Memoirs of the Faculty of Engineering, Okayama University, Vol. 36, No. 1, pp. 79–90, December 2001

Evaluation and Selection of Models for Motion Segmentation

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

(Received November 9, 2001)

We first present an improvement of Kanatani's *subspace separation* [8] for motion segmentation by newly introducing the *affine space constraint*. We point out that this improvement does not always fare well due to the *effective noise* it introduces. In order to judge which solution to adopt if different segmentations are obtained, we present two criteria: one is the standard F test; the other is model selection using the *geometric AIC* of Kanatani [7] and the *geometric MDL* of Matsunaga and Kanatani [13]. We test these criteria doing real image experiments.

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 seeing their motions without knowing their identities. To solve this problem, Costeira and Kanade [2] presented a segmentation algorithm based on feature tracking. Since then, various extensions have been proposed: Gear [3] used the reduced row echelon form and graph matching; Ichimura [4] applied the discrimination criterion of Otsu [14] and the QR decomposition for feature selection [5]; Inoue and Urahama [6] introduced fuzzy clustering.

Costeira and Kanade [2] attributed their algorithm to the Tomasi-Kanade factorization [17], but the underlying principle is a simple fact of linear algebra: the image motion of points moving rigidly in the scene belongs to a *4-dimensional subspace* [3, 8]. Directly taking advantage of this *subspace constraint*, Kanatani [8] introduced model selection to it and showed that his method, which he called *subspace separation*, far outperforms the Costeira-Kanade algorithm and Ichimura's method. His method has been applied to separating the effect of illumination for moving objects [11, 12].

In reality, the subspace constraint is a *weak* condition: the data points belong to a *3-dimensional affine space* within that 4-dimensional subspace [3]. Using this stronger *affine space constraint* is expected to lead to more accurate segmentation, but all the Costeira-Kanade-type algorithms [2, 3, 4, 5, 6] are unable to exploit this constraint, because they rely on zero/nonzero discrimination of the elements of a matrix computed from the data. In contrast, Kanatani's subspace separation [8] is based on the analysis in the *original data space* rather than particular matrix elements, so it can easily incorporate this constraint.

In this paper, we first present a new segmentation algorithm based on the affine space constraint, which we call *affine space separation*, and demonstrate by simulation that it indeed outperforms the subspace separation. *This is not the whole story, however*.

We next show a rather surprising fact that *the affine space separation does not always fare well in real situations*. We assert that this is because for real data the degree of deviation from the affine space constraint is different from

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

the degree of deviation from the subspace constraint: one may be larger or smaller than the other depending on situations. We clarify this relationship by analysis and simulation.

Then, a question arises: if the subspace separation and the affine space separation produce different solutions for the same image sequence, *which should we believe?* To answer this question, one needs to evaluate the reliability of individual segmentations. Here, we present two criteria: one is the standard F test; the other is model selection using the *geometric AIC* of Kanatani [7] and the *geometric MDL* of Matsunaga and Kanatani [13]. We test these criteria doing real image experiments.

In Sec. 2, we summarize Kanatani's subspace separation [8]. In Sec. 3, we describe our affine space separation and compare its performance with the subspace separation. In Sec. 4, we explain the performance difference in terms of the *effective noise* associated with the constraint. In Sec. 5, we present criteria for a posteriori evaluation of segmentation. In Sec. 6, we test these criteria doing real image experiments. In Sec. 7, we give our conclusion.

2. SUBSPACE SEPARATION

We first summarize Kanatani's *subspace separation* [8], on which our new algorithm is built.

2.1 Subspace Constraint

We track N rigidly moving feature points over M images and let $(x_{\kappa\alpha}, y_{\kappa\alpha})$ be the image coordinates of the αth point in the κth 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 αth 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 κth frame. Let $(a_\alpha, b_\alpha, c_\alpha)$ be the object coordinates of the αth point. Its position in the κth frame with respect to the world coordinate system is given by

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

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 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 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 *4-dimensional* (*linear*) *subspace* spanned by the vectors $\{\boldsymbol{m}_0, \boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3\}$. This *subspace constraint* holds for all *affine camera* models including weak perspective and paraperspective [15], because eq. (2) holds irrespective of the *metric condition* [17, 15] that demands $\{\boldsymbol{i}_\kappa, \boldsymbol{j}_\kappa, \boldsymbol{k}_\kappa\}$ to be orthonormal.

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 4-dimensional subspaces. Inspired by the Tomasi-Kanade factorization

[17], Costeira and Kanade [2] found that this can be done by zero/nonzero discrimination of the elements of a matrix computed from the data. We reiterate this fact according to Kanatani [8].

Let $\{\boldsymbol{p}_\alpha\}$ be N points that belong to an r-dimensional subspace $\mathcal{L} \subset \mathcal{R}^n$. Define an $N \times N$ 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 an $N \times N$ 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 ith set $\{\boldsymbol{p}_\alpha\}$, $\alpha \in \mathcal{I}_i$. If the m subspaces $\mathcal{L}_i$, $i = 1, ..., m$, are linearly independent, the following holds (see [8] for the formal proof):

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

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

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 [2] proposed this type of strategy, known as the *greedy algorithm*.

