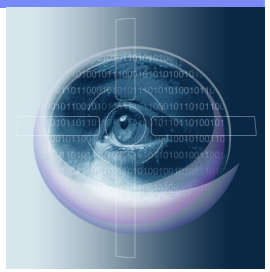


Behavior Analysis

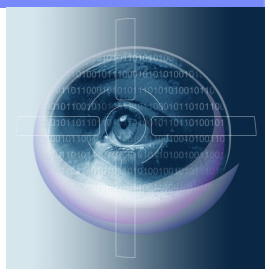
Rogerio Feris

IBM TJ Watson Research Center

rsferis@us.ibm.com

http://rogerioferis.com

Outline

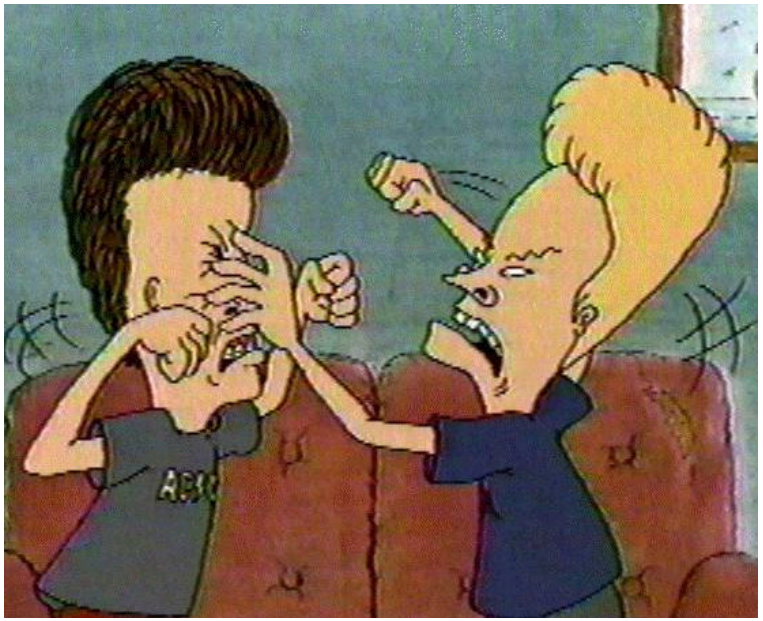


- Motivation
- Action Recognition
  - Template-Based Approaches
  - State-Space Approaches
- Detecting Suspicious Behavior

Motivation

- Action Recognition in Surveillance Video

Detecting people fighting



Falling person detection

Motivation

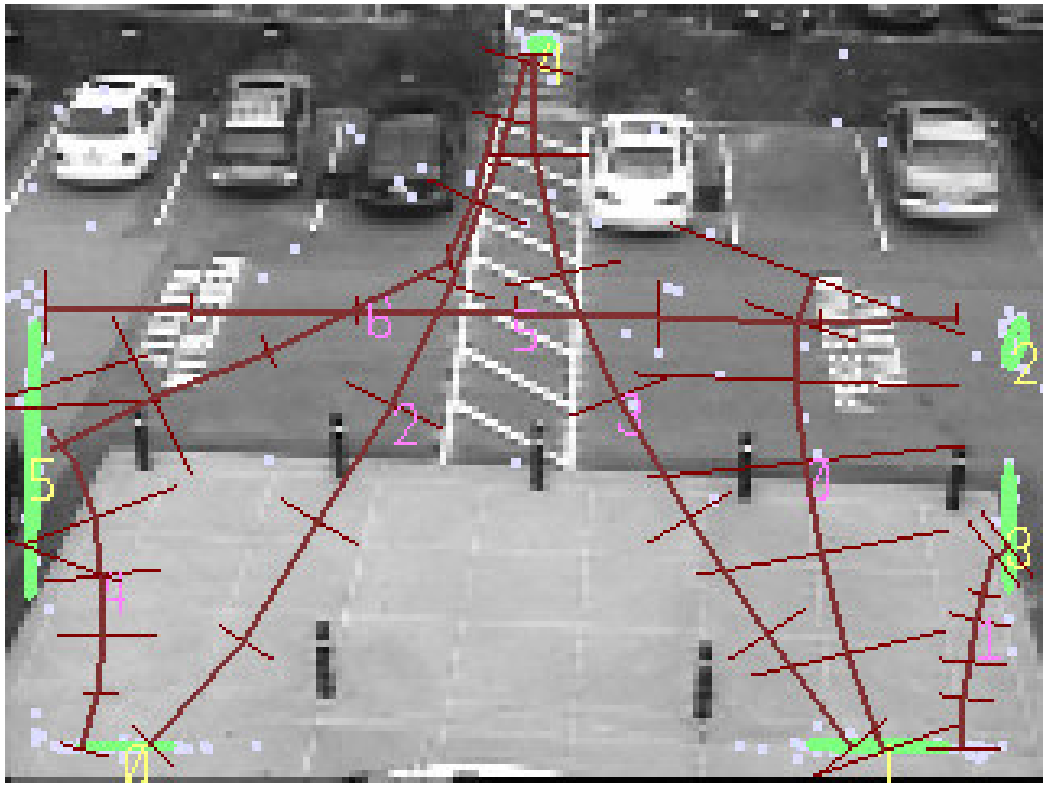
- Detecting suspicious behavior

Fence Climbing

[Boiman and Irani, 2005]

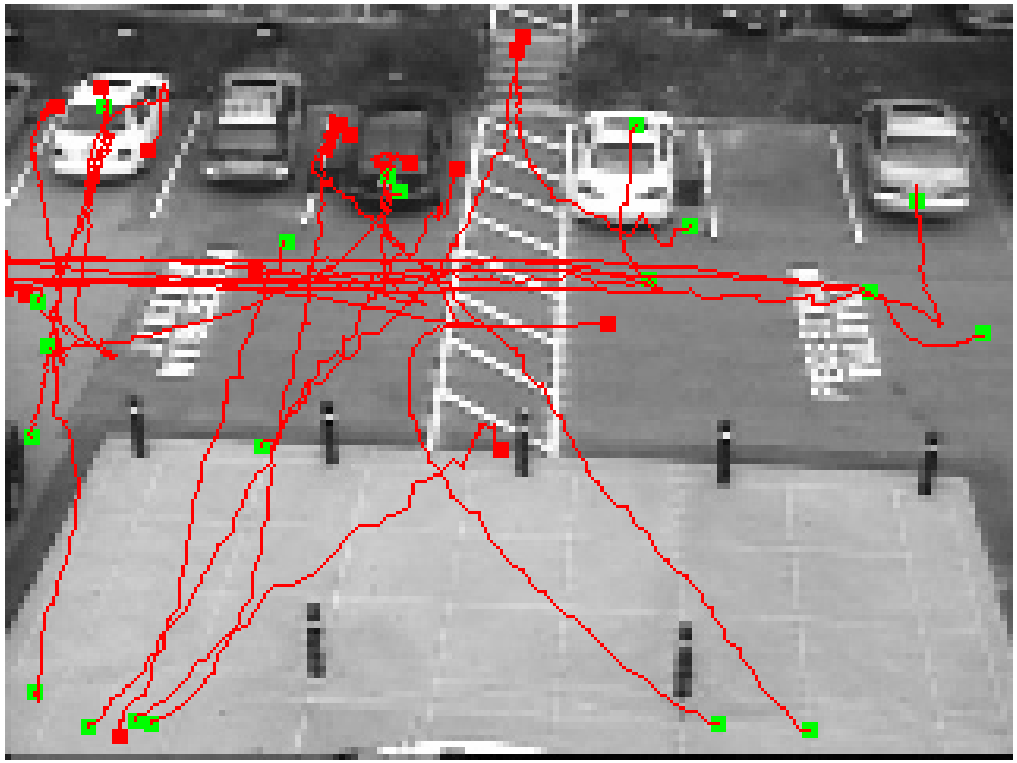
Motivation

- Find all locations where objects enter or exit (green)
- Find all ‘normal’ routes between these locations- average path and observed deviations.



