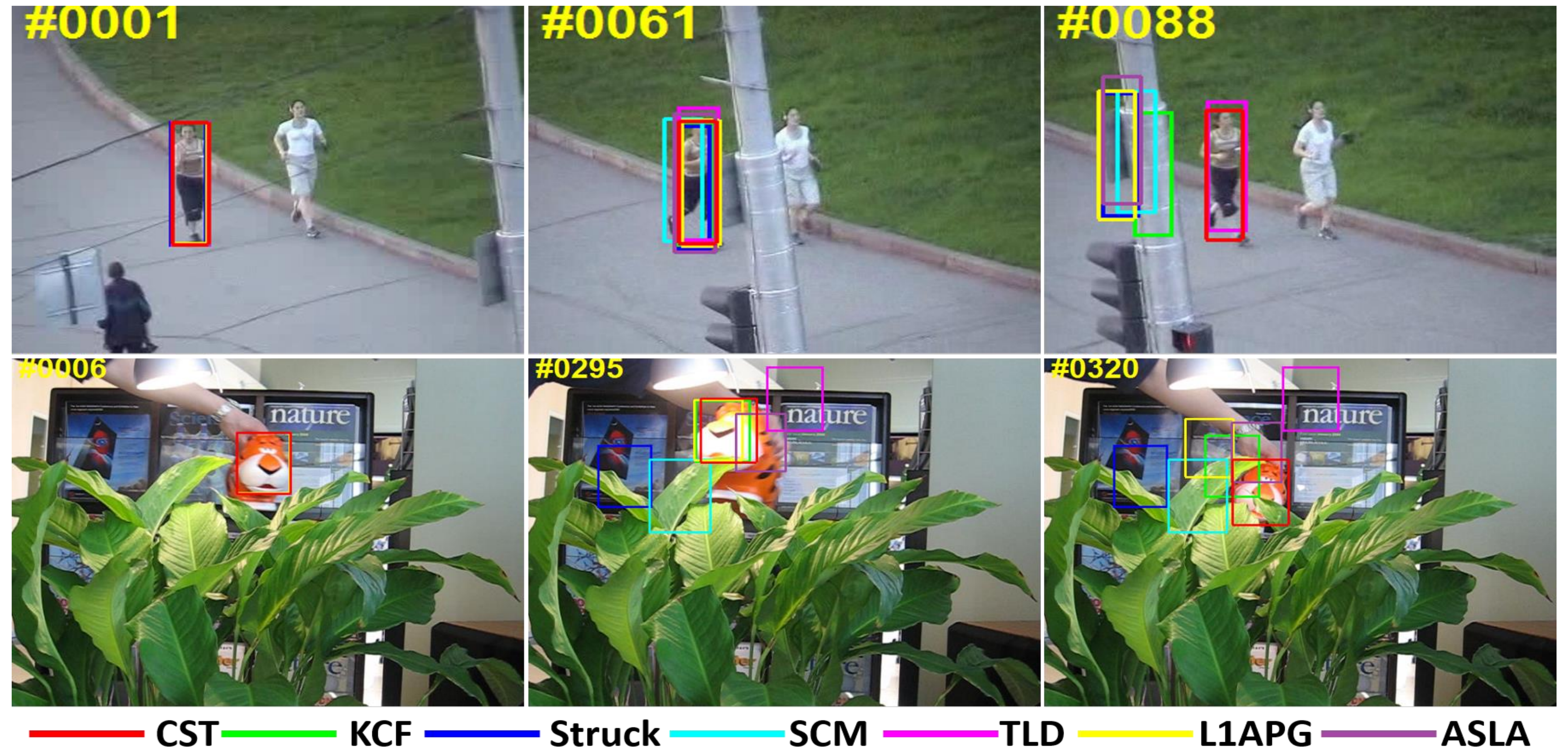
In Defense of Sparse Tracking: Circulant Sparse Tracker

Tianzhu Zhang[1,2], Adel Bibi[1], Bernard Ghanem[1]

IEEE 2016 Conference on Computer Vision and Pattern Recognition
CVPR 2016

Motivation

- Most sparse representation based trackers within the particle filter framework have **high computational cost**, **less than promising tracking performance**, and **limited feature representation**.



- To deal with the above issues, we propose a novel **circulant sparse tracker** (**CST**), which exploits circulant target templates. Because of the circulant structure property, CST can **refine and reduce particles**, be solved efficiently in the **Fourier domain**, and make use of **high dimensional features**.

