BIT: Bio-inspired Tracker

Bolun Cai, Xiangmin Xu, Xiaofen Xing, Chunmei Qing
School of Electronic and Information Engineering
South China University of Technology, Guangzhou, China
caibolun@gmail.com,{xmxu,xfxing,qchm}@scut.edu.cn

ABSTRACT

Visual tracking is a challenging problem due to various factors such as deformation, rotation and illumination. As is well known, given the superior tracking performance of human vision, bio-inspired model is expected to improve the computer visual tracking. According to the ventral stream in visual cortex, a novel bioinspired tracker (BIT) is proposed, which simulates shallow neurons (S1 and C1) to extract low-level bio-inspired feature for target appearance and imitates senior learning mechanism (S2 and C2) to combine generative and discriminative model for position estimation. In addition, Fast Fourier Transform (FFT) is adopted for real-time learning and detection in this framework.

Introduction

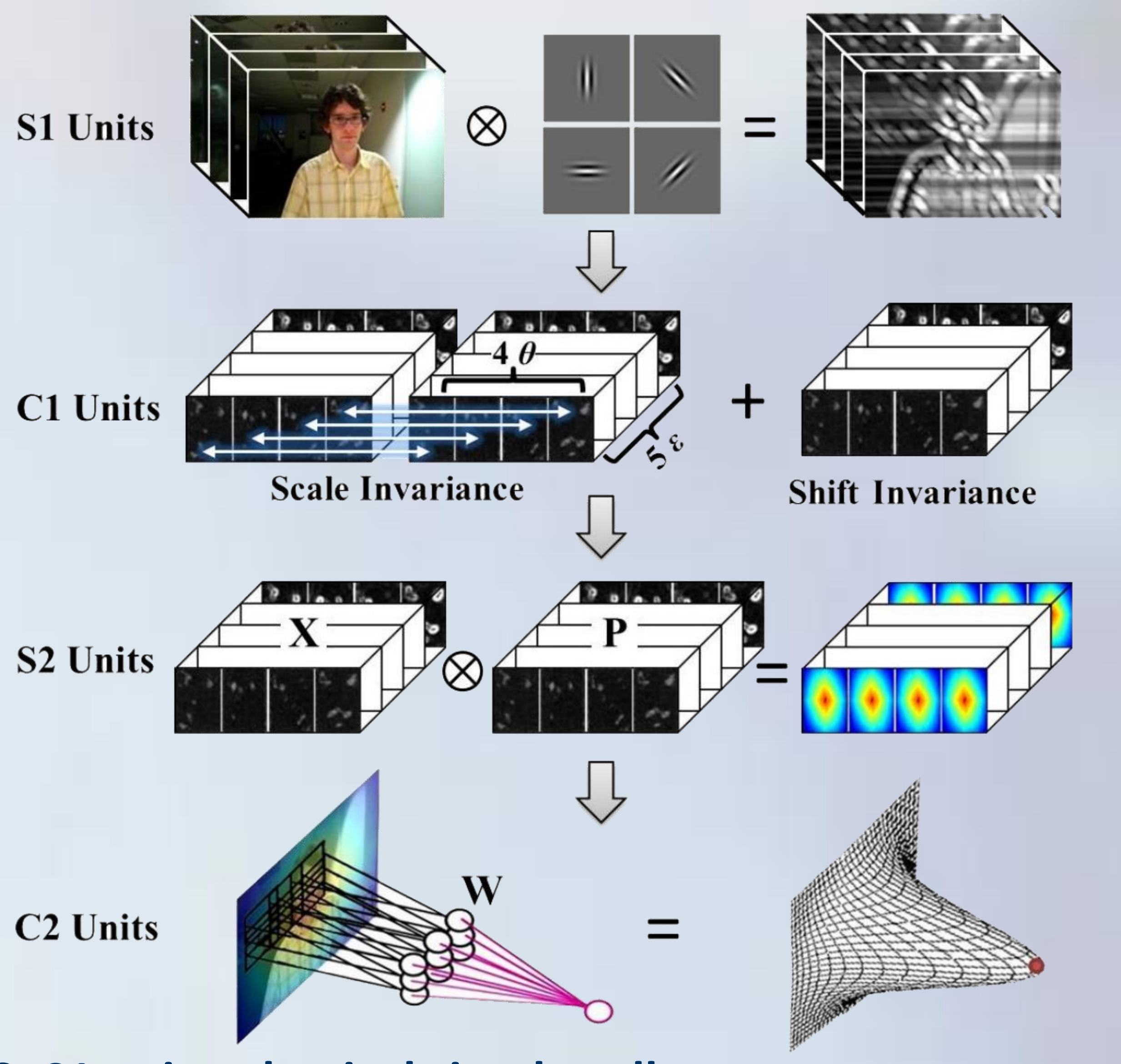
Visual object tracking is an important problem of computer vision, with wide-ranging applications including video surveillance, human-machine interfaces and robot perception. Recent tracking algorithms can be split into two main modules generally: ***feature extraction*** and ***tracking model***.

