Memoirs of the Faculty of Engineering, Okayama University, Vol. 37, No. 1, pp. 29–36, November 2002

Robust Image Matching under a Large Disparity

Yasushi KANAZAWA
Department of Knowledge-based Information Engineering
Toyohashi University of Technology
Toyohashi, Aichi 441-8580 Japan

Kenichi KANATANI*
Department of Information Technology
Okayama University
Okayama 700-8530 Japan

(Received September 27, 2002)

We present a new method for detecting point matches between two images without using any combinatorial search. Our strategy is to impose various local and non-local constraints as "soft" constraints by introducing their "confidence" measures via "mean-field approximations". The computation is a cascade of evaluating the confidence values and sorting according to them. In the end, we impose the "hard" epipolar constraint by RANSAC. We also introduce a model selection procedure to test if the image mapping can be regarded as a homography. We demonstrate the effectiveness of our method by real image examples.

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 feature detector [3, 15], 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 in the scene, the corresponding parts in the images appear differently with local deformations that depend on the 3-D shape of that part of the scene and its orientation relative to the two camera positions. The template correlation is very much affected by such local deformations. In particular, the correlation significantly diminishes if the camera is rotated or zooming takes place during the image capturing process.

It follows that local correlations alone are not sufficient for establishing correspondences. If the scene is a planar surface or in the distance, the two images are related by an image transformation called *homography* [4]. This strong constraint can be combined with voting techniques such as LMedS [14] and RANSAC [2] to match the images robustly. Analyzing the template residual [7], the authors have proposed a hierarchical voting scheme called *stratified matching* for robustly matching two images even in the presence of a large image deformations [10].

For a general scene, however, the only available constraint is what is known as the *epipolar equation* [4], and various voting schemes using it have been proposed [1, 4, 19]. This gives a strong constraint if the *fundamental matrix* is known (e.g., when a calibrated stereo rig is used). For an unknown fundamental matrix, however, it is a very weak constraint, and a lot of wrong matches, which humans could easily detect, satisfy this equation. As a result, the combination of local image correlations and the epipolar constraint works only when the disparity is relatively small with small camera rotations and small zooming changes.

To improve the matching performance, many techniques based on combinatorial optimization have been proposed: we assign various attributes to each feature point, define a similarity measure between them, and search for an optimal match that maximizes the total similarity. The difficulty of this approach is that the optimization needs to be done with respect to a permutation matrix with integer elements that define one-to-one correspondence with some unmatched points admitted. Various efforts have been made to relax the integer matrix to real variables [12, 18]. Beside the attribute similarities, local spatial proximity is also incorporated in the form of con-

*E-mail kanatani@suri.it.okayama-u.ac.jp

straint relaxation [19], tensor voting [11] and multiresolution search coupled with the distance transform [13]. However, it is difficult to impose non-local constraints among mutually far apart multiple pairs.

In this paper, we present a robust matching algorithm that does not involve any combinatorial search: the computation is a cascade of evaluating the *confidence values* and *sorting* according to them. We also impose *non-local constraints* such that correct matches be fairly smooth and consistent, assuming that the scene does not have an extraordinary 3-D shape. However, they should not be imposed definitively, since seemingly inconsistent matches can be correct. Let us say such violable constraints are *soft* while inviolable constraints such as the epipolar equation are *hard*.

In order to impose soft constraints, we introduce *confidence measures* to *all* potential matches in such a way that those that satisfy the constraints well have high confidence, yet none of them is definitively rejected. To describe non-local constraints, we introduce *mean-field approximations* similar to what statistical physicists do to deal with many body problems. On the other hand, the hard epipolar constraint is strictly imposed by RANSAC. In other words, we favor, among those matches that strictly satisfy the epipolar equation, those that have high confidence of the soft constraints.

We first introduce the soft constraints that we require and define their confidence measures via mean-field approximations. Then, we describe the voting procedure that combines the soft constraints and the hard epipolar equation. We also describe a model selection procedure to test if the image mapping can be regarded as a homography. Finally, we show real image examples to demonstrate that our method is very effective even when conventional methods fail.

