Inference of Multiple Subspaces from High-Dimensional Data and Application to Multibody Grouping[*]

Zhimin Fan
Department of Automation
Tsinghua University, China
fzm01@mails.tsinghua.edu.cn

Jie Zhou
Department of Automation
Tsinghua University, China
jzhou@tsinghua.edu.cn

Ying Wu
Department of ECE
Northwestern University, IL
yingwu@ece.northwestern.edu

Abstract

Multibody grouping is a representative of applying subspace constraints in computer vision tasks. Under linear projection models, feature points of multibody reside in multiple subspaces. We formulate the problem of multibody grouping as multiple subspace inference from high-dimensional data space. The theoretical value and practical advantage of this formulation come from the relaxation of the motion independency assumption which has to be enforced in most factorization based methods. In the proposed method, an Oriented-Frame (OF), which is a multi-dimensional coordinate frame, is associated with each data point indicating the point's preferred subspace structure. Then, a similarity measurement of these OFs is introduced and a novel mechanism is devised for conveying the information of the inherent subspace structure among the data points. In contrast to the existing factorization-based algorithms that can not find correct segmentation of correlated motions such as articulated motion, the proposed method can robustly handle motion segmentation of both independent and correlated cases. Results on controlled and real experiments show the effectiveness of the proposed subspace inference method.

1. Introduction

Most realistic vision tasks involve multibody motions. Among many proposed segmentation techniques, factorization method, which was originally developed by Tomasi and Kanade [1] for structure from motion of a single object, is particularly interesting. It is revealed that under linear projection models, points trajectories of a single body over a sequence of frames 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 by enforcing the constraint that the trajectory subspaces (defined in Section 2), spanned by objects' feature trajectories, are independent: $\mathcal{T}_p \cap \mathcal{T}_q = Null$ for $\forall p \neq q$, where $\mathcal{T}_p$ corresponds to trajectory subspace of object p. Gear [2] formulated the problem as weighted graph matching. Costeira and Kanade [3] presented a multibody factorization method in which a shape interaction matrix $\mathbf{Q}$ is introduced. If any features i and j are from different objects, $\mathcal{Q}_{ij}$ will be zero, otherwise, non-zero. Several extensions have also been made [4-6].

However, the problem of motion segmentation with correlated motions, such as segmenting an articulated structure, is left unaddressed. For example, a simple scenario of a moving arm involves two dependently moving objects, the upper arm and the lower arm. Here, the dependence of motions is revealed by the intersections of trajectory subspaces, i.e. $\mathcal{T}_p \cap \mathcal{T}_q \neq Null$, $\exists p \neq q$, where $\mathcal{T}_p$ corresponds to trajectory subspace of object p. In the presence of multiple correlated motions, the existing methods [2-6] cannot find the correct segmentation for the reason as explained in Section 2. In [7], Zelnik-Manor and Irani described the problem of correlation (or dependency) between motions and suggest an eigenvector segmentation algorithm.

In this paper, we treat the segmentation of multibody as a problem of multiple subspaces inference: To grasp the inherent multiple subspace structure within a high-dimensional data space and classify each data point into its own subspace no matter whether these subspaces are independent or correlated.

Firstly, referring to the tensor representation, each input point is associated with a multi-dimensional coordinate frame which we call Orientated-Frame (**OF**), encoding the most preferred subspace of each point. Then, after defining a similarity measurement of these **OF**s, data points communicate with each other in the process of evolution and voting about the information of multiple subspace structure. For any two points in the same subspace, their consistency of **OF**s can be enforced. Therefore, outliers are effectively rejected due to their incompatibilities with the multiple subspace structure. The strength of majority inliers is accumulated to overcome the interference of noise and outliers resulting in a more robust performance. Exact multiple sub-

space structure may emerge from the set of filtered inliers. The process of classifying data points to these subspaces is equivalent to multibody grouping, in which the independence of the subspaces (the independence of motions) is not assumed. So, the algorithm can segment correlated motions as well as independent motions.

