Adaptive Color Attributes for Real-Time Visual Tracking

Martin Danelljan[1], Fahad Khan[1], Michael Felsberg[1], Joost van de Weijer[2]

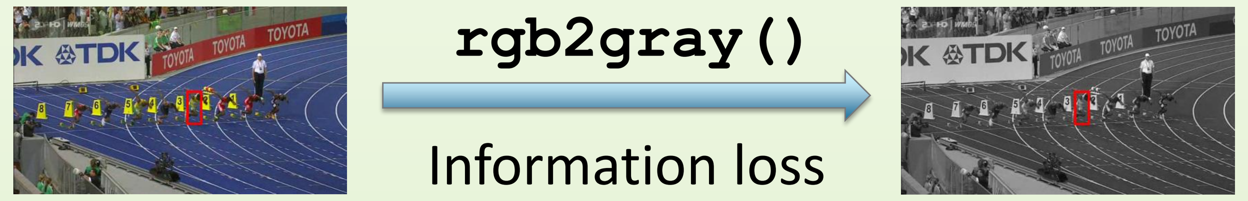
[1]Computer Vision Laboratory, Linköping University, Sweden • [2]Computer Vision Center, CP Dept. Universitat Autonoma de Barcelona, Spain

IEEE 2014 Conference on Computer Vision and Pattern Recognition

Introduction

Problem

- Color has been largely ignored in the tracking community.

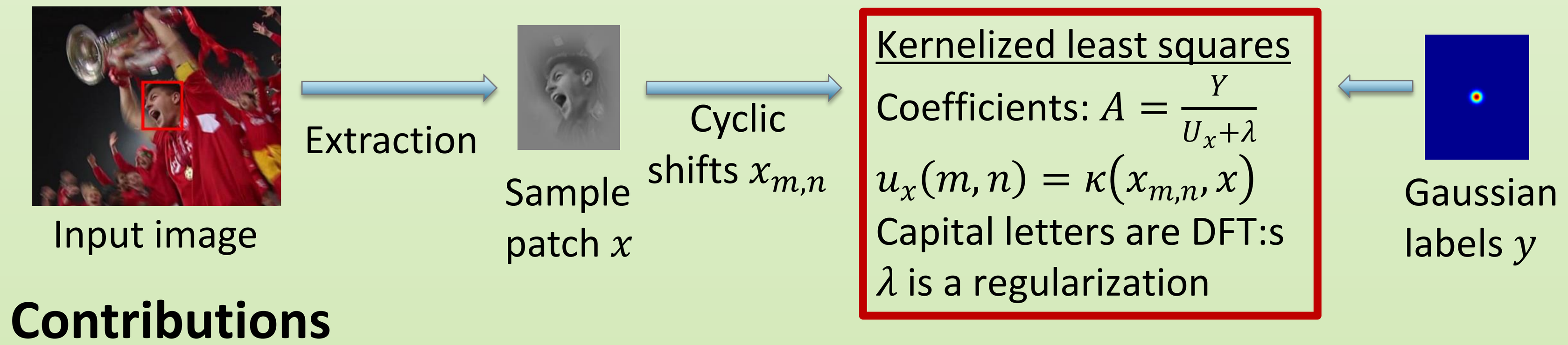


Motivation

- Recently color representations has been successfully applied to several related areas in computer vision, e.g. object detection.

Baseline

- We start from a baseline tracker, called CSK [1].
- Fastest among the top 10 in the recent evaluation [2].
- Learns a kernelized least squares classifier on the appearance.



Contributions

1. Improved classifier model update scheme.
2. Incorporation of color information into the tracker.
3. Dynamical selection of color features for tracking.

