Multibody Grouping via Orthogonal Subspace Decomposition

Ying Wu†, Zhengyou Zhang‡, Thomas S. Huang♯, John Y. Lin♯

† Northwestern University, 2145 Sheridan, Evanston, IL 60208

‡ Microsoft Research, One Microsoft Way, Redmond, WA 98052

♯ University of Illinois at Urbana-Champaign, 405 N. Mathews, Urbana, IL 61801

† yingwu@ece.nwu.edu, ‡ zhang@microsoft.com, ♯ {huang,jy-lin}@ifp.uiuc.edu

Abstract

Multibody structure from motion could be solved by the factorization approach. However, the noise measurements would make the segmentation difficult when analyzing the shape interaction matrix. This paper presents an orthogonal subspace decomposition and grouping technique to approach such a problem. We decompose the object shape spaces into signal subspaces and noise subspaces. We show that the signal subspaces of the object shape spaces are orthogonal to each other. Instead of using the shape interaction matrix contaminated by noise, we introduce the shape signal subspace distance matrix for shape space grouping. Outliers could be easily identified by this approach. The robustness of the proposed approach lies in the fact that the shape space decomposition alleviates the influence of noise, and has been verified with extensive experiments.

1 Introduction

Most realistic vision tasks involve multibody motions. A simple scenario of tracking a car with a moving camera already involves two moving objects (ignoring the wheels). Multibody tracking and segmentation are essential for many applications including structure from motion, human-computer interaction, surveillance, and video coding. In this paper, we only consider the problem of segmentation, assuming tracking is done, e.g., with the KLT feature tracker [9]. We allow outliers in the matches, though.

Among many techniques proposed in the literature, factorization is particularly interesting for three reasons: no knowledge of the number of objects is required; no initial segmentation is necessary; and a measurement matrix is *globally* factorized into two matrices (one for motion, and the other for structure), achieving higher robustness to data noise. Factorization was originally developed by Tomasi and Kanade for structure from motion of a single object under orthographic projection [11], and was later extended to paraperspective or affine cameras in [8]. A sequential version was proposed in [7]. Attempts were made to generalize the technique for full perspective [10], but due to the inherent nonlinearity of camera projection, some preprocessing (especially depth estimation) is necessary, which leads to a sub-optimal solution.

Costeira and Kanade proposed a first algorithm for multibody segmentation based on factorization [2]. Similar approaches were later developed for linearly moving objects [4] and for deformable objects [1]. In this paper, we only consider Costeira and Kanade's original problem: Given p feature points tracked over T frames with an affine camera, determine the number of moving objects in the scene, their motions and their structures. This is a formidable problem because of the inherent combinatorial property and data noise. Costeira and Kanade [2] based their segmentation algorithm on a so-called shape interaction matrix $\mathbf{Q}$ (see below). If two features belong to two different objects, their corresponding element in $\mathbf{Q}$ should be zero; otherwise, the value should be non-zero. They then grouped features into objects by thresholding and sorting $\mathbf{Q}$. Gear [3] formulated the problem as graph matching by placing appropriate weights on the graph edges, which are difficult to determine. Unfortunately, the performance of both techniques degrades quickly when data points are corrupted with noise; the reason is that the relationship between data noise and the coefficients of $\mathbf{Q}$ (or weights of the graph edges) is very complicated, making it hard to determine an appropriate threshold. Ichimura [5] proposed an improved algorithm by applying a discriminant criterion for thresholding, but the discriminant analysis is still performed on the elements of $\mathbf{Q}$, resulting in a similar degradation with noise. To avoid this problem, Kanatani [6] proposed to work in the original data space by incorporating such techniques as dimension correction (fitting a subspace to a group of points and replacing them with their projections onto the fitted subspace) and model selection

0-7695-1272-0/01 $10.00 © 2001 IEEE

(using a geometric information criterion to determine whether two subspaces can be merged).

In this paper, we propose a new grouping technique based on orthogonal subspace decomposition. After performing a singular value decomposition (SVD) of the measurement matrix, we decompose the object shape spaces into signal subspaces and noise subspaces. We show the signal subspaces of the object shape spaces are orthogonal to each other. Instead of using the shape interaction matrix contaminated by noise, we introduce the *shape signal subspace distance matrix* (or *subspace distance matrix* for short), $\mathbf{D}$, for shape space grouping, based on a distance measure defined over the subspaces. The values of most entries of $\mathbf{D}$ are around 0 or 1, making the grouping procedure much easier. Outliers are easily identifiable because their distances to all object subspaces are comparable. The robustness of the proposed approach lies in the fact that the shape space decomposition alleviates the influence of noise. This has been verified with extensive experiments.

