to appear in ACCV 2004, Jeju Island, Korea, January, 2004

MOTION SEGMENTATION BASED ON INDEPENDENT SUBSPACE ANALYSIS

Zhimin Fan†, *Jie Zhou*‡ *and Ying Wu*♯

† ‡ Department of Automation, Tsinghua University, Beijing 100084, China
♯ Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208
† fzm01@mails.tsinghua.edu.cn, ‡ jzhou@tsinghua.edu.cn, ♯ yingwu@ece.northwestern.edu

ABSTRACT

In this paper, we propose a novel method to address the segmentation problem of multiple independently moving objects. Based on the fact that multiple objects' trajectories correspond to multiple independent subspaces, first, bases of these subspaces are extracted by applying independent subspace analysis (ISA). Then, these bases are grouped properly after evaluating the correlation coefficients of them. Feature grouping and outlier rejection are effectively performed by calculating the data point's membership functions to these subspaces. A reasonable energy function is also introduced to facilitate optimal segmentation. The geometrical essence of the method is regarded as a global constraint added in the segmentation process resulting in a considerable increase in error tolerance, without either prior knowledge of the number of objects or prior assumption about existence of degeneracy. The experimental results on synthetic and real data both demonstrate the effectiveness of our algorithm.

1. INTRODUCTION

Multibody motion segmentation is an important issue in computer vision. A number of algorithms have been proposed to address this problem. Ref. [1] presents a method based on evaluating the similarities of some image-level cues. In Ref. [2], rigidity constraints are introduced. Hence, the segmentation process is under the control of optimizing an energy function or maximizing a posteriori (MAP) criterion. However, only using the image-based or model-based algorithms alone has limited applicability because of the ambiguities produced by noise and outliers.

Recently, factorization method, a novel algorithm originally developed by Tomasi and Kanade [3], has attracted much popularity. It reveals the fact that under linear projection models, trajectories of a single body lie in a low dimensional subspace of frame space. And 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 **C**, whose columns correspond to the features' trajectories imaged in a sequence of frames, the segmentation of independently moving objects can be achieved by grouping columns of **C** into a set of independent subspaces. However, when noise or outliers present, the number and dimensions of these subspaces are quite hard to estimate.

Gear [4] formulated this problem as probability analysis of the bipartite graph. Both the partition and the dimensions of the subspaces are determined by maximum likelihood estimation, which may involve local minima and high computational cost.

Costeira and Kanade [5] presented a multibody factorization method, in which a shape interaction matrix **Q** is introduced. $\mathbf{Q}=\mathbf{V}\mathbf{V}^T$ where **V** comes from the singular value decomposition (SVD) of **C**. Elements of **Q** has a nice 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 **Q**. Ichimura [6] applied a discriminant criterion to select the most representative vectors in **Q** for grouping features.

Unfortunately, the performance of algorithms based on **Q** degrades quickly when noise or outliers exist. This degradation results from that **Q** only records the relationships between individual features. A single spike of noise could jeopardize the estimates of both the number of objects and the dimensions of objects' subspaces.

Wu et al. [7] devised a method by performing orthogonal subspace decomposition on **Q** together with a subspace grouping technique. More robust performance is obtained. However, degeneracy is not discussed.

It will be shown in Section 2 that columns of $\mathbf{V}^T$, which comes from the SVD of **C**, span a set of mutually orthogonal subspaces called objects' shape subspaces. And the segmentation of feature points to objects is equivalent to grouping $\mathbf{V}^T$'s columns to those shape subspaces. In this paper, we present a novel approach towards this problem. Bases of the space spanned by $\mathbf{V}^T$ are extracted and properly grouped to build up shape subspaces. A suitable membership function is introduced for grouping features as well as detecting outliers. Influence of noise can be considerably alleviated by utilizing the mutual orthogonality of those sub-

spaces. Estimations of both the number and dimensions of those subspaces are efficiently and simultaneously performed without knowing the prior knowledge of either the number of objects or the existence of degeneracy.

