Motion Segmentation Using Feature Selection and Subspace Method Based on Shape Space

Naoyuki ICHIMURA

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

Abstract

Motion segmentation using feature correspondences can be regarded as a combinatorial problem. A motion segmentation algorithm using feature selection and subspace method is proposed to solve the combinatorial problem. Feature selection is carried out as computation of a basis of the linear space that represents the shape of objects. Features can be selected from "each" object "without segmentation information" by keeping the correspondence of basis vectors to features. Only four or less features of each object are used; the combination in segmentation is reduced by feature selection. Thus the combinatorial problem can be solved without optimization. The remaining features in selection are classified using the subspace method based on the segmentation result of selected features. Experiments are done to consider the usefulness of the proposed method.

1 Introduction

Segmentation is fundamental processing in computer vision. Motion segmentation has attracted great attention, because it can be used for applications such as modeling by shape from motion, video coding, and the analysis of movement. Many algorithms have been proposed for motion segmentation based on Hough transformation [1], clustering [2], random fields model [3], etc. Although motion parameters are used in these methods, segmentation information is needed to estimate the motion parameters correctly. This situation leads to a chicken and egg problem.

Factorization method [4]-[6] has been used to avoid the chicken and egg problem. A measurement matrix with the coordinates of feature correspondences as entities is decomposed into two matrices only once. Initial segmentation is not needed, since these matrices contain motion and shape information of multiple objects.

The algorithm proposed in the present paper is inspired by the work of Costeira and Kanade [6]. A *shape interaction matrix* which is a kind of proximity matrix of motions is used for segmentation. If objects are segmented correctly, shape interaction matrix becomes block-diagonal. Hence the optimization method that minimizes the total energy of all possible off-diagonal blocks of shape interaction matrix was used. If the number of objects and the number of features are large, however, the exhaustive search of all possible off-diagonal blocks causes a combinatorial explosion. A nonlinear optimization has been applied in such case, but it may involve local minima and high computational cost.

A motion segmentation algorithm using feature selection and subspace method is proposed in this paper. Only four or less features per object are selected to reduce the number of the combinations in segmentation. Thus stable and simple numerical computation can be used to solve the combinatorial problem; the nonlinear optimization is not needed. Classification of the remaining features in selection is done by the subspace method based on the segmentation result of selected features.

The problem in feature selection is how to select features from "each" object "without segmentation information". Feature selection is systematically carried out as computation of the basis of shape space which is the linear space that represents the shape of objects. Features can be selected from each object without segmentation information by keeping the correspondence of basis vectors to features. That is, the chicken and egg problem is also avoided in feature selection; this is an important advantage of our method.

Section 2 shows the summary of shape interaction matrix and the problems of the conventional method. Section 3 describes the proposed method. Section 4 presents experimental results. Section 5 shows conclusions.

2 Shape Interaction Matrix and Segmentation

The summary of shape interaction matrix[6] is shown. The features, e.g., points, in image sequence are tracked to obtain feature correspondence. Feature correspondences obtained from P features and F frames are collected in measurement matrix $\boldsymbol{W}$ $(2F \times P)$. Under the affine projection model, measurement matrix is decomposed by the SVD [7].

$$\underset{2F\times P}{\boldsymbol{W}} = \underset{2F\times r}{\boldsymbol{U}_r}\ \underset{r\times r}{\boldsymbol{\Sigma}_r}\ \underset{r\times P}{\boldsymbol{V}_r^t} \tag{1}$$

where $\boldsymbol{\Sigma}_r$ is diagonal matrix with singular values and r is the rank of measurement matrix. For full-3D case where the shapes of objects are not degenerate, $r = 4N$ for N objects; the feature correspondences of one object are represented by four dimensional subspace corresponding to submatrix of $\boldsymbol{W}$ whose rank is four. This condition is broken in shape degenerate case where planes and lines are included in scene.

The orthogonal basis in $\boldsymbol{U}_r$ and $\boldsymbol{V}_r$ contain motion and shape information of objects[7]. Thus the linear space spaned by the basis in $\boldsymbol{V}_r$ represents the shape of objects; the *shape space* is spaned by $\boldsymbol{V}_r$. The number of dimension of shape space for multiple objects is equal to the rank of measurement matrix r and one for single object is four or less: 3D shape is represented by four dimensional shape space, plane is represented by three dimensional shape space, etc.