Section 2 reviews the factorization method. Section 3 describes our orthogonal subspace approach. Section 4 provides experimental results with both simulated and real data.

2 The Factorization Method

Suppose there are n independently moving objects in a scene, and the structure of each object is represented by a set of p_k 3D points, i.e.,

$$S_k = \begin{bmatrix} x_{k1} & x_{k2} & \cdots & x_{kp_k} \\ y_{k1} & y_{k2} & \cdots & y_{kp_k} \\ z_{k1} & z_{k2} & \cdots & z_{kp_k} \end{bmatrix} \quad (1)$$

So, the 3D structure of the whole scene could be represented by

$$S = \begin{bmatrix} S_1 & & \\ & \cdots & \\ & & S_n \end{bmatrix} \quad (2)$$

where the off-block-diagonal elements are equal to zero. When we assume affine projection (orthographic, weak perspective or paraperspective), the projection of the scene on the image plane is:

$$\begin{bmatrix} u_1 & \cdots & u_p \\ v_1 & \cdots & v_p \end{bmatrix} = [M_1\ M_2\ \ldots\ M_n]S + \mathbf{t} \quad (3)$$

where M_k ($k = 1, \ldots, n$) is the projection matrix related to object k, $\mathbf{t}$ is the camera translation, and $p = \sum_k p_k$. $\mathbf{t}$ could be eliminated by subtracting the mean of the 2D projections. When considering T frames, we have:

$$W = \begin{bmatrix} u_1^{(1)} & \ldots & u_p^{(1)} \\ v_1^{(1)} & \ldots & v_p^{(1)} \\ u_1^{(2)} & \ldots & u_p^{(2)} \\ v_1^{(2)} & \ldots & v_p^{(2)} \\ \ldots & \ldots & \ldots \\ u_1^{(T)} & \ldots & u_p^{(T)} \\ v_1^{(T)} & \ldots & v_p^{(T)} \end{bmatrix} \quad (4)$$

And we can also write:

$$W = \begin{bmatrix} M_1^{(1)} & \ldots & M_n^{(1)} \\ \ldots & \ldots & \ldots \\ M_1^{(T)} & \ldots & M_n^{(T)} \end{bmatrix} \begin{bmatrix} S_1 & & \\ & \cdots & \\ & & S_n \end{bmatrix} \quad (5)$$

For each rigid object, its structure and motion could be solved by the factorization method [11] based on SVD decomposition, i.e.,

$$W_k = U_k \Sigma_k V_k^T \quad (6)$$

And its motion $\hat{M}$ and structure $\hat{S}$ could be factorized by:

$$\hat{M} = U_k \Sigma_k^{\frac{1}{2}} A \quad \text{and} \quad \hat{S} = A^{-1} \Sigma_k^{\frac{1}{2}} V_k^T \quad (7)$$

where A is an invertible matrix and can be solved using the fact that $\hat{M}$ must have certain properties. Therefore, we could write:

$$W = [U_1 \ldots U_n] \begin{bmatrix} \Sigma_1 & & \\ & \cdots & \\ & & \Sigma_n \end{bmatrix} \begin{bmatrix} V_1^T & & \\ & \cdots & \\ & & V_n^T \end{bmatrix}$$

Let

$$\mathbf{V} = \begin{bmatrix} V_1 & & \\ & \cdots & \\ & & V_n \end{bmatrix}$$

For multibody structure from motion problem, the identities of the set of feature points are unavailable, except the correspondences are given. Therefore, $\mathbf{V}$ would not be a block diagonal matrix, instead, structure vectors of different objects would be mixed up. In order to solve the structure and motion, we have to reveal the identities of each feature points, i.e., solve the multibody grouping problem.

