# Accurate Scale Estimation for Robust Visual Tracking

**Winner of VOT 2014**



Martin Danelljan
Gustav Häger

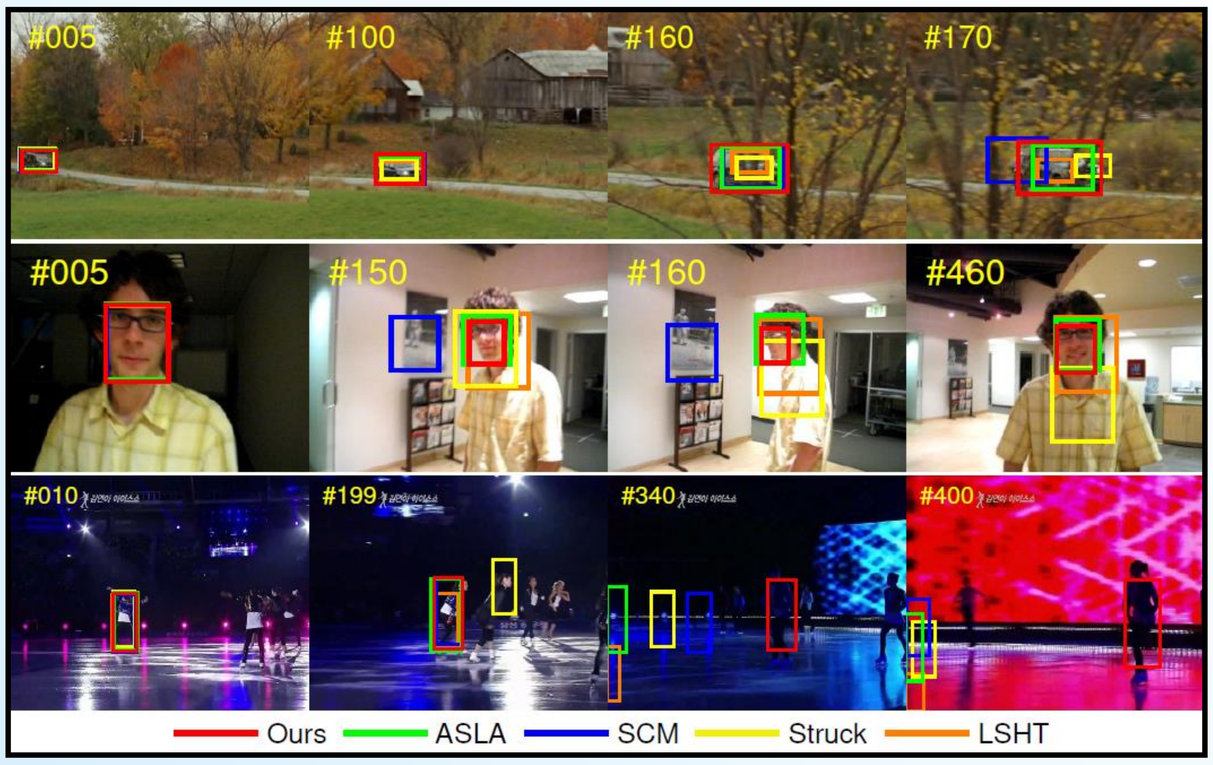

Fahad Khan
Michael Felsberg


Computer Vision Laboratory, Linköping University, Sweden


## Introduction

**Problem**

Accurate scale estimation for visual tracking.




**Motivation**

- Most trackers do not estimate scale.
- Those that do are usually slow.

**Contribution**

- Fast and accurate scale-adaptive tracker.
- A generic scale estimation component.

## Multi-channel DCF:s

We use discriminative correlation filters (DCF:s).

- Local target patch of feature dimension $l$: $f^l$
- Desired correlation output: $g$
- Sought correlation filter for dimension $l$: $h^l$

Find $h^l$ that minimizes:

$$\epsilon = \left\| \sum_{l=1}^{d} h^l \star f^l - g \right\|^2 + \lambda \sum_{l=1}^{d} \left\| h^l \right\|^2$$

Solved by: $H^l = \frac{A^l}{B} = \frac{\bar{G} F^l}{\sum_{k=1}^{d} \overline{F^k} F^k + \lambda}$

Capital letters denote DFT:s, e.g. $H^l = \mathcal{F}\{h^l\}$.

The model is updated similarly to [1]:

Numerator: $A_t^l = (1-\eta) A_{t-1}^l + \eta \bar{G} F^l$

Denominator: $B_t = (1-\eta) B_{t-1} + \eta \sum_{k=1}^{d} \overline{F^k} F^k$

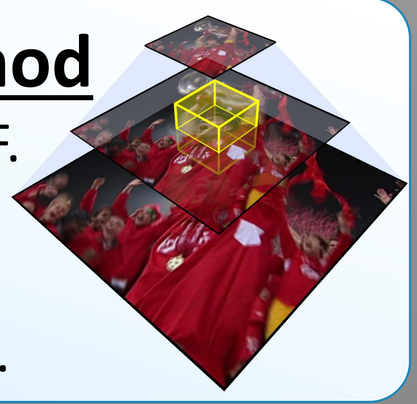
## Exhaustive Method

**First approach:** 3D scale space DCF.

**+** Joint translation-scale tracking.

**-** High computational complexity.

**-** Scale space shearing distortions.

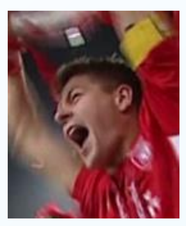


## Discriminative Scale Space Tracker

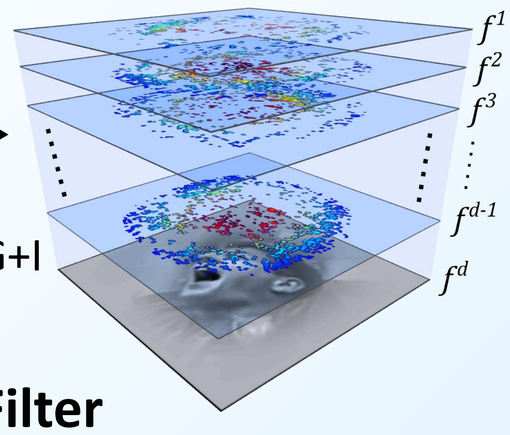
- First apply a translation filter.
- Use a separate 1D filter for scale estimation.

