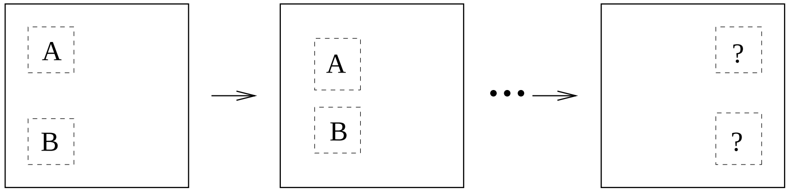
NORMAL TARGET TRACKING

General problem: in first image have R targets $\{F_i\}$
In next image have L targets $\{N_i\}$
How to pair the targets out of the $(R+1)^L$ possibilities?

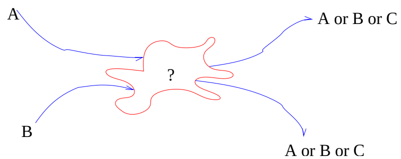


Video rate fast → targets don't move much
Kalman Filter predicts position
Overlap with detected target makes correspondences

MAINTAINING TARGET PERSISTENCE

Issue: tracking targets breaks down when close or occluded

Solution: need identity persistence thru occlusion



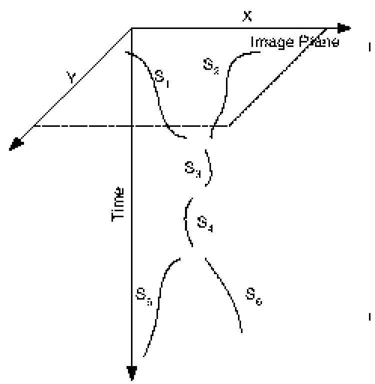
Jorge et al: Bayesian network

PROBLEM REPRESENTATION

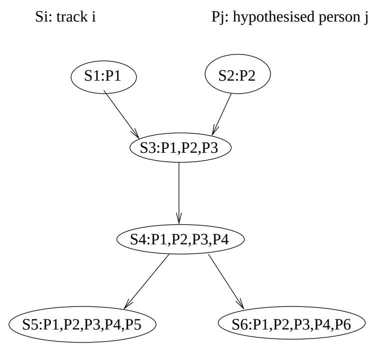
Break trajectories at occlusions

Create label nodes X_i for each track S_i, also inheriting previous tracks

Target tracks



Initial hypothesis Nodes



MATCHING COLOUR TARGETS I

Assume binary mask of target

Use mask to select target pixels $\{(r_i, g_i, b_i)\}$

Compute RGB histogram over all pixels in region and all frames in segment (eg. maybe 20K values):

$$\forall i \; h_1(r_i, g_i, b_i) = h_1(r_i, g_i, b_i) + 1$$

How well does distribution $h_1(r_i, g_i, b_i)$ match distribution $h_2(r_i, g_i, b_i)$?

MATCHING COLOUR TARGETS II

Normalize:

$H_j(r, g, b) = h_j(r, g, b)/\Sigma_{r,g,b} h_j(r, g, b)$

Use Bhattacharyya distance:

$$d(H_1, H_2) = 1 - \sum_{r,g,b} \sqrt{H_1(r, g, b) \times H_2(r, g, b)}$$

If $d(h_1, h_2)$ small then likely same target

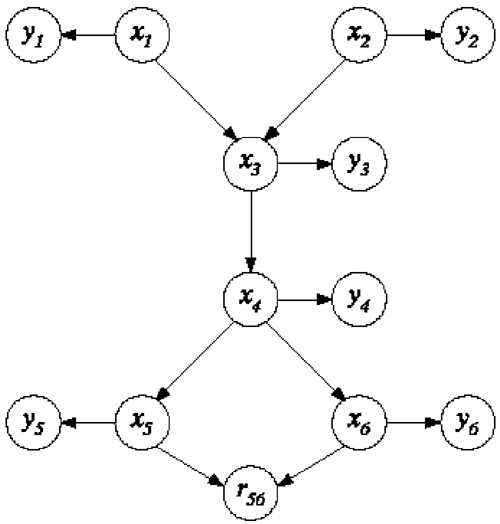
Group colour levels together (eg. 0-31, 32-64, ... 224-255) because so few pixels

EXTENDING NETWORK

Create label nodes X_i for each segment section S_i

Add data matching nodes Y_i (color histogram of target appearance)

