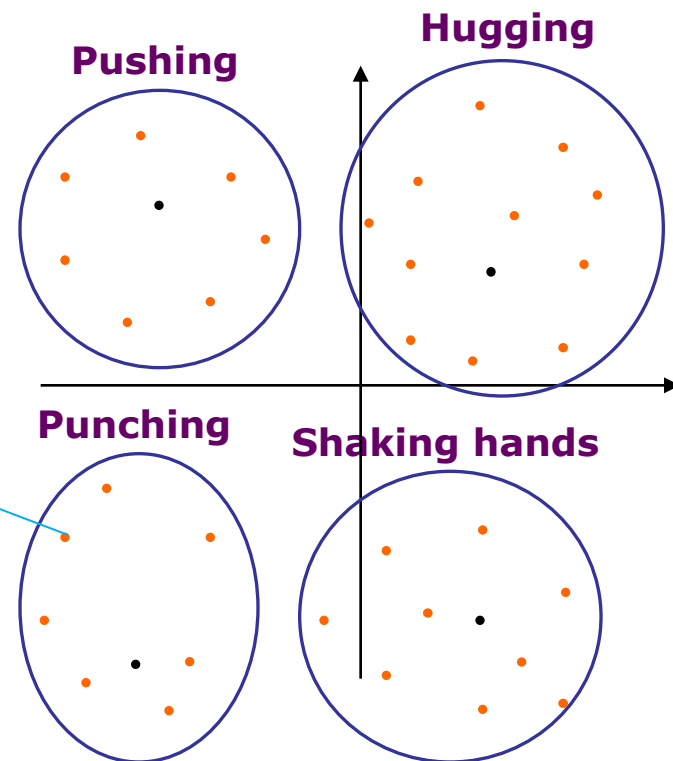
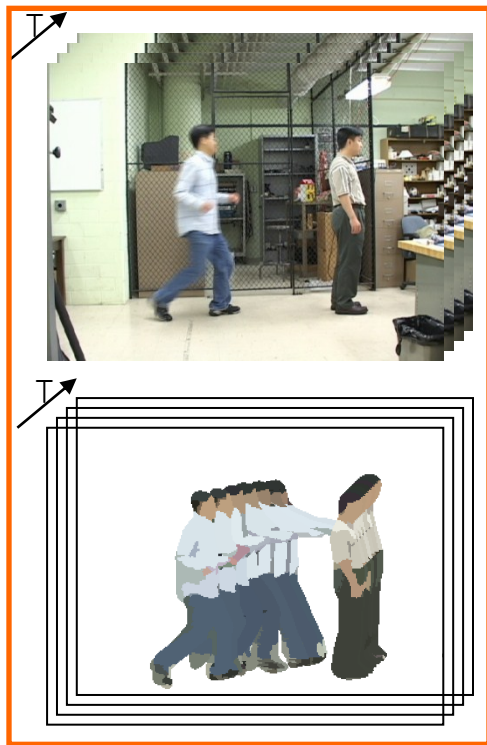

Frontiers of ***Human Activity Analysis***

J. K. Aggarwal
Michael S. Ryoo
Kris M. Kitani

Overview

Machine point of view

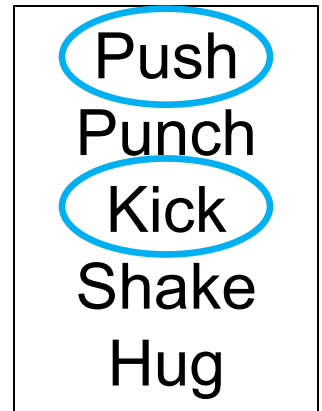
- Activities as videos
 - Activity = a particular set of videos



video space (640*480*100 D)

