Multibody Motion Segmentation Based on Simulated Annealing[*]

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

The problem of multibody motion segmentation is an important and challenging issue in computer vision. In this paper, a novel segmentation technique based on simulated annealing (SA) is proposed. According to the fact that under linear projection models, feature points of multibody reside in multiple subspaces, firstly, a meaningful energy function is proposed, which favors the correct formation of those subspaces, and some subspaces are generated as the initial state. Then, two strategies of subspace evolution and transformation are developed to optimize the energy function in a manner of simulated annealing. The ultimate configuration of these subspaces will reveal the inherent multiple subspace structure embedded in the data space. The classification of data points to these subspaces is equivalent to multibody grouping. The global optimization process results in an increase of robustness with noise tolerance. The method is also effective in degenerate cases. Promising results on synthetic and real data are presented.

1. Introduction

In various real applications in the computer vision field, multibody motions are frequently encountered. Therefore, the techniques for motion segmentation are of fundamental importance and have gained wide-spreading concerns.

A number of algorithms have been proposed to address this problem. Recently, factorization method, which was originally developed by Tomasi and Kanade [1] for structure from motion of a single object, has attracted much popularity. It reveals that under linear projection models, trajectories of a single body lie in a low dimensional subspace of the data space. In the case of shape degeneracy (object has less than three independent dimensions such as a line or a plane) or motion degeneracy (object performs pure rotation or pure translation), the dimension of that subspace would be even lower. So, feature points of multibody actually reside in multiple subspaces. Started with a data matrix $\mathbf{W}$,

whose columns correspond to the features' trajectories imaged over a sequence of frames. The segmentation of independently moving objects can be equivalently achieved by grouping columns of $\mathbf{W}$ into a set of independent subspaces.

In the framework of factorization, Gear [2] formulated the problem as weighted graph matching which may be stuck in some local minima. Costeira and Kanade [3] presented a multibody factorization method in which a shape interaction matrix $\mathbf{Q}$ is introduced. $\mathbf{Q}=\mathbf{VV}^T$ where $\mathbf{V}$ comes from the singular value decomposition (SVD) of $\mathbf{W}$. Elements of $\mathbf{Q}$ has a property that if any features i and j are from different objects, Q_{ij} will be zero, otherwise, non-zero. They then grouped features by thresholding and sorting $\mathbf{Q}$. Ichimura [4] applied a discriminant criterion to select the most representative vectors in $\mathbf{Q}$ for feature grouping. Kanatani [5] developed a method through dimension correction and model selection.

Unfortunately, the performance of algorithms based on $\mathbf{Q}$ degrades quickly when noise exists; the reason is that $\mathbf{Q}$ only records the relationships between individual features and is vulnerable to noise. The work in [2] also suffers from this problem. Presence of noise or outliers will pose additional challenge to this problem.

Wu et al. [6] decomposed $\mathbf{Q}$ into orthogonal subspaces and grouped these fragment subspaces for motion segmentation. More robust performance is obtained because the elements to be grouped are not individual features but several subspaces formed by groups of the feature points. However, the subspace distance defined in [6] for subspace grouping could only be applied to subspaces with same dimensions and therefore, will fail in degenerate cases in which subspaces of each object may have different dimensions.

Fan et al. [7] applied Independent Subspace Analysis (ISA), an approach totally based on the statistical properties of the data, for independent motion segmentation. Vidal et al. [8] proposed a powerful tool, called General Principal Component Analysis (GPCA), for subspace clustering and motion segmentation. In [9], L. Zelnik-Manor et al. proposed an segmentation algorithm by applying subspace constraints on the flow-field.