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].

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 [3, 15]. Computing 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. The computation can be done efficiently if the residuals $\{J(p_\alpha, q_\beta)\}$ are sorted in ascending order in the beginning. 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 a confidence measure to all possible 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 assigned to small residuals $J(p, q)$. Physicists 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 (*condensation*).

Here, we determine the value of s 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 according to their residuals $J(p_\alpha, q_\beta)$ in ascending order 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.

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

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 so that correct matches are fairly consistent.

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

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

We enumerate the resulting matches by the index μ = 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. Let us 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

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

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

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

where $(\vec{a}, \vec{b})$ designates the inner product of vectors $\vec{a}$ and $\vec{b}$.

Unlike $P_\lambda^{(0)}$, this is a non-local measure: a flow vector $\vec{r}_\lambda$ has low confidence if it largely deviates from the average of high-confidence optical flow $\{\vec{r}_\mu\}$. This *mean-field approximation* is in the same spirit of what statistical physicists do to deal with many body problems.

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.

As in the case of spatial consistency, we first 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

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

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

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) be the two points of 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. In this vector representation, 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 values (in the absence of noise) of $\{\boldsymbol{x}_\mu\}$ and $\{\boldsymbol{x}'_\mu\}$ be, respectively, $\{\bar{\boldsymbol{x}}_\mu\}$ and $\{\bar{\boldsymbol{x}}'_\mu\}$ (their third components are identically 1). 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 the optimization technique called *renormalization*[3] [8]

Now, 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.

Again, the confidence $P_\lambda^{(2)}$ is a non-local measure: a pair has high confidence if it conforms to the homography that approximates high-confidence matches. This is also a mean-field approximation.

[3] The C++ program is publicly available at `http://www.suri.it.okayama-u.ac.jp/`

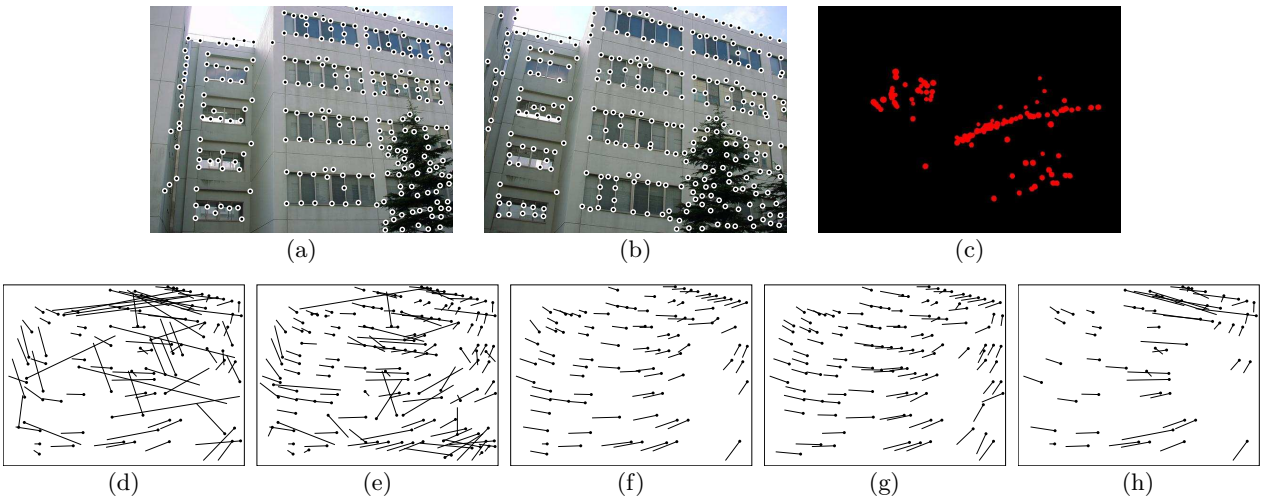

Figure 1: (a), (b) Input images and detected feature points. (c) 3-D reconstruction computed from the final matches (top view). (d) Initial matches based on local correlations. (e) Matches with spatial consistency incorporated. (f) Matches with global smoothness added. (g) Final matches with the epipolar constraint imposed. (h) Matches obtained from (d) by direct RANSAC.