Add restriction nodes R_i enforcing mutual exclusion between sibling nodes



EVALUATING PERSISTENCE

Find labeling $\vec{X}$ that maximizes

$$p(\vec{X} \mid \vec{Y}, \vec{R})$$

Probability of labeling $\vec{X}$ given data $\vec{Y}$ and restrictions $\vec{R}$

Gives probability that each person P_i is observed in track X_j

Use standard conditional probability propagation algorithm

PROBLEM 1: REAL-TIME ANSWERS

Full network evaluation is expensive

Incremental evaluation, using Bayes rule after the k^{th} block of T frames:

$$p(x_i \mid \vec{Y}_0^t, \vec{R}_0^t) = \alpha p(\vec{Y}_{kT}^t, \vec{R}_{kT}^t \mid x_i) p(x_i \mid \vec{Y}_0^{kT}, \vec{R}_0^{kT})$$

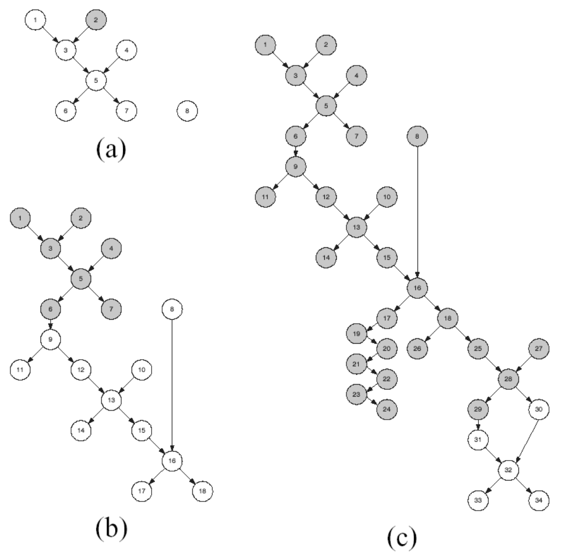
PROBLEM 2: GROWTH OF NETWORK

Each new track inherits labels plus adds a new one: network grows large

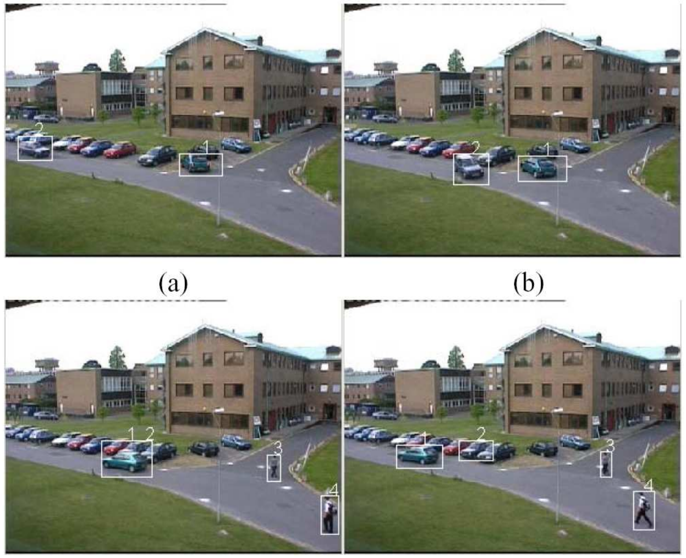
Solution: freeze all but N most recent nodes

Freeze: fix most probable identity for track S_i rather than keep all possible ids (which could change with future evidence)