In this paper, a novel algorithm based on simulated annealing is proposed, which has a robust performance against

noise and outliers and is also able to cope with degeneracies. It will be shown in Section 2 that columns of $\mathbf{V}^T$, which come from the SVD of $\mathbf{W}$, span a set of mutually orthogonal subspaces called objects' *shape subspaces*. Our approach is to discover this mutually-orthogonal-subspace configuration by optimizing an energy function in a simulated annealing (SA) manner. Several potential shape subspaces are generated as the initial solution state. A novel technique for subspace transformation and a conventional subspace evolution technique are used for state transition towards the minimum of the energy which coincides exactly with the subspace structure embedded in the data set. All feature points are involved in this global optimization, which results in an increase of robustness with noise tolerance. An effective technique for detecting and coping with degeneracies is also described. Compared with previous work, the SA based approach in this paper is more intuitive by giving an algebraic description of shape subspaces and exhibits a plain scenario for the procedure of shape subspace finding.

The paper is organized as follows. In Section 2, the problem of multibody motion segmentation is briefly reviewed. Section 3 presents our simulated annealing based approach. Experimental results on synthetic and real image sequences are demonstrated in Section 4. Conclusion is summarized in Section 5.

2. The problem of multibody motion segmentation

Suppose there are N independently moving objects in the scene, each object contains n_i 3D points. Their homogeneous coordinates is represented by a $4 \times n_i$ matrix $\mathbf{S}_i$,

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

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

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

where

$$\mathbf{W}_i = \begin{bmatrix} u_{11} & \cdots & u_{1n_i} \\ v_{11} & \cdots & v_{1n_i} \\ u_{21} & \cdots & u_{2n_i} \\ v_{21} & \cdots & v_{2n_i} \\ \cdots & \cdots & \cdots \\ u_{F1} & \cdots & u_{Fn_i} \\ v_{F1} & \cdots & v_{Fn_i} \end{bmatrix} \text{ and } \mathbf{M}_i = \begin{bmatrix} \mathbf{M}_{i1} \\ \mathbf{M}_{i2} \\ \cdots \\ \mathbf{M}_{iF} \end{bmatrix}.$$

Each column of the $\mathbf{W}_i$ contains the observations for a single point over F frames, while each row contains the ob-

served u-coordinates or v-coordinates for a single frame. $\mathbf{M}_i$ is a $2F \times 4$ matrix and $\mathbf{M}_{if}$ ($f{=}1, \ldots, F$) is the 2×4 projection matrix related to object i in the fth frame. Assume at least four non-coplanar feature points are chosen from each object, thus, the n_i columns of $\mathbf{W}_i$ reside in a 4D subspace spanned by the columns of $\mathbf{M}_i$.

All feature points across all frames can be compactly written into a $2F \times P$ matrix $\mathbf{W}$,

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

where $P = \Sigma n_i$ is the total number of features in the scene.

Since the motions of all objects are independent, the rank of $\mathbf{W}$ is $4N$ (degeneracies will be discussed later). By singular value decomposition $\mathbf{W}{=}\mathbf{U}\Sigma\mathbf{V}^T$, where $\mathbf{U}{\in}2F{\times}4N$, $\Sigma{\in}4N{\times}4N$ and $\mathbf{V}{\in}P{\times}4N$, the shape interaction matrix $\mathbf{Q}$ can be computed by $\mathbf{Q}{=}\mathbf{V}\mathbf{V}^T$ and

$$Q_{ij} \begin{cases} = 0, & \text{if feature } i \text{ and } j \text{ belong to different objects,} \\ \neq 0, & \text{if feature } i \text{ and } j \text{ belong to the same object.} \end{cases} \tag{4}$$

