International Journal of Image and Graphics Vol. 2, No. 2 (2002) 179–197
© World Scientific Publishing Company

# MOTION SEGMENTATION BY SUBSPACE SEPARATION: MODEL SELECTION AND RELIABILITY EVALUATION

KENICHI KANATANI

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

Received 18 June 2001
Revised 14 October 2001
Accepted 27 November 2001

Reformulating the Costeira-Kanade algorithm as a pure mathematical theorem, we present a robust segmentation procedure, which we call *subspace separation*, by incorporating model selection using the geometric AIC. We then study the problem of estimating the number of independent motions using model selection. Finally, we present criteria for evaluating the reliability of individual segmentation results. Again, model selection plays an important role. We confirm the effectiveness of our method by experiments using synthetic and real images.

*Keywords*: Motion Segmentation; Model Selection; Reliability Evaluation; Geometric AIC; Geometric MDL.

## 1. Introduction

Segmenting individual objects from backgrounds is one of the most important of computer vision tasks. A significant clue is provided by motion; humans can easily discern independently moving objects by seeing their motions without knowing their identities. To solve this problem, Costeira and Kanade[1] presented a segmentation algorithm based on feature tracking. Since then, various extensions have been proposed: Gear[2] used the reduced row echelon form and graph matching; Ichimura[3] applied the discrimination criterion of Otsu[4] and the QR decomposition for feature selection[5]; Inoue and Urahama[6] introduced fuzzy clustering.

Costeira and Kanade[1] attributed their algorithm to the Tomasi-Kanade factorization.[7] In this paper, we show that their principle is a simple fact of linear algebra: the image motion of points that belong to an object moving rigidly in the scene is constrained to be in a four-dimensional subspace.[2,8] We then present a new segmentation procedure that directly exploits this fact and demonstrate that our method, which we call *subspace separation*, far outperforms the Costeira-Kanade algorithm[1] and Ichimura's method.[5] Our method can also be used for analyzing the effect of illumination on moving objects.[9,10]

The subspace separation method assumes that the number of independent motions is given. It has been known that estimating the number of motions is more difficult than segmentation itself.[1,2] Gear[2] attempted a complicated statistical analysis for estimating the number of motions and concluded that individual points were likely to be judged as independently moving. This is a natural consequence, because saying that each point is moving independently (but coincidentally in unison) is, as long as the judgment is based on statistical likelihood alone, always a more likely interpretation than saying that the motions of different points are constrained. It follows that we need a criterion that favors a small number of motions. In this paper, we show that the problem can be solved by *model selection* without introducing any empirical thresholds.

Once the number $m$ of motions is given, the subspace separation algorithm segments the points into $m$ groups, but the result may not always be correct. In this paper, we also study the method for evaluating the correctness of the segmentation *a posteriori*. Again, model selection plays an important role.

Sec. 2 describes the mathematical structure of the problem. The details of the segmentation procedure are given in Sec. 3. In Sec. 4, we discuss the problem of estimating the number of independent motions. In Sec. 5, we present criteria for evaluating the reliability of individual segmentation results. In Sec. 6, we give our conclusion.

## 2. Subspace Separation

### 2.1. *Motion subspaces*

We track $N$ rigidly moving feature points over $M$ images and let $(x_{\kappa\alpha}, y_{\kappa\alpha})$ be the image coordinates of the $\alpha$th point in the $\kappa$th frame. If all the image coordinates are stacked vertically into a $2M$-dimensional vector in the form

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

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

The $XYZ$ camera coordinate system is regarded as the world coordinate system with the $Z$-axis taken along the optical axis. 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 $\kappa$th frame. Let $(a_\alpha, b_\alpha, c_\alpha)$ be the object coordinates of the $\alpha$th point. Its position in the $\kappa$th frame with respect to the world coordinate system is given by

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

If orthographic projection is assumed, we have

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

where $\tilde{\boldsymbol{t}}_\kappa$, $\tilde{\boldsymbol{i}}_\kappa$, $\tilde{\boldsymbol{j}}_\kappa$, and $\tilde{\boldsymbol{k}}_\kappa$ are the two-dimensional vectors obtained from $\boldsymbol{t}_\kappa$, $\boldsymbol{i}_\kappa$, $\boldsymbol{j}_\kappa$, and $\boldsymbol{k}_\kappa$, respectively, by chopping off the third components. If the vectors $\tilde{\boldsymbol{t}}_\kappa$, $\tilde{\boldsymbol{i}}_\kappa$, $\tilde{\boldsymbol{j}}_\kappa$, and $\tilde{\boldsymbol{k}}_\kappa$ are stacked over the $M$ frames vertically into $2M$-dimensional vectors $\boldsymbol{m}_0$, $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, and $\boldsymbol{m}_3$, respectively, the vector $\boldsymbol{p}_\alpha$ has the form

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

This implies that the $N$ points $\{\boldsymbol{p}_\alpha\}$ should belong to the *four-dimensional subspace* spanned by the vectors $\{\boldsymbol{m}_0, \boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3\}$. This constraint holds for all *affine camera* models including weak perspective and paraperspective,[11] because Eq.(2) holds irrespective of the *metric condition*[7,11] that demands $\{\boldsymbol{i}_\kappa, \boldsymbol{j}_\kappa, \boldsymbol{k}_\kappa\}$ be orthonormal.

For *planar* motions, i.e., when 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 $\boldsymbol{i}_\kappa$, $\boldsymbol{j}_\kappa$, and $\boldsymbol{k}_\kappa$ are taken to be in the $X$, $Y$, and $Z$ directions, respectively. This means that the $N$ points $\{\boldsymbol{p}_\alpha\}$ belong to the *three-dimensional* subspace spanned by $\{\boldsymbol{m}_0, \boldsymbol{m}_1, \boldsymbol{m}_2\}$.

