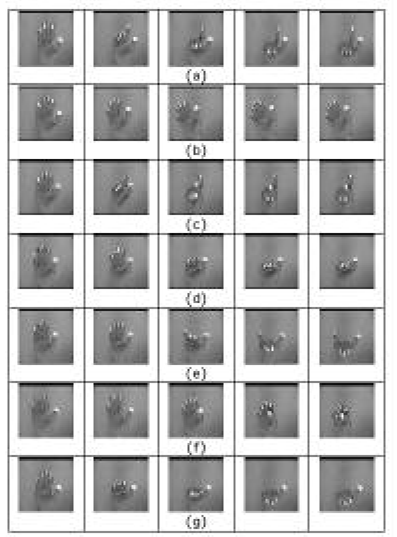
Lecture-19

Hand Gesture Recognition

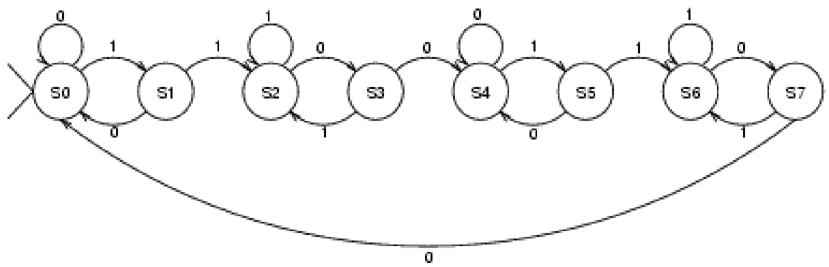
Seven Gestures



Gesture Phases

- Hand fixed in the start position.
- Fingers or hand move smoothly to gesture position.
- Hand fixed in gesture position.
- Fingers or hand return smoothly to start position.

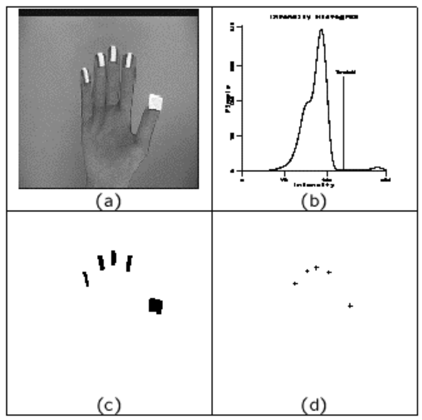
Finite State Machine



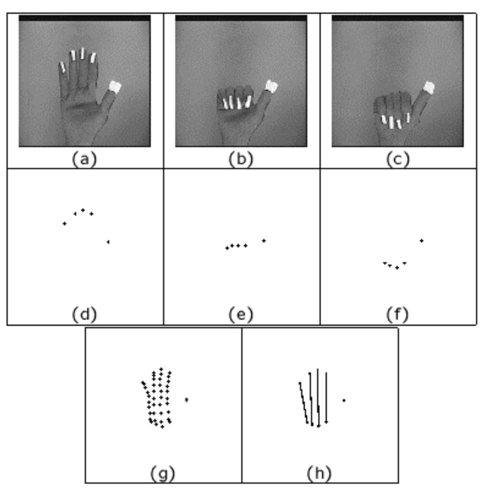
Main Steps

- Detect fingertips.
- Create fingertip trajectories using motion correspondence of fingertip points.
- Fit vectors and assign motion code to unknown gesture.
- Match

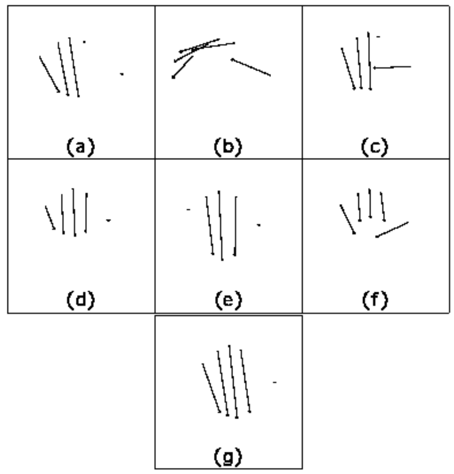
Detecting Fingertips



Vector Extraction



Vector Representation of Gestures



Results

Results

Run	Frames	L	R	U	D	T	G	S
1	200	√	√	√	√	√	√	√
2	250	√	√	√	√	√	√	√
3	250	√	√	√	X	√	√	√
4	250	√	√	√	√	√	√	√
5	300	√	√	√	√	√	√	√
6	300	√	√	√	√	√	√	√
7	300	√	√	√	√	√	√	√
8	300	√	√	√	√	√	√	√
9	300	√	√	√	√	*	*	*
10	300	√	√	√	√	√	√	√