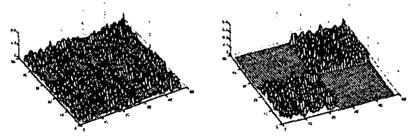
Fortunately, it is easy to show that the shape interaction matrix $\mathbf{Q}$, defined by

$$\mathbf{Q} = \mathbf{V}\mathbf{V}^T \quad (8)$$

is motion invariant [6]. More interestingly, $\mathbf{Q}$ has a very nice property:

$$\mathbf{Q}_{ij} = \begin{cases} 0 & \text{if point } i \text{ and } j \text{ belong to different objects} \\ * & \text{if point } i \text{ and } j \text{ belong to the same object} \end{cases} \quad (9)$$

where $*$ indicates any possible value. Such a property provides a clue for the segmentation of multiple objects, i.e., if $V_i^T V_j \neq 0$, the i-th and j-th feature points should be grouped together; otherwise, they may belong to different objects. Therefore, the segmentation could be achieved by permuting $\mathbf{V}$ to make $\mathbf{Q}$ block diagonal. This was the basic idea in [2] for multibody structure from motion. It could be illustrated in Figure 1, where (a) displays the original $\mathbf{Q}$, while (b) displays the $\mathbf{Q}$ after permutation.



(a) (b)

Figure 1: The shape interaction matrix $\mathbf{Q}$ before and after permutation

It is noticed that the nice property of the shape interaction matrix $\mathbf{Q}$ is valid only for the ideal case where there are no noise and outliers. Unfortunately, in practice, the extraction and tracking of feature points would incur some inaccuracy, thus the measurement noise is unavoidable. Thus, even if two feature points belong to different objects, $\mathbf{Q}_{ij}$ may not be equal to zero. This will be illustrated in Figure 3.

3 Orthogonal Subspace Approach

This section describes our approach to this problem based on orthogonal subspace decomposition and grouping methods.

3.1 Orthogonal Subspace Decomposition

Any vector x in a Hilbert space $\mathcal{R}^n$ could be represented by a summation of two projection vectors from two subspaces M and $M^\perp$, i.e., $x = P_M x + (I - P_M)x$, where P_M is the projection matrix of the subspace M and $I - P_M$ is that of the subspace $M^\perp$. At the same time, if a Hilbert space is spanned by a set of m linearly independent vectors $X = [x_1 \dots x_m]$, i.e., $M = span\{x_1, \dots, x_m\}$, then the projection matrix onto the space M will be given by:

$$P_M = X\langle X, X\rangle^{-1} X^T \quad (10)$$

where $\langle X_1, X_2\rangle$ defines the inner product of X_1 and X_2 in the Hilbert space, and the projection is given by $P_M x = X\langle X, X\rangle^{-1} X^T x, \forall x \in \mathcal{R}^n$. Usually, we could write $P_M = X(X^T X)^{-1} X^T$. Obviously, if x_k's are orthogonal to each other, P_M will be given by $P_M = XX^T$. It is also very easy to show some properties of P_M, such as $P_M P_M = P_M$, $P_M^T = P_M$, $P_M^\perp = I - P_M$ and $P_M^\perp P_M = \mathbf{0}$.

It is easy to verify that the projection matrix P_M is unique for any linearly independent vector set which spans M, since $P_M = XA(A^T X^T X A)^{-1} A^T X = X(X^T X)^{-1} X^T$. Therefore, we can use the projection matrix P_M to characterize the space M. In this sense, we could define the distance between two subspaces. Suppose S_1 and S_2 are two subspaces of $\mathcal{R}^n$, and $dim(S_1) = dim(S_2)$, then the distance between S_1 and S_2 could be defined by:

$$\mathcal{D}(S_1, S_2) = ||P_1 - P_2||_2 \quad (11)$$

where P_1 and P_2 are two projection matrix onto S_1 and S_2, respectively.

3.2 Signal Subspace vs. Noise Subspace

Given a set of vectors $X = [x_1 \dots x_m]$, where $x_k \in \mathcal{R}^n$, we could decompose X by SVD, i.e., $X = U\Sigma V^T$, where $\Sigma = diag(\sigma_1, \dots, \sigma_r, \sigma_{r+1}, \dots, \sigma_m)$, and $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m$.

