The 8th International Conference on Computer Vision, July 2001, Vancouver, Canada, Vol.2, pp. 586–591.

Motion Segmentation by Subspace Separation and Model Selection

Kenichi Kanatani
Department of Information Technology, Okayama University, Okayama 700-8530 Japan
kanatani@suri.it.okayama-u.ac.jp

Abstract

Reformulating the Costeira-Kanade algorithm as a pure mathematical theorem independent of the Tomasi-Kanade factorization, we present a robust segmentation algorithm by incorporating such techniques as dimension correction, model selection using the geometric AIC, and least-median fitting. Doing numerical simulations, we demonstrate that our algorithm dramatically outperforms existing methods. It does not involve any parameters which need to be adjusted empirically.

1. Introduction

Segmenting individual objects from backgrounds is one of the most important of computer vision tasks. An important clue is provided by motion; humans can easily discern independently moving objects by simply seeing their motions without knowing their identities. Costeira and Kanade [1] presented an algorithm for segmentation from image point motions captured by feature tracking. They associated their method with the Tomasi-Kanade factorization [11], but a close examination reveals that the underlying principle is a simple fact of linear algebra, as pointed out by Gear [2], who also presented an alternative method.

In this paper, we first state the principle as *subspace separation* with the intention of applying it to a wider range of problems not limited to motion segmentation or even to computer vision. In fact, Maki and Wiles [6] have pointed out that the same principle applies to separating illumination sources by observing multiple images.

The biggest drawback of the Costeira-Kanade algorithm [1], and the essentially equivalent method of Gear [2] as well, is that the performance severely deteriorates in the presence of noise. This is because segmentation is based on the decision if particular elements of a matrix computed from the data are zero. In the presence of noise, a small error in one datum can affect all the elements of the matrix in a complicated manner, and finding a suitable threshold is difficult even if the noise is known to be Gaussian with a known variance.

To avoid this difficulty, one needs to analyze the *original data* rather than a matrix derived from them. In this paper, we present a robust segmentation algorithm by working in the original data space, where we incorporate the geometric AIC [4, 5] and least-median fitting [7, 10]. Doing numerical simulation, we demonstrate that our method dramatically outperforms existing methods. We also derive a bound on the accuracy, with which our method is compared. Our algorithm has a notable feature that *no parameters need to be adjusted empirically*.

2. Motion Subspaces

Suppose we track N rigidly moving feature points over M images. Let $(x_{\kappa\alpha}, y_{\kappa\alpha})$ be the image coordinates of the αth point in the κth frame. If we stack the image coordinates over the M frames vertically into a $2M$-dimensional vector in the form

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

the image motion of the αth point is represented by a single point $\boldsymbol{p}_\alpha$ in a $2M$-dimensional space.

We regard the XYZ camera coordinate system as the world coordinate system with the Z-axis along the optical axis. We fix an arbitrary object coordinate system to the object and let $\boldsymbol{t}_\kappa$ and $\{\boldsymbol{i}_\kappa, \boldsymbol{j}_\kappa, \boldsymbol{k}_\kappa\}$ be, respectively, its origin and orthonormal basis in the κth frame. Let $(a_\alpha, b_\alpha, c_\alpha)$ be the coordinates of the αth point with respect to the object coordinate system. Its position in the κth frame with respect to the world coordinate system is given by

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

If we assume orthographic projection, we have

$$\begin{pmatrix} x_{\kappa\alpha} \\ y_{\kappa\alpha} \end{pmatrix} = \tilde{\boldsymbol{t}}_\kappa + a_\alpha \tilde{\boldsymbol{i}}_\kappa + b_\alpha \tilde{\boldsymbol{j}}_\kappa + c_\alpha \tilde{\boldsymbol{k}}_\kappa, \qquad (3)$$

where $\tilde{\boldsymbol{t}}_\kappa$, $\tilde{\boldsymbol{i}}_\kappa$, $\tilde{\boldsymbol{j}}_\kappa$, and $\tilde{\boldsymbol{k}}_\kappa$ are the 2-dimensional vectors obtained from $\boldsymbol{t}_\kappa$, $\boldsymbol{i}_\kappa$, $\boldsymbol{j}_\kappa$, and $\boldsymbol{k}_\kappa$, respectively, by chopping the third components.

If we stack the vectors $\tilde{\boldsymbol{t}}_\kappa$, $\tilde{\boldsymbol{i}}_\kappa$, $\tilde{\boldsymbol{j}}_\kappa$, and $\tilde{\boldsymbol{k}}_\kappa$ over the M frames vertically into $2M$-dimensional vectors $\boldsymbol{m}_0$, $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, and $\boldsymbol{m}_3$, respectively, in the same way as eq. (1), the vector $\boldsymbol{p}_\alpha$ has the form

$$\boldsymbol{p}_\alpha = \boldsymbol{m}_0 + a_\alpha \boldsymbol{m}_1 + b_\alpha \boldsymbol{m}_2 + c_\alpha \boldsymbol{m}_3. \qquad (4)$$

Thus, the N points $\{\boldsymbol{p}_\alpha\}$ belong to the 4-dimensional subspace spanned by the vectors $\{\boldsymbol{m}_0, \boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3\}$. This fact holds for all affine camera models including weak perspective and paraperspective [9].

