Robust Motion Segmentation by Spectral Clustering

Hongbin Wang and Phil F. Culverhouse
Centre for Robotics Intelligent Systems
University of Plymouth
Plymouth, PL4 8AA, UK
{hongbin.wang, P.Culverhouse}@plymouth.ac.uk

Abstract

Multibody motion segmentation is important in many computer vision tasks. One way to solve this problem is factorization. But practically segmentation is difficult since the shape interaction matrix is contaminated by noise. This paper presents a novel approach to robustly segment multiple moving objects by spectral clustering. We introduce two new affinity matrixes. One is based on the shape interaction matrix and the other one is based on the motion trajectory. By computing the sensitivities of the larger eigenvalues of a related Markov transition matrix with respect to perturbations in the affinity matrix, we improve the piecewise constant eigenvectors condition dramatically. The feature points are mapped into a low dimensional subspace and clustered in this subspace using a graph spectral approach. This makes clustering much more reliable and robust, which we confirm with experiments.

1 Introduction

Motion segmentation is one of the important tasks in computer vision. It has many applications including structure from motion, video coding and human computer interaction. Among many techniques discussed in the literature, Costeira and Kanade[3]proposed an algorithm for multibody motion segmentation based on factorization. Given tracked feature points, the technique defines a shape interaction matrix Q and groups the points into different moving clusters without motion estimation. Gear[4]presented an alternative method by exploring the reduced row echelon form of measurement matrix.

The drawback of these two techniques is that the performance degrades quickly in the presence of noise. The reason is that the shape interaction matrix loses its discriminative ability when noise is present. An improved approach, provided by Ichimura[5], set threshold for Q, but suffered the same degradation. Wu[14]presented a method to separate points in subgroup level, and the subgroups are obtained using Ichimura's method by setting high threshold. In extreme case, it becomes point-by-point merging. Kanatani[7] proposed to work in the original data space by subspace merging, and improved the results using dimension correction and robust fitting. The subspace merging criterion and the number of objects are determined by model selection. The subspace merging technique does not guarantee the globally optimal segmentation, because it is based on local

BMVC 2003 doi:10.5244/C.17.66

point-by-point interaction. The number of objects is critical to whole process, but it cannot be reliably estimated by model selection.

The most related work is proposed by Inoue[6]. The absolute value of the shape interaction matrix is used as the affinity matrix. The feature points are mapped into a low dimensional subspace. Clusters are then extracted by a graph spectral method. We know that, the success of spectral clustering can guaranteed by a proposition: that the leading eigenvectors of a related Markov transition matrix must be roughly piecewise constant[8]. Practically, in the presence of noise, this piecewise constant eigenvectors condition breaks down. Inoue's method doesn't address this problem so this method degrades when noise is present.

In this paper, we provide a novel approach to robust segmentation of multiple moving objects by spectral clustering. Firstly, we introduce two new affinity matrixes. One is based on the shape interaction matrix and the other one is based on the motion trajectory. Secondly, after mapping the feature points into a low dimensional subspace, we compute the sensitivities of the larger eigenvalues of a related Markov transition matrix when the affinity matrix changes. By selecting appropriate affinity matrix and computing the sensitivity of the eigenvalues with respect to changes in affinity matrix, we improve the piecewise constant eigenvectors condition dramatically. This makes clustering procedure much more reliable and well conditioned. Our approach is robust to noise due to the preservation of the piecewise constant eigenvectors condition. This is verified by extensive experiments.

2 Background and Basic Definitions

2.1 Factorization

Suppose n feature points are tracked in f frames under an affine camera model and there are N independently moving objects in the scene. The coordinate of the ith point in jth frame is (u_i^j, v_i^j). The coordinates of all points may be collected into a $2f \times n$ matrix

$$W = \begin{bmatrix} X \\ \hline Y \end{bmatrix}_{2f\times n} \tag{1}$$

Where $X = \begin{bmatrix} u_1^1 & u_2^1 & \cdots & u_n^1 \\ u_1^2 & u_2^2 & \cdots & u_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ u_1^f & u_2^f & \cdots & u_n^f \end{bmatrix}_{f\times n}$ and $Y = \begin{bmatrix} v_1^1 & v_2^1 & \cdots & v_n^1 \\ v_1^2 & v_2^2 & \cdots & v_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ v_1^f & v_2^f & \cdots & v_n^f \end{bmatrix}_{f\times n}$

