ADAPTIVE VISUAL TARGET TRACKING BASED ON LABEL CONSISTENT K-SVD SPARSE CODING AND KERNEL PARTICLE FILTER

Jinlong Yang[1], Xiaoping Chen[1], Yu Hen Hu[2], Jianjun Liu[1]

1.Jiangnan University 2.University of Wisconsin–Madison

ABSTRACT

We propose an adaptive visual target tracking algorithm based on label consistent K-SVD(LC-KSVD) dictionary learning:

- LC-KSVD is applied to local patches to simultaneously estimate a set of low-dimension dictionary and classification parameters.
- To track the target over time, a kernel particle filter is proposed to integrates both local and global motion information of the target.
- An adaptive template updating scheme is also developed to improve the robustness of the tracker.

ALGORITHM FRAMEWORK

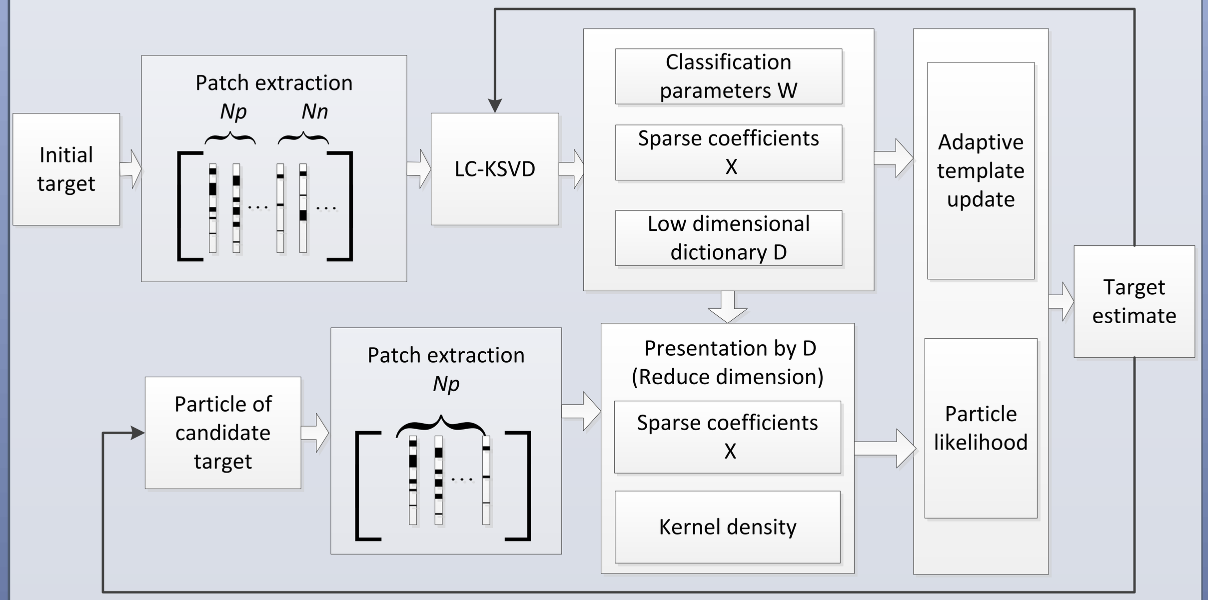


Image patches of both the foreground and background image will be extracted from target image and then a kernel density based particle filter is applied to deduce a sparse representation using the low dimension dictionary. To account for potential occlusion of the target, we introduce a detection scheme of sparse coefficient histogram matrix and design a adaptive parameter model for the proposed template update scheme, which can improve the robustness of the tracker.

PROPOSED METHOD

1 LC-KSVD Learning and Kernel Particle Filter

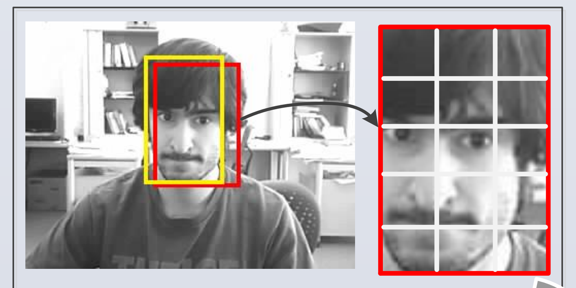
1.1 LC-KSVD

The LC-KSVD dictionary learning algorithm can simultaneously train an over-complete dictionary and a linear classifier:

