



电子科技大学  
University of Electronic Science and Technology of China

# 相关滤波和VOT和目标跟踪

Correlation Filter, VOT and Object Tracking

郭正奎

2018.5.17



# 目录

**1 相关滤波**

**2 VOT(Visual Object Tracking)**



# 1 相关滤波

## 相关滤波基础

两个信号  $f$  和  $g$  的相关性(correlation):

$$(f \otimes h)[n] = \sum_{m=-\infty}^{+\infty} f^*[m]h[m+n]$$

相关性衡量两个信号在某个时刻的相似程度

在目标跟踪领域，就是找到与跟踪目标响应值最大的部分

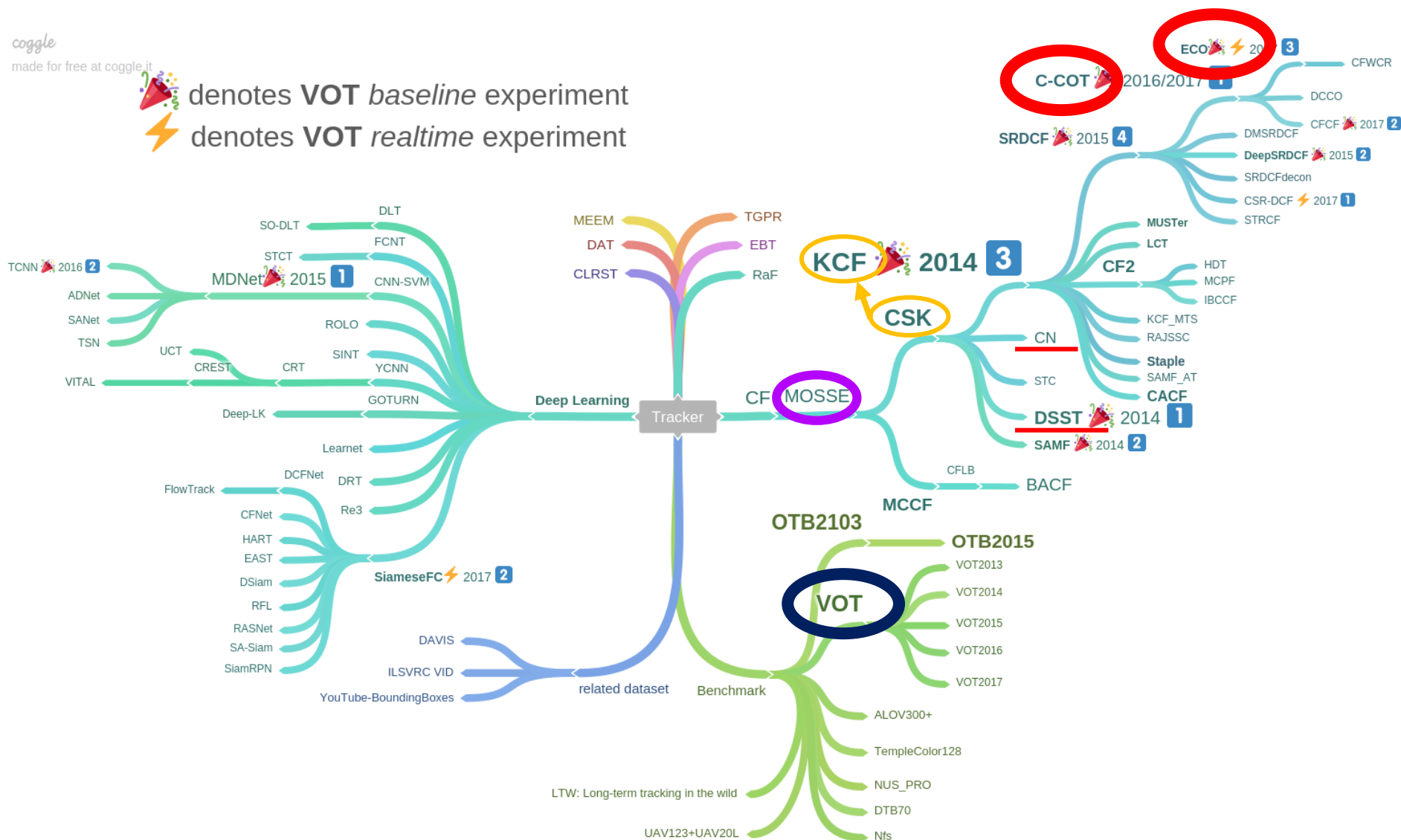
$$G = \mathcal{F}(f \otimes h) = F \odot H^*$$

$$\Rightarrow H^* = \frac{G}{F}$$



coggle  
made for free at coggle.it

denotes VOT baseline experiment  
 denotes VOT realtime experiment





# 1 相关滤波

## MOSSE: 相关滤波方法的始祖

(Minimum Output Sum of Squared Error)