Improved Update Scheme

- The original CSK update scheme is: $A^p = (1-\gamma)A^{p-1} + \gamma A$.
- This lacks temporal consistency, and is unstable for high dimensional color representations.
- We consider all target samples simultaneously in one cost to derive a consistent update scheme.
- The constrained weighted sum of errors are minimized by choosing:
  - Numerator: $A_N^p = (1-\gamma)A_N^{p-1} + \gamma Y^p U_x^p$
  - Denominator: $A_D^p = (1-\gamma)A_D^{p-1} + \gamma U_x^p(U_x^p + \lambda)$
  - Template appearance: $\hat{x}^p = (1-\gamma)\hat{x}^{p-1} + \gamma \hat{x}$

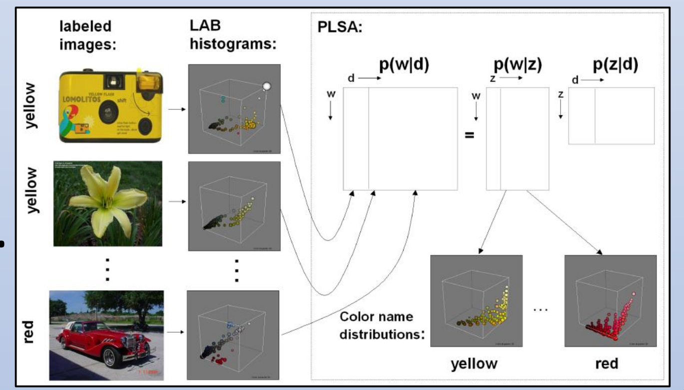
Coloring Visual Tracking

Color Attributes

- We propose to use color attributes [3] for tracking.
- A probabilistic representation of the 11 basic color names in the English language.
- Successfully applied to object detection, object recognition and action recognition.

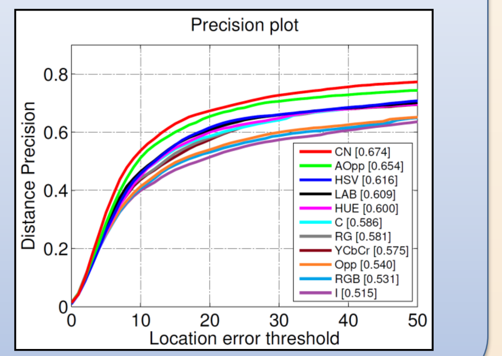
Properties

- Certain degree of photometric invariance.
- Discriminative power due to achromatic colors.
- Compactness due to only 11-D histogram.



Color Evaluation

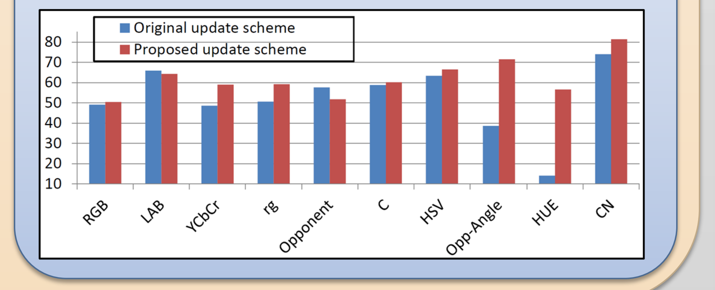
- 10 color representations.
- Intensity augmented.
- Color attributes performs best.
- 16 % improvement over using intensity alone.



Demo

Update Evaluation

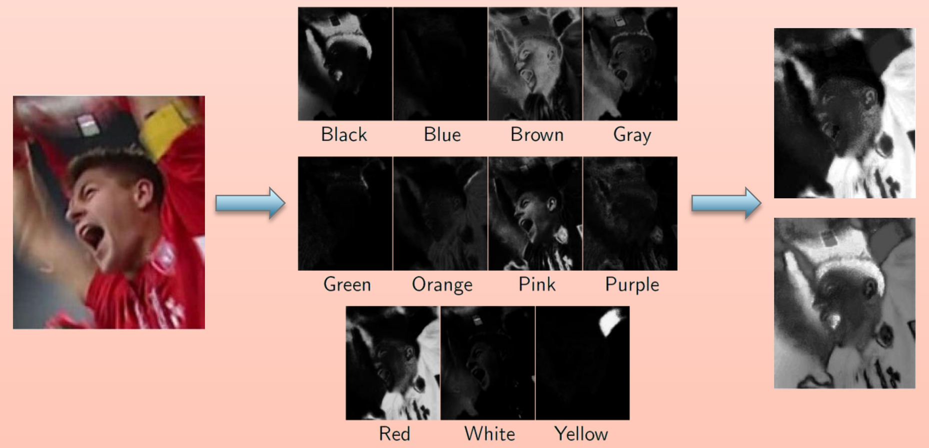
- Improves with most color features.