A shape interaction matrix is defined as follows:

$$\underset{P\times P}{\boldsymbol{X}} = \boldsymbol{V}_r\boldsymbol{V}_r^t = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_P)^t \tag{2}$$

The size of this matrix is $P \times P$; both columns and rows correspond to P features. The entities x_{ij} of $\boldsymbol{X}$ have the following property:

$$x_{ij}\begin{cases} \neq 0, & \text{If features corresponding to the } i\text{-th row and } j\text{-th column belong to the same object.} \\ = 0, & \text{If features corresponding to the } i\text{-th row and } j\text{-th column belong to a different object.} \end{cases} \tag{3}$$

From above property, shape interaction matrix is block-diagonalized under correct segmentation result. Thus Costeira and Kanade used the optimization method that minimizes the total energy of all possible off-diagonal blocks over all sets of the permutations of the rows and the columns of shape interaction matrix for segmentation[6].

The conventional method has following problems.

(i) the number of combinations in segmentation

Before segmentation, the number of features of each object, that is, the sizes of block matrices in shape interaction matrix corresponding to segmentation result are unknown (Fig.1). Thus the optimization method that minimizes the total energy of all possible off-diagonal blocks over all sets of the permutations is needed.

For N objects and P features, the number of combinations for dividing P features into N subsets is $\frac{1}{N!}\sum_{k=1}^{N}(-1)^{N-k}\,{}_N C_k k^P$. For $N = 4$ and $P = 1230^*$, the number of combinations is approximately 2×10^{740}. The exhaustive search of this combination is very difficult, or may be impossible. A nonlinear optimization can be applied in such case. However, it may involve local minima and high computational cost.

(ii) difficulty in degenerate analysis

*This is a condition of the experiments shown in section 4.

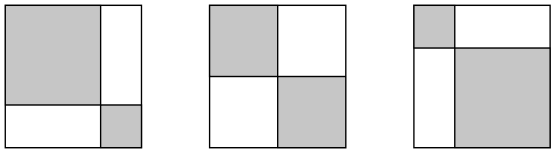


Figure 1: The shape interaction matrix corresponding to segmentation result. The sizes of the block matrices in the shape interaction matrix vary with the number of features of each object as shown above.

The method using shape interaction matrix can be used to segment the degenerate shape such as plane and line. However shape interaction matrix does not give the information about degenerate shape; there is no way to know which block matrix corresponds to degenerate shape without extra processing such as the SVD. An exceptional treatment should be used to reconstruct degenerate shape. Thus the algorithm that can discriminate degenerate object is useful.

3 Segmentation Using Feature Selection and Subspace Method

3.1 Overview of Proposed Algorithm

To solve the above-mentioned problem, the segmentation algorithm using feature selection and subspace method is proposed. Figure 2 shows its overview.

In the first step, measurement matrix is decomposed using the SVD. The rank of measurement matrix r can be estimated using singular values. The second step is an essential part: r features are selected using shape space. The computation of basis of shape space using the QR decomposition can select features from "each" object "without segmentation information". Thus there is no the chicken and egg problem in feature selection. In the third step, the selected features are segmented. The problem on combination does not occur, since only r features are used. Estimation of the number of objects and discrimination of degenerate shape are simultaneously carried out in this segmentation. In the fourth step, the remaining $P - r$ features are classified using subspace method based on the segmentation result of selected features.

The details of each step are described in the subsequent sections.

3.2 Feature Selection Based on Basis of Shape Space

Since the rank of measurement matrix is r, only r features are linear independent. Thus r features can be selected from P ones to eliminate the redundancy, e.g. four features are selected for each object under full-3D case.

For N objects, r should be decomposed as follows:

$$r = r_1 + r_2 + \ldots + r_N, \qquad 1 \leq r_i \leq 4 \qquad (i = 1, \ldots, N) \tag{4}$$

Singular value decomposition and estimation of rank of measurement matrix

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

$\Downarrow$

Feature selection based on basis of shape space

$$\underset{r\times P}{\boldsymbol{V}_r^t}\,\underset{P\times P}{\boldsymbol{\Pi}} = \underset{r\times r}{\boldsymbol{Q}}\,\underset{r\times P}{\boldsymbol{R}}$$

$\Downarrow$

Segmentation of selected features, estimation of the number of objects, and discrimination of degenerate shape

$\Downarrow$

Classification of remaining features

Figure 2: The overview of the proposed algorithm.

