resulting in almost all correct matchers in the end.

The upper right of Fig. 1(c) is the 3-D shape reconstructed from the computed fundamental matrix. We used the method described in [9].

Fig. 2 shows another example similarly arranged. A small camera rotation exists between the left and right images, and the scene has many similar textures, so template matching based on local correlations produces considerable mismatches, as we can see in Fig. 2(a). However, each step adds to correct matches. In the end, denser correct matches are obtained (Fig. 2(d)) than by the method of Zhang et al. [25] (Fig. 2(e)). The upper right of Fig. 2 is the panoramic image generated by the computed homography.

We then examined the effects of camera rotations. The upper row of Figs. 3 shows the left image and two right images. Right images 1 and 2 are rotated approximately by 5 and 10 degrees, respectively, relative to the left image. These images consist in large part of almost identical periodic patterns with very similar textures, so matching by local correlation alone is extremely difficult. The middle row shows the results using the left image and the right image 1; The bottom row shows the results using the left image and the right image 2. In both, (a)∼(e) correspond to (a)∼(e) in Figs. 1 and 2. As we can see, our method successfully generated sufficiently many correct matches even in the presence of camera rotations, but the method of Zhang et al. [25] failed.

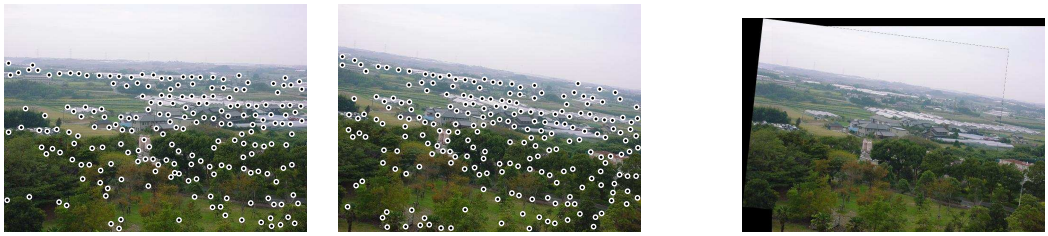
We also examined the effects of zooming changes, and the results are similarly arranged in Figs. 4. This time, right images 1 and 2 are zoomed out approximately by 80% and 65%, respectively, relative to the left image. Again, our method produced a sufficient number of correct matches, while the method of Zhang et al. [25] failed.

For our examples, the total computation time (including loading image files, feature point extraction, and outputting debug information) was 23 sec on average. We used Pentium III 700MHz for the CPU with 768MB main memory and Linux for the OS.

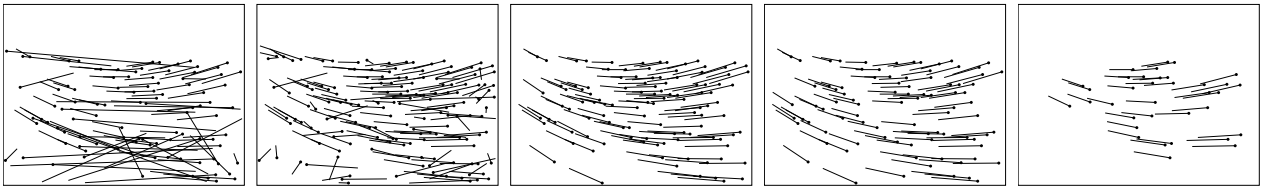
8. Conclusions

We have presented a new method for detecting point matches between two images. Our strategy for preserving the global consistency is to impose nonlocal "soft" constraints on all potential matches via their "confidence values" that normalize the degree of satisfaction of different types of constraint. The confidence values are progressively updated by "mean-field approximation". Finally, the "hard" epipolar constraint is imposed by RANSAC. Using real images, we have demonstrated that our method is robust to

[6] We used the program from placed at `http://www-sop.inria.fr/robotvis/personnel/zzhang/softwares.html`.



Left image. Right image. Panoramic image.



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

Figure 2: Upper row: Input images and the generated panoramic image. Bottom row: (a) Initial matches based on local correlations. (b) Matches with spatial consistency incorporated. (c) Matches with global smoothness added. (d) Final matches with the epipolar constraint imposed. (e) The method of Zhang et al. [25]

camera rotations and zooming changes[7].

Acknowledgments: 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.

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] J. R. Bergen and P. Anandan, K. J. Hanna and R. Higorani, Hierarchical model-based motion estimation, *Proc. 2nd Euro. Conf. Comput. Vision*, May 1992, Santa Margerita, Italy, pp. 237–252.

