Spectral Clustering for Robust Motion Segmentation

JinHyeong Park[1], Hongyuan Zha[1], and Rangachar Kasturi[2]

[1] Department of Computer Science and Engineering,
The Pennsylvania State University,
111 IST Building, University Park, PA 16802 USA
{jhpark,zha}@cse.psu.edu

[2] Department of Computer Science and Engineering,
University of South Florida,
4202 East Fowler Ave., ENB118,
Tampa, FL 33620-5399 USA
chair@csee.usf.edu

Abstract. In this paper, we propose a robust motion segmentation method using the techniques of matrix factorization and subspace separation. We first show that the shape interaction matrix can be derived using **QR** decomposition rather than Singular Value Decomposition(**SVD**) which also leads to a simple proof of the shape subspace separation theorem. Using the shape interaction matrix, we solve the motion segmentation problems by the spectral clustering techniques. We exploit multi-way Min-Max cut clustering method and provide a novel approach for cluster membership assignment. We further show that we can combine a cluster refinement method based on subspace separation with the graph clustering method to improve its robustness in the presence of noise. The proposed method yields very good performance for both synthetic and real image sequences.

1 Introduction

The Matrix factorization methods proposed by Tomasi, Costeira and Kanade [1] [2] have been widely used for solving the motion segmentation problems [3] [4] [5] [6] [7] [8] and the 3D shape recovering problems [9] [10] [11]. The basic idea of the methods is to factorize the feature trajectory matrix into the motion matrix and the shape matrix, providing the separation of the feature point trajectories into independent motions. In this paper, we develop a novel robust factorization method using the techniques of spectral clustering.

Given a set of N feature points tracked through F frames, we can construct a feature trajectory matrix $\mathbf{P} \in R^{2F \times N}$ where the rows correspond to the x or y coordinates of the feature points in the image plane and the columns correspond to the individual feature points. Motion segmentation algorithms based on matrix factorization [6] first construct a shape interaction matrix, $\mathbf{Q}$ by applying the singular value decomposition (**SVD**) to the feature trajectory matrix $\mathbf{P}$. Under the noise-free situation, the shape interaction matrix $\mathbf{Q}$ can be transformed

T. Pajdla and J. Matas (Eds.): ECCV 2004, LNCS 3024, pp. 390–401, 2004.

to a block diagonal matrix by a symmetric row and column permutation thereby grouping the feature points of the same object into a diagonal block.

If the trajectory matrix $\mathbf{P}$ is contaminated by noise, however, the block diagonal form of Q no longer holds, and the methods such as the greedy technique proposed in[2] tend to perform rather poorly. Recently there have been several research proposed specifically addressing this problem [7] [5] [3] [4] [5] [6] [8]. We will give a brief review of these methods in Section 2.

In this paper we deal with the issues related to the robustness of the factorization methods. We first show that the shape interaction matrix can be extracted from the trajectory matrix using **QR** decomposition with pivoting, an idea that was briefly mentioned in [2]. As a by-product we give a simple and clean proof of the subspace separation theorem described in [6]. We then observe that the shape interaction matrix is very similar to the weight matrix used for graph partitioning and clustering [12] [13] [14] [15], and the motion segmentation problem can be cast as an optimal graph partitioning problem. To this end, we apply the spectral k-way clustering method [13] [14] to the shape interaction matrix to transform it into near-block diagonal form. In particular, we propose a novel **QR** decomposition based technique for cluster assignment. The technique at the same time also provides confidence levels of the cluster membership for each feature point trajectory. The confidence levels are explored to provide a more robust cluster assignment strategy: we assign a feature point directly to a cluster when it has a very confidence level for the cluster compared to those for other clusters. Using the assigned feature points in each cluster, we compute a linear subspace in the trajectory space. The cluster memberships of other feature points having lower confidence levels, and are therefore not assigned to a cluster, are determined by their distances to each of the linear subspaces. Our experiments on both synthetic data sets and real video images have shown that this method are very reliable for motion segmentation even in the presence of severe noise.