In Section 2, the problem of multibody grouping and previous work on subspace inference are briefly reviewed. Section 3 presents the multiple subspaces inference technique with its application to multibody grouping. Section 4 provides the experimental results on both synthetic and real images. Feature points were detected and tracked by using KLT tracker [8]. Section 5 presents conclusions.

2. Background

2.1. The problem of multibody grouping

Suppose there are m moving objects in the scene, each object contains p_i 3D 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 $\mathbf{W}_i$,

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

where

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

$\mathbf{M}_i$ is a $2F \times 4$ matrix and $\mathbf{M}_{ji}$ ($j = 1, \ldots, F$) is the 2×4 projection matrix related to object i in the jth frame.

The columns of $\mathbf{W}_i$ reside in a 4D *trajectory subspace* $\mathcal{T}_i$ spanned by the columns of $\mathbf{M}_i$. In fact, these columns lie in a 3D manifold when they are considered to be points in the $2F$-D space. Let $\mathbf{r}_p$ and $\mathbf{r}_q$ denote the positions of any two columns of $\mathbf{W}_i$. 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 3 columns of $\mathbf{M}_i$, because the last row of $\mathbf{S}_i$ is all 1's. Let $\mathcal{W}_i$ represent this 3D *motion subspace* formed by vectors $\mathbf{r}_{pq}$ ($p, q \in [1, p_i]$, $p \neq q$).

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

$$\begin{aligned} \mathbf{W} &= [\mathbf{W}_1 \mathbf{W}_2 \ldots \mathbf{W}_m] = \mathbf{M}\mathbf{S} \\ &= [\mathbf{M}_1 \mathbf{M}_2 \ldots \mathbf{M}_m] \begin{bmatrix} \mathbf{S}_1 & & & \\ & \mathbf{S}_2 & & \\ & & \cdots & \\ & & & \mathbf{S}_m \end{bmatrix}, \end{aligned} \tag{3}$$

where $P = \sum_{i=1}^m p_i$ is the total number of feature points.

The block diagonal form of $\mathbf{S}$ is the basic assumption of previous factorization based methods. If the motions are not independent, trajectory subspaces spanned by columns of $\mathbf{M}_i$s will have intersections, destroying the block diagonal form of $\mathbf{S}$. Thus, both the reduced row echelon form in [2] and the nice property of $\mathbf{Q}$ in [3-6] will vanish. Therefore, those algorithms wouldn't find the correct segmentation.

2.2. Previous work on subspace inference

Actually, multibody grouping is one of the representatives of applying subspace constraints in computer vision tasks. These constraints show that some concerned information compactly resides in some low-dimensional subspaces.

Inference of surfaces and curves in 3D vision can be considered as the exploitation of low-dimensional subspace structures in the 3D space. Szeliski and Tonnesen [9] have proposed an oriented particle system associating each particle an oriented trihedral coordinate frame. Interaction potentials are designed which favor meaningful structural arrangements of the particles. Guy and Medioni [10] used tensor voting to obtain 3D structural information, which is achieved by a nonlinear voting process.

Beyond the 3D space, subspace constraints also exist in various vision problems involving higher-dimensional data structure. Irani [11] showed that multi-frame subspace constraints can be used for constraining the 2D correspondence estimation. Huynh and Heyden [12] utilized a motion's 4D subspace constraint for outlier detection. Tang et.al. [13] formulated the problem of epipolar geometry estimation as inferring hyperplane (a 7D manifold) in an 8D space analogous to plane detection in a 3D space. They extended the idea of tensor voting [10] to N-D and achieved robust performance in hypersurface and hyperjunction detection.

Besides [13], 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. The main difficulty stems from the unavoidable outliers inherent in the data set. If significant portion of data is corrupted by noise, the detection of subspaces will be difficult and the result will be inexact, which will severely damage the efficiency of applying subspace constraints to practical applications. One recent work is Generalized Principal Component Analysis (GPCA), a geometric method proposed by Vidal et al. [14]. GPCA shows a sound theoretical approach to the identification of mixture of subspaces

in noise-free case. However, how to embed techniques to enhance the performance of GPCA when large amount of noise presents remains an open issue.

