MOSSE

CS 510

Lecture #13

March 6, 2019

Colorado State University

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 if track correlation too low.
- Image window at t+1 position becomes new target; iterate for frame t+1

Colorado State University

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.

US Patent 8,520,956. David S. Bolme, J. Ross Beveridge and Bruce A. Draper, Optimized Correlation Filters for Signal Processing, August 27, 2013.

Colorado State University

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 Dx & Dy for every pixel
  - Edge magnitude is $\sqrt{(Dx^2 + Dy^2)}$
- Remember: linear filter

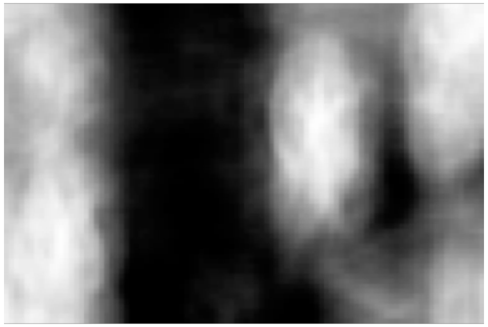
Edge Detection in Practice

Simple Templates

- Cutting a template from an example doesn't always work…

Edge Image

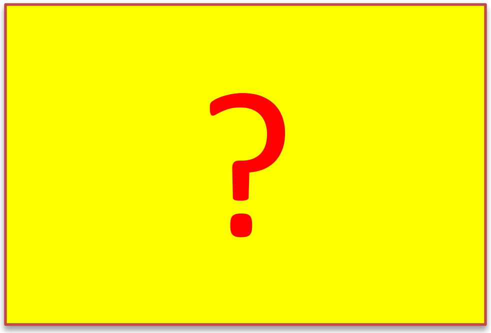
Edge Template



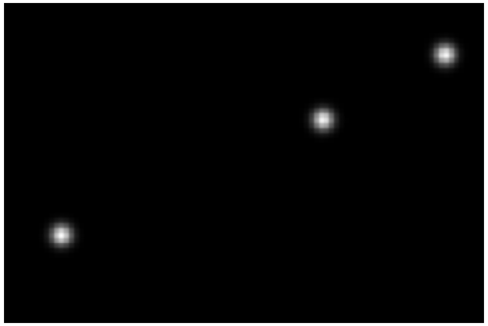
Correlation

Training

Edge Image (f)



Unknown Template (h)



Desired Output (g)

$$f \otimes h = g$$

$$F \cdot H^{*} = G$$

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."

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikimedia Shop

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools

Article Talk

Convolution theorem

From Wikipedia, the free encyclopedia

? This article includes a list of references, related reading or external links, but **its sources remain unclear because it lacks inline citations**. Please improve this article by introducing more precise citations. *(October 2013)*

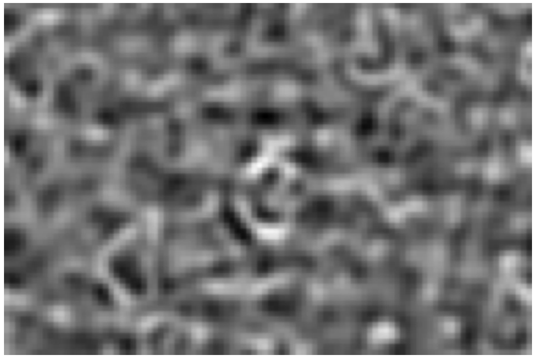
In mathematics, the **convolution theorem** states that under suitable conditions the Fourier transform of a convolution is the pointwise product of Fourier transforms. In other words, convolution in one domain (e.g., time domain) equals point-wise multiplication in the other domain (e.g., frequency domain). Versions of the convolution theorem are true for various Fourier-related transforms. Let f and g be two functions with convolution $f * g$. (Note that the asterisk denotes convolution in this context, and not multiplication. The tensor product symbol $\otimes$ is sometimes used instead.) Let $\mathcal{F}$ denote the Fourier transform operator, so $\mathcal{F}\{f\}$ and $\mathcal{F}\{g\}$ are the Fourier transforms of f and g, respectively. Then

$$\mathcal{F}\{f * g\} = \mathcal{F}\{f\} \cdot \mathcal{F}\{g\}$$

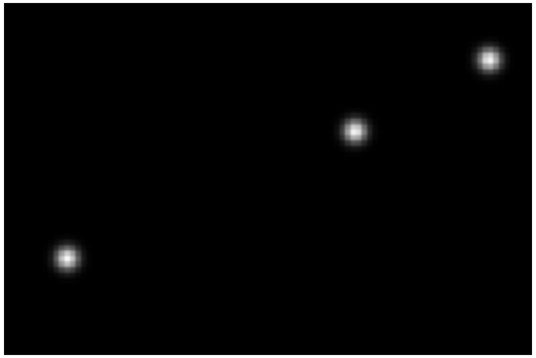
where $\cdot$ denotes point-wise multiplication. It also works the other way around:

Exact Filter

Edge Image (f)



Exact Filter (h)



Output (g)

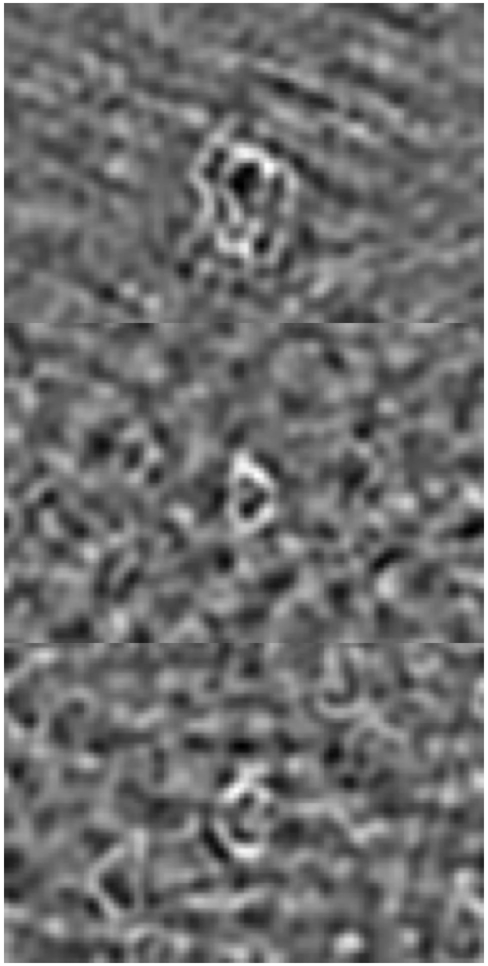
$$F \cdot H^* = G$$

$$H^* = \frac{G}{F}$$

Colorado State University

Average of Exact Synthetic Filters

Exact filters in spatial domain



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