L = Left, R = Right, U = Up, D = Down, T = Rotate, G = Grab, S = Stop, √ - Recognized, X - Not Recognized, * - Error in Sequence.

Action Recognition Using Temporal Templates

Jim Davis and Aaron Bobick

Main Points

- Compute a sequence of difference pictures from a sequence of images.
- Compute Motion Energy Images (MEI) and Motion History Images (MHI) from difference pictures.
- Compute Hu moments of MEI and MHI.
- Perform recognition using Hu moments.

MEI and MHI

Motion-Energy Images (MEI)

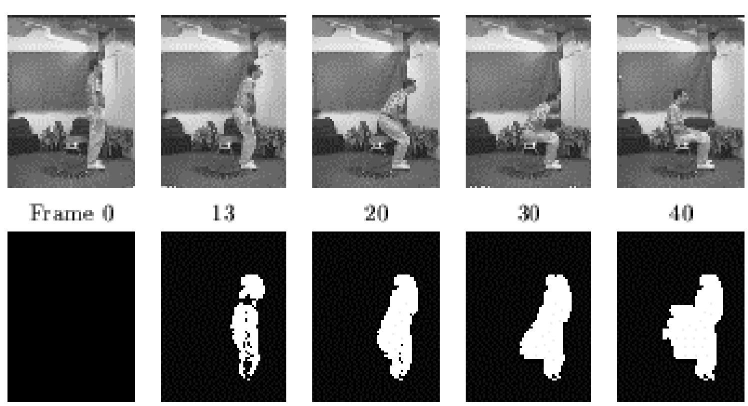
$$E_t(x, y, t) = \bigcup_{i=0}^{t-1} D(x, y, t-i)$$

Change Detected Images

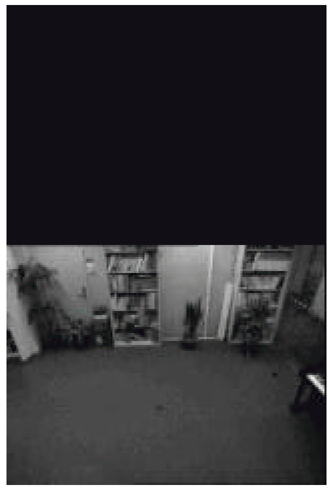
Motion History Images (MHI)

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

MEIs



Color MHI Demo



Summary

- Use seven Hu moments of MHI and MEI to recognize different exercises.
- Use seven views (-90 degrees to +90 degrees in increments of 30 degrees).
- For each exercise several samples are recorded using all seven views, and the mean and covariance matrices for the seven moments are computed as a model.
- During recognition, for an unknown exercise all seven moments are computed, and compared with all 18 exercises using Mahalanobis distance.
- The exercise with minimum distance is computed as the match.
- They present recognition results with one and two view sequences, as compared to seven view sequences used for model generation.

Moments

Binary image

General Moments

$$m_{pq} = \int \int x^p y^q r(x, y) dxdy$$

Central Moments (Translation Invariant)

$$m_{pq} = \int\int (x - \overline{x})^p (y - \overline{y})^q \, r(x, y) \; d(x - \overline{x}) d(y - \overline{y})$$

$$\overline{x} = \frac{m_{10}}{m_{00}}, \overline{y} = \frac{m_{01}}{m_{00}}$$

centroid

Central Moments

$$m_{00} = m_{00} \equiv m$$
$$m_{01} = 0$$
$$m_{10} = 0$$
$$m_{20} = m_{20} - m\overline{x}^2$$
$$m_{11} = m_{11} - m\overline{xy}$$
$$m_{02} = m_{02} - m\overline{y}^2$$
$$m_{30} = m_{30} - 3m_{20}\overline{x} + 2m\overline{x}^3$$
$$m_{21} = m_{21} - m_{20}\overline{y} - 2m_{11}\overline{x} + 2m\overline{x}^2 y$$
$$m_{12} = m_{12} - m_{02}\overline{x} - 2m_{11}\overline{y} + 2m\overline{x}y^2$$
$$m_{03} = m_{03} - 3m_{02}\overline{y} + 2m\overline{y}^3$$