Dynamic Selection of Color Features

Idea

Reduce the number of color feature dimensions by projecting onto a linear subspace, parameterized by an ON-basis B_p ($D_1 \times D_2$ matrix). Dynamically adaptive.



Total cost

$$\eta_{\text{tot}}^p = \alpha_p \eta_{\text{data}}^p + \sum_{j=1}^{p-1} \alpha_j \eta_{\text{smooth}}^j$$

Data term

Reconstruction error of the appearance $\hat{x}^p$.

$$\eta_{\text{data}}^p = \frac{1}{MN} \sum_{m,n} \left\| \hat{x}^p(m,n) - B_p B_p^T \hat{x}^p(m,n) \right\|^2$$

Smoothness term

Amount of subspace change between frames.

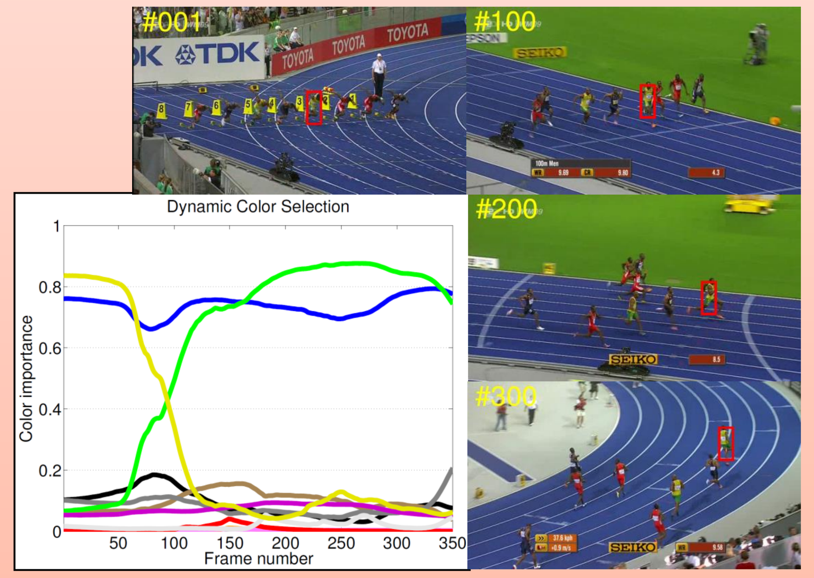
$$\eta_{\text{smooth}}^j = \sum_{k=1}^{D_2} \lambda_j^{(k)} \left\| b_j^{(k)} - B_p B_p^T b_j^{(k)} \right\|^2$$

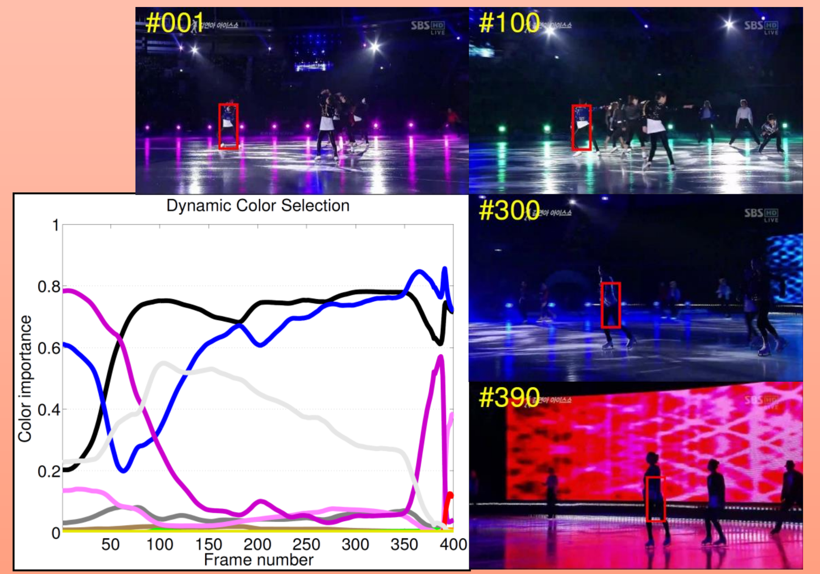
Algorithm 1 Adaptive projection matrix computation.

