9th Symposium on Sensing via Image Information (SSII2003), June 2003, Yokohama, Japan, pp. 519–524.

Extending Interrupted Feature Point Tracking for 3-D Affine Reconstruction

Yasuyuki Sugaya Takayuki Tsubouchi Kenichi Kanatani
Okayama University, Okayama 700-8530 Japan
E-mail: {sugaya,tubo,kanatani}@suri.it.okayama-u.ac.jp

Abstract

Feature point tracking over a video sequence fails when the points go out of the field of view or behind other objects. Motivated by 3-D reconstruction applications, we extend such interrupted tracking by imposing the constraint that under the affine camera model feature trajectories should be in an affine space in the parameter space. Our method consists of iterations for optimally extending the trajectories so that they are compatible with the estimated affine space and optimally estimating the affine space from the extended trajectories, coupled with an outlier removal process based on a statistical model of image noise. Using real video images, we demonstrate that our method can restore sufficiently many trajectories for accurate and detailed 3-D reconstruction.

1. Introduction

The factorization method of Tomasi and Kanade [15] can reconstruct the 3-D shape of the scene from feature point trajectories tracked over a video sequence. The computation is very efficient, requiring only linear operations. The solution is sufficiently accurate for many practical purposes and can be used as an initial value for iterations of a more sophisticated reconstruction procedure [3].

However, the feature point tracking fails when the points go out of the field of view or behind other objects. This can be avoided if we use a short sequence, but then we cannot obtain sufficient information for 3-D reconstruction. In order to resolve this problem, it is necessary to extend interrupted trajectories so that they cover the entire image sequence. There have been several such attempts in the past.

Tomasi and Kanade [15] first reconstructed the 3-D positions of partly visible feature points from their visible image positions and then reprojected them onto the frames in which they are invisible. The camera positions were estimated from other visible feature points.

Kamijima and Saito [7] projectively reconstructed the tentative 3-D positions of missing points by sampling two frames in which they are visible and reprojected them onto the frames in which they are invisible. The camera positions were computed up to projectivity.

Using the knowledge that the trajectories of feature points should be in a 4-dimensional subspace in the parameter space, Jacobs [5] randomly sampled four trajectories, constructed a high dimensional subspace from them by letting the missing data have free values, and computed its orthogonal complement. He repeated this many times and fitted by least squares a 4-dimensional subspace that was best orthogonal to the resulting orthogonal complements[1]. Partial trajectories were extended so that they were compatible with the estimated 4-dimensional subspace. A similar method was also used by Kahl and Heyden [6].

Brandt [1] estimated the centroid of the feature points from their incomplete coordinates by assuming a camera model, reconstructed their 3-D positions and reprojected them onto all the frames. Iterating this, he optimized the camera model and the feature positions, starting from the solution given by the method of Jacobs [5].

For all these methods, we should note the following:

- We need not reconstruct a tentative 3-D shape. 3-D reconstruction is made possible by some geometric constraints over multiple frames, so one can directly map 2-D point positions to other frames by using those constraints[2].
- If a minimum number of frames are sampled for tentative 3-D reconstruction, the accuracy depends on the sampled frames. Rather, one should make full use of all information contained in all frames.
- All existing methods are based on the assumption that the observed trajectories are correct, but this is not always the case. Incorrect trajectories are useless even if they are of full length, and extension of partial trajectories is meaningless if they are incorrect.

In this paper, we present a new scheme for extending partial trajectories based on the constraint that under the affine camera model all trajectories should be in a 3-dimensional affine space in the parameter space [10], which is a stronger requirement than that used by Jacobs [5]. No reprojection of tentative 3-D reconstruction is involved.

Like the method of Brandt [1], our method consists of iterations for optimally extending the trajectories

[1] In actual computation, he interchanged the roles of points and frames: he sampled two frames, i.e., two lists of x coordinates and two lists of y coordinates. The mathematical structure is the same.

[2] The projective reconstruction of Kamijima and Saito [7] is equivalent to the use of what is known as the *trilinear* (or *trifocal*) *constraint* [3].

so that they are compatible with the estimated affine space and optimally estimating the affine space from the extended trajectories.

Exact maximal likelihood estimation may be possible, e.g., by using the method of Shum, et al. [13], but it involves complicated iterations. Here, we simplify the procedure for efficiency by introducing to each partial trajectory a weight that reflects its length.