The rest of the paper is organized a s follows: Previous works are discussed in Section 2. Section 3 is devoted to a simple proof that the shape interaction matrix can be computed using **QR** decomposition. Motion segmentation based on spectral relaxation k-way clustering and subspace separation is described in Section 4. Experiment results are shown in Section 5 and conclusion is given in Section 6.

2 Previous Work

The factorization method was originally introduced by Tomasi and Kanade [1]. The method decomposes a matrix of image coordinates of N feature points tracked through F frames into two matrices which, respectively, represent object shape and camera motion. The method deals with a single static object viewed by a moving camera. Extending this method, Costerira and Kanade [2] proposed a multibody factorization method which separates and recovers the shape and motion of multiple independently moving objects in a sequence of images. To achieve this, they introduce a shape interaction matrix which is invariant to both the object motions and the selection of coordinate systems, and suggest a greedy algorithm to permute the shape interaction matrix into block diagonal form. Gear [3] exploited the reduced row echelon form of the shape interaction matrix to group the feature points into the linearly independent subspaces. For Gear's method, in the noise-free case, any two columns of the echelon form which have nonzero elements in the same row correspond to feature points belonging to the same rigid body. The echelon form matrix can be represented by a weighted bipartite graph. Gear also used a statistical approach to estimate the grouping of feature points into subspaces in the presence of noise by computing which partition of the graph has the maximum likelihood.

Ichimura [4] suggested a motion segmentation method based on discriminant criterion [16] features. The main idea of the method is to select useful features for grouping noisy data. Using noise-contaminated shape interaction matrix, it computes discriminant criterion for each row of the matrix. The feature points are then divided into two groups by the maximum discriminant criterion, and the corresponding row gives the best discriminant feature. The same procedure is applied recursively to the remaining features to extract other groups. Wu et. al. [5] proposed an orthogonal subspace decomposition method to deal with the noisy problem of the shape interaction matrix. The method decomposes the object shape space into signal subspaces and noise subspaces. They used the shape signal subspace distance matrix, D, for shape space grouping rather than the noise-contaminated shape interaction matrix.

Kanatani [6] [7] reformulated the motion segmentation problems based on the idea of subspace separation. The approach is to divide the given N feature points to form m disjoint subspaces $\mathcal{I}_i$, $i = 1, \cdots, m$. A rather elaborated proof was given showing that provided that the subspaces are linearly independent, the elements Q_{ij} in the shape interaction matrix Q is zero if the point i and the point j belong to different subspaces. Kanatani also pointed out that even a small noise in one feature point can affect all the elements of Q in a complicated manner. Based on this fact, Kanatani proposed noise compensation methods using the original data rather than the shape interaction matrix Q.

Zelnik-Manor and Irani [8] showed that different 3D motions can also be captured as a single object using previous methods when there is a partial dependency between the objects. To solve the problem, they suggested to use an affinity matrix $\bar{Q}$ where $\bar{Q}_{ij} = \sum_k exp(v_k(i) - v_k(j))^2$, where v_k's are the largest eigenvectors of Q. They also dealt with the multi-sequence factorization problems for temporal synchronization using multiple video sequences of the same dynamic scene.

3 Constructing the Shape Interaction Matrix Using QR Decomposition

In this section, we exhibit the block diagonal form of the shape interaction matrix using **QR** decomposition with pivoting [17], this also provides a simpler proof of the shape subspace separation theorem (Theorem 1 in [6]). Assume we have N rigidly moving feature points, $p_1, \ldots, p_N$, which are on image plane corresponding 3D points over the F frames. Motion segmentation can be interpreted as dividing the feature points p_i into S groups [6] each spanning a linear subspace corresponding to feature points belonging to the same object. We denote the grouping as follows,

$$\{1, \ldots, N\} = \bigcup_{i=1}^{S} \mathcal{I}_i, \quad \mathcal{I}_i \cap \mathcal{I}_j = \emptyset.$$

Now define $l_i = |\mathcal{I}_i|$ which is the number of the points in the set $\mathcal{I}_i$, and k_i=dim $span\{p_j\}_{j \in \mathcal{I}_i} \leq l_i$ and $P_i = \{p_j\}_{j \in \mathcal{I}_i}$.

