Real-Time Visual Tracking: Promoting the Robustness of Correlation Filter Learning

Yao Sui
University of Kansas
Lawrence, KS 66045, USA

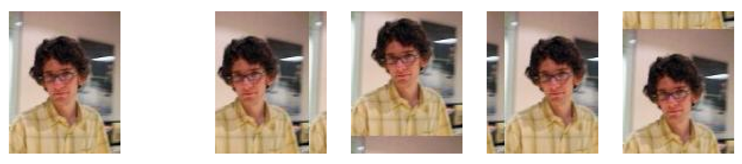
Ziming Zhang
Boston University
Boston, MA 02215, USA

Guanghui Wang
University of Kansas
Lawrence, KS 66045, USA

Yafei Tang
China Unicom Research
Beijing 100032, China

Li Zhang
Tsinghua University
Beijing 100084, China

Problem Statement



Correlation filtering from circulant structure of tracking (Henriques et al. 2015 T-PAMI)

- **Isotropic Response** (the Gaussian shaped response) may fail to reveal the circulant structure due to the discontinuity from the cyclic shifts.
- **Squared Loss** is unable to reliably respond to the drastic appearance changes, e.g., in the presence of occlusions.

Solution: Anisotropy

Basic Idea: Using more robust loss function to learn the correlation filter, resulting in an anisotropic response.

$$\min_{\mathbf{w}} \sum_i \ell\left(f\left(\mathbf{x}_i\right) - y_i\right) + \lambda \left\|\mathbf{w}\right\|_2^2$$

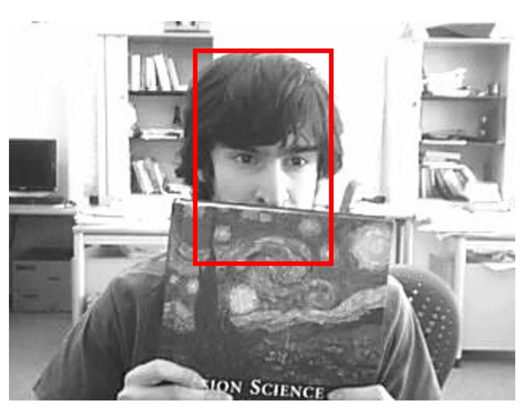
where the loss function $\ell \in \{\ell_1, \ell_1\ell_2, \ell_{2,1}\}$.

Implementation: Introducing relax variables

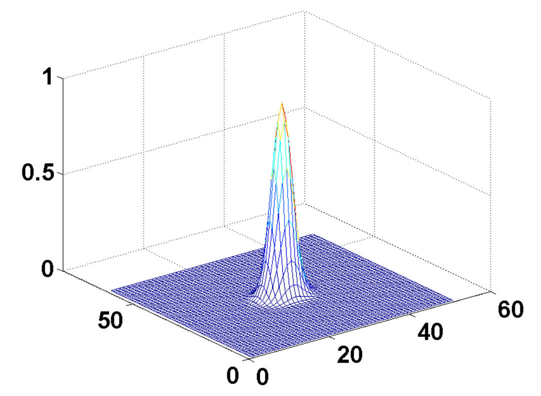
$$\min_{\mathbf{w},\mathbf{e}} \sum_i \ell\left(e_i\right) + \lambda \left\|\mathbf{w}\right\|_2^2$$
$$s.t.\ e_i = y_i - f\left(\mathbf{x}_i\right)$$

and alternately optimizing the variables.