According to [3], without noise and outliers, every column of W lies in a 4-dimensional subspace and the rank of measurement matrix W is $4N$, where N is the number of objects. W can be decomposed by SVD.

$$W = U\Sigma V^T$$

If the features from same objects are grouped together, U, Σ and V will have a block diagonal structure.

$$W = \begin{bmatrix} U_1 & \cdots & U_N \end{bmatrix} \cdot \begin{bmatrix} \Sigma_1 & & \\ & \ddots & \\ & & \Sigma_N \end{bmatrix} \cdot \begin{bmatrix} V_1 & & \\ & \ddots & \\ & & V_N \end{bmatrix}$$

This is because every $U_k \Sigma_k V_k$ is the result of a single object factorization[12].

In real situations, we do not know which feature belongs to which object. The feature points from different objects are mixed in the columns of W . To permute and group the columns of W , Costeira and Kanade[3]define a shape interaction matrix

$$Q = VV^T \tag{2}$$

Q is motion invariant and has a property: $Q_{ij} = 0$, if points i, j belong to different objects; $Q_{ij} \neq 0$, if points i, j belong to same objects.

2.2 Problem Definition

Unfortunately, the rank of W is difficult to estimate even with a small noise component[6], which makes computing the interaction matrix Q very difficult without prior knowledge of the number of objects. Even if the interaction matrix Q has been obtained, the elements of Q are nonzero in general. This makes Q lose its zero/nonzero discriminative ability.

In this paper, we want to solve the following problem: suppose points are tracked in many frames under the affine camera model, given the number of objects, how can we compute the interaction matrix Q and use Q to reliably segment feature points into multiple moving objects in the presence of noise?

3 Our Approach

3.1 A Basic Spectral Clustering Method

Suppose an affinity matrix encodes pairwise interaction information of . We propose to use this information to map the original feature points to a low dimensional subspace and group the points in this subspace. Previous work in image segmentation has implemented this idea to do bipartite graph segmentation[10, 13] and extended to multipartite segmentation[8]. This can be done by casting the problem into a spectral graph clustering problem[2].

Given a $n \times n$ pairwise affinity matrix A, where A is symmetric and $A_{ij} = 0$ if points i, j belong to different clusters, then following the formulation in [1], we consider an undirected graph G with vertices $v_i, i = 1, \ldots, n$, and edges $e_{ij} = A_{ij}$ which represent the affinity between vertices v_i and v_j. A Markov chain is defined by setting the transition probability $m_{ij} = d_j^{-1} A_{ij}$ where $d_j = \sum_{i=1}^{n} A_{ij}$ gives the normalizing factor which ensures that $\sum_{i=1}^{n} m_{ij} = 1$. The matrix form of the definition above is:

$$M = AD^{-1}, D = diag(d_1, \cdots, d_n) \tag{3}$$

In practice, we consider the matrix

$$L = D^{-1/2} M D^{1/2} = D^{-1/2} A D^{-1/2} \tag{4}$$

Where L is symmetric and computationally more stable in eigen-decomposition.

Spectral clustering can be done using following simple algorithm[9]:

1. Find the leading K eigenvectors of L, if the number of clusters is known. Form the matrix $X = [V_1, \cdots, V_K]$.

2. Form the matrix Y from X by normalizing each row of X.

3. Treating each row of Y as a point in R^K , use K-means to cluster them into K clusters.

4. Assign the original point w_i(one column in W) to clusters according to the assignment of ith row of Y.

3.2 Improving the Piecewise Constant Eigenvectors Condition

The algorithm above is only valid in ideal case. In the presence of noise, the affinity matrix $A_{ij} \neq 0$ if points i, j belong to different clusters. How can we group the points into correct clusters in the noisy case? It has been shown that, if the points can group into K clusters, then the leading K eigenvectors of M must be roughly piecewise constant[8]. We also found that, if the leading K eigenvectors of M are roughly piecewise constant, the leading K eigenvalues of M all are 1. That is, if we can preserve the piecewise constant eigenvectors condition, the points can be grouped into several clusters without difficulty. So we propose two ways to improve the piecewise constant eigenvectors condition in the presence of noise.