Input:
Frame number p; Learned object appearance $\hat{x}^p$
Previous covariance matrix Q_{p-1}; Parameters μ, D_2

Output:
Projection matrix B_p; Current covariance matrix Q_p

1: Set $\bar{x}^p = \frac{1}{MN} \sum_{m,n} \hat{x}^p(m,n)$
2: Set $C_p = \frac{1}{MN} \sum_{m,n} (\hat{x}^p(m,n) - \bar{x}^p)(\hat{x}^p(m,n) - \bar{x}^p)^{\mathrm{T}}$
3: **if** $p = 1$ **then**
4: Set $R_1 = C_1$
5: **else**
6: Set $R_p = (1-\mu)Q_{p-1} + \mu C_p$
7: **end if**
8: Do EVD $R_p = E_p S_p E_p^{\mathrm{T}}$, with sorted eigenvalues in S_p
9: Set B_p to the first D_2 columns in E_p
10: Set $[\Lambda_p]_{i,j} = [S_p]_{i,j}$, $1 \le i, j \le D_2$
11: **if** $p = 1$ **then**
12: Set $Q_1 = B_1 \Lambda_1 B_1^{\mathrm{T}}$
13: **else**
14: Set $Q_p = (1-\mu)Q_{p-1} + \mu B_p \Lambda_p B_p^{\mathrm{T}}$
15: **end if**

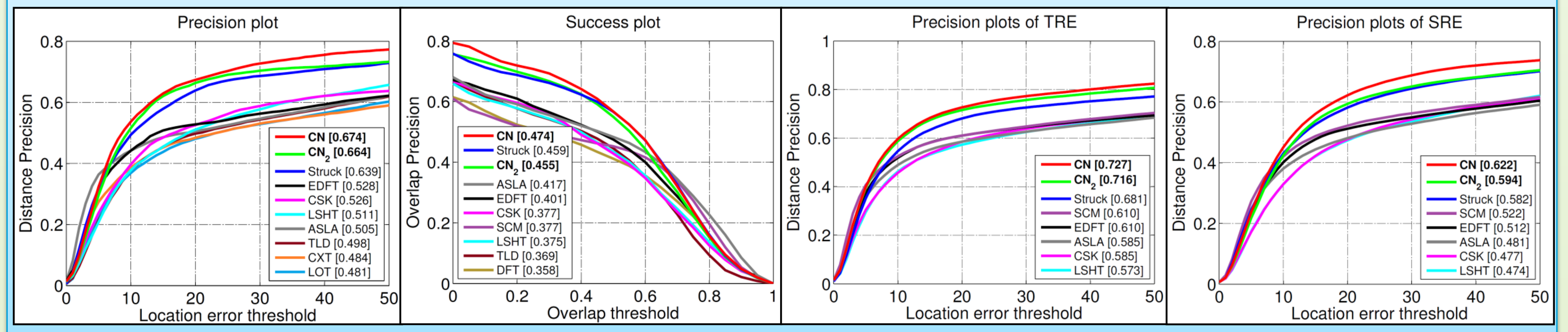




Evaluation

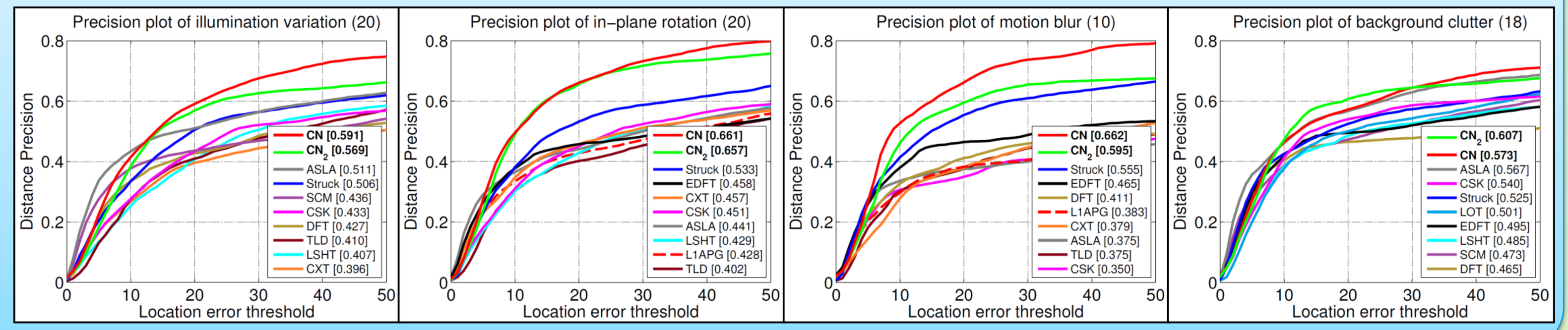
- We use the benchmark protocol of Wu et al. [2].
- 41 videos including all 35 color videos from [2].
- Compared with 15 state-of-the-art trackers.

