VALSE 2016 武汉

亮风台 格灵深瞳 百度 大疆创新 学习宝 图普科技 通甲优博 top+ 商汤科技 搜狗 驭势科技

BIT: Biologically Inspired Tracker

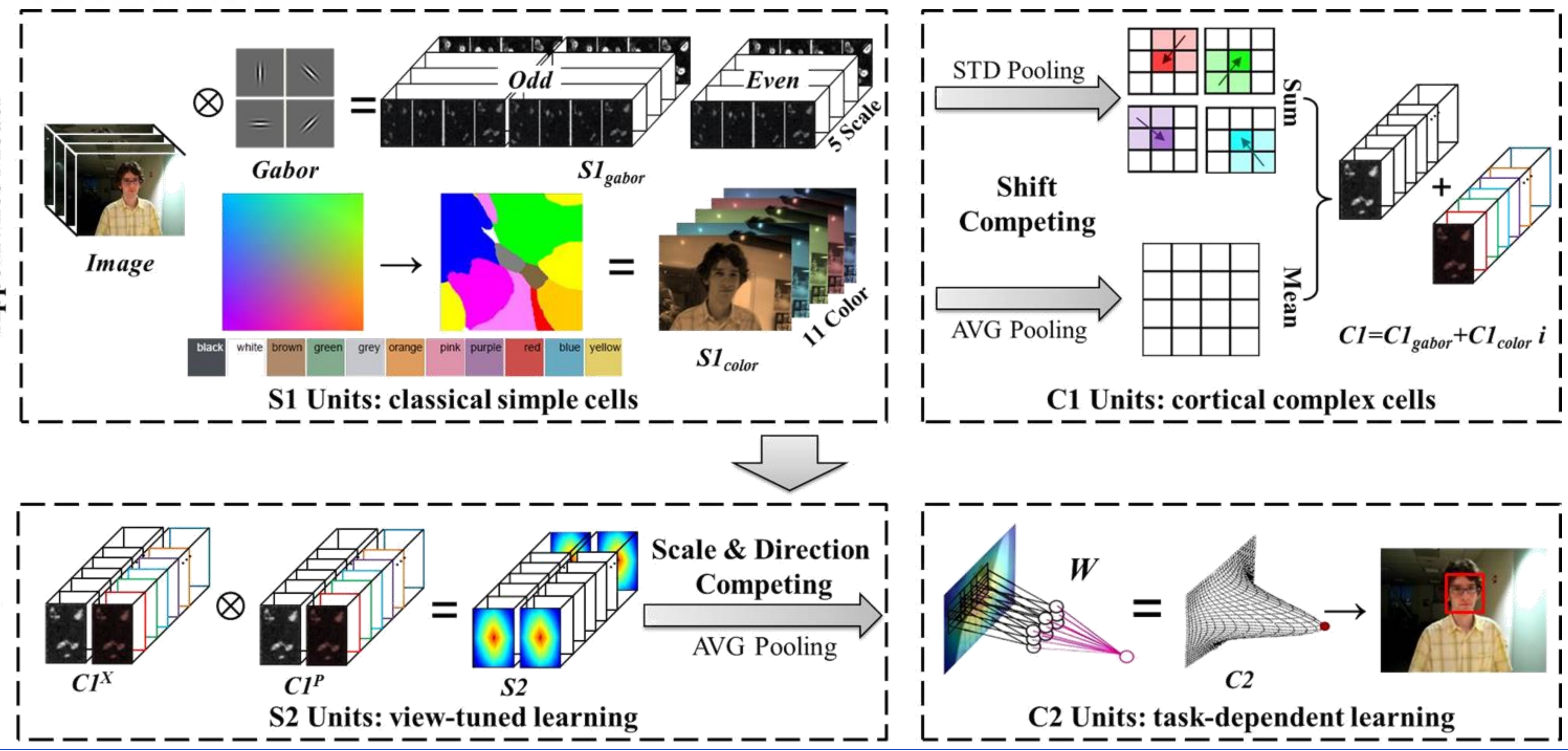
Bolun Cai, Xiangmin Xu, Xiaofen Xing, Kui Jia, Jie Miao, Dacheng Tao

South China University of Technology

IEEE Transactions on Image Processing, 2016

Abstract

Visual tracking is challenging due to various factors. Given the superior tracking performance of human visual system, an ideal design of ***Biologically Inspired Model*** is expected to improve visual tracking. Based on the analysis of the ventral stream in the visual cortex, the ***biologically inspired tracker (BIT)*** simulates shallow neurons (S1 units and C1 units) to extract low-level features for the target appearance and imitates an advanced learning mechanism (S2 units and C2 units) to combine generative and discriminative models for target location. In addition, ***Fast Gabor Approximation (FGA)*** and ***Fast Fourier Transform (FFT)*** are adopted for real-time learning and detection in this framework.