The first is to choose an appropriate affinity matrix. The affinity matrix reflects pairwise interaction information of W. We propose two new affinity matrixes to improve the ability of segmentation. One of new affinity matrixes is built from the shape interaction matrix. In our problem, given the number of objects is N, the shape interaction matrix can be constructed by using first $r = 4N$ column of V as (2). We can construct an affinity matrix based on Q. One simple way is let $A_{ij} = |Q_{ij}|$[6]. But we found that it is vulnerable to noise and easily violates the piecewise constant eigenvectors condition. We propose a new affinity matrix

$$A_{ij} = \exp(-(\frac{1}{|Q_{ij}|}) \cdot \frac{1}{2\delta^2}) \tag{5}$$

Where δ is a scale parameter. The Gaussian function introduces δ into affinity matrix to control the scale of interactions between points. Taking the reciprocal of the absolute value of Q should make the affinity matrix positive and $A_{ij} = 0$, if points i, j belong to different objects.

The other new affinity matrixes is based on the motion trajectory. We can see that each column of W contains the coordinates of a single point across all the frames in the sequence. We define the ith column of W as P_i, $i = 1, \dots, n$. P_i and P_j must belong to the same moving object if they undergo similar movements. So we define the motion trajectory of P_i as displacements between adjacent frames in all the frames. $V_i = ((u_i^1 - u_i^0), (v_i^1 - v_i^0), \cdots, (u_i^n - u_i^{n-1}), (v_i^n - v_i^{n-1}))^T$, $i = 1, \dots, n$. Based on above, we build an affinity matrix A as:

$$A_{ij} = |V_i^T V_i| \tag{6}$$

We can see that A is a symmetric and positive matrix. It encodes pairwise motion interaction information between points.

The second improvement is computing the sensitivities of the larger eigenvalues of L with respect to perturbations in the edge weights. Consider symmetric matrix L, its eigen decomposition is:

$$L = U \Lambda U^T$$

Where $U = [\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_n]$ are eigenvectors. Λ is a diagonal matrix which is composed by eigenvalues $[\lambda_1, \cdots, \lambda_n]$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then the Markov transition matrix $M = D^{1/2} U \Lambda U^T D^{-1/2}$. Consider the Markov chain in graph G, it propagates t iterations. The Markov transition matrix after t iterations is:

$$M^t = D^{1/2} U \Lambda^t U^T D^{-1/2}$$

It can be found that M^t is completely characterized by Λ^t. In other words, the changes of L's eigenvalues reflect the changes of transition probabilities in the edges of graph G. We called this COP(Changes Of Probabilities).

For the vertices i, j belong to different clusters, the COP between them is small. In contrast, the COP within each cluster is large. This is because the connected edges between different clusters are sparse and have small weight values, and the connected edges within clusters are dense and have high weight values.

If the edge weight of a single edge between different clusters changes, the COP in this edge will more sensitive to this change because it has fewer alternative routes to take. In contrast, the COP in the edge within cluster will less sensitive to this change because it has many alternative routes to take. If we can find the edge in which the COP is more sensitive to the change of edge weight, then cut the edge. This is because, in the ideal case of well-separated clusters, the weight of this edge must be zero. In the presence of noise, the well-separated clusters become weakly coupled. If the linked edge generated by noise can be identified, we can cut the edge and recover the original well-separated clusters.

Because the changes of L's eigenvalues reflect the COP in the edges of graph G, we compute the sensitivity of eigenvalues of L with respect to the edge weight A_{ij}, which represents the sensitivity of the COP with respect to edge weight[1].

$$S_{ij} = \frac{d\lambda}{dA_{ij}} = \vec{u}^T \frac{dL}{dA_{ij}} \vec{u} = 2 \frac{u_i u_j}{\sqrt{d_i d_j}} - \lambda \left(\frac{u_i^2}{d_i} + \frac{u_j^2}{d_j} \right) \tag{7}$$

Here (u_i, u_j) are the (i, j) elements of eigenvector $\vec{u}$. (d_i, d_j) are degrees of nodes (i, j). The proof is omitted here due to space limit or one can refer to a similar one[1]. In practice, we need consider only larger eigenvalues of L(smaller eigenvalues have few impact on clustering), we set a threshold ε to select them except 1(The eigenvalue 1 correspond to well separated clusters and does not need to be considered). We take $\varepsilon = 0.9$. If $|S_{ij}| > \sigma \cdot median(S)$, then cut the edge between i, j. σ takes a high value in order to cut only edges with the highest sensitivities.