Let the **SVD** of P_i be $P_i = U_i \Sigma_i V_i^t$, where $\Sigma_i \in R^{k_i \times k_i}, i = 1, \ldots, S$. Then $\mathbf{P} = [P_1, P_2, \ldots, P_s]$ can be written as,

$$\mathbf{P} = [P_1, P_2, \ldots, P_s] = [U_1\Sigma_1, U_2\Sigma_2, \ldots, U_s\Sigma_s] \begin{bmatrix} V_1^T & 0 & \cdots & 0 \\ 0 & V_2^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V_s^T \end{bmatrix}, \quad (1)$$

where rank(V_i)=k_i for $i = 1, \cdots, s$. We assume the S subspaces $span\{p_j\}_{j \in \mathcal{I}_i}, i = 1, \ldots, S$ are linearly independent, then the matrix $[U_1\Sigma_1, U_2\Sigma_2, \ldots, U_s\Sigma_s]$ has full column rank of $k = k_1 + \cdots + k_s$. Therefore, an arbitrary orthonormal basis for the row space of $\mathbf{P}$ can be written as $\Phi diag(V_1, \cdots, V_s)^T$ for an arbitrary orthogonal matrix $\Phi \in R^{k \times k}$. Now the shape interaction matrix can be written as

$$\mathbf{Q} = diag(V_1, \cdots, V_s)\Phi^T \Phi diag(V_1, \cdots, V_s)^T = diag(V_1 V_T, \cdots, V_s V_s^T).$$

This clearly shows that $Q_{ij} = 0$ if i and j belong to different subspaces, i.e., if the corresponding feature points belong to different objects.

A cheaper way to compute an orthonormal basis for the row-space of $\mathbf{P}$ than using **SVD** is to apply **QR** decomposition with column pivoting to $\mathbf{P}^T$,

$$\mathbf{P}^T E = \hat{Q} R \quad (2)$$

where E is a permutation matrix, and $\hat{Q}$ has k columns. It is easy to see that $\hat{Q}\hat{Q}^T = \mathbf{Q}$. In the presence of noise, $\mathbf{P}$ will not exactly have rank k, but **QR** decomposition with column pivoting will in general generate an R matrix that can reliably revealing the numerical rank of $\mathbf{P}$. We can truncate R by deleting rows with small entries.

4 Motion Segmentation

4.1 Spectral Multi-way Clustering

In the last section, we have shown that the shape interaction matrix, $\mathbf{Q} \in R^{N \times N}$ has the block diagonal form when the feature points are grouped into independent subspaces corresponding to S different objects. In general, this grouping is unknown, and we need to find row and column permutations of the matrix $\mathbf{Q}$ to exhibit this block diagonal form, and thus assigning the feature points to different objects. A greedy algorithm has been proposed in [2] for this problem, but it performs poorly in the presence of noise. We now present a more robust method based on spectral graph clustering [12] [13] [14] [15]. We propose a novel technique for cluster assignment in spectral clustering and show that it provides a confidence level that can be used for further refining the cluster memberships of the feature points, thus improving the robustness of the spectral clustering method.

We consider the absolute value of the (i, j) element of the shape interaction matrix $\mathbf{Q}$ as a measure of the similarity of feature points i and j with feature points belonging to the same object more similar than those of other points. In fact, in the noise-free case, feature points in different objects will have zero similarity. Our goal is then to partition the feature points into S groups so that feature points are more similar within each group than across different groups. Let $W = (w_{ij})$ with $w_{ij} = |\mathbf{Q}_{ij}|$. For a given partition of the feature points into S groups, we can permute the rows and columns of W so that rows and columns corresponding to the feature points belonging to the same objects are adjacent to each other, i.e., we can re-order the columns and rows of the W matrix accordingly such that

$$W = \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1S} \\ W_{21} & W_{22} & \cdots & W_{2S} \\ \cdots & \cdots & & \vdots \\ W_{S1} & W_{S2} & \cdots & W_{SS} \end{bmatrix}. \tag{3}$$