Sparse based Tracking

P1: Have **high computational cost**	P2: **Less** than promising tracking **performance**	P3: **Limited** feature representation

Circulant Sparse Tracker

A1: Reduce particles using **circular shifts**	A2: Solve efficiently in the **Fourier domain**	A3: Adopt **high dimensional features**

Circulant Sparse Tracker (CST)

- **Problem Formulation**

Primal Formulation

$$\min_{\mathbf{c}} \frac{1}{2}\|\mathbf{x}-\mathbf{A}\mathbf{c}\|_2^2+\lambda\|\mathbf{c}\|_1$$

$$\mathbf{A}=[\mathbf{A}_1,\ldots,\mathbf{A}_k,\ldots,\mathbf{A}_K]$$

$\mathbf{A}_k \in \mathbb{R}^{d\times d}$ **Circulant**

Dual Formulation

$$\min_{\mathbf{z}} \quad \frac{1}{2}\mathbf{z}^\top\mathbf{z}+\mathbf{z}^\top\mathbf{x}$$

$$s.t. \quad \left\|\mathbf{A}^\top\mathbf{z}\right\|_\infty \leq \lambda$$

$$\mathbf{A}\in\mathbb{R}^{d\times Kd}$$

$$\mathcal{L}(\mathbf{c},\mathbf{z},\theta)=\frac{\mathbf{z}^\top\mathbf{z}}{2}+\mathbf{z}^\top\mathbf{x}+\mathbf{c}^\top(\mathbf{A}^\top\mathbf{z}-\theta)+\frac{u}{2}\left\|\mathbf{A}^\top\mathbf{z}-\theta\right\|_2^2+\mathbb{1}_{\{\|\theta\|_\infty\leq\lambda\}}$$

- **Optimization**

Time Domain

$$\mathbf{z}=(\mathbf{A}\mathbf{A}^\top+\frac{1}{u}\mathbf{I})^{-1}(\mathbf{A}\theta-\frac{1}{u}\mathbf{x}-\frac{1}{u}\mathbf{A}\mathbf{c})$$

$$\theta=\arg\min_{\theta}\frac{u}{2}\left\|\mathbf{A}^\top\mathbf{z}-\theta\right\|_2^2-\mathbf{c}^\top\theta$$

$$\Rightarrow \theta=\mathcal{P}_{\mathcal{B}_\lambda^\infty}(\mathbf{A}^\top\mathbf{z}+\frac{\mathbf{c}}{u})$$

$$\mathbf{c}=\mathbf{c}+u(\mathbf{A}^\top\mathbf{z}-\theta)$$

Fourier Domain

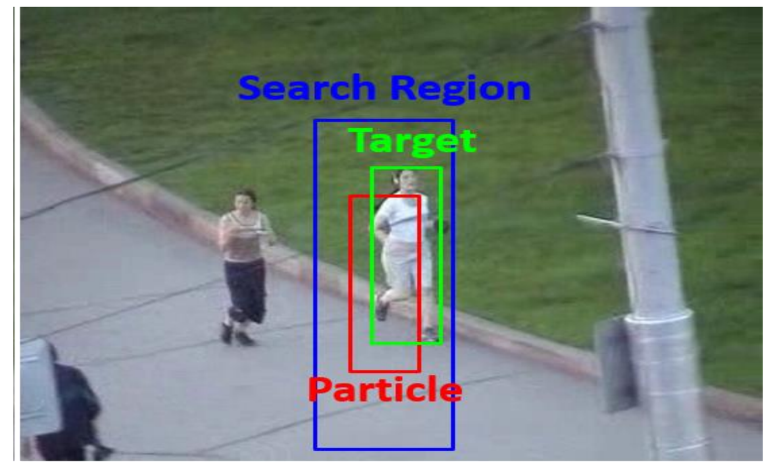
$$\hat{\mathbf{z}}=\frac{\sum_{k=1}^{K}(\hat{\mathbf{a}}_k\odot\hat{\theta}_k-\frac{1}{u}\hat{\mathbf{a}}_k\odot\hat{\mathbf{c}}_k)-\frac{1}{u}\hat{\mathbf{x}}}{\sum_{k=1}^{K}\hat{\mathbf{a}}_k\odot\hat{\mathbf{a}}_k^*+\frac{1}{u}}$$

$$\theta=\mathcal{P}_{\mathcal{B}_\lambda^\infty}(\mathcal{F}^{-1}[\hat{\mathbf{a}}_1^*\odot\hat{\mathbf{z}}+\frac{1}{u}\hat{\mathbf{c}}_1;\ldots;\hat{\mathbf{a}}_K^*\odot\hat{\mathbf{z}}+\frac{1}{u}\hat{\mathbf{c}}_K])$$

$$\mathbf{c}=\mathbf{c}+u(\mathcal{F}^{-1}[\hat{\mathbf{a}}_1^*\odot\hat{\mathbf{z}}-\hat{\theta}_1;\ldots;\hat{\mathbf{a}}_K^*\odot\hat{\mathbf{z}}-\hat{\theta}_K])$$

Here, the inverse Fourier transform is for each k.