Whatever realigning strategy is taken, however, all such algorithms [2, 3, 4, 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, and finding a suitable threshold is difficult even if the noise is known to be Gaussian with a known variance [3]. To avoid this difficulty, Kanatani [8] worked in the *original data space*, rather than the matrix elements derived from it, using model selection and robust fitting. Doing synthetic and real-image experiments, he demonstrated that his method far outperforms the Costeira-Kanade algorithm [2] and Ichimura's method [4].

3. AFFINE SPACE SEPARATION

We now present a new segmentation algorithm based on a stronger constraint.

3.1 Affine Space Constraint

We have overlooked a crucial fact about eq. (4): the coefficient of $\boldsymbol{m}_0$ is *identically 1*, implying that the N points $\{\boldsymbol{p}_\alpha\}$ should belong to a *3-dimensional affine space* within the 4-dimensional subspace. This *affine space constraint* has long been known [3] but so far has not been utilized in any way. This is because no theorem corresponding to Theorem 1 is available for this constraint. Hence, as long as the segmentation is based on zero/nonzero discrimination of matrix elements, there is no way to exploit this constraint.

However, it can immediately be incorporated into Kanatani's subspace separation, in which model selection and robust fitting are introduced in the original data space. We now replace the subspace constraint involved there by the affine space constraint; we call the resulting algorithm the *affine space separation*. Doing simulation, we will demonstrate that it indeed outperforms the subspace separation.

3.2 Affine Space Merging

Initially regarding individual points as groups consisting of one element, we successively merge two groups by fitting a 3-dimensional affine space to them. Let $\mathcal{A}_i$ and $\mathcal{A}_j$ be candidate affine spaces to merge, and let N_i and N_j be the respective numbers of points in them.

Let J_i^A and J_j^A be the individual residuals, i.e., the sums of the squared distances of the data points to the fitted affine spaces $\mathcal{A}_i$ and $\mathcal{A}_j$. Let $J_{i\oplus j}^A$ be the residual that would result if a single affine space is fitted to the $N_i + N_j$ points. It is reasonable not to merge the two groups if $J_{i\oplus j}^A$ is much larger than $J_i^A + J_j^A$. But how large should $J_{i\oplus j}^A$ be? In fact, we always have $J_{i\oplus j}^A \geq J_i^A + J_j^A$ because a single affine space has fewer degrees of freedom to adjust than two affine spaces. It follows that we must balance the increase in the residual against the decrease in the degree of freedom. For this purpose, Kanatani used his *geometric AIC* [7]. His method is translated into the affine space constraint as follows.

We assume that the tracked feature points are perturbed from their true positions by independent Gaussian noise of mean zero and standard deviation ϵ, which we call the *noise level*. Since a 3-dimensional affine space has $4(n-3)$ degrees of freedom[1], the geometric AIC has the following form [7]:

$$\text{G-AIC}_{i\oplus j}^A = J_{i\oplus j}^A + 2\Big(3(N_i + N_j) + 4(n-3)\Big)\epsilon^2. \tag{8}$$

If two 3-dimensional affine spaces are fitted to the N_i points and the N_j points separately, the degree of freedom is the sum of those for individual affine spaces. Hence, the geometric AIC is given as follows [7]:

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

Merging $\mathcal{A}_i$ and $\mathcal{A}_j$ is reasonable if $\text{G-AIC}_{i\oplus j}^A < \text{G-AIC}_{i,j}^A$. However, this criterion can work only for $N_i + N_j > 4$. Also, all the affine spaces are included in subspaces of higher dimensions, so the matrix $\boldsymbol{Q}$ still provides information about the possibility of merging. Following Kanatani [8], we integrate these two criteria to define the following similarity measure between the affine spaces $\mathcal{A}_i$ and $\mathcal{A}_j$:

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

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

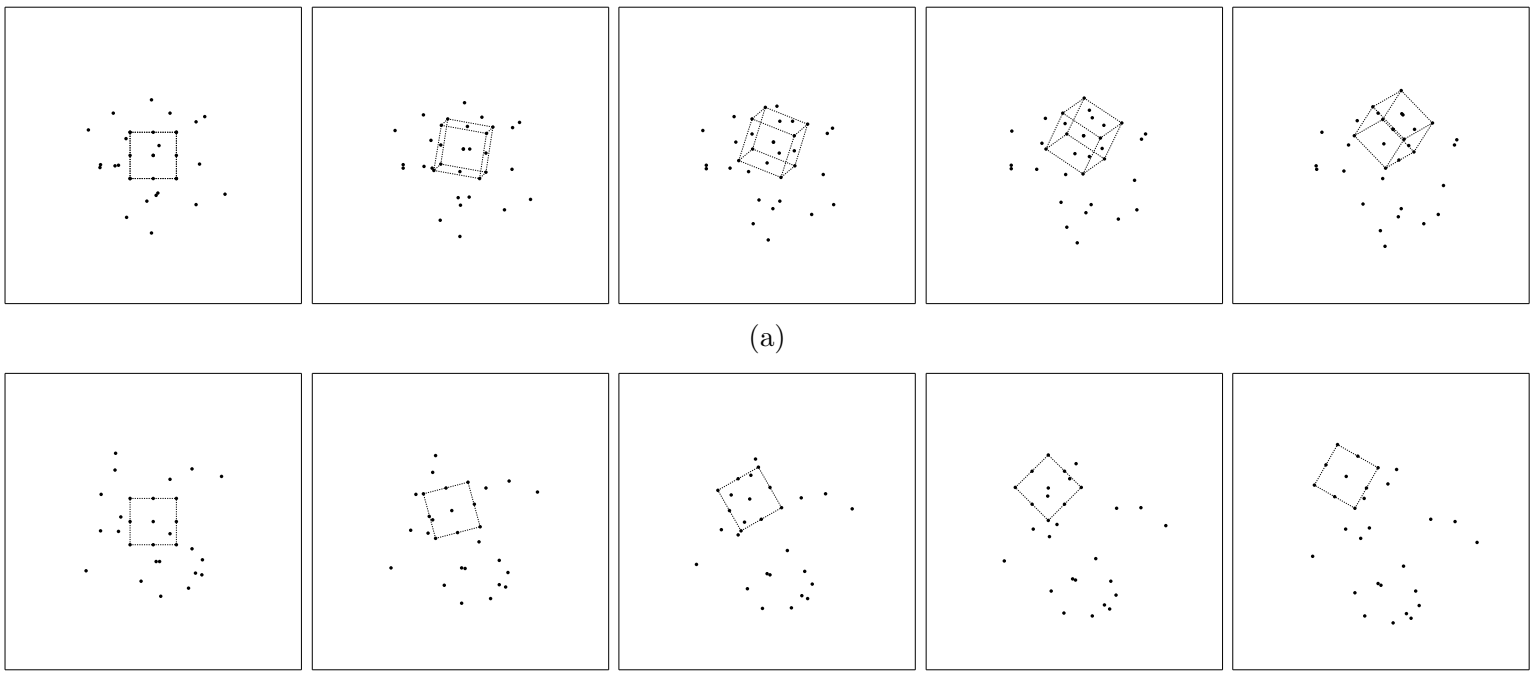
For evaluating the geometric AIC, we need to estimate the noise level ϵ. This can be done if we note that the points $\{\boldsymbol{p}_\alpha\}$ should belong to a $(4m-1)$-dimensional affine space in $\mathcal{R}^n$ in the absence of noise. Let J_t^A be the residual after fitting a $(4m-1)$-dimensional affine space to $\{\boldsymbol{p}_\alpha\}$. Then, J_t^A/ϵ^2 is subject to a χ^2 distribution with $(n-4m+1)(N-4m)$ degrees of freedom [7], so we obtain the following unbiased estimator of ϵ^2:

$$\hat{\epsilon}_A^2 = \frac{J_t^A}{(n-4m+1)(N-4m)}. \tag{11}$$

3.3 Dimension Correction and Robust Fitting

We also incorporate two techniques which were proved to be very effective in the subspace separation [8]. The first is *dimension correction*: as soon as more than four elements are grouped together, we optimally fit a 3-dimensional affine space to them and replace the points with their projections onto the fitted affine space for computing the matrix $\boldsymbol{Q}$. This effectively reduces the noise in the data if the local grouping is correct.

[1] It is specified by four points in $\mathcal{R}^n$, but they can move within that affine space into three directions. So, the degree of freedom is $4n - 4 \times 3$.



(a)

(b)

Figure 1: Image sequences of points in (a) 3-dimensional motion and (b) 2-dimensional motion.

The second technique is an a posteriori reallocation. Since a point once misclassified in the course of merging never leaves that class, we attempt to remove outliers from the m resulting classes $\mathcal{A}_1$, ..., $\mathcal{A}_m$ by robust fitting. Points near the origin may be easily misclassified, so we select from each class $\mathcal{A}_i$ half (but not less than four) of the elements that have large norms. We fit 3-dimensional affine spaces $\mathcal{A}'_1$, ..., $\mathcal{A}'_m$ to them again and select from each class $\mathcal{A}_i$ half (but not less than four) of the elements whose distances to the closest affine space $\mathcal{A}'_j$, $j \neq i$, are large. We fit 3-dimensional affine spaces $\mathcal{A}''_1$, ..., $\mathcal{A}''_m$ to them again and allocate each data point to the closest one. Finally, we fit 3-dimensional affine spaces $\mathcal{A}'''_1$, ..., $\mathcal{A}'''_m$ to the resulting point sets by *LMedS* [16]. Each data point is reallocated to the closest one.

3.4 Experiments

Fig. 1(a) is a sequence of five consecutive simulated images (512×512 pixels) of 20 points in the background and 14 points in an object; they are independently moving in three dimensions. Fig. 2(b) is a sequence of five consecutive simulated images (512×512 pixels) of 20 points in the background and 5 points in an object; they are independently moving in two dimensions. In both sequences, the object is given a wireframe for the ease of visualization.

We added Gaussian noise of mean 0 and standard deviation ϵ (pixels) to the coordinates of all the points and classified them into two groups. If the motion is 2-dimensional, the vector $\boldsymbol{m}_3$ in eq. (4) can be taken to be identically zero, so the N points $\{\boldsymbol{p}_\alpha\}$ should belong to a *2-dimensional affine space* within the 3-dimensional subspace spanned by $\{\boldsymbol{m}_0, \boldsymbol{m}_1, \boldsymbol{m}_2\}$. The algorithm described in the preceding sections can easily be tailored to this case (we omit the details).