Assume we have grouped feature points of different objects, we could express $\mathbf{V}^T$ as $\mathbf{V}^T{=}[\mathbf{V}_1 \; \mathbf{V}_2 \ldots \mathbf{V}_N]$, where $\mathbf{V}_i{=}[\mathbf{V}_i^1 \; \mathbf{V}_i^2 \ldots \mathbf{V}_i^{n_i}]$. Denote $\mathcal{SP}_i{=}\text{span}\{\mathbf{V}_i^1 \; \mathbf{V}_i^2 \ldots \mathbf{V}_i^{n_i}\}$ as the *shape subspace* for object i. According to the nice property of $\mathbf{Q}$ in (4), it is proved [6] that in the noise-free case,

$$\mathcal{SP}_i \perp \mathcal{SP}_j, \forall i \neq j. \tag{5}$$

It means that the shape subspaces are mutually orthogonal. In reality, with no information about feature grouping, we might obtain a $\mathbf{W}^\star$ whose columns are permutation of $\mathbf{W}$, as well as $\mathbf{V}^{\star T}$, a permuted version of $\mathbf{V}^T$. But this does not violate the mutual orthogonality of shape subspaces.

3. Multibody motion segmentation based on simulated annealing

Following the above descriptions, the key to solve the problem of multibody motion segmentation is clear. Given a $2F \times P$ matrix $\mathbf{W}^\star$, we manage to find a set of appropriate mutually orthogonal shape subspaces $\mathcal{SP}_i$ from the $r \times P$ matrix $\mathbf{V}^{\star T}$. Then, multibody segmentation can be equivalently achieved by grouping the columns of $\mathbf{V}^{\star T}$ to those subspaces, where $\mathbf{V}^{\star T}$ comes from the SVD of $\mathbf{W}^\star$, r is the rank of $\mathbf{W}^\star$ and P is the number of imaged features.

In this section, we develop a novel and robust method to address this problem. A meaningful energy function is introduced and then is minimized in a simulated annealing (SA) manner. The most important parts in the SA algorithm

are: the objective energy function, the initial state and the strategy of state transition. Description of these three elements will be presented in the following subsections.

3.1. Membership function and energy function

We first introduce a membership function, based on which the objective energy function is constructed.

Recall the formation of shape subspaces in Section 2, columns of $\mathbf{V}^{\star T}$ actually come from the mutually orthogonal subspaces. We can define a membership function to describe the degree to which the ith column of $\mathbf{V}^{\star T}$ belongs to a certain shape subspace $\mathcal{SP}_j$, as

$$\text{mem}(i,j) = \|\pi_j \cdot \mathbf{V}_i^{\star T}\| / \|\mathbf{V}_i^{\star T}\|, \qquad (6)$$

where $\mathbf{V}_i^{\star T}$ is the ith column vector of matrix $\mathbf{V}^{\star T}$, $\pi_j = \mathbf{B}_j(\mathbf{B}_j^T\mathbf{B}_j)^{-1}\mathbf{B}_j^T$ is the projection matrix onto $\mathcal{SP}_j$ and $\mathbf{B}_j$ consists of basis vectors of subspace $\mathcal{SP}_j$. Due to the mutual orthogonality of those $\mathcal{SP}_j$, this membership function has a nice property that

$$\text{mem}(i,j) \begin{cases} = 1, & \text{if vector } i \text{ lies in } \mathcal{SP}_j, \\ = 0, & \text{if vector } i \text{ is orthogonal to } \mathcal{SP}_j. \end{cases} \tag{7}$$

Suppose there are N shape subspaces (The determination of N will be detailed in Section 3.7), based on the membership function, we then define a energy function as

$$E(r) = A\sum_{i=1}^{P}\sum_{j=1}^{N}\sum_{k=1,k\neq j}^{N}\text{mem}(i,j)\text{mem}(i,k)$$

$$+B(P - \sum_{i=1}^{P}\sum_{j=1}^{N}\text{mem}(i,j))^2 \quad (A=B=1), \quad (8)$$

where P is the total number of imaged features. The first term of $E(r)$ is minimized if for $\forall i$, at most one of the membership functions $\text{mem}(i,j)\ j \in [1,N]$ is nonzero. The second term is minimized if the summation of all $\text{mem}(i,j)$ is equal to P. Reflect the definition and property of the membership function in (6) and (7). The minimum energy $E(r)$ favors the case that the shape subspaces are mutually orthogonal and all columns (features) are correctly classified into their own subspaces. So, (8) is the objective energy function to be minimized in our algorithm.