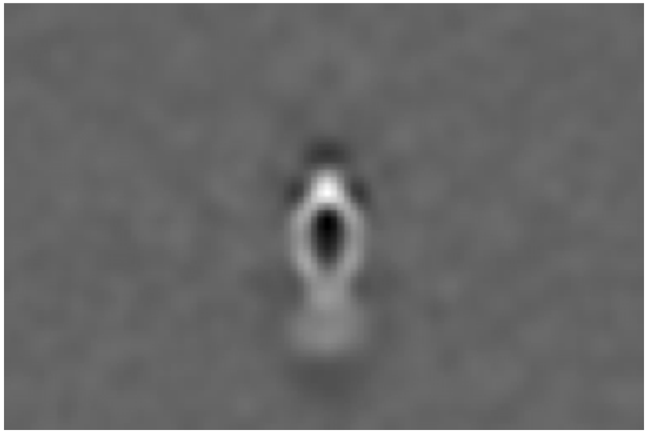


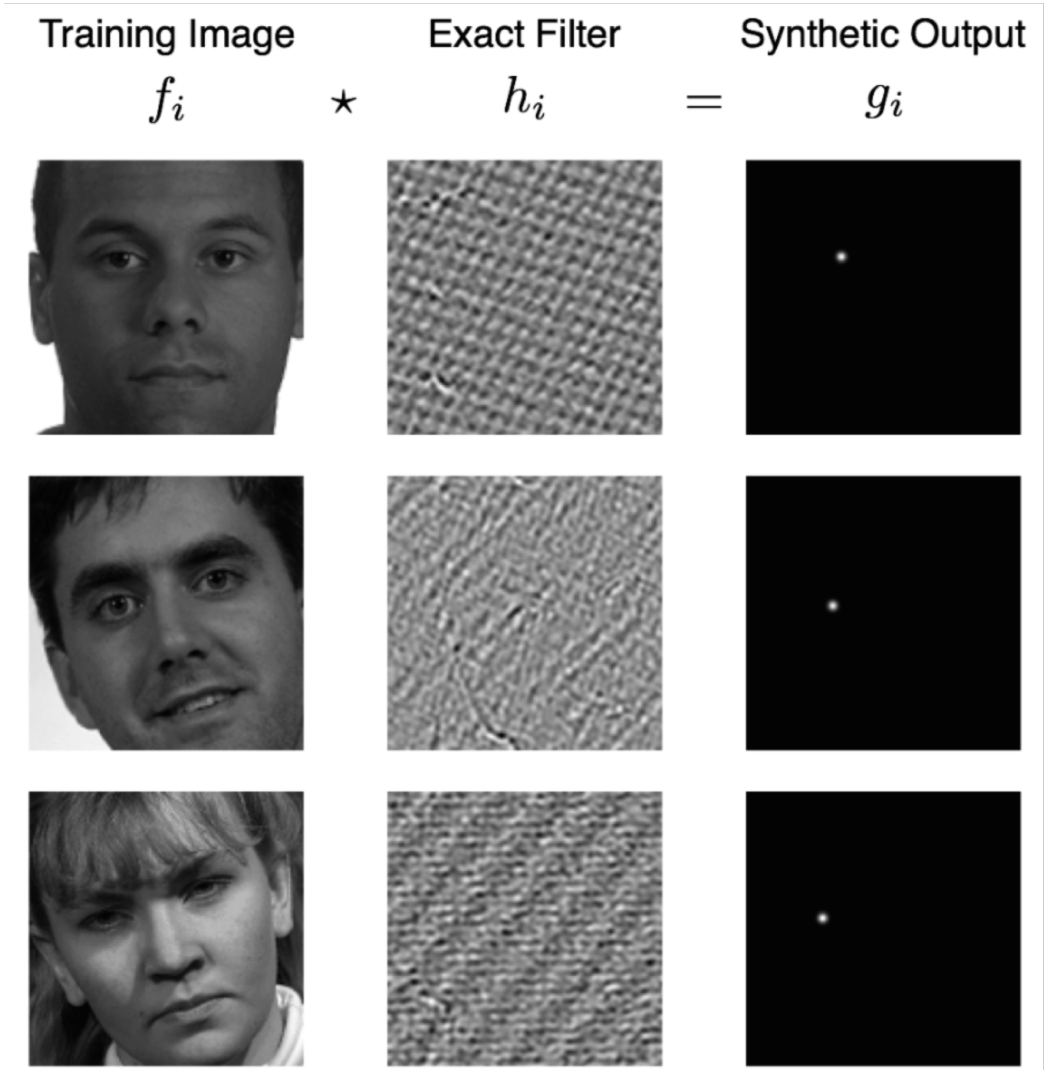
Image is of h, not H

Detection

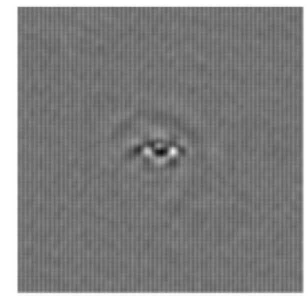
- Correlate trained template to every video frame
  - Use frequency domain to speed computation
- Find peaks in correlation images
  - Keep peeks that exceed a threshold