Independent subspace analysis (ISA) [8] is applied to estimate the relationship between those extracted bases of the objects' shape subspaces. ISA, an extension of ICA [9], is used to find a partition of independent subspaces within a data set that the projection norms of the data on those subspaces have maximally sparse distribution. All data points are involved in a global optimization framework to obtain the description of the intrinsic structure of the data space. This attempt is geometrically meaningful, which makes our approach different from the existing factorization based methods. And more robust performance is expected than the clustering process which is merely based on individual similarity measurement matrix **Q**. Besides, the proposed method can handle any case of degeneracy in the context of independent motion segmentation.

The paper is organized as follows. In Section 2, factorization method and Independent Subspace Analysis (ISA) are briefly reviewed. In Section 3, our algorithm for multibody motion segmentation is described. Experimental results are shown in Section 4. And conclusion is summarized in Section 5.

2. BACKGROUND

2.1. Factorization method

We assume that a camera is observing a scene composed of independently moving, rigid bodies and the correspondence and trajectories of feature points in the image sequence have already been discovered.

Suppose N objects are contained in F frames, each has n_i features. The homogeneous coordinates of features on object i is represented by a $4 \times n_i$ matrix $\mathbf{S}_i$ as follows,

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

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 $\mathbf{C}_i$,

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

$$\mathbf{C}_i = \mathbf{M}_i\mathbf{S}_i$$

where $\mathbf{M}_i$ is a $2F\times 4$ matrix and $\mathbf{M}_{ij}$ ($j=1,\ldots,F$) is the 2×4 projection matrix related to object i in the j^{th} frame. Assume at least four non-coplanar feature points are chosen from each object, it is obvious that n_i columns of $\mathbf{C}_i$ reside in a 4D subspace spanned by columns of $\mathbf{M}_i$.

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

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

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 $\mathbf{C}$ is $4N$ (degenerate case will be discussed later). By singular value decomposition $\mathbf{C}=\mathbf{U}\Sigma\mathbf{V}^T$, 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

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

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 Eq. (4), it is proved [7] that in noise-free case,

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

It means that the shape subspaces of objects are mutually orthogonal. In reality, with no information about feature grouping, we might obtain a $\mathbf{C}^\star$ whose columns are permutation of $\mathbf{C}$, 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 $\mathcal{SP}_i$.

2.2. Independent subspace analysis (ISA)

Independent subspace analysis is an extended version of independent component analysis (ICA). In ICA model, an observed n-D vector $\mathbf{x}$ is represented as a linear combinations of n bases vectors $\mathbf{A}=[\mathbf{a}_1,\ldots,\mathbf{a}_n]$,

$$\mathbf{x} = \mathbf{A}\mathbf{s} = \sum_{i=1}^{n}\mathbf{a}_i s_i \tag{6}$$

where $\mathbf{s}=(s_1,\ldots,s_n)^T$ is different for each observed vector $\mathbf{x}$ and the components $s_i=<\mathbf{w}_i,\mathbf{x}>$ are statistically independent. '<>' denotes dot-product and $\mathbf{W}=[\mathbf{w}_1,\ldots,\mathbf{w}_n]^T$ is the inverse of matrix $\mathbf{A}$. The ICA model can be reformulated as $\mathbf{x}=\sum_{i=1}^{n}\mathbf{x}^{(i)}$ by defining $\mathbf{x}^{(i)}=\mathbf{a}_i s_i$. This shows that $\mathbf{x}$ is a sum of n independent vectors $\mathbf{x}^{(1)}$, ..., $\mathbf{x}^{(n)}$.

While in ISA model, the decomposition into L independent subspaces, $\mathcal{R}^n$=$\mathcal{S}_{ISA-1} \oplus \cdots \oplus \mathcal{S}_{ISA-L}$, is realized. And vector $\mathbf{x}$ here admits a unique decomposition as $\mathbf{x} = \sum_{l=1}^{L} \mathbf{x}^{(l)}$ with $\mathbf{x}^{(l)} \in \mathcal{S}_{ISA-l}$. The vectors $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(L)}$ are called the *linear components* of $\mathbf{x}$ and they are independent [10],

$$\mathbf{x}^{(l)} = \sum_{i \in \mathcal{S}^l} \mathbf{a}_i s_i, \quad \mathbf{x} = \sum_{l=1}^{L} \mathbf{x}^{(l)} \tag{7}$$

The $\mathcal{S}_{ISA-l}$ are spanned by a set of bases vectors $\{\mathbf{a}_i | i \in \mathcal{S}^l\}$, where $\mathcal{S}^l$ denotes the set of indices of $\mathbf{a}_i$ belonging to that subspace. This decomposition also indicates a partition of s_i into the corresponding L groups.