BIT: Biologically Inspired Tracker

BIT

- **S1 units：classical simple cells**

In the primary visual cortex (V1), a simple cell has the characteristics of multi-orientation, multi-scale and multi-frequency selection. and can be described as Gabor filters:

$$\begin{cases} G_{even}\left(x,y,\theta,s\left(\sigma,\lambda\right)\right)=\exp\left(-\frac{X^2+Y^2}{2\sigma^2}\right)\cos\left(\frac{2\pi}{\lambda}X\right) \\ G_{odd}\left(x,y,\theta,s\left(\sigma,\lambda\right)\right)=\exp\left(-\frac{X^2+Y^2}{2\sigma^2}\right)\sin\left(\frac{2\pi}{\lambda}X\right) \end{cases}$$

$$S1_{gabor}\left(x,y,\theta,s\right)=I\left(x,y\right)\otimes G_{even/odd}\left(x,y,\theta,s\right)$$

The color units are inspired by the color double-opponent system in the cortex, and are defined by Color Names:

$$S1_{color}\left(x,y,c\right)=Map\left(R\left(x,y\right),G\left(x,y\right),B\left(x,y\right),c\right)$$

- **C1 units：cortical complex cells**

The cortical complex cells (V2) receive the response from V1 and have the function of linear feature integration.

$$C1_{gabor}\left(x,y\right)=\sum_{(x,y)\in\Sigma}\frac{S1_{gabor}\left(x,y\right)}{N_{\delta_x,\delta_y}\left(x,y\right)} \quad N_{\delta_x,\delta_y}\left(x,y\right)=\left(S1^2_{gabor}\left(x,y\right)+S1^2_{gabor}\left(x+\delta_x,y+\delta_y\right)+S1^2_{gabor}\left(x+\delta_x,y\right)+S1^2_{gabor}\left(x,y+\delta_y\right)\right)^{0.5}$$

- **S2 units：view-tuned learning**

View-tuned learning from V2 to IT as a generative model, in which S2 units is RBF distance between new input X and stored prototype P.

$$r_{S2}=\exp(-\frac{1}{2\sigma^2}\|X-P\|^2)=\exp(-\frac{1}{2}(X^TX+P^TP-2X^TP))\sim\exp(X^TP)\sim X^TP$$

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

An CNN corresponding to task-dependent learning from IT to PFC for the discrimination between target and background as

$$C2\left(x,y\right)=W\left(x,y\right)\otimes S2(x,y)$$

Real-time BIT

- **Fast Gabor Approximation (FGA)**

Using several pairs of 1-D orthogonal Gabor filters G_x, G_y, the approximate response of S1 units is defined as

$$\begin{cases} D_x\left(x,y,s\left(\sigma,\lambda\right)\right)=I\left(x,y\right)\otimes G_x\left(x,s\left(\sigma,\lambda\right)\right) \\ D_y\left(x,y,s\left(\sigma,\lambda\right)\right)=I\left(x,y\right)\otimes G_y\left(y,s\left(\sigma,\lambda\right)\right) \end{cases} \Rightarrow \begin{cases} \Theta\left(\cdot\right)=\tan^{-1}\left(\frac{D_y\left(x,y,s\left(\sigma,\lambda\right)\right)}{D_x\left(x,y,s\left(\sigma,\lambda\right)\right)}\right) \\ A\left(\cdot\right)=\sqrt{D_x^2\left(x,y,s\right)+D_y^2\left(x,y,s\right)} \end{cases}$$

$$S1_{odd}\left(\cdot\right)=\begin{cases} A\left(\cdot\right), if\ \Theta\left(\cdot\right)\in\left[\theta-\pi/8,\theta+\pi/8\right) \\ 0, otherwise \end{cases} \quad S1_{even}\left(\cdot\right)=\begin{cases} A\left(\cdot\right), if\ \Theta\left(\cdot\right)\in\left[\theta-\pi/8,\theta+\pi/8\right)\cup\left[\theta+7\pi/8,\theta+9\pi/8\right) \\ 0, otherwise \end{cases}$$