Motivation

Tracks anomalies (not matching trained routes)



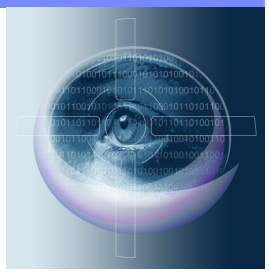
Motivation

- Long-term reasoning / object interaction

Car/person interactions (e.g., car picking up a person)

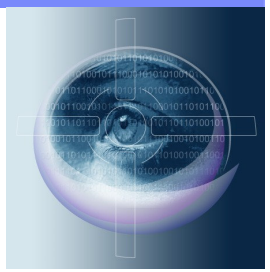
[Ivanov and Bobick, 2000]

Challenges



- Strong appearance variation in semantically similar events (e.g., people performing actions with different clothing
- Viewpoint Variation
- Duration of the action / frame rate
- Action segmentation – determining beginning and end of the action

Outline



- Motivation
- Action Recognition
  - **Template-Based Approaches**
  - State-Space Approaches
- Detecting Suspicious Behavior

Action Recognition – Template-Based

Temporal Templates [Bobick and Davis, 1996]

- Motion History Image (MHI): Scalar-valued image where brighter pixels correspond to more recently moving pixels

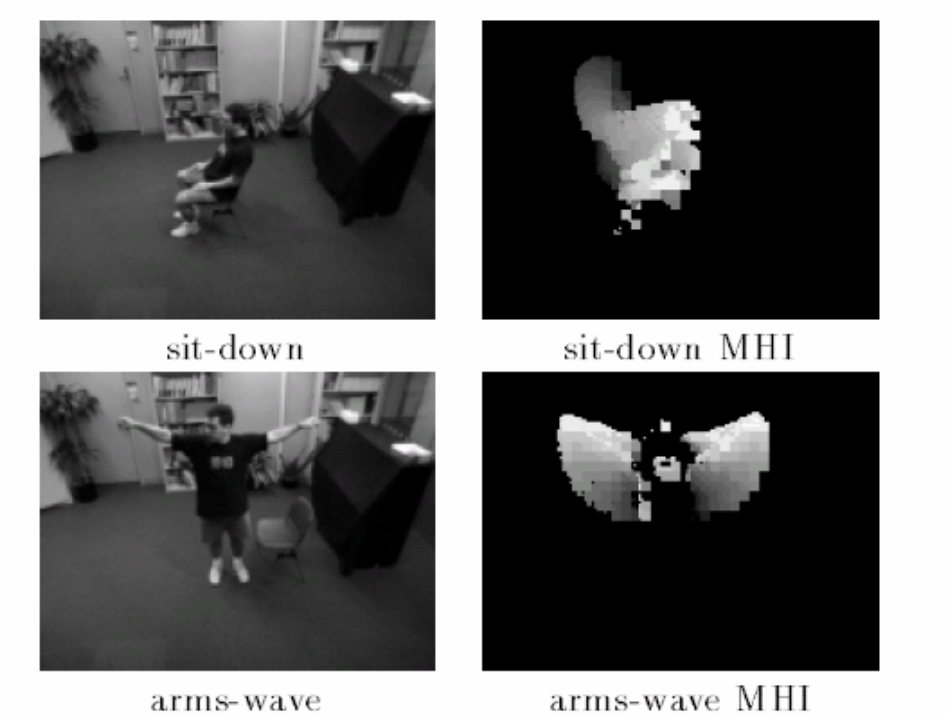
Binary image indicating regions of motion

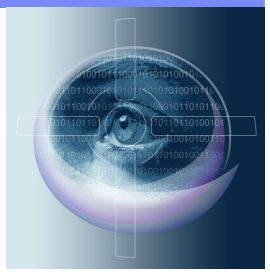
$$H_\tau(x,y,t) = \begin{cases} \tau & \text{if } D(x,y,t) = 1 \\ \max\ (0, H(x,y,t-1) - 1) & \\ & \text{otherwise} \end{cases}$$

Action Recognition – Template-Based

Temporal Templates [Bobick and Davis, 1996]

- Motion History Image (MHI): Scalar-valued image where brighter pixels correspond to more recently moving pixels





Action Recognition – Template-Based

Temporal Templates [Bobick and Davis, 1996]

- At the current frame, statistical descriptors based on moments (translation and scale invariant) are extracted from the current MHI and matched against stored exemplars for classification

- Three actions: sitting, arm waving , and crouching. View-based approach to handle camera view changes.

- Problems with ambiguities, occlusions, poor motion segmentation

Action Recognition – Template-Based

Recognizing Action at a Distance [Efros et al, ICCV'03]

- 300-pixel man
- Limb tracking
  - e.g. Yacoob & Black, Rao & Shah, etc.

- 3-pixel man
- Blob tracking
  - vast surveillance literature

Action Recognition – Template-Based

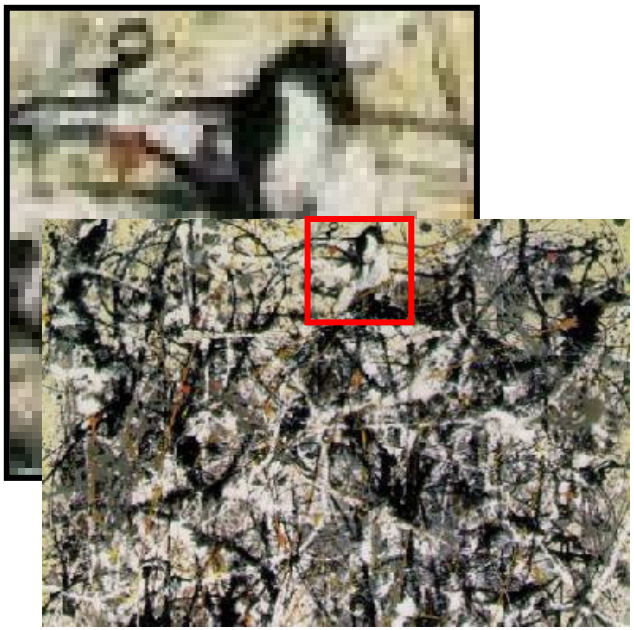
Recognizing Action at a Distance [Efros et al, ICCV'03]

The 30-Pixel Man

Action Recognition – Template-Based

Recognizing Action at a Distance [Efros et al, ICCV'03]