According to the invariant feature subspace theory [8], the independence of *linear components* $\mathbf{x}^{(l)}$ can be interpreted as the independence of projection norms of $\mathbf{x}$ on these subspaces. In other words, the independence of (squared) norms of the projections on these subspaces, $\sum_{i \in \mathcal{S}^l} s_i^2$ (l= $1, \ldots, L$), are to be maximized. Thus, given P observed n-D vectors $\mathbf{x}$. The ISA model can be formulated as a maximum likelihood estimation problem [8].

$$\log L(\mathbf{W}) = \sum_{k=1}^{P} \sum_{l=1}^{L} \log p(\sum_{i \in S^l} s_{i,k}^2) + P \log |\det \mathbf{W}| \tag{8}$$

where $p(\cdot)$ is some known density functions (often assumed to be exponential) of the projection norms. Gradient ascent algorithm [8] or Newton method [11] can be used to estimate both the basis matrix $\mathbf{A}$=[$\mathbf{a}_1$,. . .,$\mathbf{a}_n$], $\mathbf{W}$=[$\mathbf{w}_1$,. . .,$\mathbf{w}_n$]^T and the realization of $s_{i,k}$=<$\mathbf{w}_i$,$\mathbf{x}_k$> for $i = 1, \ldots, n$, $k = 1, \ldots, P$. Compared with ICA, this model specifies the information that s_i within the same group are dependent, but those s_i in different groups are independent.

In ICA applications, the correlation coefficient between ABS(s_i) and ABS(s_j) is zero, which is calculated as,

$$\mathbf{I}(s_i, s_j) = < \mathrm{ABS}(s_i), \mathrm{ABS}(s_j) > / (\|s_i\| \cdot \|s_j\|) \tag{9}$$

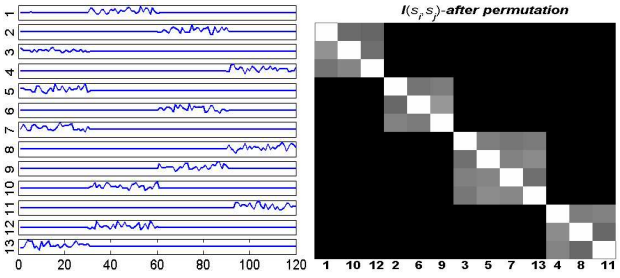
where ABS(Y) returns the absolute value of the elements of Y, $s_i = (s_{i,1}, s_{i,1}, \ldots, s_{i,P})$ is a P-dimensional vector and $\|\cdot\|$ denotes norm function. But, in the case of ISA,

$$\mathbf{I}(s_i, s_j) \begin{cases} > 0 & \text{if } s_i \text{ and } s_j \text{ are dependent} \\ = 0 & \text{if } s_i \text{ and } s_j \text{ are independent} \end{cases} \tag{10}$$

According to $\mathbf{I}(s_i, s_j)$, a partition of L groups within s_i can be found easily: Only if $\mathbf{I}(s_i, s_j)$>0, we group s_i and s_j to the same group. A *component similarity matrix* for grouping components s_i is then defined as $\mathcal{I}$,

$$\mathcal{I} = \{\mathcal{I}_{ij} : \mathcal{I}_{ij} = \mathbf{I}(s_i, s_j), \forall i, j \le n\} \tag{11}$$

Note that the process of clustering components s_i dose not need the prior knowledge of the number or dimensions of those subspaces.



(a) (b)

Fig. 1. (a) Components s_i. (b) The result of $\mathcal{I}$ permuted according to the formation of the independent groups.

