Video Segmentation

Wednesday 9 Nov 2006

overview

- scene change detection
  spatial-temporal change detection
  motion segmentation (optical flow)
  clustering in motion parameter space
  (k-means test)
  semantic video object segmenation
  chroma-keying

scene change detection

- frame difference between k-th frame and reference frame at pixel location **x**:

$$FD_{k,r}[\boldsymbol{x}] = I_k[\boldsymbol{x}] - I_r[\boldsymbol{x}]$$

Thresholded by T, segmentation label on each pixel

$$z_{k,r}[\boldsymbol{x}] = \begin{cases} 1 \text{ if } |FD_{k,r}[\boldsymbol{x}]| > T \\ 0 \text{ otherwise} \end{cases}$$

Problems:

- a uniform intensity region may be interpreted as stationary
- FD is affected by spatial gradient in the direction of motion

Gaussian pyramid

- Multi-resolution representation of image
  1, Original (highest resolution) image at bottom level
  2. Lowpass filter (e.g. Gaussian filter)
  3. Subsample by factor 2
  4. Place result in second level

Change Detection v. 0.2

- 1. Gaussian pyramid, start at lowest resolution.

  2. Compute at each pixel, normalized frame difference:

$$FDN_{k,r}[\boldsymbol{x}] = \frac{\sum_{x \in \mathcal{N}} |I_k[\boldsymbol{x}] - I_r[\boldsymbol{x}]||\nabla I_r[\boldsymbol{x}]|}{\sum_{x \in \mathcal{N}} |\nabla I_r[\boldsymbol{x}]|^2 + c}$$

  where N is a local neighborhood of x,
  gradient of image, c is fudge addend to avoid divde by 0.

  3. If FDN is high (pixel is moving), then replace FDN from previous level with this one, else retain lower res value.

  4. Repeat 2-3 for all resolution levels.

temporal integration 1

- Warp map W[A,B]: warp image A toward B using motion model parameters estimated between A and B.

  Compute internal representation image:

(*) $$\bar{I}_k[\boldsymbol{x}] = (1-\alpha) I_k[\boldsymbol{x}] + \alpha W[\bar{I}_{k-1}[\boldsymbol{x}], I_k[\boldsymbol{x}]] \qquad 0 \leq \alpha \leq 1$$

  Result: unchanged regions retain sharpness (less noise), changed regions blur

temporal integration 2

- 1. Compute motion parameters between internal representation $\bar{I}_k[\boldsymbol{x}]$ and new frame $I_k[\boldsymbol{x}]$ within support M_{k-1} of dominant object in previous frame.