We do not assume that the observed trajectories are correct. In every iteration of the optimization, we test if each trajectory, extended or not, is statistically reliable, removing unreliable ones as outliers.

Sec. 2 describes our affine space constraint. Sec. 3 describes our initial outlier removal procedure. In Sec. 4, we describe how we extend partial trajectories and test their reliability. In Sec. 5, we show real video examples and demonstrate that our method can restore sufficiently many trajectories for accurate and detailed 3-D reconstruction. Sec. 6 is our conclusion.

2. Affine Space Constraint

We first describe the geometric constraints on which our method is based.

2.1 Trajectory of feature points

Suppose we track N feature points over M frames. Let $(x_{\kappa\alpha}, y_{\kappa\alpha})$ be the coordinates of the αth point in the κth frame. We stack all the coordinates vertically and represent the entire trajectory by the following $2M$-dimensional *trajectory vector*:

$$\boldsymbol{p}_\alpha = \begin{pmatrix} x_{1\alpha} & y_{1\alpha} & x_{2\alpha} & y_{2\alpha} & \cdots & x_{M\alpha} & y_{M\alpha} \end{pmatrix}^\top. \quad (1)$$

Taking the XYZ camera coordinate system as a reference, we express the 3-D scene coordinate system with respect to the camera coordinate system. Let $\boldsymbol{t}_\kappa$ and $\{\boldsymbol{i}_\kappa, \boldsymbol{j}_\kappa, \boldsymbol{k}_\kappa\}$ be, respectively, the origin and the orthonormal basis of the scene coordinate system expressed with respect to the camera coordinate system for the κth frame. If the αth point has scene coordinates $(a_\alpha, b_\alpha, c_\alpha)$, its position with respect to the camera coordinate system for the κth frame is

$$\boldsymbol{r}_{\kappa\alpha} = \boldsymbol{t}_\kappa + a_\alpha \boldsymbol{i}_\kappa + b_\alpha \boldsymbol{j}_\kappa + c_\alpha \boldsymbol{k}_\kappa. \quad (2)$$

2.2 Affine camera model

If an affine camera model (orthographic, weak perspective, or paraperspective projection [11]) is assumed, the image position of $\boldsymbol{r}_{\kappa\alpha}$ is

$$\begin{pmatrix} x_{\kappa\alpha} \\ y_{\kappa\alpha} \end{pmatrix} = \boldsymbol{A}_\kappa \boldsymbol{r}_{\kappa\alpha} + \boldsymbol{b}_\kappa, \quad (3)$$

where $\boldsymbol{A}_\kappa$ and $\boldsymbol{b}_\kappa$ are, respectively, a 2×3 matrix and a 2-dimensional vector determined by the position and orientation of the camera and its internal parameters for the κth frame. From Eq. (2), we can write Eq. (3) as

$$\begin{pmatrix} x_{\kappa\alpha} \\ y_{\kappa\alpha} \end{pmatrix} = \tilde{\boldsymbol{m}}_{0\kappa} + a_\alpha \tilde{\boldsymbol{m}}_{1\kappa} + b_\alpha \tilde{\boldsymbol{m}}_{2\kappa} + c_\alpha \tilde{\boldsymbol{m}}_{3\kappa}, \quad (4)$$

where $\tilde{\boldsymbol{m}}_{0\kappa}$, $\tilde{\boldsymbol{m}}_{1\kappa}$, $\tilde{\boldsymbol{m}}_{2\kappa}$, and $\tilde{\boldsymbol{m}}_{3\kappa}$ are 2-dimensional vectors determined by the position and orientation of the camera and its internal parameters in the κth frame. From Eq. (4), the trajectory vector $\boldsymbol{p}_\alpha$ in Eq. (1) can be written in the form

$$\boldsymbol{p}_\alpha = \boldsymbol{m}_0 + a_\alpha \boldsymbol{m}_1 + b_\alpha \boldsymbol{m}_2 + c_\alpha \boldsymbol{m}_3, \quad (5)$$

where $\boldsymbol{m}_0$, $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, and $\boldsymbol{m}_3$ are the $2M$-dimensional vectors obtained by stacking $\tilde{\boldsymbol{m}}_{0\kappa}$, $\tilde{\boldsymbol{m}}_{1\kappa}$, $\tilde{\boldsymbol{m}}_{2\kappa}$, and $\tilde{\boldsymbol{m}}_{3\kappa}$ vertically over the M frames, respectively.