The final algorithms is summarised as:

1. Build a measurement matrix W and an affinity matrix A.
2. Compute the graph related matrices D, L and the eigen-decomposition of L.
3. Cut the matrix A on many eigenmodes simultaneously based on S_{ij}.
4. Re-compute D, L and the eigen-decomposition of L, then invoke the basic spectral clustering method above.

The difference between our spectral clustering algorithm with eigencuts algorithm in [1] is that, our algorithm is intended to segment the points using leading eigenvectors by improving the piecewise constant eigenvectors condition in the presence of noise. The

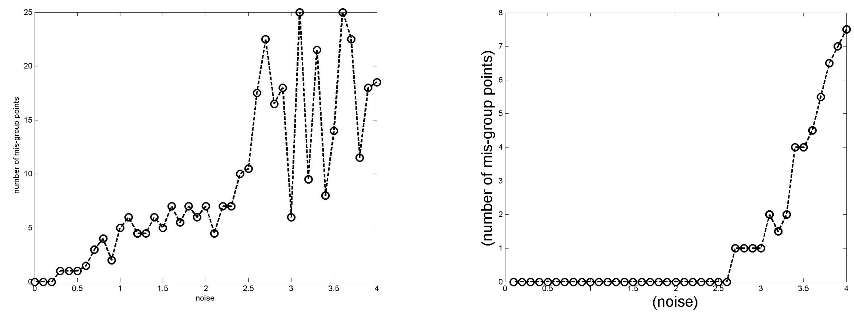

Figure 1: Shown the number of mis-grouping points under noise.

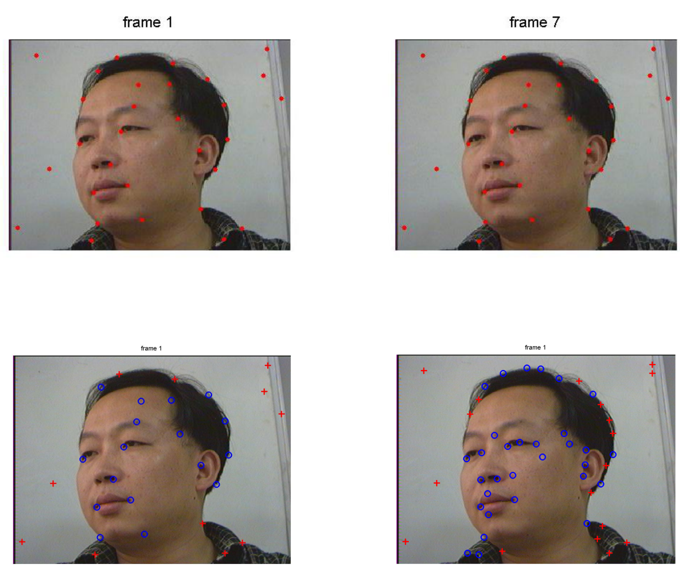

Figure 2: Segment results on a head sequence.Top row: point tracking result. Bottom left: shown the segment results with shape interaction affinity matrix in the first frame.Bottom right: shown the segment results with motion trajectory affinity matrix in the first frame.

eigencuts algorithm is proposed to solve the weakly coupled data clustering problem. It chooses an eigenmode to cut edge in an iterative way. We found that it often makes the matrix A singular. This is because matrix A has too many edges be cut. In contrast, our algorithm cut matrix A operates on many eigenmodes simultaneously. It is computationally more efficient and stable.

4 Experiments

In this section we provide experimental results with both synthetic and real data. We performed some simulations to analyse our algorithm. We build a synthetic scene that consists of two sets of points. One set of 30 points placed in a 3D cube, and the other set of 15 points represents background. These two sets of points undergo different and independent motions. We generate 20 frames and, to test the robustness of our algorithm, we also add Gaussian noise to the image points. Figure 1 shows the segmentation result when the standard deviation of the noise=1.0. The noise ranges from 0 to 4 with interval of 0.1. We perform 30 runs for each noise level and compute the mean of the mis-grouping error. The results in the left is from the algorithms using the shape interation based affinity matrix with Gaussian width $\delta = 2$. The results in the right is from the algorithms using the motion trajectory based affinity matrix. It gives superior results up to the point when noise is as large as the motion.