When the data are clear and the actual rank of X is r, then we should observe $\sigma_j = 0, j \geq r+1$. On the other hand, if x_k was contaminated by noise, there would be many small singular values, but $\sigma_r \gg \sigma_{r+1}$. In this case, $\{\sigma_1, \dots, \sigma_r\}$ belong to the signal subspace S_s, and $\{\sigma_{r+1}, \dots, \sigma_m\}$ belong to the noise subspace S_n. The singular values suggest a way to decompose the data space into signal subspace and noise subspace. Thus,

$$X = X_s + X_n = U_s \Sigma_s V_s^T + U_n \Sigma_n V_n^T \quad (12)$$

And the projection matrix P_s onto the signal subspace S_s would be given by $P_s = \hat{X}_s(\hat{X}_s^T \hat{X}_s)^{-1} \hat{X}_s^T$, where $\hat{X}_s$ is a set of independent vectors from X_s. Similarly, the projection matrix P_n onto the noise subspace S_n could be written as $P_n = \hat{X}_n(\hat{X}_n^T \hat{X}_n)^{-1} \hat{X}_n^T$. It is easy to show that S_s and S_n are orthogonal to each other ($S_s \perp S_n$), and $P_s + P_n = I$,

3.3 Our Multibody Grouping Approach

Let us first assume that we have grouped feature points of different objects, we could write $W = U\Sigma V^T$ by SVD, where $V = [V^{(1)} \dots V^{(n)}]$ and $V^{(i)} = [v_1^{(i)} \dots v_{p_i}^{(i)}]$. $S^{(i)} = span\{v_1^{(i)}, \dots, v_{p_i}^{(i)}\}$ defines the shape space for object i.

Theorem I If no noise is present, we have $S^{(i)} \perp S^{(j)}, \forall i \neq j$. It means that the shape spaces of each object spanned by $V^{(i)}$ are orthogonal to each other.

This theorem is easy to show. $\forall v_m^{(i)} \in V^{(i)}$ and $\forall v_n^{(j)} \in V^{(j)}$, we have $v_m^{(i)T} v_n^{(j)} = 0$ according to the nice property of the shape interaction matrix $\mathbf{Q}$ in equation 9. Thus, suppose we have selected $r = rank(V^{(k)})$ independent vectors $\hat{V}^{(k)} = [\hat{v}_1^{(k)} \dots \hat{v}_r^{(k)}]$ for the k-th object, we could write the projection matrix onto the k-th shape space by

$$P^{(k)} = \hat{V}^{(k)} (\hat{V}^{(k)T} \hat{V}^{(k)})^{-1} \hat{V}^{(k)T}$$

So, we easily have $P^{(i)} P^{(j)} = \mathbf{0}$, and

$$\sum_{k=1}^{n} P^{(k)} = I$$

Theorem II If measurements contain noise, the signal subspaces of each shape space are still orthogonal to each other, i.e., $S_s^{(i)} \perp S_s^{(j)}, \forall i \neq j$, where the signal subspace $S_s^{(i)}$ is separated by SVD from equation 12.

Again, when a set of independent vectors is selected for each signal subspace, we can describe the projection matrix onto the signal subspace of the k-th object by

$$P_s^{(k)} = \hat{V}_s^{(k)} (\hat{V}_s^{(k)T} \hat{V}_s^{(k)})^{-1} \hat{V}_s^{(k)T}$$

We have $P_s^{(i)} P_s^{(j)} = \mathbf{0}$, and

$$\sum_{k=1}^{n} P_s^{(k)} = nI - \sum_{k=1}^{n} P_n^{(k)}$$

The shape interaction matrix $\mathbf{Q}$ could be used for grouping the feature points for different objects if the measurements are not contaminated by noise. Unfortunately, noisy data would make use of $\mathbf{Q}$ difficult for grouping. Noticing such a nice property that the signal subspaces of shape space for each object are orthogonal to each other even under noisy measurements, we could make use of the signal subspaces to alleviate the noise influence. Instead of using the shape interaction matrix $\mathbf{Q}$, we shall introduce the shape signal subspace distance matrix $\mathbf{D}$, which would be cleaner than $\mathbf{Q}$.

Suppose N groups of feature points have been identified. This could be done by analyzing the $\mathbf{Q}$ matrix. We can simply threshold the $\mathbf{Q}$ matrix or use the discriminant analysis method described in [5]. If $\mathbf{Q}_{ij} \geq t_Q$, the i-th and j-th feature point will be put in the same group. If the measurements contain no noise and $\mathbf{Q}$ is very clean, the threshold t_Q is easy to set. Unfortunately, this is not apparent under noisy data. But we could generally set a higher threshold, which would result in several group fragments corresponding to the same object. And these group fragments should be grouped together later. Meanwhile, there would be some feature points that may not be grouped to any other group fragments. We will also handle them later.