We want to find a partition such that W_{ii} will be large while $W_{ij}, i \neq j$ will be small, and to measure the size of a sub-matrix matrix W_{ij} we use the sum of all its elements and denoted as $\mathrm{sum}(W_{ij})$. Let x_i be a cluster indication vector accordingly partitioned with that of W with all elements equal to zero except those corresponding to rows of W_{ii},

$$x_i = [0 \cdots 0, 1 \cdots 1, 0 \cdots 0]^T.$$

Denote $D = diag(D_1, D_2, \cdots, D_S)$ such that $D_i = \sum_{j=1}^{S} W_{ij}$. It is easy to see that

$$\mathrm{sum}(W_{ii}) = x_i^T W x_i, \quad \sum_{j \neq i} \mathrm{sum}(W_{ij}) = x_i^T (D - W) x_i.$$

Since we want to find a partition which will maximize $\mathrm{sum}(W_{ii})$ while minimizing $\mathrm{sum}(W_{ij}), i \neq j$, we seek to minimize the following objective function by finding a set of indicator vectors x_i. The objective function is called min-max cut in [13] [14] which is a generalization of the normalized cut objective function [12] to the multi-way partition case.

$$\begin{aligned} MCut &= \frac{x_1^T(D-W)x_1}{x_1^TWx_1} + \frac{x_2^T(D-W)x_2}{x_2^TWx_2} + \cdots + \frac{x_S^T(D-W)x_S}{x_S^TWx_S} \\ &= \frac{x_1^TDx_1}{x_1^TWx_1} + \frac{x_2^TDx_2}{x_2^TWx_2} + \cdots + \frac{x_S^TDx_S}{x_S^TWx_S} - S. \end{aligned}$$

If we define $y_i = D^{1/2}x_i/||D^{1/2}x_i||_2$ and $Y_S = [y_1, \cdots, y_S]$, we have

$$MCut = \frac{1}{y_1^T\hat{W}y_1} + \frac{1}{y_2^T\hat{W}y_2} + \cdots + \frac{1}{y_S^T\hat{W}y_S} - S \tag{4}$$

where $\hat{W} = D^{-1/2}WD^{-1/2}$ and $y_i = \frac{D^{1/2}x_i}{\|D^{1/2}x_i\|_2}$. It is easy to see that the y_i are orthogonal to each other and normalized to have Euclidean norm one. If we insist that the y_i be constrained to inherit the discrete structure of the indicator vectors x_i, then we are leading to solve a combinatorial optimization problem which has been proved to be NP-hard even when $S = 2$ [12]. The idea of spectral clustering instead is to relax this constraints and allows the y_i to be an arbitrary set of orthonormal vectors. In this case, the minimum of Eq. 4 can be shown to be achieved by orthonormal basis $y_1, \cdots, y_S$ of the subspace spanned by the eigenvectors corresponding to the largest S eigenvalues of $\hat{W}$. Next we discuss how to assign the feature points to each clusters based on the eigenvectors.

We should first mention that the cluster assignment problem in spectral clustering is not well-understood yet. Here we follow the approach proposed in [15]. Denote $\hat{Y} = [\hat{y}_1, \cdots, \hat{y}_S]^T$ as the optimal solution of Eq. 4. The vectors $\hat{y}_i$ can be used for cluster assignment because $\hat{y}_i \approx D^{1/2}\hat{x}_i/||D^{1/2}\hat{x}_i||_2$, where $\hat{x}_i$ is the cluster indicator vector of $i-th$ cluster. Ideally, if W is partitioned perfectly into S clusters, then, the columns in $\hat{X} = [\hat{x}_i, \cdots, \hat{x}_S]^T$ of the $i-th$ cluster are the same, one for the $i-th$ row and zeros for the others. Two columns of different clusters are orthogonal to each other. This property is approximately inherited by $\hat{Y}$: two columns from two different clusters are orthogonal to each other, and those from one cluster are the same. We now pick a column of $\hat{Y}$ which has the largest norm, say, it belongs to cluster i, we orthogonalized the rest of the columns of $\hat{Y}$ against this column. We assign the columns to cluster i whose residual is small. We then perform this process S times. As discussed in [15], it is exactly the same procedure of **QR** decomposition with column pivoting applied to $\hat{Y}$. In particular, we compute **QR** decomposition of Y^T with column pivoting