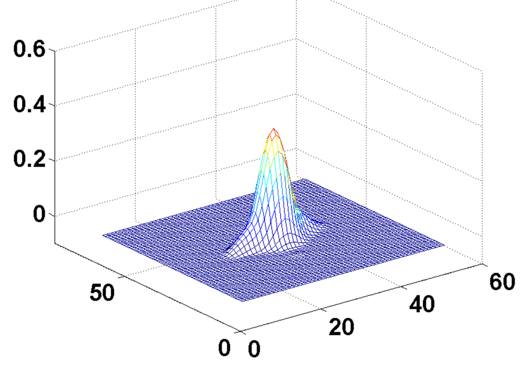
An Example of the Anisotropic Response



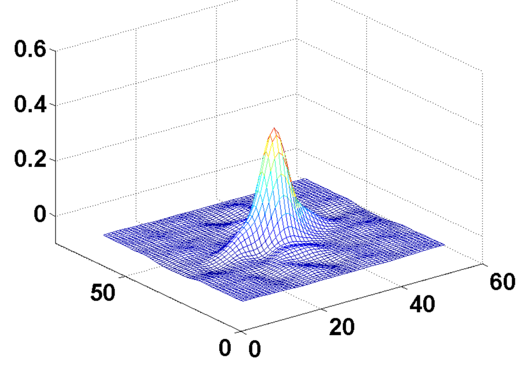
A frame



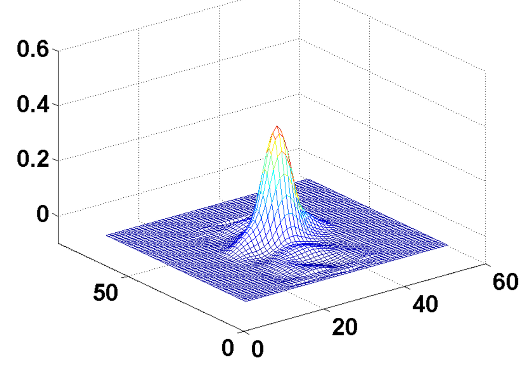
ℓ_2-loss (Gaussian shaped)



ℓ_1-loss (sparsity)

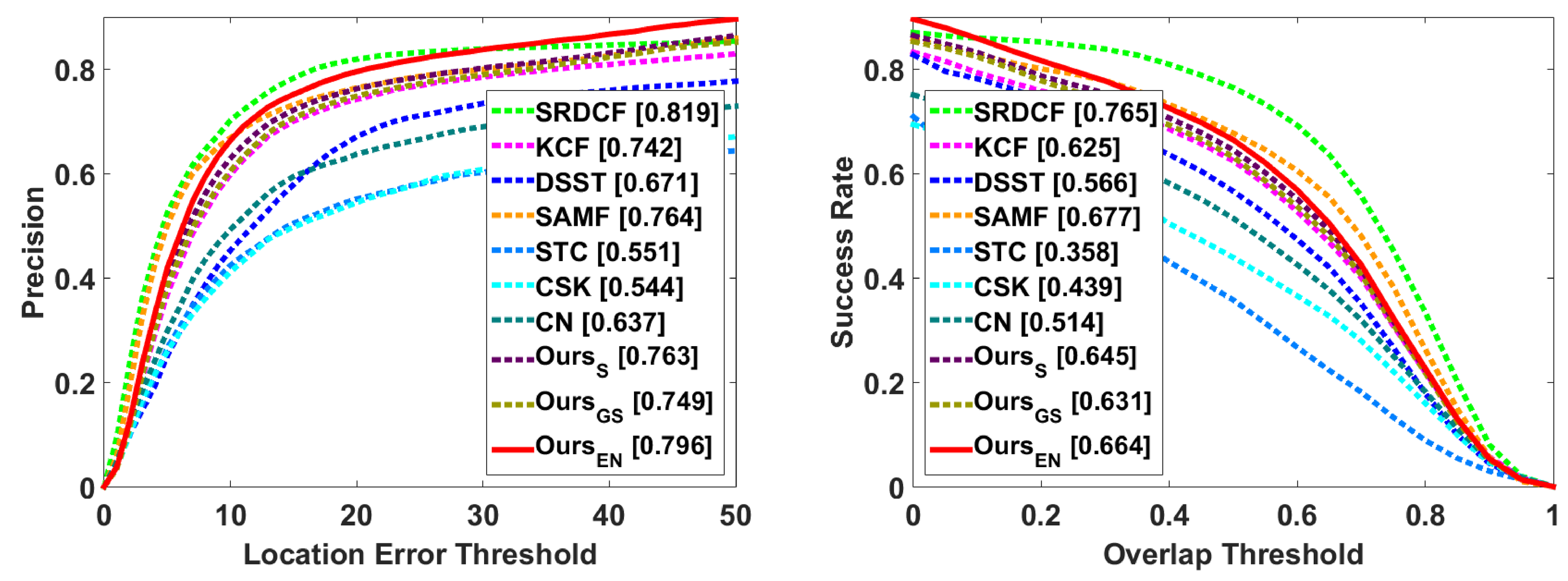


$\ell_1\ell_2$-loss (elastic net)