Appearance versus Motion



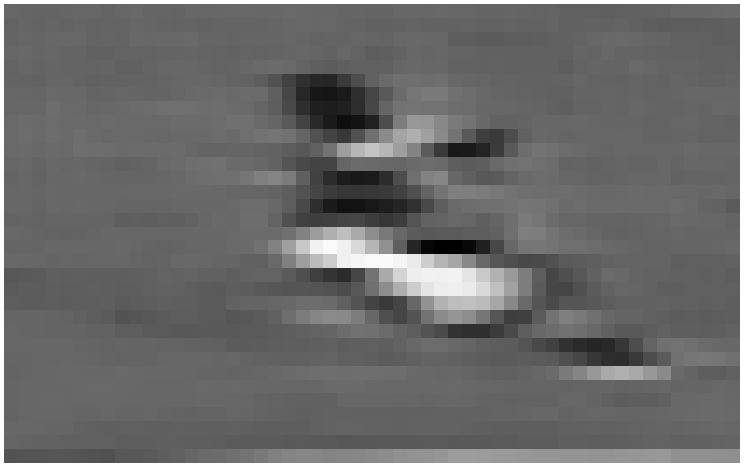
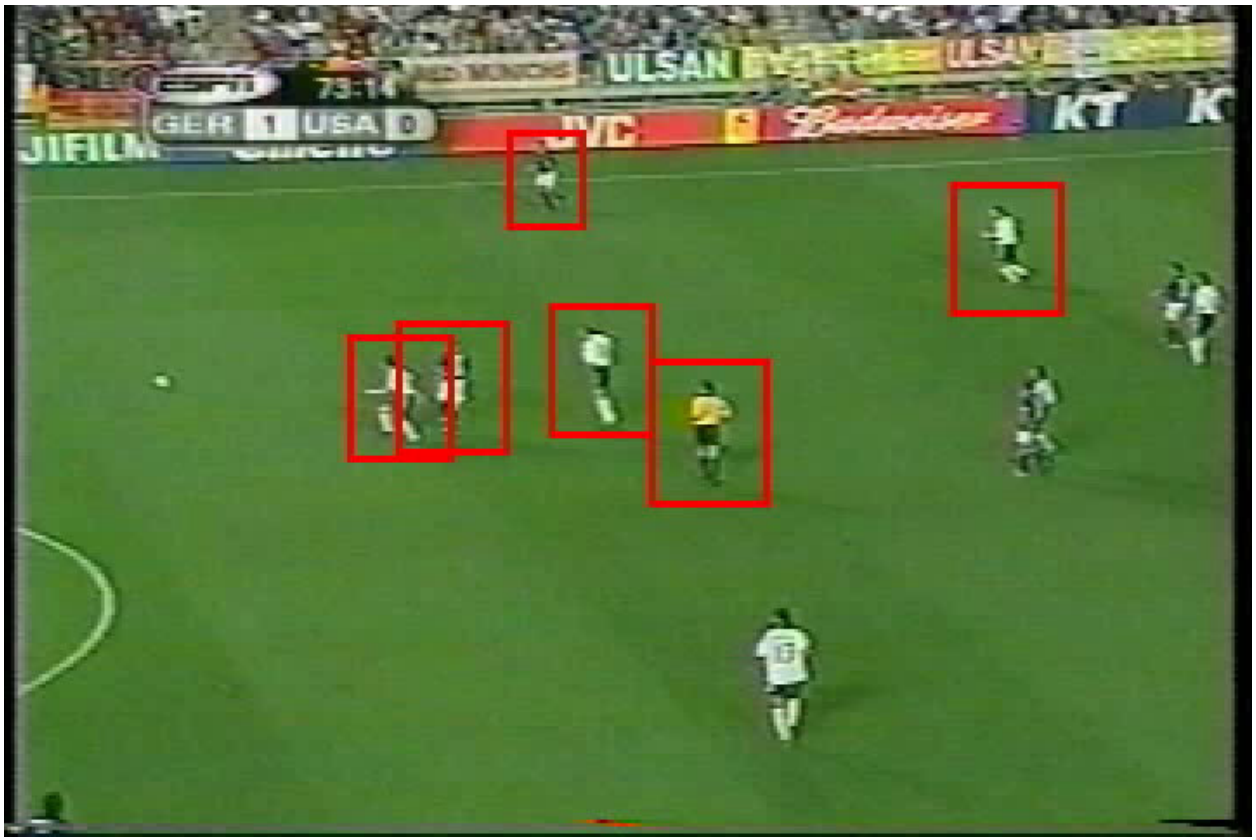


Figure-centric Representation

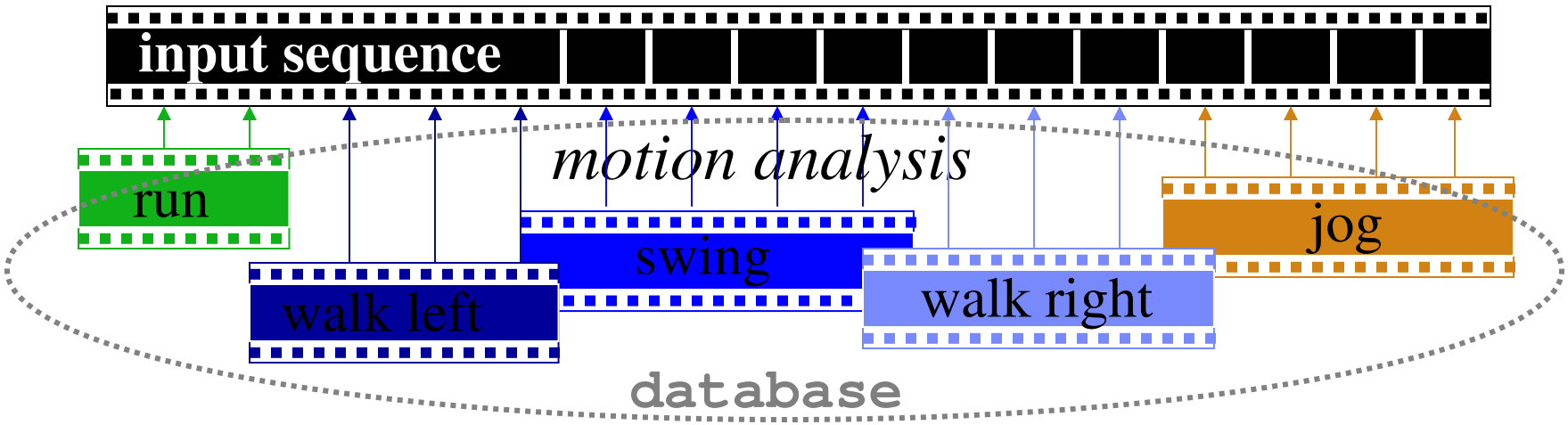


- Tracking
  - Simple correlation-based tracker
  - User-initialized

Action Recognition – Template-Based

Recognizing Action at a Distance [Efros et al, ICCV'03]

- "Explain" novel motion sequence by matching to previously seen video clips
  - For each frame, match based on some temporal extent



Challenge: how to compare motions?

Spatial Motion Descriptor

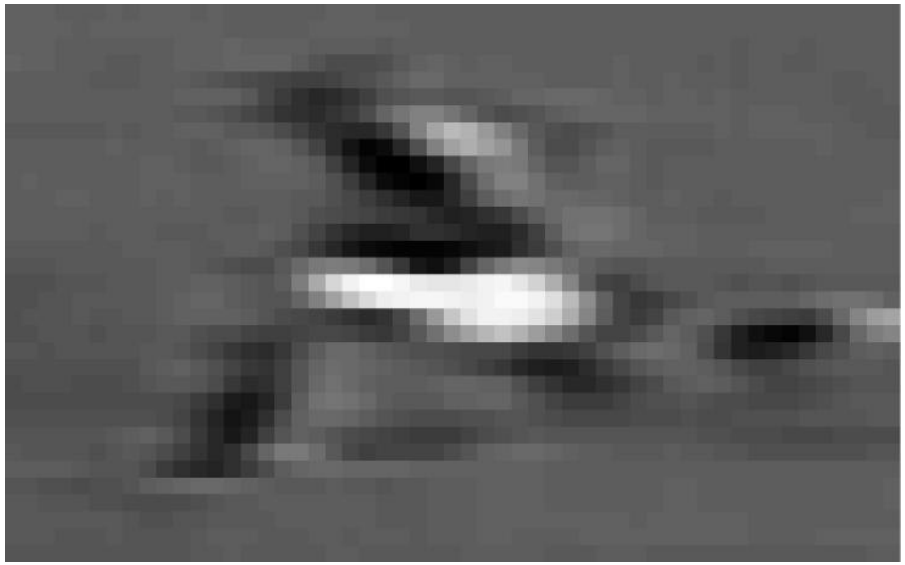
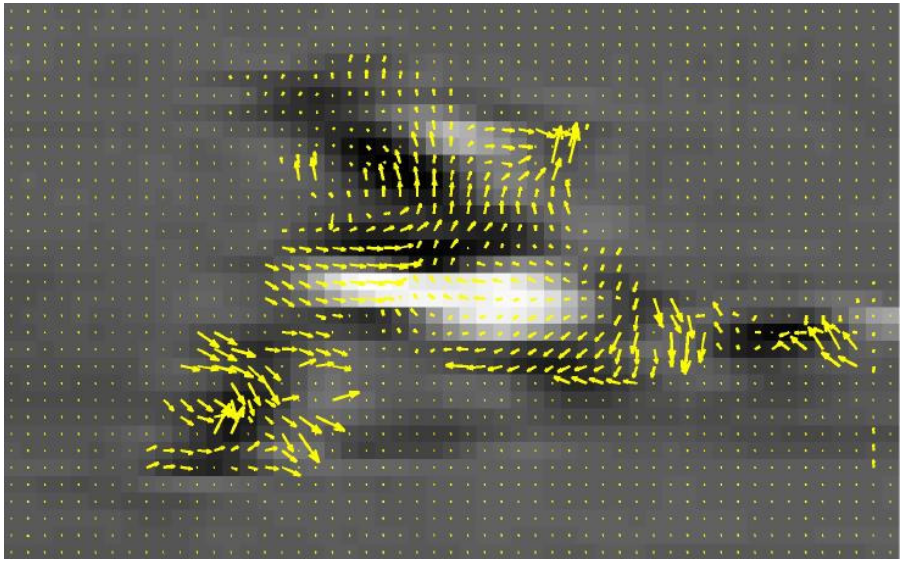
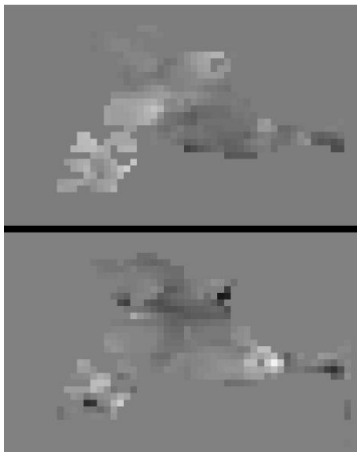