  2. Warp internal representation image at frame k-1 towards new frame.

  3. Detect stationary reqgions between registered images, using M_{k-1} as initial estiamte to compute new mask M_k .

  4. Update internal representation using (*)

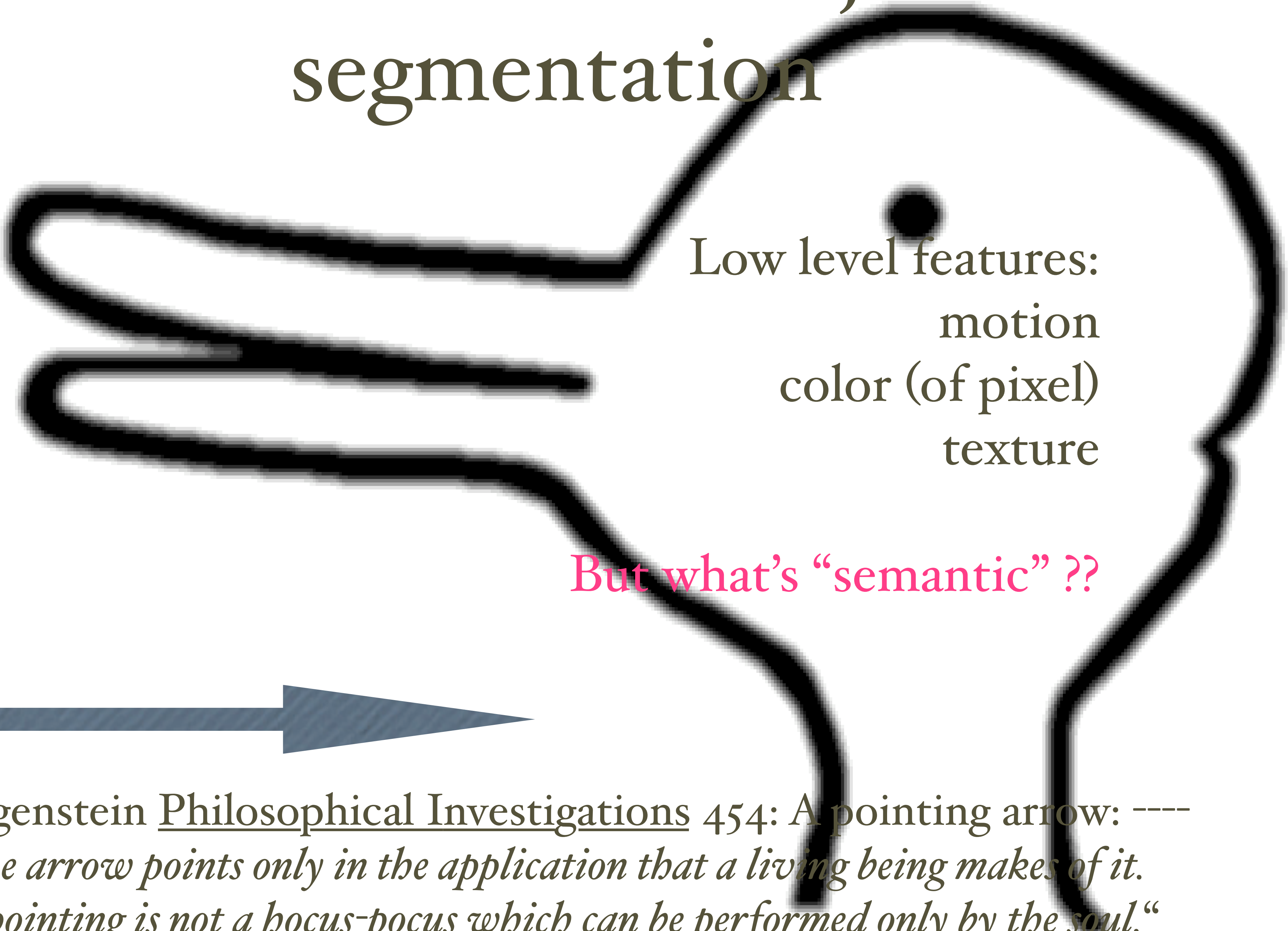
$$\bar{I}_k[\boldsymbol{x}] = (1-\alpha)I_k[\boldsymbol{x}] + \alpha W[\bar{I}_{k-1}[\boldsymbol{x}], I_k[\boldsymbol{x}]]$$

temporal integration 3

- Advantages:
  Comparing each frame with internal representaiton -- weighted by motion warp -- rather than previoous frame, tracks (dominant) moving object.

  - noise in tracked object is lower &
  - image gradients elsewhere are blurred (lower)

semantic video object segmentation



Low level features:
motion
color (of pixel)
texture

But what's "semantic" ??