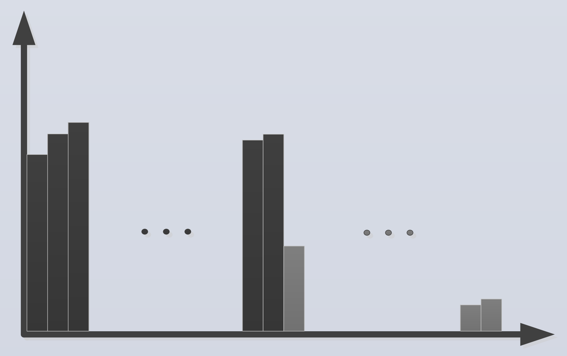
$$\langle D,W,A,X\rangle = \underset{D,W,A,X}{\arg\min} \|Y-DX\|_2^2 + \alpha\|Q-AX\|_2^2 + \beta\|H-WX\|_2^2,\ \ s.t.\ \forall i, \|x_i\| \le T$$

Where Y is the observation matrix, D is the dictionary matrix, Q is the sparse codes with discriminative power of Y for classification, W is the binary classifier(correspondes to target and background).

After extracting patches along the template border, we get the **dictionary response histogram** and **patch-based classification results** by using dictionary and classifier above.

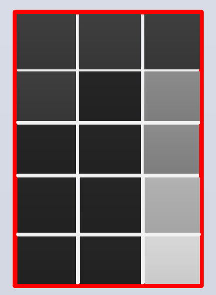


Candidate Target Patch



Dictionary Response Histogram

The darker color represents the response of target template base and the other one corresponds to the background template base.



Classification Results of Patches

The darker the patch is, the closer it is to the target.

1.2 Kernel particle filter

The likelihood function of candidate target is

$$p = \sum_{i=1}^{N_p} k\left(\left\|(y-c_i)/h\right\|^2\right) L_i M_i$$

- $L_i = \sum \min(\varphi_i, \psi^i)$ is the similariity function of the **response histogram** between the candidate and the template, φ_i and ψ^i are the sparse coefficient histogram matrices of the candidate target and the target template.
- $M_i = \cos\langle W\varphi_i, \Gamma\rangle$ is **classification results** in the previous section, $\Gamma = [1,0]^{\mathrm{T}}$ is the base vector of target classification.
- $k\left(\left\|(y-c_i)/h\right\|^2\right)$ is the patch-based **Gaussian Kernel Density**, which is used to assign smaller weights to the patches far away from the center of the target. c_i denotes the center of the ith patch, y denotes the center of the candidate target.

2 Adaptive Template Update

The template is updated using a linear combination of the histogram of the old template ψ and the latest estimated tracking result $\hat{\varphi}_n$:

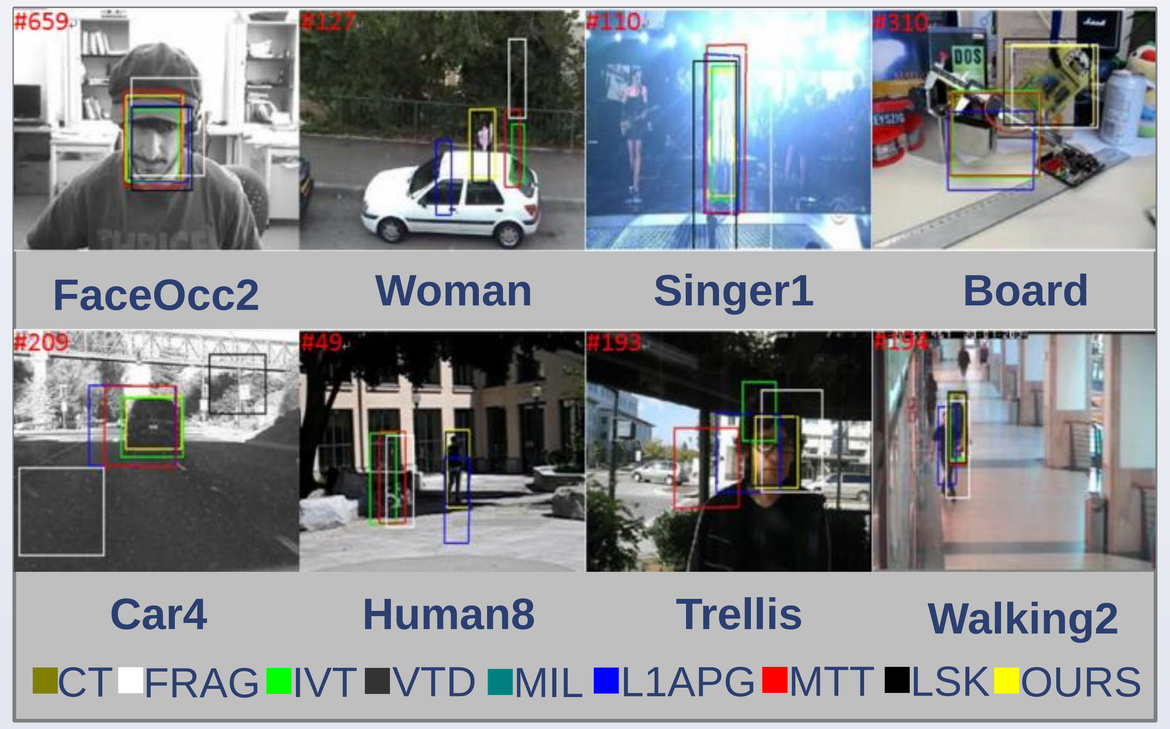
$$\hat{\psi}_n = \begin{cases} \mu\psi + (1-\mu)\hat{\varphi}_n, & O_n < O_0 \\ \hat{\psi}_{n\text{-}1}, & otherwise \end{cases}$$