$$Y^TE = \hat{Q}R = \hat{Q}[R_{11}, R_{12}]$$

where $\hat{Q}$ is a $S \times S$ orthogonal matrix, R_{11} is a $S \times S$ upper triangular matrix, and E is a permutation matrix. Then we compute a matrix $\hat{R}$ as

$$\hat{R} = R_{11}^{-1}[R_{11}, R_{12}]P^T = [I_S, R_{11}^{-1}R_{12}],$$

The matrix $\hat{R} \in R^{S \times N}$ can be considered as giving the levels of confidence of a point to be assigned to each cluster. Notice that the columns correspond to the feature points and the rows correspond to the clusters. The cluster membership of each feature point is determined by the row index of the largest element in absolute value of the corresponding column of $\hat{R}$. This provide us with a baseline spectral clustering method for motion segmentation which are quite robust in the presence of noise. Further improvement can be achieved as we discuss next.

We can assign a point to a cluster with high confidence if there is a very dominantly high confidence value in the corresponding column, however, we are not able to do this if two or more values in a column are very close to each other. Table 4.1 shows an example of the matrix $\hat{R} \in R^{3\times 10}$ that has 10 points extracted from 3 objects. The last row of the table shows the cluster membership of each point assigned by the row index of the highest absolute value. For instance, the point p_1 is assigned to cluster 2 because the second row value (0.329), is greater than the other row values (0.316 and 0.203). However, we cannot have much confidence of its membership because there is no dominant values in the corresponding column.

Table 1. An example of the matrix $\hat{R}$. There are 10 points extracted from 3 objects. The last row shows the assigned cluster

Cluster ID	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10
k = 1	0.316	0.351	0.876	0.331	0.456	0.562	0.086	0.275	0.072	0.119
k = 2	0.329	0.338	0.032	0.372	0.013	0.060	0.186	0.706	0.815	0.831
k = 3	0.203	0.017	0.031	0.173	0.566	0.556	0.775	0.126	0.094	0.113
Assigned Cluster	2	1	1	2	3	1	3	2	2	2

4.2 Refinement of Cluster Assignment for Motion Segmentation

The baseline spectral clustering shows its robustness for a noisy environment in spite of its hard clustering (it assigns each point to a cluster even though it does not have high confidence for it). The method alone, however, can sometimes fail in presence of severe noise. In this section, we discuss a two-phase approach whereby in phase one we assignment the cluster memberships for those feature points with high confidence levels, and in phase two we construct linear subspaces for each clusters based on the high confidence feature points, and assign the rest of the feature points by projecting onto these subspaces.

Our approach proceeds as follows. After computing $\hat{R}$ discussed in the previous section, the points of high confidence of each clusters are selected. Let's define $P_i = [p_{i1}, \cdots, p_{iN_i}]$ as the trajectory points in the cluster i. One of the easiest methods is to apply threshold to the values of each column, and if the highest value in the column is greater than the threshold, the point is assigned to the corresponding cluster. if it does not, let's categorize the point to cluster 0 which is in the state of temporarily pending to decide its cluster. Let's

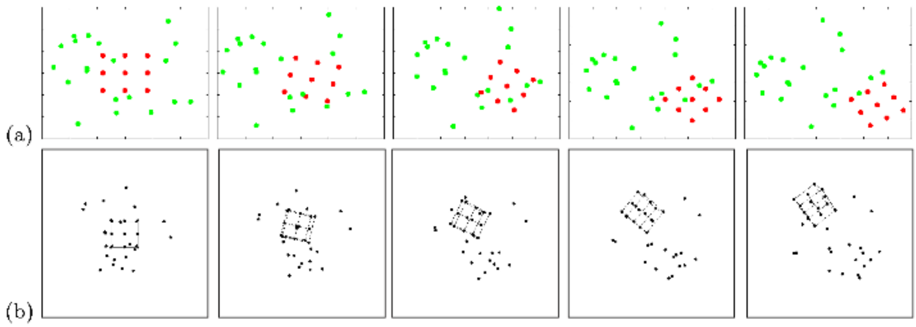


Fig. 1. Two synthetic video sequences used in [7] and [18] respectively. (a). 9 Red dots are foreground points and 20 green dots are background pixels (b). 24 background points and 14 foreground points. The foreground pixels are connected with lines.