3.2. Initial state of the simulated annealing

The initial state in the process of SA is a set of randomly generated mutually orthogonal subspaces as the potential shape subspaces in the r-dimensional space spanned by $\mathbf{V}^{\star T}$. Because none of them will have a rank more than four, we assign each of them four r-dimensional orthonormal basis vectors as,

$$\begin{cases} \mathcal{SP}_1 & = & \text{span}\{\mathbf{SP}_1^{(1)}, \mathbf{SP}_1^{(2)}, \mathbf{SP}_1^{(3)}, \mathbf{SP}_1^{(4)}\}, \\ \cdots & & \cdots \\ \mathcal{SP}_N & = & \text{span}\{\mathbf{SP}_N^{(1)}, \mathbf{SP}_N^{(2)}, \mathbf{SP}_N^{(3)}, \mathbf{SP}_N^{(4)}\}. \end{cases} \tag{9}$$

3.3. Strategy for state transition

Reasonable scheme for state transition can facilitate a fluent and consistent movement to the ideal state with the minimum energy. Totally, we design two strategies for state transition which exhibit the most novel feature of our algorithm and are confirmed to be pertinent and effective to escape local minima.

3.3.1. Strategy 1: subspace evolution. Besides the input data of P column vectors from the $r \times P$ matrix $\mathbf{V}^{\star T}$, we have N potential subspaces. Aimed to find the inherent relationship between the column vectors and those potential subspaces, the simplest but straightforward strategy that can be implemented is to rotate a subspace towards a vector. Therefore, referring to the learning scheme in subspace classifier [10], one of the strategies for state transition can be defined as

$$\pi_j^{'} = (\mathbf{I} + \frac{\mu\mathbf{V}_i^{\star T}(\mathbf{V}_i^{\star T})^T}{\|\mathbf{V}_i^{\star T}\|^2})\pi_j, \qquad (10)$$

where π_j is the orthogonal projection onto the jth subspace, $\mathbf{V}_i^{\star T}$ is the ith column vector of matrix $\mathbf{V}^{\star T}$ and $0 < \mu < 1$ denotes the learning rate decreasing when the temperature of SA drops. This manipulation will induce a rotation of subspace $\mathcal{SP}_j$ towards the vector $\mathbf{V}_i^{\star T}$ and therefore, the projection result of $\mathbf{V}_i^{\star T}$ onto $\mathcal{SP}_j$ will become greater. i and j are arbitrarily chosen from [1,P] and [1,N].

3.3.2. Strategy 2: subspace transformation. However, only performing (10) is insufficient to achieve global extremum. By investigation, we find that the solution may be trapped in some local minima that the basis components of a subspace might consist of basis vectors from different subspaces. For instance, at a state in the simulated annealing process, one subspace may be represented as $\mathcal{SP}_i = \{\mathbf{B}_i^{(1)}, \mathbf{B}_i^{(2)}, \mathbf{B}_i^{(3)}, \mathbf{B}_j^{(1)}\}$, $j \neq i$, where $\mathbf{B}_i^{(1)}$, $\mathbf{B}_i^{(2)}$ and $\mathbf{B}_i^{(3)}$ are three basis vectors of the ideal shape subspace $\mathcal{SP}_i$, while $\mathbf{B}_j^{(1)}$ is a basis vector of the ideal shape subspace $\mathcal{SP}_j$. Therefore, there is a need to swap basis vectors between subspaces in order to find the appropriate grouping of basis vectors and the optimal subspace formation. Inspired by this idea, we propose an alternative strategy of state transition as *subspace transformation*.

The concept of principal angles and principal vectors [11] is well-suited for the investigation of the essential relationship between a pair of subspaces.