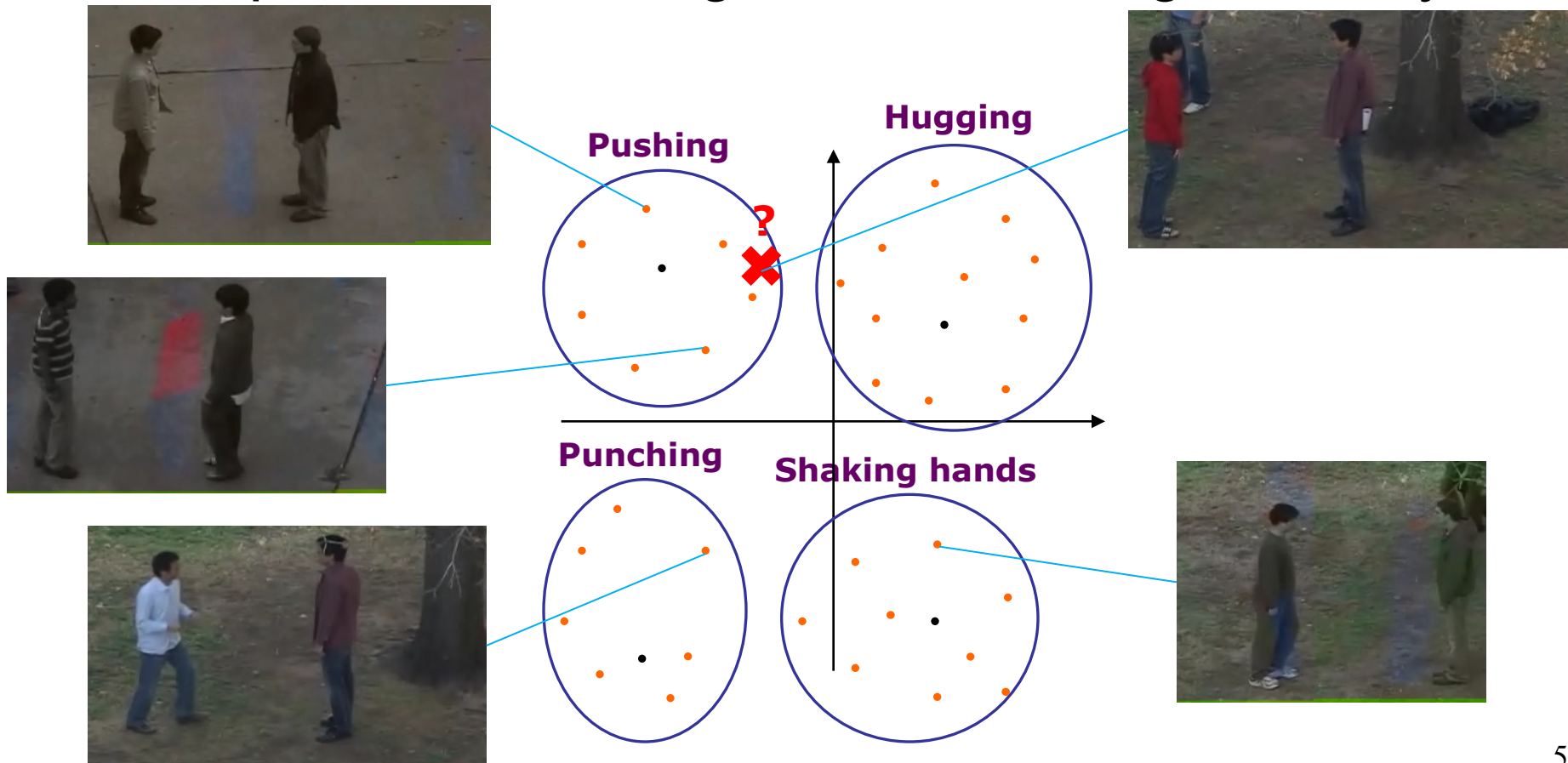
Activity classification

- Simple task of identifying videos
 - Categorize given videos into their types.
 - Known, limited number of classes
 - Assumes that each video contains a single activity