define the pending points as $P_0 = [p_{01}, \cdots, p_{0N_0}]$. The next step is to compute subspace(2D) for $p_{i1}, \cdots, p_{iN_i}$, $i = 1, \cdots, S$ using Principal Component Analysis (**PCA**). Let's denote U_i as a subspace basis for the cluster i. We finally determine the cluster membership of each pending point by computing the minimum distance from the point to subspaces.

$$\hat{\theta}_j = arg \min_i ||p_{0j} - (c_i + U_i U_i^T (p_{0j} - c_i))||^2,$$

where $j = 1, \cdots, k$ and $c_i = \sum_{j=1}^{N_i} p_i j$.

The point p_{0j} is assigned to the cluster $\hat{\theta}_j$.

5 Experimental Results

Figure 1 shows two synthetic image sequences used for performance evaluation. Actually these images are used in [7] and [18]. Figure 1-(a), denoted as *Synthetic 1*, has 20 background points(green dots) and foreground points(red dots), and Figure 1-(b),denoted as *Synthetic 2*, has 20 background points and 14 foreground points. The foreground points are connected by lines for visualization purpose.

We performed experiments using not only the original tracking data but also the data added by independent Gaussian noise of mean 0 and standard deviation σ to the coordinates of all the points. For the noise data, we generate 5 sets for each $\sigma = 1, 2, 3, 4$, and compute the misclassification rate by simply averaging the 5 experiment results. We compare two methods proposed in this paper (One is k-way Min-Max cut clustering in Sec. 4.1 denoted as **Method 1**, and the other is a combination of the k-way Min-Max cut clustering and clustering refinement using subspace projection in Sec. 4.2 denoted as **Method 2**) to the Multi-stage optimization proposed in [18] denoted as **Multi-Stage**. Table 2 shows that the misclassification rates of the three methods over the different noise levels ($\sigma = 0, 1, 2, 3, 4$). **Method 2** and **Multi-Stage** yields better performance than **Method 1**. The two methods performs almost perfect for the sequences.

Table 2. Misclassification rate (%) for two synthetic sequences. The values in parenthesis are standard deviation. **Method 1** is k-way Min-Max cut clustering in Sec. 4.1 and **Method 2** is the k-way Min-Max cut clustering + clustering refinement using subspace projection in Sec. 4.2. **Multi-Stage** is the Multi-stage optimization proposed in [18].

Video Sequence	noise	$\sigma = 0$	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 4$
Synthetic 1	**Method 1**	0.0	1.4	1.4	0.7	0.7
	Method 2	0.0	0.0	0.0	0.0	0.0
	Multi-Stage	0.0	0.0	0.7	0.0	0.0
Synthetic 2	**Method 1**	8.2	10.6(1.6)	11.7(2.1)	11.7(3.6)	13.2(1.7)
	Method 2	0.0	0.0	0.0	0.0	0.59(1.3)
	Multi-Stage	0.0	0.0	0.0	0.0	0.59(1.3)

We experimented with the real video sequences used in [18]. In all the sequences, one object is moving while background is simultaneously moving because of the camera moving. Let's denote the video sequences as *video1*, *video2* and *video3* respectively. We synthesize one more test video sequence by overlaying the foreground feature points in *video1* to *video2*, which has 2 moving objects and background. Let's denote the video sequences as *video4*. Figure 2 shows selected 5 frames of the four sequences.

We also performed experiments using not only the original tracking data but also the data added by independent Gaussian noise of mean 0 and standard deviation σ to the coordinates of all the points. For the noise data, we generate 5 sets for each $\sigma = 3, 5, 7, 10$, and compute the misclassification rate by simply averaging the 5 experiment results.