Moments

Hu Moments: translation, scaling and rotation invariant

$$u_1 = m_{20} + m_{02}$$

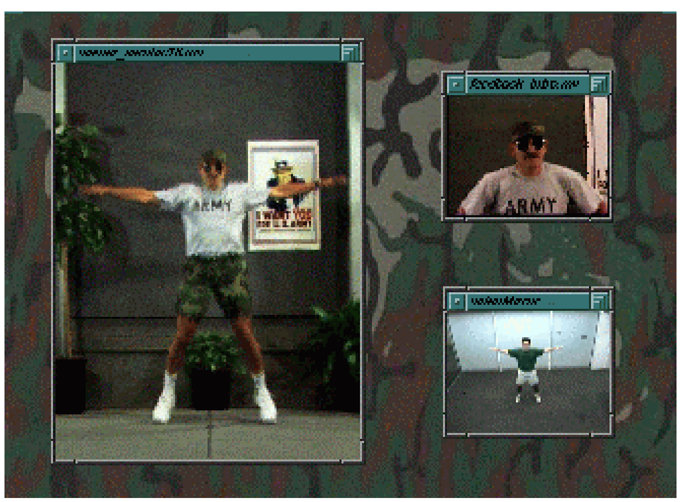
$$u2 = (m_{20} - m_{02})^2 + m_{11}{}^2$$

$$u_3 = (m_{30} - 3m_{12})^2 + (3m_{12} - m_{03})^2$$

$$u_4 = (m_{30} + m_{12})^2 + (m_{21} + m_{03})^2$$

$$\vdots$$

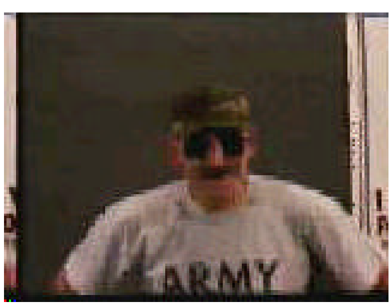
PAT (Personal Aerobic Trainer)

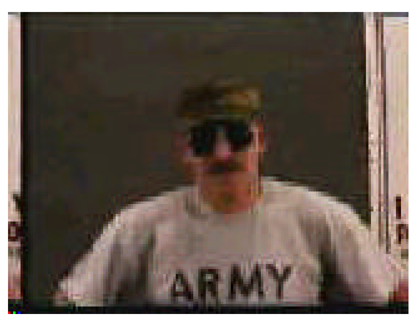


PAT (Personal Aerobic Trainer)

http://vismod.www.media.mit.edu/vismod/demos/actions/mhi_generation.mov

PAT (Personal Aerobic Trainer)