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

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

$$\begin{cases} [\mathbf{u}_1,\ldots,\mathbf{u}_p] & = & \mathbf{AY}, \\ [\mathbf{v}_1,\ldots,\mathbf{v}_p] & = & \mathbf{BZ}, \\ \cos(\theta_k) & = & \sigma_k, \qquad k=1,\ldots,p. \end{cases} \quad (11)$$

The computation of principal angles, $[\mathbf{u}_1,\ldots,\mathbf{u}_p]^T[\mathbf{v}_1,\ldots,\mathbf{v}_p] = \mathbf{Y}^T(\mathbf{A}^T\mathbf{B})\mathbf{Z} = \text{diag}(\cos(\theta_k))$ $(k=1,\ldots,p)$, indicates that the angle between the ith pair of principal vectors $(\mathbf{u}_i,\mathbf{v}_i)$ is just the ith principal angle of the subspace pair $(\mathcal{A},\mathcal{B})$. In a sense, the principal angles can fully represent and totally determine the relative configuration of the two subspaces.

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

$[\mathbf{u}_1,\ldots,\mathbf{u}_p]^T[\mathbf{v}_1,\ldots,\mathbf{v}_p] = \text{diag}(\cos(\theta_k))$, $k=1,\ldots,p$ 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.

This mutual orthogonality can offer an appropriate method for subspace transformation. Consider two subspaces, $\mathcal{SP}_i$ and $\mathcal{SP}_j$, out of the current state. We calculate their principal vectors as

$$\begin{cases} \mathcal{SP}_i = \text{span}\{\mathbf{u}^{(p1)}, \mathbf{u}^{(p2)}, \mathbf{u}^{(p3)}, \mathbf{u}^{(p4)}\}, \\ \mathcal{SP}_j = \text{span}\{\mathbf{v}^{(p1)}, \mathbf{v}^{(p2)}, \mathbf{v}^{(p3)}, \mathbf{v}^{(p4)}\}. \end{cases}$$

Within $\mathcal{SP}_i\oplus\mathcal{SP}_j$, the four subspaces spanned by $\{\mathbf{u}^{(pk)}, \mathbf{v}^{(pk)}\}$ $(k=1,\ldots,4)$ are mutually orthogonal. These pairs of principal vectors forms the corresponding principal angles, respectively. Without loss of generality, we assume that one pair of the principal vectors, e.g. $\{\mathbf{u}^{(p1)}, \mathbf{v}^{(p1)}\}$, is swapped to realize the transformation of these two subspaces. So, the configuration of these two subspaces in the new state will be

$$\begin{cases} \mathcal{SP}_i^{'} = \text{span}\{\mathbf{v}^{(p1)}, \mathbf{u}^{(p2)}, \mathbf{u}^{(p3)}, \mathbf{u}^{(p4)}\}, \\ \mathcal{SP}_j^{'} = \text{span}\{\mathbf{u}^{(p1)}, \mathbf{v}^{(p2)}, \mathbf{v}^{(p3)}, \mathbf{v}^{(p4)}\}. \end{cases}$$

However, the mutual orthogonality of the four subspaces spanned by $\{\mathbf{u}^{(pk)}, \mathbf{v}^{(pk)}\}$ $(k=1,\ldots,4)$ remains unchanged. Although the configurations of the two subspaces are really altered, $\{\mathbf{u}^{(pk)}, \mathbf{v}^{(pk)}\}$ $(k=1,\ldots,4)$ are still principal vector pairs of the new subspaces $(\mathcal{SP}_i^{'}, \mathcal{SP}_j^{'})$ (Observation 1). Therefore, the four principal angles of $(\mathcal{SP}_i^{'}, \mathcal{SP}_j^{'})$ are exactly the same as the principal angels of $(\mathcal{SP}_i, \mathcal{SP}_j)$, which further implies the invariability of the relative configuration of these two subspaces through the transformation. Furthermore, since the manipulation of the transformation is limited in $\mathcal{SP}_i\oplus\mathcal{SP}_j$, no impact has been made on other shape subspaces. In execution of SA, the two subspaces and the one pair of principal vectors to be swapped are randomly selected from the current solution state.