6. Voting the Epipolar Constraint

Finally, we strictly enforce the *epipolar constraint*. In vector representation, we have for a matching pair $\{\boldsymbol{x}, \boldsymbol{x}'\}$ the following *epipolar equation* [4]:

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

Here, $\boldsymbol{F}$ is a singular matrix with rank 2, called the *fundamental matrix* [4]. It is defined only up to an arbitrary scale factor.

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 $\mu = 1, ..., n_2$ in an arbitrary order. From these candidate matches, we robustly fit the epipolar equation (17) using RANSAC [2, 4]. 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 $\boldsymbol{F}$. Since the scale of $\boldsymbol{F}$ is indeterminate, the nine elements of $\boldsymbol{F}$ are easily obtained from the eight pairs by solving linear equations.
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). It can be shown that $f_0^2 D_\mu^F$ equals the sum of square distances of the points $\boldsymbol{x}_\mu$ and $\boldsymbol{x}'_\mu$ to their epipolar lines defined by $\boldsymbol{F}$ if higher order terms are ignored [4, 5].

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 to find the matrix $\boldsymbol{F}_m$ that gives the largest total confidence S_m.

Now, we go back to the original NM potential matches. The degree of fit to the epipolar equation is measured by D_λ^F in eq. (19) if $\boldsymbol{x}_\mu$ and $\boldsymbol{x}'_\mu$ are replaced, respectively, by $\boldsymbol{x}_\lambda$ and $\boldsymbol{x}'_\lambda$ that represent the λth pair, $\lambda = 1, ..., NM$. 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.

7. Model Selection

If the scene is a planar surface or in the distance, the image mapping between the two images is a homography, which allows us to determine the image mapping pixelwise. This information can be used to generate a panoramic image [10], whereas we cannot reconstruct the 3-D structure of the scene without additional information. So, it is crucial to see if the image mapping is a homography.

A naive idea is to fit the homography relation and the epipolar constraint to the obtained matches and decide that the mapping is a homography if the residual for the homography is smaller than for the epipolar constraint. This does not work, however, because