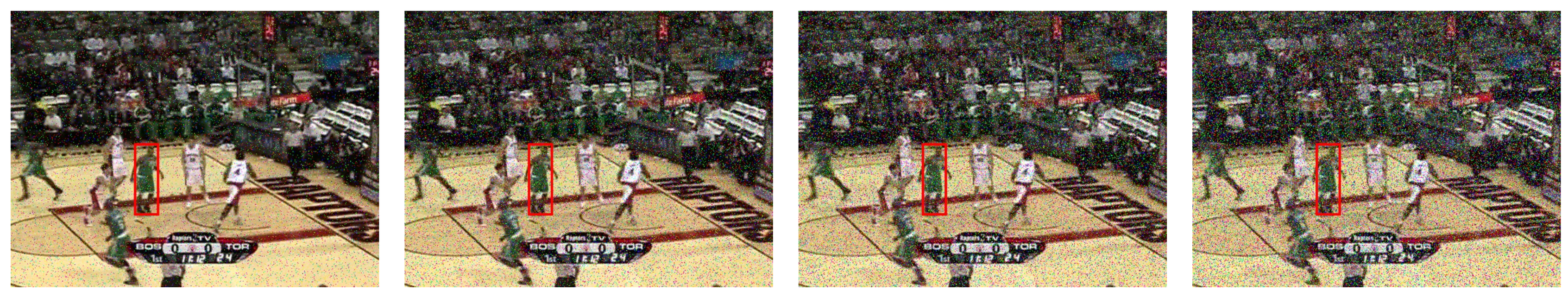


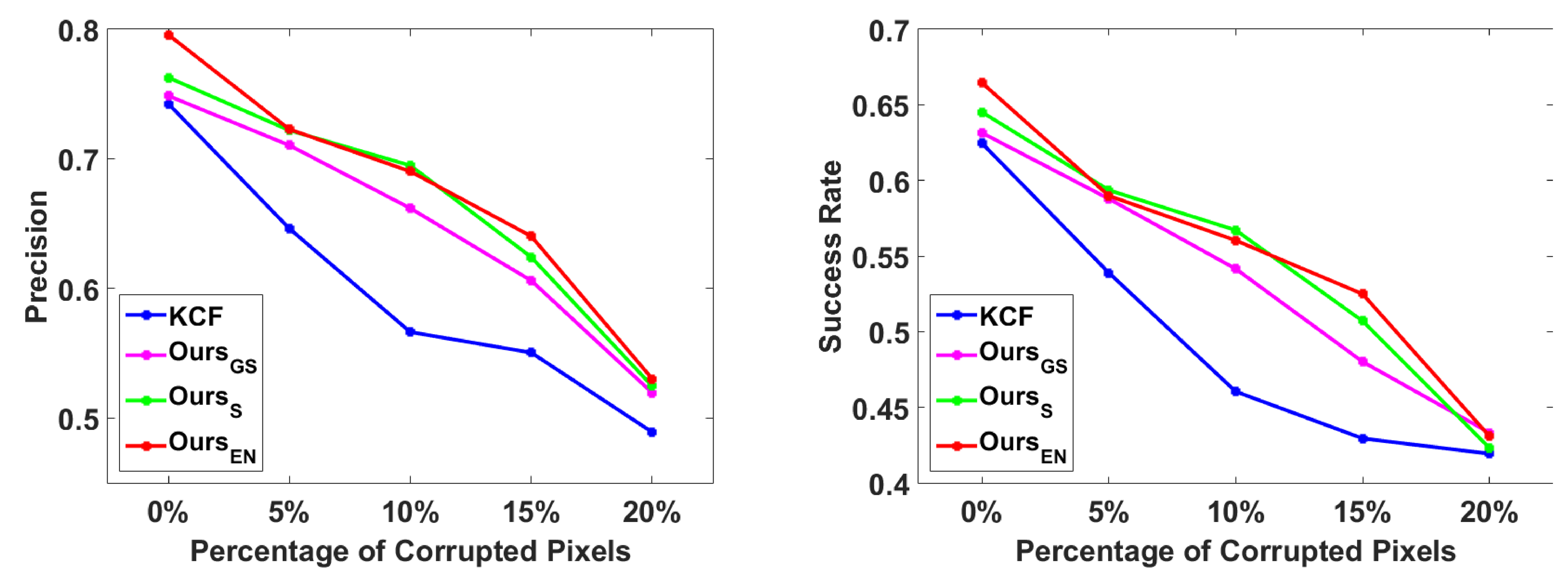
$\ell_{2,1}$-loss (group sparsity)