3.4. Degeneracies

In real applications, two types of degeneracies, shape degeneracy and motion degeneracy, are possible, both of which will make the dimension of the object's shape subspace less than four.

So, the total number of basis vectors $(4N)$ of those potential subspaces may be larger than the actual dimension (r) of the data space. The intersection of shape subspaces caused by dimension redundance can be checked in the final phase of the SA process, when the subspace configurations are fairly stable and does not vary significantly. We examine the first principal angle θ_{ij} between every pair of shape subspaces $\mathcal{SP}_i$ and $\mathcal{SP}_j$ $(i,j=1,\ldots,N, i\neq j)$. If $\cos\theta_{ij} > 0.97$, the overlapped basis vector, i.e., the first principal vector, will be removed from one of the two subspaces. The reduction of dimension is done to the very subspace which can offer a greater decrease in the energy function (8). This technique is practical and can always cope with degeneracies, as shown in the section of experiments.

3.5. Feature grouping and outlier rejection

The membership function (6) is used in this section to discard outliers and to group features. We observe that in real applications, noise and outliers will distort the mutual orthogonality of the obtained shape subspaces. In our extensive experiments, the inlier's membership to its own subspace will be mostly around 1, but is nearly 0 to other subspaces. Thus, outliers could be simply identified if memberships of such feature points to all the shape subspaces are nearly equal or comparable.

If $\text{mem}(i,j)<0.99$, feature i will not be classified to the subspace $\mathcal{SP}_j$. The benefit of this thresholding is that outliers are detected and discarded by multi-pass formed by these subspaces. We then classify the filtered inlier column i to the shape subspace $\mathcal{SP}_{j'}$ which produces the largest membership value, $j' = \arg\max_j \text{mem}(i,j)$. Thus, the result of multibody grouping is equivalently achieved.

3.6. Summary of SA based algorithm

The steps of the algorithm are summarized as below:

1. Given $\mathbf{W}^\star$, obtain the $r\times P$ matrix $\mathbf{V}^{\star T}$ by SVD, where r is the rank of $\mathbf{W}^\star$.
2. Set a start temperature T (parameter of SA). Randomly generate N mutually orthogonal subspaces as the initial state of SA.
3. Randomly choose one of strategies, subspace evolution and subspace transformation, to generate new state while decreasing the temperature.
4. Accept/reject the new state based on $\Delta E(r)$ and T till the temperature is rather low.
5. Check if any redundant dimensions to be reduced.
6. Use the membership function (6) to group features into multibodies and discard outliers.

3.7. Determination of parameter r and N

In Section 3.2, the value r, rank of $\mathbf{W}^{\star}$, is essential since the rank specifies the rows of singular vectors $\mathbf{V}^{\star T}$. The value N in Section 3.1 is of the same importance which is the number of shape subspaces. It is difficult to give them a correct estimation when noise present. In our experiment, let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{r_{\text{noise}}}$ be singular values of $\mathbf{W}^{\star}$, where $\sigma_{r_{\text{noise}}} = \min(2F, P)$. We define the rank of $\mathbf{W}^{\star}$ as the smallest r_{est} that satisfies

$$\sum_{i=1}^{r_{\text{est}}} \sigma_i^2 \Big/ \sum_{i=1}^{r_{\text{noise}}} \sigma_i^2 \geq 0.98. \qquad (12)$$

For more accuracy, the algorithm will be executed for several times with the rank of $\mathbf{W}^{\star}$ in some range, e.g. $r \in [r_{\text{est}}\text{-}3, r_{\text{est}}+3]$. For every rank possibility r, the number of potential moving objects N is restricted within the range $[\text{round}(r/4)\text{-}1, \text{round}(r/4)+1]$ ($\text{round}(\mathbf{X})$ rounds the elements of $\mathbf{X}$ to the nearest integers). Each pair of (r, N) conducts a full SA process for energy minimization. Finally, the parameter pair (r, N), which yields the minimum of energy over all pairs, is chosen as the optimal solution. If the value of the current (r, N) is incorrect, disharmony with the original data structure makes it less likely to yield the minimum energy over all r and N.

