Multibody Grouping by Inference of Multiple Subspaces from High-Dimensional Data Using Oriented-Frames

Zhimin Fan, *Student Member*, *IEEE Computer Society*,
Jie Zhou, *Senior Member*, *IEEE*, and Ying Wu, *Member*, *IEEE*

Abstract—Recently, subspace constraints have been widely exploited in many computer vision problems such as multibody grouping. Under linear projection models, feature points associated with multiple bodies reside in multiple subspaces. Most existing factorization-based algorithms can segment objects undergoing independent motions. However, intersections among the correlated motion subspaces will lead most previous factorization-based algorithms to erroneous segmentation. To overcome this limitation, in this paper, we formulate the problem of multibody grouping as inference of multiple subspaces from a high-dimensional data space. A novel and robust algorithm is proposed to capture the configuration of the multiple subspace structure and to find the segmentation of objects by clustering the feature points into these inferred subspaces, no matter whether they are independent or correlated. In the proposed method, an Oriented-Frame (**OF**), which is a multidimensional coordinate frame, is associated with each data point indicating the point's preferred subspace configuration. Based on the similarity between the subspaces, novel mechanisms of subspace evolution and voting are developed. By filtering the outliers due to their structural incompatibility, the subspace configurations will emerge. Compared with most existing factorization-based algorithms that cannot correctly segment correlated motions, such as motions of articulated objects, the proposed method has a robust performance in both independent and correlated motion segmentation. A number of controlled and real experiments show the effectiveness of the proposed method. However, the current approach does not deal with transparent motions and motion subspaces of different dimensions.

Index Terms—Computer vision, motion segmentation, subspace constraints.

1 Introduction

In various computer vision problems, multibody motions are frequently encountered and segmenting the scene into multiple entities is of fundamental importance. Motion segmentation techniques have been broadly employed in many applications such as shape from motion, video coding, surveillance, etc.

Among many proposed segmentation techniques, the factorization method, originally developed by Tomasi and Kanade [23], is particularly interesting. It is revealed that under linear projection models, points trajectories of a single body lie in a three or less dimensional linear manifold. Therefore, feature points of multibody reside in multiple subspaces. Most existing methods cope with independent multibody motion segmentation by enforcing the constraint that the trajectory subspaces, spanned by objects' feature trajectories, must be independent: i.e., $\forall p \neq q, \mathcal{T}_p \cap \mathcal{T}_q = \{0\}$, where $\mathcal{T}_p$ and $\mathcal{T}_q$ correspond to the trajectory subspace of object p and q. (Here, the trajectory of each feature is a column vector concatenating feature's image coordinates through successive frames. The detailed description of trajectory subspace will be given in Section 2.) Under this assumption, Boult and Brown [2] recursively grouped columns of trajectory matrix into independent subspaces. Gear [6] used the reduced row echelon form of the trajectory matrix and formulated the problem as weighted graph matching. Costeira and Kanade [3] presented a multibody factorization method in which a shape interaction matrix, $\mathbf{Q} = \{\mathcal{Q}_{ij}\}$, is introduced, where $\mathbf{Q} = \mathbf{V}\mathbf{V}^T$ and $\mathbf{V}$ comes from the SVD of the trajectory matrix $\mathbf{W}$, $\mathbf{W} = \mathbf{U}\Sigma\mathbf{V}^T$. In a noise-free case, if any features i and j are from different objects, $\mathcal{Q}_{ij}$ will be zero, otherwise, $\mathcal{Q}_{ij}$ may have nonzero values. They then grouped features by thresholding and sorting $\mathbf{Q}$. Based on the property of $\mathbf{Q}$, some extensions have also been made. Ichimura [10] applied a discriminant criterion to select the most representative vectors in $\mathbf{Q}$ for feature grouping. Wu et al. [29] decomposed $\mathbf{Q}$ into orthogonal subspaces and then grouped these fragment subspaces. Kanatani [12] developed a method through dimension correction and model selection.

However, the problem of multibody grouping with correlated motions, such as segmenting an articulated moving structure, poses a great challenge. For instance, considering a simple scenario of a moving arm involving two dependently moving objects: the upper arm and the lower arm, the dependence of motions is revealed by the intersections of trajectory subspaces, i.e., $\exists p \neq q,\ \mathcal{T}_p \cap \mathcal{T}_q \neq \{0\}$, where $\mathcal{T}_p$ and $\mathcal{T}_q$ denote the trajectory subspace of object p and q. In the presence of multiple correlated motions, the

- *Z. Fan is with the Department of Automation, Tsinghua University, Beijing, 100084, China and the Electrical and Computer Engineering Department, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208. E-mail: z-fan@northwestern.edu.*
- *J. Zhou is with the Department of Automation, Tsinghua University, Beijing, 100084, China. E-mail: jzhou@tsinghua.edu.cn.*
- *Y. Wu is with the Electrical and Computer Engineering Department, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208. E-mail: yingwu@ece.northwestern.edu.*

Manuscript received 13 Jan. 2004; revised 19 Nov. 2004; accepted 12 May 2005; published online 11 Nov. 2005.
Recommended for acceptance by S. Soatto.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0035-0104.

existing methods [2], [6], [3], [10], [29], [12] will be led to erroneous segmentation for the reason explained in Section 2.

In [30], Zelnik-Manor and Irani clearly describe the problem of dependencies among motions and suggest an algorithm to separate them. They use the information extracted from the most dominant eigenvectors, $\mathbf{V}$, of $\mathbf{W}^T\mathbf{W}$, which indicates the feature point's identity, to build a $\hat{\mathbf{Q}}$ instead of $\mathbf{Q}$ for segmentation. It can handle correlated motions, of different dimensions, but performs poorly in the presence of noise.

Kanatani and Sugaya proposed a multistage optimization algorithm [13]. By explicitly modeling the noise and assuming the *known* number of objects, the method classifies the feature points into objects in an EM-like manner, which has a robust performance against noise for correlated and transparent motion segmentation. However, because their method assumes the number of objects to be known and fixed, the presence of outliers, i.e., wrong data points, will significantly deteriorate its performance by possibly taking up an object identity and forcing the true points to be clustered into fewer objects.

A method combining Generalized Principle Component Analysis (GPCA) [25] and PowerFactorization for motion segmentation with missing data is proposed in [24]. With perfect data, i.e., in noise-free case, GPCA is provably correct for subspace structure identification and, thus, can handle transparent and correlated motions, with the motion subspaces being of different dimensions. However, this method alone does not handle outliers or noise properly. Additional pre and postprocessing is indispensable.

Compared with [30], [13], [24], our method does not deal with the transparent motions and motions of different dimensions and we shall say that the proposed approach is not provably correct in the absence of noise, unlike [13], and is not optimal in the presence of Gaussian noise, unlike [24]. This is because [13], [24] treat this problem as fitting a set of subspaces onto the data, which turns out to be theoretically sound. However, practically, these methods are largely confronted with the issue of handling the noise, especially the outliers, which violate the subspace structures and up to now cannot be perfectly addressed by these subspace fitting methods alone. As for [13], even if one outlier takes up an object identity by chance, the result will be poor because the number of objects is assumed known and fixed, thus the true points have to be classified into fewer objects instead.

Considering the correspondence between the multibody motions and their trajectory subspaces, if we can extract those subspaces no matter whether they are independent or correlated, both independent and correlated motion segmentation can be unified, which will lead to a broader application scope of multiple motion segmentation. However, dealing with noise and outliers in subspace identification is a hard problem. It is helpful to have several types of smoothing methods. In this paper, we propose a novel method to exploit the spatial relationship among the data points, which can properly discard outliers due to their subspace incompatibility.

The remainder of the paper is organized as follows: In Section 2, the problem of multibody grouping is presented. Related work on inference of subspaces is reviewed. Section 3 details the proposed Oriented-Frame (**OF**)-based multiple subspaces inference technique. Section 4 provides the comparative results between our method and some conventional and recent algorithms on both synthetic and real image sequences. Conclusions are summarized in Section 5.

2 Background

2.1 The Problem of Multibody Grouping

Suppose there are m moving objects in the scene, each object contains $p_i (i = 1, \ldots, m)$ points. Their homogeneous coordinates is represented by a $4 \times p_i$ matrix $\mathbf{S}_i$

$$\mathbf{S}_i = \begin{bmatrix} x_i^1 & x_i^2 & \cdots & x_i^{p_i} \\ y_i^1 & y_i^2 & \cdots & y_i^{p_i} \\ z_i^1 & z_i^2 & \cdots & z_i^{p_i} \\ 1 & 1 & \cdots & 1 \end{bmatrix}. \tag{1}$$

When a linear projection (orthographic, affine, etc.) is assumed, we collect the projected image coordinates (u, v) of these p_i points over F frames into a $2F \times p_i$ matrix, $\mathbf{W}_i$, i.e.,

$$\mathbf{W}_i = \mathbf{M}_i \mathbf{S}_i, \tag{2}$$

where

$$\mathbf{W}_i = \begin{bmatrix} u_{11} & \cdots & u_{1p_i} \\ v_{11} & \cdots & v_{1p_i} \\ u_{21} & \cdots & u_{2p_i} \\ v_{21} & \cdots & v_{2p_i} \\ \cdots & \cdots & \cdots \\ u_{F1} & \cdots & u_{Fp_i} \\ v_{F1} & \cdots & v_{Fp_i} \end{bmatrix}, \quad \mathbf{M}_i = \begin{bmatrix} \mathbf{M}_{1i} \\ \mathbf{M}_{2i} \\ \cdots \\ \mathbf{M}_{Fi} \end{bmatrix},$$

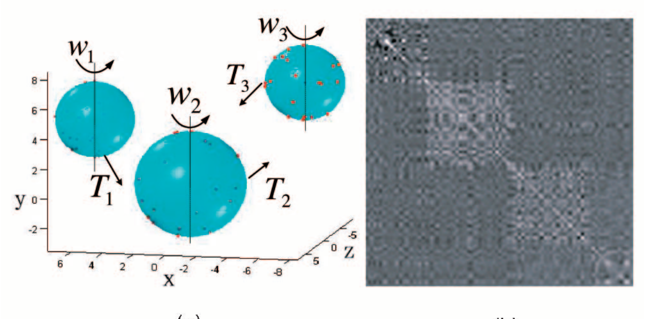
and $(\mathrm{u}_{fj}, \mathrm{v}_{fj})$ $(j = 1, \ldots, p_i)$ are the image coordinates of the feature points in the fth frame. $\mathbf{M}_i$ is a $2F \times 4$ matrix with $\mathbf{M}_{fi} (f = 1, \ldots, F)$ being the 2×4 projection matrix related to object i in the fth frame.

The p_i columns of $\mathbf{W}_i$ reside in a 4D *trajectory subspace* $\mathcal{T}_i$, which is spanned by the four columns of $\mathbf{M}_i$, i.e., $\mathcal{T}_i = \mathrm{span}(\mathbf{M}_i)$. Here, each column of $\mathbf{W}_i$ can be regarded as trajectory of the corresponding feature point.

Then, let $\mathbf{W}_1$ (of size $2F \times p_1$) and $\mathbf{W}_2$ (of size $2F \times p_2$) be the image coordinate matrices of two objects. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be the two corresponding trajectory subspaces. Given $[\mathbf{W}_1 | \mathbf{W}_2]$ up to a permutation of its columns, we would like to classify feature trajectories, i.e., columns of $[\mathbf{W}_1 | \mathbf{W}_2]$, according to objects. Denote $dim(\bullet)$ as the dimension of a subspace. Then, $\mathcal{T}_1$ and $\mathcal{T}_2$ can have different configurations:

1. **Independent trajectory subspaces**: When $\mathcal{T}_1 \cap \mathcal{T}_2 = \{0\}$, then
$$dim(\mathcal{T}_1 \cup \mathcal{T}_2) = dim(\mathcal{T}_1) + dim(\mathcal{T}_2).$$
This happens when the motions $\mathbf{M}_1$ and $\mathbf{M}_2$ of the two objects are independent.
2. **Correlated trajectory subspaces**: When $\mathcal{T}_1 \cap \mathcal{T}_2 \neq \{0\}$, $\mathcal{T}_1 \not\subseteq \mathcal{T}_2$, and $\mathcal{T}_2 \not\subseteq \mathcal{T}_1$, then
$$\begin{aligned}\max(dim(\mathcal{T}_1), dim(\mathcal{T}_2)) &< dim(\mathcal{T}_1 \cup \mathcal{T}_2) \\ &< dim(\mathcal{T}_1) + dim(\mathcal{T}_2),\end{aligned}$$
which means that intersections occur between subspaces $\mathcal{T}_1$ and $\mathcal{T}_2$. This happens when the motions $\mathbf{M}_1$ and $\mathbf{M}_2$ of the two objects are correlated.