We have also applied our algorithm to some real video sequences. The first sequence contains a moving head in front of camera. We observe that the head undergoes rotation out of plane and introduces a large perspective effect. We detect and track 30 feature points in 14 frames using a KLT tracker[11] and apply our motion segmentation algorithm. Figure 2 shows our segmentation results. In the bottom left of Figure 2, the two points(cross) on the border between the hair and the background are grouped as background. If combined with other cues such as colour, our algorithm will segment these two points correctly. In the bottom right of Figure 2, the result is not so good. This is because the affinity matrix based on the motion trajectory does not encode the 3d information contained in the matrix W and so is sensitive to perspective ”noise”.

We also compare three approaches: one is our approach based on the shape interaction affinity matrix. One is the approach presented in[9]. The last one is the approach presented in[6] which based on the affinity matrix $|Q_{ij}|$. Figure 3 shows the result. We can see from it that our approach performs the best.

Another sequence contains a moving hand and background. The result of segmentation is shown in Figure 4. The performance of our algorithm is excellent. But some points in the background are clustered into the same group as the moving hand. The reason is that, the KLT tracker makes the static points in background move with hand when hand passes by. The points share the same motion with the hand in many frames, which forces them to be clustered into same group. Segmentation of such points against the hand is very difficult.

5 Conclusions

The factorization approach to motion segmentation is based on the shape interaction matrix but noise makes segmentation difficult. In this paper, we proposed a spectral clustering approach to segment multiple moving objects robustly. By introducing two new affinity matrixes and computing the sensitivities of the larger eigenvalues of L with respect to perturbations in the edge weights, we improve the piecewise constant eigenvectors condition dramatically. The feature points are mapped into a low dimensional subspace and are clustered using a spectral clustering method. The robustness of our approach is verified using synthetic and real data.

In the future, we would like to extend the work to deal with unknown number of objects case and investigate the rules behind affinity matrix design. And we also want to apply the spectral clustering method developed in this paper to other problems such as stereo correspondence.

Acknowledgements

The authors wish to thank Richard Harvey for his help. This work is part funded by EC project Q5CR-2002-71699.

References

[1] Chakra Chennubhotla and Allan Jepson. Half-lives of eigenflows for spectral clustering. In *Advances in Neural Information Processing Systems 15*. MIT Press, 2003.

[2] Fan R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

[3] J. Costeira and T. Kanade. A multibody factorization method for independently moving objects. *IJCV*, 29(3):159–179, 1998.

[4] C. W. Gear. Multibody grouping from motion images. *IJCV*, 29(2):133–150, 1998.

[5] N. Ichimura. A robust and efficient motion segmentation based on orthogonal projection matrix of shape space. In *IEEE CVPR00*, pages 446–452, June 2000.

[6] K. Inoue and K. Urahama. Separation of multiple objects in motion images by clustering. In *Proc. 8th Int. Conf. Computer Vision*, volume 1, pages 219–224, Vancouver, Canada,, July 2001.

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

[8] M. Meila and J. Shi. A random walks view of spectral segmentation. In *International Workshop on AI and Statistics*, 2001.

[9] Ng, M. Jordan, and Y. Weiss. On spectral clustering: analysis and an algorithm. In *Advances in Neural Information Processing Systems 14*. MIT Press, 2002.

[10] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

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

[12] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *IJCV*, 9(2):137–154, 1992.

[13] Y. Weiss. Segmentation using eigenvectors: a unifying view. In *Proc. 7th ICCV*, pages 975–982, 1999.

[14] Ying Wu, Zhengyou Zhang, Thomas S. Huang, and John Y. Lin. Multibody grouping via orthogonal subspace decomposition. In *IEEE CVPR01*, volume 2, pages 252–257, 2001.