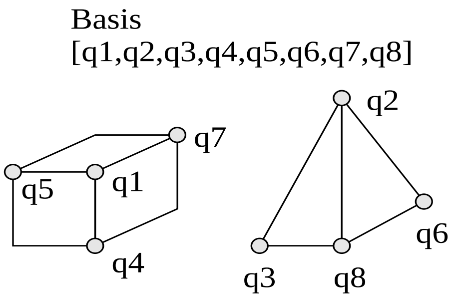


Figure 3: The correspondence of basis vectors to features. Each basis vector of shape space corresponds to a feature of object. The basis vectors are assigned to objects without redundancy. Thus the features needed to represent the shape of objects can be selected through the computation of the basis.

where r_i is the number of dimensions of the shape space of each object. That is, r_i features should be selected from the i-th object.

The problem in feature selection is how to select r_i features from the i-th object without segmentation information. We propose feature selection through the computation of the basis of shape space; the basis of shape space is computed in the manner that each basis vector corresponds to a feature of object (Fig.3).

The basis vectors of shape space are assigned to each object without redundancy. Hence r features selected by the proposed feature selection can be decomposed as shown in Eq.(4). Because, if $k > r_i$ vectors correspond to the i-th object, the representation of the i-th object is redundant and one of another object is insufficient; these vectors are not proper to represent shape space of multiple objects. This contradicts the condition of basis. The proposed feature selection based on the computation of basis, therefore, can select r_i features from the i-th object without segmentation information.

To keep the correspondence of basis vectors to features, basis vectors should be computed from r columns of $\boldsymbol{V}_r^t$ corresponding to r features. This procedure can be realized by the QR decomposition[8].

The QR decomposition is represented as follows:

$$\underset{r\times P}{\boldsymbol{V}_r^t}\,\underset{P\times P}{\boldsymbol{\Pi}} = \underset{r\times r}{\boldsymbol{Q}}\,\underset{r\times P}{\boldsymbol{R}} \tag{5}$$

where $\boldsymbol{Q}$, $\boldsymbol{\Pi}$ and $\boldsymbol{R}$ are the matrices whose columns are the orthonormal basis of the columns of $\boldsymbol{V}_r^t$, the permutation matrix of the columns of $\boldsymbol{V}_r^t$, and upper triangular matrix. Since the columns of the matrix $\boldsymbol{Q}$ are the orthonormal basis of the columns of $\boldsymbol{V}_r^t$, these are the basis vectors of shape space: the redundancy of P dimensional basis vectors in $\boldsymbol{V}_r$ is eliminated in $\boldsymbol{Q}$.

The following is an example of the permutation of the columns of $\boldsymbol{V}_r^t$ by $\boldsymbol{\Pi}$.

$$\begin{aligned} \boldsymbol{V}_r^t &= (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4, \boldsymbol{v}_5, \ldots) \\ & \quad\downarrow \\ \boldsymbol{V}_r^t\boldsymbol{\Pi} &= (\boldsymbol{v}_{140}, \boldsymbol{v}_{11}, \boldsymbol{v}_{30}, \boldsymbol{v}_{259}, \boldsymbol{v}_{182}, \ldots) \end{aligned}$$

The first five basis vectors in $\boldsymbol{Q}$ are derived from the column vectors in $\boldsymbol{V}_r^t$ corresponding to the 140th, 11th, 30th, 259th and 182th features. That is, the correspondence between the basis vectors and features is kept by the permutation matrix $\boldsymbol{\Pi}$. Thus we can see that r features corresponding to the permutated columns are the set of r_i features obtained from each object.

3.3 Segmentation of Selected Features

The selected features are segmented using shape interaction matrix. The matrix $\boldsymbol{V}_r^t$ is divided by the QR decomposition as follows:

$$\underset{r\times P}{\boldsymbol{V}_r^t}\,\underset{P\times P}{\boldsymbol{\Pi}} = \left(\underset{r\times r}{\boldsymbol{V}_{11}^t} \,\middle|\, \underset{r\times(P-r)}{\boldsymbol{V}_{12}^t} \right) \tag{6}$$

The shape interaction matrix of the selected feature is calculated by the matrix $\boldsymbol{V}_{11}$.

$$\underset{r\times r}{\boldsymbol{X}_{11}} = \boldsymbol{V}_{11}\boldsymbol{V}_{11}^t = \{x_{ij}\} \qquad i,j \quad = \quad 1,\ldots,r \tag{7}$$

The size of this matrix is always $r \times r$ regardless the number of features P.

