Discriminative Correlation Filters for Visual Tracking
Martin Danelljan
Computer Vision Laboratory, Dept. of EE, Linköping University

Introduction:

Visual Tracking:
Tracking a target object in a video.

Applications:
Robotics, autonomous driving, 3D reconstruction, surveillance.

Challenges:
Complex appearance changes, environmental factors (occlusions etc.).

Discriminative Correlation Filters (DCF):
DCF is a learning technique that efficiently exploits all cyclic shifts of the training samples. Recently DCF methods have shown state-of-the-art results for visual object tracking.

Goal:
Research and improve the DCF framework, with a particular focus on the Visual Tracking problem.

Spatially Regularized DCF
Problem: The periodic assumption in standard DCF.
- Limits training data
- Restricts search region
- Inaccurate negative samples

Our Approach:
Learn convolution filter \( f \) from samples \( x_k, y_k \). We add spatial regularization \( w \).

\[ \varepsilon_t(f) = \sum_{k=1}^{t} \alpha_k \| S_f(x_k) - y_k \|^2 + \sum_{l=1}^{d} \| w \cdot f^l \|^2 \]
- Penalizes filter coefficients in background regions
- Enables larger samples for training and detection
- Increases discriminative power, more robust tracking
- Efficient optimization in the Fourier domain

Top rank in TIR Object Tracking Challenge 2015.

Publication:
Learning Spatially Regularized Correlation Filters for Visual Tracking.

Unified Discriminative Tracking
Problem: Discriminative trackers rely on the quality of the training set used for learning the classifier.
- Training samples labelled by the tracker
- Corrupted samples due to occlusions, clutter, etc.
- Leads to drift and tracking failure

Our approach: a unified learning formulation. Jointly optimize model \( f \) and the sample weights \( \alpha_k \).

Publication:
In the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

Continuous Convolution Operators

DCF:
Learns a set of discrete filters for target localization
Outputs discrete detection scores

Our Approach:
Posing the learning problem in the continuous spatial domain

Multi-resolution (deep) feature map
Learns continuous filters
Outputs continuous detection scores

Advantages:
- Integration of multi-resolution (deep) features
- Accurate sub-pixel (or sub-grid) localization
- Sub-pixel supervision in the learning
- Efficient processing of all available information
- Avoids artefacts caused by explicit resampling


Publication:
Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking.

Experiments:
OTB100 and TC128 datasets.