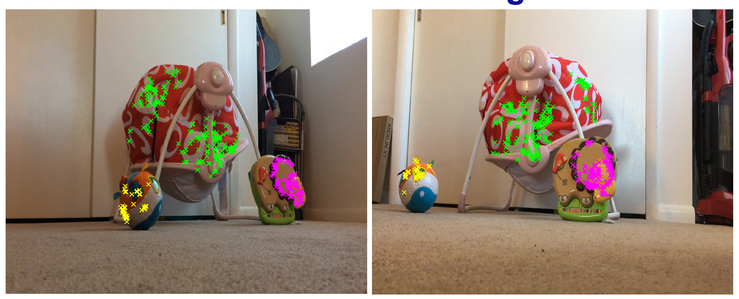
Dynamics Enhanced Multi-Camera Motion Segmentation from Unsynchronized Videos

Xikang Zhang, Bengisu Ozbay, Mario Sznaier, Octavia Camps
Robust Systems Lab, Northeastern University, Boston

ICCV17 International Conference on Computer Vision 2017

Robust Systems Lab

Problem: Multi-camera motion segmentation



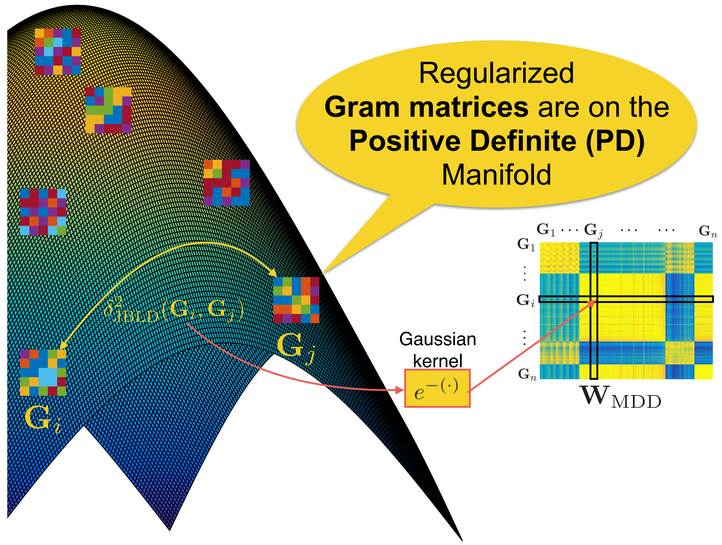
➢ Uncalibrated cameras
➢ Without correspondences across cameras
➢ No synchronization is required

Background

➢ The dynamic information is captured in its Gram matrix

Manifold Dynamics Distance (MDD)

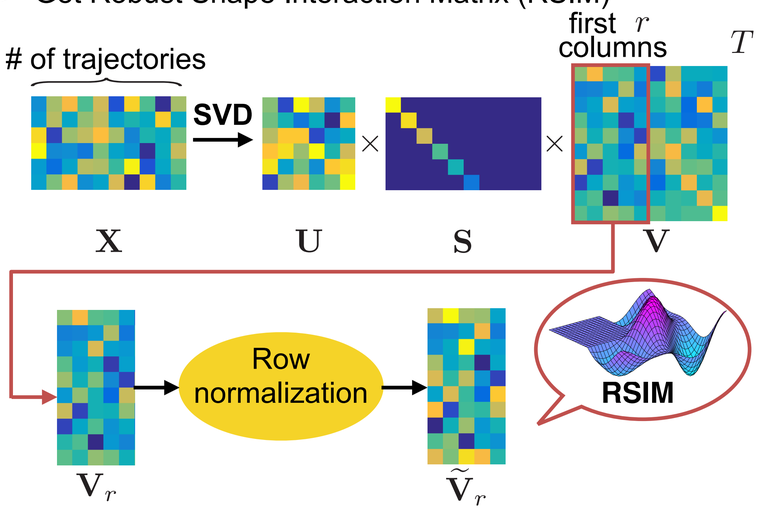
➢ MDD measures distance/similarity between Gram matrices (dynamics).
➢ JBLD is one of the most efficient MDDs.



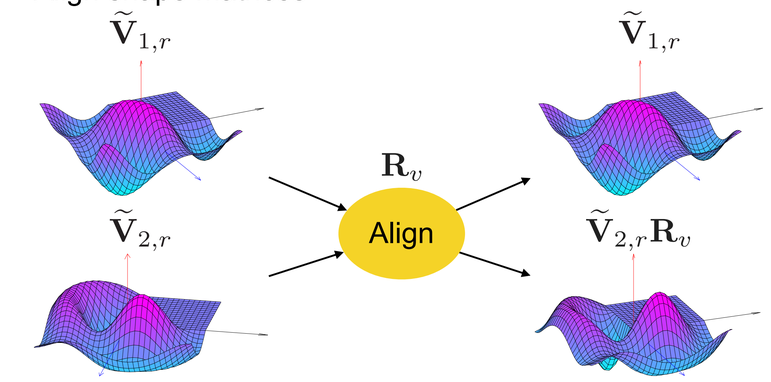
$$\delta^2_{\text{JBLD}}(\mathbf{X}, \mathbf{Y}) = \log\left|\frac{\mathbf{X}+\mathbf{Y}}{2}\right| - \frac{1}{2}\log|\mathbf{XY}|$$

Multi-camera RSIM (McRSIM)

➢ Get Robust Shape Interaction Matrix (RSIM)