- **Particle Refinement Strategy**



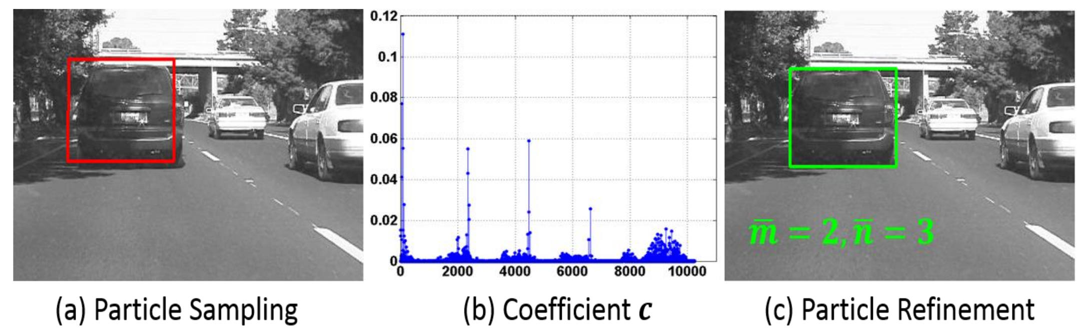
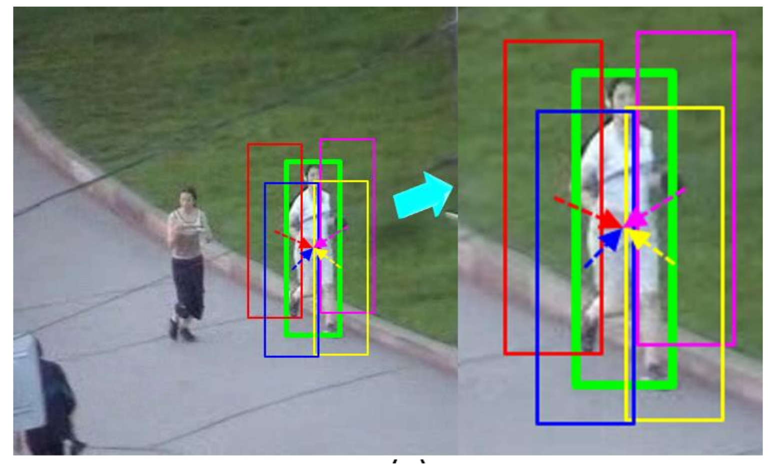
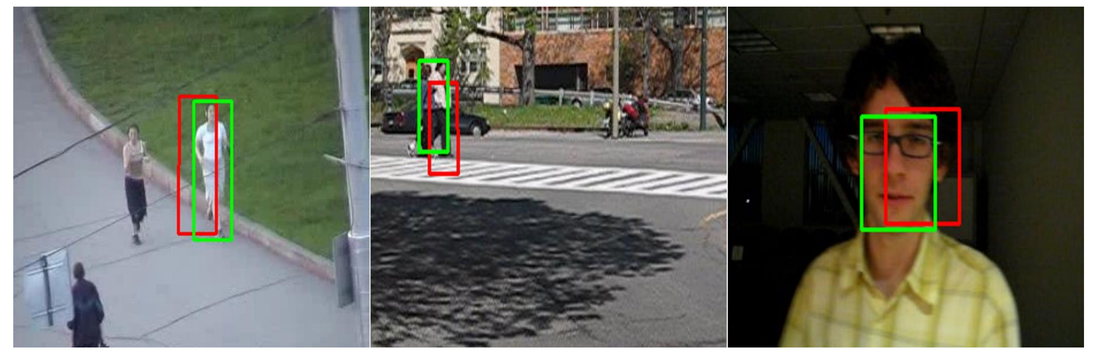


Illustration for particle refinement: (a) the original sampled particle in red, (b) the learned coefficient c, and (c) the refined particle in green with a circular shift (m = 2; n = 3). Here, K = 5 and x is 41x50x31 using HOG.