Activity classification

- Activity categorization
 - Input = a video segment containing 1 activity

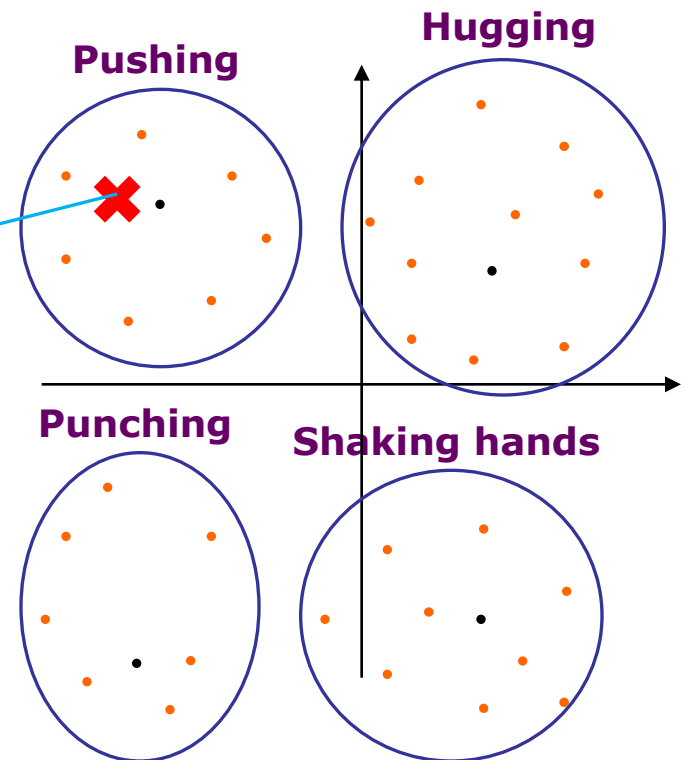


Activity detection

- **Search** for the particular time interval
 - <starting time, ending time>
 - Video segment containing the activity

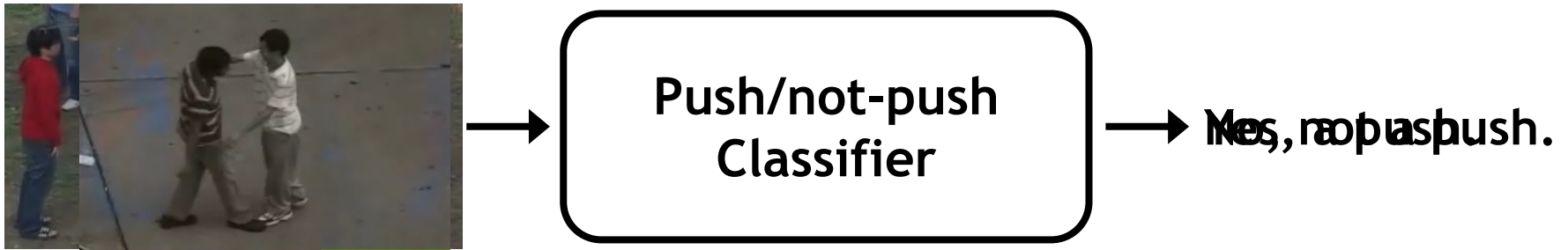
Input:

continuous video stream



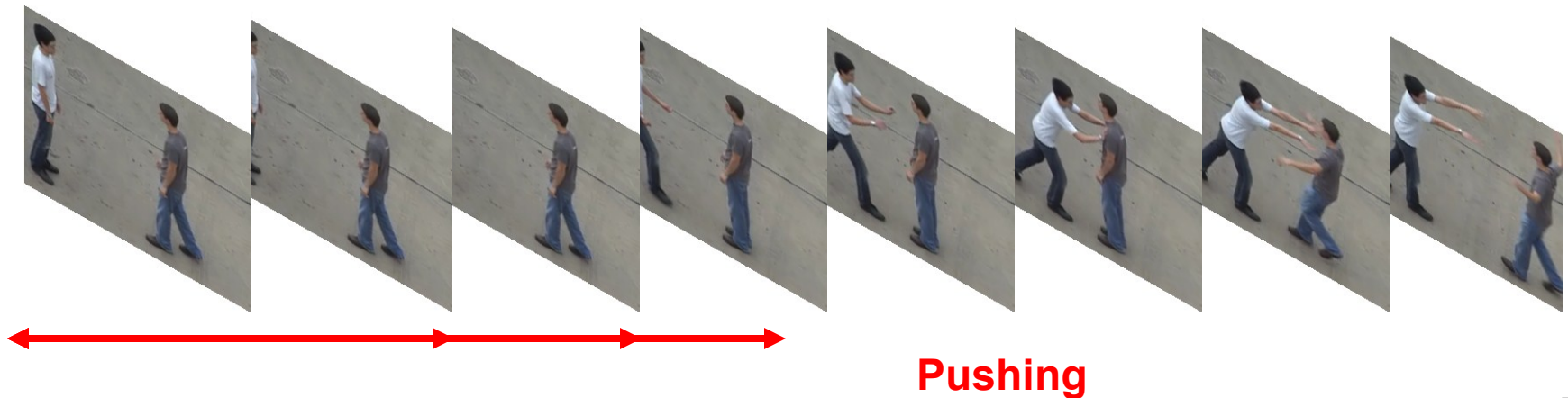
Activity detection by classification

- Binary classifier



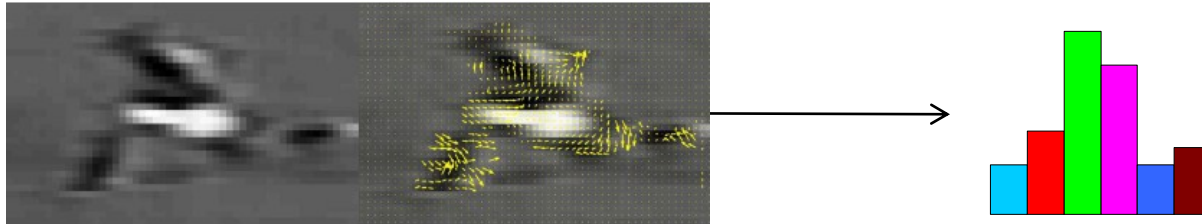
- Sliding window technique

- Classify all possible time intervals

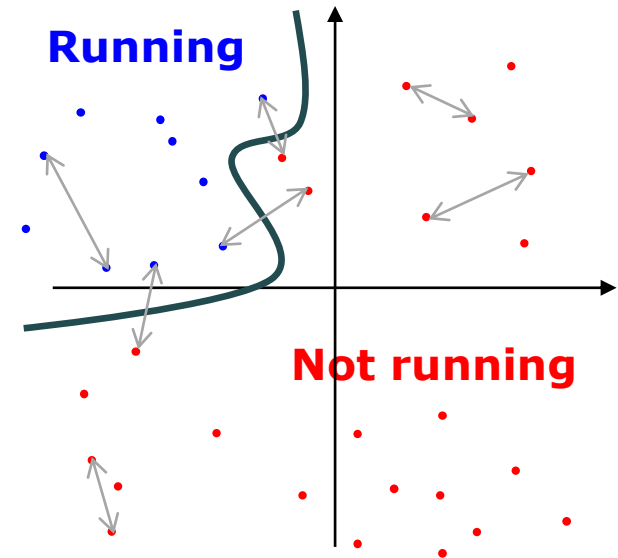


Recognition process

- Represent videos in terms of features
 - Captures properties of activity videos