Table 3 shows the misclassification rates of the three methods over the different noise levels ($\sigma = 0, 3, 5, 7, 10$). The table shows that **Method 2** can classify motion perfectly even for the severe noise presence. It is very robust and stable to noise. **Method 1** performs very well for noise-free environment, but it misclassifies some points in the presence of noise.

Multi-Stage performs very well for *video1* through *video3* which have one moving foreground object and background. It, however, does not yield good performance for *video4* which has two moving foreground objects and background in the presence of noise. Based on our experiments, the method also suffer from local minima problem. Using the same data, it yields different results based on the initialization. That is the reason the standard deviation of the method is too high shown in Table 3.

6 Conclusions

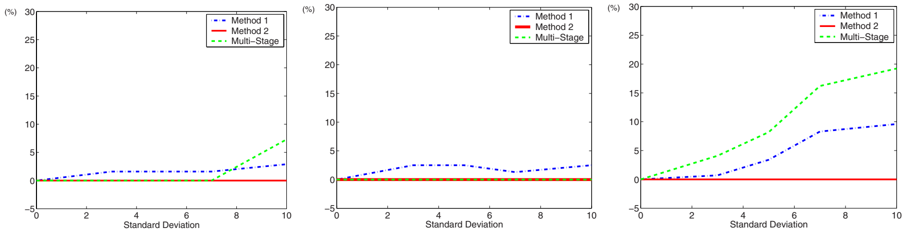
In this paper, we mathematically prove the shape interaction matrix can be computed using **QR** decomposition which is more effective than **SVD**. We solve the motion segmentation problem using spectral graph clustering technique because the shape interaction matrix has a very similar form to the weight matrix of graph. We apply the Spectral Relaxation K-way Min-Max cut clustering

Fig. 2. Real Video sequences with the feature points. 1^{st} row: *video1*, 2^{nd} row: *video2*, 3^{rd} row: *video3*, 4^{th} row: *video4* (foreground feature points in *video1* are overlaid in *video4*). Red dots correspond to background while green dots correspond to foreground. The yellow cross marks in *video4* represent the foreground feature points of *video1*

Table 3. Misclassification rate (%) for the real video sequences. The values in parenthesis are standard deviation.

Video Sequence	noise	$\sigma = 0$	$\sigma = 3$	$\sigma = 5$	$\sigma = 7$	$\sigma = 10$
video1	**Method 1**	0.0	0.0	0.0	0.0	0.0
	Method 2	0.0	0.0	0.0	0.0	0.0
	Multi-Stage	0.0	0.0	0.0	0.0	0.0
video2	**Method 1**	0	1.6(1.2)	1.6(1.2)	1.6(1.2)	2.9(1.7)
	Method 2	0.0	0.0	0.0	0.0	0.0
	Multi-Stage	0.0	0.0	0.0	0.0	7.3(16.3)
video3	**Method 1**	0.0	2.5(0.01)	2.5	1.3	2.5
	Method 2	0.0	0.0	0.0	0.0	0.0
	Multi-Stage	0.0	0.0	0.0	0.0	0.0
video4	**Method1**	0.0	0.7(1.6)	3.4(4.7)	8.3(5.0)	9.6(6.5)
	Method2	0.0	0.0	0.0	0.0	0.7(1.6)
	Multi-Stage	0.0	4.1(9.3)	8.2(9.6)	16.2 (13.2)	19.23 (9.87)

method [13] [14] to shape interaction matrix. It provides a relaxed cluster indication matrix. **QR** decomposition is applied to the matrix, which generate a new cluster indication matrix, to determine the cluster membership of each point. The values of the new cluster indication matrix reflect confidence level for each point to be assigned to clusters. This method yields a good performance in noise free environment, but it is, sometimes, sensitive to noise. We propose a



(a) *video2* (b) *video3* (c) *video4*

Fig. 3. Graph for misclassification rate. Graph of *video1* is not depicted here because all the three methods performs perfectly. **Method 1**: Dashed-dot blue line, **Method 2**: Red line and **Multi-Stage**: Dashed green line

robust motion segmentation method by combining the spectral graph clustering and subspace separation to compensate noise problem. Initially, we assign only points of high confidence to clusters based on the cluster indication matrix. We compute subspace for each cluster using the assigned points. We finally determine the membership of the other points, which are not assigned to a cluster, by computing the minimum residual when they are projected to the subspace.

