Temporal Action Co-Segmentation in 3D Motion Capture Data and Videos

Konstantinos Papoutsakis[1,3], Costas Panagiotakis[2,3], Antonis A. Argyros[1,3]

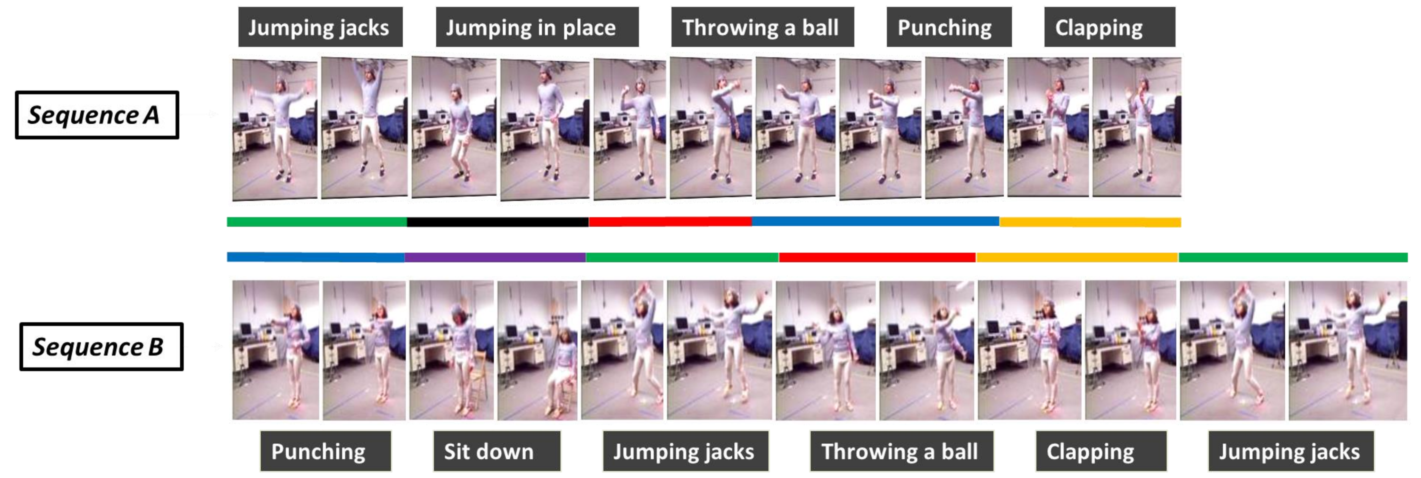
[1]Computer Science Department, University of Crete, Greece [2]Business Administration Department, TEI of Crete, Greece [3]Institute of Computer Science, FORTH, Greece

IEEE 2017 Conference on Computer Vision and Pattern Recognition

CVPR July 21-26 HONOLULU 2017

PROBLEM

- Given two action sequences (of motion capture or video data), we are interested in spotting & co-segmenting all pairs of sub-sequences (commonalities [1]) that represent the same action.



✓No a-priori model or labels of the actions are available.
✓The number of common subsequences may be unknown.
✓The sub-sequences can be located anywhere in the long sequences, may differ in duration and the corresponding actions may be performed by a different person, in different style.

MOTIVATION

The discovery of common action patterns in two or more sequences provides an efficient bottom-up way to:

- ✓ **segment** action sequences,
- ✓ **identify a set of elementary actions,**
- ✓ **build models** of the performed activities in an **unsupervised manner**.

MAIN IDEA

We propose a totally unsupervised solution to the problem of temporal action co-segmentation using **stochastic optimization by employing Particle Swarm Optimization (PSO)**.
The objective function that is minimized by PSO capitalizes on **Dynamic Time Warping (DTW)** to compare two action sequences.