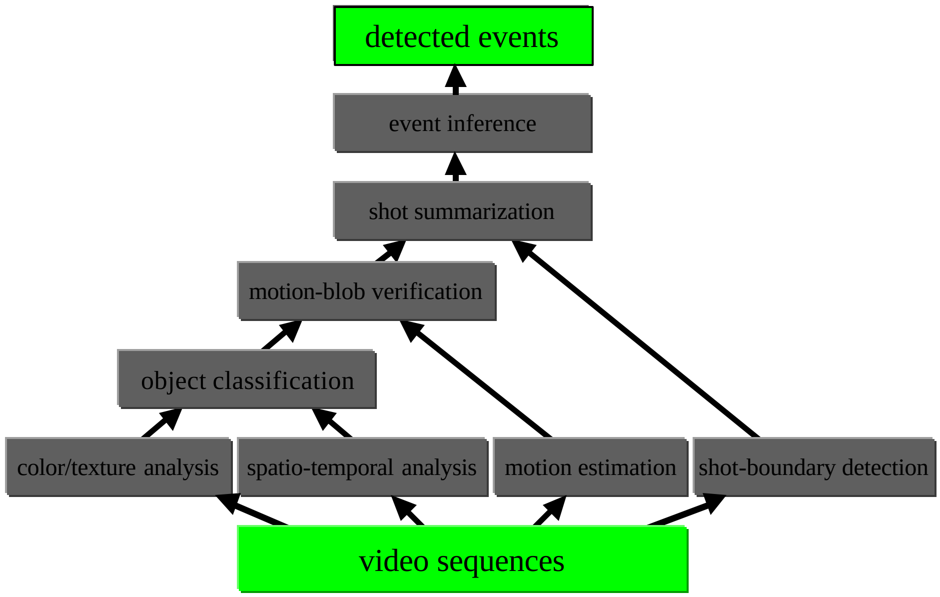




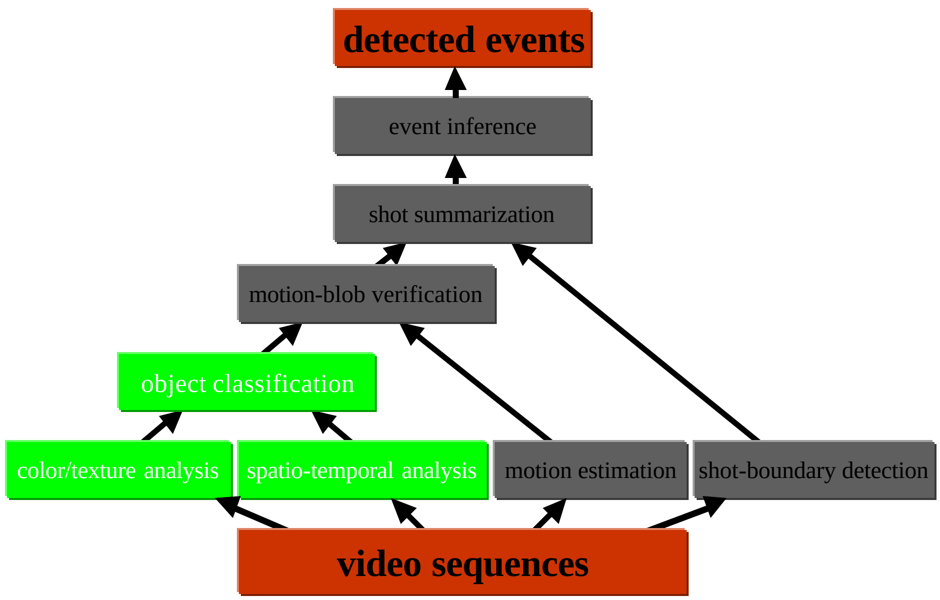
A Framework for the Design of Visual Event Detectors