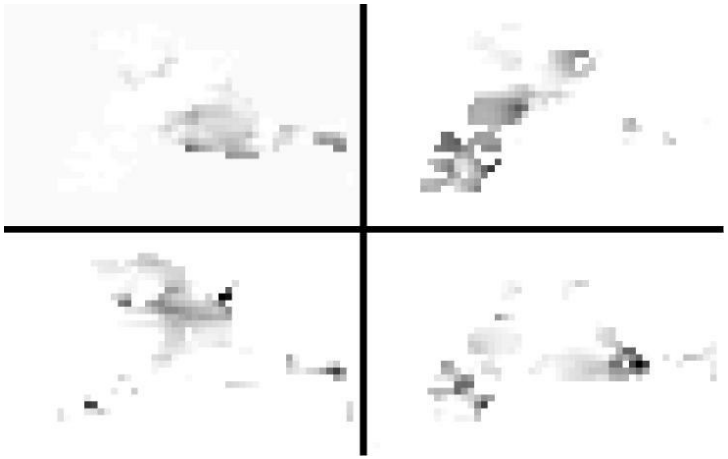
Image frame



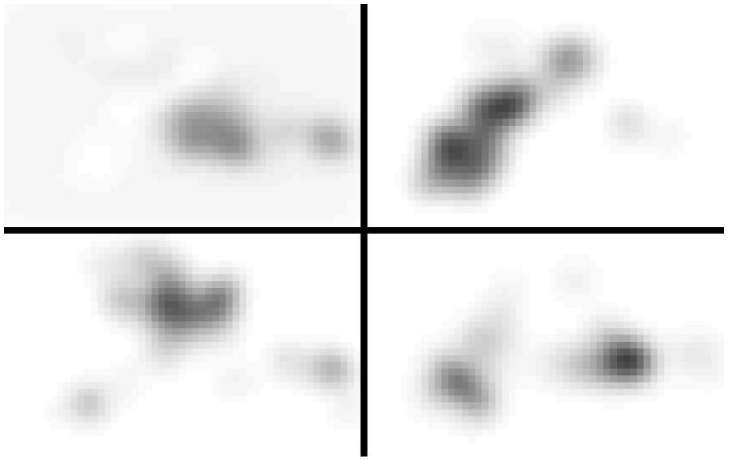
Optical flow $F_{x,y}$



F_x, F_y

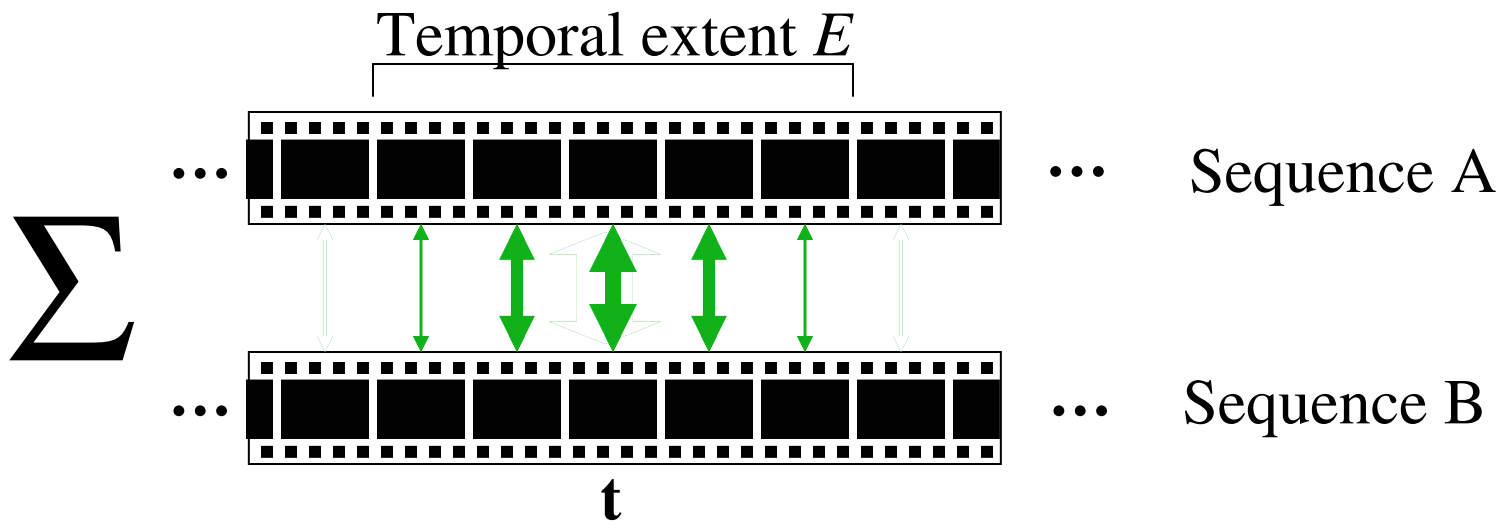


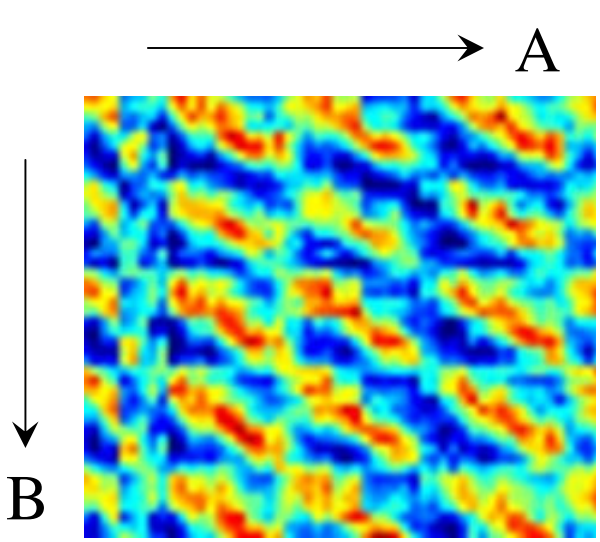
$F_x^-, F_x^+, F_y^-, F_y^+$



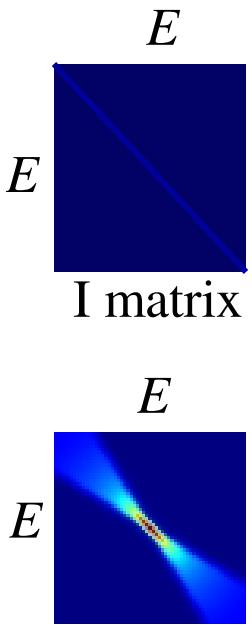
blurred $F_x^-, F_x^+, F_y^-, F_y^+$

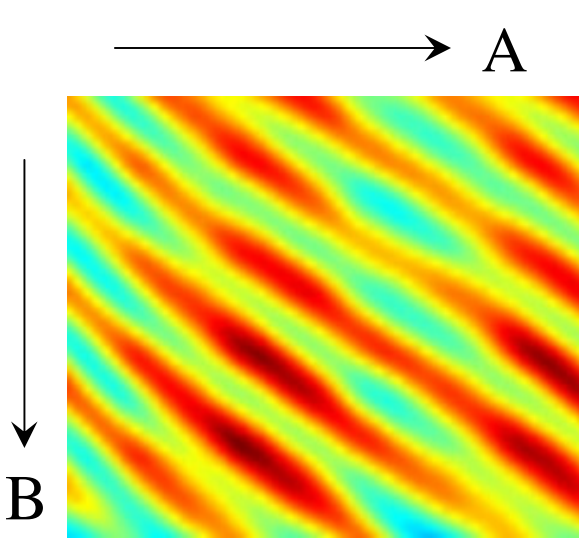
Two ‘person running’ sequences - periodic behavior



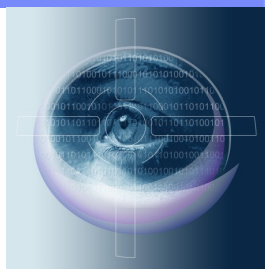


frame-to-frame
similarity matrix





motion-to-motion
similarity matrix



Action Recognition – Template-Based

Recognizing Action at a Distance [Efros et al, ICCV'03]

- Classification is done for each frame. The spatial-temporal descriptor centered at the current frame is matched against the database of actions (previously stored spatial-temporal descriptors).

- For each frame of the probe sequence, the maximum score in the corresponding row of the motion-to-motion similarity matrix (between probe and one sequence of the database) will indicate the best match to the spatial-temporal descriptor centered at this frame.