After the SVD of $[\mathbf{W}_1 | \mathbf{W}_2]$, i.e., $[\mathbf{W}_1 | \mathbf{W}_2] = \mathbf{U}\Sigma\mathbf{V}^T$, Costeira and Kanade [3] constructed the "shape interaction matrix"



(a) (b)

Fig. 1. An example of correlated motions: (a) A view of the synthetic scene. (b) The "shape interaction matrix" $\mathbf{Q}$ [3], used by previous algorithms for segmentation. The number of columns and rows in $\mathbf{Q}$ is equal to the number of feature points. The values of $\mathcal{Q}_{ij}$ are transformed to the gray scale between [0, 255]. Darker color represents lower value. The ideal block-diagonal structure of $\mathbf{Q}$ is lost, which makes it difficult for previous methods to find the correct segmentation of correlated motions.

$$\mathbf{Q} = \mathbf{V}\mathbf{V}^T = \begin{bmatrix} \mathbf{S}_1^T \Lambda_1^{-1} \mathbf{S}_1 & 0 \\ 0 & \mathbf{S}_2^T \Lambda_2^{-1} \mathbf{S}_2 \end{bmatrix},$$

whose block diagonal structure is the foundation of most factorization-based algorithms. In [30], Zelnik-Manor and Irani pointed out that since columns of $\mathbf{V}$ are the eigenvectors of

$$[\mathbf{W}_1|\mathbf{W}_2]^T[\mathbf{W}_1|\mathbf{W}_2] = \begin{bmatrix} \mathbf{S}_1^T\mathbf{M}_1^T\mathbf{M}_1\mathbf{S}_1 & \mathbf{S}_1^T\mathbf{M}_1^T\mathbf{M}_2\mathbf{S}_2 \\ \mathbf{S}_2^T\mathbf{M}_2^T\mathbf{M}_1\mathbf{S}_1 & \mathbf{S}_2^T\mathbf{M}_2^T\mathbf{M}_2\mathbf{S}_2 \end{bmatrix},$$

$\mathbf{Q}$ will have a block diagonal structure only if the motion matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ are independent. When $\mathbf{M}_1$ and $\mathbf{M}_2$ are correlated, the off-diagonal blocks $\mathbf{S}_1^T\mathbf{M}_1^T\mathbf{M}_2\mathbf{S}_2$ and $\mathbf{S}_2^T\mathbf{M}_2^T\mathbf{M}_1\mathbf{S}_1$ are nonzero. Hence, the block diagonal form of $\mathbf{Q}$ and the basic assumptions of algorithms like [2], [6], [3], [10], [29], [12] will vanish, which may lead them to erroneous segmentations.

We give an example of correlated motions in Fig. 1a. Three transparent spheres are generated. Each of them undergoes random translation and different rotation around its own axis which is parallel to the y-axis. Assume that the camera's optical axis is along the z-axis and orthographic projection is employed, the motion matrices will have the form

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \cdots \\ \mathbf{M}_F \end{bmatrix} = \begin{bmatrix} R_{11}^{(1)} & 0 & R_{13}^{(1)} & t_x^{(1)} \\ 0 & 1 & 0 & t_y^{(1)} \\ R_{11}^{(2)} & 0 & R_{13}^{(2)} & t_x^{(2)} \\ 0 & 1 & 0 & t_y^{(2)} \\ & \cdots & & \cdots \\ R_{11}^{(F)} & 0 & R_{13}^{(F)} & t_x^{(F)} \\ 0 & 1 & 0 & t_y^{(F)} \end{bmatrix}. \tag{3}$$

Note that the second column of $\mathbf{M}$ are same for all the spheres so that their motions are *correlated*, i.e., the trajectory subspaces spanned by columns of $\mathbf{M}$s have intersections. Fig. 1b shows the matrix $\mathbf{Q}$, where the number of columns and rows in $\mathbf{Q}$ is equal to the number of feature points. The values of $\mathcal{Q}_{ij}$ are transformed to the gray scale between [0, 255]. Darker color represents lower value. It is seen that the $\mathbf{Q}$ loses its block diagonal form and, therefore, it is very difficult for most existing factorization-based methods to find the correct segmentation of correlated motions.

We now introduce our formulation of multibody grouping as a multiple subspace inference problem. Recall the formation of $\mathbf{W}_i$ in (2), let $\mathbf{r}_p$ and $\mathbf{r}_q$ denote the positions of any two columns of $\mathbf{W}_i$, which are vectors in a $2F$-dimensional space. Their relative position denoted by vector $\mathbf{r}_{pq} = \mathbf{r}_p - \mathbf{r}_q$ actually resides in a 3D subspace spanned by the first three columns of $\mathbf{M}_i$ because the last row of $\mathbf{S}_i$ is all 1s. We denote $\mathcal{M}_i$ to represent this 3D subspace formed by vectors $\mathbf{r}_{pq}(p, q \in [1, p_i], p \neq q)$, which we call the *motion subspace*.

Tracking all features of the m objects through F frames, we obtain a $2F \times P$ matrix, $\mathbf{W}$, i.e.,

$$\mathbf{W} = [\mathbf{W}_1\mathbf{W}_2 \ldots \mathbf{W}_m], \tag{4}$$

where $P = \sum_{i=1}^{m} p_i$ is the total number of feature points. Now, each object i has its own 3D *motion subspace*. Then, multibody grouping can be equivalently achieved by inference of the structure of the m motion subspaces and then classifying columns of $\mathbf{W}$ into their own motion subspaces.[1] In this paper, we propose a novel and robust Oriented-Frame (**OF**)-based subspace inference algorithm toward this goal, regardless of the independencies or correlations of the motion subspaces.

2.2 Subspace Constraint and Subspace Structure Inference

Subspace constraints exist in various vision problems, which has come into being an active and interesting research topic recently. Tomasi and Kanade [23] developed a factorization method based on subspace constraints to recover both the shape and the motion of an object from a sequence of images. Irani [11] showed that multiframe subspace constraints can be used for constraining the 2D correspondence estimation. In the problem of structure from motion, iterative methods incorporating subspace constraints have also been developed by Soatto and Perona [20] and Heyden et al. [9]. An interesting application incorporating subspace constraints is epipolar geometry estimation. Tang et al. [21] formulated the problem of epipolar geometry estimation as one of inferring hyperplane (a 7D manifold) in an 8D space analogous to plane detection in a 3D space. They extended the idea of tensor voting [8] to N-dimensional and achieved a robust performance.

The solutions to the space share an *essential character* that the low-dimensional subspace structure embedded in the high-dimensional data space must be properly revealed and estimated. Outliers should be discriminated and inliers should be classified to its own subspaces.

To the best of our knowledge, the algorithmic issues on the inference of subspaces in high-dimensional space have remained largely unexplored in the literature. A probabilistic approach, Mixture Probabilistic Principal Component Analysis (MPPCA), is proposed in [22] by estimating the maximum a posteriori for a distribution model of a mixture of subspaces. This is usually done in an iterative EM algorithm, which suffers from the sensitivity to the initialization and the unknown number of potential subspaces. A geometric approach is proposed by Vidal et al. [25], [24], called Generalized Principal Component Analysis (GPCA). GPCA

1. The motion subspace $\mathcal{M}_i$, which is to be inferred, is spanned by the first three columns of $\mathbf{M}_i$. While the trajectory subspace $\mathcal{T}_i$, which is introduced for convenience of analysis, is spanned by all the four columns of $\mathbf{M}_i$.

shows a sound theoretical approach to the identification of mixture of subspaces in noise-free case. However, how to enhance the performance of GPCA when a large amount of noise is present still remains an open issue.

In general, there are two main difficulties in identifying the structure of multiple subspaces. *One difficulty* stems from the unavoidable outliers inherent in the data set. If a significant portion of the data is corrupted by noise, the detection of subspaces will be difficult and the result will be so imprecise thus damaging the efficiency of applying subspace constraints to practical applications. For approaches like MPPCA or GPCA, a preprocessing of outlier rejection is indispensable to achieve a good performance. LMedS [16] and RANSAC [4] are considered to be some of the most robust methods. However, these methods require a majority of the data be correct. *Another difficulty* is that when the number of clusters exceeds two (e.g., the inliers reside in multiple subspaces), these robust methods may fail or become less attractive.

As for the problem of subspace inference for multibody grouping, our method can pertinently address and overcome these two difficulties: 1) disturbance caused by heavy noise and outliers and 2) existence of multiple clusters (subspaces), which will be detailed in Section 3 and validated in Section 4.

3 Multiple Subspace Inference and Multibody Grouping

In this section, we present a novel technique for multiple subspaces inference and apply it to multibody grouping. The input data is a $2F \times P$ matrix $\mathbf{W}$ without prior knowledge of the number of moving objects. Each data point is a $2F$-dimensional vector denoted by $\mathbf{r}_i (i = 1, \ldots, P)$. Our purpose is to extract multiple *motion subspaces* $\mathcal{M}_k$s out of $\mathbf{W}$. Each $\mathcal{M}_k$ is a 3D subspace formed by the vectors $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ $(i, j \in \text{Object}_k, i \neq j)$ or, equivalently, the subspace spanned by the first three columns of $\mathbf{M}_k$. Multibody grouping is equivalently achieved by classifying $\mathbf{r}_i$ to these motion subspaces $\mathcal{M}_k$s. Our technique for inferring these subspaces mainly consists of four stages:

1. conversion to Oriented-Frames (OF),
2. evolution of **OF**s,
3. voting, and
4. outlier rejection and subspace inference.

3.1 Conversion to Oriented-Frames

Initially, the set of P data points does not possess any information about their own subspace configurations. There is a need to give each point a configuration which can facilitate the communication of data points to exchange information about their subspace structures.

For any points, i and j, of the same object k, $\mathbf{r}_{ij} (= \mathbf{r}_i - \mathbf{r}_j)$ resides in the same 3D motion subspace $\mathcal{M}_k$. Therefore, the unit vector, $\hat{\mathbf{r}}_{ij} = \mathbf{r}_{ij}/r_{ij}$, can be used as point j's contribution for the inference of point i's motion subspace, where r_{ij} is the norm of $\mathbf{r}_{ij}$. Each point collects this information from all other points. If many of the $\hat{\mathbf{r}}_{ij}$ reside in (roughly) the same 3D subspace, we say that there is a high likelihood the 3D subspace accommodating the majority votes ($\hat{\mathbf{r}}_{ij}$) has a similar configuration to the current point's motion subspace. To estimate the agreement (or coherence) of the votes collected at point i, we compute the second order moment, $\mathbf{O}_i$, of these $2F$-dimensional vectors $\hat{\mathbf{r}}_{ij}$, or equivalently, a $2F$-dimensional hyperellipsoid having the same moments and principal axes.

Balancing the contributions from different points should also be considered. We refer to the motion's smoothness constraint, which has been successfully used in motion analysis such as layered motion representation [26] and Markov random field motion modeling [27]. The motion smoothness constraint claims that objects are usually composed of spatially contiguous regions in real scenes. So, we set the strength of the contributions inversely proportional to the distance between the "contribution-caster" and the "contribution-collector." In practice, the decay of the contribution takes the form of $\exp(-r_{ij}^2/\sigma_d^2)$, where σ_d is a scale factor. In our experiment, we take $\sigma_d = 0.3 \times r_{\text{med}}$, where r_{med} is the median value of all r_{ij} for $i \neq j$.