### 2.2. *Subspace separation theorem*

It follows that the motions of the feature points are segmented into independently moving objects if the $N$ points in $\mathcal{R}^n$ $(n = 2M)$ are grouped into distinct four-dimensional subspaces (or three-dimensional subspaces for planar motions). Inspired by the Tomasi-Kanade factorization,[7] Costeira and Kanade[1] found that this can be done by zero/nonzero discrimination of the elements of a matrix computed from the data. We reiterate this fact from a new viewpoint.[8]

Let $\{\boldsymbol{p}_\alpha\}$ be $N$ points that belong to an $r$-dimensional subspace $\mathcal{L} \subset \mathcal{R}^n$. Define the $N \times N$ *metric matrix* $\boldsymbol{G} = (G_{\alpha\beta})$ 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 $\boldsymbol{b}$. Let $\lambda_1 \geq \cdots \geq \lambda_N$ be the eigenvalues of $\boldsymbol{G}$, 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 $\mathcal{L}_i$ be the subspace defined by the $i$th set $\{\boldsymbol{p}_\alpha\}$, $\alpha \in \mathcal{I}_i$. If the $m$ subspaces $\mathcal{L}_i$, $i = 1, ..., m$, are linearly independent, we have the following:

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

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

This theorem is the essence of the principle on which the Costeira-Kanade algorithm[1] is built. Costeira and Kanade described this result in reference to the Tomasi-Kanade factorization,[7] but this 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 $\sum_{\alpha=1}^{N} c_\alpha \boldsymbol{p}_\alpha = \mathbf{0}$, 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" $\{c_\alpha\}$ (the "null space" to be precise) is generated by those for which $\sum_{\mathbf{p}_\alpha \in \mathcal{L}_1} c_\alpha \boldsymbol{p}_\alpha = \mathbf{0}$ and those for which $\sum_{\mathbf{p}_\alpha \in \mathcal{L}_2} c_\alpha \boldsymbol{p}_\alpha = \mathbf{0}$. A formal proof is given in the Appendix.

## 3. Separation Procedure

### 3.1. *Basic principle*

In the presence of noise, all the elements of $\boldsymbol{Q}$ are nonzero in general. But if we progressively group points $\boldsymbol{p}_\alpha$ and $\boldsymbol{p}_\beta$ for which $|Q_{\alpha\beta}|$ is large and interchange the corresponding rows and columns of $\boldsymbol{Q}$, we should end up with an approximately block-diagonal matrix. Costeira and Kanade[1] proposed this type of strategy, known as the *greedy algorithm*.

Whatever realigning strategy is taken, however, all such algorithms[1,2,3,5,6] are severely vulnerable to noise, because segmentation is based on zero/nonzero discrimination of matrix elements, for which we do not know how small the nonzero elements might be in the absence of noise, yet small nonzero elements and large nonzero elements have the same meaning as long as they are nonzero.

In the presence of noise, a small error in one datum can affect all the elements of the matrix in a complicated manner, so finding a suitable threshold is difficult even if the noise is known to be Gaussian with a known variance.[2] This difficulty can be avoided if we work in the *original data space*, rather than the matrix elements derived from it.

### 3.2. *Subspace merging*

Initially regarding individual points as groups consisting of one element, we successively merge two groups by fitting a $d$-dimensional subspace to them ($d = 4$ for general motions and $d = 3$ for planar motions). Let $\mathcal{L}_i$ and $\mathcal{L}_j$ be candidate subspaces to merge, and $N_i$ and $N_j$ the respective numbers of points in them.

Let $J_i$ and $J_j$ be the individual residuals, i.e., the sums of the squared distances of the data points to the fitted subspaces $\mathcal{L}_i$ and $\mathcal{L}_j$. Let $J_{i\oplus j}$ be the residual that would result if a single subspace is fitted to the $N_i + N_j$ points. It is reasonable not to merge the two groups if $J_{i\oplus j}$ is much larger than $J_i + J_j$. But how large should $J_{i\oplus j}$ be for this judgment? In fact, we always have $J_{i\oplus j} \geq J_i + J_j$ because a single subspace has fewer degrees of freedom to adjust than two subspaces. It follows that we need to balance the increase in the residual against the decrease in the degrees of freedom. For this purpose, we use the *geometric AIC*.[12,13]

We assume that the tracked feature points are perturbed from their true positions by independent Gaussian noise of mean zero and standard deviation $\epsilon$, which we call the *noise level*. Since a $d$-dimensional subspace in $\mathcal{R}^n$ has $d(n-d)$ degrees of freedom,[a] the geometric AIC has the following form[13]:

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

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 given as follows[13]:

$$\text{G-AIC}_{i,j} = J_i + J_j + 2\Big(3(N_i + N_j) + 8(n-3)\Big)\epsilon^2. \tag{9}$$

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, all the subspaces are included in subspaces of higher dimensions, so the interaction matrix $\boldsymbol{Q}$ still provides information about the possibility of merging. Here, we integrate these two criteria to define the following similarity measure between the subspaces $\mathcal{L}_i$ and $\mathcal{L}_j$:

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

Two subspaces with the largest similarity are merged successively until the number of subspaces becomes a specified number $m$. If some of the resulting subspaces contain less than $d$ elements, they are taken as first candidates to be merged.

To evaluate the geometric AIC, we need to estimate the noise level $\epsilon$. This can be done if we note that the points $\{\boldsymbol{p}_\alpha\}$ should belong to an $r$-dimensional subspace in $\mathcal{R}^n$ in the absence of noise ($r = md$). Let $J_t$ be the residual after fitting an $r$-dimensional subspace to $\{\boldsymbol{p}_\alpha\}$. Then, $J_t/\epsilon^2$ is subject to a $\chi^2$ distribution with $(n-r)(N-r)$ degrees of freedom,[13] so we obtain the following unbiased estimator of $\epsilon^2$:

$$\hat{\epsilon}^2 = \frac{J_t}{(n-r)(N-r)}. \tag{11}$$

### 3.3. *Dimension correction and robust fitting*

We also incorporate two further techniques which prove very effective. The first is *dimension correction*: as soon as more than $d$ elements are grouped together, we optimally fit a $d$-dimensional subspace to them and replace the points with their projections onto the fitted subspace for computing the matrix $\boldsymbol{Q}$ (the projection is done always using the original data at each merging stage). 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 $\boldsymbol{Q}$.

[a]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 $dn - d^2$.

