Frontiers of
Human Activity Analysis

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

THE UNIVERSITY OF TEXAS AT AUSTIN
UT ECE
ELECTRICAL & COMPUTER ENGINEERING

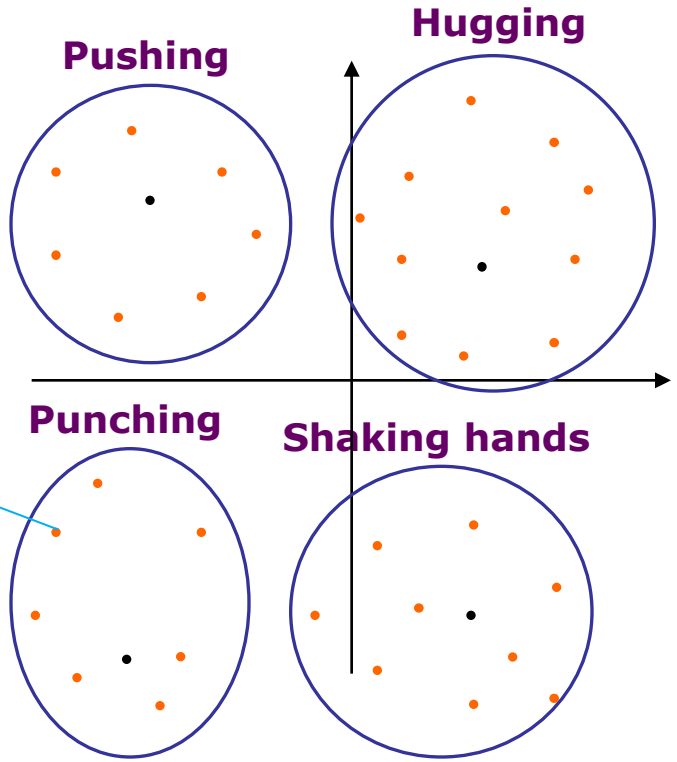
ETRI

Carnegie Mellon
THE ROBOTICS INSTITUTE

Overview

Machine point of view

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



video space (640*480*100 D)

Activity classification

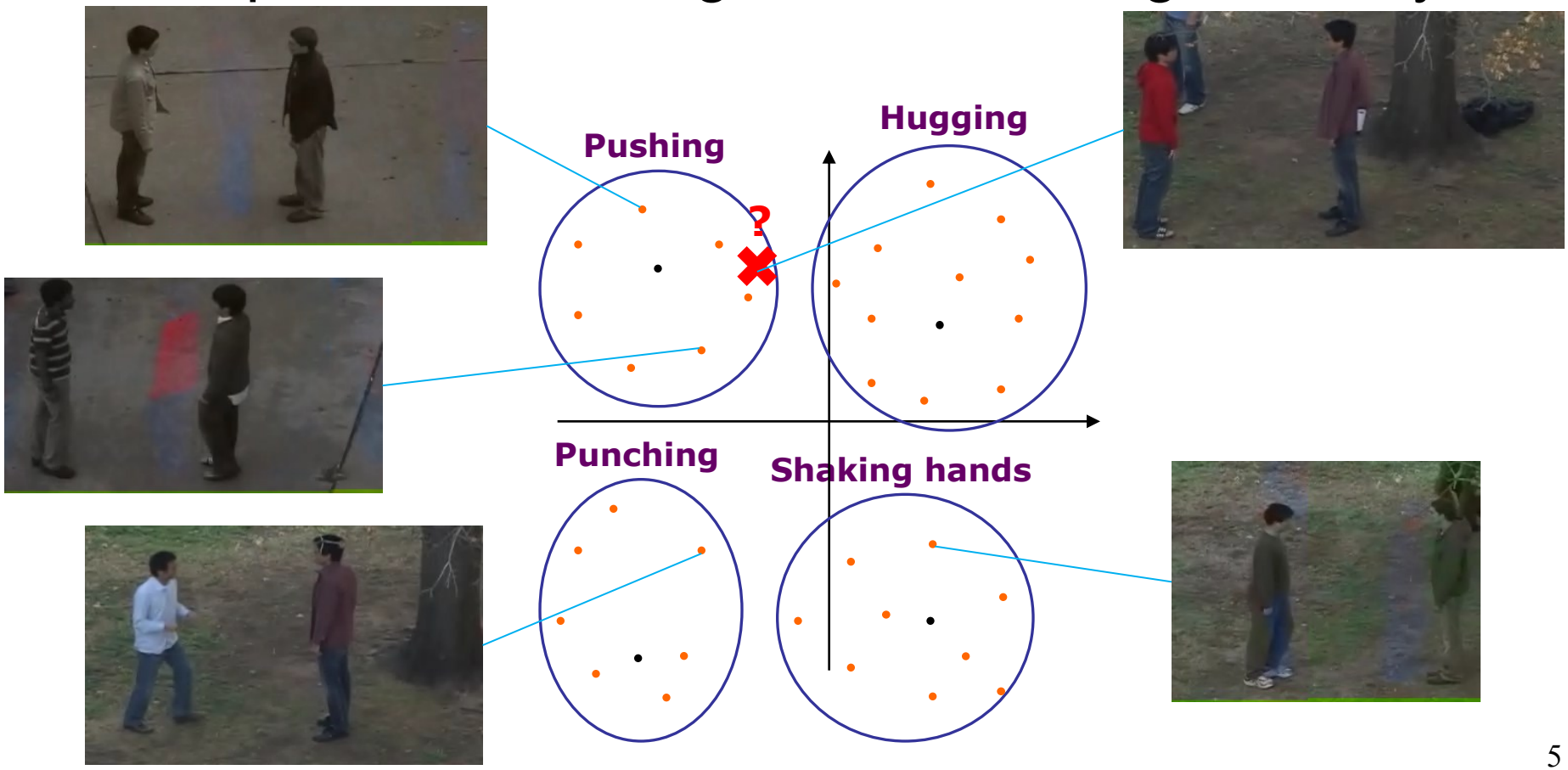
- Simple task of identifying videos
  - Categorize given videos into their types.
    - Known, limited number of classes
    - Assumes that each video contains a single activity