The property of shape interaction matrix shown in section 2 and the fact that the number of selected features for one object is four or less are used in the following segmentation algorithm.

[Step 1] Initialization:$i = 0$,$k = 0$, $S = \{1, 2, \ldots, r\}$.

[Step 2] Sort the entities in the rows of shape interaction matrix in the descending order.

[Step 3] Let $\{x_1, x_2, \ldots, x_r\}$ be the sorted entities of the row of $\boldsymbol{X}_{11}$ with the row number indicated by the first entity of S. The entity which satisfies the following rule is extracted from these entities.

$$\begin{aligned} |x_{j+1}| \quad < \quad \theta \quad \text{and} \quad & x_{j+1}/x_j < \theta \\ \text{with the smallest } & j+1,\ 1 \le j \le r \end{aligned} \tag{8}$$

where θ is threshold.
[Step 4] Construct the subset $S_k = (s_{k1}, \ldots, s_{kr_k})$, where s_{kj}, $j = 1, \ldots, r_k$ are the original column numbers of the first r_k entities which have greater values than the extracted entity in Step 3. S_k shows the permutation result of sorting in Step 2.
[Step 5] Collect the r_k rows whose permutation is equal to the set S_k. The row numbers of the collected rows should be identical to the entities of S_k.
[Step 6] Remove the row numbers of the collected r_k rows from S, and add them into the set S_{seg}. $i := i + r_k$.
[Step 7] If $i < r$, then $k := k + 1$ and go to Step 3. Otherwise, go to Step 8.
[Step 8] The segmentation result is the set $S_{seg} = \{S_1, \ldots, S_k\}$ and the number of objects is k. The number of dimensions of the shape space of each object is r_k. The permutation of shape interaction matrix by the set S_{seg} leads to the block-diagonal matrix corresponding to the segmentation result.

The rule in Step 3 shown by Eq.(8) is for search of the position of the first entity which can be regarded as zero. Thus the value of θ should be nearly zero such as 10^{-5}. The entity which satisfies the rule is in the first five entities of row, because the maximum number of features for each object is four. Since the number of features is always four for full-3D case, the above rule is not needed; it is only needed for degenerate case.

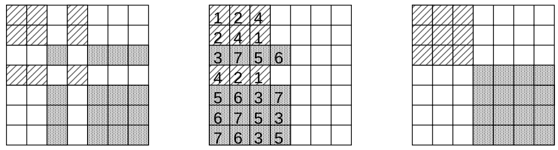
The degeneracy of object shape can be detected from the size of block matrix r_k, and the number of block matrices k represents the number of objects; discrimination of degenerate shape and estimation of the number of objects are simultaneously carried out in segmentation.

Figure 4 shows an example of segmentation of selected features. The white blocks in Fig.4(a) show the entities regarded as zero. In this example, $S = \{1, 2, 3, 4, 5, 6, 7\}$. The rows of the shape interaction matrix are sorted by Step 2 (Fig.4(b)). The result of Step 3 and Step 4 is $S_1 = \{1, 2, 4\}$. The dimension of the shape space corresponding to S_1 is three: this shape space represents the degenerate shape, i.e. plane. Since the second and fourth rows have the same permutation results, the sets are updated in Step 5 and Step 6 as follows:$S = \{3, 5, 6, 7\}$ and $S_{seg} = \{1, 2, 4\}$. After Step 3 to Step 6 applied for the updated S, $S_2 = \{3, 7, 5, 6\}$ and segmentation is done with the result $S_{seg} = \{1, 2, 4, 3, 7, 5, 6\}$. The shape interaction matrix is block-diagonalized using the permutation represented by S_{seg} (Fig.4(c)).