Quantitative Evaluation

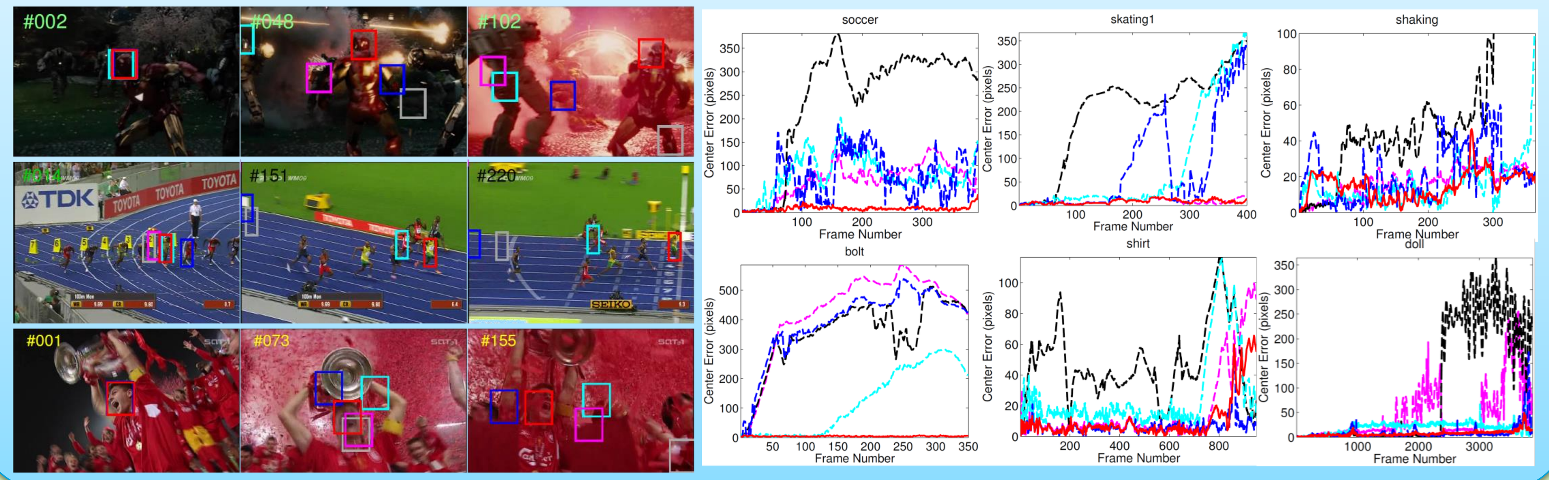


	TLD	EDFT	ASLA	CSK	SCM	LOT	CXT	Struck	LSHT	Adaptive CN
Median FPS	20.7	19.7	0.946	151	0.0862	0.467	11.3	10.4	12.5	*105*

Attribute-based Evaluation



Qualitative Evaluation



Conclusions

- Our approach outperforms state-of-the-art at over 100 fps.
- Color significantly improves tracking performance.
- The choice of color representation is crucial for tracking.

References

[1] J. Henriques, R. Caseiro, P. Martins, and J. Batista. Exploiting the circulant structure of tracking-by-detection with kernels. In ECCV, 2012.
[2] Y. Wu, J. Lim, and M.-H. Yang. Online object tracking: A benchmark. In CVPR, 2013.
[3] J. van de Weijer, C. Schmid, J. J. Verbeek, and D. Larlus. Learning color names for real-world applications. TIP, 18(7):1512–1524, 2009.