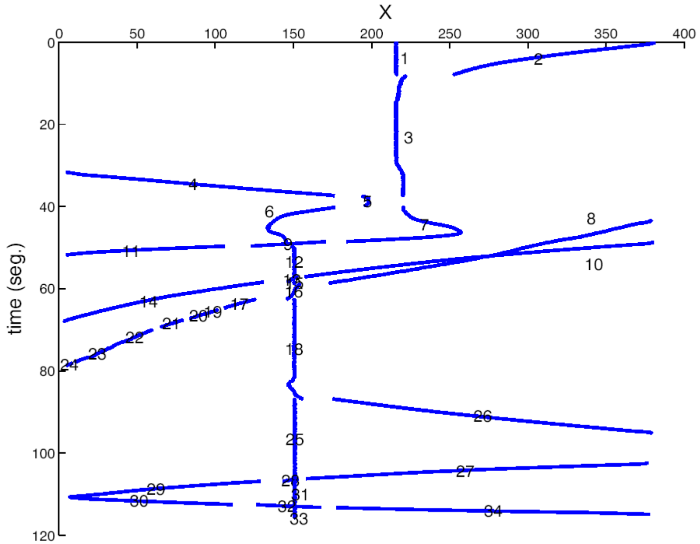
ACTIVE NODES AT 3 TIMES



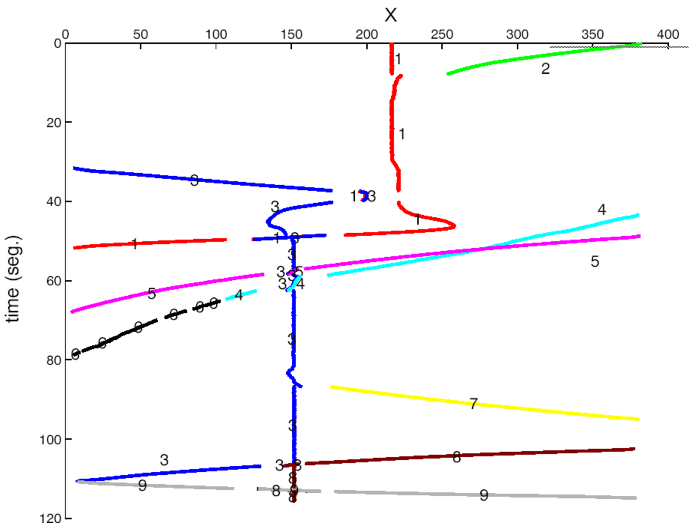
TARGET 1 & 2 PERSISTENCE



TRACKS BEFORE PERSISTENCE



TRACKS AFTER PERSISTENCE



WHAT WE HAVE LEARNED

Probabilistic approach to maintaining persistence across occlusions

Future: target separation when in close proximity

Lecture Problem

These Bayesian network graphs do not have any loops. How could we handle a person walking in a loop, meeting the same person several times?

SHORT TERM ACTION RECOGNITION

Action primitives, not sequences, eg. a hand wave

Temporally local/short-term image analysis/instantaneous

Appearance based/viewpoint specific

Efros, Berg, Mori & Malik

PROBLEM

Near field (bodies 300 pixels high)
→ Use geometric model

Far field (bodies 3 pixels high)
→ Use tracking

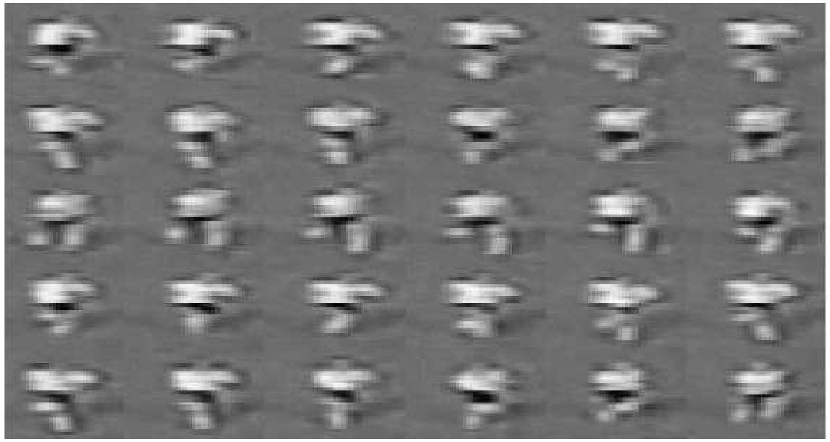
Here: medium field (bodies 30 pixels high)

GOAL

Select matching activities from database of video clips

KEY CONCEPT

Pattern of stabilized optical flow



PROCESS OVERVIEW

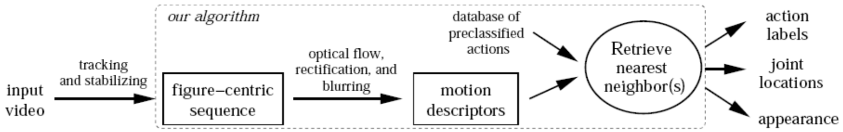


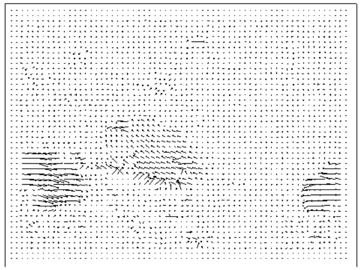
IMAGE STABILIZATION

Threshold temporal difference for regions of interest

Normalized cross-correlation inside region of interest for localization

OPTICAL FLOW

Image velocity (u, v) at every pixel: where each pixel's data is moving to in next image