训练过程

$$F^* = \operatorname{argmin}_{F^*} \sum_i |X_i \odot F^* - G_i|^2$$

求解  
 $\Rightarrow$

$$\frac{\partial}{\partial F^*} \sum_i |X_i \odot F^* - G_i|^2 = 0$$

$$\Rightarrow \boxed{F^* = \frac{\sum_i G_i \odot X_i^*}{\sum_i X_i \odot X_i^*}}$$

$f$ : 相关滤波器  $\Rightarrow F$

$g$ : ground truth  $\Rightarrow G$

$x_i$ : 输入序列的第 $i$ 帧  $\Rightarrow X_i$





# 1 相关滤波

## MOSSE: 相关滤波方法的始祖

跟踪过程

$$Y_i = X_i \odot F_{i-1}^*$$

在线更新

$$F_i^* = \frac{A_i}{B_i}$$

$$A_i = \eta G_i \odot X_i^* + (1 - \eta) A_{i-1}$$

$$B_i = \eta F_i \odot F_i^* + (1 - \eta) B_{i-1}$$

$f_i$ : 相关滤波器  $\Rightarrow F$

$g$ : ground truth  $\Rightarrow G$

$x_i$ : 输入序列的第 $i$ 帧  $\Rightarrow X_i$

$y_i$ : 与 $x_i$ 对应的输出  $\Rightarrow Y_i$

# 1 相关滤波

## HCF(Hierarchical Convolutional Features): 对C-COT的启发





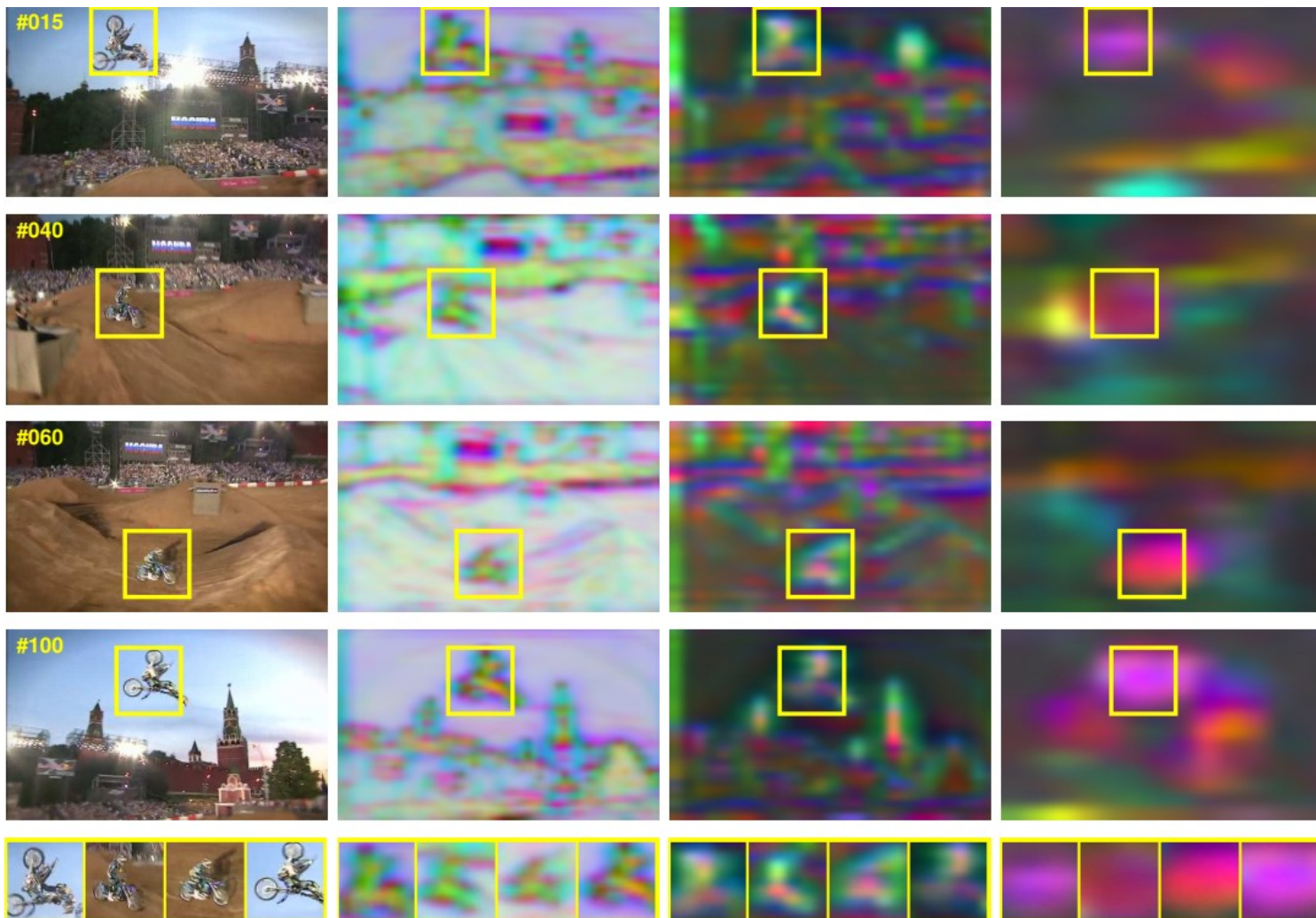
# 1 相关滤波

## HCF(Hierarchical Convolutional Features): 对C-COT的启发

### 1. 插值

- CNN的pooling操作, 会导致分辨率逐层降低
  - VGG-Net pool5 层输出feature map分辨率为 $7*7$ , 原始输入分辨率是 $224*224$
- 双线性插值 (bilinear interpolation)
- 但不符合CNN的coarse-to-fine的层级(hierarchy)





(a) Input

(b) *conv3-4*

(c) *conv4-4*

(d) *conv5-4*