- K-nearest neighbors is used to determine the action.

- Good results were demonstrated in sequences related to tennis, soccer, and dancing.

Action Recognition – Template-Based

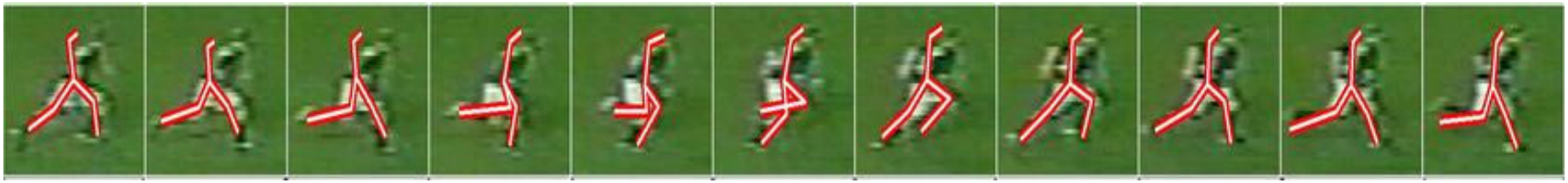
Recognizing Action at a Distance [Efros et al, ICCV'03]

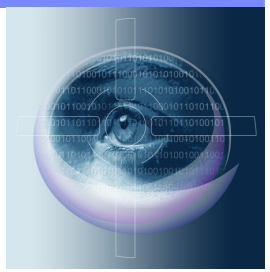
2D Skeleton Transfer

- The database is annotated with 2D joint positions
- After matching, data is transfered to novel sequence

Input sequence:

Transferred 2D skeletons:





Action Recognition – Template-Based

Recognizing Action at a Distance [Efros et al, ICCV'03]

Actor Replacement

Show Video GregWordCup.avi

http://graphics.cs.cmu.edu/people/efros/research/action/

Action Recognition – Template-Based

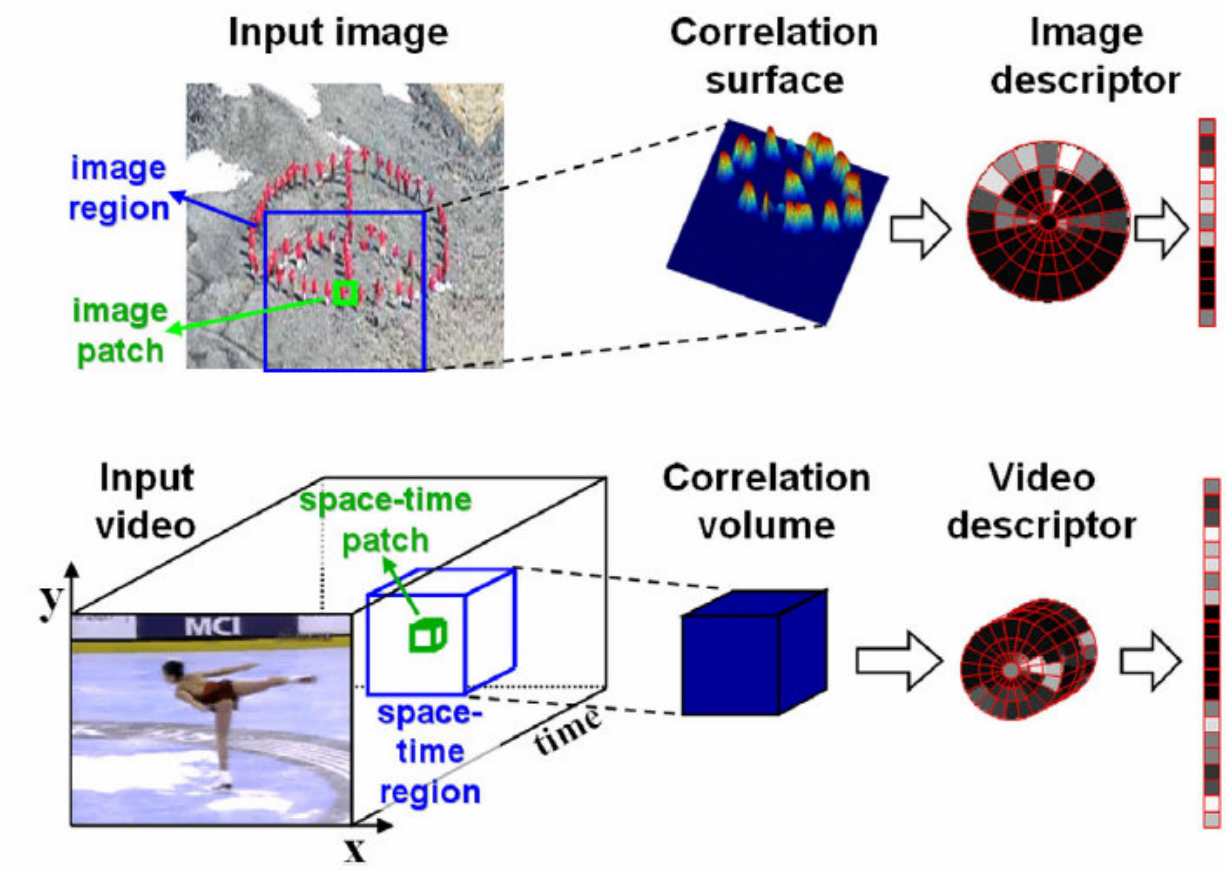
Local Self-Similarities [Shechtman and Irani, CVPR'07]

- Proposed for image similarity. Action detection is a particular application

How to measure similarity in these images?