If the motion is planar, i.e., if the object translates only in the X and Y directions and rotates only around the Z-axis, the vector $\tilde{\boldsymbol{k}}_\kappa$ vanishes if we take $\boldsymbol{i}_\kappa$, $\boldsymbol{j}_\kappa$, and $\boldsymbol{k}_\kappa$ to be in the X, Y, and Z directions, respectively. This means that the N points $\{\boldsymbol{p}_\alpha\}$ belong to the 3-dimensional subspace spanned by $\{\boldsymbol{m}_0, \boldsymbol{m}_1, \boldsymbol{m}_2\}$.

It follows that the motions of the feature points are segmented into independently moving objects by grouping the N points in $\mathcal{R}^n$ $(n = 2M)$ into distinct 4-dimensional subspaces for general motions and distinct 3-dimensional subspaces for planar motions.

3. Subspace Separation Theorem

Let $\{\boldsymbol{p}_\alpha\}$ be N points that belong to an r-dimensional subspace $\mathcal{L} \subset \mathcal{R}^n$. Define an $N \times N$ $\boldsymbol{G} = (G_{\alpha\beta})$ by

$$G_{\alpha\beta} = (\boldsymbol{p}_\alpha, \boldsymbol{p}_\beta), \tag{5}$$

where $(\boldsymbol{a}, \boldsymbol{b})$ denotes the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. This matrix gives the information about the lengths of the vectors $\{\boldsymbol{p}_\alpha\}$ and their mutual angles, so we call it the *metric matrix*.

Let $\lambda_1 \geq \cdots \geq \lambda_N$ be its eigenvalues, and $\{\boldsymbol{v}_1, ..., \boldsymbol{v}_N\}$ the orthonormal system of the corresponding eigenvectors. Define the $N \times N$ *interaction matrix* $\boldsymbol{Q} = (Q_{\alpha\beta})$ by

$$\boldsymbol{Q} = \sum_{i=1}^{r} \boldsymbol{v}_i \boldsymbol{v}_i^\top. \tag{6}$$

Divide the index set $\mathcal{I} = \{1, ..., N\}$ into m disjoint subsets $\mathcal{I}_i$, $i = 1, ..., m$, and let r_i be the dimension of the subspace $\mathcal{L}_i$ defined by the ith set $\{\boldsymbol{p}_\alpha\}$, $\alpha \in \mathcal{I}_i$. If the m subspaces $\mathcal{L}_i$, $i = 1, ..., m$, are linearly independent, we have

Theorem 1 *The $(\alpha\beta)$ element of $\boldsymbol{Q}$ is zero if the αth and βth points belong to different subspaces:*

$$Q_{\alpha\beta} = 0, \quad \alpha \in \mathcal{I}_i, \ \ \beta \in \mathcal{I}_j, \ \ i \neq j. \tag{7}$$

This theorem is the essence of the principle on which the Costeira-Kanade algorithm [1] relies. Costeira and Kanade described this result in reference to the Tomasi-Kanade factorization [11], but it can be proved purely mathematically as follows. For N $(> n)$ vectors $\{\boldsymbol{p}_\alpha\}$, there exist infinitely many sets of numbers $\{c_1, ..., c_N\}$, not all zero, such that $\sum_{\alpha=1}^{N} c_\alpha \boldsymbol{p}_\alpha = \boldsymbol{0}$, but if the points $\{\boldsymbol{p}_\alpha\}$ belong to two subspaces $\mathcal{L}_1$ and $\mathcal{L}_2$ such that $\mathcal{L}_1 \oplus \mathcal{L}_2 = \mathcal{R}^n$, the set of such "annihilating coefficients" $\{c_\alpha\}$ ("null space" to be precise) is generated by those for which $\sum_{\mathbf{p}_\alpha \in \mathcal{L}_1} c_\alpha \boldsymbol{p}_\alpha = \boldsymbol{0}$ and those for which $\sum_{\mathbf{p}_\alpha \in \mathcal{L}_2} c_\alpha \boldsymbol{p}_\alpha = \boldsymbol{0}$. A formal proof is given in the Appendix.

4. Separation Procedure

4.1 Greedy algorithm

In the presence of noise, all the elements of $\boldsymbol{Q} = (Q_{\alpha\beta})$ are nonzero in general. A straightforward method is to successively group points $\boldsymbol{p}_\alpha$ and $\boldsymbol{p}_\beta$ for which $|Q_{\alpha\beta}|$ is large. If we progressively interchange the corresponding rows and columns of $\boldsymbol{Q}$, it ends up with an approximate block-diagonal matrix [1]. Formally, we define the *similarity measure* between the ith subspace $\mathcal{L}_i$ and the jth subspace $\mathcal{L}_j$ by $s_{ij} = \max_{\mathbf{p}_\alpha \in \mathcal{L}_i, \mathbf{p}_\beta \in \mathcal{L}_j} |Q_{\alpha\beta}|$ and repeatedly merge two subspaces for which s_{ij} is large.