?

Push
Punch
Kick
Shake
Hug

Activity classification

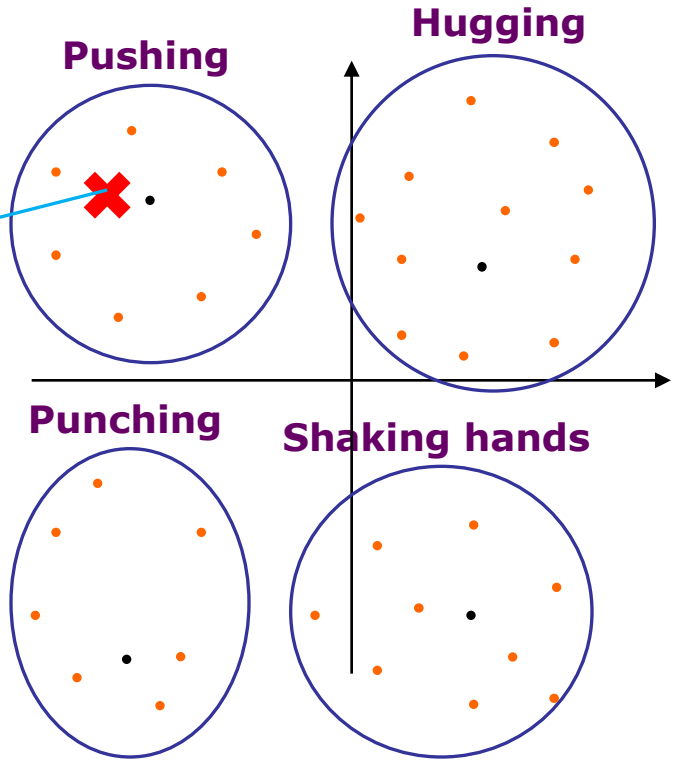
- Activity categorization
  - Input = a video segment containing 1 activity



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

Push/not-push Classifier

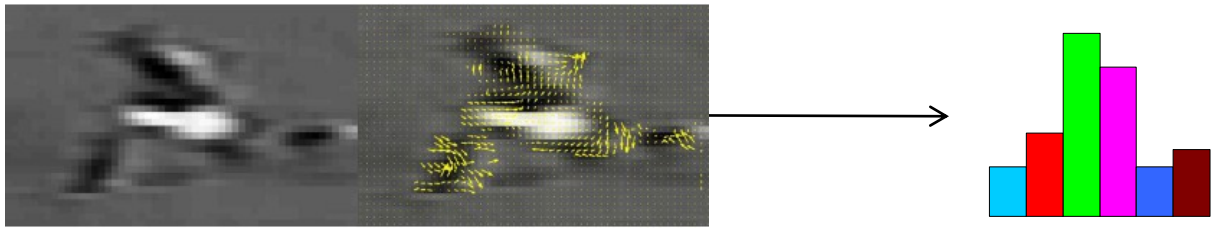
Yes, a push.