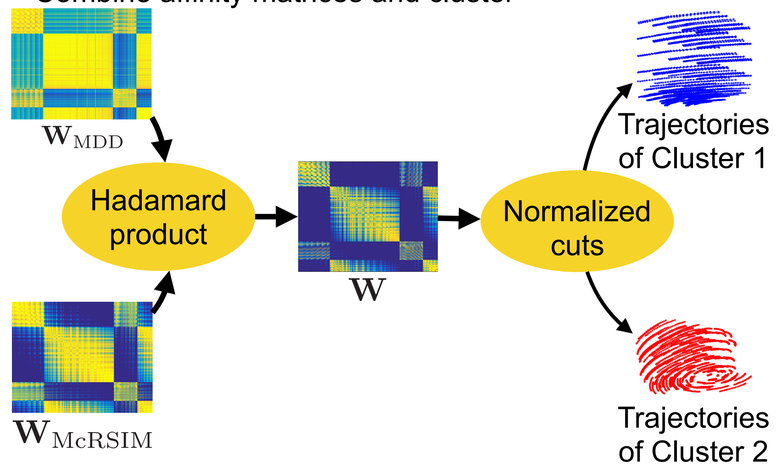
➢ Align shape matrices



➢ Compute McRSIM affinity matrix

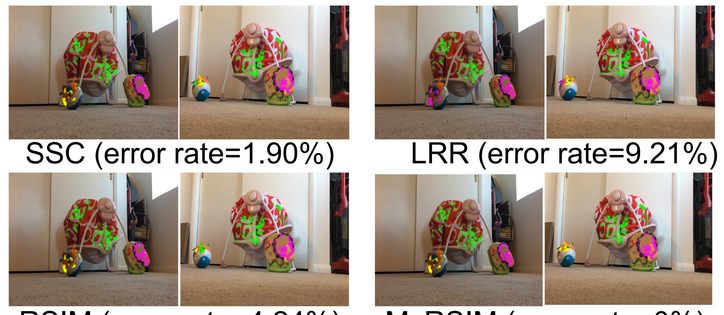
$$\mathbf{W}_{\text{McRSIM}} = \left(\begin{bmatrix}\widetilde{\mathbf{V}}_{1,r} \\ \widetilde{\mathbf{V}}_{2,r}\mathbf{R}_v\end{bmatrix}\begin{bmatrix}\widetilde{\mathbf{V}}_{1,r}^T & \mathbf{R}_v^T\widetilde{\mathbf{V}}_{2,r}^T\end{bmatrix}\right)^{\gamma}$$

➢ Combine affinity matrices and cluster



Experimental results

➢ New benchmark data sets RSL12 and RSL60
- Lengths of sequences are 0.5s ~ 4s
- # of trajectories in each sequence is 25 ~ 406



➢ Multi-camera motion segmentation

Table 1: Clustering error (in %) on RSL 12 data set. Parameters are $r_{\min} = 1$, $r_{\max} = 2$ for all RSIM and McRSIM related methods.

Methods	SSC	LRR-H2	RSIM	RSIM-MDD-LA	McRSIM	McRSIM-MDD
2 motions clustering error (in %)						
Mean	2.78	9.54	8.72	0.00	0.03	0.00
Median	0.00	2.91	0.20	0.00	0.00	0.00
3 motions clustering error (in %)						
Mean	8.36	29.55	24.08	0.00	0.00	0.00
Median	2.70	33.29	32.41	0.00	0.00	0.00
Overall clustering error (in %)						
Mean	5.57	19.54	16.40	0.00	0.02	0.00
Median	0.95	25.70	7.72	0.00	0.00	0.00
Running time (in seconds)						
Total	24.46	19.22	3.25	20.50	3.03	34.56
Avg	2.04	1.60	0.27	1.71	0.25	2.88

➢ With missing entries and gross contamination

Table 2: Clustering error (in %) on Hopkins 12 Real Motion Sequences With Incomplete Data. Half of the trajectories are rotated 45° and serve as Camera 2.

Methods	RSIM-M	RSIM-M-MDD	McRSIM-M	McRSIM-M-MDD
Mean	35.78	35.58	7.45	7.15
Median	38.06	37.96	1.89	1.69
Max	47.40	47.63	31.83	31.83
Std	11.63	11.45	11.45	11.58

Table 3: Clustering error (in %) on Hopkins 155 sequence with grossly contaminated entries (%5 corrupted). Half the trajectories are rotated 45° and serve as Camera 2.

Methods	RSIM	RSIM-MDD	McRSIM	McRSIM-MDD
2 motions				
Mean	31.76	31.39	21.79	21.64
Median	34.66	34.05	22.52	22.19
3 motions				
Mean	45.05	44.77	36.13	36.02
Median	47.42	47.32	35.66	35.41
Overall				
Mean	34.76	34.41	25.03	24.89
Median	38.67	38.19	25.95	25.97

code & data

Conclusion

➢ **Dynamics-based information** enhances performance of state-of-the-art motion segmentation methods.
➢ **McRSIM** is shown to be robust to rotation, translation and time delay, and performs well in multi-camera motion segmentation. It is efficient since the computational complexity is similar to that of an SVD.

This work was supported in part by NSF grants IIS–1318145, ECCS– 1404163, CMMI–1638234; AFOSR grant FA9550-15-1-0392; and the Alert DHS COE under Award Number 2013-ST-061-ED0001.