Thus, the votes are aggregated as follows:

$$\mathbf{O}_i = \sum_{j \neq i} \exp(-r_{ij}^2/\sigma_d^2) \cdot \hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{r}}_{ij}^T. \tag{5}$$

Then, we decompose matrix $\mathbf{O}_i$ into its corresponding eigensystems, i.e.,

$$\mathbf{O}_i = [\mathbf{V}_1 \quad \mathbf{V}_2 \cdots \mathbf{V}_{2F}] \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_{2F}) [\mathbf{V}_1 \quad \mathbf{V}_2 \cdots \mathbf{V}_{2F}]^T, \tag{6}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{2F}$ represent the sorted eigenvalues of $\mathbf{O}_i$, $\mathbf{V}_1, \mathbf{V}_1, \ldots, \mathbf{V}_{2F}$ are the corresponding eigenvectors and the symbol T denotes the transpose. These eigenvectors represent the $2F$ principal axes of the hyperellipsoid while the eigenvalues describe the strength and agreement measures on the corresponding axis. Then, each data point i will be associated with a two-tuple ($\mathbf{r}_i$, $\mathbf{OF}_i$). $\mathbf{r}_i$ is the position of that point in the $2F$-dimensional space and $\mathbf{OF}_i (= \{\mathbf{OF}_{i1}, \mathbf{OF}_{i2}, \mathbf{OF}_{i3}\})$ consists of the three dominant eigenvectors of $\mathbf{O}_i$, which is called *Oriented-Frame* in this paper, representing the *preferred* 3D motion subspace configuration.

3.2 Evolution of Oriented-Frames

Having associated each data point with an **OF** characterizing the point's own preferred subspace structure, **OF**s of the points in the same motion subspace are expected to have similar configurations. However, the initial **OF**s may not be accurate enough due to the ambiguities caused by noise and outliers. Based on the similarity of **OF**s, a novel mechanism for subspace rotation is proposed to eliminate the intraclass variabilities to provide a desirable property (a low intercluster similarity and a uniformly distributed, high intracluster similarity) for this clustering problem. In the following discussion, uppercase calligraphic letters represent subspaces, e.g., $\mathcal{A}$, and an uppercase boldface letter will represent a matrix, e.g., $\mathbf{A}$.

3.2.1 *Similarity Measurement for Subspace Comparison*

This subspace similarity measurement is derived from principal angles and principal vectors [7].

Definition 1. *Let $\mathcal{A}$ and $\mathcal{B}$ be two p-dimensional subspaces in an l-dimensional space. $\mathbf{A}$ and $\mathbf{B}$ are $l \times p$ matrices consisting of the orthonormal bases of $\mathcal{A}$ and $\mathcal{B}$. The principal angles, $0 \leq \theta_1 \leq \cdots \leq \theta_p \leq \pi/2$, and the principal* *vectors, $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$, $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$, are defined as follows: Computing the SVD of $\mathbf{A}^T\mathbf{B}$:*

$$\mathbf{Y}^T(\mathbf{A}^T\mathbf{B})\mathbf{Z} = \mathrm{diag}(\sigma_1, \ldots, \sigma_p),$$

where $\mathbf{Y}^T\mathbf{Y} = \mathbf{Z}^T\mathbf{Z} = \mathbf{I}$ (the $p \times p$ identity matrix). Then, we have

$$\begin{cases} [\mathbf{u}_1, \ldots, \mathbf{u}_p] & = & \mathbf{AY}, \\ [\mathbf{v}_1, \ldots, \mathbf{v}_p] & = & \mathbf{BZ}, \\ \cos(\theta_k) & = & \sigma_k, \quad k = 1, \ldots, p. \end{cases} \tag{7}$$

The equation,

$$[\mathbf{u}_1, \ldots, \mathbf{u}_p]^T[\mathbf{v}_1, \ldots, \mathbf{v}_p] = \mathbf{Y}^T(\mathbf{A}^T\mathbf{B})\mathbf{Z} = \mathrm{diag}(\cos(\theta_1), \ldots, \cos(\theta_p)),$$

indicates that the angle between the ith pair of principal vectors $(\mathbf{u}_i, \mathbf{v}_i)$ is the ith principal angle of the subspace pair. Although the orthonormal bases in matrices $\mathbf{A}$ and $\mathbf{B}$ can be arbitrary, they are aligned by the SVD to discover the essential relationship between these two subspaces. In this sense, a similarity measurement between the subspaces, $\mathcal{A}$ and $\mathcal{B}$, can be defined as,

$$\phi(\mathcal{A}, \mathcal{B}) = \prod_{k=1}^{p} \cos(\theta_k). \tag{8}$$

It is obvious that identical subspaces have the maximum similarity measure of value 1. However, since this similarity can only deal with a subspace pair of equal dimension, our method cannot handle motion subspaces of different dimension. For similarity, or distance, between subspaces, the Martin distance considering subspace angles is known to be an initial work, which measures the distance between ARMA models [15]. In [1], the subspace angles are used to recognize gait patterns described by a dynamic system. Recently, a formulation of kernelized subspace angle for nonlinear and complex data pattern analysis is proposed by Wolf and Shashua [28].

Observation 1. *The p subspaces spanned by $\{\mathbf{u}_i, \mathbf{v}_i\}(i = 1, \ldots, p)$ are mutually orthogonal.*

The computation of the principal angles, $[\mathbf{u}_1, \ldots, \mathbf{u}_p]^T [\mathbf{v}_1, \ldots, \mathbf{v}_p] = \mathrm{diag}(\cos(\theta_1), \ldots, \cos(\theta_p))$, indicates that vector $\mathbf{u}_i$ is orthogonal to vector $\mathbf{v}_j$ $(j \neq i)$. Since

$$[\mathbf{u}_1, \ldots, \mathbf{u}_p]^T[\mathbf{u}_1, \ldots, \mathbf{u}_p] = \mathbf{Y}^T(\mathbf{A}^T\mathbf{A})\mathbf{Y} = \mathbf{I},$$

$\mathbf{u}_i$ is also orthogonal to $\mathbf{u}_j (j \neq i)$. Analogously, $\mathbf{v}_i$ is orthogonal to both $\mathbf{u}_j$ and $\mathbf{v}_j$ for $j \neq i$. Therefore, the subspace spanned by $\{\mathbf{u}_i, \mathbf{v}_i\}$ is orthogonal to the subspace spanned by $\{\mathbf{u}_j, \mathbf{v}_j, j \neq i\}$, which further implies that the p subspaces spanned by $\{\mathbf{u}_i, \mathbf{v}_i\}(i = 1, \ldots, p)$ are mutually orthogonal.

3.2.2 Mechanism of Rotating Subspaces

Now comes the main topic of this section, mechanism of subspace rotation. As for the feature points, whose **OF**s have similar preferences of subspace configurations as defined in (8), this mechanism is used to rotate their **OF**s to strengthen their agreement on the underlying subspace structure, and equivalently to eliminate the intraclass variabilities. We begin with a simple example in the 3D space for ease of visualization and then extend it to the N-dimensional space. Consider two nonparallel planes A and B (see Fig. 2), which intersect at line f. How to rotate plane A to B, until these two planes overlap?

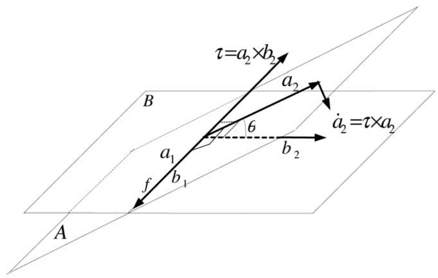


Fig. 2. An illustration of the rotation of plane A to B. $\{\mathbf{a}_1, \mathbf{a}_2\}$ and $\{\mathbf{b}_1, \mathbf{b}_2\}$ are principal vectors of this subspace pair. The principal angles between these two subspaces are 0 and $\cos^{-1}(\|\mathbf{a}_2^T\mathbf{b}_2\|)$. Rotating plane A to B is equivalent to rotating vector $\mathbf{a}_2$ toward $\mathbf{b}_2$ until the angle between them is zero. A torque $\tau = \mathbf{a}_2 \times \mathbf{b}_2$ on $\mathbf{a}_2$ is applied, which induces a clockwise rotation on $\mathbf{a}_2$ and makes the instantaneous change of $\mathbf{a}_2$ along the direction $\dot{\mathbf{a}}_2 = \tau \times \mathbf{a}_2$. This rotation can increase the similarity of these two planes according to (8).

Suppose that $\{\mathbf{a}_1, \mathbf{a}_2\}$ and $\{\mathbf{b}_1, \mathbf{b}_2\}$ are principal vectors of this subspace pair. The principal angles between these two subspaces are 0 and $\cos^{-1}(\|\mathbf{a}_2^T\mathbf{b}_2\|)$. Then, rotating plane A to B is now equivalent to rotating vector $\mathbf{a}_2$ toward $\mathbf{b}_2$ until the angle between them is zero. We can apply a torque $\tau = \mathbf{a}_2 \times \mathbf{b}_2$ on $\mathbf{a}_2$, which induces a clockwise rotation on $\mathbf{a}_2$ and makes the instantaneous change of $\mathbf{a}_2$ along the direction $\dot{\mathbf{a}}_2 = \tau \times \mathbf{a}_2$. This rotation can increase the similarity of these two planes according to (8).

Observation 2. *Let $\mathcal{S}_{ab}^{(2)}$ denote the subspace spanned by $\mathbf{a}_2$ and $\mathbf{b}_2$, and $\mathcal{S}_{ab}^{(2)^\perp}$ denote the orthogonal complement of $\mathcal{S}_{ab}^{(2)}$. In order to rotate $\mathbf{a}_2$ toward $\mathbf{b}_2$, the direction of instantaneous displacement of $\mathbf{a}_2$, i.e., $\dot{\mathbf{a}}_2$, resides in the subspace $\mathcal{S}_{ab}^{(2)}$ and is perpendicular to both $\mathcal{S}_{ab}^{(2)^\perp}$ and $\mathbf{a}_2$.*

It is worthy to note that the concept of cross product plays an essential role in the 3D case. The cross product of two 3D vectors can also be viewed as calculating the *orthogonal complement* of the subspace spanned by the concerned two vectors. Referring to this idea, rotating $\mathbf{a}_2$ toward $\mathbf{b}_2$ can be realized in two steps as follows: First, $\mathcal{S}_{ab}^{(2)^\perp}$ is calculated by cross product of $\mathbf{a}_2$ and $\mathbf{b}_2$ as the so-called "torque." Second, $\dot{\mathbf{a}}_2$ is computed (by cross product) as the orthogonal complement of subspace spanned by $\mathcal{S}_{ab}^{(2)^\perp}$ (the torque) and $\mathbf{a}_2$. So, $\dot{\mathbf{a}}_2$ is perpendicular to both $\mathcal{S}_{ab}^{(2)^\perp}$ and $\mathbf{a}_2$.

In the N-$D(N > 3)$ space, the rotation is more complex since there is no concept of cross product. Inspired by Observation 2, the mechanism for rotating an N-dimensional vector, $\mathbf{a}$, toward another N-dimensional vector, $\mathbf{b}$, narrowing the angle $\alpha = cos^{-1}(\|\mathbf{a}^T\mathbf{b}\|)$ can be deduced in a similar manner. Let $\mathcal{S}_{ab}$ denote the subspace spanned by $\mathbf{a}$ and $\mathbf{b}$. First, the orthogonal complement of $\mathcal{S}_{ab}$, denoted by $\mathcal{S}_{ab}^\perp$, is calculated using such as QR decomposition. Compared with 3D case, $\mathcal{S}_{ab}^\perp$ can be considered as a "N-dimensional torque." Second, the orthogonal complement of subspace spanned by $\mathcal{S}_{ab}^\perp$ and $\mathbf{a}$ is computed and denoted by $\mathbf{a}_r$, which can be regarded as the "high-dimensional cross product" of "N-dimensional torque" and $\mathbf{a}$ analogously in the 3D case. Then, the direction of the instantaneous displacement of $\mathbf{a}$ is obtained as $\dot{\mathbf{a}} = \mathbf{a}_r \cdot \mathbf{a}_r^T\mathbf{b}/\|\mathbf{a}_r^T\mathbf{b}\|$. It can be easily verified that the angle between $\mathbf{a} + \mu\dot{\mathbf{a}}$ and $\mathbf{b}$ is smaller than that between $\mathbf{a}$ and $\mathbf{b}$, where μ is a small scalar controlling the magnitude of rotation ($\mu = 0.01$ in our experiments). Table 1 summarizes the relationship between 3D and N-dimensional rotation.