First Example Video - YouTube

First Application: Eyes



$$\frac{1}{N}\sum_i h_i =$$

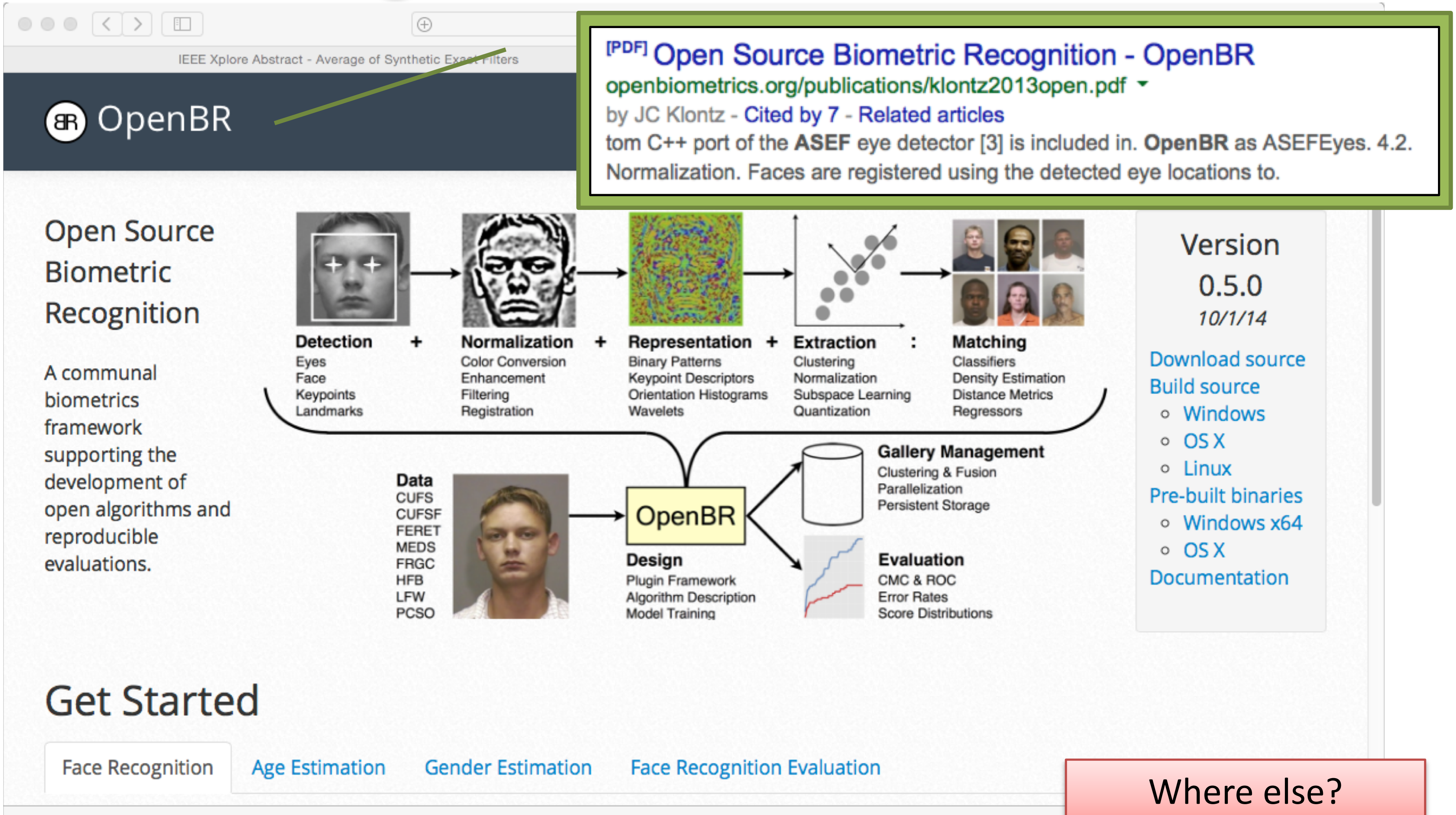


ASEF Filter

Bolme, D.S.; Draper, B.A.; Beveridge, J.R., "Average of Synthetic Exact Filters," *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* , vol., no., pp.2105,2112, 20-25 June 2009

Colorado State University

Is it being used …



Where else?