The second technique is an *a posteriori* reallocation. Since a point that is misclassified in the course of merging never leaves that class, we attempt to remove outliers from the $m$ resulting classes $\mathcal{L}_1$, ..., $\mathcal{L}_m$ by robust fitting. 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 *LMedS*.[14] Each data point is reallocated to the closest one.

### 3.4. *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 independently 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 can regard the performance of this oracle method as a bound on the accuracy.

### 3.5. *Experiments*

Figure 1 is a sequence of five consecutive simulated images ($512 \times 512$ pixels) of 20 points in the background and nine points in an object. The background and the object are independently moving in two dimensions. We added Gaussian noise of mean zero and standard deviation $\epsilon$ to the coordinates of the 29 points independently and classified them into two groups.

Figure 2(a) plots the average misclassification ratio over 500 independent trials for different $\epsilon$: we compare (i) the method using the greedy algorithm only, (ii) the method with dimension correction added, (iii) the method with model selection in addition, and (iv) the method with robust fitting further added. We can see that each added technique reduces the error further.

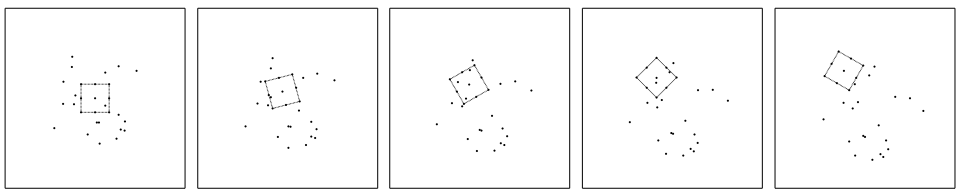


Fig. 1. An image sequence of points in planar motion.

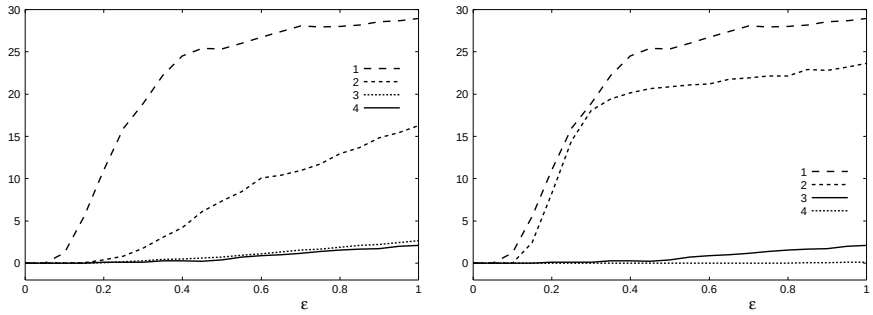



Fig. 2. Misclassification ratio for segmenting the planar motion of Fig. 1. (a) (i) Greedy algorithm. (ii) With dimension correction. (iii) With model selection. (iv) With robust fitting. (b) (i) Greedy algorithm. (ii) Ichimura's method. (iii) Our method. (iv) Lower bound.

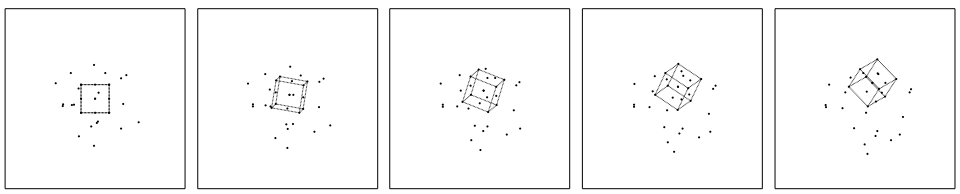


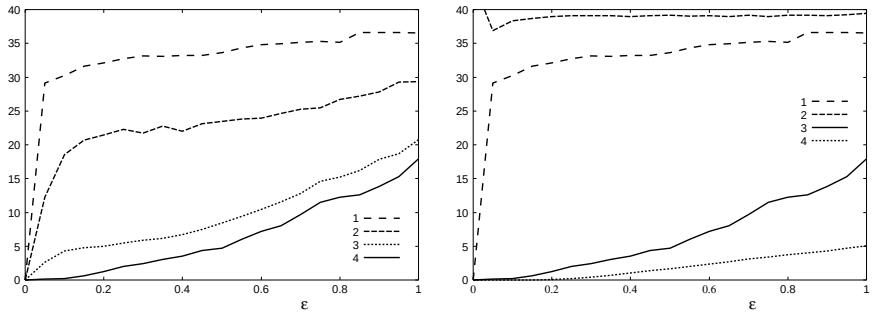
Fig. 3. An image sequence of points in 3D motion.

In Fig. 2(b), we compare (i) the greedy algorithm, (ii) our method with all the techniques combined, (iii) Ichimura's method[3] that uses the discrimination criterion of Otsu,[4] and (iv) the bound given by the oracle method. We 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.

Figure 3 is a sequence of five consecutive simulated images ($512\times512$ pixels) of 20 points in the background and 14 points in an object. The background and the object are independently moving in three dimensions. Figure 4 shows the classification results corresponding to Fig. 2. Again, we can see that our method dramatically improves the classification accuracy.

Figure 5 shows a sequence of real images (above) and feature points manually selected from them (below). Three objects are fixed in the scene, so they are moving rigidly with the scene, while one object is moving relative to the scene. The image size is $512 \times 768$ pixels.

For this data set, the greedy algorithm and our method both correctly separated an independent 3D motion from the background motion, whereas Ichimura's method




(a) (b)

Fig. 4. Misclassification ratio for segmenting the 3D motion of Fig. 3. (a) (i) Greedy algorithm. (ii) With dimension correction. (iii) With model selection. (iv) With robust fitting (b) (i) Greedy algorithm. (ii) Ichimura's method. (iii) Our method. (iv) Lower bound.