3.4 Classification of Remaining Features Using Subspace Method

The columns of the matrix $\boldsymbol{Q}$ in Eq.(5) can be permutated using the segmentation result of selected features as follows:

$$\underset{r\times r}{\tilde{\boldsymbol{Q}}} = \left[\underset{r\times r_1}{\tilde{\boldsymbol{Q}}^1}, \underset{r\times r_2}{\tilde{\boldsymbol{Q}}^2}, \ldots, \underset{r\times r_N}{\tilde{\boldsymbol{Q}}^N} \right] \tag{9}$$



(a) (b) (c)

Figure 4: An example of the segmentation of selected features.

The columns of $\tilde{\boldsymbol{Q}}^i$ $(i = 1, \ldots, N)$ are the basis vectors of shape space of the i-th object. The subspaces spaned by $\tilde{\boldsymbol{Q}}^i$ can be used to classify the remaining $P - r$ features.

We use the projection matrix $\boldsymbol{G}^i$ proposed by Noguchi[9] in classification.

$$ObjNo(j) = \underset{i}{\operatorname{argmax}} \left(\left\| \boldsymbol{G}^i \boldsymbol{v}_j \right\|^2 \right) \tag{10}$$

$$\underset{r\times r}{\boldsymbol{G}^i} = \tilde{\boldsymbol{Q}}^i \left\{ \tilde{\boldsymbol{Q}}^i \left(\tilde{\boldsymbol{Q}}^{i\,t} \tilde{\boldsymbol{Q}}^i \right)^{-1} \right\}^t \tag{11}$$

$$i = 1, \ldots, N, j = 1, \ldots, P$$

where $ObjNo(j)$ is the object number assigned to the j-th feature, and $\boldsymbol{v}_j$ is the columns of $\boldsymbol{V}_r^t = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_P)$. The projection matrix $\boldsymbol{G}^i$ has following property:

$$\begin{cases} \boldsymbol{G}^i \boldsymbol{v}_j \neq 0, & \boldsymbol{v}_j \in \operatorname{span}(\tilde{\boldsymbol{Q}}^i) \\ \boldsymbol{G}^i \boldsymbol{v}_j = 0, & \boldsymbol{v}_j \notin \operatorname{span}(\tilde{\boldsymbol{Q}}^i) \end{cases} \tag{12}$$

If the orthogonal projection matrix for each subspace is simply used, the difference of the dimension of subspace affects projection results; the norm of the projection vector for shape space of degenerate shape may be smaller than one of 3D shape. This leads to the error of classification. The projection matrix used here does not affected by the difference of the dimension of shape space due to the property shown in Eq.(12).

4 Experimental Results

To confirm the effectiveness of the proposed method, the data obtained from objects with known shape and segmentation were used.

4.1 Tracking of Objects

The objects used in the experiments are shown in Fig.5. A model-based tracking was used to track the object motion. In the experiments, the sampling points on wire-frame model of the objects were used as features (Fig.8). Each object was tracked separately, then the feature correspondences of the objects were permutated randomly to construct measurement matrix. The number of frames was 50, and the number of features of the objects were 328,326,296 and 280. Using the tracking result, we carried out two experiments: full-3D case and degenerate case.

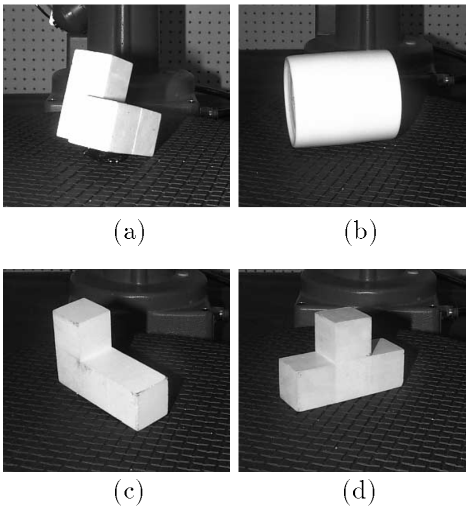


Figure 5: Four objects used in the experiments.

4.2 Full-3D Case

The four objects shown in Fig.5 were used. Figure 6 shows the correspondences of 1230 features. Table 1 shows the 18 singular values obtained by the SVD of the measurement matrix. Two smallest singular values were vanished using the threshold obtained by multiplying the maximum singular value by 10^{-6}. Thus the rank of the measurement matrix was estimated as 16, and 16 features were selected from 1230 ones.