- **Fast Fourier Transform (FFT)**

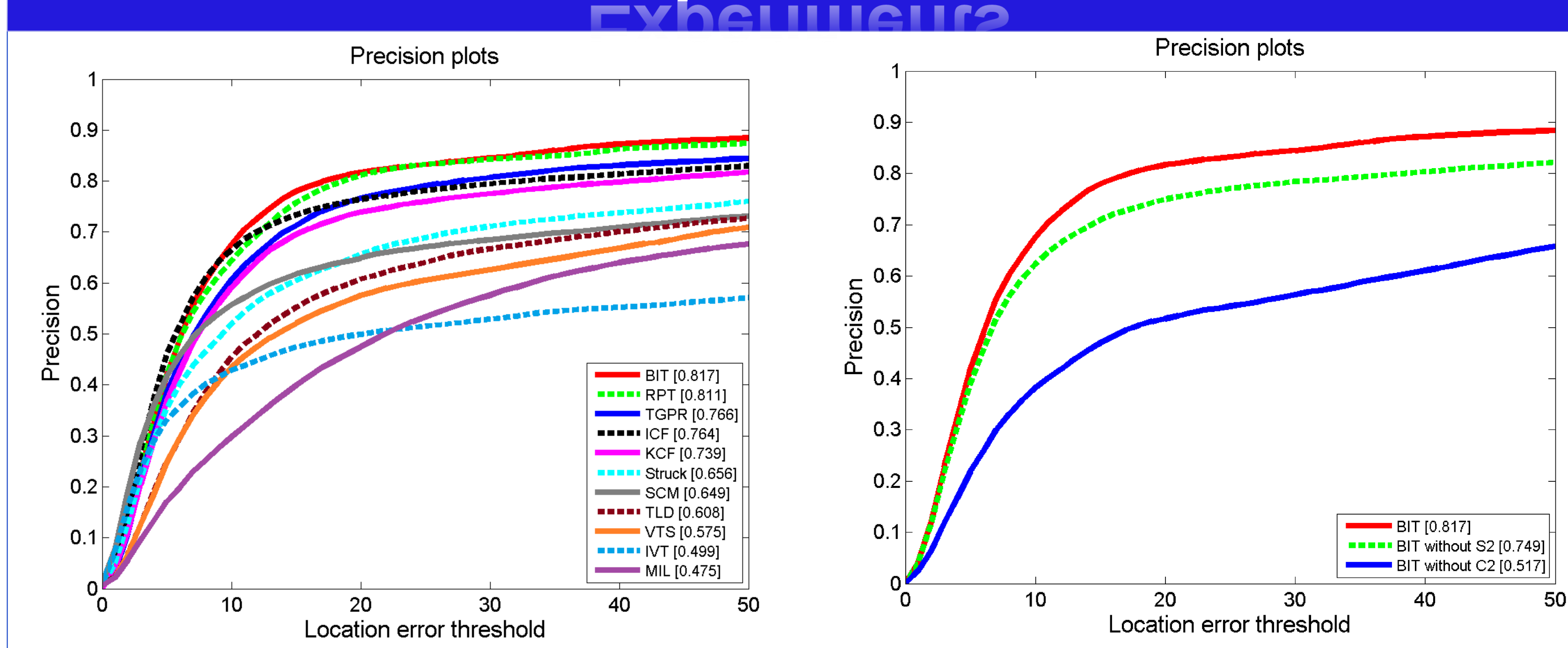
The FFT speeds up the dense sampling of S2 and C2 response calculation in the real-time BIT.

$$\mathcal{F}\left[S2_{t+1}\left(\cdot\right)\right]=\frac{1}{K}\sum_{k=1}^{K}\mathcal{F}\left[C1^X_{t+1}\left(\cdot,k\right)\right]\odot\mathcal{F}\left[C1^P_t\left(\cdot,k\right)\right]$$

$$\widetilde{C2}\left(x,y\right)=\exp\left(-\frac{1}{2\sigma_s^2}\left(\left(x-x_o\right)^2+\left(y-y_o\right)^2\right)\right)$$

Solution：$\mathcal{F}\left[W\left(x,y\right)\right]=\frac{\mathcal{F}\left[\widetilde{C2}\left(x,y\right)\right]}{\mathcal{F}\left[S2\left(x,y\right)\right]}$ Location：$\left(\hat{x},\hat{y}\right)=\arg\max_{(x,y)}C2_{t+1}\left(x,y\right)$

