Motion Segmentation by Velocity Clustering with Estimation of Subspace Dimension

Liangjing Ding*, Anke Meyer-Baese* and Adrian Barbu+

Department of Scientific Computing, Florida State University* Department of Statistics, Florida State University+

Introduction

- The task of motion segmentation is to label a set of tracked feature points from several moving objects into different groups based on their motions.
- It is an important step in many computer vision problems, such as robotics, inspection, video surveillance, etc.

The Affine Camera Model

Let $t = (x^1, y^1, x^2, y^2, \dots, x^F, y^F)^T$ be a trajectory of a tracked feature point in F frames. Given P trajectories undergoing the same rigid motion, under the affine camera model, the *measurement matrix* W could be decomposed into a *motion matrix* $M \in \mathbb{R}^{2F\times 4}$ and a *structure matrix* $S \in \mathbb{R}^{P\times 4}$ as

$$W = MS$$

$$\begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_P^1 \\ y_1^1 & y_2^1 & \cdots & y_P^1 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^F & x_2^F & \cdots & x_P^F \\ y_1^F & y_2^F & \cdots & y_P^F \end{bmatrix} = \begin{bmatrix} A^1 \\ \vdots \\ A^F \end{bmatrix} \begin{bmatrix} X_1 & \cdots & X_P \\ Y_1 & \cdots & Y_P \\ Z_1 & \cdots & Z_P \\ 1 & \cdots & 1 \end{bmatrix}.$$

where A^f is the affine motion matrix at frame f.

- $2 \le \text{rank}(W) \le 4$.
- If define $\hat{t} = \sum_{i=1}^P w_i t_i$, where $\sum_{i=1}^P w_i = 1$, the registered measurement matrix $\tilde{W} = [t_1 - \hat{t}, \tilde{t}_2 - \hat{t}, \dots, \tilde{t}_P - \hat{t}]$ is at most rank 3.

Noise Reduction

In order to reduce the effect of the accumulated error in the motion segmentation, we use the velocity vector

$$[x^1 - x^2, y^1 - y^2, \dots x^{F-1} - x^F, y^{F-1} - y^F, x^i, y^i]^T, 1 \le i \le F \qquad (1)$$

to characterize the trajectories.

- The velocities in each frame contain only the tracking error from the previous frame to the current frame.
- The rank of measurement matrix is still preseved since only simple row operations are involved.

A synthetic experiment: 242 synthetic trajectories of length 20 were generated for two different motions. Gaussian tracking errors with different signal-to-noise ratio (SNR) were introduced between consecutive frames.The trajectories were projected to a 3D subspace, and a plane was fitted for each motion. The SSE and variance of the distances from projected points to the fitted planes are shown in the following table.

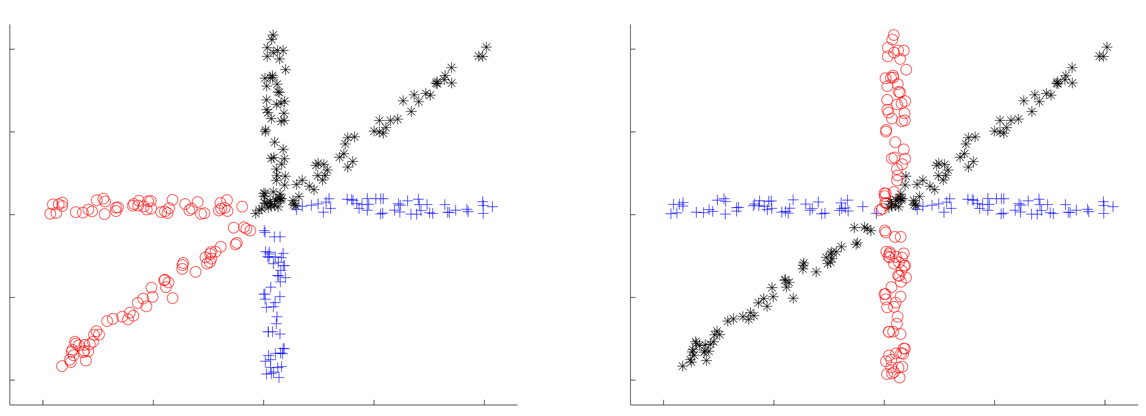
	SSE	Variance
Distance Vector (No noise added)	0	0
Velocity Vector (No noise added)	0	0
Distance Vector (SNR = 10)	0.256e-5	0.0011e-5
Velocity Vector (SNR = 10)	0.106e-5	0.0004e-5
Distance Vector (SNR = 5)	1.058e-5	0.005e-5
Velocity Vector (SNR = 5)	0.208e-5	0.001e-5