3. Multiple subspaces inference and applications to multibody grouping

In this Section, we present a novel technique for multiple subspaces inference and apply it to multibody grouping. The data input is a $2F{\times}P$ matrix $\mathbf{W}$. Each data point is a $2F$-D vector denoted by $\mathbf{r}_i$ ($i=1,\ldots,P$). The number of objects is unknown. Our purpose is to extract multiple motion subspaces $\mathcal{W}_k$s out of $\mathbf{W}$. Each $\mathcal{W}_k$ is a 3D subspace formed by the vectors $\mathbf{r}_{ij}=\mathbf{r}_i-\mathbf{r}_j$ ($i,j \in$ Object$_k$, $i{\neq}j$), or equivalently, the subspace spanned by the first 3 columns of $\mathbf{M}_k$. Multibody grouping is equivalently achieved by classifying $\mathbf{r}_i$ to these motion subspaces $\mathcal{W}_k$s.

3.1. Conversion to Oriented-Frames

There is a need to give each point a configuration which would facilitate the communication of structure information between data points.

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 $\mathcal{W}_k$. 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 subspace, r_{ij} being the norm of $\mathbf{r}_{ij}$. We use a Gaussian function to weigh the strength of these contributions. For the problem of multibody grouping, this decaying function is meaningful since feature points' trajectories of the same object are always closer than those of different objects. For practical purposes, if the distance between two points is greater than $2\sigma_d$, their mutual influence is ignored. We take $\sigma_d=0.3{\times}r_{\mathrm{med}}$, where r_{med} is the median value of all r_{ij} for $i{\neq}j$.

These votes are aggregated as the second order moment collection (second order symmetric tensor),

$$\mathbf{O}_i = \sum_{r_{ij}<2\sigma_d,j\neq i} \exp(-r_{ij}^2/\sigma_d^2) \cdot \hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{r}}_{ij}^T. \qquad (4)$$

Let $\lambda_1{\geq}\lambda_2{\geq}\cdots{\geq}\lambda_{2F}$ represent the sorted eigenvalues of $\mathbf{O}_i$ and $\mathbf{V}_1$, $\mathbf{V}_1,\ldots,\mathbf{V}_{2F}$ be the corresponding eigenvectors. Because the eigenvalues describe the strength and agreement measures on the corresponding directions, we choose 3 dominant principal axes to build up the preferred 3D motion subspace configuration of the current point. Thus, each data point i will be associated with a $\mathbf{OF}_i=\{\mathbf{OF}_{i1},\mathbf{OF}_{i2},\mathbf{OF}_{i3}\}$ consisting of the 3 dominant eigenvectors of $\mathbf{O}_i$, called an *Oriented-Frame*.

3.2. Evolution of Oriented-Frames

OFs of points in the same motion subspace are expected to have similar configurations. However, the initial **OF**s may not be accurate due to ambiguities caused by noise and outliers. A similarity measurement is introduced to
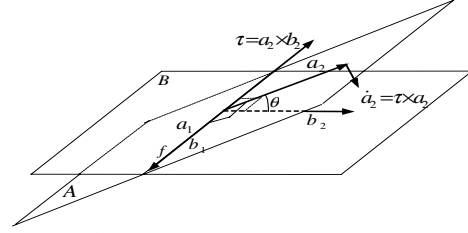

Figure 1: Rotation of plane A to B

compare these **OF**s and a novel mechanism for rotating subspace is proposed to enhance the consistency among them.

3.2.1. Similarity measurement for subspace comparison. This subspace similarity measurement is derived from the concept of principal angles and principal vectors [15].