Where $\mu = \exp((O_n/O_o)-1)$ is an adaptive weight parameter. $O_n = (\#\ \text{occluded patches})/(\#\text{all patches})$ and O_0 is a threshold of the occlusion degree.

Remark. With increasing number of occluded patches, *On* and hence μ will increase so the latest template $\hat{\varphi}_n$ will be given smaller weight because it is less reliable.

RESULTS(1)

Tracking Results



Average Center Location Error

	IVT	FRAG	L1APG	MTT	LSK	OURS
FaceOcc2	<u>6.9</u>	15.7	12.9	10.2	14.7	**4.88**
Woman	172.6	<u>109.7</u>	126.7	134.8	131.6	**4.43**
Singer1	<u>11.5</u>	91.5	53.1	35.9	21.2	**2.48**
Board	162.2	84.5	184.4	159.2	<u>45.4</u>	**12.08**
Car4	<u>4.08</u>	263.1	153.98	45.25	133.23	**3.89**
Human8	85.96	74.83	54.17	76.42	<u>2.74</u>	**2.18**
Trellis	119.57	59.51	62.30	68.99	<u>4.70</u>	**3.85**
Walking2	<u>3.04</u>	57.53	4.52	3.48	18.95	**2.84**

Success rate

	IVT	FRAG	L1APG	MTT	LSK	OURS
FaceOcc2	0.73	0.66	0.68	<u>0.75</u>	0.64	**0.82**
Woman	0.16	0.16	0.17	<u>0.18</u>	0.17	**0.74**
Singer1	<u>0.59</u>	0.23	0.32	0.37	0.37	**0.87**
Board	0.15	0.55	0.11	0.16	<u>0.65</u>	**0.83**
Car4	<u>0.88</u>	0.19	0.26	0.45	0.15	**0.89**
Human8	0.06	0.10	0.16	0.10	<u>0.69</u>	**0.74**
Trellis	0.25	0.29	0.20	0.22	<u>0.66</u>	**0.71**
Walking2	0.76	0.28	<u>0.78</u>	**0.81**	0.47	0.75

Eight challenging video sequences drawn from the public visual tracking datasets are used to examine the performance of the proposed algorithm.

RESULTS(2)

For the proposed algorithm with adaptive parameter μ, it can obtain an ideal tracking result without manually setting the parameter values.

	Criteria	μ				
		0.1	0.4	0.7	0.9	Adptive
FaceOcc2 1~200(f)	ACLE	**4.48**	4.53	4.65	4.81	4.59
	SR	**0.85**	0.85	0.84	0.84	0.84
Woman 1~170(f)	ACLE	15.66	5.80	4.39	4.32	**2.87**
	SR	0.58	0.80	0.82	0.82	**0.84**

Remark. Two evaluation criteria are <u>average center location error</u> (ACLE) and tracking <u>success rate</u> (SR).

CONCLUSION

- The template sets constructed by the local patch features from both foreground and background of the target are used to learn the low dimensional dictionary and classification parameters
- Propose an effective kernel particle filter to extract target
- An adaptive template update scheme is designed