MOSSE filter

- Minimize Output Sum of Squared Errors

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

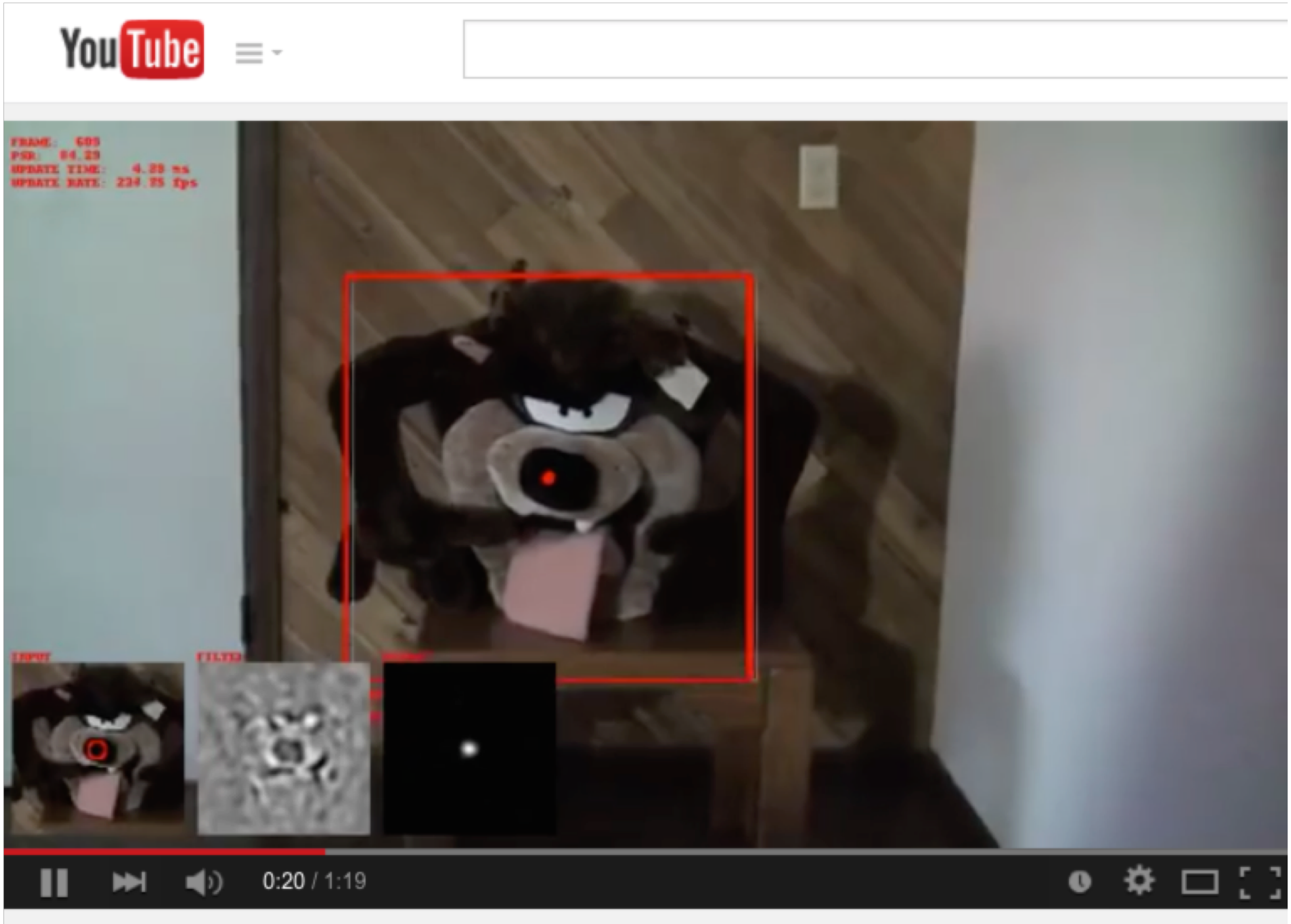
- This form is more stable for small numbers of training samples

$$H = \frac{\sum_i G_i \cdot F^*{}_i}{\sum_i F_i \cdot F^*{}_i + \varepsilon}$$

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

Colorado State University



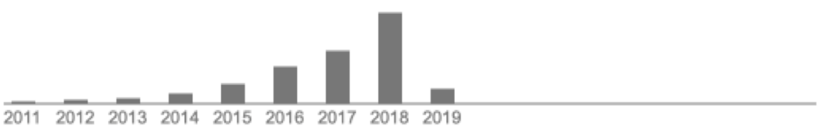
MOSSE Track : Correlation filter based tracking