Costeira and Kanade [1] adopted this type of strategy, known as the *greedy* algorithm. They used $\sum_{\mathbf{p}_\alpha \in \mathcal{L}_i, \mathbf{p}_\beta \in \mathcal{L}_j} |Q_{\alpha\beta}|^2$, but according to our experience the choice of the measure does not affect the result very much. Since noise exists in the *data* $\{\boldsymbol{p}_\alpha\}$, not in the elements of $\boldsymbol{Q}$, and no information is available about the magnitude of the nonzero elements of $\boldsymbol{Q}$, it is difficult to obtain an appropriate criterion.

Gear [2] formulated the same problem as graph matching, which he solved by a greedy algorithm, but it is difficult to weigh the graph edges appropriately. Gear [2] did a complicated statistical analysis for this, but the result does not seem very successful. Ichimura [3] applied the discrimination criterion of Otsu [8] for thresholding.

4.2 Dimension correction

Theorem 1 is based on the existence of "locally closed annihilating coefficients". In the presence of noise, no such coefficients exist, so we create them. Let d be the dimension of the subspaces to be separated ($d = 4$ for general motions and $d = 3$ for planar motions). As soon as more than d points are grouped together, we optimally fit a d-dimensional subspace to them, replace the points with their projections onto the fitted subspace, and recompute the interaction matrix $\boldsymbol{Q}$. This effectively reduces the noise in the data if the local grouping is correct. Continuing this process, we end up with an *exact* block-diagonal matrix $\boldsymbol{Q}$.

4.3 Model selection

The fundamental criterion in the data space is the *residual* J, i.e., the sum of the square distances of the data points to the fitted subspace. It is reasonable not to merge two groups of points if the resulting residual would be large compared with the sum of the residuals of separately fitting two subspaces to them. But how large should the residual be for this judgment? In general, the residual always increases after two groups of points are merged, because a single subspace has fewer

degrees of freedom to adjust than two subspaces. It follows that we must balance the increase of the residual against the decrease of the degree of freedom. For this purpose, we use the *geometric AIC* [4, 5]. A similar idea was used for motion segmentation by Torr [12] though his approach is different from ours.

Let $\mathcal{L}_i$ and $\mathcal{L}_j$ be candidate subspaces of dimension d to merge, and let N_i and N_j be the respective numbers of points in them. The corresponding residuals $\hat{J}_i$ and $\hat{J}_j$ are computed in the course of the dimension correction. We assume that each point is perturbed from its true position by independent Gaussian noise of mean zero and standard deviation ϵ, which is referred to as the *noise level*.

Let $\hat{J}_{i\oplus j}$ be the residual that would result after fitting a single d-dimensional subspace to the $N_i + N_j$ points. Since a d-dimensional subspace has $d(n-d)$ degrees of freedom[1], the geometric AIC has the following form [4, 5]:

$$\text{G-AIC}_{i\oplus j} = \hat{J}_{i\oplus j} + 2d\Big(N_i + N_j + n - d\Big)\epsilon^2. \quad (8)$$

If two d-dimensional subspaces are fitted to the N_i points and the N_j points separately, the degree of freedom is the sum of those for individual subspaces. Hence, the geometric AIC is as follows [4, 5]:

$$\text{G-AIC}_{i,j} = \hat{J}_i + \hat{J}_j + 2d\Big(N_i + N_j + 2(n-d)\Big)\epsilon^2. \quad (9)$$

Merging $\mathcal{L}_i$ and $\mathcal{L}_j$ is reasonable if $\text{G-AIC}_{i\oplus j} < \text{G-AIC}_{i,j}$. However, this criterion can work only for $N_i + N_j > d$. Also, the information provided by the interaction matrix $\boldsymbol{Q}$ will be ignored. Here, we mix these two criteria together and define the following similarity measure between the subspaces $\mathcal{L}_i$ and $\mathcal{L}_j$:

$$s_{ij} = \frac{\text{G-AIC}_{i,j}}{\text{G-AIC}_{i\oplus j}} \max_{\boldsymbol{p}_\alpha \in \mathcal{L}_i, \boldsymbol{p}_\beta \in \mathcal{L}_j} |Q_{\alpha\beta}|. \quad (10)$$

Two subspaces with the largest similarity are merged successively until the number of subspaces becomes a specified number m. However, some of the resulting subspaces may contain less than d elements, which violates our assumption. To prevent this, we take subspaces with less than d elements as first candidates to be merged as long as they exist.

For evaluating the geometric AIC, we need to estimate the noise level ϵ. This can be done if we note that the vectors $\{\boldsymbol{p}_\alpha\}$ should be constrained to be in an r-dimensional subspace of $\mathcal{R}^n$ in the absence of noise ($r = md$). Let $\hat{J}_r$ be the residual after fitting an r-dimensional subspace to $\{\boldsymbol{p}_\alpha\}$. Then, $\hat{J}_r/\epsilon^2$ is subject to a χ^2 distribution with $(n-r)(N-r)$ degrees of freedom [4]. Hence, we obtain the following unbiased estimator of ϵ^2:

$$\hat{\epsilon}^2 = \frac{\hat{J}_r}{(n-r)(N-r)}. \quad (11)$$

[1] It is specified by d points in $\mathcal{R}^n$, but they can move within that subspace into d directions. So, the degree of freedom is $dn - d^2$.