Figure 7 shows the segmentation result. The shape interaction matrix constructed by the selected features was block-diagonalized by the algorithm described in section 3.3 (Fig.7 (a),(b)). The remaining features were classified using Eq.(10). The shape interaction matrix of all features was also block-diagonalized (Fig.7 (c),(d)). This interaction matrix shows that the four objects were correctly segmented.

The 3D shapes of the segmented objects were reconstructed by the factorization method for single object [7]. The shapes of objects were correctly reconstructed (Fig.8).

4.3 Degenerate Case

The three objects shown in Fig.5 (b),(c) and (d) were used. For the mug, only the circle of the bottom was tracked; this shape was degenerate. Figure 9 shows the correspondences of 772 features. Table 2 shows 14 singular values. The rank of the measurement matrix was estimated as 11.

The dimension of each shape space and the number of the objects were correctly estimated. Thus the shape interaction matrix constructed by 11 selected features was block-diagonalized (Fig.10 (a),(b)). The shape interaction matrix of all features was also block-diagonalized (Fig.10 (c),(d)). The 3D shapes were correctly reconstructed (Fig.11).

5 Conclusions

The motion segmentation method using feature selection and subspace method has been proposed. The features were selected through the computation of the basis of shape space. The features could be selected from each object without segmentation information by keeping the correspondence of basis vectors to features as shown in Fig.3. Feature selection was realized using the QR decomposition. Since only r features were used, segmentation could be carried out without optimization. The degeneracy analysis and estimation of the number of objects could be done simultaneously. The experimental results showed the effectiveness of the proposed method.

Acknowledgments

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

References

[1] T. Y. Tian and M. Shah:"Recovering 3D motion of multiple objects using adaptive Hough transformation," IEEE Trans. Pattern Anal. & Mach. Intell., Vol.19, No.10, pp.1178-1183, 1997

[2] T. Weiss and E. H. Adelson:"A unified mixture framework for motion segmentation: Incorporating spatial coherence and estimating the number of models," Proc. Computer Vision and Pattern Recognition, pp.321-326, 1996

[3] 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

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

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

[6] 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

[7] C. J. Poelman and T. Kanade: "A paraperspective factorization method for shape and motion recovery," IEEE Trans. Pattern Anal. & Mach. Intell., Vol.19, No.3, pp.206-218, 1997

[8] G. H. Golub and C. F. Van Loan:"Matrix Computations," The Johns Hopkins University Press, 1996

[9] Y. Noguchi:"Subspace method of feature extraction using non-symmetric projection operators," Bulletin of the Electrotechnical Laboratory, Vol.40, No.7, pp.11-27, 1976

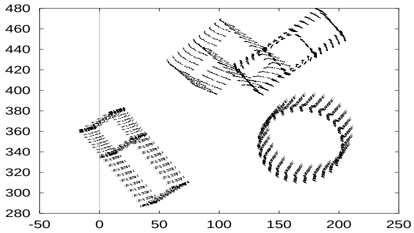


Figure 6: Feature correspondences of the four objects.

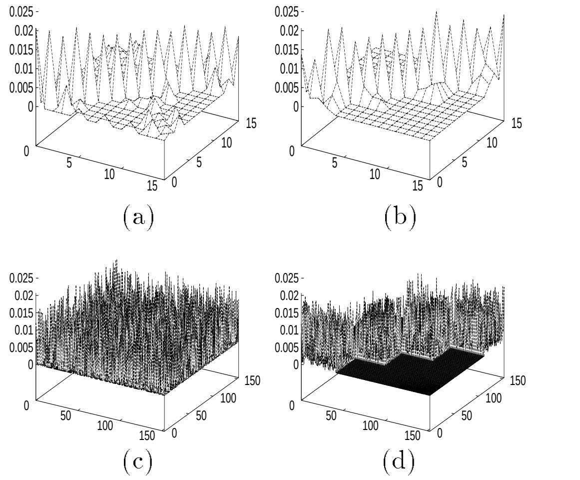


Figure 7: The segmentation result of the full-3D case. (a) the shape interaction matrix constructed by the selected features. (b) segmentation result of (a). (c) the shape interaction matrix constructed by the all features. (d) segmentation result of (c). In (c) and (d), every 8 features are sampled to display the shape interaction matrix clearly.