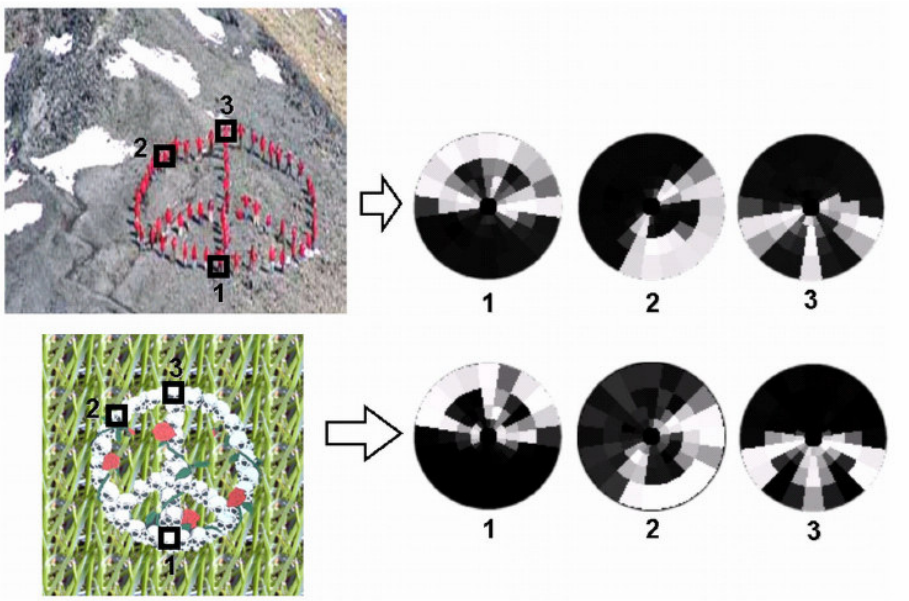
Action Recognition – Template-Based

Local Self-Similarities [Shechtman and Irani, CVPR'07]



Action Recognition – Template-Based

Local Self-Similarities [Shechtman and Irani, CVPR'07]

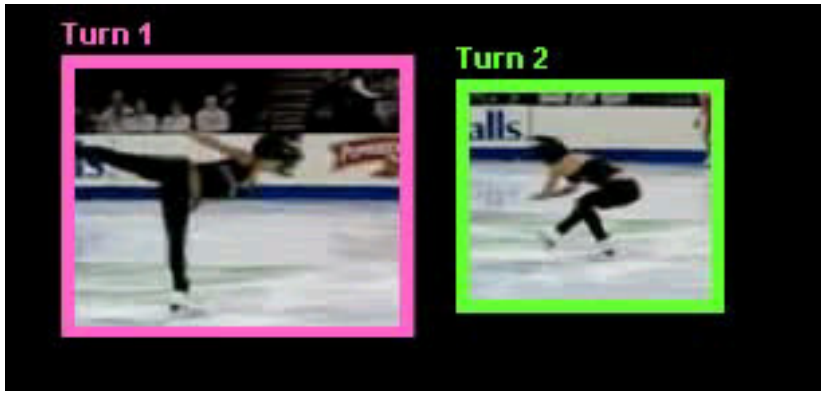


The descriptor implicitly handles the similarity between people wearing different clothes. Also, the spatial-temporal *log-polar binning* allows for better matching under different action durations / frame rate.

Action Recognition – Template-Based

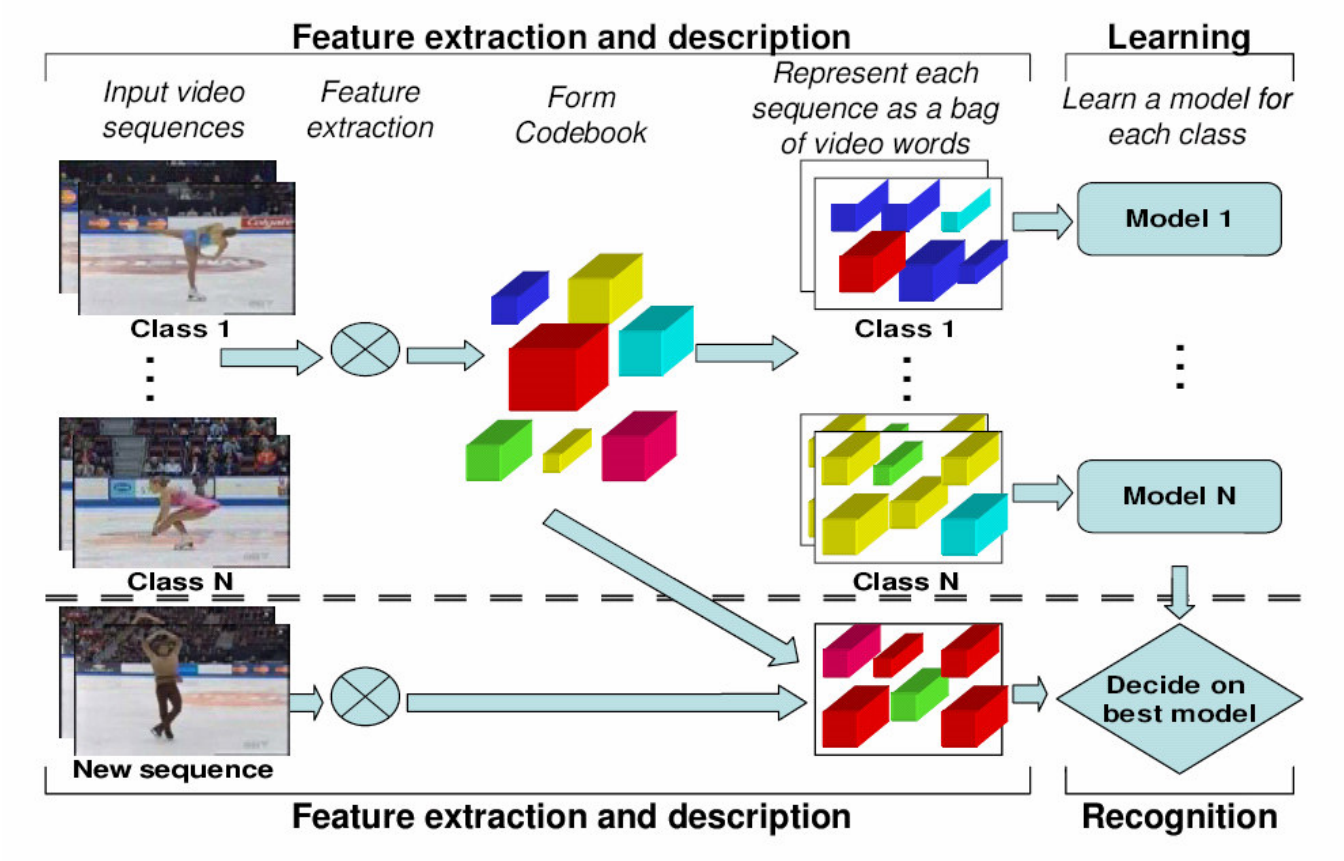
Local Self-Similarities [Shechtman and Irani, CVPR'07]

➢ Complex actions performed by different people wearing different clothes with different backgrounds, are detected with no prior learning, based on a single example clip.