### The Translation Filter

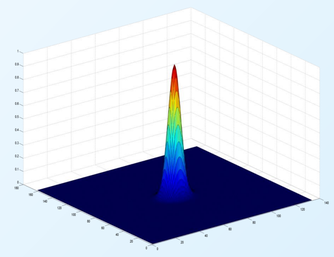
Translation sample $f$



HOG+I

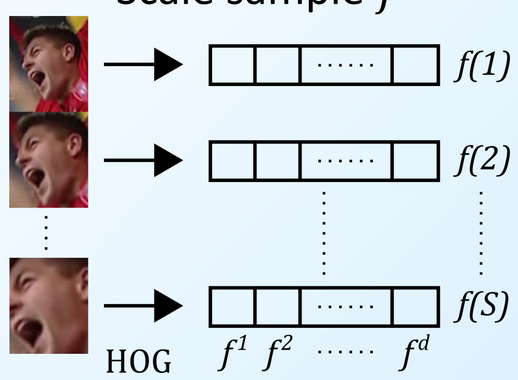



Desired translation correlation output $g$

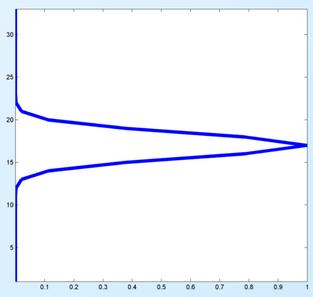


### The Scale Filter

Scale sample $f$




Desired scale correlation output $g$



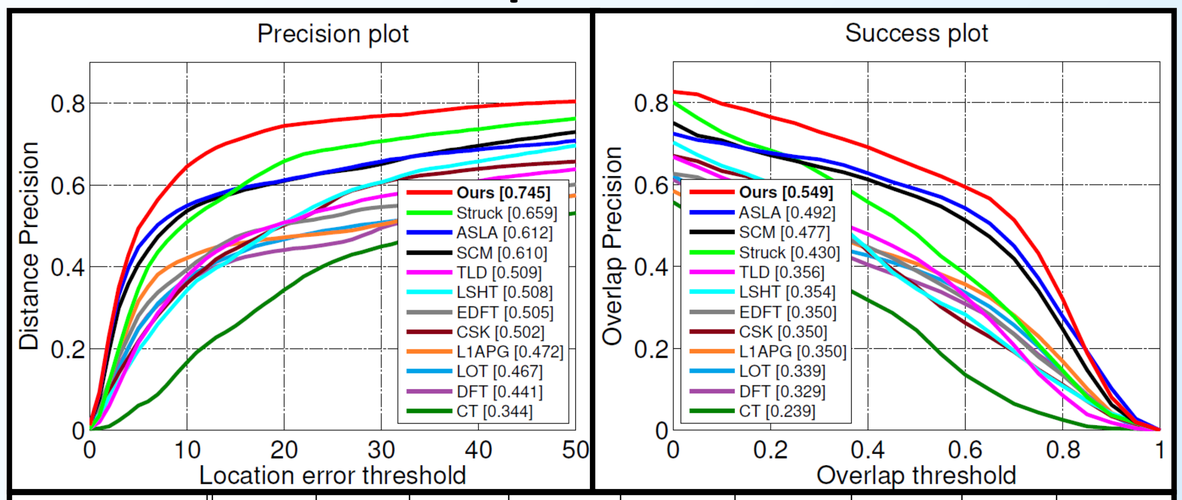
## Evaluation

- We use the tracking benchmark protocol [2].
- All the 28 “scale variation” sequences in [2].
- We compare with 11 state-of-the-art trackers.

### Scale Estimation

| Method | median OP | median DP | median CLE | median FPS |
|---|---|---|---|---|
| Baseline (no scale) | 37.8 | 74.5 | 15.9 | **44.1** |
| Exhaustive Scale Search (this paper) | 52.2 | 87.6 | 11.8 | 0.96 |
| Fast Scale Search (this paper) | **75.5** | **93.3** | **10.9** | 24.0 |

### State-of-the-Art Comparison




| | Ours | CSK | LOT | ASLA | SCM | Struck | LSHT | TLD | EDFT |
|---|---|---|---|---|---|---|---|---|---|
| Median FPS | *24* | **152** | 0.517 | 0.959 | 0.0828 | 8.96 | 12.5 | 21 | 20.6 |

### Visual Object Tracking Challenge 2014

- Winner of VOT challenge at ECCV 2014.
- 25 videos, independent verification of results.
- The same parameter settings as in the paper!


**References**

[1] Danelljan et al. Adaptive Color Attributes for Real-Time Visual Tracking. In CVPR, 2014.

[2] Y. Wu, J. Lim, and M.-H. Yang. Online object tracking: A benchmark. In CVPR, 2013.