2.3 Affine space constraint

Eq. (5) implies that all the trajectories are constrained to be in the 4-dimensional subspace spanned by $\{\boldsymbol{m}_0, \boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3\}$ in $\mathcal{R}^{2M}$. This is called the *subspace constraint* [8, 9], on which the method of Jacobs [5] is based.

In addition, the coefficient of $\boldsymbol{m}_0$ in Eq. (5) is identically 1 for all α. This means that the trajectories are in the 3-dimensional affine space within that 4-dimensional subspace. This is called the *affine space constraint* [10].

If all the feature points are tracked to the final frame, we can define the coordinate origin at the centroid of their trajectory vectors $\{\boldsymbol{p}_\alpha\}$, thereby regarding them as defining a 3-dimensional subspace in $\mathcal{R}^{2M}$. The Tomasi-Kanade factorization [15] is based on this representation, and Brandt [1] tried to find this representation by iterations. In this paper, we directly use the affine space constraint without searching for the centroid.

Unlike existing studies, our trajectory extension scheme does not assume any particular camera model (e.g., orthographic, weak perspective, or paraperspective projection) except that it is affine. A particular camera model is necessary only when we compute the 3-D shape and motion from the extended trajectories.

3. Outlier Removal

Before extending partial trajectories, we must first remove incorrectly tracked trajectories, or "outliers", from among observed complete trajectories.

This problem was studied by Huynh and Heyden [4], who fitted a 4-dimensional subspace to the observed trajectories by LMedS [12], removing those trajectories sufficiently apart from it. However, their distance measure was introduced merely for mathematical convenience without giving much consideration to the statistical behavior of image noise.

Sugaya and Kanatani [14] fitted a 4-dimensional subspace to the observed trajectories by RANSAC [2, 3] and removed outliers using a χ^2 criterion derived by modeling the error behavior of actual video tracking. In this paper, we apply their method to the affine space constraint.

3.1 Procedure

Let $n = 2M$, where M is the number of frames, and let $\{\boldsymbol{p}_\alpha\}$, $\alpha = 1, ..., N$, be the observed complete trajectory vectors. Our outlier removal procedure is as follows:

1. Randomly choose three vectors $\boldsymbol{q}_1$, $\boldsymbol{q}_2$, $\boldsymbol{q}_3$ from among $\{\boldsymbol{p}_\alpha\}$.
2. Compute the $n \times n$ moment matrix
$$\boldsymbol{M}_3 = \sum_{i=1}^{3} (\boldsymbol{q}_i - \boldsymbol{q}_C)(\boldsymbol{q}_i - \boldsymbol{q}_C)^\top, \tag{6}$$
where $\boldsymbol{q}_C$ is the centroid of $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3\}$.
3. Let $\lambda_1 \geq \lambda_2 \geq \lambda_3$ be the three eigenvalues of the matrix $\boldsymbol{M}_3$, and $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ the orthonormal system of corresponding eigenvectors.
4. Compute the $n \times n$ projection matrix
$$\boldsymbol{P}_{n-3} = \boldsymbol{I} - \sum_{i=1}^{3} \boldsymbol{u}_i \boldsymbol{u}_i^\top. \tag{7}$$
5. Let S be the number of points $\boldsymbol{p}_\alpha$ that satisfy
$$\|\boldsymbol{P}_{n-3}(\boldsymbol{p}_\alpha - \boldsymbol{q}_C)\|^2 < (n-3)\sigma^2, \tag{8}$$
where σ is an estimate of the noise standard deviation.
6. Repeat the above procedure a sufficient number of times[3], and determine the projection matrix $\boldsymbol{P}_{n-3}$ that maximizes S.
7. Remove those $\boldsymbol{p}_\alpha$ that satisfy
$$\|\boldsymbol{P}_{n-3}(\boldsymbol{p}_\alpha - \boldsymbol{q}_C)\|^2 \geq \sigma^2 \chi^2_{n-3;99}, \tag{9}$$
where $\chi^2_{r;a}$ is the ath percentile of the χ^2 distribution with r degrees of freedom.

