Simple Tracking

- To track an object from one frame to the next:
  - Find the moving object in frame $t$
  - Cross-correlation the attention window against frame $t+1$
  - The best-matching position is where the object went
    - Drop track if below minimum threshold.
  - Image window at $t+1$ position becomes new target; iterate for frame $t+1$

Can we do better?

- Is there something better we can compare to than the raw attention window?
  - Sometimes edges are better to track than pixels
  - Is there something better still?
- Do we have to search all of frame $t+1$?
  - Can we limit the search?
  - Predict where the object is headed?
  - Describe the object’s motion?
  - Exploit foreground information in frame $t+1$?
- What if we don’t find the target in frame $t+1$?

Optimized Correlation Output Filters

Ph.D. thesis by David Bolme
Colorado State University, Dec., 2010.

Step #1: Edge Detection

- To focus on structure, extract edge magnitudes
  - Convolve with Sobel edge masks
  - Compute $Dx$ & $Dy$ for every pixel
  - Edge magnitude is $\sqrt{(Dx^2 + Dy^2)}$
- Remember: linear filter

Intra-class variation

- Challenges for matching & tracking
  - Changes in shape/pose/viewpoint
  - Changes in apparent color
- Goal: learn a general template
  - Capable of matching many samples
  - Within constraints of a linear filter
Edge Detection in Practice

Simple Templates

• Cutting a template from an example doesn’t work…

Training

Convolution Theorem (review)

"In mathematics, the convolution theorem states that under suitable conditions the Fourier transform of a convolution is the pointwise product of Fourier transforms."

Exact Filter

Average of Exact Synthetic Filters
Detection

- Correlate trained template to every video frame
  - Use frequency domain to speed computation
- Find peaks in correlation images
  - Keep peaks that exceed a threshold

First Example Video - YouTube

\[
\frac{1}{N} \sum b_i = \text{AISE Filter}
\]

First Application: Eyes


Is it being used...

MOSSE filter

- Minimize Output Sum of Squared Errors
  \[
  H' = \min_H \sum \left| F_i \cdot H' - G_i \right|^2
  \]
- This form is more stable for small numbers of training samples
  \[
  H = \frac{\sum G_i \cdot F_i^*}{\sum F_i \cdot F_i^* + \epsilon}
  \]

MOSSE Filter Tracking

- User selects initial window to track
- Train filter on
  - Initial window
  - Small affine transformations of initial window
- Update filter using
  - Previous filter
  - If tracked, add small transformations of current window