Wittgenstein <u>Philosophical Investigations</u> 454: A pointing arrow: ----> *"The arrow points only in the application that a living being makes of it. This pointing is not a hocus-pocus which can be performed only by the soul.*"

examples

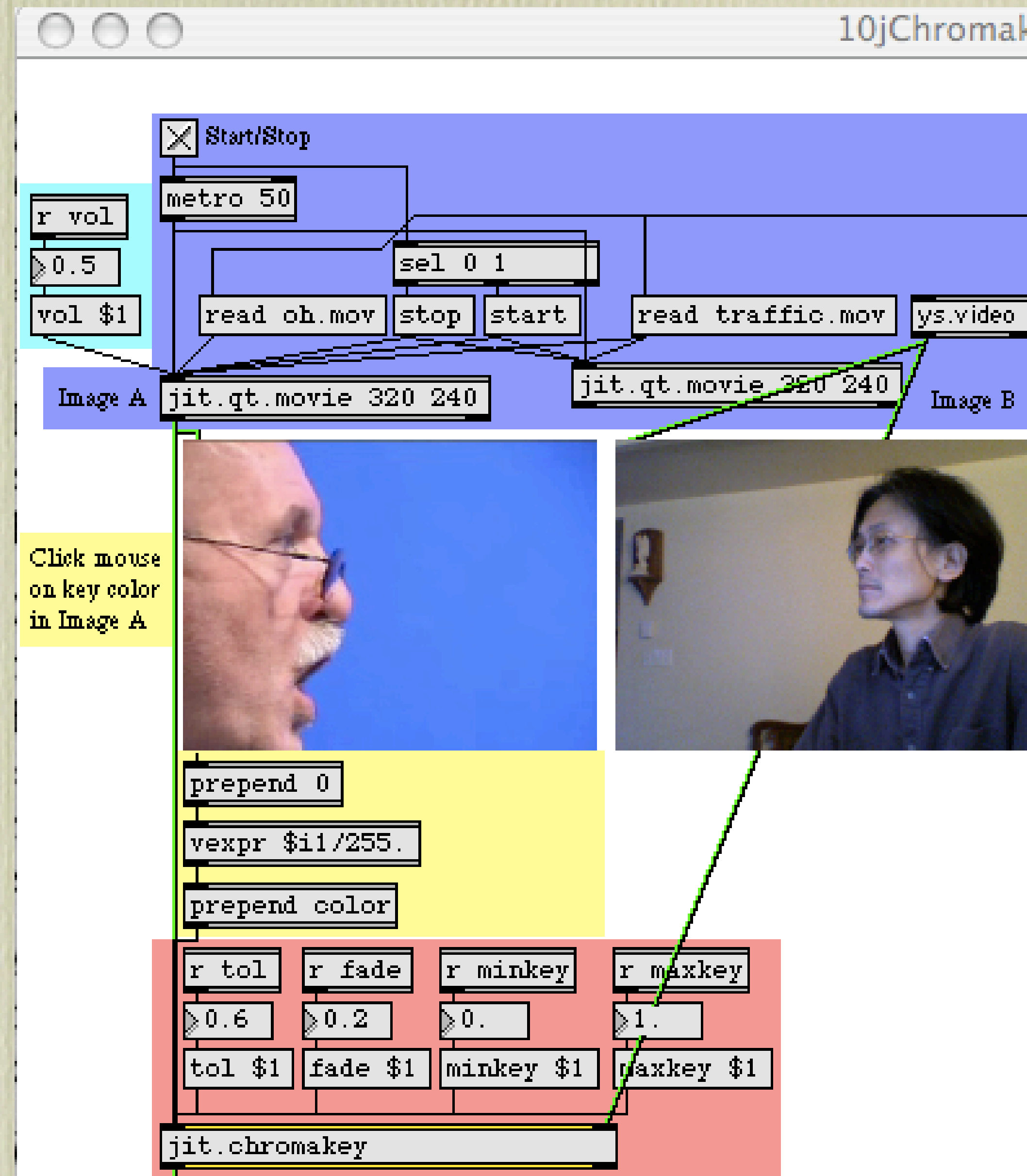
chroma-keying
cv.jit.mean
blob tracking
quennesson's conscious=camera