TABLE 1
Generalization of 3D Rotation to N-Dimensional Space

	3D	*N*-D
torque for rotating vector **a** towards **b**	$\times$: cross product	$\perp$: orthogonal complement
	τ=**a**$\times$**b**	$\mathcal{S}_{ab}^{\perp}=(\mathbf{a}\oplus\mathbf{b})^{\perp}$
result of applying torque on **a**	$\dot{\mathbf{a}}=\tau\times\mathbf{a}$	$\mathbf{a}_r=(\mathcal{S}_{ab}^{\perp}\oplus\mathbf{a})^{\perp}$
$\dot{\mathbf{a}}$: instantaneous displacement of **a**	$\dot{\mathbf{a}}$	$\mathbf{a}_r\cdot\mathbf{a}_r^T\mathbf{b}/\|\mathbf{a}_r^T\mathbf{b}\|$

(Please refer to Section 3.2.2 for details.)

Then, let us consider the rotation of a p-dimensional subspace $\mathcal{A}$ toward $\mathcal{B}$ in the N-dimensional space. Given the corresponding principal angles and principal vectors, $\{\mathbf{u}_k, \mathbf{v}_k, \theta_k, k = 1, \dots, p\}$, we can find the instantaneous change, $\dot{\mathbf{u}}_i$, for rotating $\mathbf{u}_i$ toward $\mathbf{v}_i$ using the above mechanism. Denote $\mathcal{S}_{uv}^{(i)}, \mathcal{S}_{uv}^{(i')}$, and $\mathcal{A}'$ as the subspaces spanned by $\{\mathbf{u}_i, \mathbf{v}_i\}$, $\{\mathbf{u}_i + \mu\dot{\mathbf{u}}_i, \mathbf{v}_i\}$, and $\{\mathbf{u}_1, \dots, \mathbf{u}_{i-1}, \mathbf{u}_i + \mu\dot{\mathbf{u}}_i, \mathbf{u}_{i+1}, \dots, \mathbf{u}_p\}$, respectively. Since $\dot{\mathbf{u}}_i$ is restricted in $\mathcal{S}_{uv}^{(i)}$ (see Observation 2), we have $\mathcal{S}_{uv}^{(i)} \equiv \mathcal{S}_{uv}^{(i')}$. So, after the rotation, the subspace spanned by $\{\mathbf{u}_k, \mathbf{v}_k, (k \neq i)\}$ is orthogonally complementary to $\mathcal{S}_{uv}^{(i')}$. It is important to note that the structure of mutually orthogonal subspaces spanned by $\{\mathbf{u}_i, \mathbf{v}_i\}(i = 1, \dots, p)$ is actually unaltered. In fact, $\{\mathbf{u}_i + \mu\dot{\mathbf{u}}_i, \mathbf{v}_i\}$ is still a pair of principal vectors of $\mathcal{A}'$ and $\mathcal{B}$ (see Observation 1).

Thus, the process of rotating a p-dimensional subspace to another can be divided into p steps by rotating $\mathbf{u}_1$ to $\mathbf{v}_1$, $\mathbf{u}_2$ to $\mathbf{v}_2, \dots,$ and $\mathbf{u}_p$ to $\mathbf{v}_p$, successively. After the ith step involving $\mathbf{u}_i$ and $\mathbf{v}_i$, only the corresponding ith principal angle is modified, while no impact is made on other principal angles. Through this operation, these two subspaces can gradually become identical in the sense of having the same configuration, with maximum similarity 1.

3.2.3 Evolution of Oriented-Frames

Recall that after the conversion stage as described in Section 3.1, each data point has been associated with an oriented-frame possessing the information about the point's preferred subspace structure. Now, using the above mechanism, points with similar **OF**s can now rotate their **OF**s to enhance the saliency of their underlying motion subspace structure.

We define a similarity measurement matrix of **OF** for all data points as

$$\mathbf{\Phi} = \{\phi_{(i,j)} : \phi_{(i,j)} = \prod_{k=1}^{3} \cos(\theta_k), \forall i, j \leq P\}, \tag{9}$$

where θ_k $(k = 1, 2, 3)$ are the principal angles of subspace pair spanned by $\mathbf{OF}_i$ and $\mathbf{OF}_j$. In our experiment, for points, i and j, $(j \neq i)$, if $\phi_{(i,j)} > 0.7$, we rotate $\mathbf{OF}_i$ toward $\mathbf{OF}_j$ for a appropriate magnitude to obtain a greater $\phi_{(i,j)}$. Such rotations are applied for all data points[2] and the **OF**s and $\mathbf{\Phi}$ are thus updated.

2. We first keep the current configurations of all **OF**s as $\mathbf{OF}_i^{old}, i = 1 \dots P$. Then, for each $\mathbf{OF}_i$, denote the indices of the **OF**s that $\mathbf{OF}_i$ need to be rotated toward as $\mathtt{Ind}(\mathbf{OF}_i)$. We rotate $\mathbf{OF}_i$ toward $\mathbf{OF}_j^{old}, j \in \mathtt{Ind}(\mathbf{OF}_i)$ in turn. Totally, $\mathbf{OF}_i$ is rotated rot_i times, with rot_i being the number of elements in $\mathtt{Ind}(\mathbf{OF}_i)$. The configuration of $\mathbf{OF}_i$ changes after each rotation and ends up with $\mathbf{OF}_i^{new}$. Similarly with other OF s. When we encounter a $\mathbf{OF}_j$ that needs to be rotated toward $\mathbf{OF}_i$, the subspace pair will be $\mathbf{OF}_j$ and $\mathbf{OF}_i^{old}$, but not $\mathbf{OF}_j$ and $\mathbf{OF}_i^{new}$. In brief, the kept version of $\mathbf{OF}^{old}$ remain unchanged in this batch process and serve as the reference for all $\mathbf{OF}_i$ to be rotated to $\mathbf{OF}_i^{new}$.

The evolution here has a certain resemblance to the Tangent Distance method [19]. It is known that one of the critical factors affecting the performance of pattern classification is the pattern variation. Therefore, it is desirable to develop robust classification algorithms that can tolerate small variations of the input patterns. The variations can be learned from a manifold composed of sample points from the same class [5], [14]. Modeling the variations in such a manifold actually represents the *theme* of the Tangent Distance method, that is, to *eliminate variabilities among objects of the same class*, while identifying differences among the objects of different classes.

3.3 Voting Stage

The configurations of **OF** s are important in this inference task since they are the tokens of the data points. The voting stage, another type of communication between the data points, is used to extract and accordingly encode the subspace structural information to construct **OF**. Previously, the second order moment of the normalized vector of $\mathbf{r}_{ij}$, i.e., $\hat{\mathbf{r}}_{ij}$, is used to reveal the dominant distributions of the spatial information. Besides, the **OF** itself, which represents the most probable subspace configuration of each data point, is also a collection and a compact representation of the spatial information. Since the up-to-date subspace structural information is contained in the constructing components of **OF**s, we also include the second order moment, $\sum_{k=1}^{3} \mathbf{OF}_{jk} \cdot \mathbf{OF}_{jk}^T$, into the calculation of $\mathbf{O}_i$. Recall that the dominant eigenvectors of $\mathbf{O}_i$ is used to construct $\mathbf{OF}_i$.

Furthermore, only using the relative distance to weigh the vote is also inadequate. If $\phi_{(i,j)} > \phi_{(i,k)}$, similarity of motion subspace configurations of point i and j is higher than that of point i and k, and the vote from point j to i should be more reliable than that from k to i. Therefore, the current estimated **OF** similarity measurement matrix $\mathbf{\Phi}$ can be used to weigh the vote between point pairs.

Taking these into consideration, the summation of vote collected at point i in the voting stage will be formulated as

$$\mathbf{O}_i = \sum_{j \neq i} \phi_{(i,j)} \exp(-r_{ij}^2/\sigma_d^2) \cdot \left(\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{r}}_{ij}^T + \sum_{k=1}^{3} \mathbf{OF}_{jk} \cdot \mathbf{OF}_{jk}^T \right). \tag{10}$$

The procedures of evolution and voting cooperate to enforce tight connections among inliers, to suppress noise distortions, and to make a consistent exploration toward the inherent subspace patterns for segmentation. In real situations, most data patterns are undergoing variations or perturbations more or less, being able to cope with variations in noise-corrupted data also increases the robustness of the algorithm.

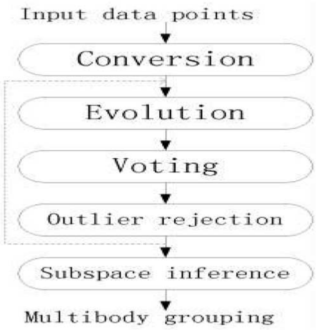


Fig. 3. The flowchart of the algorithm.

3.4 Outlier Rejection and Multibody Grouping Using Subspace Inference

Through the evolution (Section 3.2) and voting (Section 3.3) stages, the information about the underlying subspace structure can be communicated effectively among the data points. So, if a point i, its $\mathbf{OF}_i$ receives little structural agreement from others, i.e., $\sum_{j\neq i} \phi_{(i,j)} < \phi_{th}$, point i can be isolated as an outlier. Due to the discriminative similarity measurement $\phi_{(i,j)}$, which evaluates all the cosine values of the principal angles for subspace comparison, it is difficult for an outlier to be mixed within the group of inliers because of its structural incompatibility. Thus, it is allowed to choose the threshold ϕ_{th} in a relatively wide range. Typically, the value of $\phi\mathrm{th}$ is set to 20 to 30 percent of the median value of all $\sum_{j\neq i} \phi_{(i,j)}$.

To improve accuracy, we repeatedly run the evolution and voting stages to filter out outliers and to reduce the intraclass variations. The set of inliers is progressively refined as more outliers are rejected in each pass. Usually, only a few iterations are needed. We use four to five passes in our experiments.

In our experiments, the $\phi_{(i,j)}$ is usually close to 1 for pair of points i and j in the same subspace. This property considerably facilitates the grouping decision. If $\phi_{(i,j)} > 0.95$, point i and j will be put in the same group. Then, we calculate the second order symmetric tensor of the relative positions of the inliers for each group. The underlying motion subspace configuration $\mathcal{M}_k$s can be obtained by spanning the three dominant eigenvectors of these tensors. Consequently, data points $\mathbf{r}_i$s can be checked against the inferred multiple motion subspaces, producing a set of grouped inliers. Thus, the task of classifying feature points $\mathbf{r}_i$s to the motion subspaces, $\mathcal{M}_k$s, is accomplished, which is equivalent to the multibody grouping. Of course, no prior knowledge of the number of moving objects is assumed here. Furthermore, throughout the process, no assumption of independence between subspaces is made. Our method can segment correlated motions as well as independent motions.

3.5 Summary of the Algorithm

The subspace inference algorithm for multibody motion segmentation can now be summarized as below, see also Fig. 3.

In the first stage, the spatial configuration of the data space is explored, and each data point is associated with an *Oriented Frame*. Then, **OF** evolution and **OF** voting act, in turn, to abate the noise disturbance and expose the underlying multiple subspace structure. Finally, the subspace configuration obtained from the filtered data leads to the multibody motion segmentation.

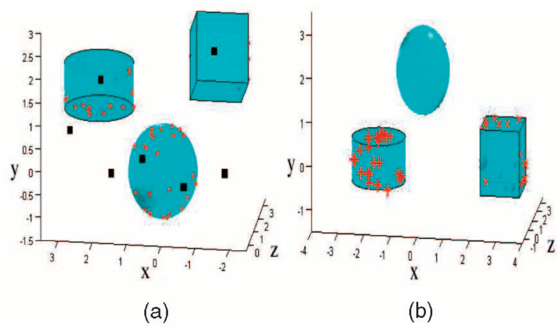


Fig. 4. An example of synthetic independent motions. (a) A view of the synthetic scene. (b) By using our method, grouped feature points on three moving objects are shown by "x," "+," and "o," respectively. Some detected outliers are shown by black "□" in (a).