4.4 Robust fitting

Once a point is misclassified in the course of the merging process, it never leaves that class. We now attempt to remove outliers from the m resulting classes $\mathcal{L}_1$, ..., $\mathcal{L}_m$.

Points near the origin may be easily misclassified, so we select from each class $\mathcal{L}_i$ half (but not less than d) of the elements that have large norms. We fit d-dimensional subspaces $\mathcal{L}'_1$, ..., $\mathcal{L}'_m$ to them again and select from each class $\mathcal{L}_i$ half (but not less than d) of the elements whose distances to the closest subspace $\mathcal{L}'_j$, $j \neq i$, are large. We fit d-dimensional subspaces $\mathcal{L}''_1$, ..., $\mathcal{L}''_m$ to them again and allocate each data point to the closest one. Finally, we fit d-dimensional subspaces $\mathcal{L}'''_1$, ..., $\mathcal{L}'''_m$ to the resulting point sets by the *least-median* (to be precise, *least median-of-squares*) *method* [7, 10]. Each data point is reallocated to the closest one.

4.5 Accuracy bound

Whatever method we use, we cannot reach 100% accuracy as long as noise exists in the data. For objective evaluation of an algorithm, we should compare its performance with an ideal method. Suppose we know by an "oracle" the true subspaces $\bar{\mathcal{L}}_1$, ..., $\bar{\mathcal{L}}_m$, from which the observed data were perturbed by independent and identically distributed Gaussian noise. Evidently, each point should be grouped into the closest subspace from it. Of course we cannot do this using real data, but we can do simulations, for which the true solution is known, and regard the performance of this oracle method as a bound on the accuracy.

5. Examples

Fig. 1 shows five consecutive images of 20 points in the background and 9 points in an object. The background and the object are independently moving in 2 dimensions; the object is given a wireframe for the ease of visualization. We added Gaussian noise of mean 0 and standard deviation ϵ to the coordinates of the 29 points independently and classified them into two groups.

Fig. 2(a) plots the average error ratio over 500 independent trials for different ϵ: we compared (1) the method using the greedy algorithm only, (2) the method with dimension correction added, (3) the method with model selection in addition, and (4) the method with robust fitting further added. We can see that each added technique reduces the error further.

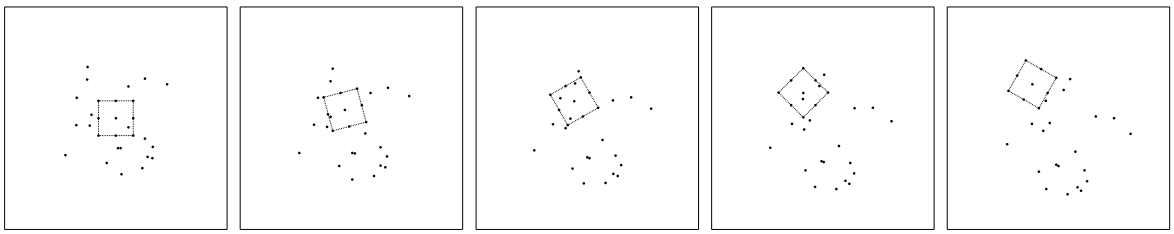
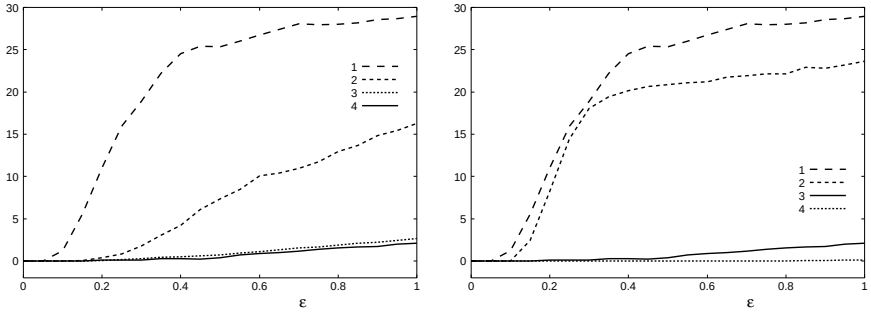


Figure 1: An image sequence of points in planar motion.



(a) (b)

Figure 2: Error ratio for segmenting the planar motion of Fig. 1. (a) 1. Greedy algorithm. 2. With dimension correction. 3. With model selection. 4. With robust fitting. (b) 1. Greedy algorithm. 2. Ichimura's method. 3. Our method. 4. Lower bound.

In Fig. 2(b), the greedy algorithm, our method with all the techniques combined, and Ichimura's method [3] that uses the discrimination criterion of Otsu [8] are compared with the bound given by the oracle method. We can observe that Ichimura's method is slightly better than the greedy algorithm but inferior to our method. This is because the Otsu criterion classifies elements in the least-squares sense and hence nonzero elements $|Q_{\alpha\beta}|$ that are close to zero are judged to be zero in the presence of noise.