Fig. 2 plots the average ratio of misclassified points over 500 independent trials for different ϵ for the 3-dimensional motion (a) and the 2-dimensional motion (b): we compared 1. the greedy algorithm, 2. Ichimura's method [4], 3. the subspace separation, and 4. our affine space separation. As we see, the subspace separation indeed outperforms both the greedy algorithm and Ichimura's method, but the affine space separation further improves the accuracy. The improvement is particularly remarkable for the 3-dimensional motion; for the 2-dimensional motion the subspace separation already gives a nearly ideal solution.

Thus, one is tempted to conclude that the affine space separation is better than the subspace separation. We now assert that *this is not necessarily so*.

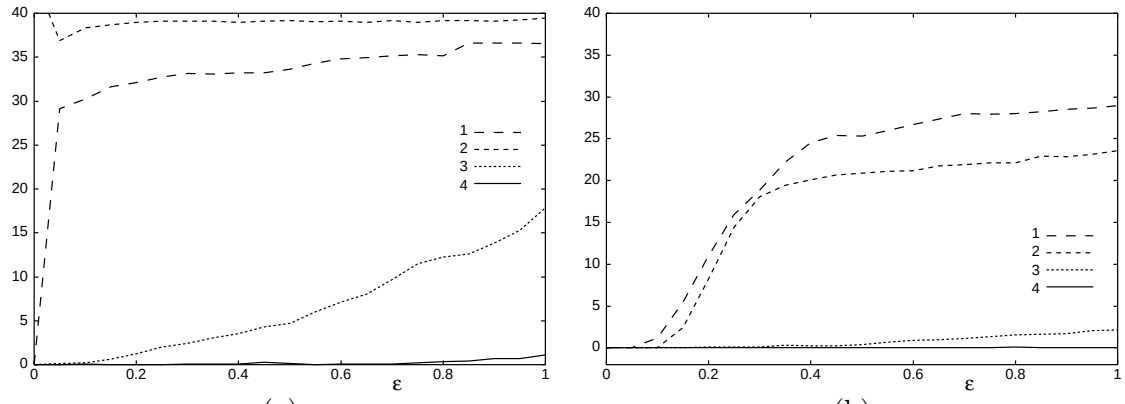


Figure 2: Misclassification ratio of segmentation vs. the noise level ϵ (pixels) for the 3-D motion of Fig. 1(a) and the 2-D motion of Fig. 1(b): 1. greedy algorithm; 2. Ichimura's method; 3. subspace separation; 4. affine space separation.

4. EFFECTIVE NOISE

4.1 Effective Noise Level

The performance gain of the affine space separation is brought about by using a stronger constraint. In general, the stronger the constraint, the better the inference based on it *provided it strictly holds*. However, any constraint is an idealization of the phenomenon, and slight deviations from it are regarded as "noise", which we call *effective noise* to distinguish it from other sources of data inaccuracy (e.g., insufficient image resolution). Generally, the stronger the constraint, *the less strictly it holds*. In other words, the effective noise for a stronger constraint is usually larger than the effective noise for a weaker constraint.

Suppose we have correctly segmented N points into m groups by subspace separation, and let J_1^S, ..., J_m^S be the residuals of fitting d-dimensional subspaces to individual groups, where $d = 3$ or 4 depending on whether the motion is 2-dimensional or 3-dimensional. The effective noise level ϵ_S is defined in such a way that random noise of that magnitude *would* produce these residuals. Using the same logic for deriving eq. (11), we obtain

$$\epsilon_S^2 = \frac{\sum_{i=1}^m J_i^S}{(n-d)(N-d)}. \tag{12}$$

Similarly, if J_1^A, ..., J_m^A are the residuals of fitting $(d-1)$-dimensional affine spaces to individual groups, the effective noise level ϵ_A is estimated from

$$\epsilon_A^2 = \frac{\sum_{i=1}^m J_i^A}{(n-d+1)(N-d)}. \tag{13}$$

4.2 Perspective Effects

Since both the subspace constraint and the affine space constraint are based on the *affine camera model* of eq. (4), the effective noise is not zero for a perspective camera even when no image noise exists. Moreover, it should be different for the subspace constraint and for the affine space constraint, since their strengths are different. To confirm this, we added perspective effects to the image sequence of Fig. 1(a) by changing the angle of view α with the field of view fixed. Fig. 3 shows the changes of the third frame of Fig. 1(a) for $\alpha = 0°$, 40°, and 80° ($\alpha = 0°$ corresponds to orthographic projection).

Fig. 4(a) plots the effective noise level vs. the angle of view α for the affine space constraint (solid line) and the subspace constraint (dashed line). We can see that the presence of perspective effects is equivalent to much larger noise for the affine space constraint than the subspace constraint, which is relatively insensitive to the angle of view. This is because the affine space constraint depends more heavily on the affine camera model than the subspace constraint.