Affinity Measure

Traditional distance-based affinity could not be used in subspace clustering. Here, we use the affinity measure based on the angular information, defined by

$$A_{ij} = \left(\frac{x_i^T x_j}{\|x_i\|_2 \|x_j\|_2}\right)^{2\alpha}, i \ne j, \alpha \in \mathbb{N} \qquad (2)$$

$\alpha > 1$ is a tuning parameter to increase the separation.



Spectral clustering of lines in 2D: the distance-based affinity mixes points from different subspaces (left), while the angle-based affinity separates them very well (right).

Dimension Selection

The main difficulty for selecting the best ambient space dimension is that the dimension of one affine subspace is not fixed. Here we employ an exhaustive strategy to search the best dimension in range $[2k, 4k]$, and a measure is proposed to select the best result.

Any registered trajectory $\tilde{t}$ in $\tilde{W}$ will have a corresponding point $\tilde{P} \in \mathbb{R}^3$ obtained by least squares $\hat{P} = \text{argmin}_{\tilde{P}} \|\tilde{t} - \tilde{M}\tilde{P}\|^2$.

We define the RMSE error of the trajectory t as $\sqrt{\min_{\tilde{P}} \|\tilde{t} - \tilde{M}\tilde{P}\|^2}$.

We could expect that the best segmentation would have the smallest sum of RMSE errors

$$E(L) = \sum_{l=1}^{k} \sum_{i=1}^{P} \sqrt{\min_{\tilde{P}} \|\tilde{t}_i - \tilde{M}^l \tilde{P}\|^2}. \qquad (3)$$

Algorithm

Input: The measurement matrix $W = [t_1, t_2, \dots, t_P] \in \mathbb{R}^{2F\times P}$ whose columns are point trajectories, and the number of clusters k.

Preprocessing: Build the velocity measurement matrix W' by row transformations of W given by eq. (1).

for $D = d_{min}$ **to** d_{max} **do**

1. Perform SVD: $W' = U\Sigma V^T$
2. Build the N-by-D data matrix $[X_D = [v_1, ..., v_D]]$ where v_i is the i-th column of V.
3. Apply spectral clustering to the N points in X_D using the affinity measure (2).
4. Compute the clustering error E_D of the segmentation result using eq. (3).

end for

Output: The segmentation result with the smallest error E_D.