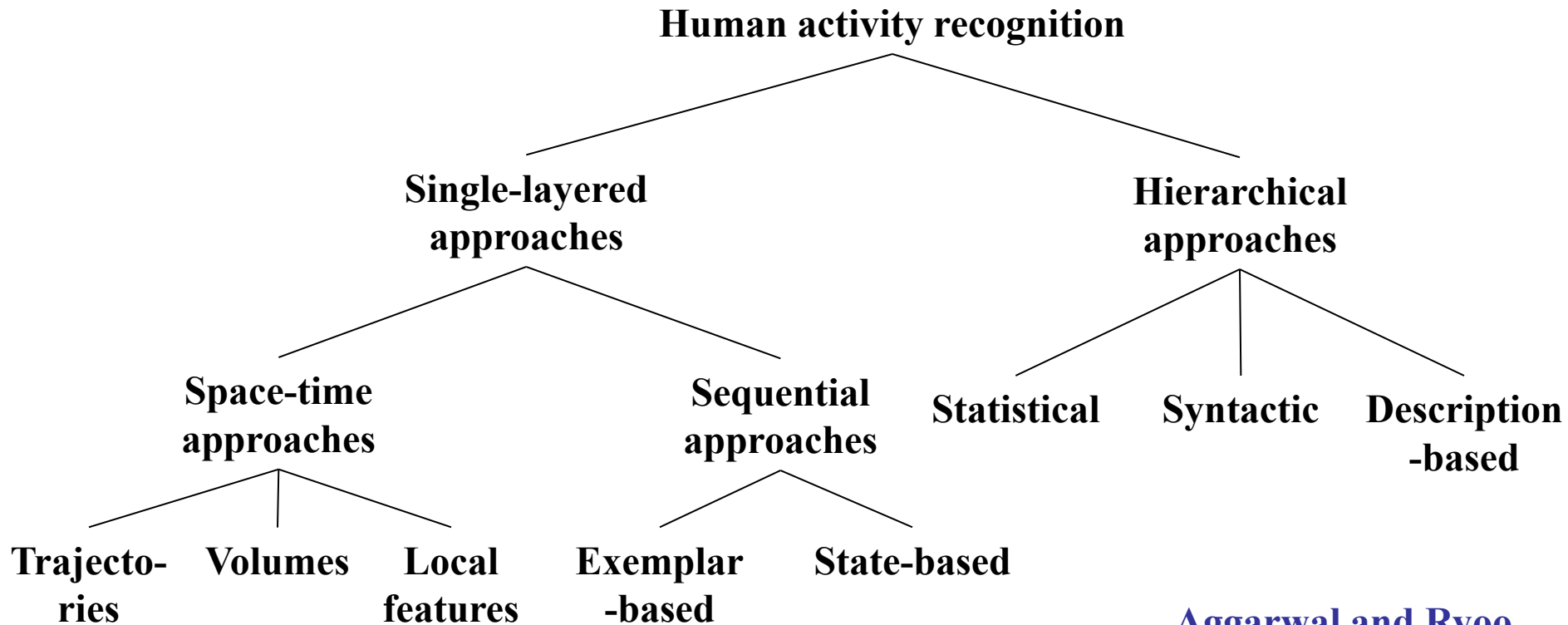


- Recognize activities by comparing video representations
 - Decision boundary



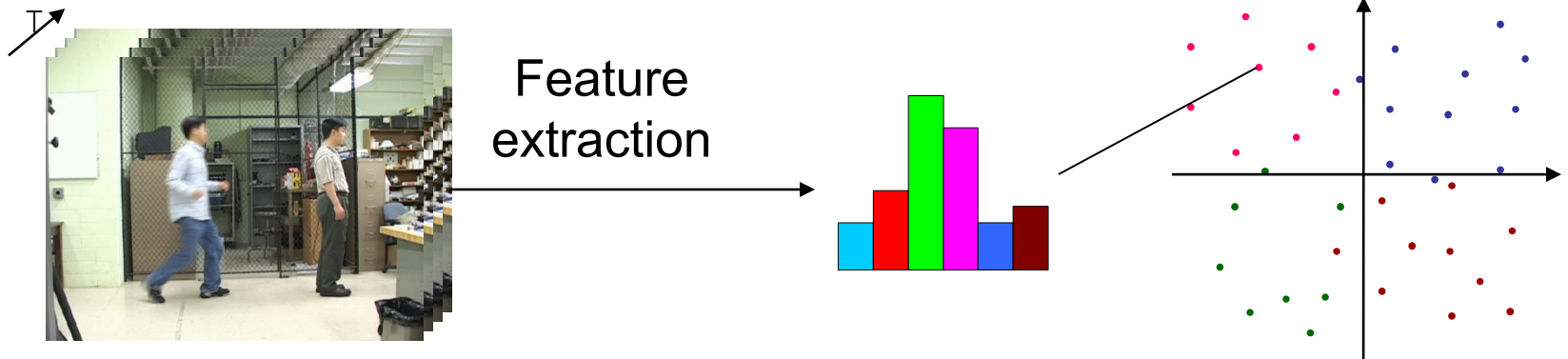
Taxonomy

- Approach based taxonomy
 - Recognition approaches can be categorized.