REFERENCES

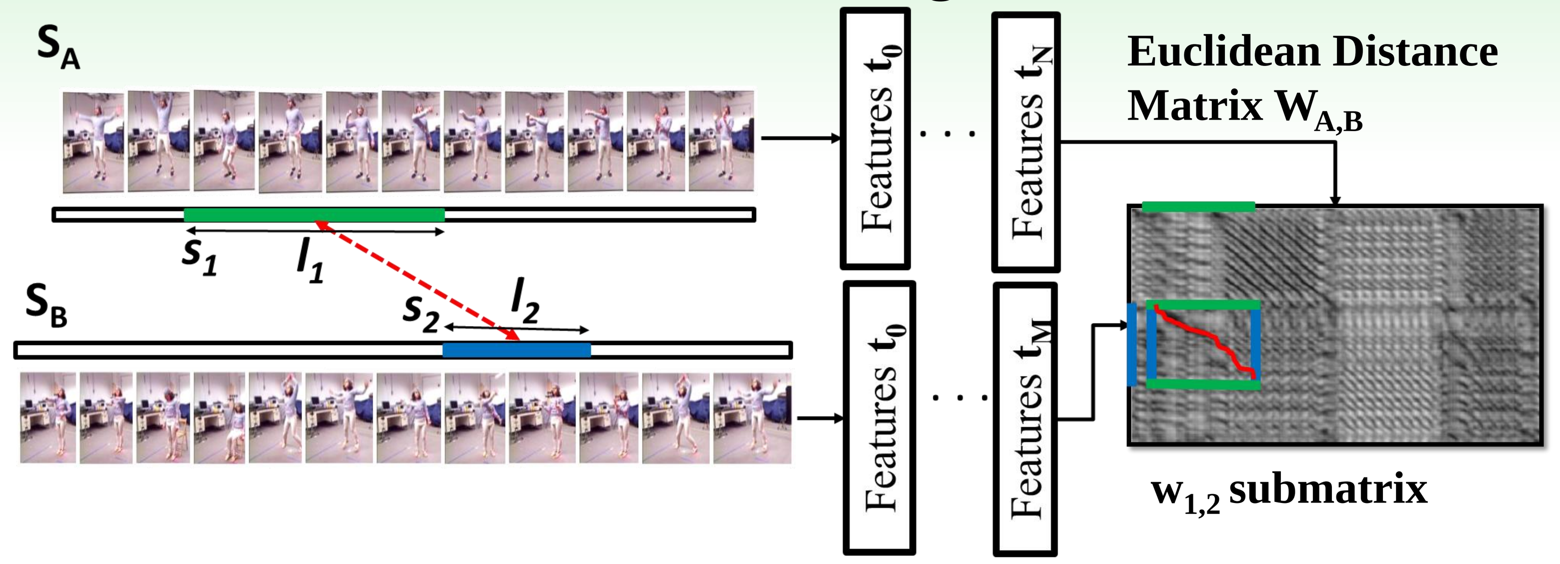
[1] W.S. Chu, et.al. Unsupervised temporal commonality discovery, ECCV 2012.
[2] A. S. Park, et.al. Unsupervised pattern discovery in speech. IEEE/ACM TASLP 2008.
[3] J. Guo, et.al. Video cosegmentation for meaningful action extraction. IEEE ICCV 2013.
[4] H. Wang, et.al. Action recognition with improved trajectories. IEEE ICCV 2013.
[5] J,Kennedy, et.al.,"Particle Swarm Optimization". IEEE ICNN 1995.
[6] R.Ofli et.al., Berkeley MHAD: A Comprehensive Multimodal Human Action Database. IEEE WACV 2013.

PROPOSED FRAMEWORK

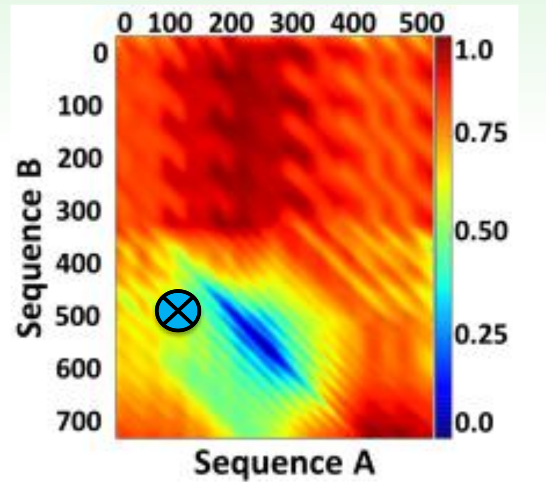
Evaluating a candidate commonality using DTW

We treat the action sequences as multivariate time-series of either:

- ✓ **3D motion capture data:** angle-based feature representation of skeletal joints.
- ✓ **or 2D video data:** motion-based features (dense trajectories [4]) in RGB videos.