averaging over time

- cv.jit.mean

chroma-keying

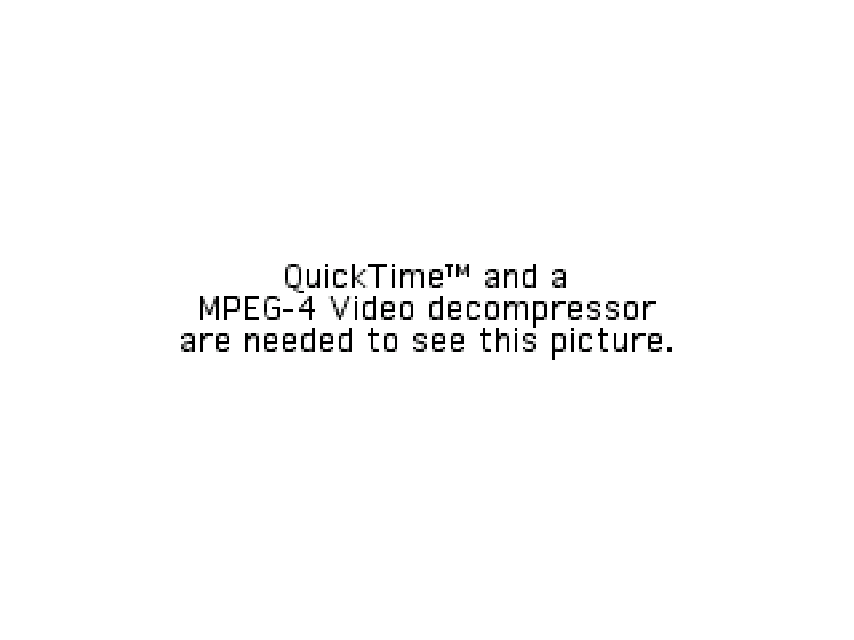
- 10jChromakey-x.pat



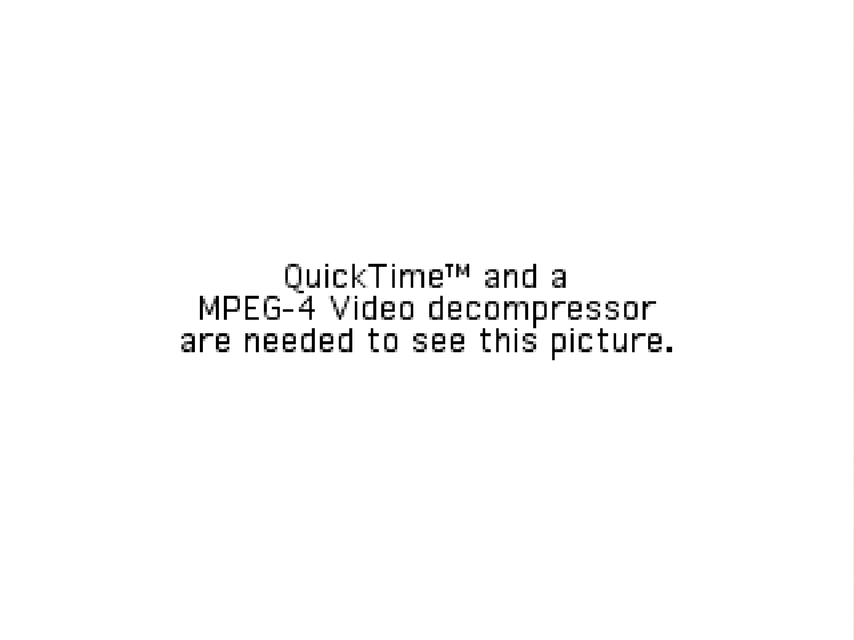
blob tracking

- cv.jit.label
  cv.jit.blobs.bounds
  cv.jit.blobs.centroids
  cv.jit.blobs.direcition
  cv.jit.blobs.elongation
  cv.jit.blobs.moments
  cv.jit.blobs.orientation
  cv.jit.blobs.recon