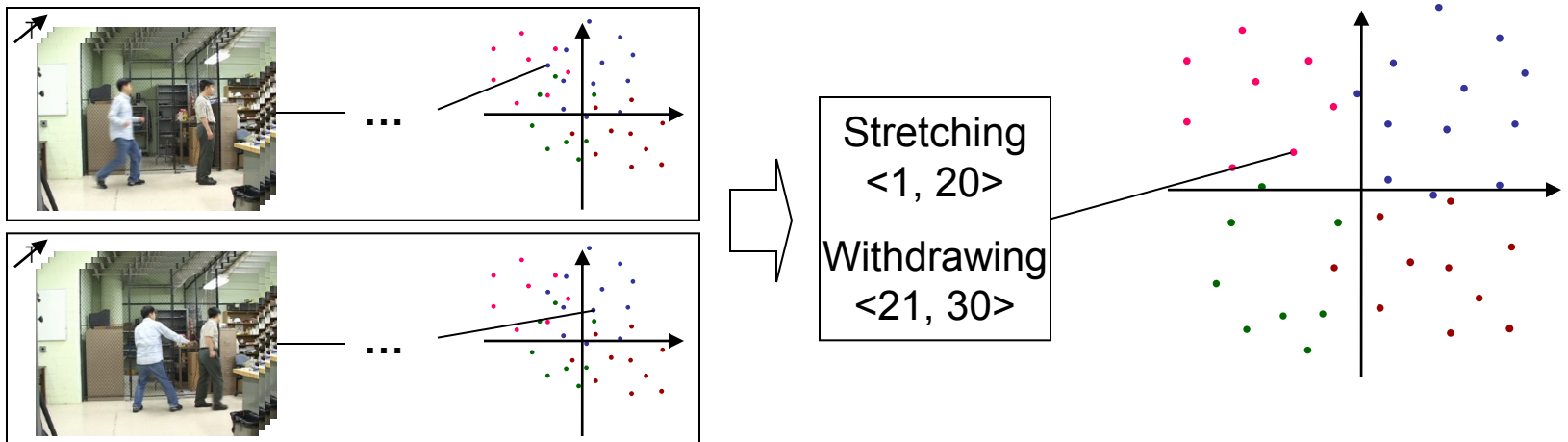


Single layered vs. hierarchical

- Single layered approaches



- Hierarchical approaches



Taxonomy – single layered

- These approaches recognize actions directly from a sequence of images.

Single-layered approaches

Space-time approaches

Sequential approaches

Trajectories

Space-time volume

Space-time features

Data-based

State model-based

Template matching

[Campbell and Bobick '95]
[Rao and Shah '01]

[Bobick and J. Davis '01]
[Shechtman and Irani '05]
[Rodriguez et al. '08]

[Zelnik-Manor '01]
[Laptev and Lindeberg '03]
[Dollar et al. '05]

[Darrell and Pentland '93]
[Gavrila and L. Davis '95]
[Yacoob and Black '98]
Ali and Aggarwal '01]

[Yamato et al. '92]
[Starner and Pentland '95]
[Bregler '97]

Neighbor-based (including SVM)

[Sato and Aggarwal '04]

[Efros et al. '03]
[Yilmaz and Shah '05]
[Ke et al. '07]

[Shuldt et al. '04]
[Blank et al. '05]
[Scovanner et al. '07]
[Laptev et al. '08]

[Veeraraghavan et al. '06]
[Lubliner et al. '06]
[Jiang et al. '06]

[Bobick and Wilson '97]
[Oliver et al. '00]
[Park and Aggarwal, '04]
[Natarajan and Nevatia '07]

[Vaswani et al. '03]^G

[Moore et al. '99]^O

Statistical matching

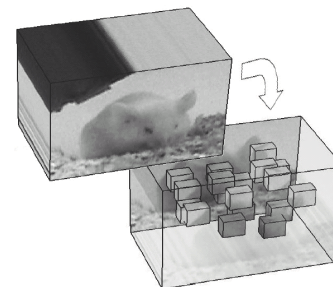
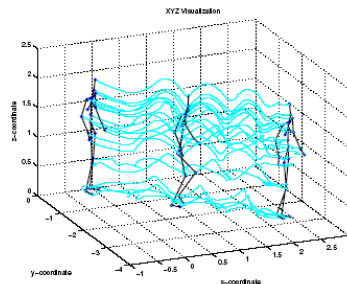
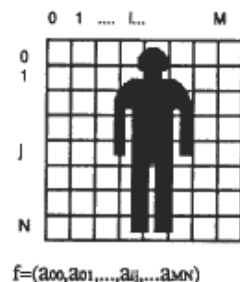
[Sheikh et al. '05]
[Khan and Shah '05]^G

[Chomat and Crowley '99]
[Niebles et al. '06, '08]
[Wong et al. '07]
[Lv et al. '04]^G

[Gupta and Davis '07]^O
[Filipovych and Ribeiro '08]^O

Single layered approaches

- Action representation
 - Video volumes themselves
 - Features directly extracted from videos