3.6 Complexity Reduction

For practical purposes, the computational complexity can be significantly reduced by incorporating *range limitation*. If there are a large amount of data points, it is prohibitively expensive to compute all interpoint communications. Therefore, a distance threshold is set. If distance between two concerned points is greater than $2\sigma_d$, the mutual influence should be ignored. So, (5) in the conversion stage can be rewritten as follows:

$$\mathbf{O}_i = \sum_{r_{ij} < 2\sigma_d, j\neq i} \exp(-r_{ij}^2/\sigma_d^2)\cdot \hat{\mathbf{r}}_{ij}\cdot \hat{\mathbf{r}}_{ij}^T, \tag{11}$$

where σ_d is the same as the one defined in Section 3.1. Likewise, (10) in the voting stage will be

$$\mathbf{O}_i = \sum_{r_{ij} < 2\sigma_d, j\neq i} \phi_{(i,j)} \exp(-r_{ij}^2/\sigma_d^2) \cdot \left(\hat{\mathbf{r}}_{ij}\cdot\hat{\mathbf{r}}_{ij}^T + \sum_{k=1}^{3} \mathbf{OF}_{jk}\cdot \mathbf{OF}_{jk}^T\right). \tag{12}$$

4 Experimental Results on Multibody Grouping

In this section, the experimental results on a variety of synthetic and real image sequences are presented. Both of the synthetic data and real images contain multiple independent motions or correlated motions. We compared our method with some conventional and recent algorithms, such as the discriminant method [10], which extracts the most representative vectors in $\mathbf{Q}$ for feature grouping, the Zelnik-Manor and Irani's method [30], which uses the indicator information of the first eigenvector of $\mathbf{W}^T\mathbf{W}$ for segmentation, and Kanatani's multistage optimization [13],[3] which groups feature points by an iterative multistage optimization.

4.1 Synthetic Data

We use two synthetic data sets consisting of independent motions and correlated motions, respectively. In both data sets, a comparison with previous methods [10], [30], [13] is given.

Independent motion segmentation. Fig. 4a shows a synthetic scene. Three transparent 3D objects, i.e., a sphere, a cylinder and a cubic, are generated and 30 points from each body are randomly chosen. Each object undergoes a

3. http://www.suri.it.okayama-u.ac.jp/e-program-separate.html.

TABLE 2
A Summary of Segmentation Results for Multibody Grouping on Synthetic Scenes

	independent motions				correlated motions			
image size	100×100				100×100			
std. of Gaussian noise	2				2			
input features	good matches: 90			outliers	good matches: 94			outliers
	sphere	cylinder	cubic		sphere 1	sphere 2	sphere 3	
	30	30	30	100	22	30	42	100
clustering result	30	28	29	103	20	30	42	102
true inliers (true outliers)	30	28	29	100	20	30	42	100
false outliers (false inliers)	0	0	0	3	0	0	0	2

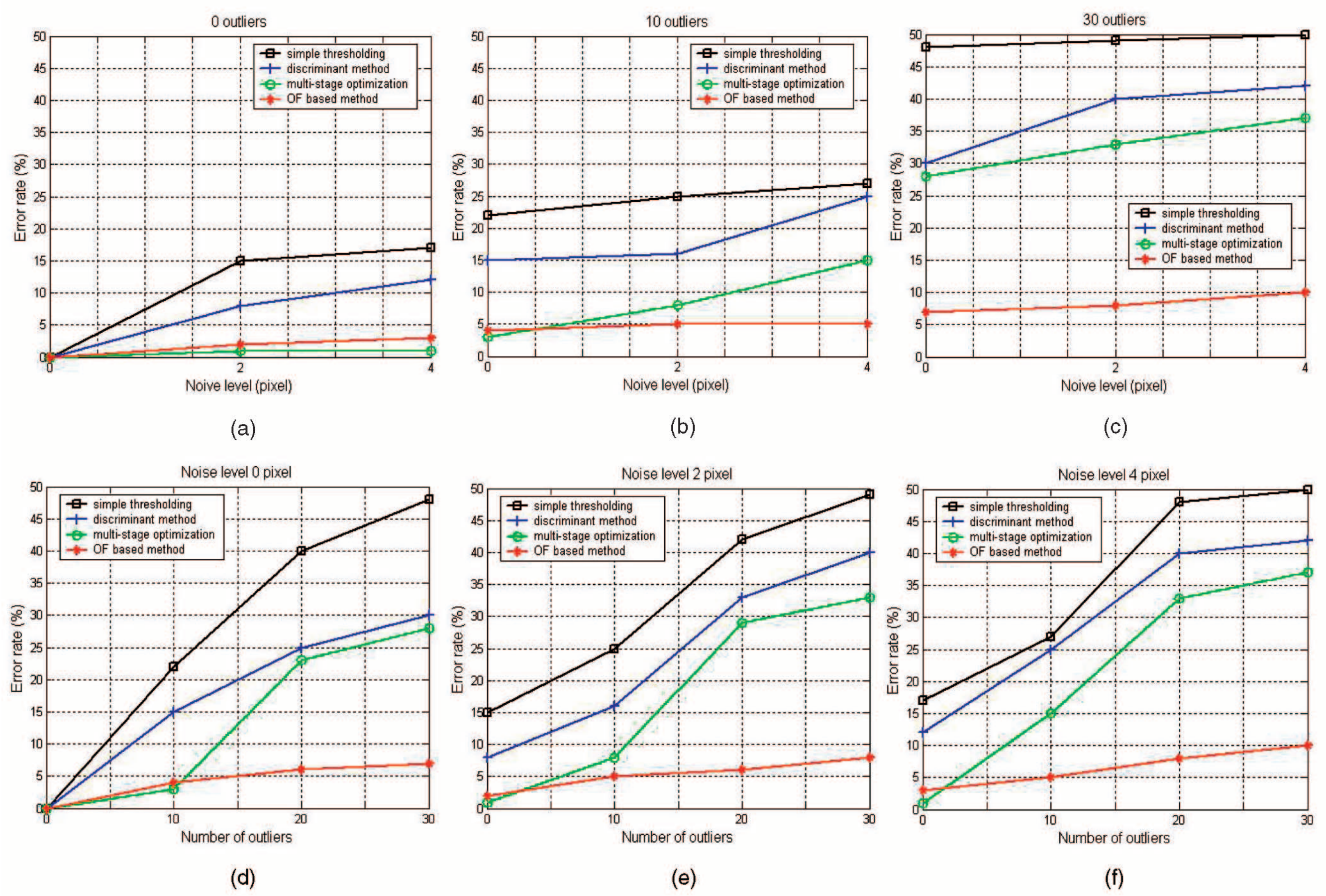


Fig. 5. Four methods, simple thresholding using matrix Q, discriminant method [10], multistage optimization [13], and Oriented-Frame-based method, are compared on independent motion segmentation. (a), (b), and (c) Error rate versus the level of noise with the fixed number of outliers. (d), (e), and (f) Error rate versus the number of outliers with the fixed standard deviation of noise. For each level of noise, the error rate is computed as the average value of 100 independent trials of each method.

sequence of independent motion. Ten image frames with a resolution of 100×100 pixels are captured. Gaussian noise with a standard deviation of 2 pixels is added to the data matrix $\mathbf{W}$. We fabricate 100 fake trajectories in the image stream as outliers. Thus, the ratio of outlier to inlier is 1.11. Our method correctly discards all outliers. The segmentation result is shown in Fig. 4b and Table 2.

Using this sequence, four algorithms, the simple thresholding using matrix $\mathbf{Q}$, the discriminant method [10], Kanatani's multistage optimization [13] and the **OF**-based method, are compared. Fifty feature points are selected, 25 from the sphere and 25 from the cylinder. We add Gaussian noise with zero-mean and a standard deviation ranging from 0 to 4 pixels to the coordinates of the feature points. We also introduced 0, 10, 20, and 30 fabricated outliers. The results are shown in Fig. 5. There are two sets of graphs. Figs. 5a, 5b, and 5c show the error rate versus the level of noise with the fixed number of outliers (0 outlier in Fig. 5a, 10 outliers in Fig. 5b, 30 outliers in Fig. 5c). Figs. 5d, 5e, and 5f show the error rate versus the number of outliers with the fixed standard deviation of the noise (0 pixel in Fig. 5d, 2 pixels in Fig. 5e, 4 pixels in Fig. 5f). For each level of noise, the error rate is computed as the average value of 100 independent trials of each method. Our algorithm performs much better than the method of simple thresholding and the discriminant method. Compared with multistage optimization, our algorithm achieves a much lower misclassification rate when there exists outliers. In contrast, because multistage optimization

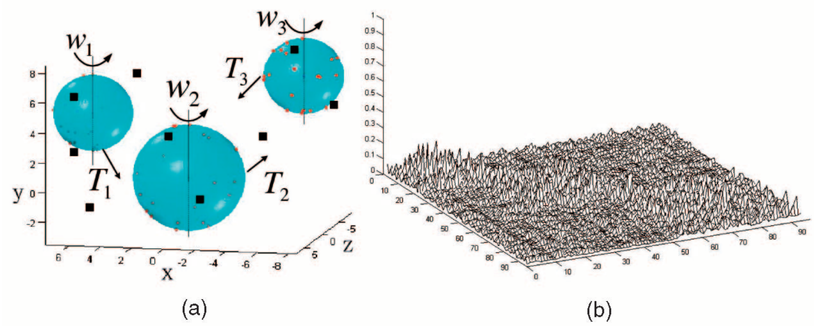


Fig. 6. An example of synthetic correlated motions. (a) A view of the synthetic scene. "□" denote some of the added outliers. (b) The "shape interaction matrix" $\mathbf{Q}$ [3], used by previous algorithms. The number of columns and rows in $\mathbf{Q}$ is equal to the number of feature points. The height represent values of entry. The ideal block-diagonal structure of $\mathbf{Q}$ is lost, which makes it difficult for previous methods to find the correct segmentation of correlated motions.

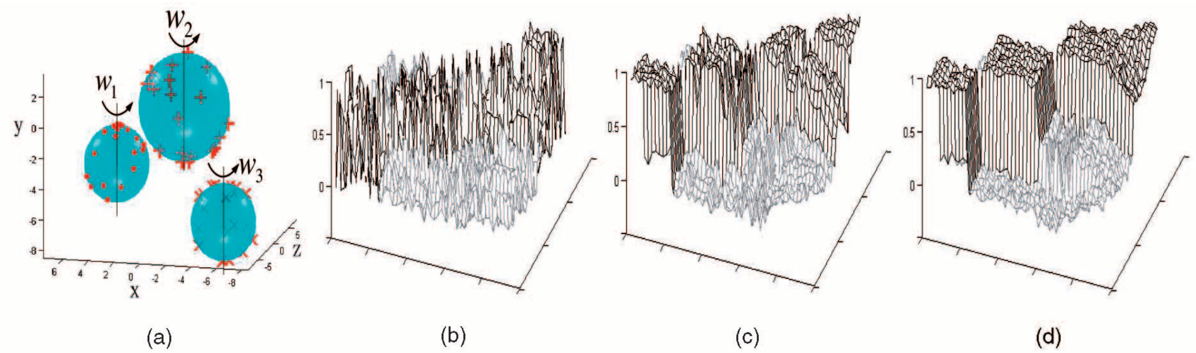


Fig. 7. The segmentation results by applying our method on the synthetic correlated motions depicted in Fig. 6a. (a) By using our method, grouped feature points on three moving spheres. (b), (c), and (d) The sequential changes of the similarity measurement matrix $\mathbf{\Phi}$, which illustrate that the consistency within the same group is effectively enhanced through evolution and voting. The height represent values of entry.

assumes a known and fixed number of the objects, if very few outliers happen to occupy an object identity, they will stick to it and make the number of objects carrying other feature points underestimated. This sensitivity to outliers impairs its robustness.

We find the advantage of the Kanatani's method is that it has a good robustness against noise because of the noise modeling in its framework. In Fig. 5a, it is seen that when there is no outlier, multistage optimization achieves a better performance.

Correlated motion segmentation. We use the example presented in Section 2.1 to demonstrate the efficacy of our method to the problem of correlated motion segmentation (Fig. 6a). Three transparent spheres are generated. Each of them undergoes random translation and different rotation around y-axis. Assuming that the camera's optical axis is along the z-axis and orthographic projection is employed, the motion matrices will have the form of (3). Note that their motions are *not independent*.[4] Fig. 6b shows the "shape interaction matrix" $\mathbf{Q}$ used by previous algorithms, in which the height represents the entry value. It can be seen that the matrix $\mathbf{Q}$ has no apparent block-diagonal structure and, therefore, as described in Section 2, the methods in [2], [6], [3], [10], [29], [12] can hardly find the correct segmentation. This dependence of motions is intentionally introduced to show the effectiveness of our method.

