Learning Spatially Regularized Correlation Filters for Visual Tracking

**Introduction**

**Application:** Visual tracking

**Goal:** Address problems caused by the periodic assumption in Discriminative Correlation Filters (DCF).

- **Limits training data**
- **Restricts search region**
- **Inaccurate** negative training samples
- **Increased sample size degrades** discriminative power

**Contributions**

- **Spatial regularization** component in the DCF learning.
- **Penalizes** filter coefficients based on the target size.
- **Alleviates** the negative effects of the periodic assumption.
- **Training and detection on larger image regions.**
- **Increases discriminative power => more robust** tracking.
- **Efficient optimization** method.
- **Unlike previous DCF trackers, no approximate learning.**
- **Fast target detection with sub-grid accuracy.**
- **Best results on 4 datasets:** OTB-13, OTB-15, ALOV, VOT2014.

**The Standard DCF Tracking**

**Idea:** Learn a convolution filter $f$ from examples $\{(x_k, y_k)\}_{k=1}^{t}$

**Convolution response:**

$S_f(x) = \{1 \big\{ \sum_{l=1}^{d} x^l \cdot f^l \big\} \}_{l=1}^{d}$

**Learning loss:**

$\mathcal{L}(f) = \sum_{k=1}^{t} \| S_f(x_k) - y_k \|^2 + \lambda \sum_{l=1}^{d} \| f^l \|^2$

**Optimization:**

Solved using approximate update rules, leading to suboptimal filters $f$.

**Detection:**

Maximize classification scores in the image sample $z$. $S_f(z) = \mathcal{F}^{-1}\{ \sum_{l=1}^{d} z^l \cdot f^l \}_{l=1}^{d}$

**Our loss:**

The regularization weights $w$ penalize filter coefficients in the background.

$\varepsilon(f) = \sum_{k=1}^{t} \alpha_k \| S_f(x_k) - y_k \|^2 + \sum_{l=1}^{d} \| w^l \cdot f^l \|^2$

**Resulting filter coefficients $f^l$:**

- **Standard regularization**
- **Our spatial regularization**

**Efficient optimization:**

Parseval: Discrete Fourier Transform (DFT)

$\varepsilon(\tilde{f}) = \sum_{k=1}^{t} \alpha_k \| D(\tilde{x}_k)^T \tilde{f} - y_k \|^2 + \sum_{l=1}^{d} \| \tilde{w}^l \cdot MN \cdot f^l \|^2$

The regularization weights $w^l$ have sparse DFT coefficients $\tilde{w}^l$, which gives a sparse convolution matrix $C(\tilde{w})$.

**Vectorization:**

$\varepsilon(\tilde{f}) = \sum_{k=1}^{t} \alpha_k \| D(\tilde{x}_k)^T \tilde{f} - y_k \|^2 + \sum_{l=1}^{d} \| C(\tilde{w})MN \cdot f^l \|^2$

Normal equations are solved in a real Fourier basis $A_t \tilde{f} = b$.]

**Efficient incremental update of the system:**

$A_t = \sum_{k=1}^{t} \alpha_k D^T_k D_k + W^T W$ \quad $b_t = \sum_{k=1}^{t} \alpha_k D^T_k y_k$

- **Iterative solution using Gauss-Seidel by exploiting sparsity.**
- **Initialization with previous solution ensures fast convergence.**

**Fast sub-grid detection**

**Fourier domain interpolation:** $s(u,v) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \hat{s}(m,n)e^{2\pi(iu/M + iv/N)}$

**Experiments**

**Baseline evaluation on OTB-2013 [2].**

<table>
<thead>
<tr>
<th>Regularization</th>
<th>Conventional sample size</th>
<th>Expanded sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean OP (%)</td>
<td>Ours</td>
<td>Standard</td>
</tr>
<tr>
<td>----------------</td>
<td>------</td>
<td>----------</td>
</tr>
<tr>
<td>OTB-2013</td>
<td>63.3</td>
<td>61.9</td>
</tr>
<tr>
<td>OTB-2015</td>
<td>30.1</td>
<td>31.9</td>
</tr>
<tr>
<td>VOT2014</td>
<td>89.2</td>
<td>89.5</td>
</tr>
<tr>
<td>VOT2015</td>
<td>88.2</td>
<td>88.5</td>
</tr>
</tbody>
</table>

**Comparison with ICCV 2015 trackers on OTB-2013**

**OTB-2015 [3] (100 videos)**

- **ALOV++**
- **59.9**
- **Efficient tracking**
- **Fahad Khan**
- **Winner of OpenCV Challenge in tracking [1].**
- **Among top 3 in VOT 2015 and VOT-TIR 2015.**
- **Best results on OTB-2013 published in ICCV.**

**References**