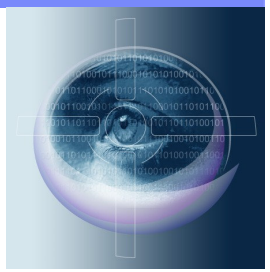


Action Recognition – Template-Based

Spatial-Temporal Bag of Words [Niebles et al, CVPR'06]



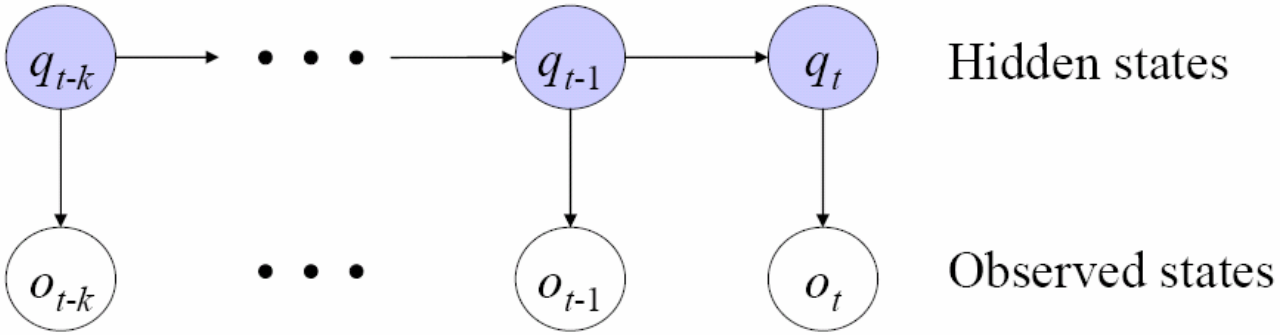
Outline



- Motivation
- Action Recognition
  - Template-Based Approaches
  - **State-Space Approaches**
- Detecting Suspicious Behavior

Action Recognition – State-Space

Hidden Markov Models [Rabiner, 1989]

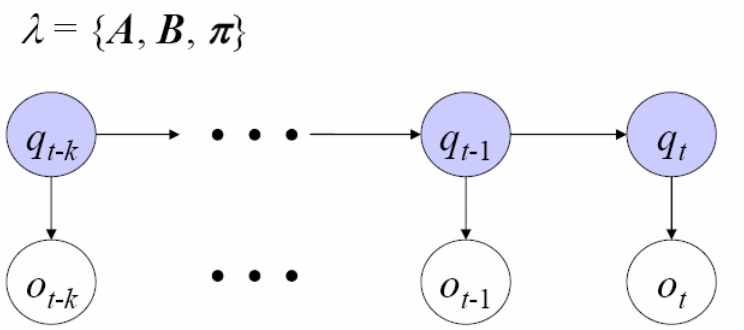


- Graphical Model
- Circles indicate states
- Arrows indicate probabilistic dependencies between states

Action Recognition – State-Space

Hidden Markov Models [Rabiner, 1989]

1. A set of states $S = \{s_1, s_2, ..., s_N\}$ of the model, where N is the number of states of the model. We will denote the state at time t as q_t.
2. A set of distinct observation symbols $V = \{v_1, v_2, ..., v_M\}$, where M is the number of observation symbols of the model.
3. A set of state transition probabilities $A = \{a_{ij}\}$, where $a_{ij} = P[q_{t+1} = s_j | q_t = s_i], 1 \leq i, j \leq N$;
4. A set of observation symbol probability distribution on state j, $B = \{b_j(k)\}$, where $b_j(k) = P[v_k | q_t = s_j]$, $1 \leq j \leq N$ and $1 \leq k \leq M$;
5. The initial state distribution $\pi = \pi_i$ where $\pi_i = P[q_1 = S_i], 1 \leq i \leq N$.

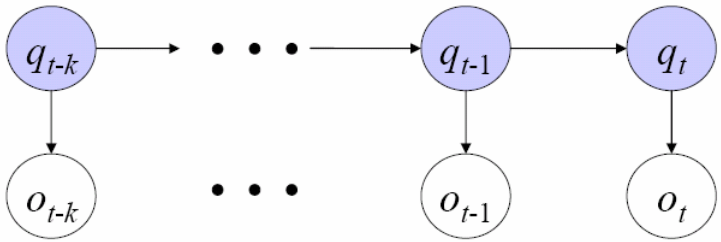


Action Recognition – State-Space

Hidden Markov Models [Rabiner, 1989]

Three Basic Problems:

$\lambda = \{A, B, \pi\}$



1. Given the observation sequence $O = o_1 o_2 \ldots o_T$ and the model λ, how to efficiently compute $P(O|\lambda)$.

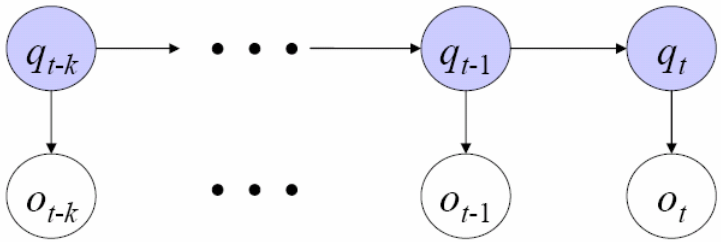
Forward-Backward Algorithm

Action Recognition – State-Space

Hidden Markov Models [Rabiner, 1989]

Three Basic Problems:

$\lambda = \{A, B, \pi\}$



2. Given the observation sequence $O = o_1 o_2 \ldots o_T$ and the model λ, how to compute the hidden state sequence $Q = q_1 q_2 \ldots q_T$ which best "explains" the observations (i.e., compute the most likely hidden state sequence)

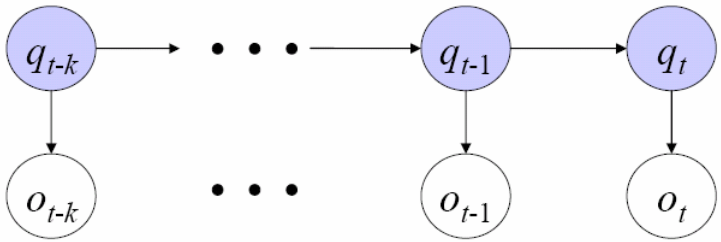
Viterbi Algorithm

Action Recognition – State-Space

Hidden Markov Models [Rabiner, 1989]

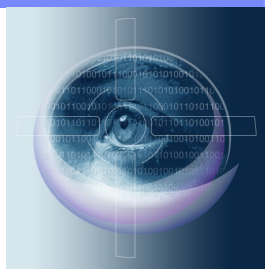
Three Basic Problems:

$\lambda = \{A, B, \pi\}$



3. How to adjust the model parameters $\lambda = \{A, B, \pi\}$ to maximize $P(O \mid \lambda)$. (adjust the model to fit the data best?)

Baum-Welch Algorithm



Action Recognition – State-Space

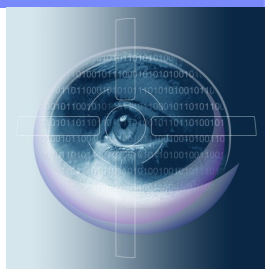
Hidden Markov Models [Rabiner, 1989]

Action Recognizer:

- Learn an HMM model for each action in the database (e.g., HMM for ‘running’, HMM for ‘fighting’, etc.) – Baum-Welch algorithm

- Given an action sequence, compare it with all HMMs in the database and select the one which best explains the probe sequence – Forward-Backward algorithm

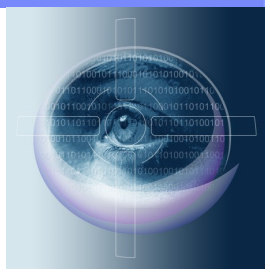
Action Recognition – State-Space



- [Yamato et al, 1992] - First application of HMMs for gesture recognition (for recognizing tennis strokes)

- From there on HMMs have been extensively applied in many gesture recognition problems (Sign Language Recognition, Head Gesture, etc.)