- Sliding window technique
  - Classify all possible time intervals

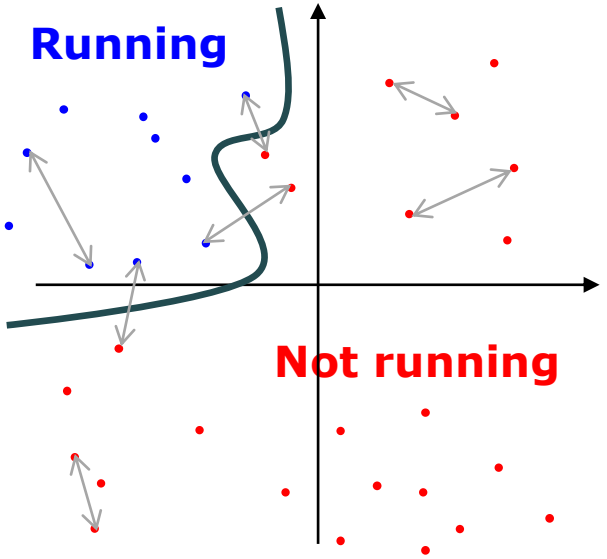
Pushing

Recognition process

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

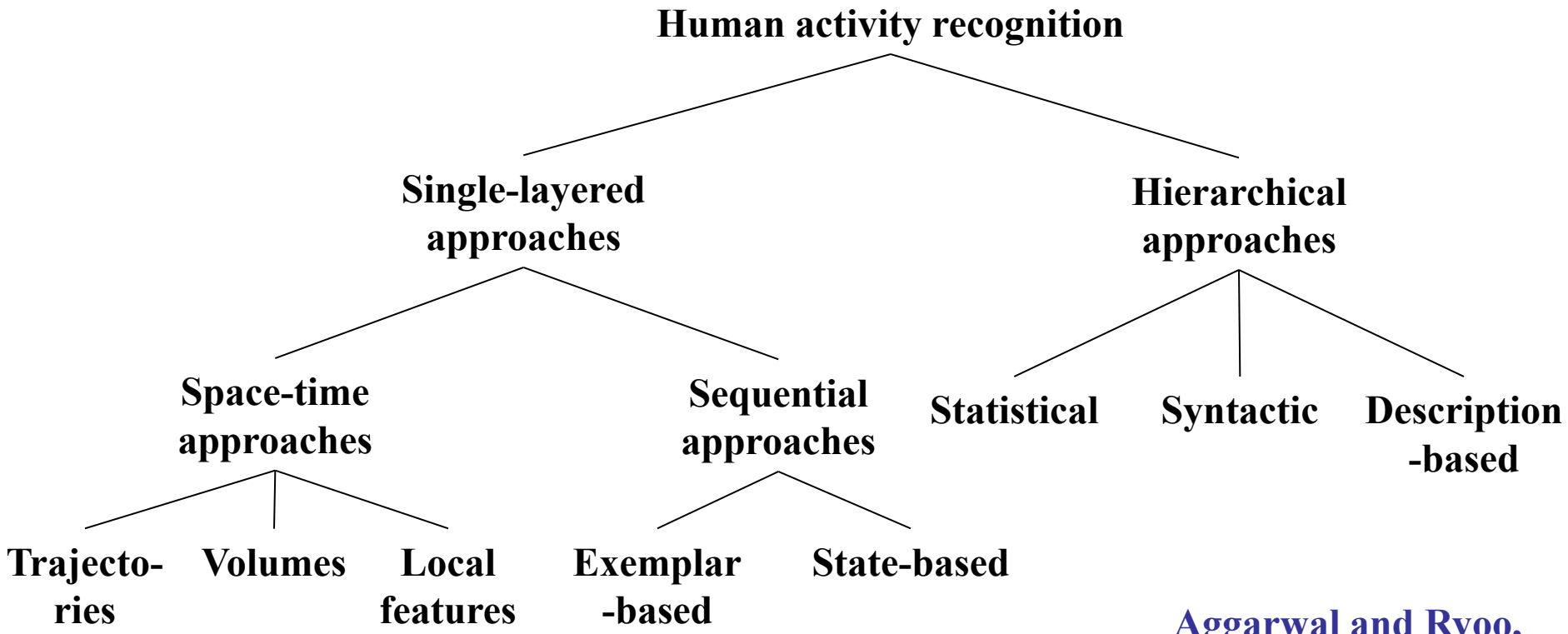