Fig. 3 shows five consecutive images of 20 points in the background and 14 points in an object. The background and the object are independently moving in 3 dimensions. Fig. 4 shows the classification results corresponding to Fig. 2. Again, we can see that our method dramatically improves the classification accuracy.

Fig. 5 shows a sequence of perspectively projected images (above) and manually selected feature points from them (below). For this data set, we could correctly separate an independent 3-D motion from the background motion by the greedy algorithm and our method, whereas Ichimura's method failed. We added independent Gaussian noise of mean 0 and standard deviation $\epsilon = 0, 1, 2, 3, \ldots$ (pixels) to the coordinates of the feature points and applied our method 10 times for each ϵ, using different noise each time. The greedy algorithm and Ichimura's method caused misclassifications, but our method was always correct up to $\epsilon = 5$ (pixels)

This image sequence captures a 3-D motion, but if we regard it as a planar motion, the greedy algorithm and our method can detect the correct motion, but Ichimura's method fails. However, the greedy algorithm fails if random noise of $\epsilon = 1$ is added, while our method works up to $\epsilon = 3$ (pixels).

6. Concluding Remarks

We have reformulated the Costeira-Kanade method as a pure mathematical theorem independent of the Tomasi-Kanade factorization and presented a robust segmentation algorithm by incorporating such techniques as dimension correction, model selection using the geometric AIC, and least-median fitting. We did numerical simulations and compared the performance of our method with a bound on the accuracy. Real image examples were also shown. We conclude that our algorithm dramatically improves the classification accuracy over existing methods.

For practical segmentation, we should incorporate multiple features such as brightness, color, texture, and shape as well as motion. Since our algorithm is based solely on feature point motion, it alone may not be sufficient. But for the same reason it is more fundamental, and it elucidates the mathematical structure of the segmentation problem.

Our algorithm does not involve any parameters which need to be adjusted empirically. This is a notable feature in a stark contrast to many of today's "intelligent" systems for which a lot of parameter must be tuned.

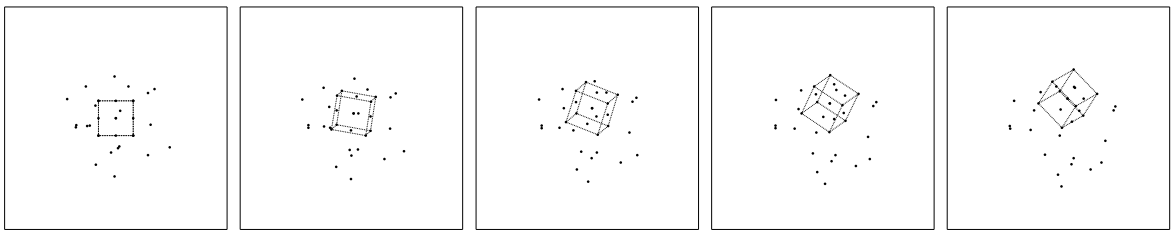
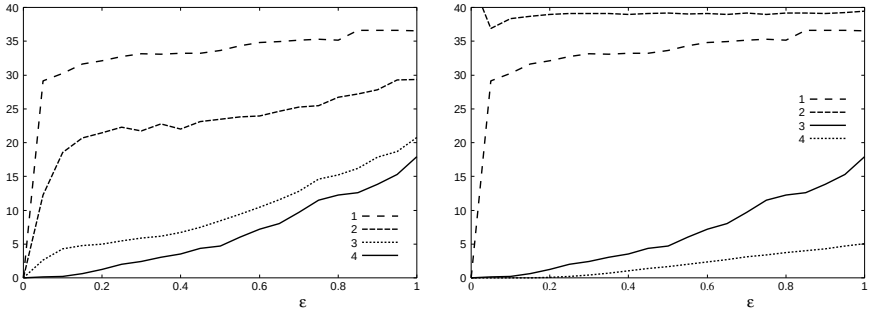


Figure 3: An image sequence of points in 3-D motion.



(a) (b)

Figure 4: Error ratio for segmenting the 3-D motion of Fig. 3. (a) 1. Greedy algorithm. 2. With dimension correction. 3. With model selection. 4. With robust fitting (b) 1. Greedy algorithm. 2. Ichimura's method. 3. Our method. 4. Lower bound.

References

[1] J. P. Costeira and T. Kanade, A multibody factorization method for independently moving objects, *Int. J. Comput. Vision*, **29**-3 (1998), 159–179.

[2] C. W. Gear, Multibody grouping from motion images, *Int. J. Comput. Vision*, **29**-2 (1998), 133–150.

[3] N. Ichimura, Motion segmentation based on factorization method and discriminant criterion, *Proc. 7th Int. Conf. Comput. Vision*, September 1999, Kerkya, Greece, pp. 600–605.

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

[5] K. Kanatani, Geometric information criterion for model selection, *Int. J. Comput. Vision*, **26**-3 (1998), 171–189.