To demonstrate the effectiveness of ISA, we generate 4 mutually orthogonal subspaces of dimensions 3, 3, 3 and 4 in a 13D space. 30 samples from each subspace are selected to form a 13×120 observation matrix $\mathbf{X}$=[$\mathbf{x}_1$, $\mathbf{x}_1$, . . . , $\mathbf{x}_{120}$]. By applying Fast-ICA [11] on $\mathbf{X}$, Fig. 1(a) shows the result of s_i, $i = 1, \ldots, 13$. Obviously, there are totally 4 groups consisting of components $\mathcal{S}^1$={1,10,12}, $\mathcal{S}^2$={2,6,9}, $\mathcal{S}^3$= {3,5,7,13}, and $\mathcal{S}^4$={4,8,11}, respectively. This waveform of s_i interprets the concept of "independence between projection norms": Assume $\mathbf{x}_k \in \mathcal{S}_{ISA-p}$, its projection result onto its own subspace $\{s_{i,k} | i \in \mathcal{S}^p\}$ has no correlation with its projection results onto other subspaces $\{s_{j,k} | j \notin \mathcal{S}^p\}$.

Fig. 1(b) represents the matrix $\mathcal{I}$ which has been permuted according to the formation of the independent groups of these s_i. The higher the $\mathcal{I}_{ij}$ is, the greater the brightness.

3. MOTION SEGMENTATION BASED ON ISA

Given a $2F \times P$ matrix $\mathbf{C}^\star$, the purpose of our algorithm is to find a set of mutually orthogonal (i.e. independent) shape subspaces $\mathcal{SP}_i$ from the $r \times P$ matrix $\mathbf{V}^{\star T}$, where r is the rank of $\mathbf{C}^\star$ and P is the number of imaged features.

3.1. Estimation of the number of subspaces

Each column of $\mathbf{V}^{\star T}$ is regarded as an observed r-D vector. After applying Fast-ICA on $\mathbf{V}^{\star T}$, a $r \times r$ basis matrix $\mathbf{A}$, the corresponding r components s_i and the *component similarity matrix* $\mathcal{I}$ are obtained.

Because $\mathcal{I}_{ij}$ is a similarity measurement of relevant bases, the greater the $\mathcal{I}_{ij}$ is, the more possible that s_i and s_j belong to one independent subspace. Due to local measurement errors, the entry $\mathcal{I}_{ij}$ may exhibit a small nonzero value for independent components s_i and s_j. However, this disturbance does not bring too much difficulty. Referring to the nearest-neighbor algorithm [12], we begin with the largest $\mathcal{I}_{ij}$, either by merging a component s_i into one set or by merging two sets into one, iteratively. Thus, the number of subspaces and dimension of each subspace can be acquired simultaneously. The only constraints in this clustering procedure are that dimension of any subspace is at most 4 and

the adopted $\mathcal{I}_{ij}$ for clustering is bigger than a certain threshold.

Note that the entry of $\mathcal{I}$ is a kind of global structure based similarity measurement, which evaluates the similarities between extracted bases of data space. All feature points are involved in the maximization process in Eq. (8) to achieve a maximally sparse distribution of their projection results on those inferred subspaces. So, more accurate description of the data space structure and more robust performance are expected than the clustering process which is merely based on individual similarity measurement matrix **Q**.

3.2. Feature grouping and outlier rejection

After grouping s_i into N groups, here N is the *automatically obtained* number of moving objects, the N subspaces $\mathcal{S}_{ISA-l}$ can be constructed by spanning basis sets $\mathbf{B}_l$=$\{\mathbf{a}_i | i \in \mathcal{S}^l\}$, l=$1, \ldots, N$. In fact, these $\mathcal{S}_{ISA-l}$ exactly correspond to the shape subspace $\mathcal{SP}_l$. Because both of them describe the inherent structure of independent subspaces within the same data set. In the following, we use $\mathcal{SP}_l$ denote these subspaces. The grouping of feature points to the bodies now is equivalent to grouping the columns of $\mathbf{V}^{\star T}$ to those derived subspaces $\mathcal{SP}_l$. For this purpose, a membership function is defined as,

$$\text{Mem}(i,l) = \|\pi_l \cdot \mathbf{V}_i^{\star T}\| \,/\, \|\mathbf{V}_i^{\star T}\| \tag{12}$$

where $\pi_l = \mathbf{B}_l(\mathbf{B}_l^T\mathbf{B}_l)^{-1}\mathbf{B}_l^T$ is the projection matrix onto $\mathcal{SP}_l$. Eq. (12) describes the degree to which the i^{th} column of $\mathbf{V}^{\star T}$ belongs to the subspace $\mathcal{SP}_l$,

$$\text{Mem}(i,l) \begin{cases} = 1 & \text{if vector } i \text{ lies in } \mathcal{SP}_l \\ = 0 & \text{if vector } i \text{ is orthogonal to } \mathcal{SP}_l \end{cases} \tag{13}$$