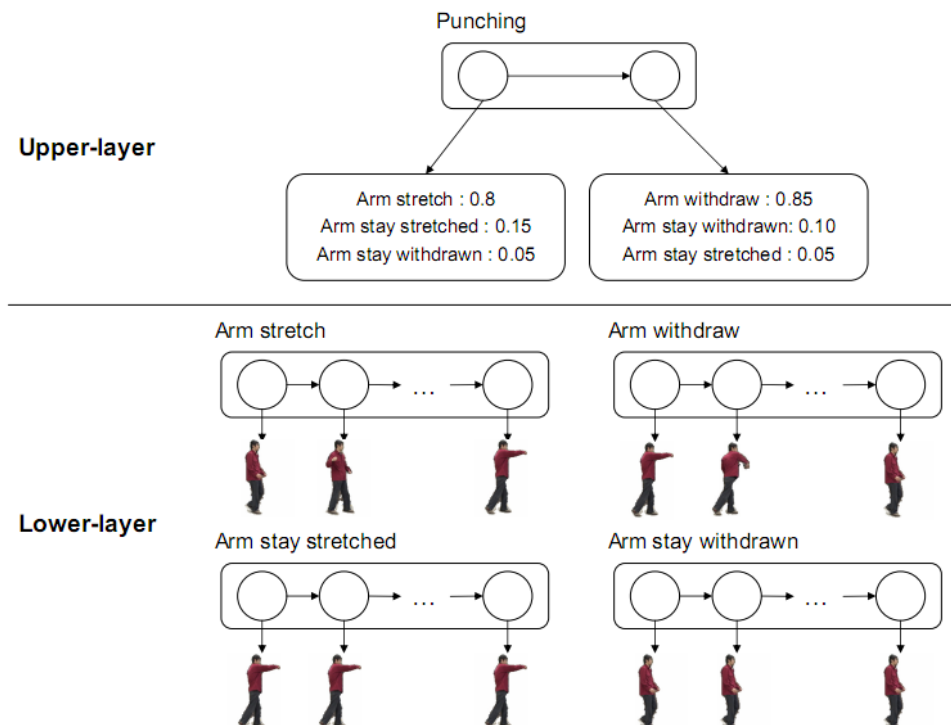
- Action classification
 - Machine learning techniques
 - Support vector machines
 - Hidden Markov models

Taxonomy – Hierarchical

Hierarchical approaches			
	Statistical approaches	Syntactic approaches	Description-based approaches
Human actions	[Nguyen et al. '05]		[Pinhanez and Bobick '98] [Gupta et al. '09]
Human-Human interactions	[Oliver et al. '02]	[Ivanov and Bobick '00] [Joo and Chellapha '06]	[Intille and Bobick '99] [Vu et al. '03] [Ghanem et al. '04] [Ryoo and Aggarwal '06, '09a]
Human-Object interactions	[Shi et al. '04] ^o [Yu and Aggarwal '06] ^o [Damen and Hogg '09] ^o	[Moore and Essa '02] ^o [Minnen et al. '03] ^o [Kitani et al. '07] ^o	[Siskind '01] ^o [Nevatia et al. '03, '04] ^o [Ryoo and Aggarwal '07] ^o
Group activities	[Cupillard et al. '02] ^G [Gong and Xiang '03] ^G [Zhang et al. '06] ^G [Dai et al. '08] ^G		[Ryoo and Aggarwal '08, '10] ^G

Hierarchical approaches

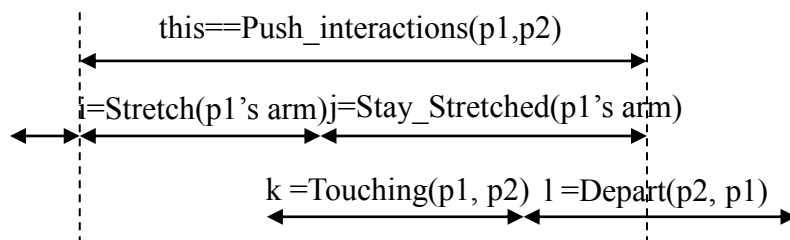
- Layered approaches
 - Activities in terms of sub-events.
 - Human interactions
 - Multiple agents
- Suitable for activity-level recognition



Hierarchical approaches

- Activities as semantic structures
 - Activity = a concatenation of its sub-events
 - Human-oriented: high-level
 - Hierarchically organized *representations*

Hand shake = “*two persons* do **shake-action**
(**stretches**, **stays stretched**, **withdraw**) **simultaneously**,
while touching”.



Fighting	->	Punching	: 0.3
		Punching Fighting	: 0.7
Punching	->	stretch withdraw	: 0.8
		stretch stay_withdrawn	: 0.1
		stay_stretched withdraw	: 0.1