- ✓ Quadruple $p = (s_1, l_1, s_2, l_2)$ defines a 2D rectangle $w_{1,2}$ in $W_{A,B}$ that is a **possible commonality p.**
- ✓ Compute the alignment DTW score D(p) in the $w_{1,2}$ submatrix.
- ✓ Define **a particle p** ⊗ in the 4D PSO search space.



Spotting a single commonality using PSO

- ✓ Define an objective function based on the D(p) score and the #diagonal steps $n_p(p)$ in the optimal path p of DTW.
- ✓ Optimize (minimize) over all possible commonalities in the 4D search space using the PSO for P particles over G generations.

$$O(p) = \frac{D(p)+c}{np(p)+1} \qquad \Omega(p_i, p_j) = \frac{|R(pi) \cap R(pj)|}{|R(pi)|}$$

Spotting multiple commonalities

- ✓ Define a new objective function using the normalized overlap among retrieved commonalities.
- ✓ Iteratively run PSO to retrieve N commonalities.

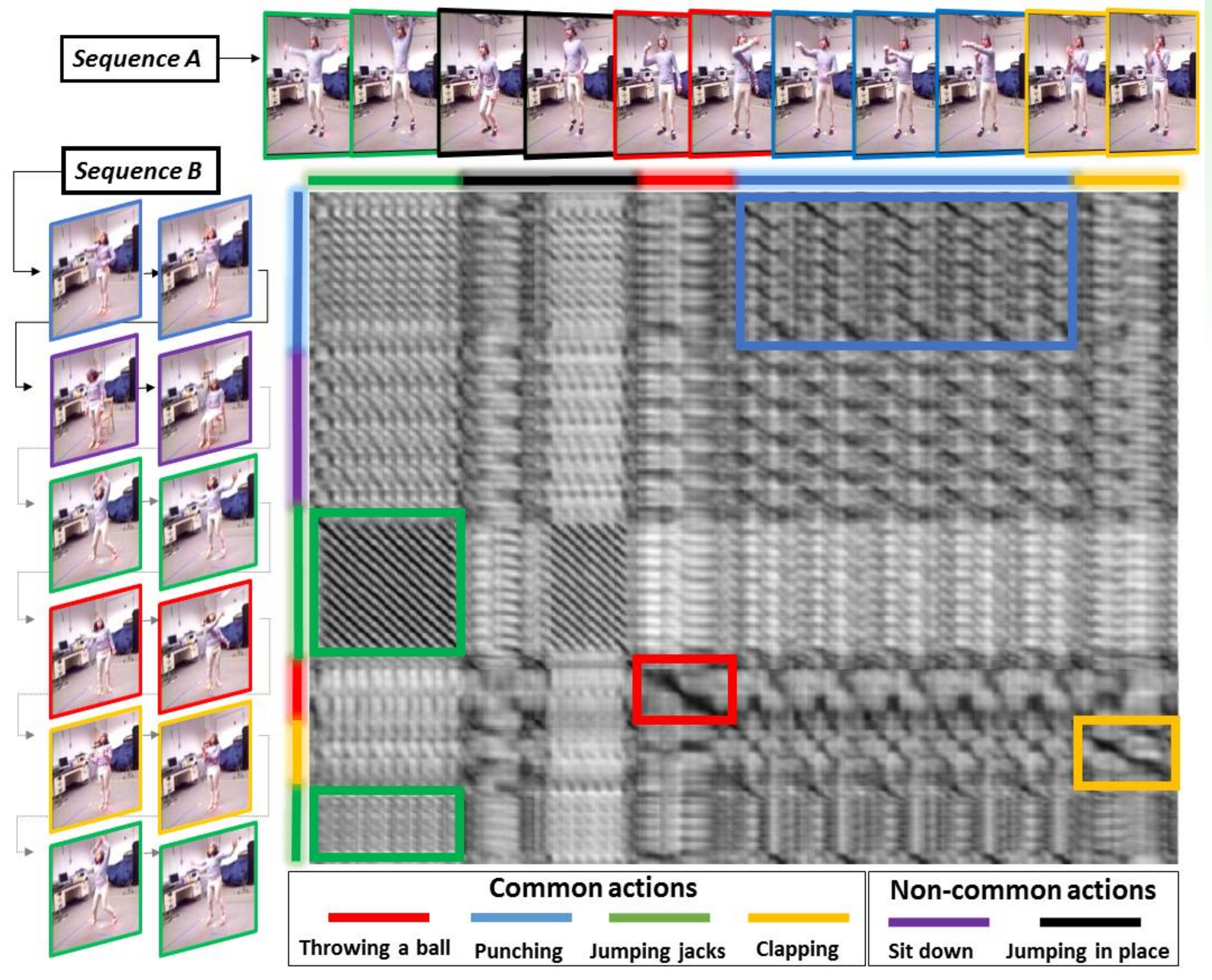
$$O(p) = \frac{D(p)+c}{n_p(p)+1} \qquad p_i^* = \arg\min_p \left(O(pi) + \lambda \sum_{j=1}^{i-1} \Omega(pi, pj) \right)$$