- Recognize activities by comparing video representations
  - Decision boundary



Taxonomy

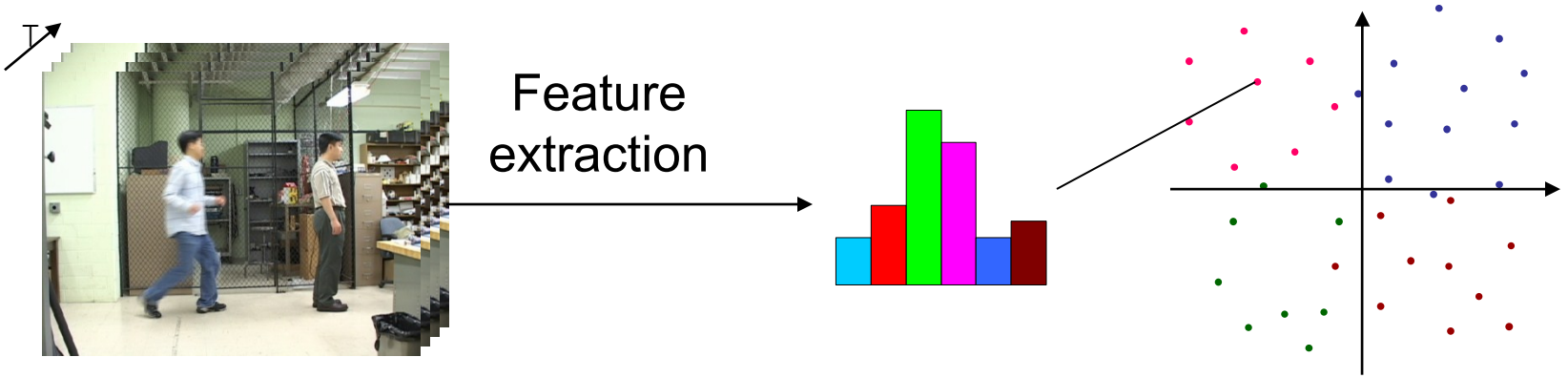
- Approach based taxonomy
  - Recognition approaches can be categorized.