Experimental Results



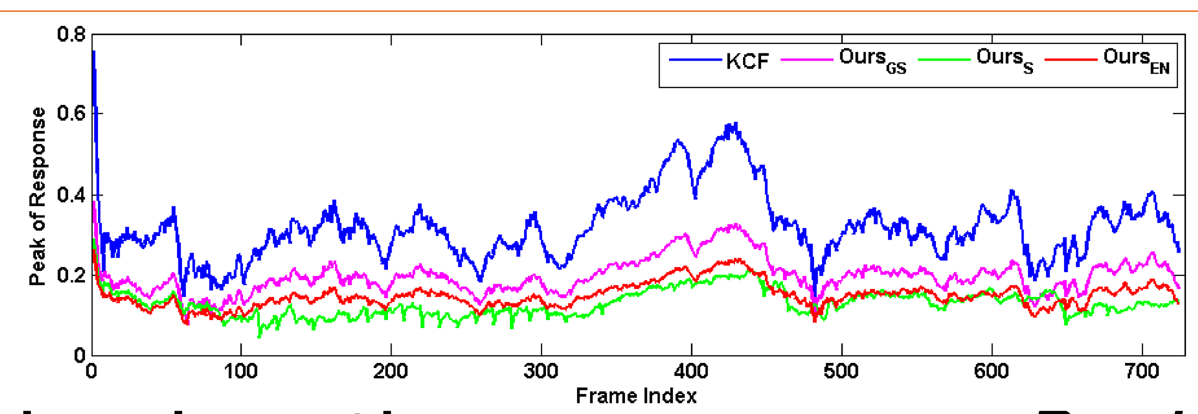
Performance on the OTB 2013 benchmark





Performance on the OTB 2013 benchmark with {5, 10, 15, 20}% corrupted pixels.

The Anisotropic Response

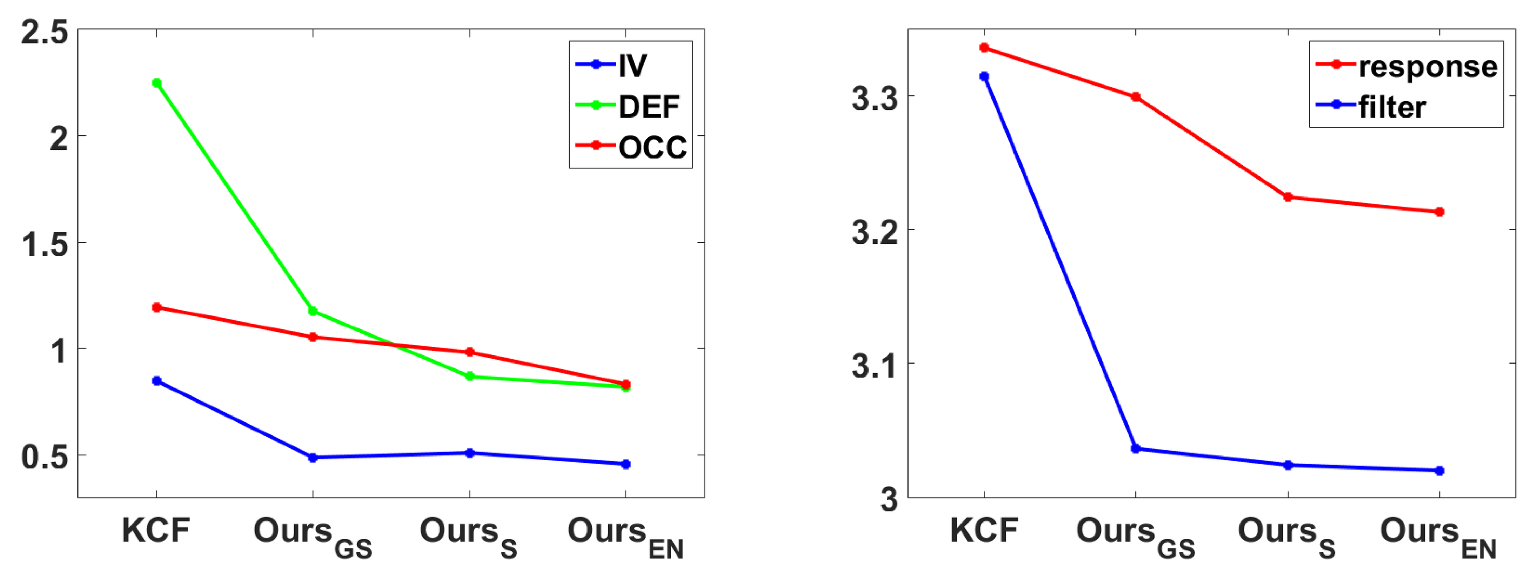


Peak values the responses on *Basketball*

Definition of the *Peak Sensitivity:*

$$s = \sum_{i=1}^{n} \left(p_i - p_m\right)^2$$

where p_i and p_m denote the peak values of the response in the i-th frame and the mean peak values in the n frames, respectively.



Peak sensitivity in 3 cases (left) and all frames of the OTB 2013 benchmark.