Feature extraction	
Handcrafted	Handcrafted feature designing is difficult, depending on the time-consuming parameter adjustment.
Automated	Automated feature requires a good underlying model and decrease the real-time performance.
Tracking model	
Generative	Generative models search for the most similar region to the target object within a neighborhood.
Discriminative	Discriminative models treat tracking as a classification problem to distinguish the target from the background.

- A bio-inspired feature extraction: a low-level feature simulate S1 and C1 units to exhibit a trade-off between invariance and discrimination.
- A bio-inspired tracking model: the S2 convolution units is a generative model and the C2 neuronal connection as a discriminative model.
- Importantly, the model exploits FFT to speed up the bio-inspired model and dense sampling.

Bio-inspired Tracker

According to the primate visual pathway, the bio-inspired tracker includes ***S1 units*** modeling primary visual cortex, ***C1 units*** simulating cortical complex cells, ***S2 units*** and ***C2 units*** corresponding to the learning mechanism of senior neurons.



- **S1 units: classical simple cells**

In the primary visual cortex (V1), simple cell (S1) has multi-directional, multiscale and multi-frequency selections, and can be described as Gabor filters:

$$G\left(x,y,\theta,s(\delta,\lambda,\gamma)\right)=\exp\left(-\frac{\left(X^2+\gamma^2Y^2\right)}{2\sigma^2}\right)\times\cos\left(\tfrac{2\pi}{\lambda}X\right)$$
$$s.t.\quad X=x\cos\theta+y\sin\theta, Y=-x\cos\theta+y\sin\theta$$
$$S1\left(x,y,\theta,s\right)=I\left(x,y\right)\otimes G\left(x,y,\theta,s\right)$$

- **C1 units: cortical complex cells**

C1 units correspond to complex cortical complex cells (V2) showing the invariance to scale and shift:

$$C1\left(x,y,\theta,\varepsilon\right)=\max_{(x,y)\in\Sigma}\left(\max_{s=\{2\varepsilon,2\varepsilon-1\}}S1\left(x,y,\theta,s\right)\right)$$

- **S2 units: shape-tuned learning**

Shape-tuned learning from V2 to IT as a generative model, in which S2 units depends in an RBF distance between a new input X and a stored prototype P:

$$r=\exp\left(-1/_{2\sigma^2}\left\|X-P\right\|^2\right)$$
$$\sim\exp\left(X^TP\right)\sim X^TP$$

- **C2 units: task-dependent learning**

An CNN corresponding to the task-specific circuits with neurons from IT to PFC for the discrimination between target objects and background clusters as

$$C2\left(x,y\right)=\sum_{\theta,\varepsilon}W\left(x,y\right)\otimes S2\left(x,y,\theta,\varepsilon\right)$$

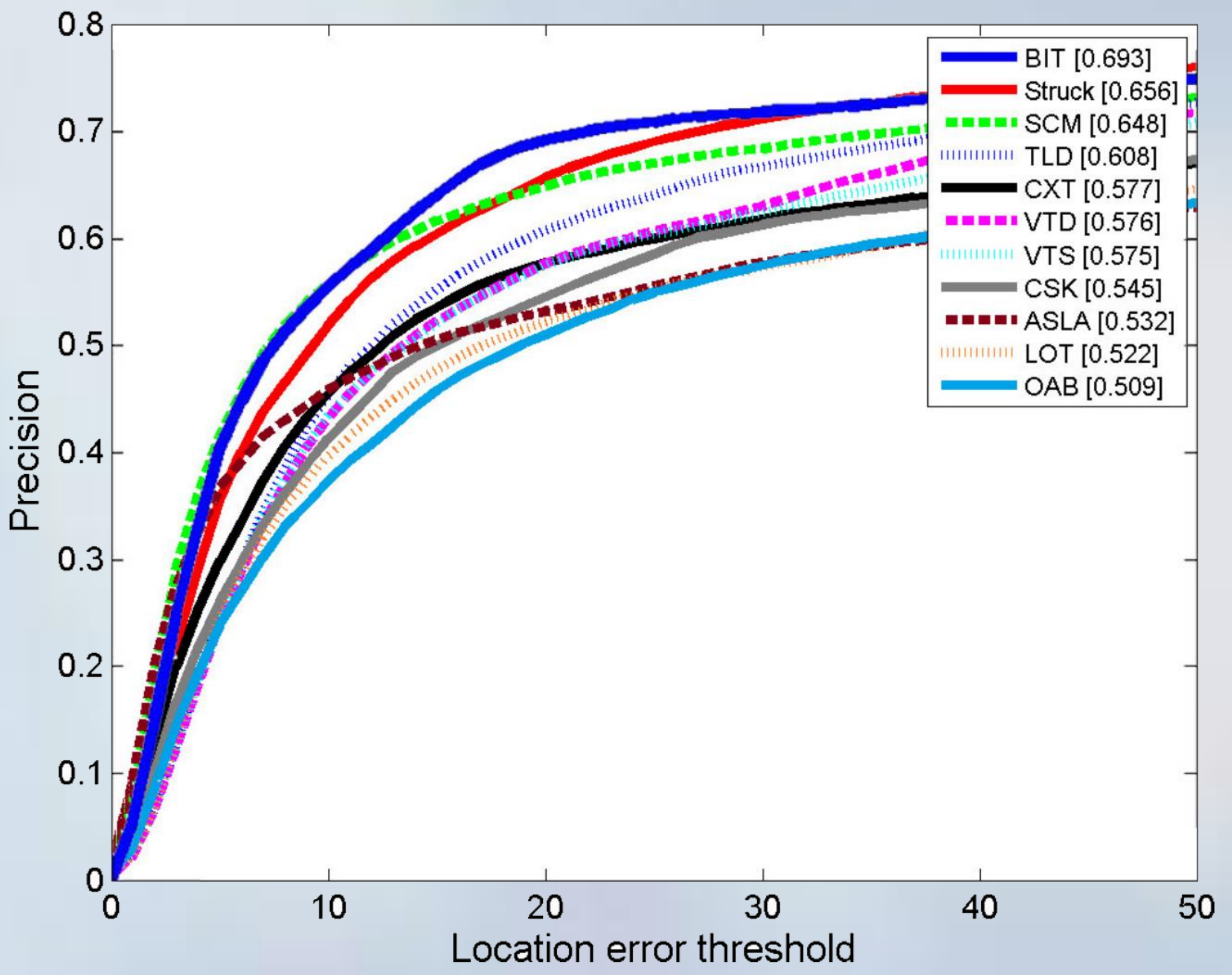
- **Real-time bio-inspired tracker via FFT**