kevin quennesson



initial test

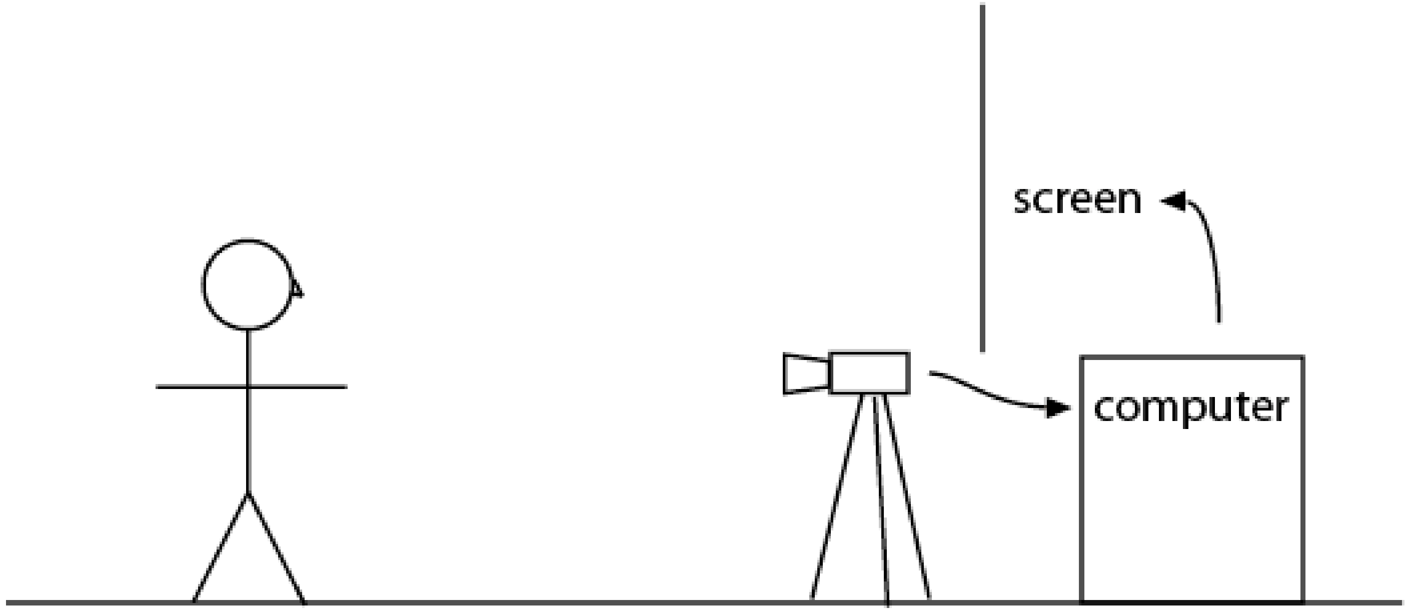


hands

conscious=camera

SIGGRAPH2005

- Interactive video installation



Consciousness of “things”

SIGGRAPH2005

- Static moments: shows face and hands
- Movement: shows body
- Motion: shows trail
- Memory: marks remain on background

Body-tracking technique

SIGGRAPH2005

- Inspired from Pfinder
  - Blob tracking (of face and hands) in YUV space
  - Difference: we use skin tone database
- Technologies used
  - **Platform**: MAC OS X Tiger
  - **Code**: C, Objective-C (Cocoa framework).
  - **Graphics**: vImage (CPU, altivec), Core Image (GPU).
  - **Other**: Core Data, …

Implications

SIGGRAPH2005

- Different work for the programmer
  - Does not know where he is going initially
- Different work for the “creator”
  - Design a function, not an fixed output
    (ie. not *y* in *f(x)=y*, but *f*)
- Different relation of users with the piece
  - What kind of consciousness does the users have of it?
  - What kind of narrative is generated?

QuickTime™ and a
MPEG-4 Video decompressor
are needed to see this picture.