We applied the proposed method to two synthetic image sequences and four real video sequences. **Method 2** and **Multi-Stage** produce almost perfect performance for the synthetic image sequences in the presence of noise. Experiments also show that the proposed method, **Method 2**, performs very well for the real video sequences even in the sever noise presence. It performs better than Multi-Stage optimization method [18] for real video sequences in which there are more than two objects.

Acknowledgments. The authors would like to thank Prof. Kanatani and his colleagues for generously providing their data for this work. The work is supported in part by NSF grant CCF-0305879.

References

1. C. Tomasi and T. Kanade. Shape and motion from image streams under orthography - a factorization method. *International Journal of Computer Vision*, 9(2):137–154, 1992.
2. J. Costeira and T. Kanade. A multi-body factorization method for motion analysis. *International Journal of Computer Vision*, 29(3):159–179, 1998.
3. C. W. Gear. Multibody grouping from motion images. *International Journal of Computer Vision*, 29(2):133–150, 1998.
4. N. Ichimura. Motion segmentation based on factorization method and discriminant criterion. In *Proc. IEEE Int. Conf. Computer Vision*, pages 600–605, 1999.
5. Y. Wu, Zhang Z., and J. Y. Huang, T. Lin. Multibody grouping via orthogonal subspace decomposition. In *Proc. IEEE Computer Vision and Pattern Recognition*, volume 2, pages 252–257, 2001.

6. K. Kanatani. Motion segmentation by subspace separation and model selection. In *Proc. IEEE Int. Conf. Computer Vision*, volume 2, pages 586–591, 2001.
7. K. Kanatani. Motion segmentation by subspace separation and model selection:model selection and reliability evaluation. *Intl. J. of Image and Graphics*, 2(2):179–197, 2002.
8. L. Zelnik-Manor and M. Irani. Degeneracies, dependencies and their implications in multi-body and multi-sequence factorizations. In *Proc. IEEE Computer Vision and Pattern Recognition*, 2003.
9. C. Bregler and H. Hertzmann, A. Biermann. Recovering non-rigid 3d shape from image streams. In *Proc. IEEE Computer Vision and Pattern Recognition*, 2000.
10. L. Torresani, D. B. Yang, E. J. Alexander, and C. Bregler. Tracking and modeling non-rigid objects with rank constraints. In *Proc. IEEE Computer Vision and Pattern Recognition*, 2001.
11. M. Brand. Morphable 3d models from video. In *Proc. IEEE Computer Vision and Pattern Recognition*, 2001.
12. J. Shi and J. Malik. Normalized cut and image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
13. M. Gu, H. Zha, C. Ding, X. He, and H. Simon. Spectral relaxation models and structure analysis for k-way graph clustering and bi-clustering. In *Technical Report,Department of Computer Science and Engineering, CSE-01-007,*, 2001.
14. C. Ding, X. He, H. Zha, M. Gu, and H. Simon. Spectral min-max cut for graph partitioning and data clustering. In *IEEE International Conference on Data Mining*, pages 107–114, 2001.
15. H. Zha, C. Ding, M. Gu, X. He, and H. Simon. Spectral relaxation for k-means clustering. In *Neural Information Processing Systems 14*, pages 1057–1064, 2002.
16. N. Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. on Systems, Man and Cybernetics*, SMC-9(1):62–66, 1979.
17. G. H. Golub and C. F. Van Loan. *Matrix Computation, 3rd Ed.* Johns Hopkins Univ. ress, 2000.
18. K. Kanatani and Y. Sugaya. Multi-stage optimization for multi-body motion segmentation. In *Proc. Australia-Japan Advanced Workshop on Computer Vision*, pages 25–31, 2003.