[6] A. Maki and C. Wiles, Geotensity constraint for 3D surface reconstruction under multiple light sources, *Proc. 6th Euro. Conf. Comput. Vision*, June–July 2000, Dublin, Ireland, Vol.1, pp. 725–741.

[7] P. Meer, D. Mintz and A. Rosenfeld, Robust regression methods for computer vision: A review, *Int. J. Compute. Vision*, **6**-1 (1990), 59–70.

[8] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Trans. Sys. Man Cyber.*, **9**-1 (1979), 62–66.

[9] C. J. Poelman and T. Kanade, A paraperspective factorization method for shape and motion recovery, *IEEE Trans. Pat. Anal. Mach. Intell.*, **19**-3 (1997), 206–218.

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

[11] C. Tomasi and T. Kanade, Shape and motion from image streams under orthography—A factorization method, *Int. J. Comput. Vision*, **9**-2 (1992), 137–154.

[12] P. H. S. Torr, Geometric motion segmentation and model selection, *Phil. Trans. Roy. Soc.*, A-**356** (1998), 1321–1340.

Appendix: Proof of Theorem 1

Let N_i be the number of elements of the set $\mathcal{I}_i$. It is sufficient to prove the theorem for $m = 2$ (the proof is the same for $m > 2$). Suppose $\{\boldsymbol{p}_\alpha\}$ are *aligned*, i.e., $\boldsymbol{p}_1, ..., \boldsymbol{p}_{N_1} \in \mathcal{L}_1$ and $\boldsymbol{p}_{N_1+1}, ..., \boldsymbol{p}_N \in \mathcal{L}_2$.

Since the subspace $\mathcal{L}_1$ has dimension r_1, the $n \times N_1$ matrix $\boldsymbol{W}_1 = \left(\boldsymbol{p}_1 \cdots \boldsymbol{p}_{N_1} \right)$ has rank r_1. Hence, $\boldsymbol{W}_1$ defines a linear mapping of rank r_1 from an N_1-dimensional space $\mathcal{R}^{N_1}$ to an n-dimensional space $\mathcal{R}^n$; its null space $\mathcal{N}_1$ has dimension $\nu_1 = N_1 - r_1$. Let $\{\boldsymbol{n}_1, ..., \boldsymbol{n}_{\nu_1}\}$ be an arbitrary orthonormal basis of $\mathcal{N}_1$, each $\boldsymbol{n}_i$ being an N_1-dimensional vector. Similarly, the $n \times N_2$ matrix $\boldsymbol{W}_2 = \left(\boldsymbol{p}_{N_1} \cdots \boldsymbol{p}_N \right)$ defines a linear mapping of rank r_2 from $\mathcal{R}^{N_2}$ to $\mathcal{R}^n$; its null space $\mathcal{N}_2$ has dimension $\nu_2 = N - r_2$. Let $\{\boldsymbol{n}'_1, ..., \boldsymbol{n}'_{\nu_2}\}$ be an arbitrary orthonormal basis of $\mathcal{N}_2$, each $\boldsymbol{n}_i$ being N_2-dimensional vector.

Let $\{\tilde{\boldsymbol{n}}_i\}$, $i = 1,.., \nu_1$, and $\{\tilde{\boldsymbol{n}}'_i\}$, $i = 1,.., \nu_2$, be the N-dimensional vectors defined by padding $\{\boldsymbol{n}_i\}$ and $\{\boldsymbol{n}'_i\}$ with zero elements as follows:

$$\tilde{\boldsymbol{n}}_i = \begin{pmatrix} \boldsymbol{n}_i \\ \boldsymbol{0} \end{pmatrix}, \quad \tilde{\boldsymbol{n}}'_i = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{n}'_i \end{pmatrix}. \tag{12}$$

As a result, the $N-r$ vectors $\{\tilde{\boldsymbol{n}}_1, ..., \tilde{\boldsymbol{n}}_{\nu_1}, \tilde{\boldsymbol{n}}'_1, ..., \tilde{\boldsymbol{n}}'_{\nu_2}\}$ are an orthonormal system of $\mathcal{R}^N$ belonging to the null space $\mathcal{N}$ of the $n \times N$ *observation matrix*

$$\boldsymbol{W} = \left(\boldsymbol{p}_1 \cdots \boldsymbol{p}_N \right). \tag{13}$$

Since the matrix $\boldsymbol{W}$ has rank $r_1 + r_2$ $(= r)$ by assumption, its null space $\mathcal{N}$ has dimension $\nu = N - r$. Hence, $\{\tilde{\boldsymbol{n}}_1, ..., \tilde{\boldsymbol{n}}_{\nu_1}, \tilde{\boldsymbol{n}}'_1, ..., \tilde{\boldsymbol{n}}'_{\nu_2}\}$ are an orthonormal basis of the null space $\mathcal{N}$.

