A Robust and Efficient Motion Segmentation Based on Orthogonal Projection Matrix of Shape Space

Naoyuki ICHIMURA

Electrotechnical Laboratory
1-1-4, Umezono, Tsukuba
Ibaraki, 305-8568 Japan
ichimura@etl.go.jp

Abstract

A novel algorithm for motion segmentation is proposed. The algorithm uses the fact that shape of an object with homogeneous motion is represented as 4 dimensional linear space. Thus motion segmentation is done as the decomposition of shape space of multiple objects into a set of 4 dimensional subspace. The decomposition is realized using the discriminant analysis of orthogonal projection matrix of shape space. Since only discriminant analysis of 1D data is needed, this analysis is quite simple. The algorithm based on the analysis is robust for data with noise and outliers, because the analysis can extract useful information for motion segmentation while rejecting useless one. The implementation results show that the proposed method is robust and efficient enough to do online task for real scenes.

1 Introduction

Motion segmentation is essential to analysis of motions using an image sequence for shape from motion, video coding, and motion based control of active camera.

Motion segmentation algorithms proposed thus far is classified into three categories. The first category uses sampling and verification [1] [2]: some feature correspondences are randomly sampled from image pair and then they are verified whether epipolar constraints for single motion are satisfied. The second one is optimization based on probability distribution [3]-[5], e.g. Bayesian estimation. The algorithms in these two categories have the same problem, the chicken and egg problem, because the former incorporates spatial proximity into random sampling to increase the possibility for extracting feature correspondences that belong to the same group and the latter requires appropriate initial guesses of motion information; these facts mean that information about "each group" is necessary in advance to obtain information used in segmentation.

The third category uses simultaneous calculation of motions to avoid the chicken and egg problem [6]-[9]. In particular, factorization based method [7]-[9] has attracted attention. Basically, orthogonal projection matrix of the linear space that represents shape of objects, that is *shape space*, has been used for segmentation in the method: the basis of shape space can be computed from the SVD of measurement matrix with the coordinates of feature correspondences as entities. The clustering for the basis of shape space that can be regarded as optimization has been used in [7]. In [8], the orthogonal projection matrix was called the shape interaction matrix, and optimization based on the energy represented by the sum of entities of it has been used in segmentation. Optimization using bipartite graph which was made based on shape space has been used in [9]. These optimizations were needed to solve the combinatorial problem in segmentation. They, however, have a common problem: the lack of robustness for feature correspondences with noise and outliers [10]; these methods were difficult to apply for practical situation.

A concept of feature selection using a discriminant criterion is introduced to give robustness and efficiency for motion segmentation based on orthogonal projection matrix of shape space. Feature selection is carried out by selecting only one row of orthogonal projection matrix corresponding to one feature point extracted in image. The discriminant analysis for the entities in each row of orthogonal projection matrix is used to select the feature with the most useful information for segmentation. A group is extracted based on the result of discriminant analysis for the selected feature. The same procedure is applied recursively to remaining features to extract other groups. Since features with no useful information are rejected automatically, the algorithm is robust against noise and outliers. No combinatorial problem occurs in grouping, since only one row of orthogonal projection matrix is used. Additionally, no prior knowledge is needed on the number of objects because groups are extracted recursively.

Section 2 shows the derivation and the properties of measurement matrix under N objects case which we reformulate. Section 3 describes the definition and property of orthogonal projection matrix of shape space. Section 4 presents the proposed motion segmentation algorithm based on feature selection using discriminant criterion. Section 5 demonstrates the experimental results of off-line and on-line implementations. Section 6 shows conclusions.

2 Measurement Matrix Obtained from Multiple Moving Objects

The measurement matrix obtained from multiple moving objects under affine projection is derived in this section. The properties of the measurement matrix which are useful for segmentation are shown from the results of the derivation. In the following sections, N is the number of objects, P^i is the number of features of the i-th object, $P = \sum_{i=1}^{N} P^i$, and F is the number of frames.