Each of such group fragments $V^{(k)} = [v_1^{(k)} \dots v_{nk}^{(k)}]$ would span a space $S^{(k)} = span\{v_1^{(k)} \dots v_{nk}^{(k)}\}$. Since the space $S^{(k)}$ contains noise, we could identify its signal subspace $S_s^{(k)}$ by the method described before, and we could also calculate the projection matrix $P_s^{(k)}$ to represent the k-th group fragment. So, the *shape signal subspace distance matrix* is defined by:

$$\mathbf{D} = \{\mathbf{D}_{ij} : \mathbf{D}_{ij} = \mathcal{D}(P_s^{(i)}, P_s^{(j)}), \forall i, j \leq N\} \qquad (13)$$

where $\mathbf{D}_{ij} = \mathcal{D}(P_s^{(i)}, P_s^{(j)})$ is the distance between the signal subspaces of the i-th and j-th group fragments. We notice that if $P_s^{(i)}$ and $P_s^{(j)}$ are orthogonal, then $\mathbf{D}_{ij} = 0$; if $P_s^{(i)}$ and $P_s^{(j)}$ characterize the same space, then $\mathbf{D}_{ij} = 1$; otherwise, $0 < \mathbf{D}_{ij} < 1$. Since the signal subspace excludes the noise, we would see that the $\mathbf{D}$ matrix is cleaner than $\mathbf{Q}$. We have observed such a property of the $\mathbf{D}$ matrix in our extensive experiments. The values of most entries of $\mathbf{D}$ are around 1 or 0. Consequently, further grouping of the fragments based on $\mathbf{D}$ would be performed easily. We simply set a threshold for $\mathbf{D}$. Our approach tolerate a large variation in the value of this threshold because it is more discriminating in $\mathbf{D}$ for different group fragments. The procedure is simple: if $\mathbf{D}_{ij} \leq t_D$, then we merge the i-th and j-th group fragments together.

After these fragments are grouped together into object groups, we calculate and update the signal subspace and its projection matrix $P_s^{(k)}$ for each object. Outliers could be simply identified if the orthogonal projections of such feature point onto all the object spaces are nearly equal or comparable, because such feature point could not be confidently classified into any of the object spaces.

The outline of the proposed multibody segmentation method is summarized below:

1. Decompose measurement matrix W by SVD and get $V = \{v_j : 1 \leq j \leq m, v_j \in \mathcal{R}^r\}$;
2. Calculate the shape interaction matrix $\mathbf{Q} = VV^T$;
3. Group v_j into group fragments $F_k (1 \leq k \leq P_f)$ based on $\mathbf{Q}$;
4. Calculate signal subspace projection matrices $P_s^{(k)} (1 \leq k \leq P_f)$ for each F_k;
5. Calculate signal subspace distance matrix $\mathbf{D} = \{d_{ij} : d_{ij} = \mathcal{D}(P_s^{(i)}, P_s^{(j)}\}$;
6. Group F_k into objects $O_i (1 \leq i \leq P)$ based on $\mathbf{D}$;
7. Calculate signal subspace of each object O_i and identify outliers, classify all feature points and update all signal subspaces;

8. Calculate motion and structure for each object O_i based on segmentation.

We shall ask the intuition why the grouping based on $\mathbf{D}$ is more robust than that based on $\mathbf{Q}$. The basic difference between $\mathbf{D}$ and $\mathbf{Q}$ is that $\mathbf{Q}$ is point-level interaction, while $\mathbf{D}$ is group-level interaction. Based on $\mathbf{Q}$, a single outlier could even fool the grouping because such grouping is based only on point similarities and the relationship between data noise and the coefficients of $\mathbf{Q}$ is very complicated. Matrix $\mathbf{D}$, however, introduces more robustness because the grouping depends on a number of feature points, instead of one. Furthermore, since signal subspaces are used, the entries of $\mathbf{D}$ are more or less around 0 or 1, which considerably facilitates the grouping decision.

4 Experiments

We provide in this section experimental results with both simulated and real data.