In this experiment, 22, 30, and 42 points from each body are randomly chosen. Seven frames with a resolution of 100×100 pixels are captured and Gaussian noise with 2 pixels of the standard deviation is added. We also imported 100 random, wrong trajectories as outliers. Thus, the ratio of outlier to inlier is 1.06. By using our method, all outliers are correctly discarded (See Fig. 7a and Table 2).

The essential and the most novel feature of our method is the mechanisms of evolution and voting. To demonstrate their effectiveness, Figs. 7b, 7c, and 7d show the changes of the **OF** similarity measurement matrix, $\mathbf{\Phi}$, in the segmentation process, where the three blocks correspond to the three spheres and the height represents the entry value. It can be observed that a low intercluster similarity and a uniformly distributed, high intracluster similarity is obtained, which considerably facilitates the subsequent grouping decision. In contrast, in matrix $\mathbf{Q}$ (See Fig. 6b), *diversities of the intraclass similarities* are present. This nonuniformly distributed intraclass similarity is undesirable in the clustering problem.

Besides the intuition that voting plays an important role here, the following experiment shows that the evolution and the inclusion of the tensor of the components of **OF**, i.e., $\sum_{k=1}^{3} \mathbf{OF}_{jk} \cdot \mathbf{OF}_{jk}^T$, into voting (10) are also necessary for a better performance. See Fig. 8.

There are three moving objects, each containing 20 points. Totally 60 outliers are fabricated into the data. For an object, say the first object, we measure the median value of the similarities between the **OF** s of its feature points and its ideal subspace, as shown in Fig. 8a. We here call our method "Full OF" for simplicity. The four curves are obtained by "Full OF," "Full OF without rotation," "Full OF without moment

4. Please refer to Section 2.1 for the description of the correlations between the sphere motions.

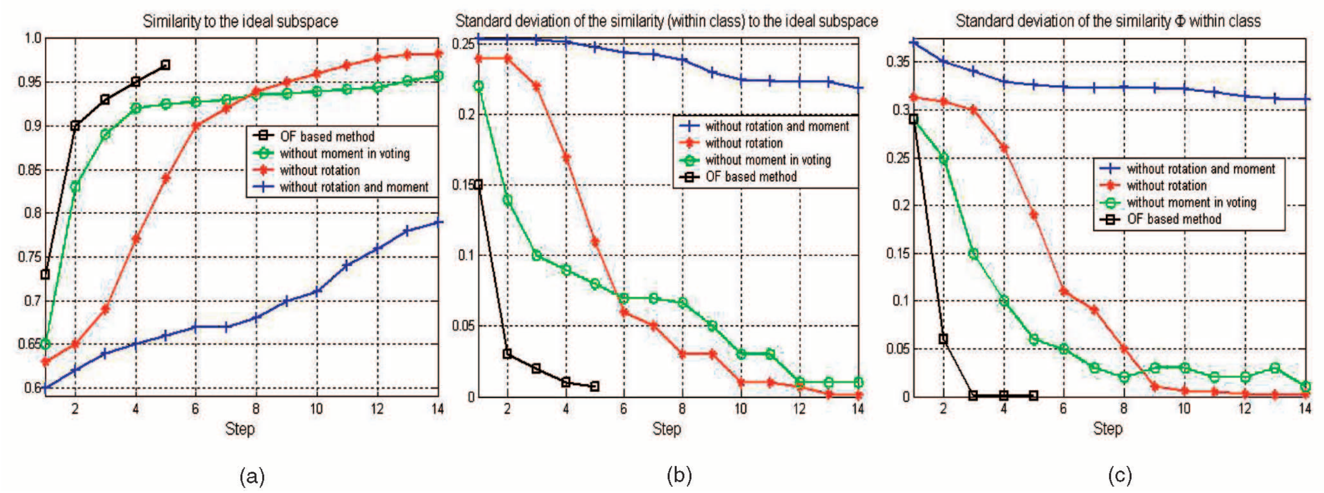


Fig. 8. For the first object, (a) the median value of the similarities between the OFs of its feature points and its ideal subspace. (b) The standard deviation of the similarities between the OFs of its feature points and its ideal subspace. (c) The standard deviation of the intraclass similarity ϕ_{ij}, $i, j \in$ object 1.

calculation in voting," and "Full OF with neither moment calculation nor rotation," respectively.

We also measure the standard deviation of the similarities between the **OF**s of its feature points and its ideal subspace, as shown in Fig. 8b and the standard deviation of the intraclass similarity ϕ_{ij}, $i, j \in$ object 1, as shown in Fig. 8c.

Compared with the curves without rotation, without moment calculation and without both of these two operations, it is seen that the "Full OF" method has a much faster approach to the ideal subspace. In addition, the "Full OF" method also helps a faster elimination of the intraclass variabilities (an undesirable property in the clustering problem). The lower the standard deviation of the intraclass property is, the tighter the intraclass connection is enforced, and the easier the grouping decision can be implemented. Seen from the curves, most of these improvements occur during the first few steps of the "Full OF" method. Though similar status as the "Full OF" method can be achieved without rotation or moment calculation, more steps are needed. The voting (including moment calculation of **OF**) and rotation, being necessary components of our algorithm, have their own contributions to this shortened iteration and the desirable clustering property. In the experiment, we observed that, for inference of a subspace structure, a moderate amount of true inliers belonging to that subspace are need. Normally, occupying 20-25 percent of the total features can produce a good segmentation result for one object.

Using the synthetic correlated motion sequence, Fig. 6a, we compare three algorithms: Zelnik-Manor and Irani's method [30], Kanatani's multistage optimization [13] and our **OF**-based method. Besides the 22, 30, and 42 points chosen from the three spheres, 0, 10, 20, and 30 fabricated outliers are added in turn. Gaussian noise with zero-mean and a standard deviation ranging from 0 to 4 pixels is also added. Following [30], the single, most dominant eigenvector of $\mathbf{W}^T\mathbf{W}$ is used to construct $\hat{\mathbf{Q}}$ for segmentation.[5]

5. Since the first eigenvector is informative for the two-class clustering problem, recursive Normalized Cuts [17] is applied on $\hat{\mathbf{Q}}$ in our experiment for multiclass clustering.

Similar to Fig. 5, we also draw two sets of graphs in Fig. 9 to show the results. In most cases, our algorithm performs much better than multi-stage optimization and Zelnik-Manor and Irani's method, except that multistage optimization has a better performance when there is no outlier.

To see further into the method of [30], for constant block diagonal similarity matrix $\mathbf{W}^T\mathbf{W}$, the first eigenvector, $\mathbf{v}$, has the property that if point i and j belong to the same object, then $\mathbf{v}(i) = \mathbf{v}(j)$ and, thus, $\mathbf{v}(i) - \mathbf{v}(j)$ can characterize the distance between these two points. The method in [30] has to deal with the nonconstant block diagonal matrix. Therefore, they use the exponential function to exaggerate the indicator information for segmentation, i.e., $\exp(-(\mathbf{v}(i) - \mathbf{v}(j))^2)$ is used as the similarity function. However, in the presence of noise, the variations in the dot product recorded in $\mathbf{W}^T\mathbf{W}$ may cause ambiguities in the eigenvectors and the nonuniform intraclass similarities, which leaves [30] open to the instability.

4.2 Real Images

We now demonstrate the applicability of our method in real situations and make comparisons with the discriminant method [10] and Kanatani's multistage optimization [13]. In this section, feature points are detected and tracked by using KLT tracker [18].

Independent motion segmentation. Figs. 10a 10b, and 10c show the segmentation result in three frames from the *flower garden* sequence. One-hundred sixteen features are tracked through 10 frames and 120 false trajectories are created as outliers. In the scene, the background and the tree exhibit distinct motions due to their different distance to the camera. Black "□" in Fig. 10a denote some detected outliers. Compared with the result obtained by the discriminant method, Fig. 10d, and the result of multistage optimization, Fig. 10e, our method achieves a superior performance.

Though the discriminant method is able to discard outliers, the drawback is that once a few outliers happen to be classified as inliers, they will stick to the "object" assigned to them, and will continually attract other feature points, either true inliers or true outliers, into this false

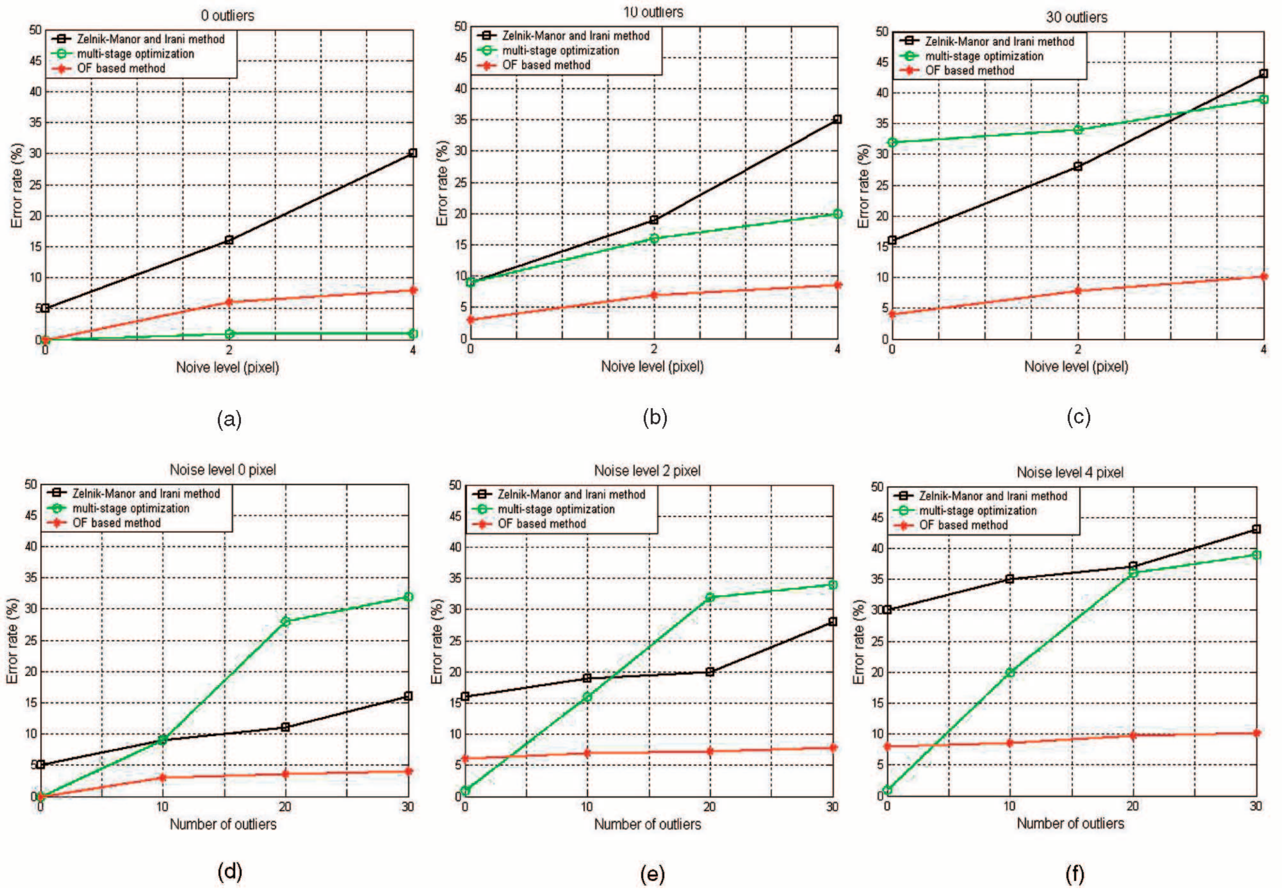


Fig. 9. Three methods, Zelnik-Manor and Irani's [30], Kanatani's multistage optimization [13], and Oriented-Frame-based method, are compared on correlated motion segmentation. (a), (b), and (c) Error rate versus the level of noise with the fixed number of outliers. (d), (e), and (f) Error rate versus the number of outliers with the fixed standard deviation of noise. For each level of noise, the error rate is computed as the average value of 100 independent trials of each method.