2.1 Derivation of Measurement Matrix

The 3D coordinates of the feature of the i-th object in world coordinates is represented by matrix $\boldsymbol{D}_s{}^i$ $(i = 1, \ldots, N)$ as follows:

$$\underset{4\times P^i}{\boldsymbol{D}_s{}^i} = \begin{pmatrix} x_1^i & x_2^i & \ldots & x_{P^i}^i \\ y_1^i & y_2^i & \ldots & y_{P^i}^i \\ z_1^i & z_2^i & \ldots & z_{P^i}^i \\ 1 & 1 & \ldots & 1 \end{pmatrix} \tag{1}$$

The motion parameters of the i-th object in the j-th frame is represented by $\boldsymbol{M}_j^i$ $(j = 1, \ldots, F)$. This matrix consists of rotation matrix $\boldsymbol{R}_j^i$ and translation vector $\boldsymbol{t}_j^i$.

$$\underset{4\times 4}{\boldsymbol{M}_j^i} = \begin{pmatrix} \boldsymbol{R}_j^i & \boldsymbol{t}_j^i \\ 0 \quad 0 \quad 0 & 1 \end{pmatrix} \tag{2}$$

The motion parameters of the i-th object through all frames are represented by matrix $\boldsymbol{M}^i$ as follows:

$$\underset{4F\times 4}{\boldsymbol{M}^i} = \left(\boldsymbol{M}_1^{i\,t}, \boldsymbol{M}_2^{i\,t}, \ldots, \boldsymbol{M}_F^{i\,t}\right)^t \tag{3}$$

The 3D coordinates of features in each frame are collected in matrix $\boldsymbol{W}_w$.

$$\underset{4F\times P}{\boldsymbol{W}_w} = \underset{4F\times 4N}{\boldsymbol{M}_w} \underset{4N\times P}{\boldsymbol{D}_s} \tag{4}$$

where

$$\underset{4N\times P}{\boldsymbol{D}_s} = \mathrm{diag}\left(\boldsymbol{D}_s^1, \boldsymbol{D}_s^2, \ldots, \boldsymbol{D}_s^N\right) \tag{5}$$

$$\underset{4F\times 4N}{\boldsymbol{M}_w} = \left(\boldsymbol{M}^1, \boldsymbol{M}^2, \ldots, \boldsymbol{M}^N\right) \tag{6}$$

The coordinates of features measured in image plane are the projection of above 3D coordinates. The following affine camera model is used in this projection.

$$\lambda\tilde{\boldsymbol{m}} = \boldsymbol{P}_a^{'}\tilde{\boldsymbol{X}} \tag{7}$$

$$\boldsymbol{P}_a^{'} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & p_{34} \end{pmatrix} \tag{8}$$

where $\tilde{\boldsymbol{m}}$ and $\tilde{\boldsymbol{X}}$ are the coordinates of features in image plane and 3D space represented by homogeneous coordinates. If $\tilde{\boldsymbol{m}}$ and $\tilde{\boldsymbol{X}}$ do not get the point at infinity, Eq.(7) can be expressed as follows:

$$\boldsymbol{m} = \boldsymbol{P}_a\left(\boldsymbol{X}^t, 1\right)^t \tag{9}$$

$$\boldsymbol{P}_a = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \end{pmatrix} \tag{10}$$

where $\boldsymbol{m}$ and $\boldsymbol{X}$ are the Euclidean coordinates whose representations in homogeneous coordinates are $\tilde{\boldsymbol{m}}$ and $\tilde{\boldsymbol{X}}$.

The feature correspondences are obtained by applying the projection matrix of each frame $\boldsymbol{P}_a^i$ $(i = 1, \ldots, F)$ to the 3D coordinates of Eq.(4) as follows:

$$\begin{aligned} \underset{2F\times P}{\boldsymbol{W}_s} &= \underset{2F\times 4F}{\boldsymbol{A}_p} \underset{4F\times 4N}{\boldsymbol{M}_w} \underset{4N\times P}{\boldsymbol{D}_s} \\ &= \underset{2F\times 4N}{\boldsymbol{M}} \underset{4N\times P}{\boldsymbol{D}_s} \end{aligned} \tag{11}$$

$$\boldsymbol{M} = \boldsymbol{A}_p\boldsymbol{M}_w \tag{12}$$

$$\boldsymbol{A}_p = \mathrm{diag}\left(\boldsymbol{P}_a^1, \boldsymbol{P}_a^2, \ldots, \boldsymbol{P}_a^F\right) \tag{13}$$

$$2F, P \geq 4N \tag{14}$$

Segmentation information is contained in the matrix $\boldsymbol{W}_s$: the first P^1 columns are the feature correspondences of object 1, and the next P^2 columns are the feature correspondences of object 2, etc. The segmentation information, however, is not given a priori. Thus the columns of $\boldsymbol{W}_s$ corresponding to the features of objects are permutated. For example, this permutation depends on the method of scan on image to measure the coordinates of features. The permutation can be represented by the permutation matrix $\boldsymbol{\Pi}_{col}$. The permutated matrix is represented as follows:

$$\boldsymbol{W} = \boldsymbol{W}_s\boldsymbol{\Pi}_{col} \tag{15}$$

The matrix $\boldsymbol{W}$ is called a measurement matrix.