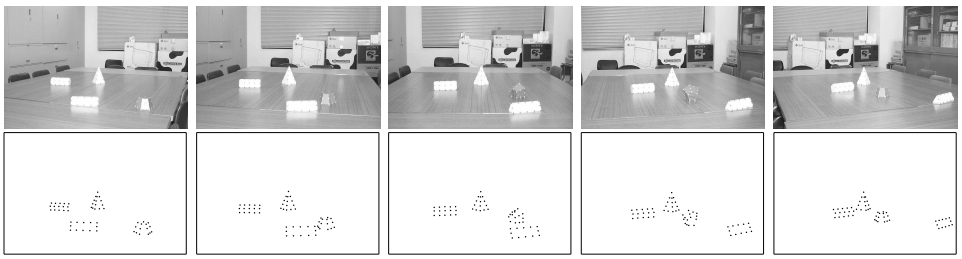


Fig. 5. Real images of moving objects (above) and the selected feature points (below).

failed. We added independent Gaussian noise of mean zero and standard deviation $\epsilon = 0, 1, 2, 3, \ldots$ (pixels) to the coordinates of the feature points and applied our method ten times for each $\epsilon$, 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)

## 4. Estimating the Number of Motions

### 4.1. *What is the true number of motions?*

We now turn to the problem of estimating the "number of motions". Here, a very subtle problem arises: what is the "true" number of motions?

Suppose six objects are moving in the scene in three ways, say, objects 1 and 2 are moving rigidly as a whole, and so are objects 3 and 4 and objects 5 and 6. What is the correct answer for the number of motions? Obviously, we expect it to be 3, but if we say that it is 6, what is wrong? After segmenting the points into six groups, we can compute their 3D motions by a structure-from-motion algorithm.

We may find *a posteriori* that the motions of some groups are very similar to others, concluding that there are three motions in total. This 3D interpretation is correct.

By the same token, saying that the number of motions is five is also correct, and so are the answers four and three. But the answer two is wrong. In other words, the correct answers form a spectrum, the lower bound being three and the upper bound being the number of points observed. The surest answer is the upper bound, but the 3D reconstruction becomes more vulnerable to noise as the number of feature points in one group becomes smaller. So, we want to minimize the overestimation, but at the same time some "safety margin" is desired. We must take this into consideration in estimating the "number of motions".

### 4.2. *Rank estimation*

From the argument in Sec. 2.1, the number $m$ of independent motions can be inferred from the *rank* $r$ of the data vectors $\{\boldsymbol{p}_\alpha\}$, i.e., the dimension of the subspace they span: $r = 4m$ for general motions; $r = 3m$ for planar motions. Since $M$ images define a $2M$-dimensional space, discerning $m$ independent motions requires more than $2m$ images for general motion and more than $1.5m$ images for planar motions.

Mathematically, there are basically three ways for computing the rank $r$ of a set $\{\boldsymbol{p}_\alpha\}$ of $N$ $n$-dimensional vectors.

- Define the $n \times n$ *moment matrix* $\boldsymbol{M}$ by

$$\boldsymbol{M} = \sum_{\alpha=1}^{N} \boldsymbol{p}_\alpha \boldsymbol{p}_\alpha^\top. \tag{12}$$

  Let $\lambda_1 \geq \cdots \geq \lambda_n$ be its eigenvalues, and $\{\boldsymbol{u}_1, ..., \boldsymbol{u}_n\}$ the corresponding orthonormal set of eigenvectors. The rank $r$ equals the number of positive eigenvalues, i.e., the rank of $\boldsymbol{M}$.

- Define the $N \times N$ *metric matrix* $\boldsymbol{G} = (G_{\alpha\beta})$ by

$$G_{\alpha\beta} = (\boldsymbol{p}_\alpha, \boldsymbol{p}_\beta). \tag{13}$$

  Let $\lambda_1 \geq \cdots \geq \lambda_N$ be its eigenvalues, and $\{\boldsymbol{v}_1, ..., \boldsymbol{v}_N\}$, the corresponding orthonormal set of eigenvectors. The rank $r$ equals the number of positive eigenvalues,[b] i.e., the rank of $\boldsymbol{G}$.

- Define the $n \times N$ *observation matrix* $\boldsymbol{W}$ by

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{p}_1 & \cdots & \boldsymbol{p}_N \end{pmatrix}. \tag{14}$$

  Let its singular value decomposition be

$$\boldsymbol{W} = \boldsymbol{U}_{n\times N} \mathrm{diag}(\sigma_1, \sigma_2, ..., \sigma_N) \boldsymbol{V}_{N\times N}^\top, \tag{15}$$

  where $\boldsymbol{U}_{n\times N}$ and $\boldsymbol{V}_{N\times N}$ are, respectively, $n \times N$ and $N \times N$ matrices having orthonormal columns. The symbol $\mathrm{diag}(\cdots)$ designates the diagonal matrix

[b]The positive eigenvalues of $\boldsymbol{G}$ are also positive eigenvalues of $\boldsymbol{M}$.

with $\cdots$ as its diagonal elements in that order. The rank $r$ equals the number of positive singular values,[c] i.e., the rank of $\boldsymbol{W}$.

However, *all these are in theory only*: they can be applied only when no noise exists in the data and the computation can be done with infinite accuracy. In the presence of noise and with finite computation accuracy, all eigenvalues and singular values are nonzero in general. Hence, we need to truncate small eigenvalues and singular values, but it is difficult to set an appropriate threshold.

### 4.3. *Model selection for rank estimation*

A naive idea for estimating the rank of a set $\{\boldsymbol{p}_\alpha\}$ of $N$ $n$-dimensional vectors is to fit subspaces of different dimensions to them and adopt the dimension of the subspace having the smallest residual. This does not work, as we pointed out in Sec. 3.2, because the residual becomes smaller as we fit a higher dimensional subspace. In particular, the residual is zero if we fit the $n$-dimensional subspace (= the entire space). Also, the residual becomes smaller if the subspace to fit has more parameters to adjust. Again, we need to balance the residual against the dimension and the degree of freedom of the subspace. First, we consider the *geometric AIC*[13] and the *geometric MDL*.[15,16]

As mentioned in Sec. 3.2, an $r$-dimensional subspace in $\mathcal{R}^n$ has $r(n-r)$ degrees of freedom, and we are assuming independent Gaussian noise of mean zero and variance $\epsilon^2$. Hence, the geometric AIC and the geometric MDL are respectively given as

$$\begin{aligned} \text{G-AIC} &= J_r + 2r(m+n-r)\epsilon^2, \\ \text{G-MDL} &= J_r - r(m+n-r)\epsilon^2 \log\Bigl(\frac{\epsilon}{L}\Bigr)^2. \end{aligned} \tag{16}$$