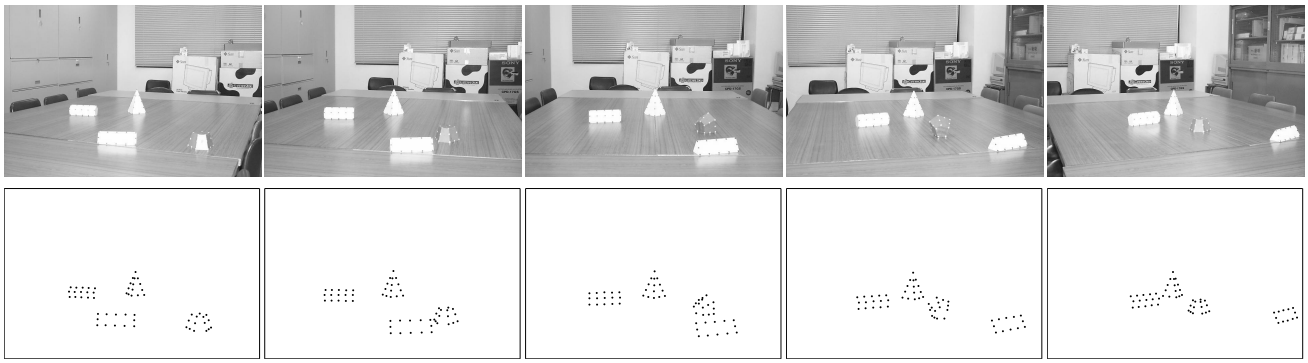


Figure 5: Real images of moving objects (above) and the selected feature points (below).

Since eq. (5) is equivalent to $\boldsymbol{G} = \boldsymbol{W}^\top \boldsymbol{W}$, we see that $\{\tilde{\boldsymbol{n}}_1, ..., \tilde{\boldsymbol{n}}_{\nu_1}, \tilde{\boldsymbol{n}}'_1, ..., \tilde{\boldsymbol{n}}'_{\nu_2}\}$ are an orthonormal system of the eigenvectors of $\boldsymbol{G}$ for eigenvalue 0. If we let $\{\boldsymbol{v}_{r+1}, ..., \boldsymbol{v}_N\}$ be an arbitrary orthonormal system of the eigenvectors of $\boldsymbol{G}$ for eigenvalue 0, there exists a $\nu \times \nu$ orthogonal matrix $\boldsymbol{C}$ such that these two are related by

$$\begin{pmatrix} \boldsymbol{v}_{r+1} \cdots \boldsymbol{v}_N \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{n}}_1 \cdots \tilde{\boldsymbol{n}}_{\nu_1} \ \tilde{\boldsymbol{n}}'_1 \cdots \tilde{\boldsymbol{n}}'_{\nu_2} \end{pmatrix} \boldsymbol{C}. \quad (14)$$

Consider the $N \times N$ matrix whose $(\alpha\beta)$ element is the inner product of the αth and βth rows of the $N \times \nu$ matrix $\begin{pmatrix} \boldsymbol{v}_{r+1} \cdots \boldsymbol{v}_N \end{pmatrix}$. We observe that

$$\begin{aligned}
&\begin{pmatrix} \boldsymbol{v}_{r+1} \cdots \boldsymbol{v}_N \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_{r+1} \cdots \boldsymbol{v}_N \end{pmatrix}^\top \\
&= \begin{pmatrix} \tilde{\boldsymbol{n}}_1 \cdots \tilde{\boldsymbol{n}}_{\nu_1} \ \tilde{\boldsymbol{n}}'_1 \cdots \tilde{\boldsymbol{n}}'_{\nu_2} \end{pmatrix} \boldsymbol{C}\boldsymbol{C}^\top \\
&\quad \begin{pmatrix} \tilde{\boldsymbol{n}}_1 \cdots \tilde{\boldsymbol{n}}_{\nu_1} \ \tilde{\boldsymbol{n}}'_1 \cdots \tilde{\boldsymbol{n}}'_{\nu_2} \end{pmatrix}^\top \\
&= \begin{pmatrix} \tilde{\boldsymbol{n}}_1 \cdots \tilde{\boldsymbol{n}}_{\nu_1} \ \tilde{\boldsymbol{n}}'_1 \cdots \tilde{\boldsymbol{n}}'_{\nu_2} \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{n}}_1 \cdots \tilde{\boldsymbol{n}}_{\nu_1} \ \tilde{\boldsymbol{n}}'_1 \cdots \tilde{\boldsymbol{n}}'_{\nu_2} \end{pmatrix}^\top \\
&= \begin{pmatrix} \boldsymbol{n}_1 \cdots \boldsymbol{n}_{\nu_1} & \boldsymbol{0} \cdots \boldsymbol{0} \\ \boldsymbol{0} \cdots \boldsymbol{0} & \boldsymbol{n}'_1 \cdots \boldsymbol{n}'_{\nu_2} \end{pmatrix} \begin{pmatrix} \boldsymbol{n}_1^\top & \boldsymbol{0}^\top \\ \vdots & \vdots \\ \boldsymbol{n}_{\nu_1}^\top & \boldsymbol{0}^\top \\ \boldsymbol{0}^\top & \boldsymbol{n}'_1{}^\top \\ \vdots & \vdots \\ \boldsymbol{0}^\top & \boldsymbol{n}'_{\nu_2}{}^\top \end{pmatrix} \\
&= \begin{pmatrix} * & \boldsymbol{O} \\ \boldsymbol{O} & \dagger \end{pmatrix}, \qquad (15)
\end{aligned}$$

where $(*)$ and $(\dagger)$ are $N_1 \times N_1$ and $N_2 \times N_2$ submatrices, respectively. This implies that the αth and βth rows of the matrix $\begin{pmatrix} \boldsymbol{v}_{r+1} \cdots \boldsymbol{v}_N \end{pmatrix}$ are mutually orthogonal if $\boldsymbol{p}_\alpha$ and $\boldsymbol{p}_\beta$ belong to different subspaces.