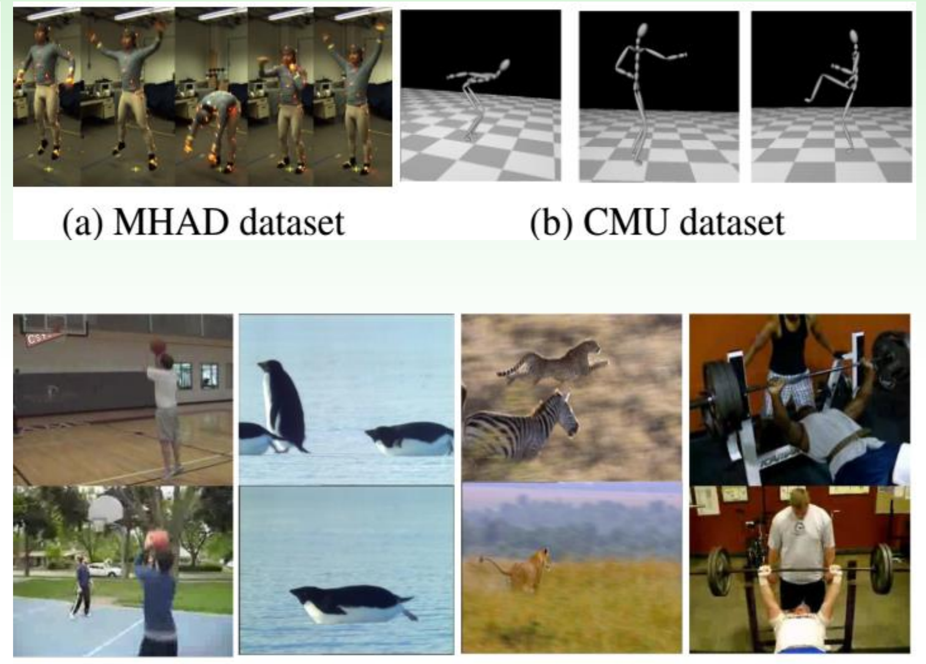
Supervised vs Unsupervised Action Co-segmentation

S-EVACO: Number of commonalities is known a-priori.
U-EVACO: The number of common actions is determined automatically by solving a model selection task.

$$j^* = \underset{j \in \{1,\ldots,K-1\}}{\operatorname{argmax}} \left| \frac{1}{j}\sum_{i=1}^{j} O(pi) - \frac{1}{K-j}\sum_{i=j+1}^{K} O(pi) \right|$$