Here, $L$ is a reference length for the data[16]; we can use for it an arbitrary value whose order is approximately the same as the data,[16] say the image size. Let $\nu = \min(n, m)$. The residual $J_r$ is given by

$$J_r = \sum_{i=r+1}^{\nu} \sigma_i^2, \tag{17}$$

where $\{\sigma_i\}$ are the singular values, in descending order, of the observation matrix $\boldsymbol{W}$ defined by Eq.(14). Evaluating Eqs.(16) for $r = 1, 2, \ldots$, we choose the value $r$ that minimizes them. If an upper bound $r_{\max}$ on the rank $r$ is known, the noise variance $\epsilon^2$ is estimated as follows (see Eq.(11)):

$$\hat{\epsilon}^2 = \frac{J_{r_{\max}}}{(n - r_{\max})(N - r_{\max})}. \tag{18}$$

[c]The squares of the positive singular values are the positive eigenvalues of $\boldsymbol{M}$ and $\boldsymbol{G}$.

The geometric AIC and the geometric MDL effectively truncate the eigenvalues and singular values without using any threshold. A well known automatic thresholding scheme is the discrimination criterion of Otsu,[4] which Ichimura[3] used for thresholding the interaction matrix $\boldsymbol{Q}$. Applying it to threshold the singular values $\{\sigma_i\}$ of $\boldsymbol{W}$, we obtain the following criterion, which we call the *Otsu-Ichimura criterion*:

$$\mathrm{OIC} = \frac{r(\nu - r)(\mu_1 - \mu_2)^2}{\sum_{i=1}^{r}(\sigma_i - \mu_1)^2 + \sum_{i=r+1}^{\nu}(\sigma_i - \mu_2)^2}. \tag{19}$$

Here, we define

$$\mu_1 = \frac{1}{r}\sum_{i=1}^{r}\sigma_i, \qquad \mu_2 = \frac{1}{\nu - r}\sum_{i=1+r}^{\nu}\sigma_i. \tag{20}$$

The number $r$ that maximizes Eq.(19) is chosen as the rank.

### 4.4. *Experiments*

We defined a $10 \times 20$ matrix with random elements uniformly generated over $[-1, 1]$. We computed its singular value decomposition in the form $\boldsymbol{V}\mathrm{diag}(\sigma_1, ..., \sigma_{10})\boldsymbol{U}^\top$. The nonzero singular values $\sigma_1$, ..., $\sigma_5$ are 3.81, 3.58, 3.09, 2.98, and 2.75, respectively. Then, we defined the matrix

$$\boldsymbol{A} = \boldsymbol{V}\mathrm{diag}(\sigma_1, ..., \sigma_5, \gamma\sigma_5, 0, ..., 0)\boldsymbol{U}^\top. \tag{21}$$

We added Gaussian noise of mean zero and standard deviation 0.05 to each element of $\boldsymbol{A}$ independently and estimated its rank with $r_{\max} = 6$. Figure 6 plots the average number of the rank over 200 trials for each $\gamma$. We used the reference length $L = 1$.

The geometric AIC predicts the rank to be six with some probability even when the true rank is five ($\gamma = 0$), and it mostly predicts the rank to be six for a small

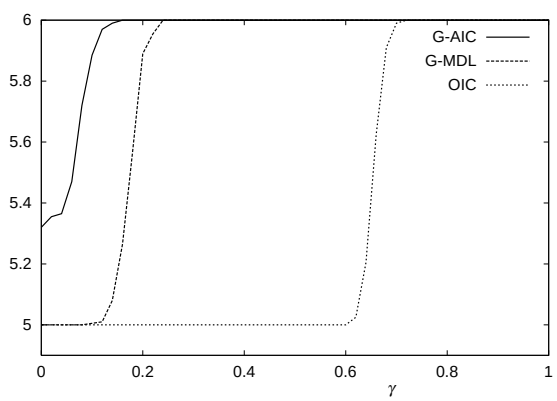



Fig. 6. Average estimated rank. The solid lines are for the geometric AIC; the broken lines are for the geometric MDL; the dotted lines are for the OIC.

value of $\gamma$. The geometric MDL almost always guesses the rank to be five when the true rank is five ($\gamma = 0$), but it keeps guessing the rank to be five for a wide range of $\gamma$. The OIC is likely to threshold smaller singular values to zero and predict the rank to be five up to a fairly large value of $\gamma$.

Figure 7 is a sequence of five consecutive simulated images ($512 \times 512$ pixels) of 20 background points, nine object points, and another nine object points, each independently moving rigidly within the image plane. We added Gaussian noise of mean zero and standard deviation $\epsilon$ to the coordinates of the 38 points independently and estimated the number of independent motions. In this case, we only need to decide if the rank is 3, 6, 9, ... corresponding to 1, 2, 3, ... motions.

Figure 8 plots the average number of detected motions over 500 trials for each $\epsilon$ with the upper bound $m_{\max} = 4$. We used the reference length $L = 600$. The geometric AIC (solid line) tends to overestimate the number of motions but is stable. The geometric MDL (broken line) estimates the correct number of motions when the noise is very small but tends to underestimate it as the noise increases. The OIC (dotted lines) always estimates the number of motions to be unity.

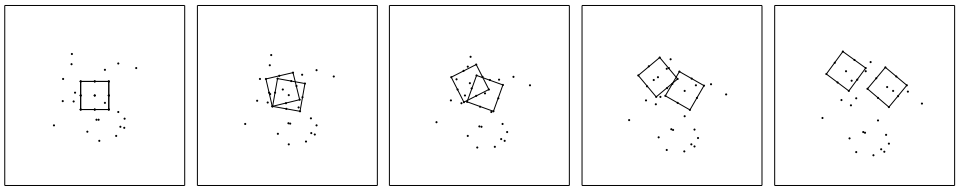


Fig. 7. Points moving two-dimensionally.