3.2 Interpretation

In Eq. (8), the term $\|\boldsymbol{P}_{n-3}(\boldsymbol{p}_\alpha - \boldsymbol{q}_C)\|^2$, which we call the *residual*, is the squared distance of point $\boldsymbol{p}_\alpha$ from the fitted 3-dimensional affine space. If the noise in the coordinates of the feature points is an independent Gaussian random variable of mean 0 and standard deviation σ, the residual $\|\boldsymbol{P}_{n-3}(\boldsymbol{p}_\alpha - \boldsymbol{q}_C)\|^2$ divided by σ^2 should be subject to a χ^2 distribution with $n-3$ degrees of freedom. Hence, its expectation is $(n-3)\sigma^2$. The above procedure effectively fits a 3-dimensional affine space that maximizes the number of the trajectories whose residuals are smaller than $(n-3)\sigma^2$. After fitting such an affine space, we remove those trajectories which cannot be regarded as inliers with significance level 1% (Fig. 1). We have confirmed that the value $\sigma = 0.5$ can work well for all image sequences we tested [14].

3.3 Final affine space fitting

After removing outlier trajectories, we optimally fit a 3-dimensional affine space to the resulting inlier trajectories. Let $\{\boldsymbol{p}_\alpha\}$, $\alpha = 1, ..., N$, be their trajectory vectors. We first compute their centroid
$$\boldsymbol{p}_C = \frac{1}{N} \sum_{\alpha=1}^{N} \boldsymbol{p}_\alpha. \tag{10}$$

[3] In our experiment, we stopped if S did not increase 200 times consecutively.

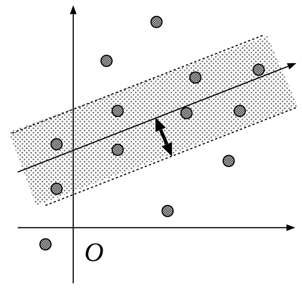

Figure 1 Removing outliers by fitting a 3-dimensional affine space.

Then, we compute the $n \times n$ moment matrix
$$\boldsymbol{M} = \sum_{\alpha=1}^{N} (\boldsymbol{p}_\alpha - \boldsymbol{p}_C)(\boldsymbol{p}_\alpha - \boldsymbol{p}_C)^\top. \tag{11}$$
Let $\lambda_1 \geq \lambda_2 \geq \lambda_3$ be the three eigenvalues of the matrix $\boldsymbol{M}$, and $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ the orthonormal system of corresponding eigenvectors. The optimally fitted 3-dimensional affine space is simply the affine space spanned by the three vectors of $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, and $\boldsymbol{u}_3$ starting from $\boldsymbol{p}_C$.

Mathematically, this affine space fitting is equivalent to the factorization operation using the singular value decomposition (SVD) in the method of Tomasi and Kanade [15].

4. Trajectory Extension

We now describe our trajectory extension scheme. It consists of several components, which we describe one by one.

4.1 Interrupted trajectories

If the αth feature point can be tracked only over κ of the M frames, its trajectory vector $\boldsymbol{p}_\alpha$ has $n-k$ unknown components (as before, we put $n = 2M$ and $k = 2\kappa$). We divide the vector $\boldsymbol{p}_\alpha$ into the k-dimensional vector $\boldsymbol{p}_\alpha^{(0)}$ consisting of the k known components and the $(n-k)$-dimensional vector $\boldsymbol{p}_\alpha^{(1)}$ consisting of the remaining $n-k$ unknown components. We also divide[4] the centroid $\boldsymbol{p}_C$ and the basis vectors $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ into the k-dimensional vectors $\boldsymbol{p}_C^{(0)}$ and $\{\boldsymbol{u}_1^{(0)}, \boldsymbol{u}_2^{(0)}, \boldsymbol{u}_3^{(0)}\}$ and the $(n-k)$-dimensional vectors $\boldsymbol{p}_C^{(1)}$ and $\{\boldsymbol{u}_1^{(1)}, \boldsymbol{u}_2^{(1)}, \boldsymbol{u}_3^{(1)}\}$ in accordance with the division of $\boldsymbol{p}_\alpha$.

