Multi-view Synchronization of Human Actions and Dynamic Scenes

Emilie Dexter, Patrick Pérez, Ivan Laptev
INRIA Rennes - Bretagne Atlantique

emilie.dexter@inria.fr

UMR IRISA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE
INRIA
centre de recherche RENNES - BRETAGNE ATLANTIQUE

Introduction

Purpose

- Synchronization of image sequences of the same dynamic event or similar dynamic events
- Generally hard problem
  - Unknown camera position or motion
  - Unknown temporal offset or time warping
- Applications
  - 3D dynamic reconstruction
  - Analysis of multi-view dynamic scenes

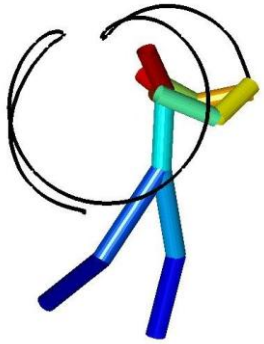
Outline

- Temporal descriptor for image sequences
- Proposed method
- Experimental Results

Temporal description of image sequences

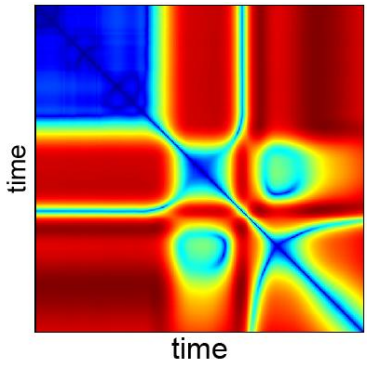
Inputs:

- Sequence of T images denoted $I=\{I_1, \ldots, I_T\}$
- Features extracted from each image: real point trajectories
- Dominant motion estimation and compensation



Self-similarity matrix definition (SSM):

$$d_{ij} = \sum_{k=1}^{N_{ij}} \|\underline{x}_i^k - \underline{x}_j^k\|_2$$



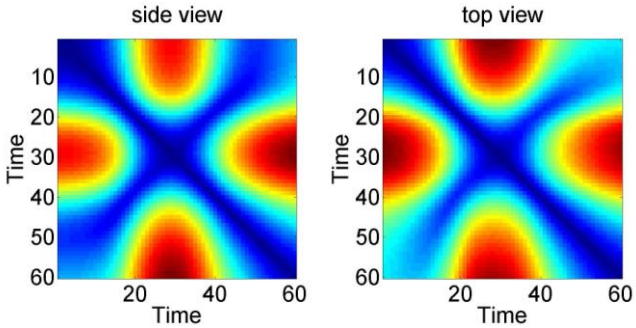
SSM for a golf swing action

where i and j are the instant indexes, k is the trajectory index and N_{ij} the number of trajectories that span the time intervals between frames I_i and I_j.

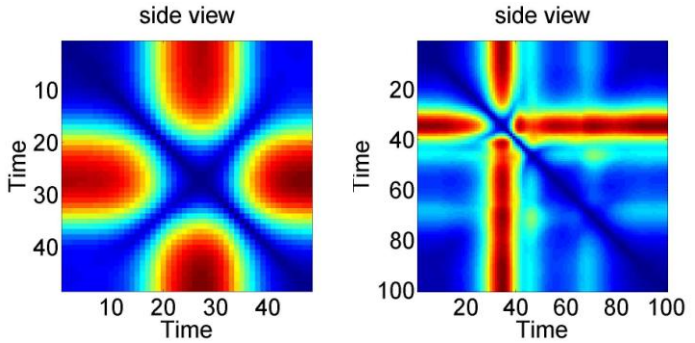
Temporal description of image sequences

SSM Characteristics:

- Stable across views
- Exhibit dynamics of the scenes



SSMs of bend action seen from side and top views



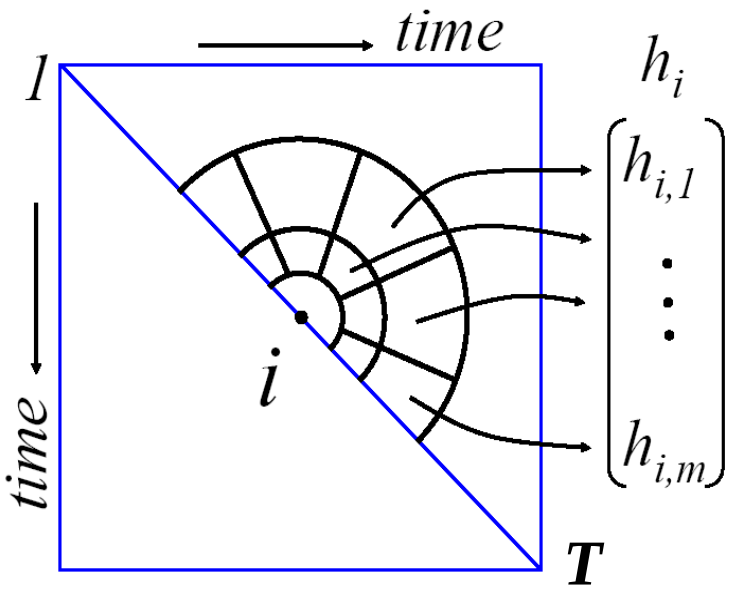
SSMs of bend and jump actions both seen from side view