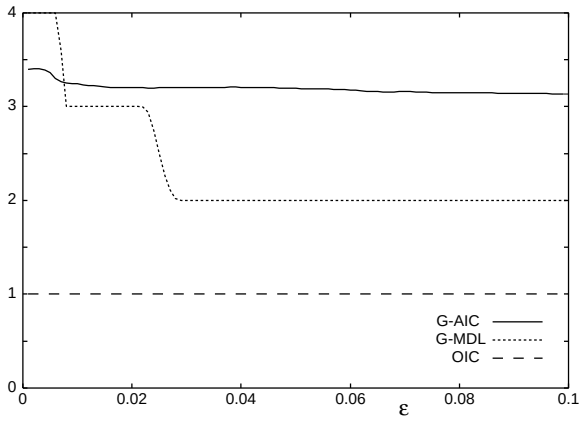



Fig. 8. Average estimate number of motions. The solid lines are for the geometric AIC; the broken lines are for the geometric MDL; the dotted lines are for the OIC.

We then applied our criteria to the image sequence of Fig. 5. Letting[d] $m_{\max} = 2$, we found that the number $m$ of independent motions was two according to both the geometric AIC and the geometric MDL but was one according to the OIC. Using the subspace separation method with $m = 2$, we obtained a correct segmentation.

The five images Fig. 5 were actually decimated from a motion sequence. Inserting five intermediate images, we estimated the number of motions with $m_{\max} = 3$, 4. The geometric AIC and the geometric MDL both predicted $m = 3$, while the OIC predicted $m = 1$. Using the subspace separation algorithm with $m = 3$, we detected the independently moving object correctly. In this sense, the estimations by the geometric AIC and the geometric MDL are both correct, while the estimation by the OIC cannot be correct (see Sec. 4.1).

From the above results, we can conclude that the geometric AIC tends to be faithful to noise and overestimate the rank, while the geometric MDL tends to ignore noise and slightly underestimate the rank. The OIC inadmissibly underestimates the rank.

In doing these experiments, we have found that if we use a large upper bound for $r_{\max}$, the residual $J_{r_{\max}}$ in Eq.(18) becomes so small that the lack of significant digits causes underestimation of $\hat{\epsilon}^2$, often regarding it as zero. In fact, the irregular behavior at the left end of the plot of Fig. 8 is due to the lack of significant digits for computing very small residuals. Hence, the upper bound $r_{\max}$ should be taken to be as small as possible.

## 5. Reliability Evaluation for Segmentation

Although the subspace separation algorithm performs very well, there is no guarantee that a particular segmentation is correct. So, we need some means for an *a posteriori* reliability evaluation.

### 5.1. *F test*

Suppose the $N$ points $\{\boldsymbol{p}_\alpha\}$ that represent the feature motion history are segmented into $m$ groups having $N_i$ points, $i = 1, ..., m$. Let $J_i$ be the residual of fitting a $d$-dimensional subspace $\mathcal{L}_i$ to the $i$th group, and $J_t$ the residual of fitting an $md$-dimensional subspace $\mathcal{L}_t$ to the entire $N$ points $\{\boldsymbol{p}_\alpha\}$. As argued in Sec. 3.2, $J_i/\epsilon^2$ should be subject to a $\chi^2$ distribution with

$$\phi_i = (n-d)N_i - d(n-d) = (n-d)(N_i - d) \tag{22}$$

degrees of freedom *if the segmentation is correct*. On the other hand, $J_t/\epsilon^2$ should be subject to a $\chi^2$ distribution with

$$\phi_t = \sum_{i=1}^{m}(n-md)N_i - md(n-md) = (n-md)(N-md) \tag{23}$$

degrees of freedom *irrespective of the correctness of the segmentation*.

[d]In order to let $m_{\max} = 3$, we need seven or more images, cf. Sec. 4.2.

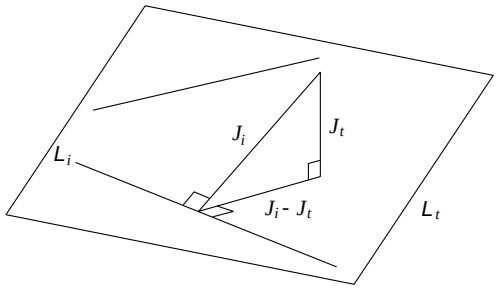



Fig. 9. The residual of subspace fitting.

The residual $J_i$ is the sum of squared distances of the points of the $i$th group to the subspace $\mathcal{L}_i$, which is the sum of their squared distances to the subspace $\mathcal{L}_t$ and the squared distances of their projections onto $\mathcal{L}_t$ to the subspace $\mathcal{L}_i$ (Fig. 9). Let us call the former the *external distances*, and the latter the *internal distances*. The sum of the squared internal distances for all the points is $\sum_{i=1}^m J_i - J_t$; the sum of the squared external distances is $J_t$.

If the segmentation is correct (the *null hypothesis*), $(\sum_{i=1}^m J_i - J_t)/\epsilon^2$ should also be subject to a $\chi^2$ distribution. The noise that contributes to the internal distances and the noise that contributes to the external distances are *orthogonal* to each other and hence independent. So, $(\sum_{i=1}^m J_i - J_t)/\epsilon^2$ has

$$\sum_{i=1}^m \phi_i - \phi_t = (m-1)d(N-md) \tag{24}$$

degrees of freedom.[13] It follows that

$$F = \frac{(\sum_{i=1}^m J_i - J_t)/(m-1)d(N-md)}{J_t/(n-md)(N-md)} \tag{25}$$

should be subject to an $F$ distribution with $(m-1)d(N-md)$ and $(n-md)(N-md)$ degrees of freedom. If this segmentation is not correct (the *alternative hypothesis*), the internal distances will increase on average while the external distances are not affected. It follows that if

$$F_{(n-md)(N-md)}^{(m-1)d(N-md)}(\alpha) < F, \tag{26}$$

this segmentation is rejected with significance level $\alpha$, where $F_{(n-md)(N-md)}^{(m-1)d(N-md)}(\alpha)$ is the upper $\alpha$th percentile of the $F$ distribution with $(m-1)d(N-md)$ and $(n-md)(N-md)$ degrees of freedom.

### 5.2. *Model selection*

The result of an $F$ test depends on the significance level, which we can arbitrarily set. Model selection can dispense with any thresholds.