4.2 Reliability test

We test if each of the partial trajectories is sufficiently reliable. Let $\boldsymbol{p}_\alpha$ be a partial trajectory vector. If image noise does not exist, the displacement of $\boldsymbol{p}_\alpha$

[4] This is merely for the convenience of description. In real computation, we treat all data as n-dimensional vectors after multiplying them by an appropriate diagonal matrix consisting of 1s for the known component positions and 0s for the rest.

from the centroid $\boldsymbol{p}_C$ is expressed as a linear combination of $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, and $\boldsymbol{u}_3$. Hence, there exist some constants c_1, c_2, and c_3 such that

$$\boldsymbol{p}_\alpha^{(0)} - \boldsymbol{p}_C^{(0)} = c_1\boldsymbol{u}_1^{(0)} + c_2\boldsymbol{u}_2^{(0)} + c_3\boldsymbol{u}^{(0)} \tag{12}$$

for the k known components. In the presence of image noise, this equality does not hold. If we let $\boldsymbol{U}^{(0)}$ be the $k \times 3$ matrix consisting of $\boldsymbol{u}_1^{(0)}$, $\boldsymbol{u}_2^{(0)}$, and $\boldsymbol{u}_3^{(0)}$ as its columns, Eq. (12) is replaced by

$$\boldsymbol{p}_\alpha^{(0)} - \boldsymbol{p}_C^{(0)} \approx \boldsymbol{U}^{(0)}\boldsymbol{c}, \tag{13}$$

where $\boldsymbol{c}$ is the 3-dimensional vector consisting of c_1, c_2, and c_3. Assuming that $k \geq 3$, we estimate the vector $\boldsymbol{c}$ by least squares in the form

$$\hat{\boldsymbol{c}} = \boldsymbol{U}^{(0)-}(\boldsymbol{p}_\alpha^{(0)} - \boldsymbol{p}_C^{(0)}), \tag{14}$$

where $\boldsymbol{U}^{(0)-}$ is the generalized inverse of $\boldsymbol{U}^{(0)}$. It is computed by

$$\boldsymbol{U}^{(0)-} = (\boldsymbol{U}^{(0)\top}\boldsymbol{U}^{(0)})^{-1}\boldsymbol{U}^{(0)\top}. \tag{15}$$

The residual, i.e., the squared distance of point $\boldsymbol{p}_\alpha^{(0)}$ from the 3-dimensional affine space spanned by $\{\boldsymbol{u}_1^{(0)}$, $\boldsymbol{u}_2^{(0)}$, $\boldsymbol{u}_3^{(0)}\}$ is $\|\boldsymbol{p}_\alpha^{(0)} - \boldsymbol{p}_C^{(0)} - \boldsymbol{U}^{(0)}\hat{\boldsymbol{c}}\|^2$. If the noise in the coordinates of the feature points is an independent Gaussian random variable of mean 0 and standard deviation σ, the residual $\|\boldsymbol{p}_\alpha^{(0)} - \boldsymbol{p}_C^{(0)} - \boldsymbol{U}^{(0)}\hat{\boldsymbol{c}}\|^2$ divided by σ^2 should be subject to a χ^2 distribution with $k-3$ degrees of freedom. Hence, we regard those trajectories that satisfy

$$\|\boldsymbol{p}_\alpha^{(0)} - \boldsymbol{p}_C^{(0)} - \boldsymbol{U}^{(0)}\hat{\boldsymbol{c}}\|^2 \geq \sigma^2\chi^2_{k-3;99} \tag{16}$$

as outliers with significance level 1%.

4.3 Extension of trajectories

In accordance with Eq. (12), the optimal estimate of the unknown components in the presence of Gaussian noise as modeled above is obtained by letting

$$\boldsymbol{p}_\alpha^{(1)} - \boldsymbol{p}_C^{(1)} = c_1\boldsymbol{u}_1^{(1)} + c_2\boldsymbol{u}_2^{(1)} + c_3\boldsymbol{u}^{(1)} = \boldsymbol{U}^{(1)}\boldsymbol{c}, \tag{17}$$

where $\boldsymbol{U}^{(1)}$ is the $(n-k)\times 3$ matrix consisting of $\boldsymbol{u}_1^{(1)}$, $\boldsymbol{u}_2^{(1)}$, and $\boldsymbol{u}_3^{(1)}$ as its columns. Substituting Eq. (14) for $\boldsymbol{c}$, we obtain

$$\hat{\boldsymbol{p}}_\alpha^{(1)} = \boldsymbol{p}_C^{(1)} + \boldsymbol{U}^{(1)}\boldsymbol{U}^{(0)-}(\boldsymbol{p}_\alpha^{(0)} - \boldsymbol{p}_C^{(0)}). \tag{18}$$