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

[4] 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), pp. 381–395.

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

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

[7] K. Kanatani and Y. Kanazawa, Automatic thresholding for correspondence detection, *Int. J. Image Graphics*, **3** (2003), to appear.

[8] K. Kanatani and N. Ohta, Accuracy bounds and optimal computation of homography for image mosaicing applications, *Proc. 7th Int. Conf. Comput. Vision*, September 1999, Kerkyra, Greece, Vol. 1, pp. 73–78.

[9] K. Kanatani and N. Ohta, Comparing optimal 3-D reconstruction for finite motion and optical flow, *J. Elec. Imaging*, **12** (2003), to appear.

[10] Y. Kanazawa and K. Kanatani, Image mosaicing by stratified matching, *Workshop on Statistical Methods in Video Processing*, June 2002, Denmark, Copenhagen, pp. 31–36.

[7] The source program is available at
`http://www.img.tutkie.tut.ac.jp/programs/index-e.html`

[11] M. I. A. Lourakis, S. V. Tzurbakis, A. A. Argyros and S. C. Orphanoudakis, Feature transfer and matching in disparate stereo views through the use of plane homography, *IEEE Trans. Patt. Anal. Mach. Intell.*, **25**-2 (2003), 271–276.

[12] M.-S. Lee, G. Medioni and P. Mordohai, Inference of segmented overlapping surfaces from binocular stereo, *IEEE Trans. Patt. Anal. Mach. Intell.*, **24**-6 (2002), 824–837.

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

[14] J. Maciel and J. Costeira, A global solution to sparse correspondence problems, *IEEE Trans. Patt. Anal. Mach. Intell.*, **25**-2 (2003), 187–199.

[15] C. F. Olson, Maximum-likelihood image matching, *IEEE Trans. Patt. Anal. Mach. Intell.*, **24**-6 (2002), 853–857.

[16] Pilu, A direct method for stereo correspondence based on singular value decomposition, *Proc. IEEE Conf. Comput. Vision Patt. Recog.*, Puerto Rico, June 1997, pp. 261–266.

[17] D. Reisfeld, H. Wolfson and Y. Yeshurun, Context-free attentional operators: The generalized symmetry transform, *Int. J. Comput. Vision*, **14**-2 (1995), 119–130.

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

[19] F. Schaffalitzky and A. Zisserman, Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?", *Proc. 7th Euro. Conf. Comput. Vision*, May 2002, Copenhagen, Denmark, Vol 1, pp. 414–431.

[20] C. Schmid, R. Mohr and C. Bauckhage, Evaluation of interest point detections, *Int. J. Comput. Vision*, **37**-2 (2000), 151–172.

[21] S. M. Smith and J. M. Brady, SUSAN—A new approach to low level image processing, *Int. J. Comput. Vision*, **23**-1 (1997), 45–78.

[22] P. H. S. Torr and A. Zisserman, MLESAC: A new robust estimator with application to estimating image geometry, *Comput. Vision Image Understand.*, **78** (2000), 138–156.

[23] P. H. S. Torr and C. Davidson, IMPSAC: Synthesis of importance sampling and random sample consensus, *IEEE Trans. Patt. Anal. Mach. Intell.*, **25**-3 (2003), 354–364.

[24] M. A. van Wyk, T. S. Durrani and B. J. van Wyk, A RKHS interpolator-based graph matching algorithm, *IEEE Trans. Patt. Anal. Mach. Intell.*, **24**-7 (2002), 988–995.

[25] 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.