Definition 1. Let $\mathcal{A}$ and $\mathcal{B}$ be two p-D subspaces in an l-D space. $\mathbf{A}$ and $\mathbf{B}$ are $l{\times}p$ matrices consisting of orthonormal bases of $\mathcal{A}$ and $\mathcal{B}$. The principal angles $0{\leq}\theta_1{\leq}\cdots{\leq}\theta_p{\leq}\pi/2$ and principal vectors $\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}{\in}\mathcal{A}$, $\{\mathbf{v}_1,\ldots,\mathbf{v}_p\}{\in}\mathcal{B}$ 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

$$\begin{cases} [\mathbf{u}_1,\ldots,\mathbf{u}_p] &= \mathbf{AY}, \\ [\mathbf{v}_1,\ldots,\mathbf{v}_p] &= \mathbf{BZ}, \\ \cos(\theta_k) &= \sigma_k, \quad k=1,\ldots,p. \end{cases} \qquad (5)$$

Any arbitrary orthonormal bases in matrices $\mathbf{A}$ and $\mathbf{B}$ can be aligned by the SVD. In this sense, we can define the similarity measurement between subspaces $\mathcal{A}$ and $\mathcal{B}$ as

$$\phi(\mathcal{A},\mathcal{B}) = \prod_{k=1}^{p} \cos(\theta_k). \qquad (6)$$

It is obvious that identical subspaces have the maximum similarity measure of value 1.

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

$[\mathbf{u}_1,\ldots,\mathbf{u}_p]^T[\mathbf{v}_1,\ldots,\mathbf{v}_p] = \mathrm{diag}[(\cos(\theta_k))]$ indicates that vector $\mathbf{u}_i$ is orthogonal to vector $\mathbf{v}_j$ ($j{\neq}i$).

Moreover, 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 $\mathbf{u}_j$ ($j{\neq}i$). Similarly, $\mathbf{v}_i$ 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. The mechanism for subspace rotation is used to enhance the consistency among points whose **OF**s have high similarity measurement. We begin with a simple example in the 3D space for ease of visualization. Consider two non-parallel

planes *A* and *B*, they intersect at line *f*. How to rotate plane *A* to *B*, until these two planes overlap? See Fig. 1

Suppose that $\{\mathbf{a}_1,\mathbf{a}_2\}$ and $\{\mathbf{b}_1,\mathbf{b}_2\}$ are principal vectors of this subspace pair corresponding to principal angles 0 and $\cos^{-1}(\|\mathbf{a}_2^T\mathbf{b}_2\|)$, respectively, then rotating plane *A* to *B* is now equivalent to rotating vector $\mathbf{a}_2$ towards $\mathbf{b}_2$. 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 Eq.(6).

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 $\mathbf{a}_2$ towards $\mathbf{b}_2$, 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 $\mathbf{a}_2$.

The cross product of two 3D vectors may also be viewed as the *orthogonal complement* of the subspace spanned by the concerned two vectors. Inspired by this idea, rotating $\mathbf{a}_2$ towards $\mathbf{b}_2$ requires two steps. First, $\mathcal{S}_{ab}^{(2)\perp}$ is calculated by cross product of $\mathbf{a}_2$ and $\mathbf{b}_2$ as the so called "torque". Secondly, $\dot{\mathbf{a}}_2$ 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 $\mathbf{a}_2$, which also indicates that $\dot{\mathbf{a}}_2$ is restricted in the subspace $\mathcal{S}_{ab}^{(2)}$.

From Observation 2, the mechanism for rotating an *N*-D vector **a** towards another *N*-D vector **b** narrowing the angle $\alpha=\cos^{-1}(\|\mathbf{a}^T\mathbf{b}\|)$ can be deduced. Let $\mathcal{S}_{ab}$ denote the subspace spanned by **a** and **b**. First, the orthogonal complement of $\mathcal{S}_{ab}$, denoted by $\mathcal{S}_{ab}^\perp$, is calculated. Comparing with 3D case, $\mathcal{S}_{ab}^\perp$ can be considered as a "*N*-D torque". Secondly, the orthogonal complement of subspace spanned by $\mathcal{S}_{ab}^\perp$ and **a** is computed and denoted as $\mathbf{a}_r$, which can be regarded as the "high-dimensional cross product" of "*N*-D torque" and **a** analogously in 3D case. Then the direction of the instantaneous displacement of **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 **b** is smaller than that between **a** and **b**, where μ is a small scalar controlling the magnitude of rotation ($\mu=0.01$).