Niels Haering

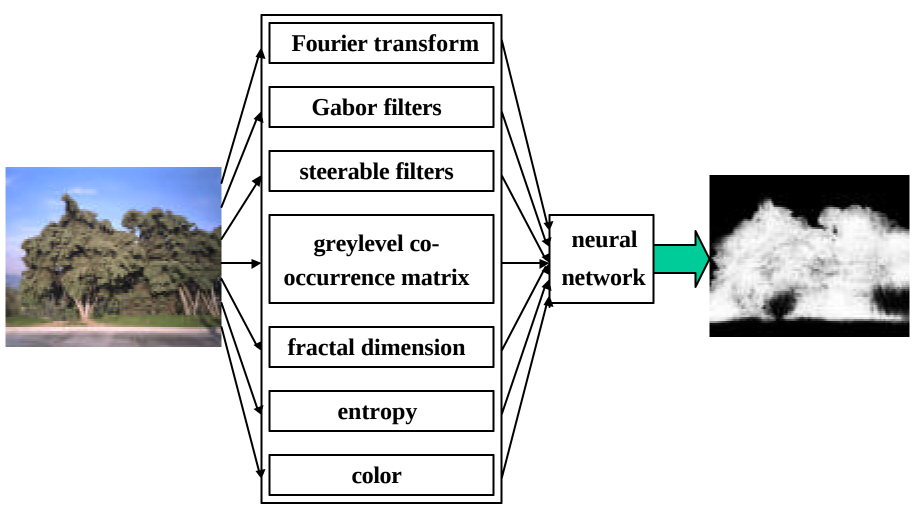
Our Framework



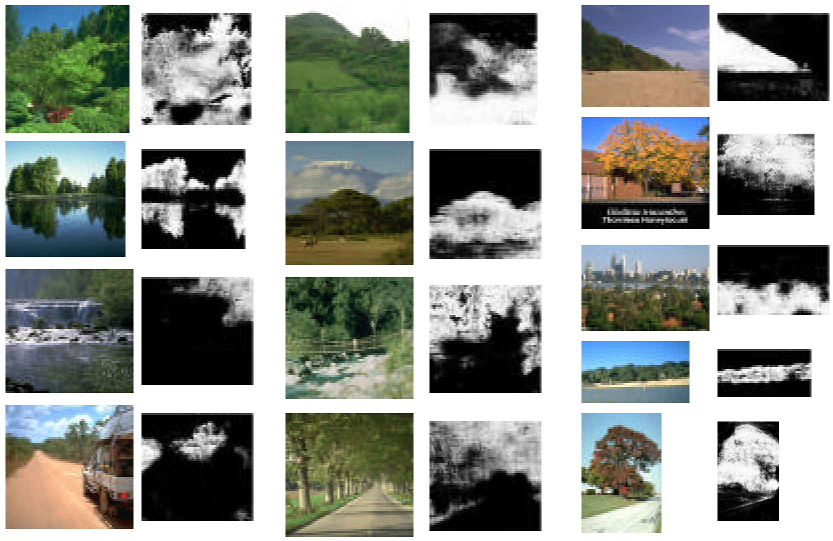
Object Classification



Object Classification

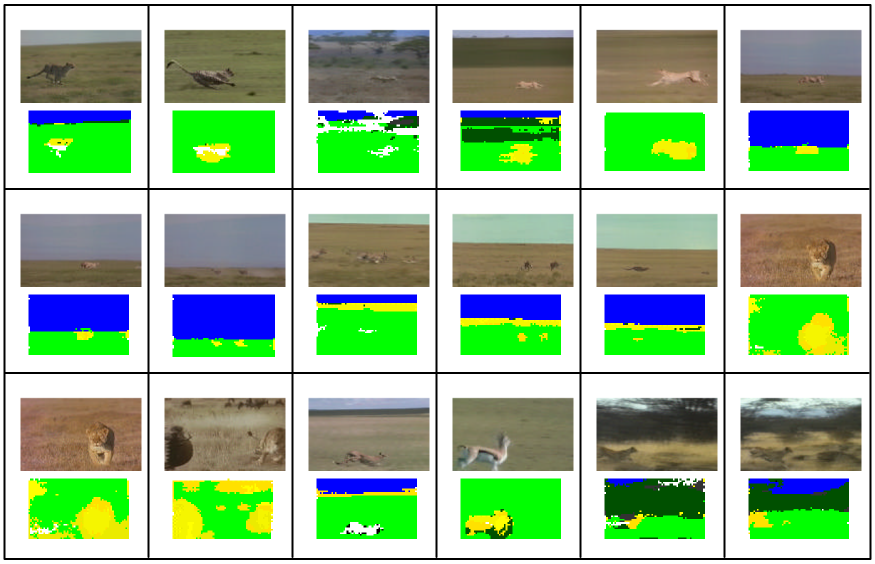


Deciduous Trees

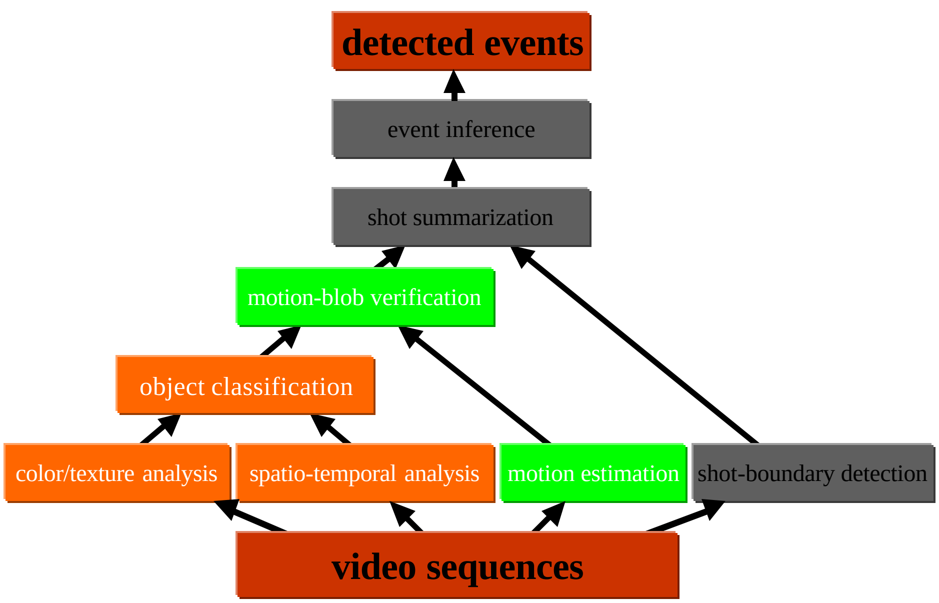


Sky

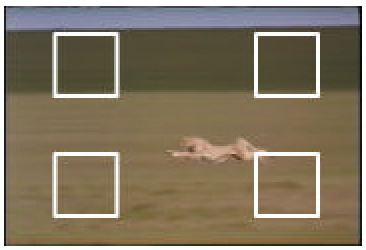
Animals, Sky, Grass, Trees, Rock

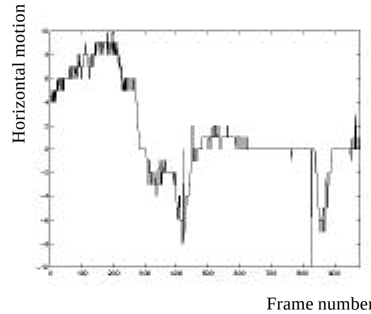


Motion-blob Verification



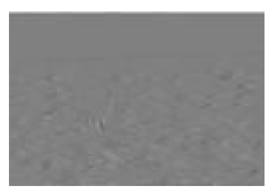
Motion Estimation



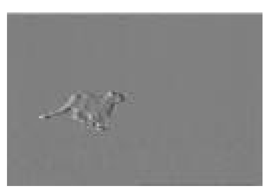


- three parameter system: x-, y-translation, and zoom,
- 4 motion estimates based on pyramid,
- 4 motion estimates based on previous best match,
- “texture” measure prevents ambiguous matches