The geometric AIC and the geometric MDL for the model that the segmentation is correct are, respectively, $\sum_{i=1}^{m}$ G-AIC$_i$ and $\sum_{i=1}^{m}$ G-MDL$_i$, where G-AIC$_i$ and G-MDL$_i$ are given as

$$\begin{aligned} \text{G-AIC}_i &= J_i + 2\Big(dN_i + d(n-d)\Big)\epsilon^2, \\ \text{G-MDL}_i &= J_i - \Big(dN_i + d(n-d)\Big)\epsilon^2 \log\Big(\frac{\epsilon}{L}\Big)^2. \end{aligned} \tag{27}$$

The geometric AIC and the geometric MDL for the model that the points $\{\boldsymbol{p}_\alpha\}$ can be somehow segmented into $m$ motions are given as

$$\begin{aligned} \text{G-AIC}_t &= J_t + 2\Big(mdN_i + md(n-md)\Big)\epsilon^2, \\ \text{G-MDL}_t &= J_t - \Big(mdN_i + md(n-md)\Big)\epsilon^2 \log\Big(\frac{\epsilon}{L}\Big)^2. \end{aligned} \tag{28}$$

Since the latter model is correct irrespective of the segmentation result, we can estimate the squared noise level $\epsilon^2$ from it, using

$$\hat{\epsilon}^2 = \frac{J_t}{(n-md)(N-md)}. \tag{29}$$

The condition that the segmentation is not correct is given by G-AIC$_t$ $<$ $\sum_{i=1}^{m}$ G-AIC$_i$ or G-MDL$_t$ $<$ $\sum_{i=1}^{m}$ G-MDL$_i$, which are rewritten, respectively, as

$$2 < F, \qquad -\log\Big(\frac{\epsilon}{L}\Big)^2 < F, \tag{30}$$

where $F$ is the $F$ statistic given by Eq.(25). Thus, model selection reduces to the $F$ test, the difference being that the threshold $F_{(n-md)(N-md)}^{(m-1)d(N-md)}(\alpha)$ is given automatically without specifying any significance level. When the noise is small, $-\log(\epsilon/L)^2$ is usually larger than two. This implies that the geometric AIC is more conservative than the geometric MDL, which is more confident of the particular result.

### 5.3. *Experiments*

We evaluated the reliability of the segmentation of the image sequence of Fig. 5. The $F$ statistic of Eq.(25) was computed to be $F = 0.893$. The upper 5th percentile is 1.346 in this case, so the correctness of the segmentation is not rejected with significance level 5%. The geometric AIC also selects this segmentation. Using the reference length $L = 600$, we have $-\log(\epsilon/L)^2 = 13.5$, so the geometric MDL also selects this segmentation, but it allows a much wider margin than the geometric AIC.

## 6. Concluding Remarks

In this paper, we first reformulated the Costeira-Kanade algorithm as a pure mathematical theorem independent of the Tomasi-Kanade factorization and presented a robust segmentation procedure, named *subspace separation*, by incorporating model

selection using the geometric AIC. We demonstrated by experiments using synthetic and real images that our algorithm dramatically improves the classification accuracy over existing methods. Our algorithm does not involve any parameters which need to be adjusted empirically.

Then, we studied the problem of estimating the number of independent motions. We elucidated the mathematical structure of the problem and presented an estimation method using model selection. We conclude that the use of geometric AIC seems to be a good practical choice.

Finally, we presented criteria for evaluating the reliability of individual segmentation results *a posteriori*, using the standard $F$ test and model selection. The use of model selection has the advantage that no significance level needs to be assigned. Again, the geometric AIC seems to be a feasible choice.

**Acknowledgements**

The author thanks Noriyoshi Kurosawa of FSAS Network Solutions Inc., Japan, and Chikara Matsunaga of FOR-A, Co. Ltd, Japan, for doing numerical simulations and real image experiments. He also thanks David Suter of Monash University, Australia, Atsuto Maki of Toshiba Co., Japan, and Naoyuki Ichimura of AIST, Japan, for helpful discussions. 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).

## Appendix A

### Appendix A. 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 $\boldsymbol{W}_1 = \begin{pmatrix} \boldsymbol{p}_1 & \cdots & \boldsymbol{p}_{N_1} \end{pmatrix}$ 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 $\boldsymbol{W}_2 = \begin{pmatrix} \boldsymbol{p}_{N_1} & \cdots & \boldsymbol{p}_N \end{pmatrix}$ 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 an $N_2$-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:

$$\tilde{\boldsymbol{n}}_i = \begin{pmatrix} \boldsymbol{n}_i \\ \boldsymbol{0} \end{pmatrix}, \qquad \tilde{\boldsymbol{n}}'_i = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{n}'_i \end{pmatrix}. \tag{1}$$

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 $\boldsymbol{W}$ defined

by Eq.(14). Since $\boldsymbol{W}$ 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 $\mathcal{N}$.

Since Eq.(13) 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

$$\begin{pmatrix} \boldsymbol{v}_{r+1} & \cdots & \boldsymbol{v}_N \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{n}}_1 & \cdots & \tilde{\boldsymbol{n}}_{\nu_1} & \tilde{\boldsymbol{n}}'_1 & \cdots & \tilde{\boldsymbol{n}}'_{\nu_2} \end{pmatrix} \boldsymbol{C}. \tag{2}$$