Then, let us consider the rotation of *p*-D subspaces $\mathcal{A}$ towards $\mathcal{B}$ in *N*-D space. Similarly, given the corresponding principal angles and principal vectors $\{\mathbf{u}_k,\mathbf{v}_k,\theta_k,k=1,\ldots,p\}$, we can find the instantaneous change, $\dot{\mathbf{u}}_i$, for rotating $\mathbf{u}_i$ towards $\mathbf{v}_i$ 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 $\{\mathbf{u}_1,\ldots,\mathbf{u}_{i-1},\mathbf{u}_i+\mu\dot{\mathbf{u}}_i,\mathbf{u}_{i+1},\ldots,\mathbf{u}_p\}$, respectively. Since $\dot{\mathbf{u}}_i$ is restricted in $\mathcal{S}_{uv}^{(i)}$ (Observation 2), we have $\mathcal{S}_{uv}^{(i)}\equiv\mathcal{S}_{uv}^{(i')}$, which means after the rotation, the structure of mutually orthogonal subspaces spanned by $\{\mathbf{u}_i,\mathbf{v}_i\}$ ($i=1,\ldots,p$) is 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 $\mathcal{B}$ (Observation 1). Since no impact is made on other principal vector pairs, rotating $\mathbf{u}_i$

towards $\mathbf{v}_i$ results in that only the ith principal angle decreases and the other principal angles remain unchanged.

This nice property indicates that the process of rotating a *p*-D subspace to another can be divided into *p* steps by rotating $\mathbf{u}_1$ to $\mathbf{v}_1$, $\mathbf{u}_2$ to $\mathbf{v}_2$, ..., and $\mathbf{u}_p$ to $\mathbf{v}_p$ successively. Through this manipulation, the two subspaces can gradually become identical in the sense of having the maximum similarity measure of value 1 according to Eq. (6).

3.2.3. Evolution of Oriented-Frames. Points with similar **OF**s can then rotate their **OF**s to enhance the saliency of their motion subspace structure.

We define a similarity measurement matrix of **OF** as

$$\Phi = \{\phi_{(i,j)} : \phi_{(i,j)} = \prod_{k=1}^{3} \cos(\theta_k), \forall i,j \leq P\}. \quad (7)$$

where θ_k (k=1,2,3) are principal angels of subspace pair spanned by $\mathbf{OF}_i$ and $\mathbf{OF}_j$. In our experiment, if $\phi_{(i,j)} >0.7$, we let $\mathbf{OF}_i$ rotate towards $\mathbf{OF}_j$ to obtain a greater $\phi_{(i,j)}$. **OF**s of all data points and Φ are thus updated simultaneously and consistency among same motion subspace can be enhanced.

3.3. Voting between data points

The voting stage comes after the evolution. For the inference of the underlying motion subspace structure of $\mathbf{r}_i$, only using the spatial relationship $\mathbf{r}_{ij}$ (Section 3.1) is not sufficient. The 3 vectors of $\mathbf{OF}_j$, which build up the potential underlying motion subspace of $\mathbf{r}_j$, could also be utilized to propagate the jth point's structural information to other points. Furthermore, if $\phi_{(i,j)}>\phi_{(i,k)}$, the vote from point j to i should be more reliable than the one from k to i.