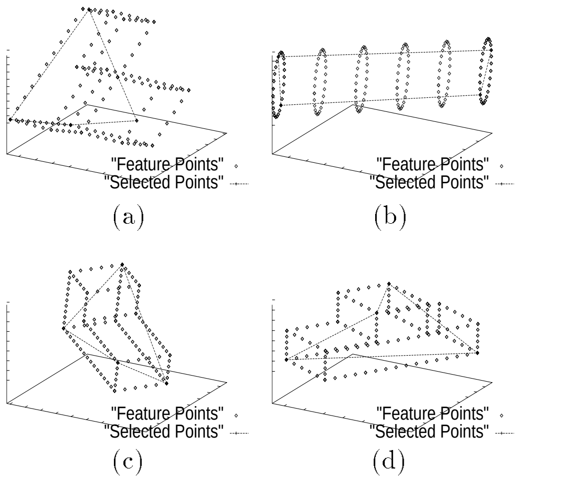


Figure 8: Results of the shape reconstruction in the full-3D case. The “Selected Points” in the figures are the features selected in segmentation.

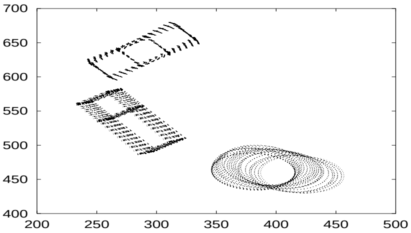


Figure 9: Feature correspondences of the three objects. The object with degenerate shape, plane, is included in this data.

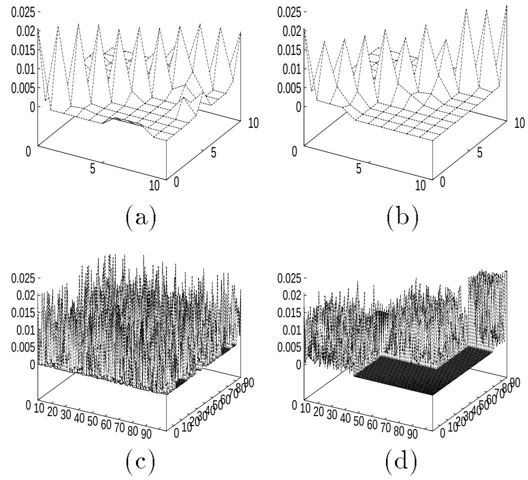


Figure 10: The segmentation result of the degenerate case. (a) the shape interaction matrix constructed by the selected features. (b) segmentation result of (a). (c) the shape interaction matrix constructed by the all features. (d) segmentation result of (c).

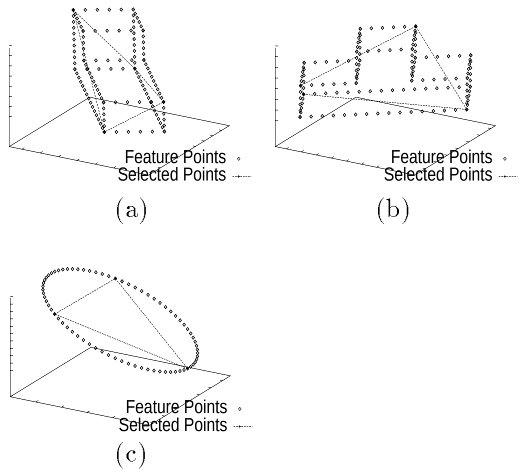


Figure 11: Results of the shape reconstruction in the degenerate case. The “Selected Points” in the figures are the features selected in segmentation.

Table 1: Singular values of the full-3D case.

Order	Singular Value	Order	Singular Value
1	130739	10	15.9511
2	35807.2	11	13.3373
3	313.306	12	11.8829
4	201.673	13	11.1419
5	47.1147	14	9.89105
6	37.6064	15	7.08395
7	23.8859	16	5.34026
8	21.7266	17	8.30671e-11
9	18.9427	18	3.5091e-11

Table 2: Singular values of the degenerate case.

Order	Singular Value	Order	Singular Value
1	110092	8	14.0002
2	28609.3	9	11.3941
3	229.167	10	8.44566
4	132.309	11	6.61284
5	41.211	12	7.45128e-11
6	27.9543	13	2.05956e-11
7	19.3066	14	1.30156e-11