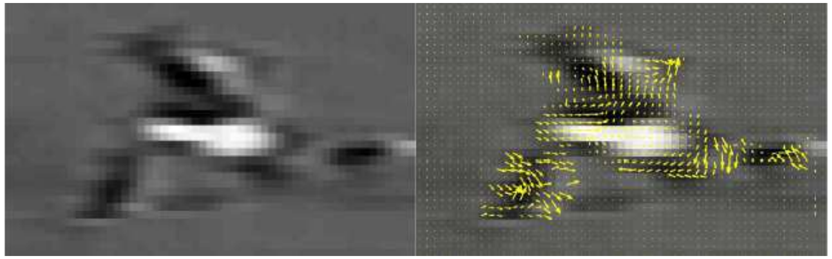


Computed by standard algorithms that match local gradients to temporal gradients

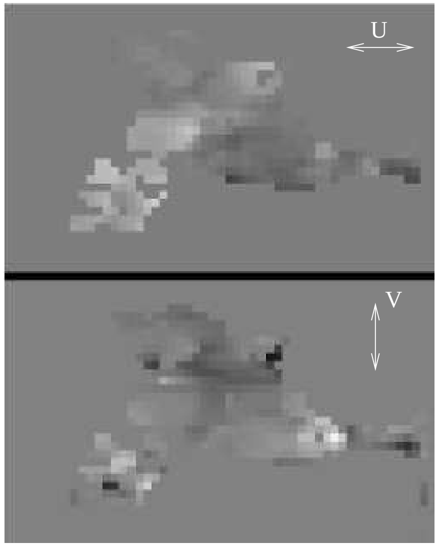
OPTICAL FLOW DESCRIPTORS 1

Goal: aggregate spatial pattern of noisy relative O.F.