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

$$\mathbf{O}_i = \sum_{r_{ij}<2\sigma_d,j\neq i} \phi_{(i,j)}\exp(-r_{ij}^2/\sigma_d^2)\cdot \\ \left(\hat{\mathbf{r}}_{ij}\cdot\hat{\mathbf{r}}_{ij}^T + \sum_{k=1}^{3} \mathbf{OF}_{jk}\cdot\mathbf{OF}_{jk}^T\right). \quad (8)$$

Thus, the information involved in the voting process is apparently enriched and additional efficiency is brought to the communication between data points.

3.4. Outlier rejection and subspaces inference

After the evolution and voting, consistency among points with similar **OF**s is enhanced. Because outliers are unlikely to accidentally form salient structured arrangements. So, if a point i, its $\mathbf{OF}_i$ receives little structural agreement from others, i.e. $\sum_{r_{ij}<2\sigma_d,j\neq i} \phi_{(i,j)} < \phi_{\text{thres}}$, point i can be isolated as an outlier. The threshold ϕ_{thres} has no significant effect on the filtered result if kept within a reasonable range.

To improve accuracy, we need to run several passes of evolution and voting stages to filter out outliers progressively. Typically, only four to five passes are needed.
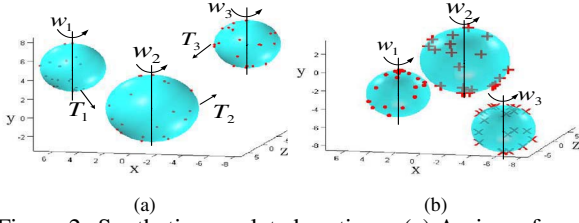
Figure 2: Synthetic correlated motions. (a) A view of synthetic scene. (b) Detected feature points on three moving spheres are shown by "x", "+" and "o", respectively.

In our extensive experiments, the $\phi_{(i,j)}$ is mostly close to 1 for pair of points i and j in the same subspace. This property considerably facilitates the grouping decision. If $\phi_{(i,j)} > 0.99$, 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. And the underlying 3D motion subspace configuration $\mathcal{W}_k$s can be obtained by spanning the 3 dominant eigenvectors of this tensors. Then, data points can be checked against the inferred multiple motion subspaces, producing a set of grouped inliers. For the problem of multibody grouping, both segmentation of feature points and the number of moving objects can be discovered simultaneously.

4. Experimental results

Robust performance of our algorithm on both synthetic and real data containing multiple independent motions and correlated motions are shown in this section.

4.1. Synthetic data

We carried out two experiments on synthetic data. One of them is correlated motion segmentation. The other shows our method's robustness against noise and outliers.

In the first experiment, 22, 30 and 42 points are randomly chosen from three transparent spheres as shown in Fig. 2(a). 7 frames with resolution 100×100 are captured. Gaussian noise with 2 pixels of the standard deviation is added and 100 false trajectories are randomly fabricated into the data matrix $\mathbf{W}$. Each sphere undergoes random translation and different rotation around y-axis. Using orthographic projection, the motion matrices will have the form

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \cdots \\ \mathbf{M}_F \end{bmatrix} = \begin{bmatrix} R_{11}^{(1)} & 0 & R_{13}^{(1)} & t_x^{(1)} \\ 0 & 1 & 0 & t_y^{(1)} \\ R_{11}^{(2)} & 0 & R_{13}^{(2)} & t_x^{(2)} \\ 0 & 1 & 0 & t_y^{(2)} \\ & \cdots & & \cdots \\ R_{11}^{(F)} & 0 & R_{13}^{(F)} & t_x^{(F)} \\ 0 & 1 & 0 & t_y^{(F)} \end{bmatrix}. \quad (9)$$

Note that the second column of $\mathbf{M}$ are same for all spheres, so, their motions are *correlated*. i.e. The motion subspaces spanned by the first 3 columns of $\mathbf{M}$s have intersections. Therefore, the methods in [2-6] cannot find
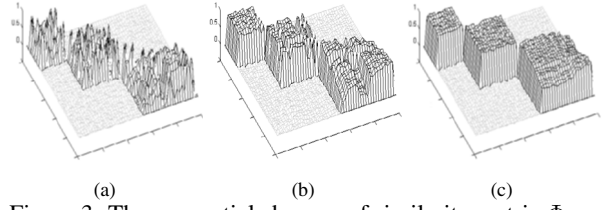