The real-time performance is an important index of object visual tracking method. Therefore, the FFT speeds up the dense sampling of S2 and C2 response calculation in the real-time BIT.

$$\mathcal{F}\left[S2_{t+1}\left(\cdot,\theta,\varepsilon\right)\right]=\mathcal{F}\left[C1_{t+1}\left(\cdot,\theta,\varepsilon\right)\right]\odot\mathcal{F}\left[C1_t^P\left(\cdot,\theta,\varepsilon\right)\right]$$
$$C2\left(x,y\right)=\exp\left(-\frac{1}{2\sigma_s^2}\left(\left(x-x_o\right)^2+\left(y-y_o\right)^2\right)\right)$$
$$\mathcal{F}\left[W\left(x,y\right)\right]=\sum_{\theta,\varepsilon}\frac{\mathcal{F}\left[C2\left(x,y\right)\right]}{\mathcal{F}\left[S2\left(x,y,\theta,\varepsilon\right)\right]}$$
$$\left(\hat{x},\hat{y}\right)=\arg\max_{(x,y)}C2_{t+1}\left(x,y\right)$$

EXPERIMENTAL RESULTS

	BIT	Struck[11]	SCM[22]	TLD[12]	CXT[13]	VTD[8]	VTS[9]	CSK[23]	ASLA[10]
ALL	**0.693**	0.656	0.648	0.608	0.577	0.576	0.575	0.545	0.532
IV	**0.607**	0.558	0.592	0.537	0.505	0.557	0.572	0.481	0.516
SV	0.652	0.639	**0.672**	0.606	0.550	0.597	0.582	0.503	0.552
OCC	0.631	0.565	**0.639**	0.563	0.494	0.546	0.533	0.500	0.460
DEF	**0.612**	0.521	0.586	0.512	0.422	0.501	0.487	0.476	0.445
MB	0.537	**0.551**	0.339	0.518	0.509	0.375	0.375	0.342	0.278
FM	0.502	**0.604**	0.331	0.551	0.519	0.353	0.351	0.381	0.253
IPR	**0.620**	0.617	0.596	0.584	0.612	0.600	0.578	0.547	0.511
OPR	**0.629**	0.597	0.617	0.596	0.576	0.620	0.603	0.540	0.518
OV	0.388	0.539	0.429	**0.576**	0.510	0.462	0.455	0.379	0.333
BC	**0.689**	0.585	0.578	0.428	0.443	0.571	0.578	0.585	0.496
LR	0.516	**0.545**	0.305	0.349	0.371	0.168	0.187	0.411	0.156



Multi-direction Gabor filters in ***S1*** contribute to the robustness of illumination (IV) and rotation (IPR and OPR). ***C1*** provide the scale and shift competitions for scale variation (SV) and deformation (DEF). The generative model in ***S2*** and the discriminative model in ***C2*** rise to the challenges of occlusion (OCC) and background clutters (BC) respectively.

This work is supported by the National Natural Science Foundation of China (#61171142, #61401163), the Science and technology Planning Project of Guangdong Province of China (#2011A010801005, #2012B061700102).