Temporal description of image sequences

SSM description:

- Matrix as an image:
  - HOG-based description along diagonal
  - Point scale automatic selection
- Sequence of local descriptors

$H = (h_1, \ldots, h_T)$



Synchronization Method

Common framework for:

- A same scene seen from different views
- Similar dynamic events as actions

Proposed Framework

Do for each sequence
    Extract input data : trajectories
    Build SSM
    Compute descriptor H^i
end
Apply DTW between H^1 et H^2

Fixed vs. Adaptive Size of Local Descriptor

- Evaluation based on synchronization error estimation: the average distance of points on the estimated warp to the ground truth

- Descriptors :
  - [6] = fixed and identical size of the support for both sequences **[Junejo et al. ECCV08]**
  - [6]* = fixed size of the support but different for each sequence
  - Proposed = automatic adapted size

Fixed vs. Adaptive Size of Local Descriptor

Table 1: Mean synchronization error for MoCap data for different frame rate ratios, R_{FR}

Sequence name	$R_{FR}=2$			$R_{FR}=3$			$R_{FR}=4$		
	[6]	[6]*	proposed	[6]	[6]*	proposed	[6]	[6]*	proposed
golf seq1	1.43	1.24	1.08	1.92	5.30	0.70	3.02	2.15	0.77
golf seq2	0.98	1.46	0.78	2.10	3.70	0.74	2.83	3.16	0.69
golf seq3	0.98	0.97	0.85	1.51	3.74	0.90	3.97	2.85	0.63

Table 2: Mean synchronization error for natural image sequences

Sequence name	$R_{FR}=2$			$R_{FR}=3$			$R_{FR}=4$		
	[6]	[6]*	proposed	[6]	[6]*	proposed	[6]	[6]*	proposed
seq1 (81)	3.72	4.67	4.55	3.73	8.75	3.66	8.6	6.41	2.17
seq2 (−9)	3.53	3.91	1.83	7.57	8.36	4.92	5.5	5.19	1.4
seq3 (−27)	2.93	15.42	2.48	5.77	23.13	4.77	16.56	17.79	2.14

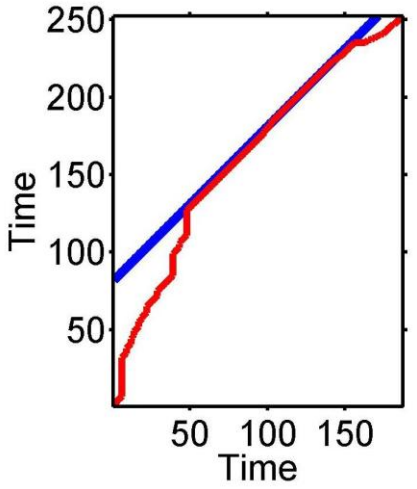
Natural Sequences – static cameras

View1

Unsynchronized View2

View1

Synchronized View2



Estimation (red) and Gr. Truth (blue)

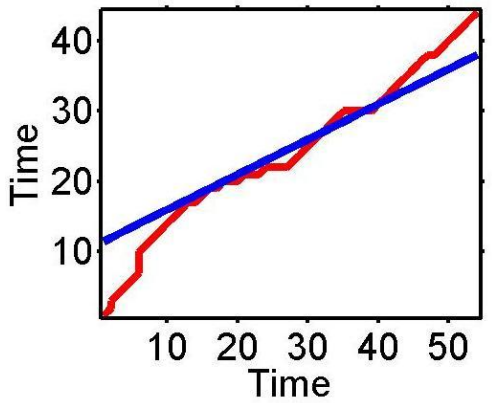
Unconstrained Natural Sequences

View1

Unsynchronized View2

View1

Synchonized View2



Estimation (red) and Gr. Truth (blue)

Drinking action from *Coffee and Cigarettes*





Smoking action from *Coffee and Cigarettes*





Conclusion

- Automatic method for synchronizing image sequences
  - Generic solution (few constraints)
  - Descriptors stable under view-changes
  - DTW can handle non-linear time warping

- Comparison with others state-of-art methods
- Address higher level problem: matching of sequences

Thank you

Previous Works and Characteristics

- Caspi and Irani [PAMI 2002]
  - Linear time warping
  - Correspondences between views
- Rao et al. [ICIP 2003]
  - Correspondences between trajectories across views
- Tuytelaars and Van Gool [CVPR 2004]
  - Tracked point manually chosen
- Wolf and Zomet [IJCV 2006]
  - Time-shift assumption

Adaptive size descriptor

For each diagonal point :

- Scale detection : maximizing the normalized Laplacian (σi)
- Size choice : $2\sigma i$