Figure 3: The sequential changes of similarity matrix Φ are shown from (a) to (c). These illustrate that the consistency within the same group is effectively enhanced through evolution and voting
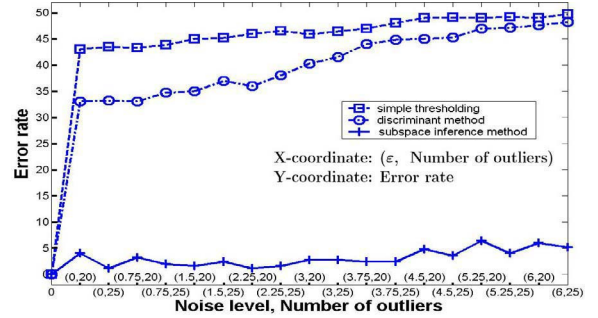

Figure 4: Comparison of the error rate.

the correct segmentation. While by using our method, all outliers are correctly discarded and 20 out of 22 points of sphere1, 30 out of 30 points of sphere2, 42 out of 42 points of sphere3 are properly classified as shown in Fig. 2(b).

The essential character of our method is the mechanism of evolution and the scheme of voting. To demonstrate their effectiveness, Fig. 3 shows the change of **OF** similarity matrix Φ. These are the forms of Φ after being thresholded and grouped. The three blocks correspond to three spheres.

The second experiment tested our method's robustness to noise and outliers. Comparison is made by applying the simple thresholding method and the discriminant method similar to [4] to the same data set. There are 2 arbitrarily moving objects, each has 25 feature points. In the presence of 20 or 25 fabricated outliers and a zero-mean Gaussian noise with a standard deviation ε ranging from 0 to 6 pixels added to the coordinates of feature points, Fig 4 plots the results of the three methods. The first part of the X-coordinate is ε. The second part of the X-coordinate is the number of outliers (NOutlier). The Y-coordinate is the average error rate of 10 trials for each pair of (ε, NOutlier). From this comparison, it shows that our algorithm, which works up to $\varepsilon=6$ (pixels), performs the best and is robust to noise and outliers.

4.2. Real images

In this Section, random wrong outliers are added and all the outlier to inlier ratios in the real data are greater than one. Fig. 5(a)-(c) show the scene containing a moving background and a moving vehicle. 68 tracked features within 8
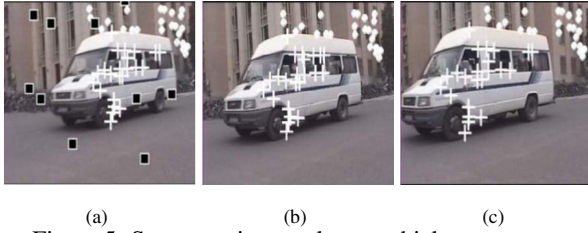
(a)   (b)   (c)

Figure 5: Segmentation results on vehicle sequences.

frames are used and 70 fake trajectories are imported. The detected two moving groups are shown by "o" and "+". Black "□" in (a) denote some detected outliers.

The algorithm is also applied on articulated motion sequences. By analyzing the "ground truth" obtained by manually picked true feature trajectories, the motion matrices **M**s of the linked moving parts indeed have intersections and are *correlated*. i.e motion subspaces $\mathcal{W}_i \cap \mathcal{W}_j \neq Null$. Fig. 6(a)-(c) show the segmentation result on a sequence of 8 frames, which contains 47 tracked features and 50 fabricated outliers. Correctly grouped features of the upper arm, lower arm and book are shown by "o", "+" and "x", respectively. Squares in (a) denote some outliers. Fig. 6(d)-(f) show the result on another sequence of 8 frames, which contains 36 tracked features and 50 random, wrong outliers. "x" and "+" represent properly classified features of the lower arm and fingers. Squares in (d) are some detected outliers.