# 1 相关滤波

## HCF(Hierarchical Convolutional Features): 对C-COT的启发

### 2. 相关滤波的应用

$$w^* = \underset{w^*}{\operatorname{argmin}} \sum_{m,n} \|w \cdot x_{m,n} - y(m,n)\|^2 + \lambda \|w\|_2^2$$

$$W^d = \frac{Y \odot (X^d)^*}{\sum_{i=1}^D X^i \odot (X^i)^* + \lambda}$$

$w$ : 相关滤波器

$\lambda$ : 正则化系数



# 1 相关滤波

## HCF(Hierarchical Convolutional Features): 对C-COT的启发

### 3. Coarse-to-fine Translation Estimation

$$(\hat{m}, \hat{n})_{l-1} = \underset{m, n}{\operatorname{argmax}} f_{l-1}(m, n) + \gamma f_l(m, n),$$
$$s. t. \quad |m - \hat{m}| + |n - \hat{n}| \leq r$$

从深层逐渐往浅层反向传播(back propagation)

$(\hat{m}, \hat{n})$ : 相关滤波器响应最大处的坐标  
 $f_l$ : 第 $l$ 层相关滤波响应图



# 1 相关滤波

## C-COT: ECO方法的基础

### 创新点

1. 提出continuous convolution operators, 将离散空间域扩展到连续时间域
2. 将多分辨率特征图融合



# 1 相关滤波

## C-COT: ECO方法的基础

### 符号说明

$x_j$ : 训练样本

$x_j^1, x_j^2 \dots x_j^D$ :  $x_j$ 中抽取的 $D$ 个feature channel

$N_d$ :  $x_j^d$ 中的采样数

$x_j^d[n]$ :  $x_j^d$ 的第 $n$ 个采样点

$T$ : 特征扩展到连续域后, 自变量范围为 $[0, T]$

(In practice, however,  $T$  is arbitrary since it represents the scaling of the coordinate system.)



# 1 相关滤波

## C-COT: ECO方法的基础

### 1. 插值

$$J_d\{x^d\}(t) = \sum_{n=0}^{N_d-1} x^d[n] b_d(t - \frac{T}{N_d} n)$$

$b_d$ : 插值函数(interpolation function)