2.2 Properties of Measurement Matrix

The properties of the measurement matrix derived from Eq.(11) are as follows:
(i) The measurement matrix can be decomposed into two matrices $\boldsymbol{M}$ and $\boldsymbol{D}_s$ which are contained motion and shape information.
(ii) The rank of the measurement matrix is $4N$ for N objects case.

These are fundamental properties for motion segmentation using measurement matrix.

3 Orthogonal Projection Matrix of Shape Space

To utilize the properties of measurement matrix shown in 2.2 for motion segmentation, decomposition of measurement matrix is considered. Given measurement matrix, we can decompose it using the SVD as follows:

$$\underset{2F\times P}{\boldsymbol{W}} = \underset{2F\times r}{\boldsymbol{U}_r} \underset{r\times r}{\boldsymbol{\Sigma}_r} \underset{r\times P}{\boldsymbol{V}_r^t} \tag{16}$$

where $\boldsymbol{U}_r$ and $\boldsymbol{V}_r$ are the orthogonal basis of column and row space of measurement matrix and $r = 1, \ldots, \min(2F, P)$. The diagonal matrix $\boldsymbol{\Sigma}_r$ consists of singular values. If r is identical to the number of nonzero singular values, it is the rank of the measurement matrix.

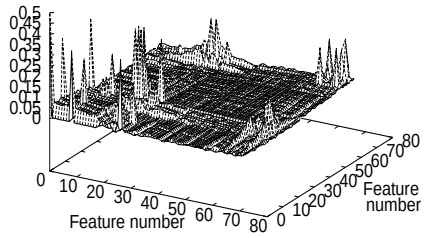
From the first property shown in 2.2, we can see that the row space of measurement matrix, a linear space spaned by the orthogonal basis in matrix $\boldsymbol{V}$ in Eq.(16), represents the shape of objects. Thus the row space is called *shape space* in this paper.

The dimension of shape space is $4N$ and shape space can be decomposed into a set of 4 dimensional subspace corresponding to one object. These facts can be understood from the second property shown in 2.2. Motion segmentation, therefore, can be carried out through the decomposition of shape space.

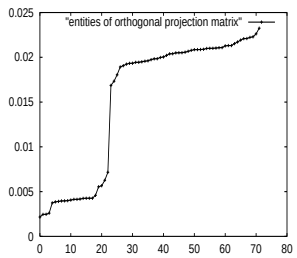
The following orthogonal projection matrix of shape space is utilized in the decomposition:

$$\underset{P\times P}{\boldsymbol{X}} = \boldsymbol{V}_r\boldsymbol{V}_r^t = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_P)^t \tag{17}$$

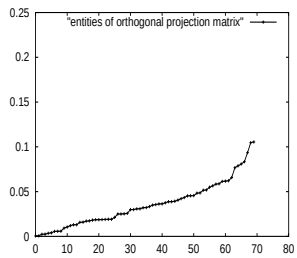
The size of this matrix is $P\times P$; both columns and rows of this matrix correspond to P features extracted in image. If data are not contaminated by noise, entities x_{ij} of $\boldsymbol{X}$



(a)



(b)



(c)

Figure 1: An example of data with information useful for segmentation and without. (a) Orthogonal projection matrix obtained from real image sequence. (b) Change in the sorted entities of a row with useful information. (c) Change in the sorted entities of a row without useful information.

have the following property [8] [12]:

$$x_{ij} \begin{cases} \neq 0, & \text{If features corresponding to the } i\text{-th row and } j\text{-th column belong to the same object.} \\ = 0, & \text{If features corresponding to the } i\text{-th row and } j\text{-th column belong to a different object.} \end{cases} \tag{18}$$

Using this property, we can easily find the group of features corresponding to homogeneous motions. The property, however, is not exactly satisfied for data with noise and outliers. The lack of robustness for such data is a problem for the conventional methods using orthogonal projection matrix [7]-[10].