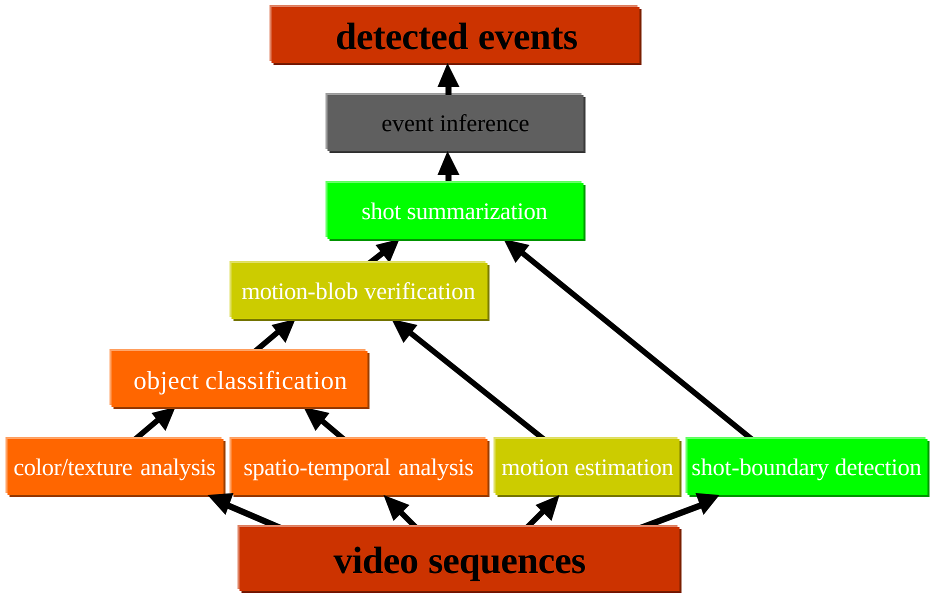
Motion-blob detection



Motion estimate
Äx = -7
Äy = 0
zoom = 1.0



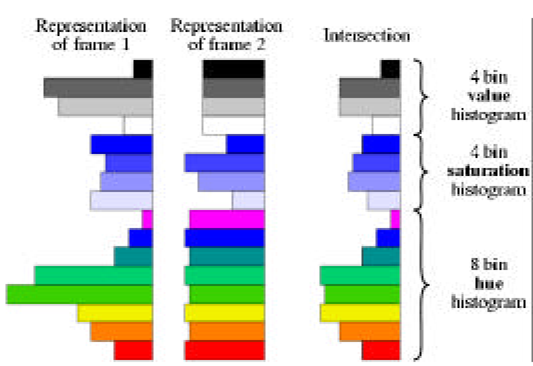
Shot Summarization



Shot Detection

Characteristics of shot boundaries:

- Change of camera/viewpoint
- Change of color characteristics



4 Bins for Value
4 Bins for Saturation
8 bins for hue

= 0.79

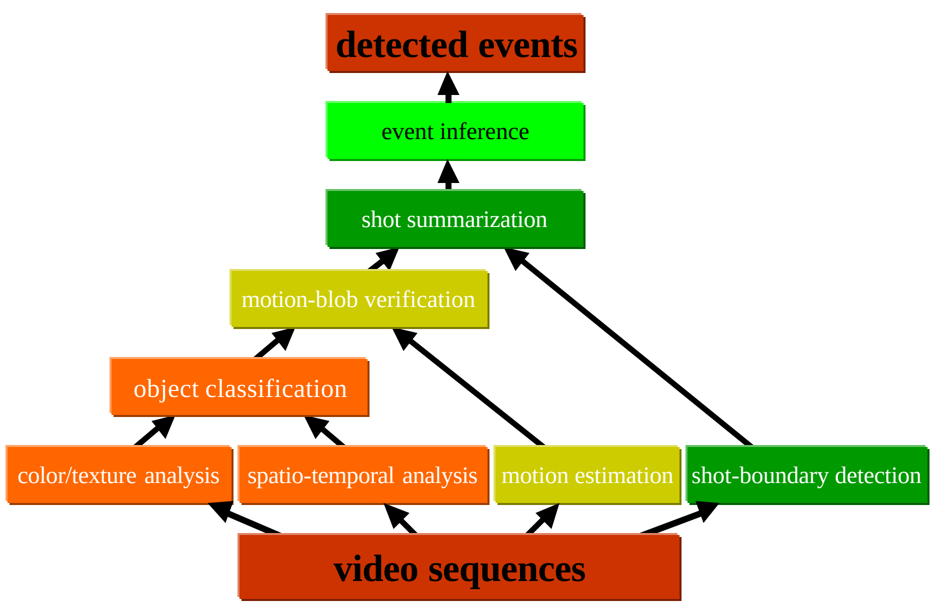
$$\cap = \frac{\sum_{n=0}^{15} \min(f_1(n), f_2(n))}{\min(\sum_{n=0}^{15}(f_1(n), \sum_{n=0}^{15} f_2(n))}$$