4.1 Simulation

We have performed some simulations and quantitative analysis on a synthetic scene. The scene consists of two sets of 3D points. One set of 60 points describes a 3D cube, and the other set of 40 points represents the background. We permute the orders of the data points, and we know their identities. The image resolution is 160×120 pixels. Figure 2 shows two views of the synthetic scene. These two sets of points undergo different and independent motions, and we capture 4 frames. We introduce noise to the measurement data by adding a zero-mean Gaussian noise to the coordinates of the projected points.

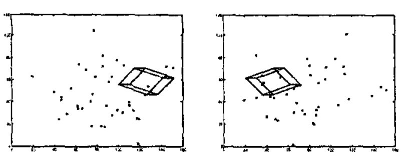


Figure 2: Two views of a synthetic scene for simulation. The scene consists of a set of points representing background, and another set of points from a 3D cube.

Our first simulation set the standard deviation of the Gaussian noise to 2 pixels. Figure 3(a) shows the $\mathbf{Q}$ matrix of the shape space after the factorization. Obviously, the data points are not clustered. We perform a segmentation method based on a linear discriminate analysis, which is similar to the method in [5], then permute the order of $\mathbf{Q}$. From Figure 3(b), we can see that the data are partially clustered, but not all. Figure 3(c) illustrates the $\mathbf{Q}$ matrix after the grouping by our proposed method. The two groups representing two objects become pretty clear in the permuted $\mathbf{Q}$ matrix. We can see that many $\mathbf{Q}_{ij}$ (i and j belong to different objects) are not zeros. Furthermore, we project the data of each object onto its signal subspace. We have verified that the signal subspaces of the two objects are indeed orthogonal ($P_s^{(1)}P_s^{(2)} = 0$), which is illustrated by Figure 3(d). Almost all $\mathbf{Q}_{ij}$ (i and j belong to different objects) are zeros.

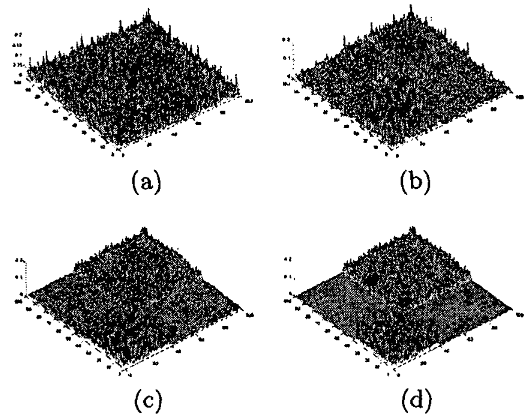


Figure 3: Experiments. (a) The noisy $\mathbf{Q}$ matrix; (b) The permuted $\mathbf{Q}$ matrix based on Ichimura's method; (c) The permuted $\mathbf{Q}$ matrix of our method; (d) The permuted $\mathbf{Q}$ matrix of de-noised data.

The second simulation compares three methods under noisy measurements. The first one is a simple thresholding method for grouping based on $\mathbf{Q}$. The second one is the method based on a linear discriminant analysis similar to [5]. And the third one is our method. The noise level ranges from 0 to 5 pixels with interval of 0.1 pixels (50 noise levels in total). We perform 30 runs for each noise level and compute the average of the mis-grouping error. The result is shown in Figure4. It clearly shows that the proposed subspace decomposition method performs the best. It is very robust to noise measurements. Only when the noise level goes up to more than 4.5 pixels, our method outputs small mis-grouping error. The other two methods are not at all robust to noise.

We have also constructed another synthetic scene with 3 independent moving objects, and we have observed a similar result.

4.2 Real Video

We have also applied our algorithm to some real video sequences. The first sequence contains a moving hand taken by a moving camera. We detect and track 20 feature points in 15 frames. The segmentation of the hand and the background is shown in Figure 5.

Another sequence contains two independent moving objects. The camera motion is not large. We detect and track 26 feature points in 20 frames. A couple

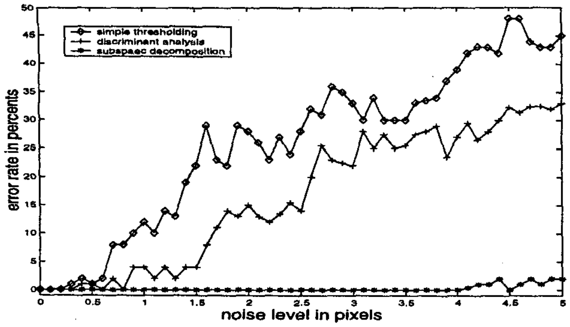


Figure 4: Comparison results. We compare the mis-grouping error rate against noise levels of three methods. Our method works the best.