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

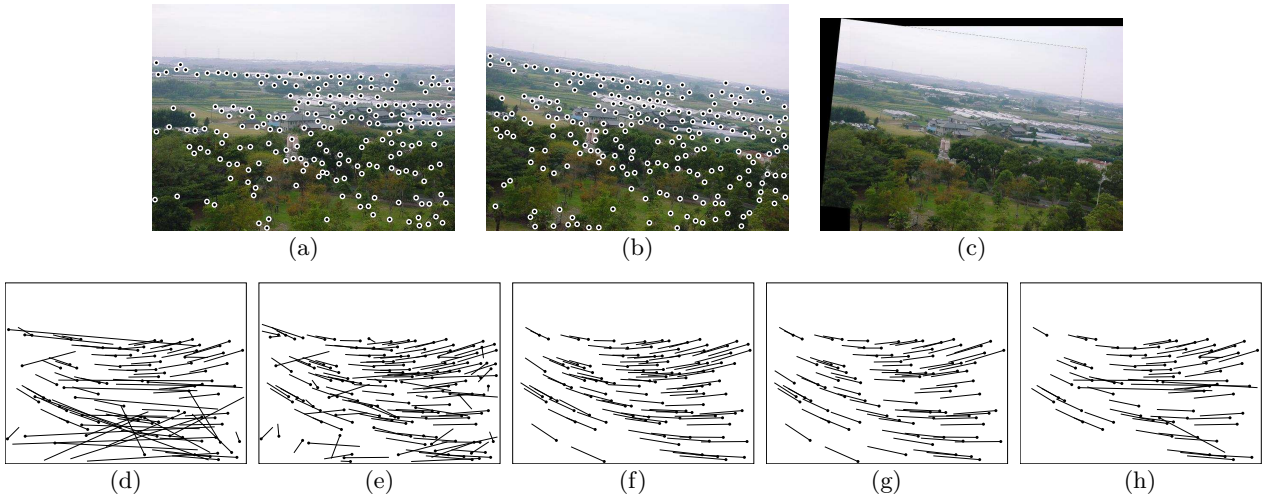

Figure 2: (a), (b) Input images and detected feature points. (c) Panoramic image generated from the final matches. (d) Initial matches based on local correlations. (e) Matches with spatial consistency incorporated. (f) Matches with global smoothness added. (g) Final matches with the epipolar constraint imposed. (h) Matches obtained from (d) by direct RANSAC.

the homography is a stronger constraint than the epipolar equation, which is always satisfied. As a result, the residual for the homography is never smaller than for the epipolar constraint.

Thus, we need to balance the strength of the constraint with the increase in the residual. This can be done by *model selection* using, among other measures, the *geometric AIC* [6] (see [16, 17] for other criteria). Let n be the number of the finally obtained matches. We fit the homography relation and the epipolar constraint to the n matches in an optimal manner and compute their respective residuals J^H and J^F (see [5] for the details of the computation). Their geometric AICs are respectively given by

$$\begin{aligned} \text{G-AIC}^H &= J^H + 2(2n+8)\epsilon^2, \\ \text{G-AIC}^F &= J^F + 2(3n+7)\epsilon^2, \end{aligned} \tag{21}$$

where ϵ is a constant that specifies the degree of image noise. It is estimated from the residual of the general epipolar constraint as follows [5]:

$$\hat{\epsilon}^2 = \frac{J^F}{n-7}. \tag{22}$$

The image relationship is regarded as a homography if $\text{G-AIC}^H < \text{G-AIC}^F$.

8. Real Image Examples

Figs. 1(a) and (b) show two images of an outdoor scene. We detected 300 feature points from each image using the Harris operator [3], as marked there. Fig. 1(d) shows the "optical flow" of the initial candidate matches based on local correlations. As we can see, this scene has many periodic patterns, so the template matching based only on local correlations produce many mismatches.

Fig. 1(e) is the matches after spatial consistency is imposed, and Fig. 1(f) is the matchers after global smoothness is added. We see that the accuracy increases as we impose more constraints. Doing RANSAC to the matches in Fig. 1(f), we obtained the final matches in Fig. 1(g).

For comparison, the optical flow obtained by directly doing RANSAC to the initial matches in Fig. 1(d) is shown in Fig. 1(h). As we can see, many wrong matches are included.

Thus, our procedure is very effective in narrowing down the correct matches by gradually incorporating various soft constraints through their confidence measures. As a result, we can obtain correct matches accurately and robustly even when the initial matches contain a large number of mismatches.

Comparing the geometric AICs computed from the final matches, we can conclude that in this case the image mapping cannot be regarded as a homography. Fig. 1(c) shows the 3-D shape reconstructed from the computed fundamental matrix $\boldsymbol{F}$ (top view). We used the method described in [9].

Fig. 2 shows another example similarly arranged. There is a slight camera rotation between Figs. 2(a) and (b), and the scene has many similar textures, so a lot of mismatches occurred by template matching based on local correlations alone. However, our method successfully generated many correct matches as compared with direct RANSAC. This time, the image transformation can be regarded as a homography. Fig. 2(c) shows the panoramic image generated by the computed homography.

We also examined the effects of camera rotations and zooming changes using the images in Figs. 3 and 4. These images consist in large part of almost identical periodic patterns with very similar textures, so