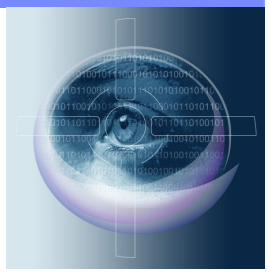
- Many variations have been proposed (see e.g., coupled HMMs). More recently, **Conditional Random Fields (CRFs)** have proven to be very successful to model human motion [Sminchisescu et al, ICCV 2005]



Action Recognition – State-Space

Modeling Interactions with Stochastic Grammars
[Ivanov and Bobick, 2000]

- Recognize actions with larger temporal range
- Two-Stage Approach:
  - Detection of low-level discrete events (e.g., using HMMs or tracking)
  - Action Recognition using Stochastic Grammars



Action Recognition – State-Space

Modeling Interactions with Stochastic Grammars
[Ivanov and Bobick, 2000]

Background: Earley Parsing for Context-free Grammars

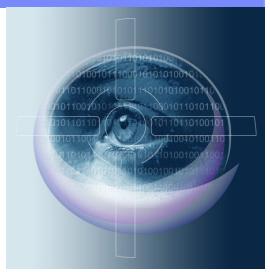
- See description in wikipedia
- Three main steps: **Prediction**, **Scanning**, **Completion**

Earley Parsing Example

S	→	NP VP	Det	→	a
NP	→	Det N	N	→	circle|square|triangle
VP	→	VT NP	VT	→	touches
VP	→	VI PP	VI	→	is
PP	→	P NP	P	→	above|below

	a	circle	touches	a	square
$_0$ → .S	*scanned*	*scanned*	*scanned*	*scanned*	*scanned*
predicted	$_0$Det → a.	$_1$N → circle.	$_2$VT → touches.	$_3$Det → a.	$_4$N → triangle.
$_0$S → .NP VP	*completed*	*completed*	*completed*	*completed*	*completed*
$_0$NP → .Det N	$_0$NP → Det.N	$_0$NP → Det N.	$_2$VP → VT.NP	$_3$NP → Det.N	$_4$NP → Det N.
$_0$Det → .a	*predicted*	$_0$S → NP.VP	*predicted*	*predicted*	$_3$VP → VT NP.
	$_1$N → .circle	*predicted*	$_3$NP → .Det N	$_5$N → .circle	$_0$S → NP VP.
	$_1$N → .square	$_2$VP → .VT NP	$_3$Det → .a	$_4$N → .square	$_0$ → S.
	$_1$N → .triangle	$_2$VP → .VI PP		$_4$N → .triangle	
		$_2$VT → .touches			
		$_2$VI → .is			
State set 0	1	2	3	4	5

Action Recognition – State-Space



Modeling Interactions with Stochastic Grammars
[Ivanov and Bobick, 2000]

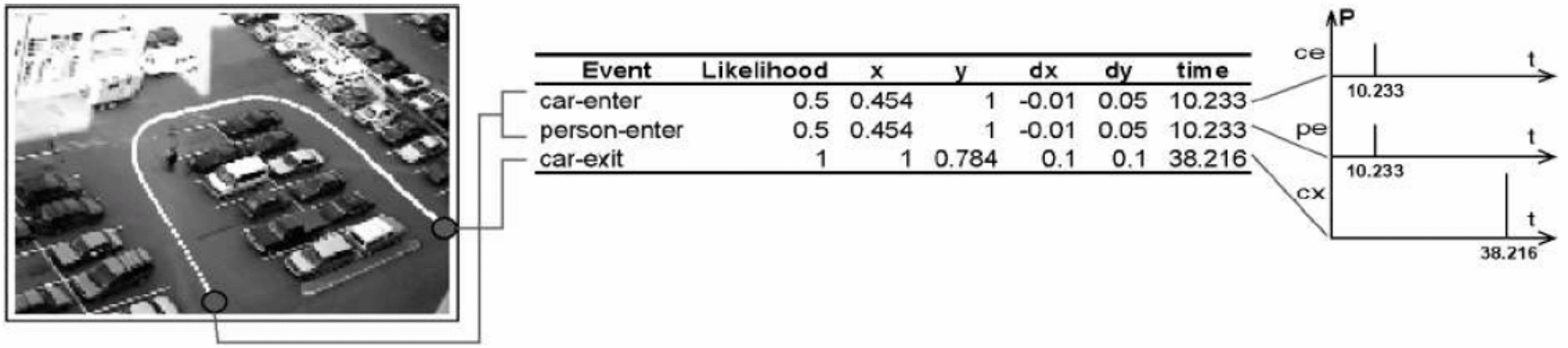
Probabilistic Earley Parsing

- Production rules are augmented with probabilities
- Parse tree with highest probability is generated [Stolcke, Bayesian Learning of Probabilistic Language Models,1994]

Action Recognition – State-Space

Modeling Interactions with Stochastic Grammars
[Ivanov and Bobick, 2000]

Car/Person Interaction

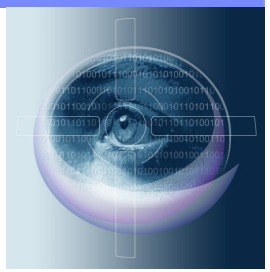


Low-level discrete event detection

- Track moving blobs
- Generate events: {person,car}+{enter,found,exit,lost,stopped}

Modeling Interactions with Stochastic Grammars

[Ivanov and Bobick, 2000]

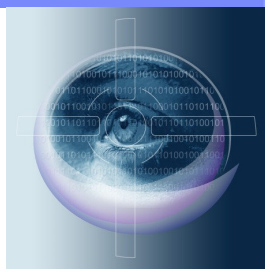


G_p :

TRACK	→	CAR-TRACK	[0.5]
	|	PERSON-TRACK	[0.5]
CAR-TRACK	→	CAR-THROUGH	[0.25]
	|	CAR-PICKUP	[0.25]
	|	CAR-OUT	[0.25]
	|	CAR-DROP	[0.25]
CAR-PICKUP	→	ENTER-CAR-B CAR-STOP PERSON-LOST B-CAR-EXIT	[1.0]
ENTER-CAR-B	→	CAR-ENTER	[0.5]
	|	CAR-ENTER CAR-HIDDEN	[0.5]
CAR-HIDDEN	→	CAR-LOST CAR-FOUND	[0.5]
	|	CAR-LOST CAR-FOUND CAR-HIDDEN	[0.5]
B-CAR-EXIT	→	CAR-EXIT	[0.5]
	|	CAR-HIDDEN CAR-EXIT	[0.5]
CAR-EXIT	→	car-exit	[0.7]
	|	SKIP car-exit	[0.3]
CAR-LOST	→	car-lost	[0.7]
	|	SKIP car-lost	[0.3]
CAR-STOP	→	car-stop	[0.7]
	|	SKIP car-stop	[0.3]
PERSON-LOST	→	person-lost	[0.7]
	|	SKIP person-lost	[0.3]

A CAR-PICKUP branch of a simplified grammar describing interactions in a parking lot.

Outline



- Motivation
- Action Recognition
  - Template-Based Approaches
  - State-Space Approaches
- **Detecting Suspicious Behavior**

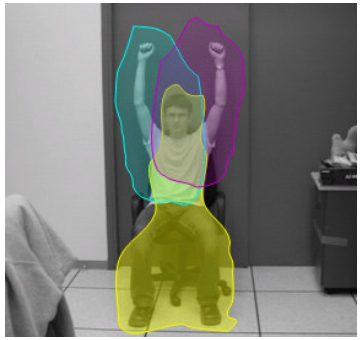
Suspicious Behavior

Detecting Irregularities [Boiman and Irani, ICCV 2005]

- Problem: given a few "regular" examples, compute the likelihood of a new observation

Database

Query



- Construct the likelihood using chuncks of data from the examples. Large matching chunks imply large likelihood.

Suspicious Behavior

Detecting Irregularities [Boiman and Irani, ICCV 2005]

- Problem: given a few "regular" examples, compute the likelihood of a new observation

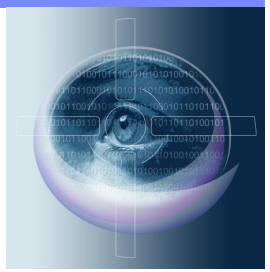
Database

Query

- Construct the likelihood using chuncks of data from the examples. Large matching chunks imply large likelihood.

Suspicious Behavior

Detecting Irregularities [Boiman and Irani, ICCV 2005]



Suspicious Behavior

See Also:

- [Zhong et al, Detecting Unusual Activity in Video, CVPR'04]

Motion Trajectory Behavior:

- [Stauffer and Grimson, Learning patterns of activity using real-time tracking, 2000]

- [Lei Chen et al, Robust and fast similarity search for moving object trajectories, 2005]