EXPERIMENTAL RESULTS



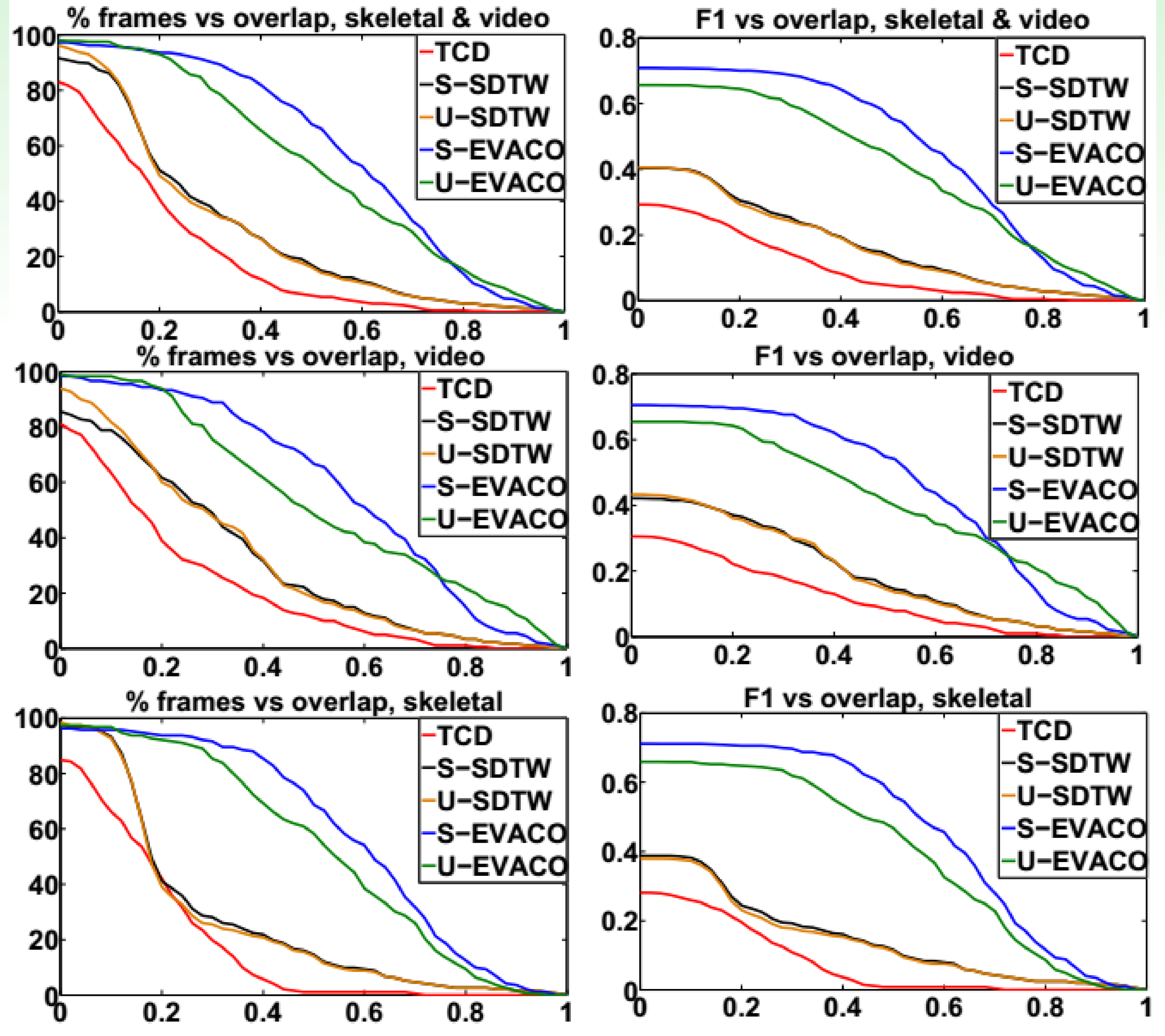


(a) MHAD dataset (b) CMU dataset

(c) 80-Pair dataset

Datasets/Features/Info		#Pairs	#Common (per pair)	#Common (total)
MHAD101-s	3D skeletal	101	1 – 4	203
CMU86-91 [1]	3D skeletal	91	3 – 18	609
MHAD101-v	MBH,IDT HOF, HOG	101	1 – 4	203
80-pair [3]	MBH	80	1	80

Overlap scores %	3D skeletal data		Video data	
Methods / Datasets	MHAD 101-s	CMU86 -91	MHAD 101-v	80-pair
TCD [1]	08.5	24.1	19.3	21.5
S-SDTW [2]	35.9	16.1	37.7	21.6
U-SDTW [2]	35.1	16.1	35.5	25.6
S-EVACO	**59.4**	**57.5**	**56.2**	**64.5**
U-EVACO	*50.3*	*51.0*	*45.9*	*54.2*
Guo et.al [3]	-	-	-	51.6



- **Datasets available online:** www.ics.forth.gr/cvrl/evaco
- **Contact**: papoutsa@ics.forth.gr http://users.ics.forth.gr/~papoutsa/
- This research received funding from the EU H2020 projects Co4Robots, ACANTO.