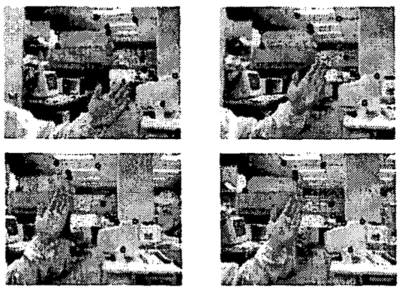


Figure 5: Motion segmentation. 20 feature points are detected and tracked in 15 frames. Feature points that belong to the background and the hand are shown by black "o" and red "+".

of points were occluded during tracking. The segmentation of two hands and the background is shown in Figure 6. Our algorithm performed very well on these two sequences.

5 Conclusions

The factorization method proposed by Costeira and Kanade provides a flexible way for multibody structure from motion and motion segmentation. The segmentation is based on the shape interaction matrix $\mathbf{Q}$ that indicates whether two feature points belong to the same object or not. However, their method is plagued by measurement noise. Measurement noise contributes to the distortion of the coefficients of $\mathbf{Q}$ in a very complicated way. It is not robust for grouping in point level. In this paper, we have proposed a grouping based on the shape signal subspace distance matrix $\mathbf{D}$ that describes the relations among different groups of feature points. The shape signal subspace for a group of points is obtained by the subspace decomposition technique. We have shown that the signal subspaces for different objects are orthogonal to each other. Since signal subspaces are separated from noise, $\mathbf{D}$ is more robust for grouping. Extensive experiments have confirmed the robustness of our approach.

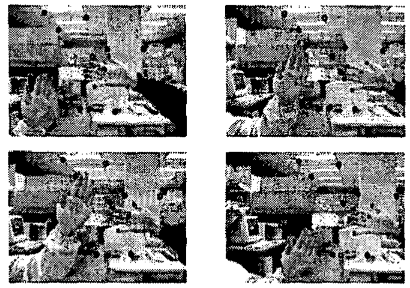


Figure 6: The scene contains two moving object and a moving background represented by 26 feature points. Background is shown by black "o", and the two moving hands by red "+" and blue "x" respectively.

It would be interesting to investigate the incremental method for multibody motion analysis. We shall also extend our approach to multiple persons tracking and articulated object analysis.

Acknowledgments

This work was supported by Northwestern Faculty Startup Funds, NSF Grants CDA-96-24396 and EIA-99-75019 and NSF Alliance Program.

References

[1] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 690–696, Hilton Head Island, SC, June 2000.

[2] J. Costeira and T. Kanade. A multibody factorization method for independently moving objects. *Int'l Journal of Computer Vision*, 29:159–179, 1998.

[3] C. Gear. Multibody grouping from motion images. *Int'l Journal of Computer Vision*, 29:133–150, 1998.

[4] M. Han and T. Kanade. Reconstruction of a scene with multiple linearly moving objects. In *IEEE Conf. on Computer Vision and Pattern Recognition*, volume II, pages 542–549, Hilton Head Island, SC, June 2000.

[5] N. Ichimura. Motion segmentation based on factorization method and discriminant criterion. In *Proc. IEEE Int'l Conf. on Computer Vision*, pages 600–605, Greece, Sept. 1999.

[6] K. Kanatani. Motion segmentation by subspace separation and model selection. In *Proc. IEEE Int'l Conf. on Computer Vision*, Vancouver, Canada, July 2001.

[7] T. Morita and T. Kanade. A sequential factorization method for recovering shape and motion from image streams. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19:858–867, 1997.

[8] C. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19:206–218, 1997.

[9] J. Shi and C. Tomasi. Good feature to track. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 593–600, Seattle, June 1994.

[10] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *Proc. European Conf. on Computer Vision*, volume II, pages 709–720, 1996.

[11] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography – a factorized method. *Int'l Journal of Computer Vision*, 9:137–154, 1992.