Fig. 4(b) shows the average misclassification ratio of the affine space separation and the subspace separation when Gaussian noise of standard deviation ϵ is added to image coordinates for the angles of view $\alpha = 0°$, 40°,

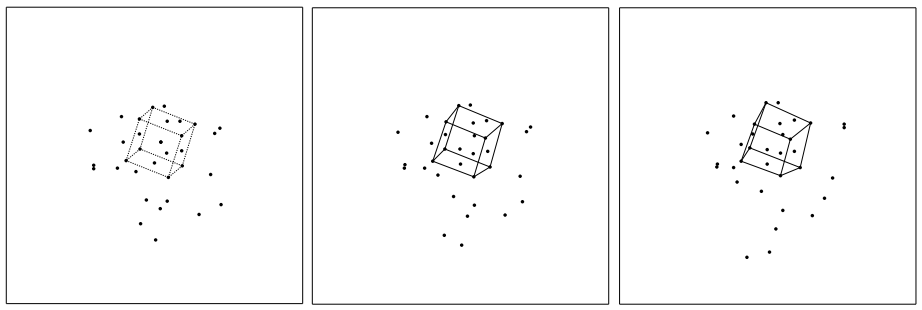


Figure 3: The perspective effects of the third frame of Fig. 1(a) for the angles of view $\alpha = 0°$ (orthographic projection), 40°, and 80°.

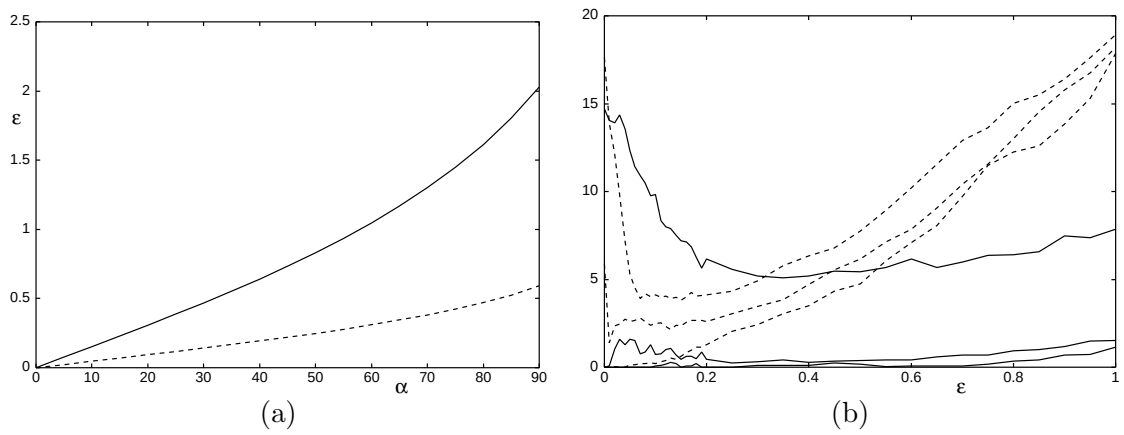


Figure 4: (a) Effective noise level vs. the angle of view α (degrees) for the affine space constraint (solid line) and the subspace constraint (dashed line). (b) Misclassification ratio of segmentation vs. the noise level ϵ (pixels) for affine space separation (solid lines) and subspace separation (dashed lines). For each, the angles of view are $\alpha = 0°$, 40°, and 80° from bottom to top.

and 80°. The result is quite different from Fig. 2(a), because in the presence of the perspective effects the noise level ϵ on the horizontal axis of Fig. 2(a) should be read to be the *sum* of the effective noise and the random noise. For the affine space constraint, the effective noise is already high even when random noise is small, so the misclassification ratio is very high as compared with the subspace constraint. As random noise increases, however, the misclassification ratio of the subspace separation grows more rapidly than the affine space separation, because the latter has a higher capability of canceling random noise.

Thus, we are compelled to conclude that *neither* the affine space separation nor the subspace separation is universally superior: their performance depends on the *balance* between the perspective effects and the data accuracy.

5. A POSTERIORI EVALUATION OF SEGMENTATOPM

If neither of the two methods is always better than the other, and if the two methods produce *different* solutions, which should we believe? This is a very difficult yet very important question in practice. We now present two criteria for evaluating the reliability of individual segmentations.

5.1 F Test for Segmentation

Suppose the N points $\{\boldsymbol{p}_\alpha\}$ are segmented into m groups having N_i points, $i = 1, ..., m$. We first consider the subspace separation.

Let J_i^S be the residual of fitting a d-dimensional subspace $\mathcal{L}_i$ to the ith group. The subspace $\mathcal{L}_i$ has codimension $n - d$ and $d(n - d)$ degrees of freedom. Hence, J_i^S/ϵ^2 should be subject to a χ^2 distribution with

$$\phi_i^S = (n - d)N_i - d(n - d) = (n - d)(N_i - d) \tag{14}$$

degrees of freedom [7].

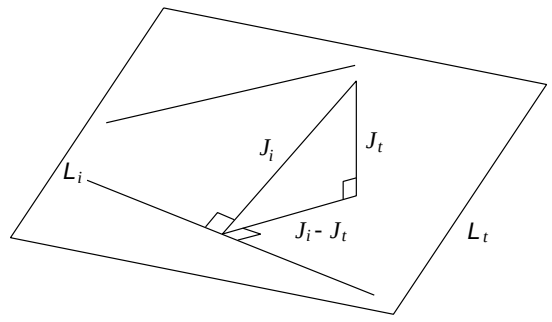


Figure 5: The residual of subspace fitting.