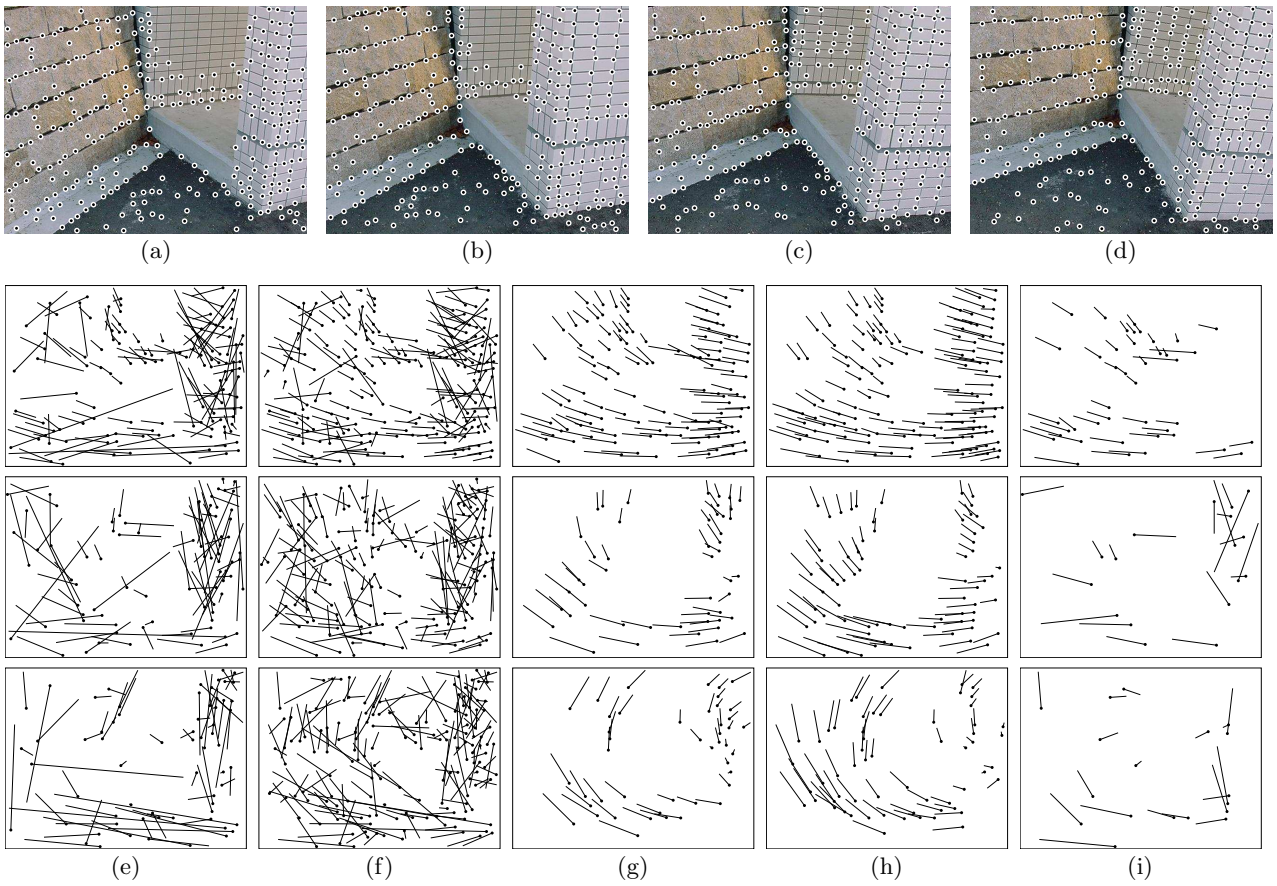

Figure 3: Effects of image rotations: (a) Left image. (b) Right image. (c) Rotation of (b) by 5°. (d) Rotation of (b) by 10°. The results from the pair (a) and (b) (top row), the pair (a) and (c) (middle row), and the pair (a) and (d) (bottom row): (e) Initial matches based on local correlations. (f) Matches with spatial consistency incorporated. (g) Matches with global smoothness added. (h) Final matches with the epipolar constraint imposed. (i) Matches obtained from (e) by direct RANSAC.

matching by local correlation alone is extremely difficult. Yet, our method successfully generated sufficiently many correct matches.