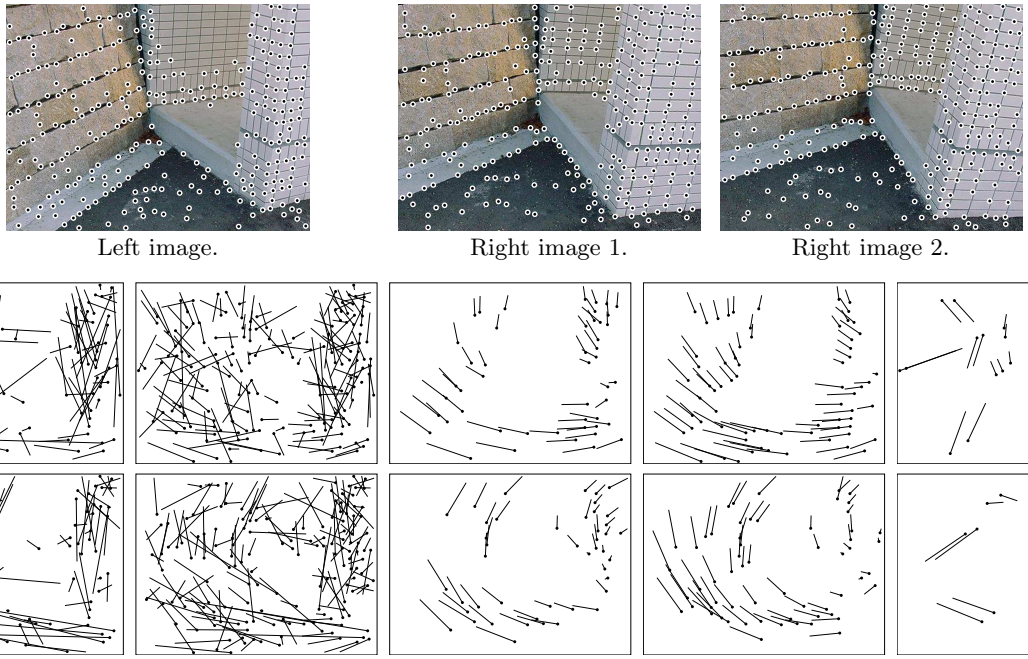


Figure 3: Upper row: Input images. The right images 1 and 2 are rotated approximately by 5° and 10°, respectively, relative to the left image. Middle row: Results using the left image and the right image 1. Bottom row: Results using the left image and the right image 2. (a) Initial matches based on local correlations. (b) Matches with spatial consistency incorporated. (c) Matches with global smoothness added. (d) Final matches with the epipolar constraint imposed. (e) The method of Zhang et al. [25].

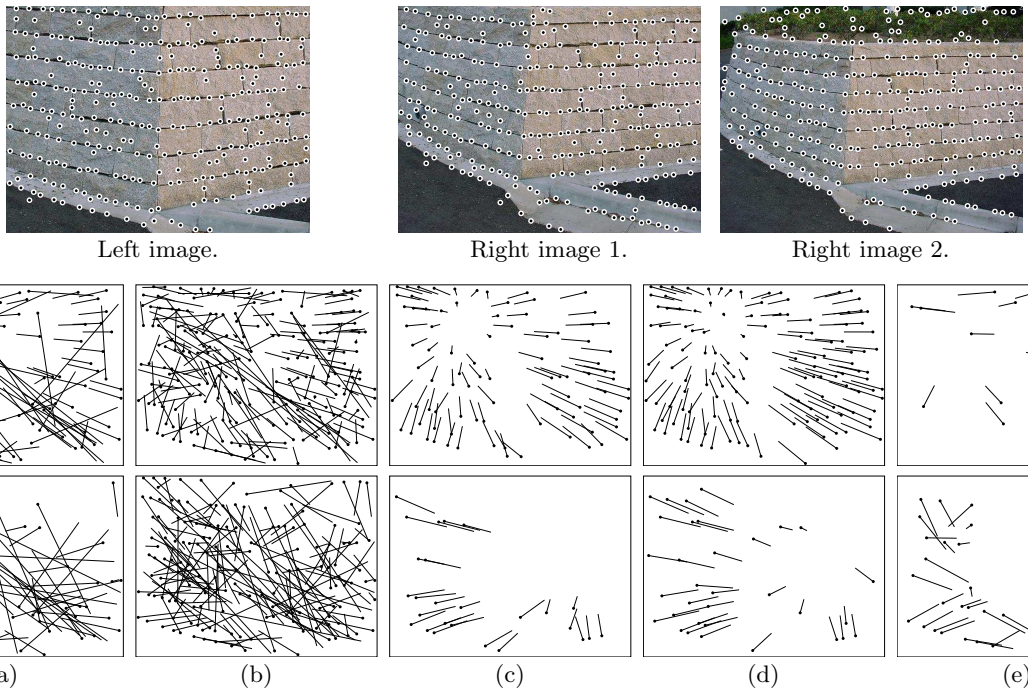


Figure 4: Upper row: Input images. The right images 1 and 2 are zoomed out approximately by 80% and 65%, respectively, relative to the left image. Middle row: Results using the left image and the right image 1. Bottom row: Results using the left image and the right image 2. (a) Initial matches based on local correlations. (b) Matches with spatial consistency incorporated. (c) Matches with global smoothness added. (d) Final matches with the epipolar constraint imposed. (e) The method of Zhang et al. [25].