If we let J_t^S be the residual of fitting an md-dimensional subspace $\mathcal{L}_t$ to the entire N points $\{\boldsymbol{p}_\alpha\}$, J_t^S/ϵ^2 should be subject to a χ^2 distribution with

$$\phi_t^S = \sum_{i=1}^{m}(n-md)N_i - md(n-md) = (n-md)(N-md) \tag{15}$$

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

The residual J_i^S is the sum of squared distances of the points of the ith 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. 5). 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 $\sum_{i=1}^m J_i^S - J_t^S$; the sum of the squared external distances is J_t^S.

If this segmentation is correct (the *null hypothesis*), $(\sum_{i=1}^m J_i^S - J_t^S)/\epsilon^2$ should also be subject to a χ^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, $(\sum_{i=1}^m J_i^S - J_t^S)/\epsilon^2$ has

$$\sum_{i=1}^{m}\phi_i^S - \phi_t^S = (m-1)d(N-md) \tag{16}$$

degrees of freedom [7]. It follows that

$$F^S = \frac{(\sum_{i=1}^m J_i^S - J_t^S)/(m-1)d(N-md)}{J_t^S/(n-md)(N-md)} \tag{17}$$

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

$$F_{(n-md)(N-md)}^{(m-1)d(N-md)}(\alpha) < F^S, \tag{18}$$

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

Next, we consider the affine space separation. Let J_i^A be the residual of fitting a $(d-1)$-dimensional affine space $\mathcal{A}_i$ to the ith group. The affine space $\mathcal{A}_i$ has codimension $n-d+1$ and $d(n-d+1)$ degrees of freedom. Hence, J_i^A/ϵ^2 should be subject to a χ^2 distribution with

$$\phi_i^A = (n-d+1)N_i - d(n-d+1) = (n-d+)(N_i-d) \tag{19}$$

degrees of freedom. If we let J_t^A be the residual of fitting an $(md-1)$-dimensional affine space $\mathcal{A}_t$ to the entire N points $\{\boldsymbol{p}_\alpha\}$, J_t^A/ϵ^2 should be subject to a χ^2 distribution with

$$\phi_t^A = \sum_{i=1}^{m}(n-md+1)N_i - md(n-md+1) = (n-md+1)(N-md) \tag{20}$$

degrees of freedom irrespective of the correctness of the segmentation. If this segmentation is correct, $(\sum_{i=1}^m J_i^A - J_t^A)/\epsilon^2$ should also be subject to a χ^2 distribution with

$$\sum_{i=1}^{m}\phi_i^A - \phi_t^A = (m-1)d(N-md) \tag{21}$$

degrees of freedom. It follows that

$$F^A = \frac{(\sum_{i=1}^m J_i^A - J_t^A)/(m-1)d(N-md)}{J_t^A/(n-md+1)(N-md)} \tag{22}$$

should be subject to an F distribution with $(m-1)d(N-md)$ and $(n-md+1)(N-md)$ degrees of freedom. Hence, if

$$F_{(n-md)(N-md)}^{(m-1)d(N-md)}(\alpha) < F^A, \tag{23}$$

this segmentation is rejected with significance level α.

5.2 Model Selection for Segmentation

The result of an F test depends on the significance level, which we can arbitrarily set. In contrast, it is known that model selection can dispense with any thresholds. Here, we apply the *geometric AIC* of Kanatani [7] and a modification [9, 10] of the *geometric MDL* of Matsunaga and Kanatani [13].

We first consider the subspace separation. The geometric AIC and the geometric MDL for the model that the segmentation is correct are, respectively, $\sum_{i=1}^m \text{G-AIC}_i^S$ and $\sum_{i=1}^m \text{G-MDL}_i^S$, where G-AIC_i^S and G-MDL_i^S are the geometric AIC and the geometric MDL of the ith group, which are given as follows:

$$\begin{aligned} \text{G-AIC}_i^S &= J_i^S + 2\Big(dN_i + d(n-d)\Big)\epsilon^2, \\ \text{G-MDL}_i^S &= J_i^S - \Big(dN_i + d(n-d)\Big)\epsilon^2 \log\Big(\frac{\epsilon}{L}\Big)^2. \end{aligned} \tag{24}$$

Here, L is a reference length, for which we can use an arbitrary value whose order is approximately the same as the data, say the image size; the model selection result is not affected as long as it has the same order of magnitude [9, 10].

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

$$\begin{aligned} \text{G-AIC}_t^S &= J_t^S + 2\Big(mdN_i + md(n-md)\Big)\epsilon^2, \\ \text{G-MDL}_t^S &= J_t^S - \Big(mdN_i + md(n-md)\Big)\epsilon^2 \log\Big(\frac{\epsilon}{L}\Big)^2. \end{aligned} \tag{25}$$

Since the latter model is correct irrespective of the segmentation result, we can estimate the noise level ϵ from it, using

$$\hat{\epsilon}^2 = \frac{J_t^S}{(n-md)(N-md)}. \tag{26}$$

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

$$2 < F^S, \qquad -\log\Big(\frac{\epsilon}{L}\Big)^2 < F^S, \tag{27}$$

where F^S is the F statistic given by eq. (17). Thus, model selection using the geometric AIC and the geometric MDL reduces to the F test, the difference being that the threshold is given automatically without specifying any significance level. When the noise is small, $-\log(\epsilon/L)^2$ is usually larger than 2, so the geometric AIC is more conservative than the geometric MDL, which is more confident of the particular result.