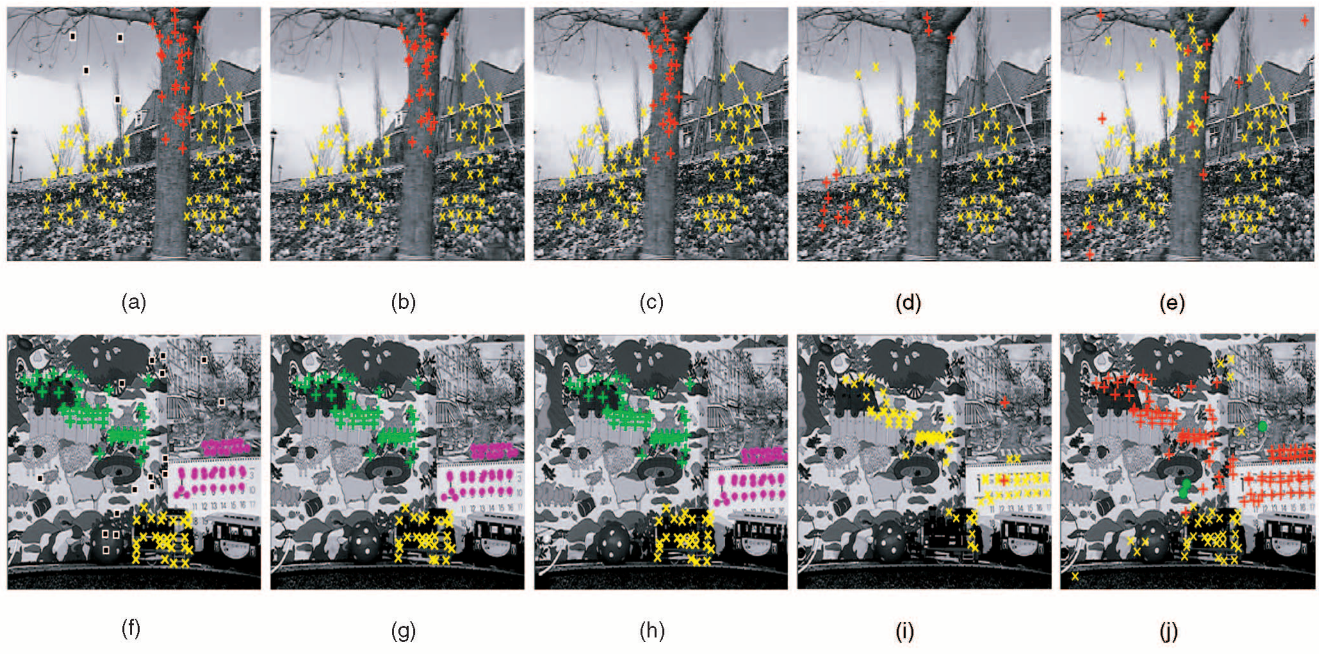


Fig. 10. Comparative results. The segmentation found by our algorithm on *flower garden* sequence are shown in (a), (b), and (c). In the scene, the background and the tree exhibit distinct motions due to their different distance to the camera. The segmentation results on *mobile and calendar* sequence are shown in (f), (g), and (h). Black "□" in (a) and (f) denote some of discarded outliers. For comparison, (d) and (i) are the results of the discriminant method [10] on these two sequences. (e) and (j) Are the results obtained by multistage optimization [13].

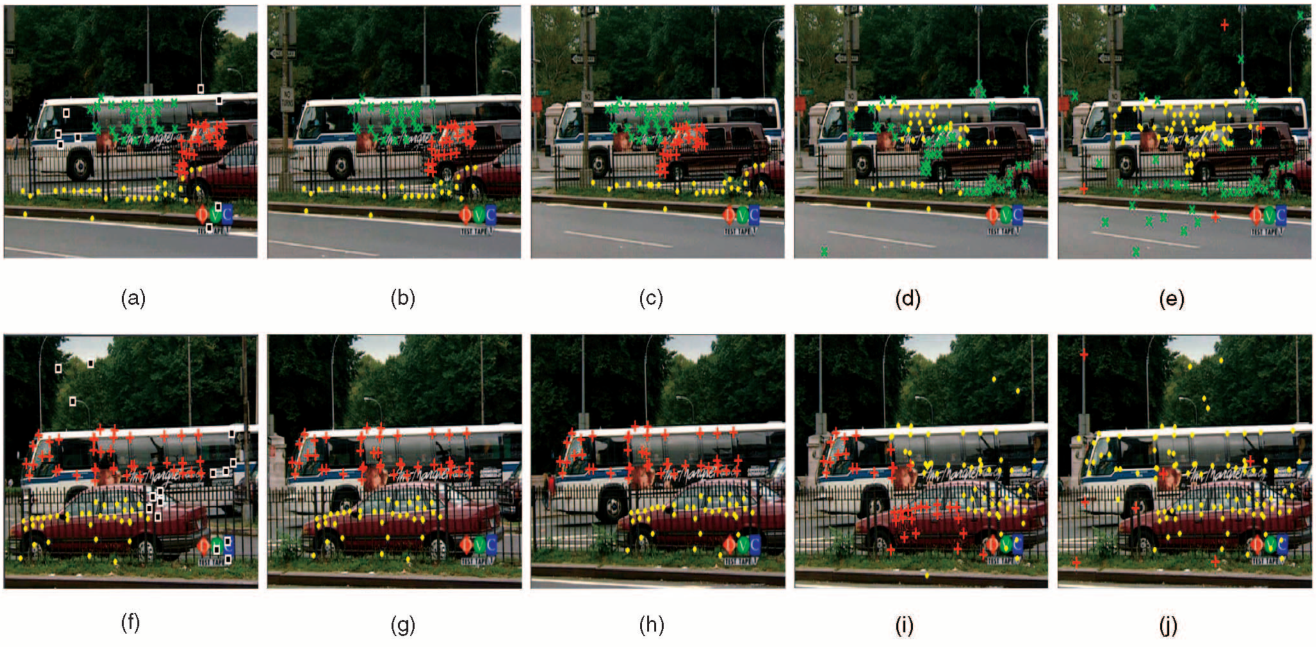


Fig. 11. Comparative results. The result of our algorithm on *bus* sequence is shown in (a), (b), and (c). There are three distinct moving objects: a bus, a van which is moving faster than the bus, and the background. Black "□" in (a) denote some of discarded outliers. (f), (g), and (h) are our result on another sequence. For comparison, (d) and (i) are the results of the discriminant method, (e) and (j) are the result obtained by multistage optimization.

object. The resulting distractions, see Fig. 10d, limits the applicability of this method.

The multistage optimization works when all the data are correct. Moreover, since the number of objects is assumed known and fixed, once a few outliers occupy an object identity, the true feature points will thus have to be classified into fewer objects instead. This leads to low percentage of correct classification. See Fig. 10e.

Figs. 10f, 10g, and 10h show the segmentation result in three frames from the *mobile and calendar* sequence. Thirteen frames are used, which contains 124 corresponding features and 125 added random, wrong trajectories. For comparison, the results given by the discriminant method and the multistage optimization are displayed in Fig. 10i and Fig. 10j, respectively. Their performance degrades due to the distortions caused by outliers.

Figs. 11a, 11b, and 11c show the segmentation result in three frames from the *bus* sequence. There are three distinct moving objects: a bus, a van which is moving faster than the bus, and the background. Thirty frames are used, which contains 115 correct features and 120 introduced random, wrong trajectories. Black "□" in Fig. 11a denote some detected outliers. The results of the discriminant method and multistage optimization are provided in Fig. 11d and Fig. 11e. It is observed that the presence of outliers deteriorates the performance of these two methods.

Figs. 11f, 11g, and 11h show the result of another three frames. A moving camera is viewing a static scene in which a bus is going from right to left. Ninety-three features are tracked through 23 frames. One hundred fake trajectories are fabricated as outliers. Fig. 11i and Fig. 11j are the results by applying the discriminant method and the multistage optimization, respectively.

Correlated motion segmentation. We have hereto performed the experiments on segmentation of multiple independent motions and then we come to show the applicability of our method on multiple correlated motion segmentation.

We choose two articulated motion sequences. One of them contains a whole moving arm and an attached book in the hand. So, there are three articulated moving parts to be segmented: the upper arm, the lower arm, and the book. Denote $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$ as the motion subspaces of these three moving parts. By analyzing the "ground truth" obtained by manually picked true feature trajectories, we found that the motion subspaces of these linked moving parts indeed have intersections and therefore, *are not independent*. In fact, the trajectory of the joint point connecting the upper arm and the lower arm resides in both of their motion subspaces. Specifically, the dimension of the subspace $\mathcal{M}_i \cup \mathcal{M}_j$ (for $i \neq j$ and $i, j = 1, 2, 3$) is higher than the dimension of $\mathcal{M}_i$ $(i = 1, 2, 3)$ but is lower than the sum of them, i.e., $dim(\mathcal{M}_1) < dim(\mathcal{M}_1 \cup \mathcal{M}_2) < dim(\mathcal{M}_1) + dim(\mathcal{M}_2)$. The dimension of $\mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3$ is also lower than the sum of individual $\mathcal{M}_i$ $(i = 1, 2, 3)$. Among the input data of this sequence, there are 47 tracked features and 50 fabricated outliers. Eight frames are captured. Figs. 12a, 12b, and 12c show the segmentation result.

Figs. 12f, 12g, and 12h show another image sequence containing two articulated moving parts: the lower arm and the fingers. The examination on dimensions of the motion subspaces obtained from the "ground truth" indicates that the motions of these two linked parts are also correlated. Thirty-six features are selected and tracked throughout the sequence of eight frames. Extra 50 random, wrong artificial outliers are added. "x" and "+" represent properly classified features of the lower arm and the fingers. Squares

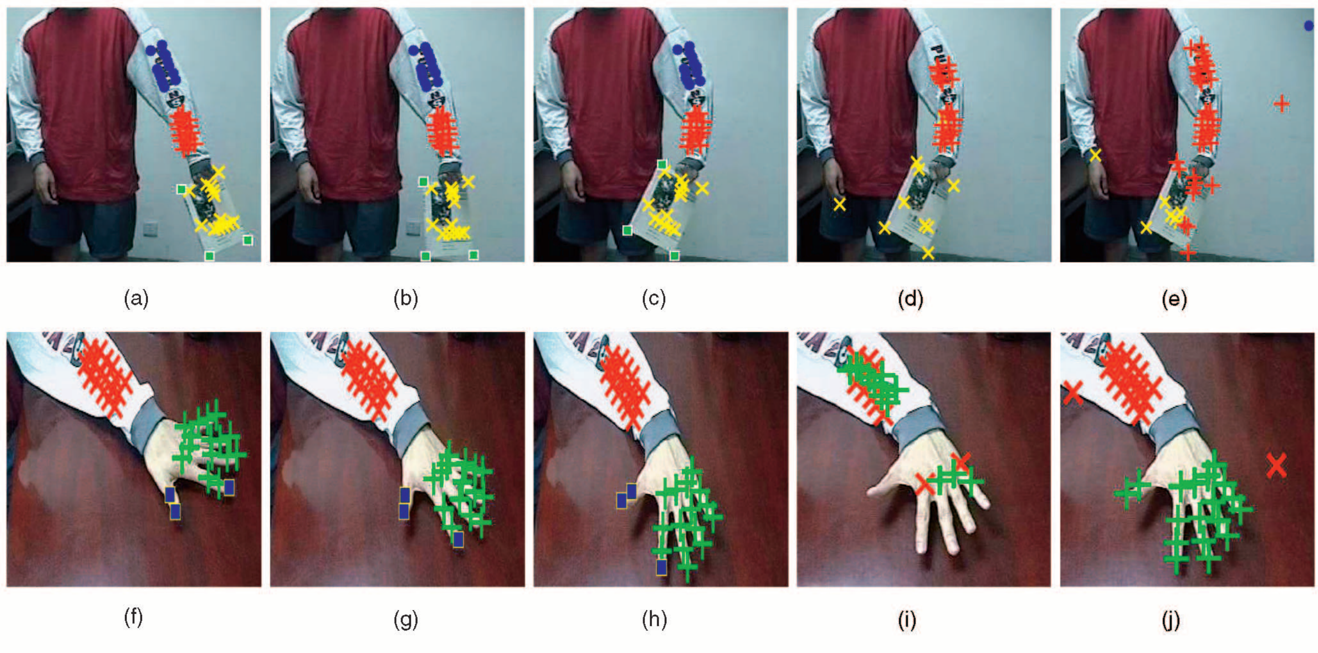


Fig. 12. Comparative results on sequences involving correlated motions (articulated motions). (a), (b), and (c) Results given by our algorithm. For another sequence, (f), (g), and (h) are the results of our algorithm. "□" in (a) and (f) represent some of the discarded outliers. For comparison, (d) and (i) are the results of the discriminant method on these two sequences. (e) and (j) Are the results obtained by multi-stage optimization.