Figs. 3(a) and (b) are the original image pair, and Figs. 3(c) and (d) are obtained by rotating the image in Figs. 3(b) by 5 and 10 degrees, respectively. The matches in the top row are obtained from the pair (a) and (b); the matches in the middle row are obtained from the pair (a) and (c); the matches in the bottom row are obtained from the pair (a) and (d). In each row, (e)∼(i) are the result corresponding to (d)∼(h) in Figs. 1 and 2.

Figs. 4(a) and (b) are another image pair, and Figs. 4(c) and (d) are obtained by zooming out the image in Figs. 4(b) by 80% and 65%, respectively. The rest is arranged in the same way as in Figs. 3.

From these, we can see that our method is robust to image rotations and zooming changes. For the above examples, the total computation time of our method (including loading image files, feature point extraction, and outputting debug information) was 23 sec on average, while direct RANSAC took 14 sec on average. We used Pentium III 700MHz for the CPU with 768MB main memory and Linux for the OS. Thus, we can gain accuracy and robustness at a relatively small computational cost.

9. Summary of the Procedure

The basic strategy of our method is that we assign the confidence value to *all* pairs but observe only those matches obtained by enforcing uniqueness to high-confidence pairs. In other words, all the pairs are divided into *visible matches* and *suppressed matches*. The confidence values of all the pairs are updated using the visible matches, but the order of confidence changes as a result. So, some of the visible matches are denied and some of the suppressed matches emerge as visible matches in the next stage. That is why we can obtain correct matches in the end even if the initial visible matches are almost entirely wrong, as we have seen in our experiments.

Some of the constraints are *non-local* and defined via *mean-field approximations*. They are treated as *soft* constraints through their confidence values. The necessary computation is a cascade of evaluating the confidence values and sorting according to them with-