4.4 Optimization

Although Eq. (18) produces the vector that optimally fits to the affine space define by $\boldsymbol{p}_C$ and $\{\boldsymbol{u}_1$, $\boldsymbol{u}_2$, $\boldsymbol{u}_3\}$, that affine space is computed only from a small number of complete trajectories; no information contained in the partial trajectories is used, irrespective of how long they are. So, we incorporate partial trajectories by iterations.

Note that if three components of $\boldsymbol{p}_\alpha$ are specified, one can place it, in general, in any 3-dimensional affine space by appropriately adjusting the remaining $n-3$ components. In view of this, we introduce the "weight" of the trajectory vector $\boldsymbol{p}_\alpha$ with k known components in the form

$$W_\alpha = \frac{k-3}{n-3}. \tag{19}$$

Let N be the number of all trajectories, complete or partial, inliers or outliers. The optimization goes as follows:

1. Set the weights W_α of those trajectories, complete or partial, that are so far judged to be outliers to 0. All other weights are set to the value in Eq. (19).
2. Fit a 3-dimensional affine space to all the trajectories. The procedure is the same as described in Sec. 3.3 except that Eq. (10) is replaced by the *weighted* centroid

$$\boldsymbol{p}_C = \frac{\sum_{\alpha=1}^N W_\alpha \boldsymbol{p}_\alpha}{\sum_{\alpha=1}^N W_\alpha}, \tag{20}$$

   and Eq. (11) is replaced by the *weighted* moment matrix

$$\boldsymbol{M} = \sum_{\alpha=1}^N W_\alpha(\boldsymbol{p}_\alpha - \boldsymbol{p}_C)(\boldsymbol{p}_\alpha - \boldsymbol{p}_C)^\top. \tag{21}$$

3. Test each trajectory if it is an outlier, using Eq. (16).
4. Estimate the unknown components of the inlier partial trajectory vectors, using Eq. (18).

These computations are repeated until the fitted affine space converges.

In the course of this optimization, trajectories once regarded as outliers may be judged to be inliers later, and vice versa. In the end, inlier partial trajectories are optimally extended with respect to the affine space that is optimally fitted to all the complete and partial inlier trajectories.

This optimization procedure starts from at least three complete trajectories that define the initial affine space. If no such initial trajectories are given, we can use, for example, the method of Jacobs [5] to compute the initial affine space for starting the optimization.

5. Experiments

We tested our method using real video sequences. Fig. 2(a) shows five decimated frames from a 50 frame sequence (320×240 pixels) of a static scene taken by a moving camera. We first detected 200 feature points and tracked them using the Kanade-Lucas-Tomasi algorithm [16]. When tracking failed at some frame, we restarted the tracking after adding a new feature point in that frame. Fig. 2(b) shows the 871 tracked trajectories thus obtained.

In the end, we obtained 29 complete trajectories, of which 11 were regarded as inliers by the procedure described in Sec. 3. The marks □ in Fig. 2(a) indicate