Next, we consider the affine space separation. The geometric AIC and the geometric MDL of the ith group are

$$\begin{aligned} \text{G-AIC}_i^A &= J_i^A + 2\Big((d-1)N_i + d(n-d+1)\Big)\epsilon^2, \\ \text{G-MDL}_i^A &= J_i^A - \Big((d-1)N_i + d(n-d+1)\Big)\epsilon^2 \log\Big(\frac{\epsilon}{L}\Big)^2. \end{aligned} \tag{28}$$

The geometric AIC and the geometric MDL for the general model are

$$\begin{aligned} \text{G-AIC}_t^A &= J_t^A + 2\Big((md-1)N_i + md(n-md+1)\Big)\epsilon^2, \\ \text{G-MDL}_t^A &= J_t^A - \Big((md-1)N_i + md(n-md+1)\Big)\epsilon^2 \log\Big(\frac{\epsilon}{L}\Big)^2. \end{aligned} \tag{29}$$

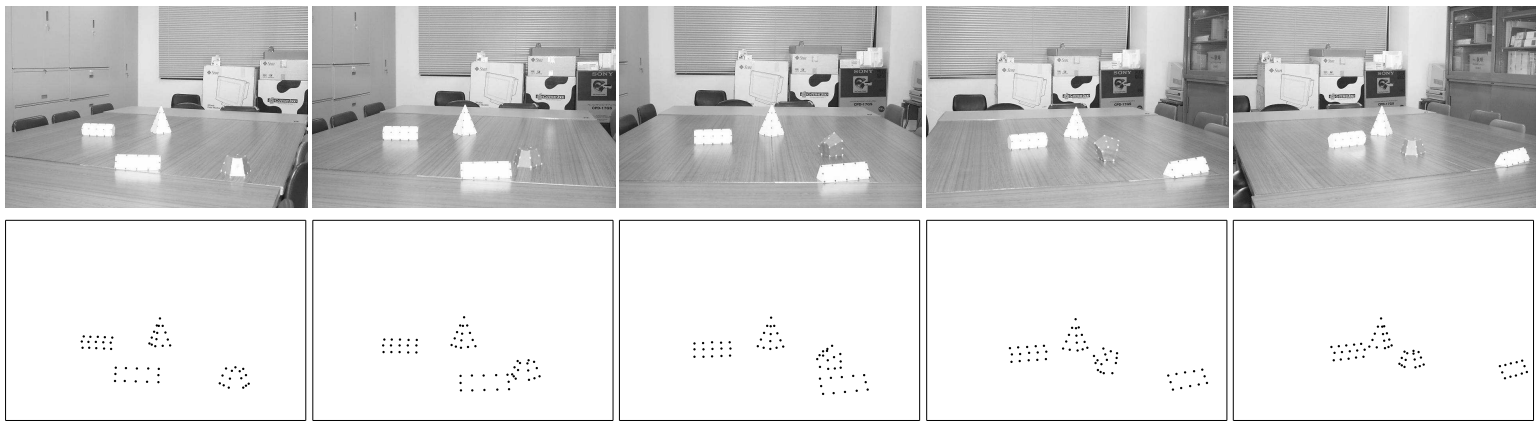


Figure 6: Real images of moving objects (above) and the selected feature points (below).

The noise level ϵ is estimated from the general model, using

$$\hat{\epsilon}^2 = \frac{J_t^A}{(n - md + 1)(N - md)}. \tag{30}$$

The condition that the segmentation is not correct is $\text{G-AIC}_t^A < \sum_{i=1}^m \text{G-AIC}_i^A$ or $\text{G-MDL}_t^A < \sum_{i=1}^m \text{G-MDL}_i^A$, which reduces to

$$2 < F^A, \qquad -\log\Big(\frac{\epsilon}{L}\Big)^2 < F^A, \tag{31}$$

where F^A is the F statistic given by eq. (22).

6. REAL IMAGE EXPERIMENTS

Fig. 6 shows a sequence of perspectively projected images (above) and manually selected feature points from them (below). Three objects are fixed in the scene, moving rigidly with the scene, while one object is moving relative to the scene. The image size is 512×768 pixels.