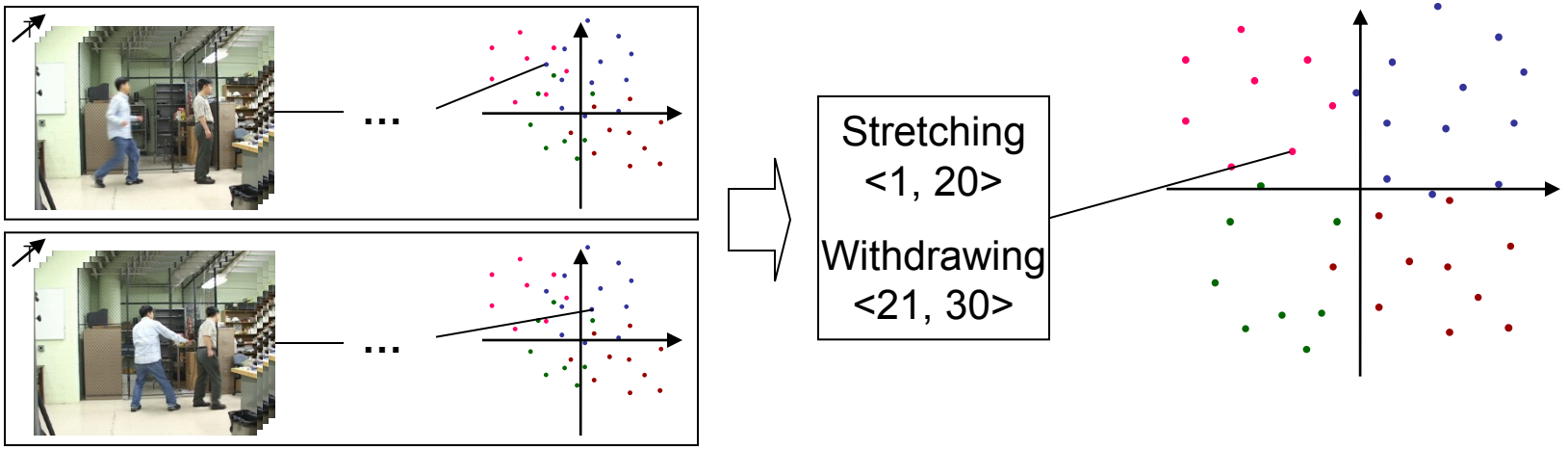
Aggarwal and Ryoo, ACM CSUR 2011

Single layered vs. hierarchical

- Single layered approaches

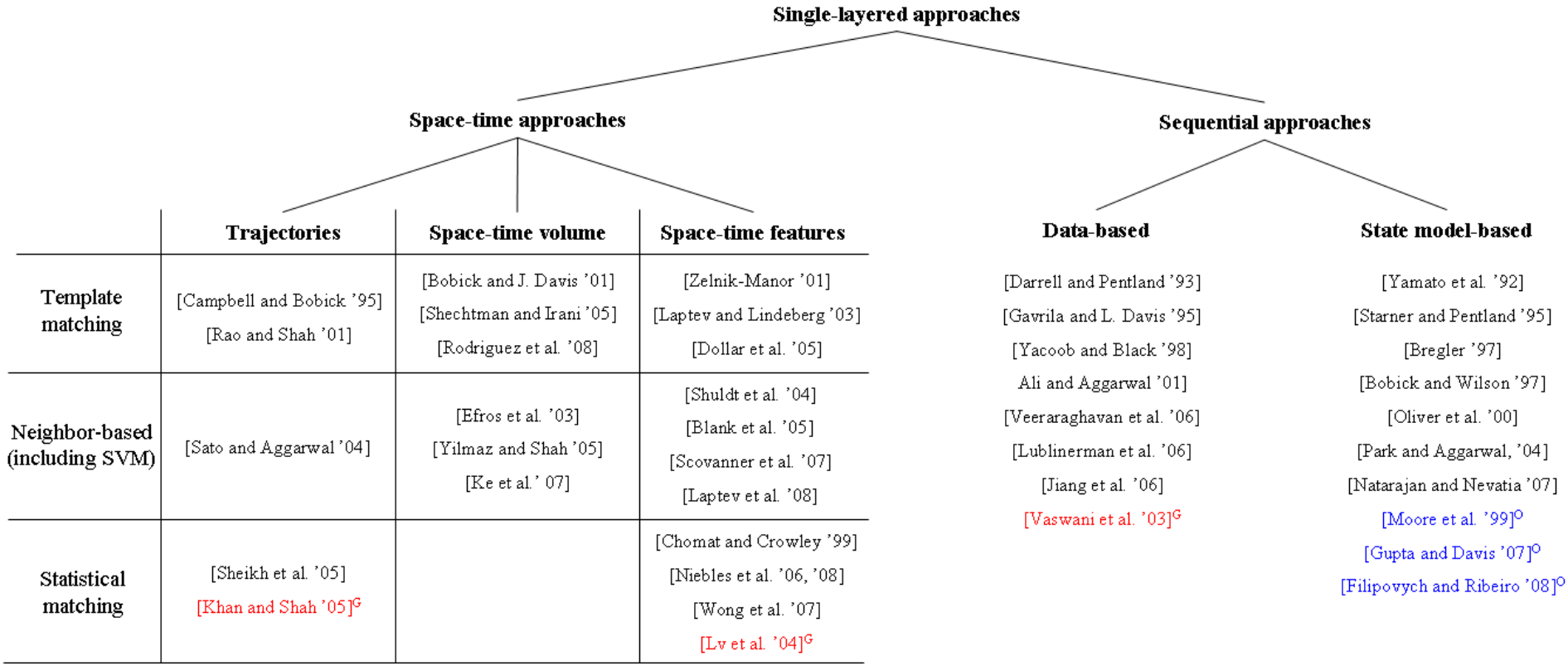


- Hierarchical approaches



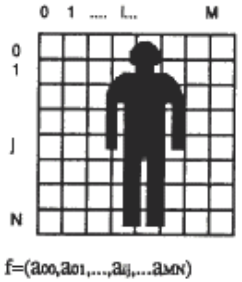
Taxonomy – single layered

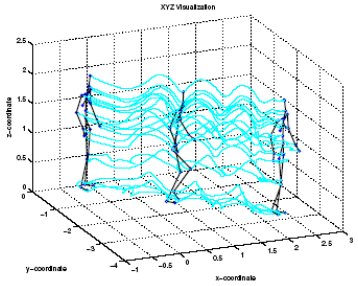
- These approaches recognize actions directly from a sequence of images.

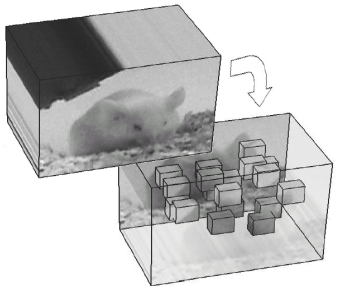


Single layered approaches