Results and Summary

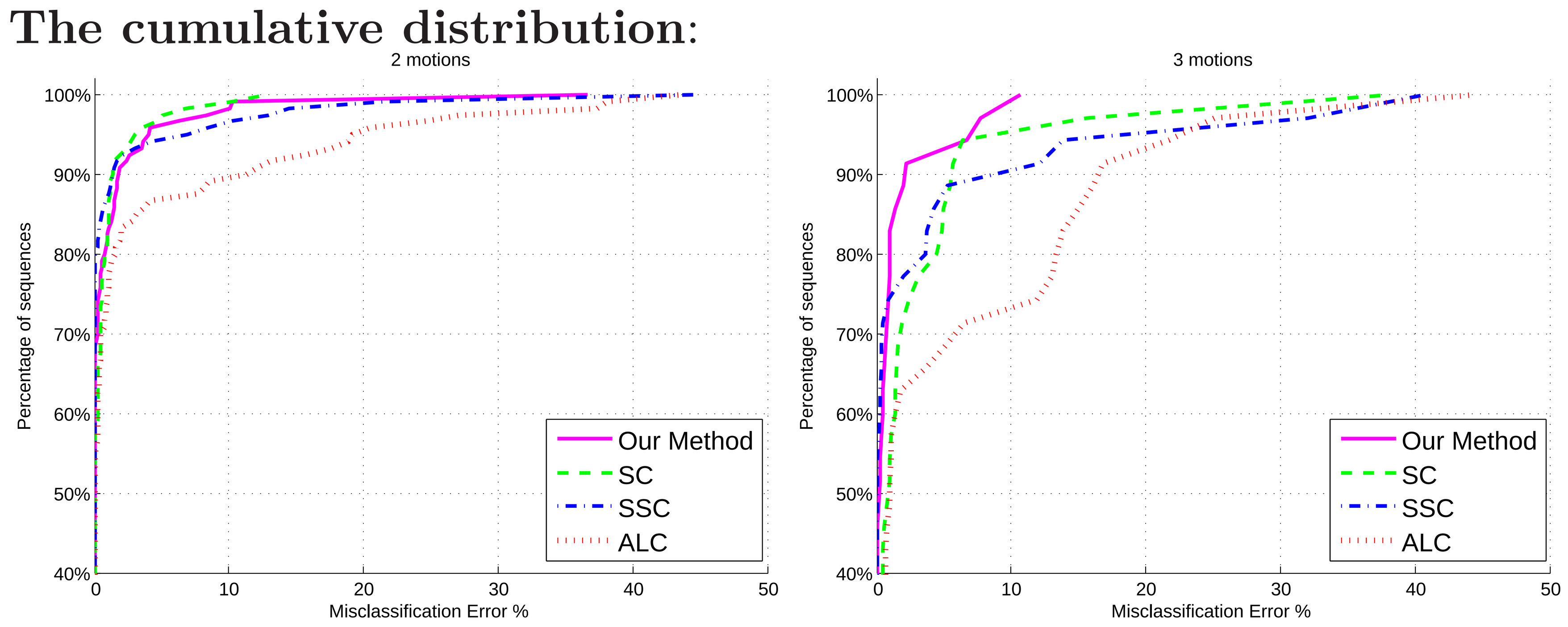
The parameter settings are $\alpha = 2$, $d_{min} = 2k$, $d_{max} = 4k$, and the locations of the last frame are kept to build the velocity matrix.

We tested our algorithm and several other state-of-the-art algorithms on the sequences in the Hopkins 155 database.

The misclassification rate (in percent):

	Method	ALC	SC	SSC	Our Method
Checkerboard (2 motion)	Average	1.55	0.85	1.12	**0.67**
	Median	0.29	0.00	0.00	0.00
Traffic (2 motion)	Average	1.59	0.90	**0.02**	0.99
	Median	1.17	0.00	0.00	0.22
Articulated (2 motion)	Average	10.70	1.71	**0.62**	2.94
	Median	0.95	0.00	0.00	0.88
All (2 motion)	Average	2.40	0.94	**0.82**	0.96
	Median	0.43	0.00	0.00	0.00
Checkerboard (3 motion)	Average	5.20	2.15	2.97	**0.74**
	Median	0.67	0.47	0.27	0.21
Traffic (3 motion)	Average	7.75	1.35	**0.58**	1.13
	Median	0.49	0.19	0.00	0.21
Articulated (3 motion)	Average	21.08	4.26	**1.42**	5.65
	Median	21.08	4.26	0.00	5.65
All (3 motion)	Average	6.69	2.11	2.45	**1.10**
	Median	0.67	0.37	0.20	0.22
All sequences combined	Average	3.37	1.20	1.24	**0.99**
	Median	0.49	0.00	0.00	0.00

The cumulative distribution:



In summary, we present a new motion segmentation method. One contribution we made is employing the velocity vectors to reduce the accumulated error, the other contribution is to propose an novel clustering error measure to select the best ambient subspace. The evaluation shows our method is competitive with current state-of-the-art methods.