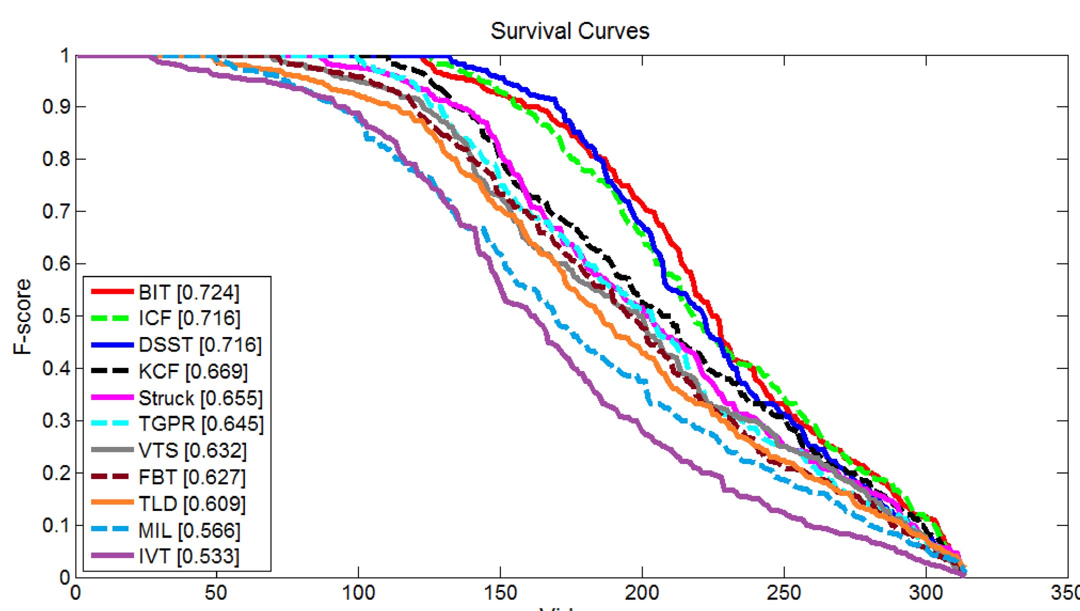
Experiments



	BIT	RPT[58]	TGPR[59]	ICF[60]	KCF[3]	Struck[23]	SCM[12]	TLD[55]	VTS[61]	MIL[25]
IV	0.764	**0.827**	0.687	0.696	0.717	0.558	0.594	0.537	0.573	0.349
SV	0.786	**0.802**	0.703	0.707	0.667	0.639	0.672	0.606	0.582	0.471
OCC	**0.854**	0.765	0.708	0.817	0.744	0.564	0.640	0.563	0.534	0.427
DEF	**0.817**	0.748	0.768	0.754	0.751	0.521	0.586	0.512	0.487	0.455
MB	0.663	**0.783**	0.578	0.654	0.621	0.551	0.339	0.518	0.375	0.357
FM	0.643	**0.745**	0.575	0.612	0.581	0.604	0.333	0.551	0.353	0.396
IPR	0.783	**0.795**	0.706	0.739	0.731	0.617	0.597	0.584	0.579	0.453
OPR	**0.831**	0.807	0.741	0.741	0.724	0.597	0.618	0.596	0.604	0.466
OV	**0.654**	0.641	0.495	0.584	0.555	0.539	0.429	0.576	0.455	0.393
BC	0.789	**0.840**	0.761	0.698	0.725	0.585	0.578	0.428	0.578	0.456
LR	0.369	0.478	0.539	0.516	0.379	**0.545**	0.305	0.349	0.187	0.171

Multi-direction Gabor filters used in S1 units contribute to the robustness of illumination (IV) and rotation (IPR and OPR). ***Pooling operations*** in C1 and S2 units provide the shift and scale competitive to deal with deformation (DEF) and scale (SV). The ***generative model*** in S2 units and the ***discriminative model*** in C2 units rise to the challenges of OCC and OV respectively.

The hybrid-model (81.7%) achieved excellent performances in comparison to single-model (74.9% and 51.7%). In addition, the performance gap between the discriminative model and the generative model in the literature is 23.2%.



Tracker	Speed (fps)	Precision (%)
BIT	44.9	**81.7**
ICF[60]	68.8	76.4
KCF[3]	**284.4**	73.9
TLD[55]	28.1	60.8
MIL[25]	38.1	49.9
IVT[9]	33.4	47.5
RPT[58]	4.1	81.1
TGPR[59]	0.7	76.6
Struck[23]	20.2	65.6
SCM[12]	0.5	64.9
VTS[61]	5.7	57.5

The survival curves and average F-scores demonstrate that the BIT achieves the best (0.724) overall performance on ALOV300++.

BIT tracks the object at an average speed of **45fps**, which is significantly faster than the second best tracker RPT (4.1 fps).

MiniEye Simple Eye 松下新加坡研发中心 三星 光电高斯 理光软件研究所 腾讯 华富睿智