Simple Tracking

To track an object from one frame to the next:

• Find window with moving object in frame \( t \)
• Correlate window to all locations in frame \( t+1 \)
• Move tracking window to best match
  – Drop track correlation too low.
• 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.

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
Step #1: Edge Detection

• To focus on structure, extract edge magnitudes
  – Convolve with Sobel edge masks
  – Compute $D_x$ & $D_y$ for every pixel
  – Edge magnitude is $\sqrt{(D_x^2 + D_y^2)}$

• Remember: linear filter
Edge Detection in Practice
Simple Templates

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

Edge Image  Edge Template  Correlation
Training

Edge Image (f)  Unknown Template (h)  Desired Output (g)

\[ f \otimes h = g \]

\[ F \cdot H^* = G \]
“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

$F \cdot H^* = G$

$H^* = \frac{G}{F}$
Average of Exact Synthetic Filters

\[ H^* = \frac{1}{N} \sum_i \frac{F_i}{G_i} \]

Image is of \( h \), not \( H \)
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

Person Detection ASEF S2L1_T1234
First Application: Eyes

Is it being used ...
MOSSE filter

• Minimize Output Sum of Squared Errors

\[ H^* = \min_{H^*} \sum_i \left| F_i \cdot H^* - G_i \right|^2 \]

• This form is more stable for small numbers of training samples

\[ H = \frac{\sum_i G_i \cdot F^*_i}{\sum_i F_i \cdot F^*_i + \varepsilon} \]
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
MOSSE Track: Correlation filter based tracking
<table>
<thead>
<tr>
<th>Input</th>
<th>Filter</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template</td>
<td><img src="image1" alt="Template Image" /></td>
<td><img src="image2" alt="Filter Image" /></td>
</tr>
<tr>
<td>MOSSE</td>
<td><img src="image4" alt="MOSSE Image" /></td>
<td><img src="image5" alt="Filter Image" /></td>
</tr>
</tbody>
</table>

**CVPR2010 Tracking Highlights**