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







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

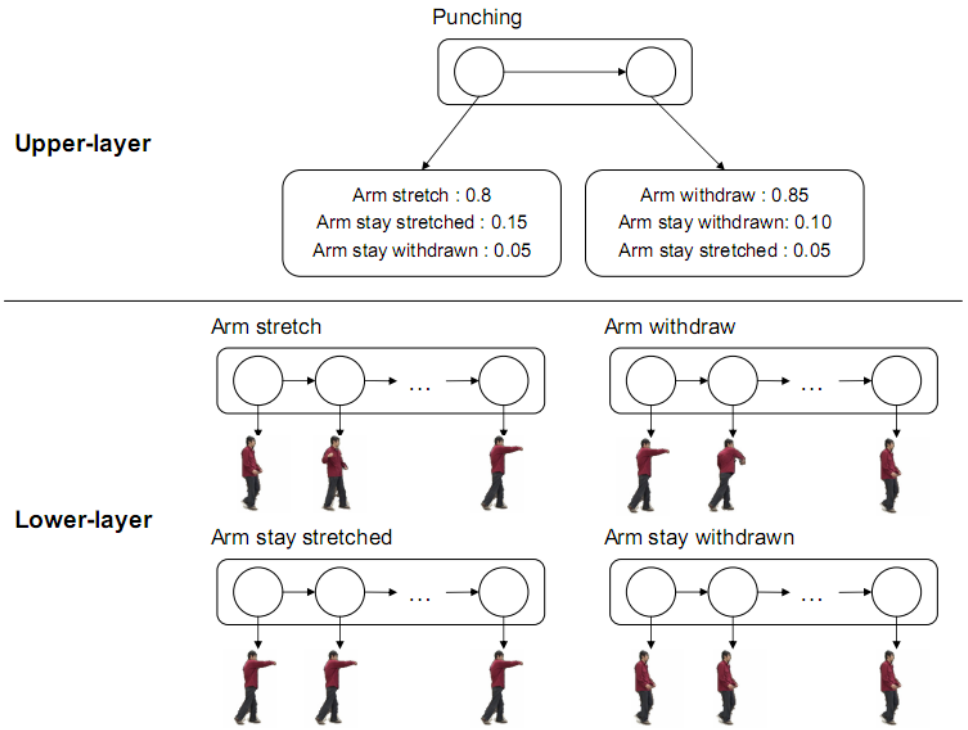
Taxonomy – Hierarchical

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

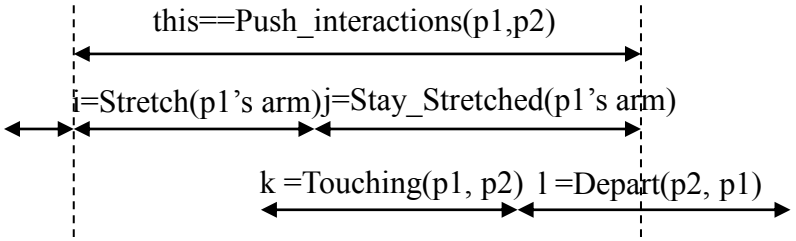
- Layered approaches
  - Activities in terms of sub-events.
  - Human interactions
    - Multiple agents
- Suitable for activity-level recognition



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