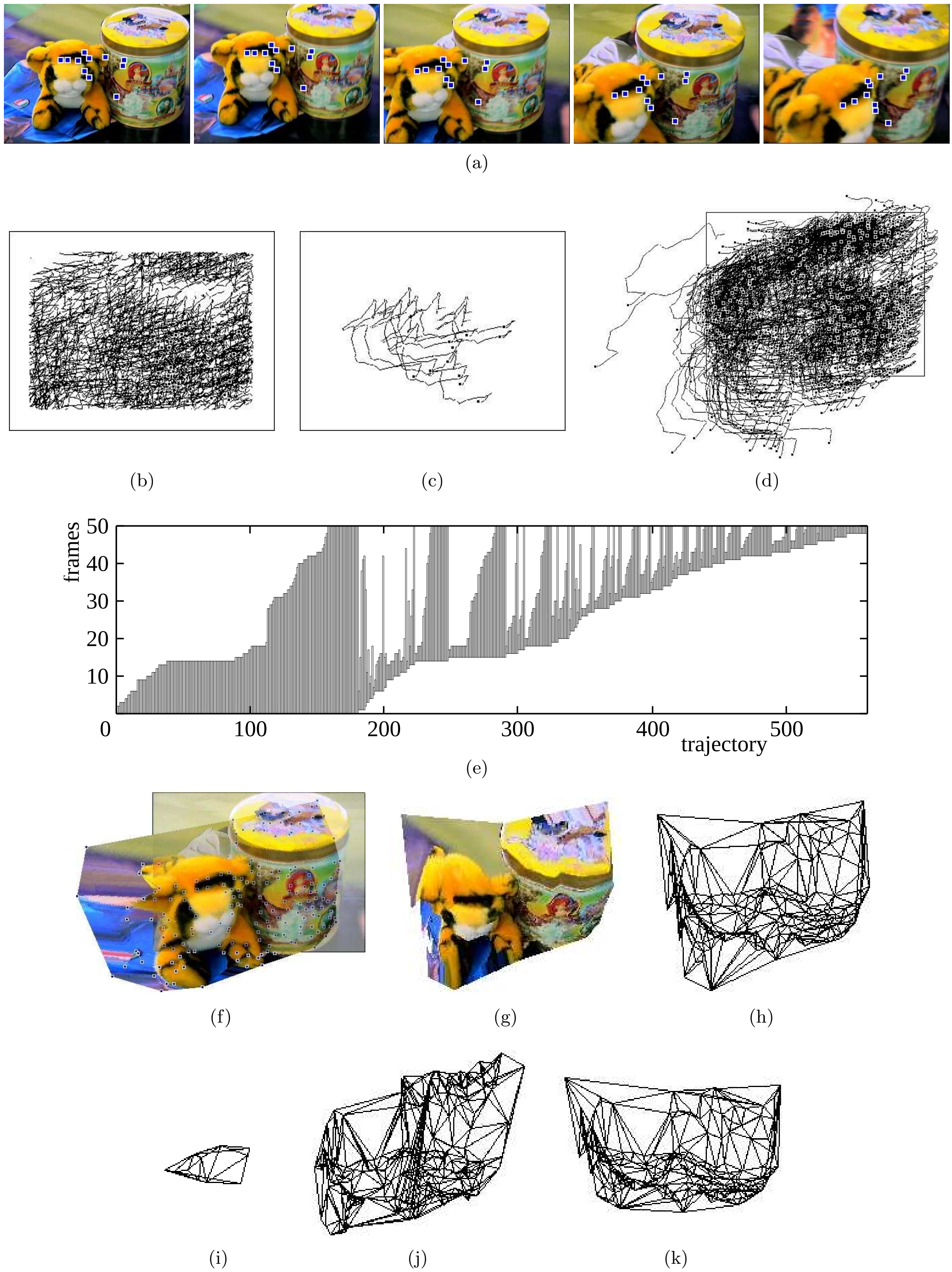

Figure 2 (a) Five decimated frames from an image sequence of 50 frames and 11 feature points correctly tracked throughout them. (b) The 871 initial trajectories obtained by the Kanade-Lucas-Tomasi algorithm. (c) The 11 complete inlier trajectories. (d) The 560 optimal extensions of the trajectories in (b). (e) The duration of the trajectories. (f) The extended texture mapped image of the 33th frame. (g) The reconstructed 3-D shape. (h) The wireframe representation of (g). (i) The wireframe reconstructed from the 11 initial complete trajectories in (c). (j) The wireframe reconstructed from all extended trajectories without optimization. (k) The wireframe reconstructed after extending and optimizing only the trajectories starting from the first frame.

their positions; Fig. 2(c) shows their trajectories. Evidently, we cannot reconstruct a meaningful 3-D shape from these trajectories alone.

Using the affine space they define, we extended the partial trajectories and optimized the affine space and the extended trajectories, testing the reliability of the extension in every iteration. The optimization converged after 11 iterations, resulting in the 560 inlier trajectories shown in Fig. 2(d). Some portions are extrapolated out of the frame. The total computation time for this optimization was 134 seconds. We used Pentium 4 2.4B GHz for the CPU with 1 Gb main memory and Linux for the OS.

Fig. 2(e) plots the durations of the 560 trajectories; they are enumerated on the horizontal axis in the order of disappearance and new appearance; the white part corresponds to missing data.