Let $\{\boldsymbol{v}_1, ..., \boldsymbol{v}_r\}$ be an arbitrary orthonormal system of the eigenvectors of the matrix $\boldsymbol{G}$ for nonzero eigenvalues. Combining these with $\{\boldsymbol{v}_{r+1}, ..., \boldsymbol{v}_N\}$, we obtain an orthonormal system of the eigenvectors of the matrix $\boldsymbol{G}$ for all the eigenvalues. It follows that the $N \times N$ matrix

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{v}_1 \cdots \boldsymbol{v}_r \ \boldsymbol{v}_{r+1} \cdots \boldsymbol{v}_N \end{pmatrix} \quad (16)$$

is orthogonal. Hence, its N rows are pair-wise orthogonal. If we let $v_{\alpha i}$ be the αth element of vector $\boldsymbol{v}_i$, the αth and βth rows of the matrix $\boldsymbol{V}$ are $(v_{\alpha 1}, ..., v_{\alpha N})$ and $(v_{\beta 1}, ..., v_{\beta N})$, respectively. It follows that for $\alpha \neq \beta$ we have

$$\begin{aligned}
&v_{\alpha 1} v_{\beta 1} + \cdots + v_{\alpha r} v_{\beta r} \\
&\quad + v_{\alpha (r+1)} v_{\beta j (r+1)} + \cdots + v_{\alpha N} v_{\beta N} = 0. \quad (17)
\end{aligned}$$

We have already shown that $v_{\alpha (r+1)} v_{\beta (r+1)} + \cdots + v_{\alpha N} v_{\beta N} = 0$ if $\boldsymbol{p}_\alpha$ and $\boldsymbol{p}_\beta$ belong to different subspaces. This means that if $\boldsymbol{p}_\alpha$ and $\boldsymbol{p}_\beta$ belong to different subspaces, we have

$$v_{\alpha 1} v_{\beta 1} + \cdots + v_{\alpha r} v_{\beta r} = 0. \quad (18)$$

This implies that if $\boldsymbol{p}_\alpha$ and $\boldsymbol{p}_\beta$ belong to different subspaces, the αth and βth rows of the $N \times r$ matrix

$$\boldsymbol{V}_r = \begin{pmatrix} \boldsymbol{v}_1 \cdots \boldsymbol{v}_r \end{pmatrix} \quad (19)$$

are mutually orthogonal. The $N \times N$ matrix whose $(\alpha\beta)$ element is the inner product of the αth and βth rows of the matrix $\boldsymbol{V}_r$ is given by

$$\boldsymbol{V}_r \boldsymbol{V}_r^\top = \begin{pmatrix} \boldsymbol{v}_1 \cdots \boldsymbol{v}_r \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_1 \cdots \boldsymbol{v}_r \end{pmatrix}^\top = \sum_{i=1}^{r} \boldsymbol{v}_i \boldsymbol{v}_i^\top = \boldsymbol{Q}. \quad (20)$$

Hence, the $(\alpha\beta)$ element of the interaction matrix $\boldsymbol{Q}$ is zero if $\boldsymbol{p}_\alpha$ and $\boldsymbol{p}_\beta$ belong to different subspaces.

We have so far assumed that $\boldsymbol{p}_1, ..., \boldsymbol{p}_{N_1} \in \mathcal{L}_1$ and $\boldsymbol{p}_{N_1+1}, ..., \boldsymbol{p}_N \in \mathcal{L}_2$. It is easy to see that the theorem holds if we arbitrarily permute $\boldsymbol{p}_1, ..., \boldsymbol{p}_N$. If $\boldsymbol{p}_\alpha$ and $\boldsymbol{p}_\beta$ are interchanged, the αth and βth rows and the αth and βth columns of the matrix $\boldsymbol{G}$ are simultaneously interchanged. As a result, its αth and βth eigenvectors are interchanged, and hence the αth and βth columns of the matrix $\boldsymbol{V} = \begin{pmatrix} \boldsymbol{v}_1 \cdots \boldsymbol{v}_r \end{pmatrix}$ are interchanged. It follows that the αth and βth rows and the αth and βth columns of the interaction matrix $\boldsymbol{Q}$ are simultaneously interchanged. Since any permutation of $\boldsymbol{p}_1, ..., \boldsymbol{p}_N$ can be generated by pair-wise interchanges, the theorem holds for an arbitrary permutation. The theorem can be straightforwardly extended to more than two subspaces. □