# 1 相关滤波

## C-COT: ECO方法的基础

### 1. 插值

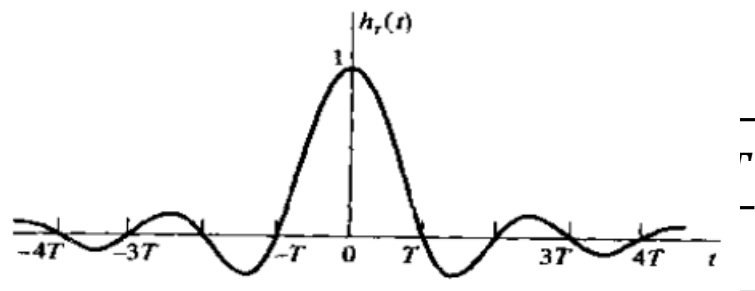
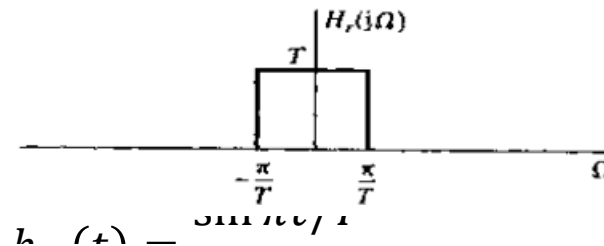
$$J_d\{x^d\}(t) = \sum_{n=0}^{N_d-1} x^d[n] b_d(t - \frac{T}{N_d} n)$$

$b_d$ : 插值函数(interpolation function)

其傅里叶变换:

$$\widehat{J_d\{x^d\}}[k] = \widehat{b_d}[k] X^d[n]$$

$$x_s(t) = \sum_{n=-\infty}^{+\infty} x[n] \sigma(t - nT)$$





# 1 相关滤波

## C-COT: ECO方法的基础

### 2. 相关运算 (Continuous Convolution Operator)

$$S_f\{x\} = \sum_{d=1}^D f^d * J_d\{x^d\}, \quad x \in \mathcal{X}$$

其中  $f^d$  是对应样本  $x$  的第  $d$  层的滤波器 (连续域上的滤波器)

其傅里叶变换:

$$S_f\{\widehat{x}\}[k] = \sum_{d=1}^D \widehat{f^d}[k] \widehat{b_d}[k] X^d[n]$$





# 1 相关滤波

## C-COT: ECO方法的基础

### 3. 训练过程

$$E(f) = \sum_{j=1}^m \alpha_j \|S_f\{x_j\} - y_j\|^2 + \sum_{d=1}^D \|w f^d\|^2 \Rightarrow \boxed{f = \operatorname{argmin} E(f)}$$

用帕斯瓦尔定理:

$$E(f) = \sum_{d=1}^m \alpha_j \left\| \sum_{d=1}^D \widehat{f^d}[k] \widehat{b_d}[k] X^d[n] - \widehat{y_j} \right\|_{l^2}^2 + \sum_{d=1}^D \|\widehat{w} * \widehat{f^d}\|_{l^2}^2$$

矩阵化:

$$E_V(\hat{\mathbf{f}}) = \sum_{j=1}^m \alpha_j \|A_j \hat{\mathbf{f}} - \hat{\mathbf{y}}_j\|_2^2 + \|W \hat{\mathbf{f}}\|_2^2$$



# 1 相关滤波

## ECO：当前效果最好的相关滤波方法

- 出发点：提升**时间效率**和**空间效率**
- 特征维度越来越高，算法越来越复杂，跟踪效果提升，但速度变慢了
- 速度降低的三个最重要的因素：
  - ① Model Size：也可以理解为特征复杂度，C-COT在模型更新的时候需要更新大约800000个参数，速度慢，且易过拟合
  - ② Training Set Size：一般算法在更新模型时，会用到之前帧的信息，且保存新样本，抛弃旧样本，但新样本也可能出错
  - ③ Model Update：模型每帧都更新，比间歇更新慢