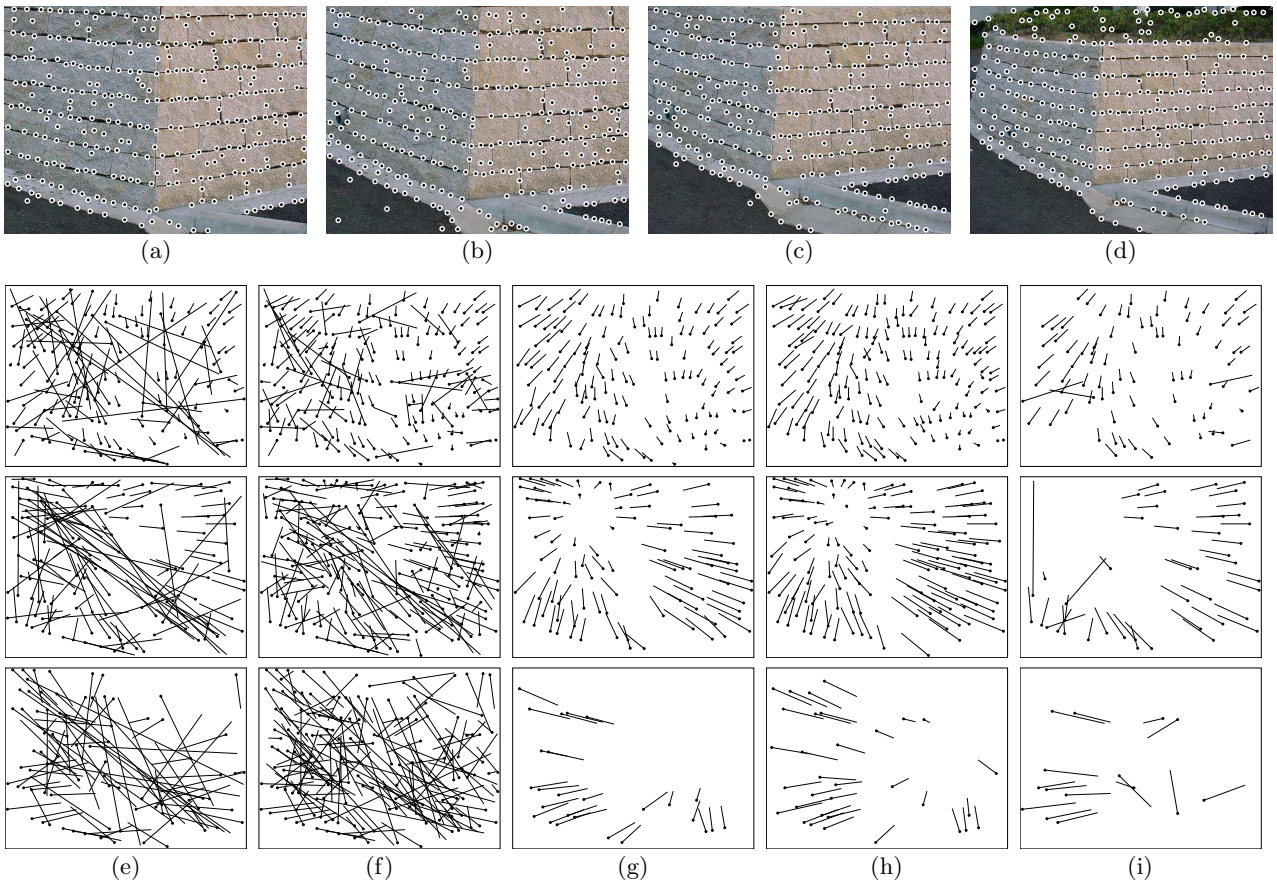

Figure 4: Effects of zooming changes: (a) Left image. (b) Right image. (c) Zooming out of (b) by 80%. (d) Zooming out of (b) by 65%. The results from the pair (a) and (b) (top row), the pair (a) and (c) (middle row), and the pair (a) and (d) (bottom row): (e) Initial matches based on local correlations. (f) Matches with spatial consistency incorporated. (g) Matches with global smoothness added. (h) Final matches with the epipolar constraint imposed. (i) Matches obtained from (e) by direct RANSAC.

out involving any combinatorial search.

In order for this strategy to work, different confidence values should be defined in a compatible way. We have carefully defined them so that the distribution of the confidence values over all the pairs have the variance of a comparative order of magnitude via the Gaussian distribution and the Gibbs distribution with a consistently defined "temperature". As a result, multiple confidence values can be multiplied to give the total confidence. In the definition of our confidence measures, we used $L = \min(N, M)$ as a reference value for the number of matching. It has been pointed out that without such a reference we might accept all the pairs or deny them altogether [7, 12].

We do not assert that the final matches obtained by our procedure are all correct. Rather, we assert that they have *high total confidence* with each match given its confidence value, allowing further refinement or selection, depending on applications.

10. Conclusions

We have presented a robust method for detecting point matches between two images without using any combinatorial search. Our strategy is to impose various local and non-local constraints as "soft" constraints by introducing "confidence" measures. The computation is a cascade of evaluating the confidence values and sorting according to them. In the end, we impose the "hard" epipolar constraint by RANSAC. We have also introduced a model selection procedure to test if the image mapping can be regarded as a homography. We have demonstrated the effectiveness of our method by real image examples.

Acknowledgments: 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 con-

sensus: 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, Geometric information criterion for model selection, *Int. J. Comput. Vision*, **26**-3 (1998), 171–189.

[7] K. Kanatani and Y. Kanazawa, Automatic thresholding for correspondence detection, *Proc. Workshop on Statistical Methods in Video Processing*, June 2002, Denmark, Copenhagen, pp. 19–24.

[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, *Memoirs of the Faculty of Engineering, Okayama University*, **36**-1 (2001), 91–106.

[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.

[11] 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.

[12] J. Maciel and J Costeira, Robust point correspondence by concave minimization, *Image Vision Comput.*, **20**-9/10 (2002-8), 683–690.

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

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

[15] 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.

[16] P. H. S. Torr, An assignment of information criteria for motion model selection, *Proc. IEEE Conf. Comput. Vision Patt. Recog.*, Puerto Rico, June 1997, pp. 47–53.

[17] P. H. Torr and A. Zisserman, Robust detection of degenerate configurations while estimating the fundamental matrix, *Comput. Vision Image Understand.*, **71**-3 (1998), 312–333.

[18] 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.

[19] 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.