In real applications, noise and outliers will make the mutual orthogonality of $\mathcal{SP}_l$ distorted. 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 Mem(i,l)<0.99, feature i will not be classified to the subspace $\mathcal{SP}_l$. 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 $\mathcal{SP}_{l'}$ which produces the largest membership value, $l' = \arg\max_l \text{Mem}(i,l)$.

3.3. Summary of algorithm

The outline of the proposed algorithm for multibody segmentation is summarized below:

1. Given $\mathbf{C}^\star$, obtain the $r \times P$ matrix $\mathbf{V}^{\star T}$ by SVD, where r is the rank of $\mathbf{C}^\star$.
2. Compute s_i (i=$1, \ldots, r$) and evaluate the *component similarity matrix* $\mathcal{I}$ by applying ISA to $\mathbf{V}^{\star T}$.
3. Use $\mathcal{I}$ to discover the grouping of these components s_i and the number (N) of these groups. Then construct the corresponding shape subspaces $\mathcal{SP}_l$, $l = 1, \ldots, N$.
4. Evaluate the membership function using Eq. (12) to group features into multibodies and discard outliers.

After this clustering procedure, the grouping of columns of $\mathbf{V}^{\star T}$ to shape subspaces $\mathcal{SP}_l$, l=$1, \ldots, N$ is discovered. This is equivalent to grouping feature points to the *N* objects. Neither prior knowledge of number of objects nor prior assumption about existence of degeneracy is assumed.

Recall that the clustering process in Section 3.1 is able to estimate the dimensions of each subspace. As long as these motions are linearly independent, the presented algorithm can robustly handle any case of shape degeneracy or motion degeneracy. This happens, for example, when the object undergoes pure rotation (corresponding to a 3D subspace) or when the object is a 2D (corresponding to a 3D subspace) or a 1D (corresponding to a 2D subspace) entity.

Another case of degeneracy is when the transformation of one object is coupled with another. This case is more complex and will be our further study.

3.4. Determination of parameter r

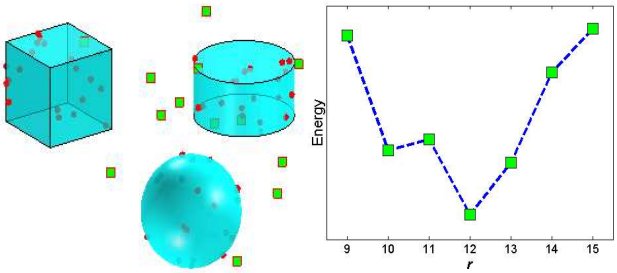
In Section 3.3, the value r, rank of $\mathbf{C}^\star$, is essential since the rank specifies the rows of singular vectors $\mathbf{V}^{\star T}$. Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{r_{\text{noise}}}$ be singular values of $\mathbf{C}^\star$, where $\sigma_{r_{\text{noise}}}$=min$(2F, P)$. In our experiment, we define the rank of $\mathbf{C}^\star$ as the smallest r_{est} that satisfies,

$$\sum_{i=1}^{r_{\text{est}}} \sigma_i^2 \Big/ \sum_{i=1}^{r_{\text{noise}}} \sigma_i^2 \geq 0.98 \tag{14}$$

For more accuracy, we run the algorithm for several times with the rank of $\mathbf{C}^\star$ in some range, e.g. r∈[r_{est}-3,r_{est}+3]. The following energy function is used to choose the optimal segmentation result which produces the minimum energy,