4 Robust Motion Segmentation Based on Feature Selection

In this section, an algorithm based on feature selection is proposed to give a robustness for motion segmentation using orthogonal projection matrix.

4.1 Concept of Feature Selection

In the proposed method, only a feature with the most useful information for segmentation is selected. This feature selection reduces the effect of noise and outliers automatically, and eliminates the combinatorial problem in segmentation.

The usefulness of feature selection can be explained using an example shown in Fig.1. Two graphs, Fig.1 (b) and (c), represent the change in the sorted entities of rows in the orthogonal projection matrix shown in Fig.1 (a) obtained from real image sequence. The entities of

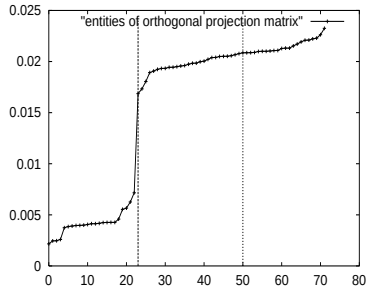


Figure 2: An example of calculation of discriminant criterion.

Fig.1 (b) have more useful information for segmentation than one of Fig.1 (c), because the entities of Fig.1 (b) can easily be divided into two groups; they can easily be divided into the entities regarded as zero and one regarded as nonzero. As this example, only part of the features may have information useful for segmentation. Thus, the proposed method selects one row in orthogonal projection matrix with entities as shown in Fig.1 (b) which corresponds to a feature with the most useful information for segmentation, while rejects features with no useful information as shown in Fig.1 (c).

4.2 Feature Selection Using Discriminant Criterion

The problem is how to select a row with the most useful information for segmentation. The discriminant criterion can be used in selection. Given r in Eq.(16), orthogonal projection matrix $\boldsymbol{X}$ is calculated. Entities of each row $\boldsymbol{x}_k$ of $\boldsymbol{X}$ $(k = 1, \ldots, P)$ are then sorted. The following discriminant criterion[13] is used to separate entities of row vector $\boldsymbol{x}_k$ into two groups:

$$\lambda = \frac{\sigma_B^2}{\sigma_W^2} \tag{19}$$

$$\sigma_B^2 = N^1 N^2 \left(\bar{\varepsilon_1} - \bar{\varepsilon_2}\right)^2 \tag{20}$$

$$\sigma_W^2 = N^1 \sigma_1^2 + N^2 \sigma_2^2 \tag{21}$$

where $\bar{\varepsilon_1}$, $\bar{\varepsilon_2}$, σ_1^2, σ_2^2, N^1 and N^2 are the mean, variance, and the number of entities of each group, σ_B^2 and σ_W^2 are the variance between groups and one within groups. σ_B^2 shows the difference between means of two groups and σ_W^2 shows the homogeneity of entities in each group.

To find the grouping that maximizes λ, x_{kl} $(l = 1, \ldots, P)$ are used as a threshold for entities of $\boldsymbol{x}_k = \{x_{k1}, x_{k2}, \ldots, x_{kP}\}$. For example, the entities of a row shown in Fig.2 are divided into two groups using line at 50 of abscissa, the discriminant criterion is small, because the variance of entities in left group, i.e., σ_W^2 is large. On the other hand, if the entities are divided using line at 23, the discriminant criterion of this case is larger than the case of line at 50, because the variance between groups is large while one within groups is small. We can, therefore, find the appropriate boundary between two groups automatically using discriminant criterion; we do not need threshold to find it.

For Fig.1 (b), discriminant criterion is large because these is a large gap at 23 of abscissa, while only small

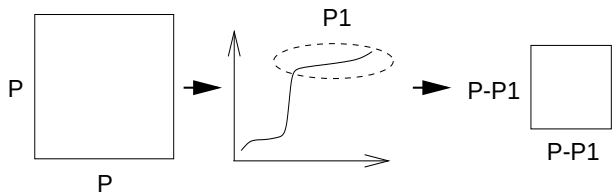


Figure 3: Recursive procedure for segmentation.