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

$$\begin{aligned}
&\begin{pmatrix} \boldsymbol{v}_{r+1} & \cdots & \boldsymbol{v}_N \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_{r+1} & \cdots & \boldsymbol{v}_N \end{pmatrix}^\top \\
&= \begin{pmatrix} \tilde{\boldsymbol{n}}_1 & \cdots & \tilde{\boldsymbol{n}}_{\nu_1} & \tilde{\boldsymbol{n}}'_1 & \cdots & \tilde{\boldsymbol{n}}'_{\nu_2} \end{pmatrix} \boldsymbol{C}\boldsymbol{C}^\top \begin{pmatrix} \tilde{\boldsymbol{n}}_1 & \cdots & \tilde{\boldsymbol{n}}_{\nu_1} & \tilde{\boldsymbol{n}}'_1 & \cdots & \tilde{\boldsymbol{n}}'_{\nu_2} \end{pmatrix}^\top \\
&= \begin{pmatrix} \tilde{\boldsymbol{n}}_1 & \cdots & \tilde{\boldsymbol{n}}_{\nu_1} & \tilde{\boldsymbol{n}}'_1 & \cdots & \tilde{\boldsymbol{n}}'_{\nu_2} \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{n}}_1 & \cdots & \tilde{\boldsymbol{n}}_{\nu_1} & \tilde{\boldsymbol{n}}'_1 & \cdots & \tilde{\boldsymbol{n}}'_{\nu_2} \end{pmatrix}^\top \\
&= \begin{pmatrix} \boldsymbol{n}_1 & \cdots & \boldsymbol{n}_{\nu_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{n}'_1 & \cdots & \boldsymbol{n}'_{\nu_2} \end{pmatrix} \begin{pmatrix} \boldsymbol{n}_1^\top & \boldsymbol{0}^\top \\ \vdots & \vdots \\ \boldsymbol{n}_{\nu_1}^\top & \boldsymbol{0}^\top \\ \boldsymbol{0}^\top & \boldsymbol{n}'^{\top}_1 \\ \vdots & \vdots \\ \boldsymbol{0}^\top & \boldsymbol{n}'^{\top}_{\nu_2} \end{pmatrix} = \begin{pmatrix} * & \boldsymbol{O} \\ \boldsymbol{O} & \dagger \end{pmatrix},
\end{aligned} \tag{3}$$

where $(*)$ and $(\dagger)$ are $N_1 \times N_1$ and $N_2 \times N_2$ submatrices, respectively. This implies that the $\alpha$th and $\beta$th 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 $\boldsymbol{G}$ for nonzero eigenvalues. Combining these with $\{\boldsymbol{v}_{r+1}, ..., \boldsymbol{v}_N\}$, we obtain an orthonormal system of the eigenvectors of $\boldsymbol{G}$ for all the eigenvalues. It follows that the $N \times N$ matrix

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_r & \boldsymbol{v}_{r+1} & \cdots & \boldsymbol{v}_N \end{pmatrix} \tag{4}$$

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

$$v_{\alpha 1} v_{\beta 1} + \cdots + v_{\alpha r} v_{\beta r} + v_{\alpha (r+1)} v_{\beta j (r+1)} + \cdots + v_{\alpha N} v_{\beta N} = 0. \tag{5}$$

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

$$v_{\alpha 1}v_{\beta 1} + \cdots + v_{\alpha r}v_{\beta r} = 0. \tag{6}$$

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

$$\boldsymbol{V}_r = \begin{pmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_r \end{pmatrix} \tag{7}$$

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

$$\boldsymbol{V}_r\boldsymbol{V}_r^\top = \begin{pmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_r \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_r \end{pmatrix}^\top = \sum_{i=1}^{r} \boldsymbol{v}_i\boldsymbol{v}_i^\top = \boldsymbol{Q}. \tag{8}$$

Hence, the $(\alpha\beta)$ element of $\boldsymbol{Q}$ 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 $\alpha$th and $\beta$th rows and the $\alpha$th and $\beta$th columns of $\boldsymbol{G}$ are simultaneously interchanged. As a result, its $\alpha$th and $\beta$th eigenvectors are interchanged, and hence the $\alpha$th and $\beta$th columns of the matrix $\boldsymbol{V} = \begin{pmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_r \end{pmatrix}$ are also interchanged. It follows that the $\alpha$th and $\beta$th rows and the $\alpha$th and $\beta$th columns of $\boldsymbol{Q}$ 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.

## References


1. J. P. Costeira and T. Kanade, *Int. J. Comput. Vision* **29**, 159 (1998).
2. C. W. Gear, *Int. J. Comput. Vision* **29**, 133 (1998).
3. N. Ichimura, in *Proc. 7th Int. Conf. Comput. Vision*, Sep. 1999, Kerkyra, Greece, p. 600.
4. N. Otsu, *IEEE Trans. Sys. Man Cyber.* **9**, 62 (1979).
5. N. Ichimura, in *Proc. 15th Int. Conf. Pattern. Recog.*, Sep. 2000, Barcelona, Spain, Vol.3, p. 858.
6. K. Inoue and K. Urahama, in *Proc. 8th Int. Conf. Comput. Vision*, Jul. 2001, Vancouver, Canada, Vol. 1, p. 219.
7. C. Tomasi and T. Kanade, *Int. J. Comput. Vision* **9**, 137 (1992).
8. K. Kanatani, in *Proc. 8th Int. Conf. Comput. Vision*, Jul. 2001, Vancouver, Canada, Vol. 2, p. 301.
9. A. Maki and C. Wiles, in *Proc. 6th Euro. Conf. Comput. Vision*, June–Jul. 2000, Dublin, Ireland, Vol.1, p. 725.
10. A. Maki and H. Hattori, in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, Dec. 2001, Kauai Marriot, Hawaii, to appear.
11. C. J. Poelman and T. Kanade, *IEEE Trans. Patt. Anal. Mach. Intell.* **19**, 206 (1997).
12. K. Kanatani, *Int. J. Comput. Vision* **26**, 171 (1998).

13. K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice* (Elsevier, Amsterdam, 1996).
14. P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection* (Wiley, New York, 1987).
15. C. Matsunaga and K. Kanatani, in *Proc. 6th Euro. Conf. Comput. Vision*, June–Jul. 2000, Dublin, Ireland, Vol. 2, p. 595.
16. K. Kanatani, in *Proc. 5th Asian Conf. Comput. Vision*, Jan. 2002, Melbourne, Australia, Vol. 1, p. xxi.

**Kenichi Kanatani** received his Ph.D. in applied mathematics from the University of Tokyo in 1979. After serving as Professor of computer science at Gunma University, Gunma, Japan, he is currently Professor of information technology at Okayama University, Okayama, Japan. He is the author of *Group-Theoretical Methods in Image Understanding* (Springer, 1990), *Geometric Computation for Machine Vision* (Oxford, 1993) and *Statistical Optimization for Geometric Computation: Theory and Practice* (Elsevier, 1996). He is an IEEE Fellow.