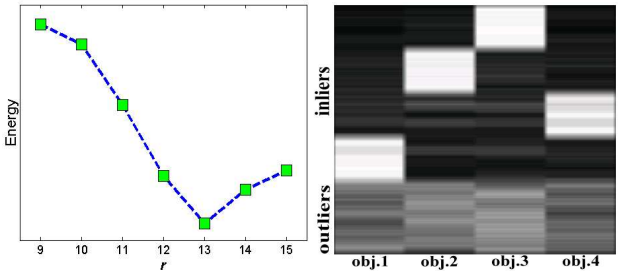
$$E(r) = A\sum_{i=1}^{P}\sum_{j=1}^{N}\sum_{k=1,k\neq j}^{N} \text{Mem}(i,j)\text{Mem}(i,k) + B(P - \sum_{i=1}^{P}\sum_{j=1}^{N}\text{Mem}(i,j))^2 \quad (A=B=1) \tag{15}$$

The first component of $E(r)$ is minimized if for $\forall i$, at most one of the membership functions Mem(i,j) $j \in [1, N]$



(a) (b)

Fig. 2. (a) A view of 1st synthetic scene containing 3 moving objects and several outliers. (b) Change in the value of $E(r)$ for different r in 1st synthetic data set.



(a) (b)

Fig. 3. (a) Change in the value of $E(r)$ for different r in 2nd synthetic data set. (b) Visualization of all feature's membership functions to each subspace. (Totally 4 objects)

is nonzero. The second component is minimized if the summation of Mem(i, j) is equal to P. Reflect the definition and property of the membership function in Eq. (12), (13). The minimum energy $E(r)$ obviously favors the case that the obtained subspaces are mutually orthogonal and all features are correctly classified. If the value of current r is incorrect, disharmony with the original data structure makes it less likely to yield the minimum energy over all r.

4. EXPERIMENTAL RESULTS

Experiments on synthetic and real data are carried out to demonstrate the effectiveness of our algorithm.

4.1. Synthetic data

In one of the synthetic experiment, three transparent full 3D objects (rank 4), a sphere, a cylinder and a cubic, are generated (r=4×3=12). Totally 90 points are randomly chosen, 30 points from each object. We let these three sets of points undergo independent motions. Using orthographic projection, 10 frames with resolution 100×100 pixels are captured. Gaussian noise with standard deviation of 2 pixels is added and 30 fake trajectories are fabricated in the image stream.

Fig. 2(a) shows a view of the synthetic scene. Points are features while squares denote outliers. Fig. 2(b) shows the minimum energy according to Eq. (15) under different rank r of $\mathbf{V}^{\star T}$. r=12 is obviously the most favorable result coinciding with the actual situation.

The 2nd experiment is conducted on more complicated synthetic data set. Image resolution is 100×100 pixels and the standard deviation of simulated Gaussian noise is 1 pixel. 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 (rank 3) across the sequence. So, both shape degeneracy and motion degeneracy are allowed (r=4+3+3+3=13). 50 outliers are imported. Fig. 3(a) shows the minimum energy according to Eq. (15). r=13 is obviously the most favorable result coinciding with the actual situation. For convenience of visualization, the membership functions of these features to the 4 extracted subspaces are transformed to gray scale between [0, 255], as shown in Fig. 3(b). We can see that each subspace's response to its inliers is very apparent while outliers have comparable impress on all subspaces.

Table 1 summarized the segmentation results on these two synthetic data. 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 sets.

4.2. Real image sequences

Adequate results are shown for real image sequences. Feature points were extracted using the corner detector proposed in [13]. Detected features are tracked by normalized correlation. Fig. 4(a)-(c) show 3 views from collected 16 frames. The background is not still due to hand-held video camera's vibration. "+" and "o" denote good features of vehicle and background correctly grouped by the algorithm. Fig. 4(d)-(f) show another sequence containing 12 frames. "+" and "o" denote properly grouped features of vehicle and background, respectively.

Fig. 5 show 3 views from an image sequence of 20 frames. Features that belong to two books and a face are properly classified and denoted by "x", "+" and "o", respectively. The results are promising.