Fig. 2(f) is the texture mapped image of the 33th frame obtained after missing feature positions are restored. Using the 180 feature points visible in the first frame, we defined triangular patches, to which the texture in the first frame is mapped, extrapolating the view outside the frame.

We reconstructed the 3-D shape using the factorization method based on weak perspective projection [11]. Fig. 2(g) is the top view of the 3-D shape of the triangular patches shown in Fig. 2(f). Fig. 2(h) is its wireframe representation.

For comparison, Fig. 2(i) shows the wireframe reconstructed from the 11 initial trajectories in Fig. 2(c) alone; Fig. 2(j) shows the wireframe reconstructed from extended trajectories without optimization. Fig. 2(k) shows the corresponding shape reconstructed after extending and optimizing only the trajectories starting from the first frame without adding new trajectories. All are viewed from the same angle.

From these results, we can see that sufficiently many trajectories can be restored for accurate and detailed 3-D reconstruction by extending and optimizing complete and partial trajectories. Increasing the number of trajectories by tracking new feature points after previous tracking has failed is also effective in improving the accuracy.

6. Concluding Remarks

We have presented a new method for extending interrupted feature point tracking for 3-D affine reconstruction. Our method consists of iterations for optimally extending the trajectories so that they are compatible with the estimated affine space and optimally estimating the affine space from the extended trajectories. In every step, the reliability of the resulting trajectories is tested, and those judged to be outliers are removed. Using real video images, we have demonstrated that sufficiently many trajectories can be restored for accurate and detailed 3-D reconstruction.

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. 15500113).

References

[1] S. Brandt, "Closed-form solutions for affine reconstruction under missing data," Proc. Statistical Methods in Video Processing Workshop, pp.109–114, Copenhagen, Denmark, June, 2002.

[2] M. A. Fischer and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Comm. ACM*, vol.24, no.6, pp.381–395, June 1981.

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

[4] D. Q. Huynh and A. Heyden, "Outlier detection in video sequences under affine projection," Proc. IEEE Conf. Comput. Vision Pattern Recog., vol.2, pp.695–701, Kauai, HI, U.S.A., Dec. 2001.

[5] D. W. Jacobs, "Linear fitting with missing data for structure-from-motion" Comput. Vision Image Understand., vol.82, no.1, pp.57–81, April 2001.

[6] F. Kahl and A. Heyden, "Affine structure and motion from points, lines and conics," Int. J. Comput. Vision, vol.33, no.3, pp.163–180, Sept. 1999.

[7] S. Kamijima and H. Saito, "Factorization method applying to measurement matrix interpolated under projective geometry," IPSJ SIG Notes, 2002-CVIM-135-15, pp.95–102, Nov. 2002 (in Japanese).

[8] K. Kanatani, "Motion segmentation by subspace separation and model selection," Proc. 8th Int. Conf. Comput. Vision, Vancouver, Canada, vol.2, pp.301-306, July 2001.

[9] K. Kanatani, "Motion segmentation by subspace separation: Model selection and reliability evaluation," Int. J. Image Graphics, vol.2, no.2, pp.179–197, April 2002.

[10] K. Kanatani, "Evaluation and selection of models for motion segmentation," Proc. 7th Euro. Conf. Comput. Vision, Copenhagen, Denmark, pp.335–349, June 2002.

[11] C. J. Poleman and T. Kanade, "A paraperspective factorization method for shape and motion recovery," IEEE Trans. Patt. Anal. Mach. Intell., vol.19, no.3, pp.206–218, March 1997.

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

[13] H.-Y. Shum, K. Ikeuchi and R. Reddy, "Principal component analysis with missing data and its application to polyhedral object modeling," IEEE Trans. Patt. Anal. Mach. Intell., vol.17, no.3, pp.854–867, Sept. 1995.

[14] Y. Sugaya and K. Kanatani, "Outlier removal for motion tracking by subspace separation," IEICE Trans. Inf. & Syst., vol.E86-D, no.6, June 2003.

[15] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography—A factorization method," Int. J. Comput. Vision, vol.9, no.2, pp.137–154, Nov. 1992.

[16] C. Tomasi and T. Kanade, Detection and Tracking of Point Features, CMU Tech. Rep. CMU-CS-91-132, April 1991;
`http://vision.stanford.edu/~birch/klt/`.