O.F. IMAGE = $[...(u_i, v_i)...]$



OPTICAL FLOW DESCRIPTORS 2



u, v components

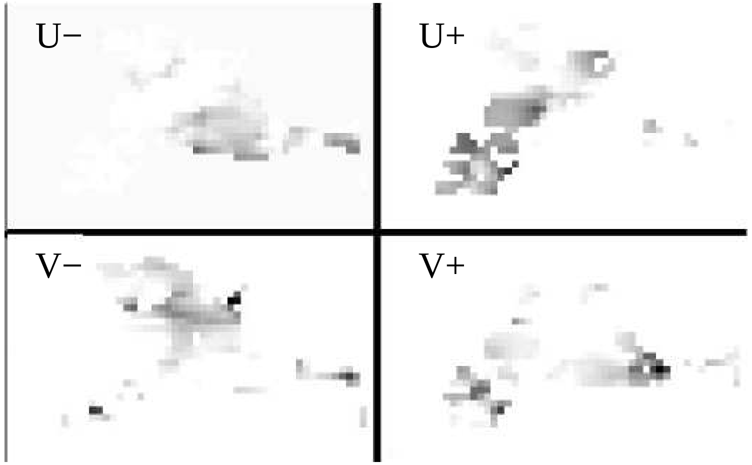
OPTICAL FLOW DESCRIPTORS 3

Noisy so smooth, but smoothing cancels +/- aspects

Solution: split +/- components

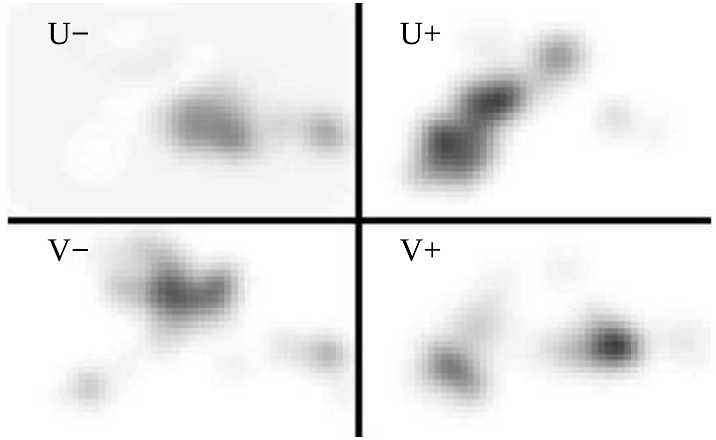
$f(x) = x$ if $x \geq 0$ else $x = 0$

$(u_i, v_i) \rightarrow (f(u_i), f(-u_i), f(v_i), f(-v_i))$



OPTICAL FLOW DESCRIPTORS 4

Convolve with gaussian to smooth



Descriptor: sequence (50 frames) of smoothed O.F. windows

MATCHING DESCRIPTORS I

Single frame matching

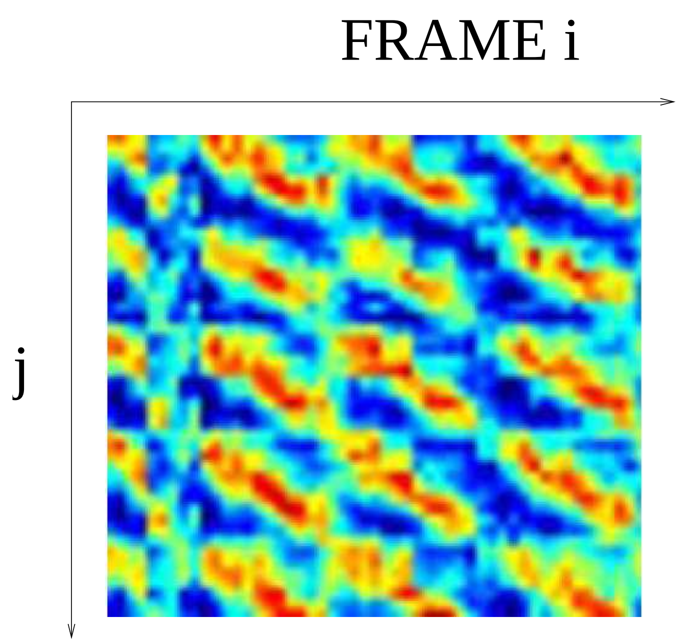
$$m(i,j) = \sum_{c=1}^{4} \sum_{x,y \in I} a_c^i(x,y) b_c^j(x,y)$$

Where

Frame i of seq. a, frame j of seq. b

c = 1,2,3,4 optical flow components

(x,y) = pixel positions

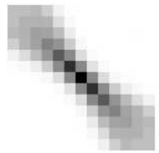


MATCHING DESCRIPTORS II

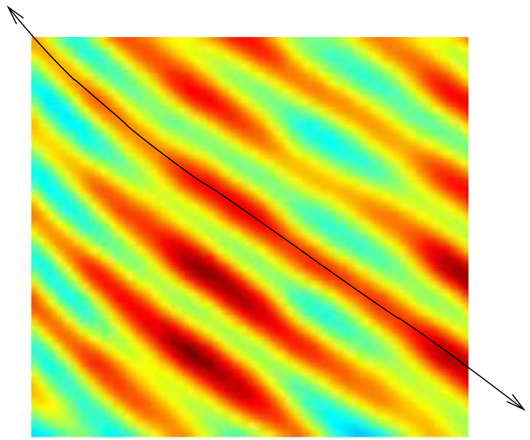
Time window of $T = 50$ frames

$$S(i,j) = \sum_{r=-T/2}^{r=+T/2} \sum_{s=-T/2}^{s=+T/2} K(r,s)m(i+r, j+s)$$

Weighted sum of nearby in time frames



K

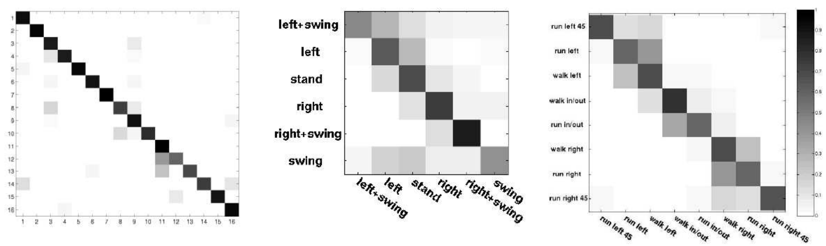


$S(i,j)$

Near diagonal is matching frames
Stripes from action periodicity

EXAMPLE MATCHING SEQUENCES

CONFUSION MATRICES



Ballet Tennis Football

WHAT WE HAVE LEARNED

1. Short term action recognition technique
2. Based on stabilized optical flow of local medium sized windows
3. Encodes temporal structure better
4. But: still somewhat viewpoint and scale dependent

Lecture Problem

Why must we stabilize the image before computing the optical flow?