discriminant criterion can be obtained for Fig.1 (c) because these is no large gap. Thus a row with the most useful information for segmentation can be selected by finding the row with maximum discriminant criterion. Let λ_k be the discriminant criterion used in grouping for the k-th row vector, e.g. the discriminant criterion at 23 in Fig.2. The feature with the most useful information for segmentation is selected as follows:

$$k_{select} = \arg\max_k \lambda_k \tag{22}$$

After feature selection, we can extract one group from the results of calculation of discriminant criterion of the k_{select}-th row (left and middle figure in Fig.3). In the next step, the matrix corresponding to the remaining features is extracted from orthogonal projection matrix (right figure in Fig.3) and discriminant criteria for rows of this matrix are calculated to extract another group. The same procedure is carried out recursively to extract other groups until the discriminant criterion of selected feature is smaller than the given threshold.

$$\lambda_{k_{select}} < \lambda_{threshold} \tag{23}$$

This equation means that there is no useful information for segmentation in the entities of orthogonal projection matrix.

The parameter r in Eq.(16) is the rank of the measurement matrix if it can be estimated correctly from the singular values obtained by the SVD. The rank, however, is difficult to estimate for feature correspondences with noise and outliers. In this research, parameter r is determined from the view on separation. The recursive algorithm explained above is applied using parameter r in some range, e.g., [3:12]. The segmentation result maximizing the following equation, sum of $\lambda_{k_{select}}$ of extracted g groups, is used as a final result:

$$\lambda(r) = \sum_{i=1}^{g} \lambda_{k_{select}}(i, r) \tag{24}$$

This equation means that the rank corresponding to the segmentation result with maximum separation of entities of orthogonal projection matrix is used.

4.3 Features of the Proposed Method

The features of the proposed method are summarized as follows:
(i) Robustness against noise and outliers due to feature selection; the data without useful information for segmentation are automatically rejected.
(ii) No combinatorial problem occurs in segmentation; the groups of features can be extracted using exhaustive calculation of discriminant criteria because only less than P combinations are contained in one row of orthogonal projection matrix.
(iii) Simple and stable numerical computation; only the SVD and the calculation of discriminant criteria are needed.
(iv) No need for prior knowledge on the number of objects; groups are extracted by the recursive procedure.

These features are useful in practical situation as shown in the experimental results.

5 Experimental Results

The results of off-line and on-line implementations are shown. The purpose of off-line implementation is to confirm the effect of outlier rejection, and the one of on-line implementation is to show the robustness and efficiency in real situations.

The threshold of Eq.(23) was 5.0 and the parameter r in Eq.(16) was changed in range [3:12] for all experiments. The feature points in images were extracted using the corner detector proposed in [14]. The matching of feature points was carried out using normalized correlation.

5.1 Off-line Implementation

For the car sequence (Fig.4), 10 frames were used. Both the camera and car were moved; the background was not still. The two groups extracted corresponded to the car and background (Fig.5 (a), (b)) and the rest corresponded to the features points with inhomogeneous motion due to matching errors (Fig.5 (c)); the proposed method could reject the outliers with no useful information while detecting the dominant motions.

For the human walk sequence (Fig.6), 10 frames were used. The camera was fixed. The two groups extracted corresponded to the human and background (Fig.7 (a), (b)). Matching errors and overlap between feature points on background and human led to the inhomogeneous motions (Fig.7 (c)). The proposed method could detect such small difference of motions and thus extract the homogeneous and dominant motions from the data with outliers.

These results of off-line implementation show the effect of outlier rejection of the proposed method.

5.2 On-line Implementation

The parallel processing by two personal computers with Pentium II 450 MHz (PC1) and Pentium III 500 MHz (PC2) was used in the implementation. Tracking of feature points was continuously carried out on PC1: the image processing board with hardware for calculation of normalized correlation was used in matching of feature points. Motion segmentation was done on PC2 every 30 frames. The measurement matrix used in motion segmentation was sent to PC2 from PC1 and the result of motion segmentation was returned to PC1 to give the group number for each feature point. The Parallel Virtual Machine (PVM) library[15] was used in the communication between two PCs. Using this parallel processing, we could segment motions without interruption of tracking process.

Two results are shown in Fig.8 and Fig.9. The rectangle in these figures shows the position of the center of gravity of feature points regarded as a group with motion larger than 10 pixels.

For the human head sequence (Fig.8), the motion of the head was continuously tracked and segmented over

800 frames (Fig.8(a)-(f)), although the feature correspondences had many matching errors, particularly in homogeneous regions such as the white wall.