We applied the greedy algorithm and Ichimura's method [4] and found that the resulting segmentations contained some errors. However, the subspace separation and the affine space separation both resulted in the correct segmentation. We then evaluated the reliability of this segmentation and observed the following:

1. The effective noise levels computed from eqs. (12) and (13) give $\epsilon_S = 0.64$ (pixels) and $\epsilon_A = 0.94$ (pixels), confirming our prediction that the affine space separation has larger effective noise than the subspace separation.
2. The F statistics of eqs. (17) and (22) are $F^S = 0.893$ and $F^A = 1.483$. The corresponding upper 5% percentiles are 1.346 and 1.293, respectively, so with significance level 5% the subspace separation is not rejected but the affine space separation is rejected.
3. From eqs. (27) and eqs. (31), the geometric AIC selects both segmentations as correct. The subspace separation passes the test with a larger margin than the affine subspace separation.
4. Letting $L = 600$, we have $-\log(\hat{\epsilon}/L) = 13.5$, 13.1 for the subspace separation and the affine space separation, respectively, so the geometric MDL again selects both segmentations as correct. The geometric MDL accepts the segmentation result much more leniently than the geometric AIC.

All these tests indicate that for *this* image sequence the subspace separation is more reliable than the affine space separation. In other words, although both segmentations happen to be correct for the *current* occurrence of noise, the subspace separation has a larger chance of being correct than the affine space separation *if noise occurred differently*.

To test this prediction, we did a *bootstrap* experiment [1]: we added random Gaussian noise of mean 0 and standard deviation 1 (pixel) to the coordinates of the detected feature points in Fig. 6 and did segmentation 500

times, each time using different noise. We found that the subspace separation was correct 100% while the affine space separation was correct only for 90.9% of the trials. We then increased the standard deviation to 2 (pixels) and found that the subspace separation was still correct 100% while the affine space separation was correct only 85.3%.

Thus, we have confirmed that if the two methods produced different segmentations we would have to believe the subspace segmentation *for this image sequence*, for which the perspective effect is relatively strong while the feature detection is relatively accurate (since we did manual selection).

This type of a posteriori reliability evaluation can be done automatically for any segmentation results.

7. CONCLUDING REMARKS

In this paper, we first presented an improvement, in theory, of Kanatani's subspace separation [8] for motion segmentation by incorporating the affine space constraint and demonstrated by simulation that it outperforms the subspace separation approach.

Next, we pointed out that this gain in performance is due to the stronger constraint based on the affine camera model and that this introduces effective noise that accounts for the deviation from the model. We have confirmed this by simulation.

Then, we presented two criteria for evaluating the reliability of a given segmentation: one is the standard F test; the other is model selection using the geometric AIC and the geometric MDL. Doing real image experiments, we have demonstrated that these criteria enable us to judge which solution to adopt if the two methods result in different segmentations.

We have observed that the affine space separation is more accurate than the subspace separation when the perspective effects are weak and the noise in the data is large, while the subspace separation is more suited for images with strong perspective effects with small noise. For images of 2-dimensional (or almost 2-dimensional) motions, the affine space separation is very powerful, since there are no (or small) perspective effects. In practice, we should apply both methods, and if the results are different, we should select the more reliable one indicated by our proposed criteria.

In this paper, we have assumed that the number of independent motions is known (basically 2 for the background and a moving object). There have been a number of attempts to estimate it [2, 3, 4, 5, 6], but all tried to estimate the *rank* of the matrix $\boldsymbol{Q}$, for which the noise in all the elements is mutually correlated in a nonlinear way, without considering the underlying linear/affine structure. The estimation should be done in the *original data space*, where we can assume the noise in the feature coordinates to be independent and identical Gaussian and take advantage of the linear/affine structure. Again, model selection provides a power tool for this (see [10] for the results).

Acknowledgements

The author thanks Noriyoshi Kurosawa of Fsas Network Solutions Inc., Japan, for doing numerical simulations and real image experiments. He also thanks Naoya Ohta of Gunma University, Japan, Atsuto Maki of Toshiba Co., Japan, and Naoyuki Ichimura of AIST, Japan, for helpful comments. 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).

References

[1] B. Efron and R. J. Tibshirani, *An Introduction to Bootstrap*, Chapman-Hall, New York, 1993.

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

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

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

[5] N. Ichimura, Motion segmentation using feature selection and subspace method based on shape space, *Proc. 15th Int. Conf. Pattern. Recog.*, September 2000, Barcelona, Spain, Vol.3, pp. 858–864

[6] K. Inoue and K. Urahama, Separation of multiple objects in motion images by clustering, *Proc. 8th Int. Conf. Comput. Vision*, July 2001, Vancouver, Canada, Vol. 1, pp. 219–224.

[7] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier Science, Amsterdam, The Netherlands, 1996.

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

[9] K. Kanatani, Model selection for geometric inference, plenary talk, *Proc. the 5th Asian Conf. Comput. Vision*, January 2002, Melbourne, Australia, to appear.

[10] K. Kanatani and C. Matsunaga, Estimating the number of independent motions for multibody segmentation, *Proc. the 5th Asian Conf. Comput. Vision*, January 2002, Melbourne, Australia, to appear.

[11] 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.

[12] A. Maki and H. Hattori, Illumination subspace for multibody motion segmentation, *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, December 2001, Kuai Marriot, Hawaii, to appear.

[13] C. Matsunaga and K. Kanatani, Calibration of a moving camera using a planar pattern: Optimal computation, reliability evaluation and stabilization by model selection, *Proc. 6th Euro. Conf. Comput. Vision*, June–July, 2000, Dublin, Ireland, Vol. 2, pp. 595–609.

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

[15] 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.

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

[17] 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.