TABLE 3
A Summary of Real Image Sequences Used in the Experiments of Multibody Grouping

		flower garden	*mobile and calendar*	*bus1*	*bus2*	*arm & book*	*arm & fingers*
correct matches		116	124	115	93	47	36
wrong matches		120	125	120	100	50	50
outlier/inlier ratios		1.03	1.01	1.04	1.08	1.06	1.39
our method	correctly classified inliers	110	115	97	78	44	33
	% of correctly classified inliers	94.8%	92.7%	84.3%	83.9%	93.6%	91.7%
	correctly classified outliers	120	122	120	95	50	50
	% of correctly classified outliers	100%	97.6%	100%	95%	100%	100%
discriminant method	correctly classified inliers	82	53	57	51	31	22
	% of correctly classified inliers	70.7%	42.7%	49.6%	54.8%	66.0%	61.1%
	correctly classified outliers	88	62	61	65	37	41
	% of correctly classified outliers	73.3%	49.6%	50.8%	65.0%	74.0%	82.0%
multi-stage optimization	correctly classified inliers	85	83	74	56	25	34
	% of correctly classified inliers	73.3%	66.9%	64.3%	60.2%	53.2%	94.4%
	correctly classified outliers	—					
	% of correctly classified outliers	—					

in Fig. 12f are some detected outliers, which are discarded due to the less structural support they collect from other points or the nonrigid movements of the finger tips.

For these two sequences, Fig. 12d and Fig. 12i are the results of the discriminant method showing its inefficiency in correlated motion segmentation. Fig. 12e and Fig. 12j show the results obtained by multistage optimization, which can segment correlated motions. However, because this method defines a fixed number of the moving objects, if a few outliers wrongly take up an object identity, shown as blue "o" in Fig. 12e, feature points belonging to the three objects have to be classified as two objects instead, denoted by red "+" and yellow "x," respectively. This sensitivity to outliers impairs its performance.

The comparison result is summarized in Table 3.[6] The presence of outliers degrade the performance of multistage optimization [13] and discriminant method [10]. In contrast, because our algorithm has properly designed schemes to discard structurally incompatible outliers, although all the ratios of outlier to inlier in these real image sequences are higher than 1, we can still achieve a *robust performance* and a high percentage of correct classification, especially in outlier rejection.[7] In addition, the two articulated motion sequences well illustrate the applicability of our method in the problem

6. Multistage optimization works when all the data are correct. This method alone has no scheme to detect outliers.

7. When the noise level is low, the theoretically sound methods [13], [24] will yield better results.

of *correlated motion segmentation*. These are two attractive advantages our method can offer in the problem of multibody grouping.

5 Conclusion and Future Work

In this paper, we proposed a novel and effective approach for inference of subspaces in high-dimensional data space. Based on the similarity measurement of subspaces, a generalized mechanism of rotation in high-dimensional space together with the scheme of voting are devised, which will facilitate the emergence of the underlying multiple subspace structure. Inliers and outliers are discriminated effectively due to the data points' structural compatibilities. The proposed Oriented-Frame (**OF**)-based subspace inference technique is a tool for information extraction under subspace constraints with promising robustness and accuracy.

We mainly investigated the applicability of this method to the problem of multibody grouping. Compared with the conventional methods, our approach possesses two attractive advantages, i.e.,

1. Multiple correlated motion segmentation as well as independent motion segmentation.
2. Robust performance against heavy noise and outliers.

The focus of our future work will be handling of missing data, transparent motion segmentation, motion subspaces of different dimensions, and the exploration of the applicability of our algorithm in other subspace constraint problems.

Acknowledgments

The authors would like to thank Dr. C.-K. Tang, Dr. D.Tonnesen, J. Lee, and the anonymous reviewers for their constructive and helpful comments. The author would like to thank Na Yue for the warm help. The authors would also like to acknowledge supports from the Natural Science Foundation of China under grants 60573062, 60205002, 60332010, and 60372020, Natural Science Foundation of Beijing under grant 4042020, and National 863 Hi-Tech Development Program of China under grant 2001AA114190. This work is also supported in part by US National Science Foundation Grants IIS-0347877, IIS-0308222, and Northwestern Murphy Fellowship.

References

[1] A. Bissacco, A. Chiuso, Y. Ma, and S. Soatto, "Recognition of Human Gaits," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 52-57, 2001.
[2] T. Boult and L. Brown, "Factorization-Based Segmentation of Motions," *Proc. IEEE Workshop Visual Motion,* pp. 179-186, 1991.
[3] J. Costeira and T. Kanade, "A Multi-Body Factorization Method for Independent Moving Objects," *Int'l J. Computer Vision,* vol. 29, pp. 159-179, 1998.
[4] M. Fischler and R. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Comm. ACM,* vol. 24, pp. 381-385, 1981.
[5] A.W. Fitzgibbon and A. Zisserman, "Joint Manifold Distance: A New Approach to Appearance Based Clustering," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* 2003.
[6] C.W. Gear, "Multibody Grouping from Motion Images," *Int'l J. Computer Vision*, vol. 29, pp. 133-150, 1998.
[7] G.H. Golub and C.F. Van Loan, *Matrix Computations,* second ed. Baltimore: The John Hopkins Univ. Press, 1989.
[8] G. Guy and G. Medioni, "Inference of Surfaces, 3D Curves, and Junctions from Sparse, Noisy, 3D Data," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 11, pp. 1265-1277, Nov. 1997.
[9] A. Heyden, R. Berthilsson, and G. Sparr, "An Interative Factorization Method for Projective Structure and Motion from Image Sequences," *Image and Vision Computing,* vol. 17, pp. 981-991, 1999.
[10] N. Ichimura, "Motion Segmentation Based on Factorization Method and Discriminant Criterion," *Proc. Int'l Conf. Computer Vision,* vol. 1, pp. 600-605, 1999.
[11] M. Irani, "Multi-Frame Correspondence Estimation Using Subspace Constraints," *Int'l J. Computer Vision*, vol. 48, no. 3, pp. 173-194, 2002.
[12] K. Kanatani, "Motion Segmentation by Subspace Separation and Model Selection," *Proc. Int'l Conf. Computer Vision,* pp. 301-306, 2001.
[13] K. Kanatani and Y. Sugaya, "Multi-Stage Optimization for Multibody Motion Segmentation," *Proc. Australia-Japan Advanced Workshop Computer Vision,* 2003.
[14] J. Lee, J. Wang, C. Zhang, and Z. Bian, "Probabilistic Tangent Subspace: A Unified View," *Proc. Int'l Conf. Machine Learning,* 2004.
[15] R.J. Martin, "A Metric for ARMA Processes," *IEEE Trans. Signal Processing,* vol. 48, no. 4, pp. 1164-1170, 2000.
[16] P. Rousseeuw and A. Leroy, *Robust Regression and Outlier Detection.* New York: John Wiley & Sons, 1987.
[17] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, pp. 888-905, 2000.
[18] J. Shi and C. Tomasi, "Good Feature to Track," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 593-600, 1994.
[19] P. Simard, Y. LeCun, and J. Denker, "Efficient Pattern Recognition Using a New Transformation Distance" *Proc. Neural Information Processing Systems Conf.,* 1993.
[20] S. Soatto and P. Perona, "Recursive 3-D Visual Motion Estimation Using Subspace Constraints," *Int'l. J. Computer Vision,* vol. 22, pp. 235-259, 1997.
[21] C.-K. Tang, G. Medioni, and M.-S. Lee, "N-Dimensional Tensor Voting and Application to Epipolar Geometry Estimation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 8, pp. 829-844, Aug. 2001.
[22] M. Tipping and C. Bishop, "Mixtures of Probabilistic Principal Component Analyzers," *Neural Computation,* vol. 11, no. 2, 1999.
[23] C. Tomasi and T. Kanade, "Shape and Motion from Image Streams under Orthography: A Factorization Method," *Int'l J. Computer Vision,* vol. 9, pp. 137-154, 1992.
[24] R. Vidal and R. Hartley, "Motion Segmentation with Missing Data Using Power Factorization and GPCA," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 310-316, 2004.
[25] R. Vidal, Y. Ma, and S. Sastry, "Generalized Principal Component Analysis (GPCA)," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 621-628, 2003.
[26] J. Wang and E.H. Adelson, "Layered Representation for Motion Analysis," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 361-366, 1993.
[27] Y. Weiss, "Smoothness in Layers: Motion Estimation Using Nonparametric Mixture Estimation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 520-526, 1997.
[28] L. Wolf and A. Shashua, "Kernel Principal Angles for Classification Machines with Applications to Image Sequence Interpretation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 635-640, 2003.
[29] Y. Wu, Z.-Y. Zhang, T.S. Huang, and J.Y. Lin, "Multibody Grouping via Orthogonal Subspace Decomposition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 252-257, 2001.
[30] L. Zelnik-Manor and M. Irani, "Degeneracies, Dependencies and Their Implications in Multi-Body and Multi-Sequence Factorizations," *Proc. IEEE Conf. Computer Vision and Pattern Recognition* pp. 287-293, 2003.

Zhimin Fan received the BS and MS degrees with honors both from the Automation Department, Tsinghua University, China, in 2001 and 2004, respectively. He is now a PhD student in the Electrical and Computer Engineering Department, Northwestern University, Evanston, Illinois. His research interests include computer vision, pattern recognition, and image processing. He is a student member of the IEEE Computer Society.

Jie Zhou (M'01, SM'04) received the BS and MS degrees both from the Department of Mathematics, Nankai University, Tianjin, China, in 1990 and 1992, respectively. He received the PhD degree from the Institute of Pattern Recognition and Artificial Intelligence, Huazhong University of Science & Technology (HUST), Wuhan, China, in 1995. From 1995 to 1997, he served as a postdoctoral fellow in the Department of Automation, Tsinghua University, Beijing, China. Currently, he is a professor and assistant chair in the Department of Automation, Tsinghua University. His research areas include computer vision, pattern recognition, information fusion, and image processing. In recent years, he has authored more than 10 papers in international journals and more than 40 papers in international conferences. He is an associate editor for the *International Journal of Robotics and Automation*. He received the Best Doctoral Thesis Award from HUST, the First Class Science and Technology Progress Award from Ministry of Education, China, and the Excellent Young Faculty Award from Tsinghua University in 1995, 1998, and 2003, respectively. He is a senior member of the IEEE.

Ying Wu (M'01) received the BS degree from the Huazhong University of Science and Technology, Wuhan, China, in 1994, the MS degree from Tsinghua University, Beijing, China, in 1997, and the PhD degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign (UIUC), Urbana, Illinois, in 2001. From 1997 to 2001, he was a research assistant at the Beckman Institute for Advanced Science and Technology at UIUC. During the summers of 1999 and 2000, he was a research intern with Microsoft Research, Redmond, Washington. Since 2001, he has been an assistant professor in the Department of Electrical and Computer Engineering of Northwestern University, Evanston, Illinois. His current research interests include computer vision, computer graphics, machine learning, multimedia, the human-computer interaction. He is an associate editor of *SPIE Journal of Electronic Imaging*. He received the Robert T. Chien Award at UIUC in 2001, and is a recipient of the US National Science Foundation CAREER award. He is a member of the IEEE.