For the two books sequence (Fig.9), the proposed method could segment the motions correctly, although the two books were separated (Fig.9(a)-(c)) and combined (Fig.9(d)-(f)), and the feature correspondences contained many outliers due to matching errors and overlap between moving regions and background. This result shows that prior knowledge on the number of objects is not needed in the proposed method. The motions were tracked and segmented over 2300 frames.

The computational time of motion segmentation and the frame rate of tracking were about 300 [ms] and 10 [frame/sec] for about 100 feature points.

These results of on-line implementation show the robustness and the efficiency of the proposed method.

6 Conclusions

A motion segmentation algorithm based on orthogonal projection matrix of shape space has been proposed. The proposed method has robustness and efficiency due to feature selection based on discriminant criterion. The usefulness of the proposed method in practical situations was confirmed by the experiments of off-line and on-line implementations using real image sequences.

The proposed method requires approximation by an affine camera model because it is based on the measurement matrix shown in Eq.(11). The development of an algorithm under perspective projection is a problem for future work.

Acknowledgments

The author would like to thank Dr.Nobuyuki Otsu, the director of the Electrotechnical Laboratory, for his encouragement.

References

[1] E. Nishimura, G. Xu and S. Tsuji: "Motion segmentation and correspondence using epipolar constraint," Proc. ACCV, pp.199-204, 1993

[2] P. H. S. Torr: "Geometric motion segmentation and model selection," Phil. Trans. R. Soc. Lond. A, Vol.356, pp.1321-1340, 1998

[3] D. W. Murray and B. F. Buxton: "Scene segmentation from visual motion using global optimization," IEEE Trans. Pattern Anal. & Mach. Intell., Vol.9, No.2, pp.220-228, 1987

[4] J. Konrad and E. Dubois: "Bayesian estimation of motion vector fields," IEEE Trans. Pattern Anal. & Mach. Intell., Vol.14, No.9, pp.910-927, 1992

[5] Y. Weiss and E. H. Adelson: "A unified mixture framework for motion segmentation: Incorporating spatial coherence and estimating the number of models," Proc. CVPR, pp.321-326l, 1996

[6] M. Shizawa: "Transparent 3D motions and structures from point correspondences in two frames: A quasi-optimal, closed-form, linear algorithm and degeneracy analysis," Proc. First Asian Conference on Computer Vision, pp.329-334, 1993

[7] T. E. Boult and L. G. Brown: "Factorization-based segmentation of motions," Proc. IEEE Workshop on Vis. Mot., pp.179-186, 1991

[8] J. P. Costeira and T. Kanade: "A multi-body factorization method for independently moving objects," Internat. J. Comp. Vis., 29, 3, pp.159-179, 1998

[9] C. W. Gear: "Multibody grouping from motion images," Internat. J. Comp. Vis., 29, 2, pp.133-150, 1998

[10] T. Kanade and D. D. Morris: "Factorization methods for structure from motion," Phil. Trans. R. Soc. Lond. A, 356, pp.1153-1173, 1998

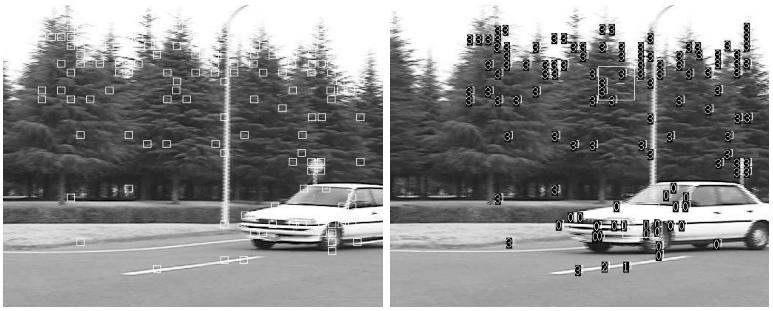
[11] C. Tomasi and T.Kanade: "Shape and motion from image streams under orthography: a factorization method," Internat. J. Comp. Vis., 9, 2, pp.137-154, 1992

[12] K. Kanatani: "Factorization without factorization: multibody segmentation," Technical Report of IEICE, PRMU98-117, 1998 (in Japanese)