Shot Summaries

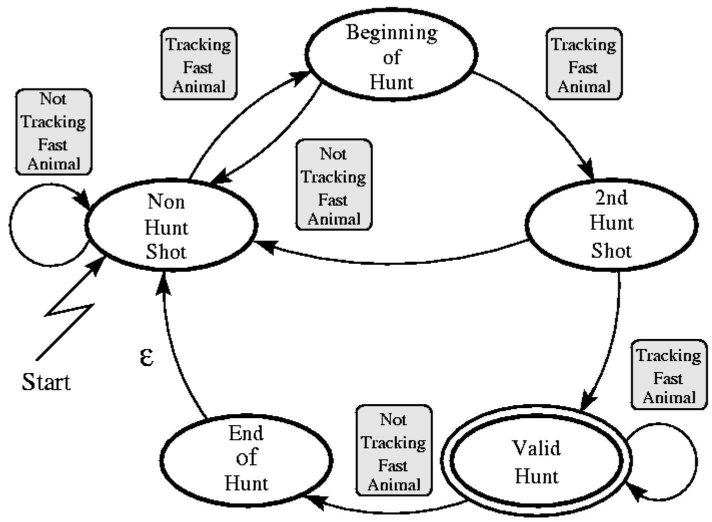
General Information	
Forced/Real Shot Summary	0
First Frame of Shot	64
Last Frame of Shot	263
Global motion estimate (x,y)	(-4.48, 0.01)
Within Frame animal motion estimate (x,y)	(-0.17, 0.23)
Initial Position (x,y)	(175, 157)
Final Position (x,y)	(147, 176)
Initial Size (x,y)	(92, 67)
Final Size (x,y)	(100, 67)
Motion smoothness throughout Shot (x,y)	(0.83, 0.75)
Precision throughout Shot	0.84
Recall throughout Shot	0.16

Hunt Information	
Tracking	1
Fast	1
Animal	1
Beginning of Hunt	1
Number of Hunt Shots	1
End of Hunt	0
Valid Hunt	0

Event Inference



Hunt events



Hunts

Hunt

Non-hunt

Hunts

Non-hunt

Hunt

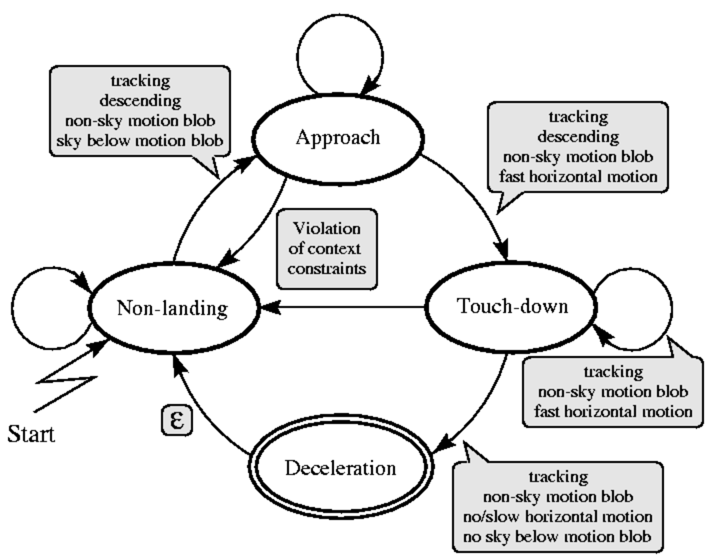
Non-hunt

Event Detection

Sequence Name	Actual Hunt Frames	Detected Hunt Frames	Precision	Recall
Hunt1	305 – 1375	305 – 1375	100%	100%
Hunt2	2472 – 2696	2472 – 2695	100%	99.6%
Hunt3	3178 – 3893	3178 – 3856	100%	94.8%
Hunt4	6363 – 7106	6363 – 7082	100%	96.8%
Hunt5	9694 – 10303	9694 – 10302	100%	99.8%
Hunt6	12763 – 14178	12463 – 13389	67.7%	44.2%
Hunt7	16581 – 17293	16816 – 17298	99.0%	67.0%
average			95.3%	86.0%

Landing Events

Landing Events



Landing Events

Non-landing

Approach

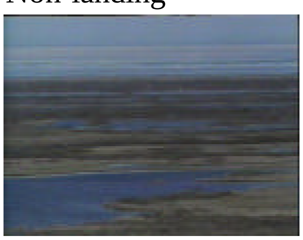
Touch-down

Deceleration

Non-landing

Landing Events

Non-landing



Approach

Touch-down

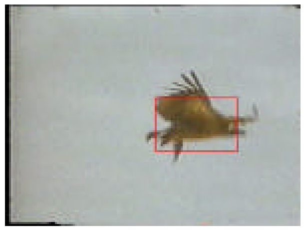
Deceleration

Non-landing

Landing Events

Non-landing

Approach



Touch-down

Deceleration

Non-landing