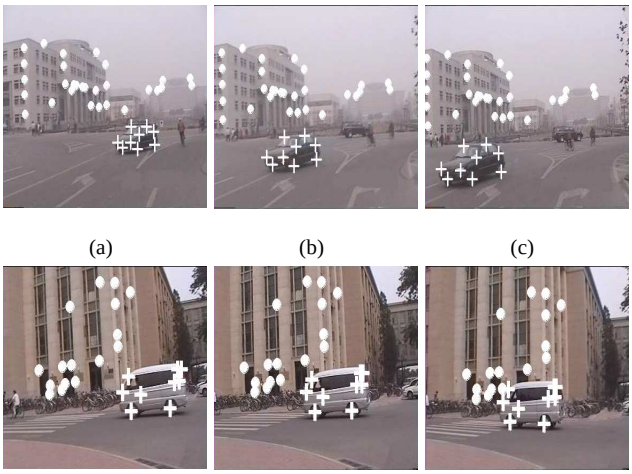
However, currently there are no available public data sets of this topic. And quantitative comparison with the existing methods is hampered by the difficulties of selecting thresholding parameters for clustering, which is the essential procedure for efficient and faithful implementation of other methods. Qualitative comparison is in Section 5.

5. CONCLUSIONS

In this paper, we have provided a novel attempt for the problem of independent multibody motion segmentation. Due to the mutually orthogonal subspace structure of $\mathbf{V}^{\star T}$, bases of these subspaces are extracted and properly grouped by using ISA. Outliers are efficiently discarded by a introduced

Table 1. Segmentation results on synthetic data

	1^{st} set of synthetic data				2^{nd} set of synthetic data				
	cubic	cylinder	sphere	outliers	sphere1	sphere2	plane1	plane2	outliers
input features	30	30	30	30	30	30	30	30	50
clustering result	29	29	29	33	31	33	22	27	57
true inliers (true outliers)	29	29	29	30	30	30	22	27	46
false outliers (false inliers)	0	0	0	3	1	3	0	0	11



(a) (b) (c)

(d) (e) (f)

Fig. 4. Two vehicle sequences. Moving vehicles and moving backgrounds are shown by "+" and "o", respectively.

(a) (b) (c)

Fig. 5. Face sequence. Three moving groups are shown by "x", "+" and "o", respectively.

membership function. A reasonable energy function is also created for selection of optimal segmentation result. The algorithm can robustly handle any degenerate cases with neither prior knowledge of the number of objects nor prior assumption about existence of degeneracy.

Unlike most previous methods based on point-level similarity measurements, in our algorithm, the evaluated *component similarity matrix* $\mathcal{I}$ can be regarded as a qualitatively group-level similarity measure. It helps achieve better and more reliable performance confirmed by experiments.

Focus of our future work will be coupled motion segmentation and quantitative analysis of motions.

6. REFERENCES

[1] Y.-T. Lin, Y.-K. Chen, and S.Y. Kung, "Object-based scene segmentation combining motion and image cues", in *Proc. International Conference on Image Processing*, 1996, pp. 957-960.

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

[3] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method", *Int'l J. Computer Vision*, vol. 9, no. 2, pp. 137-154, 1992.

[4] C.W. Gear, "Multibody grouping from motion images", *Int'l J. Computer Vision*, vol. 29, no. 2, pp. 133-150, 1998.

[5] J. Costeira and T. Kanade, "A multi-body factorization method for independent moving objects", *Int'l. J. Computer Vision*, vol. 29, no. 3, pp. 159-179, 1998.

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

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

[8] A. Hyvärinen and P. Hoyer, "Emergence of complex cell properties by decomposition of natural images into independent feature subspaces", in *International Conference on Artificial Neural Networks*, 1999, pp. 257-262.

[9] P. Comon, "Independent component analysis a new concept?", in *Signal Processing, Elsevier*, 1994, pp. 287-314.

[10] J.-F. Cardoso, "Multidimensional independent component analysis", in *Proc. International Conference on Acoustics, Speech, and Signal Processing*, 1998, pp. 1941-1944.

[11] http://www.cis.hut.fi/projects/ica/fastica/1998

[12] R.O. Duda and P.E. Hart, "*Pattern classification and scene analysis*", Wiley, New York, 1973.

[13] M. Trajkovic and M. Hedley, "Fast corner detection", *Image and Vision Computing*, vol.16, pp. 75-87, 1998.