[13] N. Otsu: "A threshold selection method from gray-level histograms," IEEE Trans. Sys., Man, and Cybern., Vol.SMC-9, No.1, pp.62-66, 1979

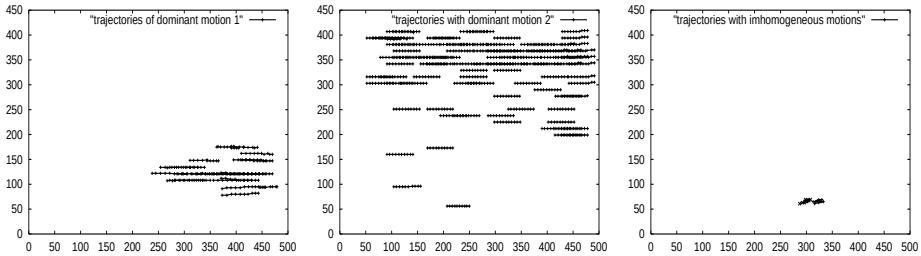
[14] F. Chabat, G.Z. Yang and D.M. Hansell:"A corner orientation detector," Im. and Vis. Comp., 17, pp.761-769, 1999

[15] http://www.epm.ornl.gov/pvm/pvm_home.html



(a) (b)

Figure 4: Car image sequence. (a) Frame 1. (b) Frame 10. The small rectangles show the extracted feature points.

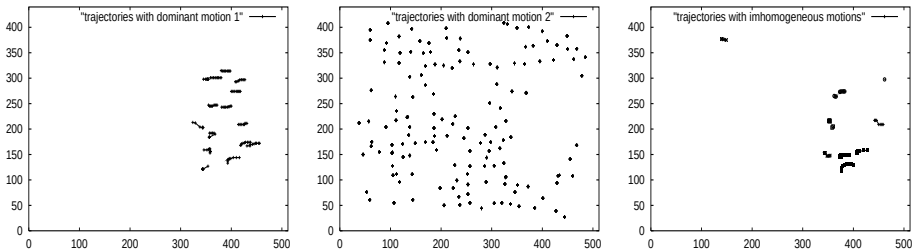


(a) (b) (c)

Figure 5: Segmentation result for a car image sequence. (a),(b) Trajectories of feature points for groups with dominant motions. (c) Trajectories of feature points regarded as outliers.

(a) (b)

Figure 6: Human walk image sequence. (a) Frame 1. (b) Frame 10.



(a) (b) (c)

Figure 7: Segmentation result for a human walk image sequence. (a),(b) Trajectories of feature points for groups with dominant motions. (c) Trajectories of feature points regarded as outliers.