4. Experimental results

In this section, experiments on synthetic and real data are carried out to demonstrate the algorithm's robust performance and the ability to cope with degeneracies. The data being analyzed in this section is the $r \times P$ matrix $\mathbf{V}^{\star T}$.

4.1. Synthetic data

In this experiment, 120 points are randomly chosen from 4 transparent entities, two spheres and two planes (rank 3), each containing 30 points. All of them move in an arbitrary way except that one of the spheres performs pure rotation across the sequence (rank 3). So, both *shape degeneracy* and *motion degeneracy* are allowed ($r=4+3+3+3=13$). 10 frames with resolution 100×100 pixels are captured and the standard deviation of simulated Gaussian noise is 1 pixel. 50 outliers are fabricated into the image stream. We let the rank r in the range [11,15] and compare the energy results (8) with different pairs of r and N, as shown in Table 1.
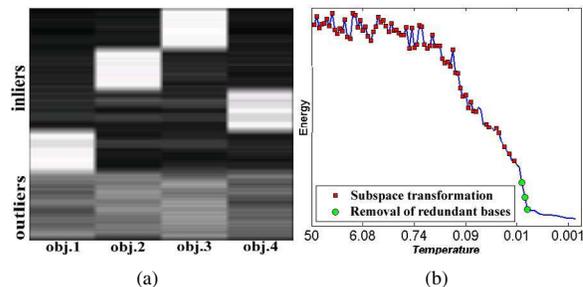

Figure 1: (a) Visualization of all feature's projection results onto each subspace (totally 4 objects). (b) Change of energy in the SA process

The results show that $r=13$ and $N=4$ is obviously the most favorable one coinciding with the actual situation. For convenience of visualization, the membership functions of these features to the 4 shape subspaces are transformed to gray scale between [0, 255] in Fig. 1(a). We can see that each subspace's response to its inliers is very apparent while outliers have comparable impress on all subspaces.

In Fig. 1(b), the energy change with different temperature in a SA process is reported. Red squares denote the occurrences of subspace transformation and green circles represent the removal of redundant bases. It is seen that when the temperature is high, swapping of basis vectors often occurs. While in the final phase of SA, the state of solution is relatively stable and no more swapping of basis vectors happens. Then, redundant bases are removed enabling a closer approach to the ideal solution.

Table 2 summarizes the segmentation result. Notice that erroneous removal of a few correct points is not catastrophic and a high percentage of the outliers are successfully discarded. Our method acts more accurately in our extensive experiments in which the inlier to outlier ratio is not so severe as in the shown synthetic data set.

In another synthetic experiment, we compared our algorithm with two methods: a discriminant method similar to [4] and a simple thresholding method. 70 points in total on two objects are chosen, 35 features from the first object and 35 points from the other. Gaussian noise with 0 mean and standard deviation of ε is added to the image feature coordinates. The ε ranges from 0 to 5 pixels with a interval of 0.25 pixels. The three methods are applied 10 times for

Table 1: The energy results of different r and N

$E(r)$	r: 11	12	13	14	15
N: 2	2.871	2.686	2.213		
3	2.130	2.165	2.050	2.359	3.053
4	0.782	0.876	*0.47*	1.063	1.573
5				0.707	1.146

Table 2: Segmentation results on synthetic data (s-sphere, p-plane, otl-outliers)

	s1	s2	p1	p2	otl
input features	30	30	30	30	50
clustering result	21	35	26	24	64
true inliers (true outliers)	21	29	26	24	44
false outliers (false inliers)	0	6	0	0	20
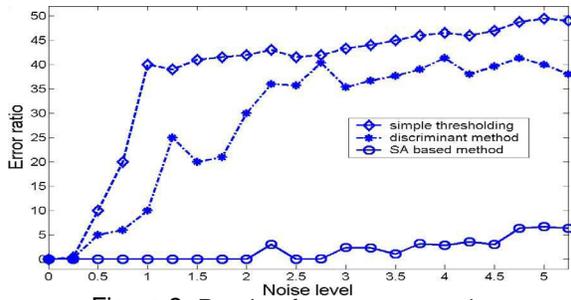
Figure 2: Results of average error ratio.