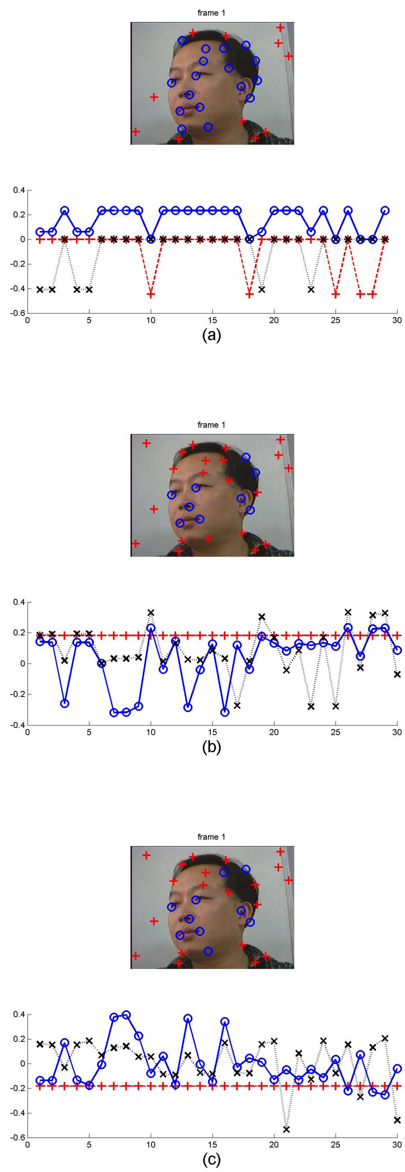

Figure 3: Segmentation results and the leading three eigenvectors (u_1, u_2, u_3) of L. u_1- red plus; u_2- blue circle; u_3- black cross. (a) our segmentation result. The correspondent eigenvalues is (1, 1, 1). (b) Segmentation result based on algorithm in[9] without computing the larger eigenvalues of with respect to perturbations in the edge weights. The correspondent eigenvalues is (1, 0.98419 , 0.97602). (c) Segmentation result based on affinity matrix $|Q_{ij}|$[6]. The correspondent eigenvalues is (1, 0.48654, 0.43743).

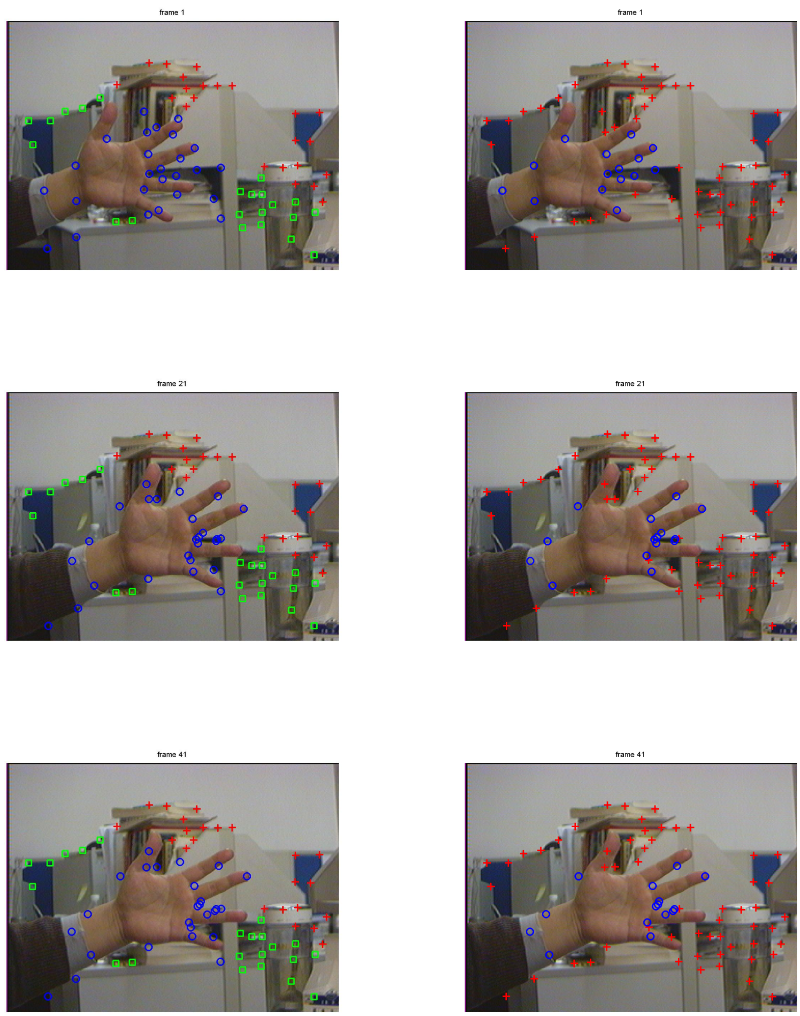

Figure 4: Segmentation result on a hand sequence. (left) Results from our algorithm using shape interaction based affinity matrix. (right)Results from our algorithm using motion trajectory based affinity matrix.