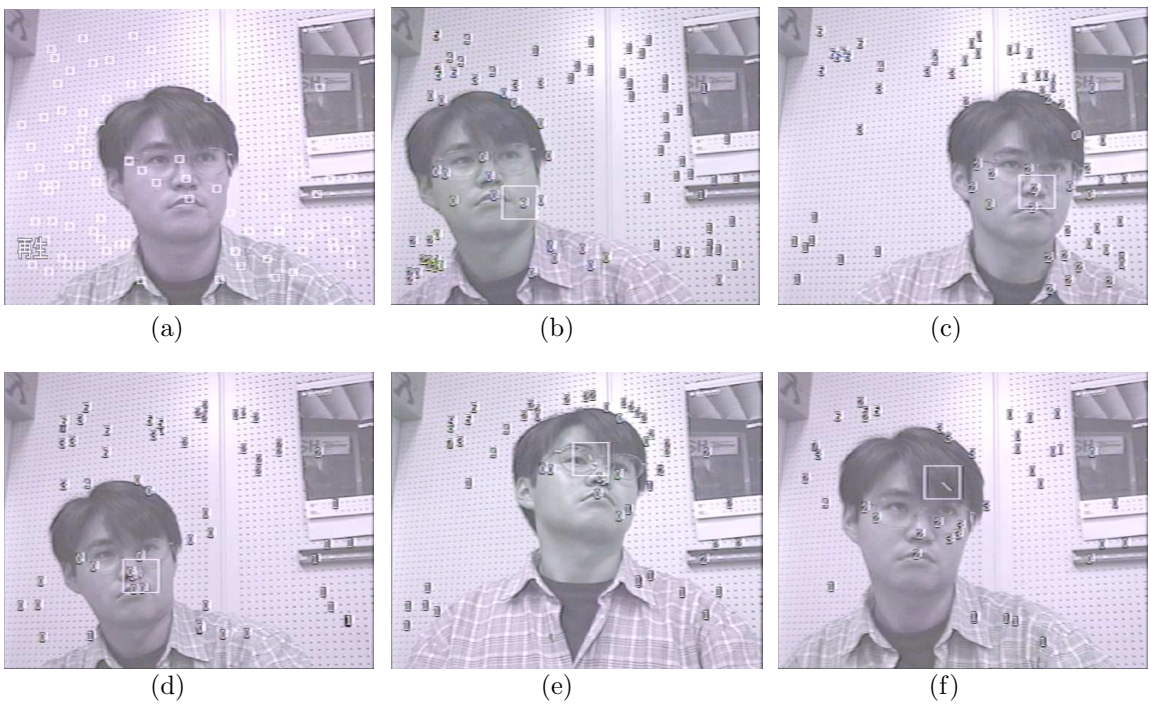


Figure 8: Segmentation result for on-line image sequence of human head. (a) 0 sec. (b) 5 sec. (c) 14 sec. (d) 18 sec. (e) 23 sec. (f) 27 sec. The rectangle shows the position of the center of gravity of feature points regarded as a group with motion larger than 10 pixels.

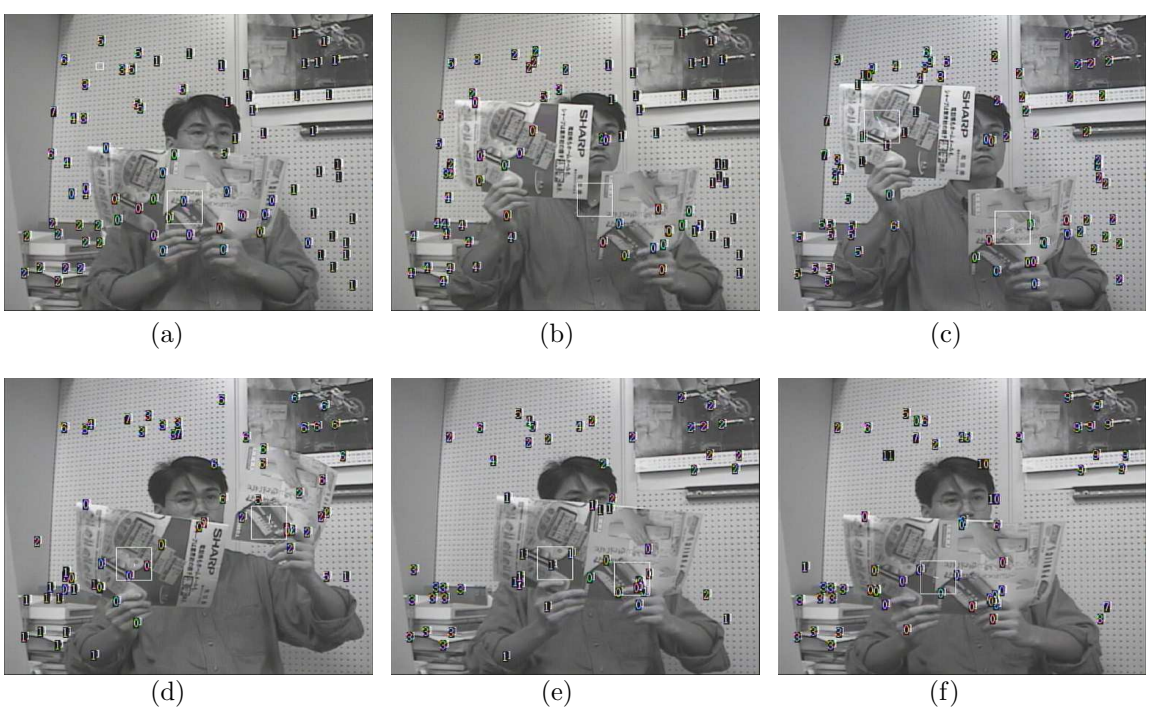


Figure 9: Segmentation result for on-line image sequence of two books. (a) 9 sec. (b) 27 sec. (c) 28 sec. (d) 41 sec. (e) 75 sec. (f) 77 sec. Two books were separated ((b),(c)) and combined ((e),(f)).