5. Conclusion

In this paper, attempts have been made to develop an effective, efficient and unconventional approach for the inference of multiple subspaces from high-dimensional data. We use this technique to address the problem of multibody grouping. Unlike other methods, independence of the motions of objects is no longer an imposed constraint in our algorithm. A generalized mechanism of rotation in high-dimensional space together with the scheme of voting facilitate the emergence of the underlying multiple subspace structure. Inliers and outliers are discriminated efficiently due to the data point's structural compatibility with this multiple subspace structure. Robust performance against a large amount of outliers and noise is achieved. In the future work, we plan to work on more difficult problem of segmenting correlated motions with partially occluded tracks.

(a)   (b)   (c)

(d)   (e)   (f)

Figure 6: Segmentation results on articulated motion sequences.

References

[1] 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.

[2] C.W. Gear, "Multibody grouping from motion images", *Int'l J. Computer Vision*, vol. 29, pp. 133-150, 1998.

[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.
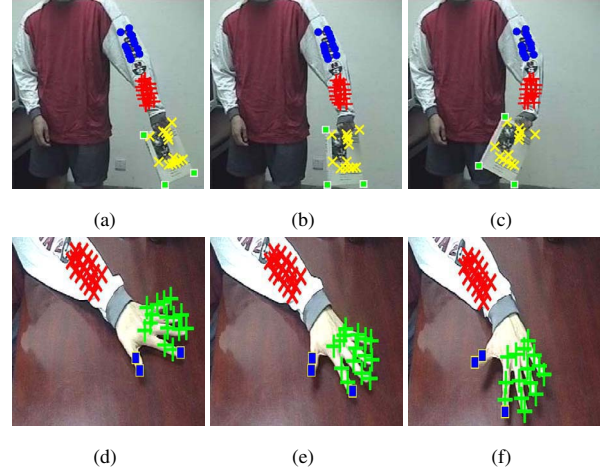
[4] N. Ichimura, "Motion segmentation based on factorization method and discriminant criterion", *Proc. International Conference on Computer Vision*, pp. 600-605, 1999.

[5] Y. Wu, Z.-Y. Zhang, T.S. Huang, and J.Y. Lin, "Multibody grouping via orthogonal subspace decomposition", *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 252-257, 2001.

[6] K. Kanatani, "Motion segmentation by subspace separation and model selection", *Proc. International Conference on Computer Vision*, pp. 301-306, 2001.

[7] L. Zelnik-Manor and M. Irani, "Degeneracies, dependencies and their implications in multi-body and multi-sequence factorizations", *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 287-293, 2003.

[8] J. Shi and C. Tomasi, "Good feature to track", *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593-600, 1994.

[9] R.Szeliski and D.Tonnesen. "Surface modeling with oriented particle systems". *Computer Graphics (SIGGRAPH '92)*, vol. 26, no. 2, pp. 185-194, 1992.

[10] G. Guy and G. Medioni, "Inference of surfaces, 3D curves, and junctions from sparse, noisy, 3D data", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 11, pp. 1265-1277, 1997.

[11] M. Irani, "Multi-frame correspondence estimation using subspace constraints", *Int'l. J. Computer Vision*, vol. 48, no. 3, pp. 173-194, 2002.

[12] D.Q. Huynh and A. Heyden, "Outlier detection in video sequences under affine projection", *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 695-701, 2001.

[13] C.-K. Tang, G.Medioni, and M.-S. Lee, "N-dimensional tensor voting and application to epipolar geometry estimation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 829-844, 2001.

[14] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis (GPCA)", *Proc. IEEE Computer Vision and Pattern Recognition*, pp.621-628, 2003.

[15] G.H. Golub, C.F. Van Loan, "*Matrix computations*", 2nd edition, The John Hopkins University Press, Baltimore, MD, 1989.