(a)   (b)   (c)

Figure 3: Vehicle sequence. Moving vehicle and moving background are shown by "+" and "o", respectively.

(a)   (b)   (c)

Figure 4: Face sequence. Three moving groups are shown by "x", "+" and "o", respectively.

each ε. Fig. 2 shows the average error ratio. The thresholding and discriminant methods caused misclassifictions, while our method greatly improves the classification accuracy and is always correct up to ε=4.5 pixels.

4.2. Real image sequences

Results are shown for real image sequences. Feature points were detected and tracked by using KLT tracker [12]. Fig. 3 show 3 views from collected 16 frames. The background is not still due to the instability of the hand-held video camera. "+" and "o" denote good features of the vehicle and the background correctly grouped by our algorithm. Fig. 4 show 3 views from an image sequence of 20 frames taken in a lab and the book's planar surface is considered as *shape degeneracy*. Features that belong to the two books and a face are properly classified and denoted by "x", "+" and "o", respectively. The results are promising.

5. Conclusions

In this paper, we have proposed a novel simulated annealing based algorithm for multibody motion segmentation, which is formulated as a non-trivial problem of finding the underlying subspace structure within the data space. A reasonable anergy function is introduced. Besides the technique for subspace evolution, a novel one for subspace transformation is developed for energy minimization.

Unlike most previous methods which are formulated as clustering problems based on similarities of either individual features or fragment subspaces, we approach this problem via a global SA optimization process. Because feature points are interacted indirectly, inliers have limited exposure to outliers and the strength of inliers are accumulated to overcome to disturbance of noise and outliers which are inconsistent with the multiple subspace structure.

This algorithm is robust against noise and is also effective in degeneracies. However, the speed of simulated annealing is comparatively slow. In our future work, we will research on developing segmentation algorithms with faster speed and the problem of correlated motion segmentation.

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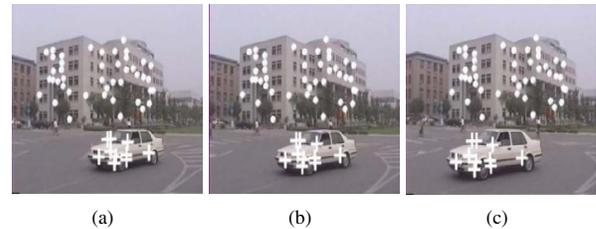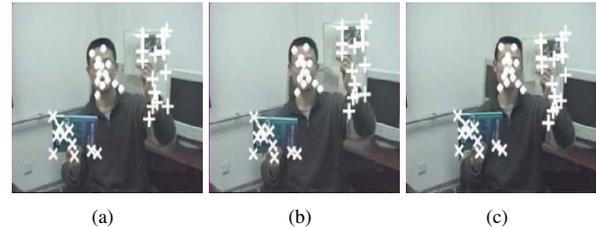
[2] C.W. Gear, "Multibody grouping from motion images", *Int'l J. Computer Vision*, vol. 29, pp. 133-150, 1998.

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

[6] 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.

[7] Z. Fan, J. Zhou and Y. Wu, "Motion Segmentation Based on Independent Subspace Analysis", *Asian Conference on Computer Vision*, pp. 997-1002, 2004.

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

[9] L. Zelnik-Manor, M. Machline and M. Irani, "Multi-body Segmentation: Revisiting Motion Consistency", *Workshop on Vision and Modeling of Dynamic Scenes, with ECCV'02*.

[10] E. Oja, "*Subspace methods of pattern recognition*", Research Studies Press, England, 1983.

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

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