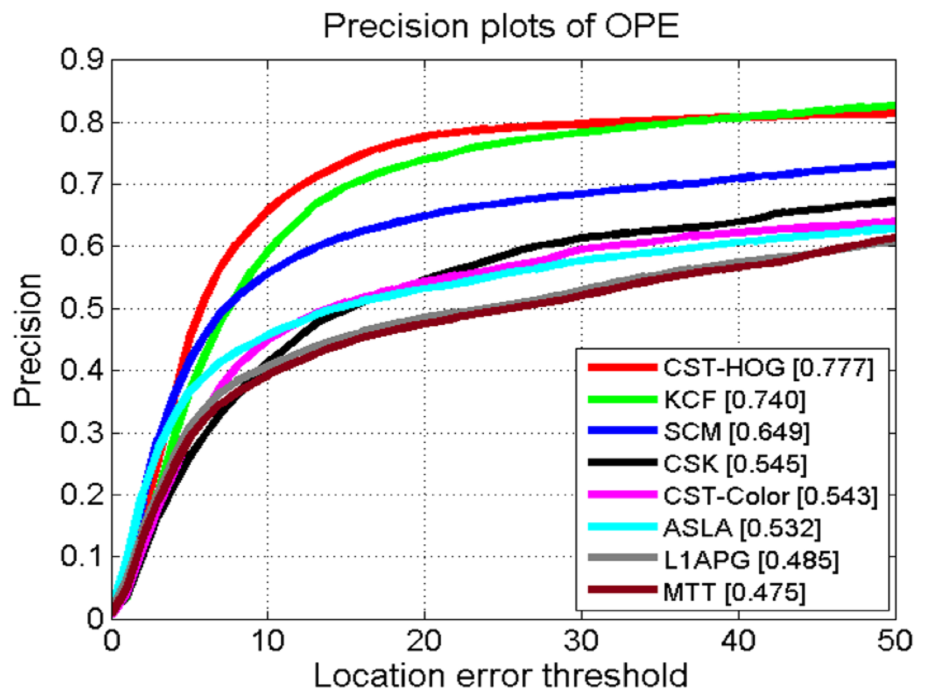
Particle reducing strategy via search region padding and particle refinement.

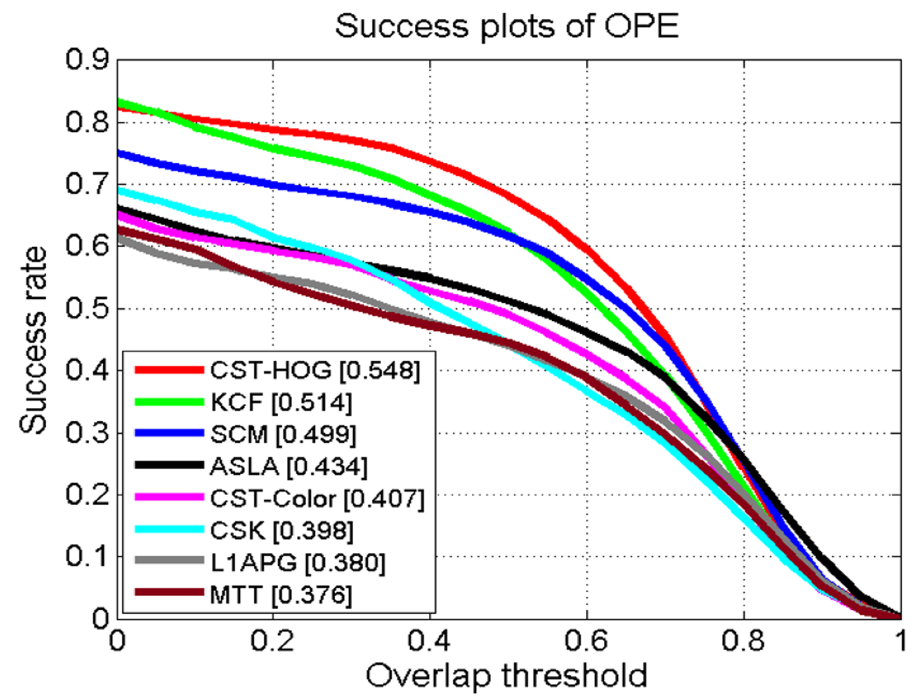


Examples on three video sequences to show particle refinement via the proposed CST. The bounding boxes with red color are the sampled particles. Due to random sampling, **the sampled particles are far away from the target object**. With the proposed CST method, these particles can be refined and **translated to better state** denoted with bounding boxes with green color.

Experiments:

- **Image Feature Evaluation**: Our results clearly suggest that the **HOG** based image representation **improves** the tracking performance, which is also demonstrated by comparing KCF to CSK.





- **Sparse Tracking Evaluation**: Our approach performs favorably against existing methods in overlap success (OS) (%), distance precision (DP) (%) and center location error (CLE) (in pixels).

	CST-HOG	CST-Color	L1APG [4]	SCM [45]	ASLA [19]	MTT [38]
OS	68.2	48.9	44.0	61.6	51.1	44.5
DP	77.7	54.3	48.5	64.9	53.2	47.5
CLE	40.4	86.2	77.4	54.1	73.1	94.5
FPS	2.2	3.0	2.4	0.4	7.5	1.0

- **Comparison with State-of-the-Art**