Is MOSSE being used - update

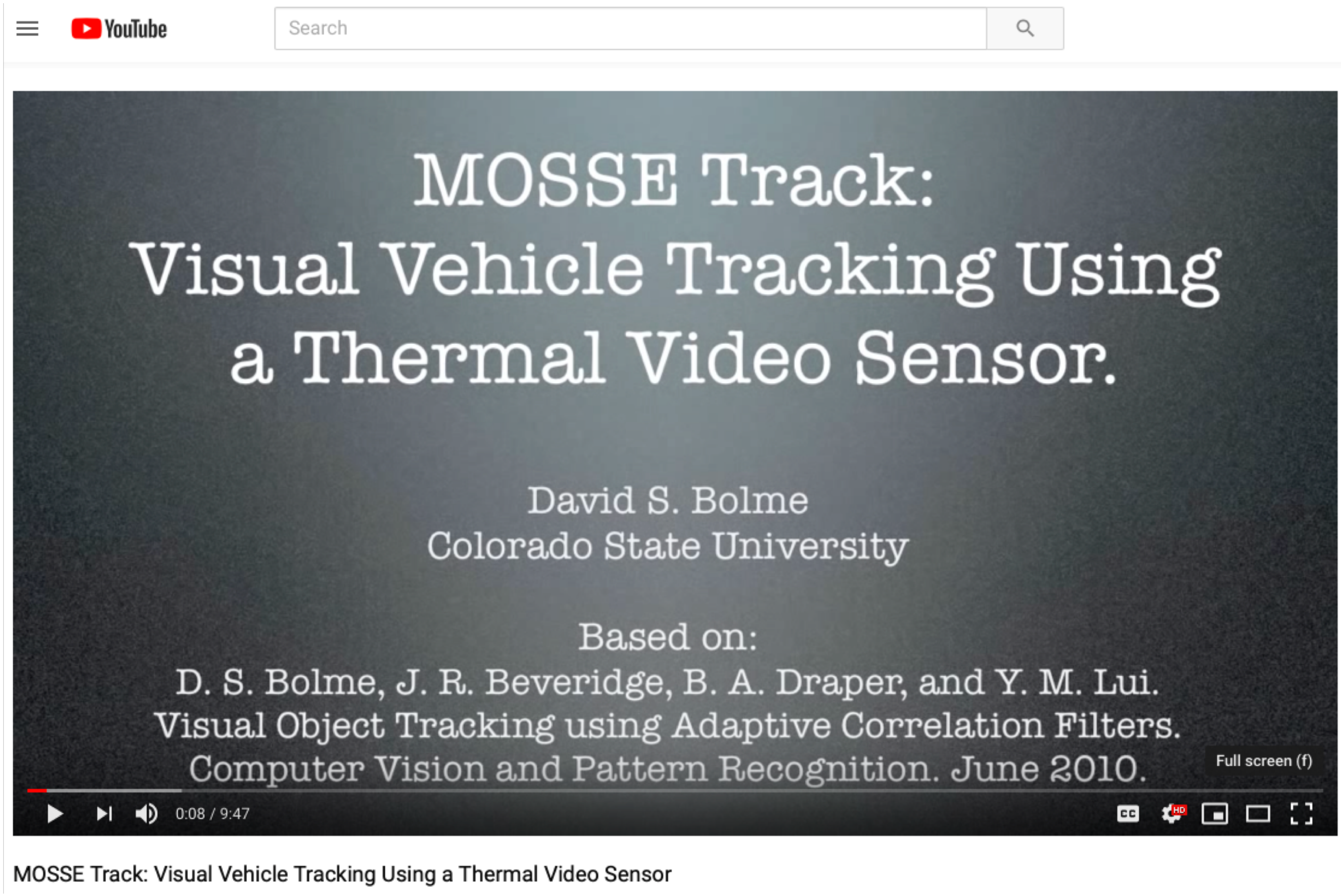
Visual object tracking using adaptive correlation filters

[PDF] from ieee.org
FindIt@CSU

Authors	David S Bolme, J Ross Beveridge, Bruce A Draper, Yui Man Lui
Publication date	2010/6/13
Conference	2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition
Pages	2544-2550
Publisher	IEEE
Description	Although not commonly used, correlation filters can track complex objects through rotations, occlusions and other distractions at over 20 times the rate of current state-of-the-art techniques. The oldest and simplest correlation filters use simple templates and generally fail when applied to tracking. More modern approaches such as ASEF and UMACE perform better, but their training needs are poorly suited to tracking. Visual tracking requires robust filters to be trained from a single frame and dynamically adapted as the appearance of the target object changes. This paper presents a new type of correlation filter, a Minimum Output Sum of Squared Error (MOSSE) filter, which produces stable correlation filters when initialized using a single frame. A tracker based upon MOSSE filters is robust to variations in lighting, scale, pose, and nonrigid deformations while operating at 669 frames per second. Occlusion is detected …
Total citations	Cited by 1129



Scholar articles
Visual object tracking using adaptive correlation filters
DS Bolme, JR Beveridge, BA Draper, YM Lui - 2010 IEEE Computer Society Conference on Computer …, 2010
Cited by 1129 Related articles All 10 versions


YouTube
Search
MOSSE Track:
Visual Vehicle Tracking Using
a Thermal Video Sensor.
David S. Bolme
Colorado State University
Based on:
D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui.
Visual Object Tracking using Adaptive Correlation Filters.
Computer Vision and Pattern Recognition. June 2010.
Full screen (f)
0:08 / 9:47
MOSSE Track: Visual Vehicle Tracking Using a Thermal Video Sensor