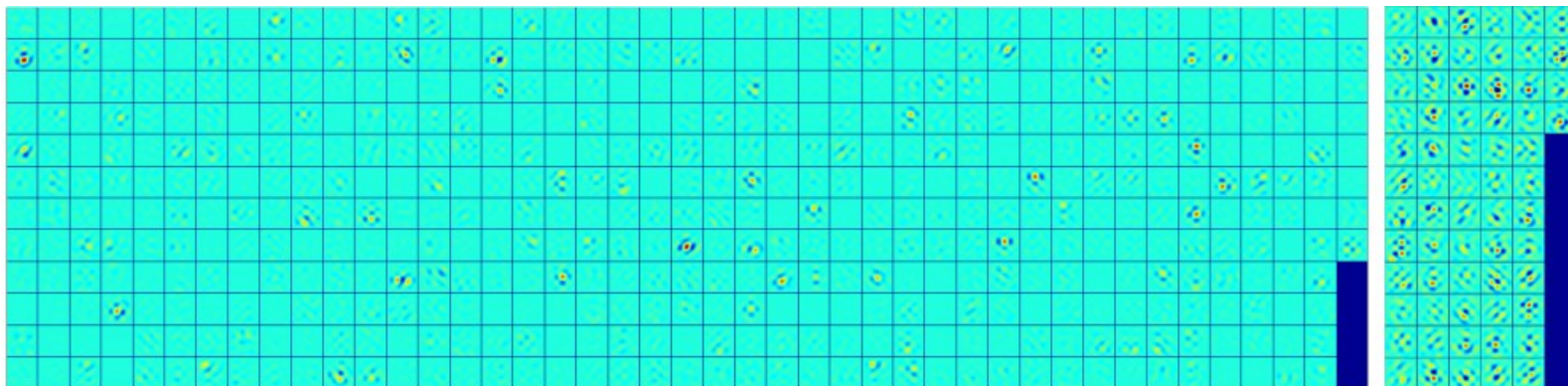


# 1 相关滤波

## ECO: 当前效果最好的相关滤波方法

创新点1 (Model Size) : Factorized Convolution Operator

$$S_{Pf}\{x\} = Pf * J\{x\} = \sum_{c,d} p_{d,c} f^c * J_d\{x^d\} = f * P^T J\{x\}$$



(a) C-COT

(b) Ours

$P$ : 尺寸为  $D \times C$  的矩阵

$D$ : 原本feature map的维数

$C$ : 校正过后feature map的维数



# 1 相关滤波

## ECO: 当前效果最好的相关滤波方法

创新点2 (Training Set Size) : Generative Sample Space Model



用高斯混合模型 (GMM) 对训练集进行分类，姿态近似的样本放在一起，称作一个component





# 1 相关滤波

## ECO：当前效果最好的相关滤波方法


创新点3 (Model Update)：模型从逐帧更新到间隔更新

- 模型逐帧更新导致计算量大，所以采取间隔更新的模型更新策略
- 间隔更新还有减少过拟合的效果
- 实验中**每6帧更新一次模型**



## 2 VOT(Visual Object Tracking)

当前单目标跟踪领域三大数据集之一，另外两个是OTB、AIOV



**VOT2018 challenge**

The VOT2018 challenge will be the 6th Visual Object Tracking challenge. Results will be presented at VOT workshop at ECCV2018. This challenge introduces a long-term subchallenge VOT-LT2018.

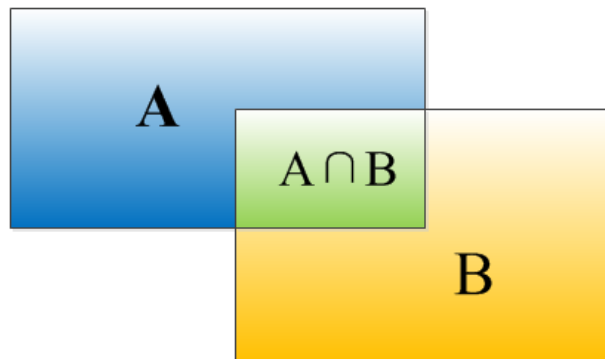
**OPEN**

## 2 VOT(Visual Object Tracking)

两个最重要的评价指标: **Accuracy**和Robustness(Failure Rate)

Accuracy用来评价tracker跟踪目标的准确度, 数值越大, 准确度越高。它借用了IoU (Intersection-over-Union, 交并比) 定义, 某序列第t帧的accuracy定义为:

$$\phi_t = \frac{A_t^G \cap A_t^T}{A_t^G \cup A_t^T}$$





## 2 VOT(Visual Object Tracking)

两个最重要的评价指标：Accuracy和**Robustness(Failure Rate)**

Robustness用来评价tracker跟踪目标的稳定性，**数值越大，稳定性越差。**

$F(i, k)$  定义为tracker-i在**第k次重复**中跟踪失败的次数





## 2 VOT(Visual Object Tracking)

### 重启 (Reset/Reinitialize)

视频中不同的因素，如光照变化、遮挡、形变等都可能影响跟踪算法的效果，tracker很可能会被其中一两个因素导致其跟踪失败，最终导致评价不全面

VOT提出在跟丢的第5帧对tracker重新初始化，初始化之后经过一段burn-in period，继续评价。



## 2 VOT(Visual Object Tracking)

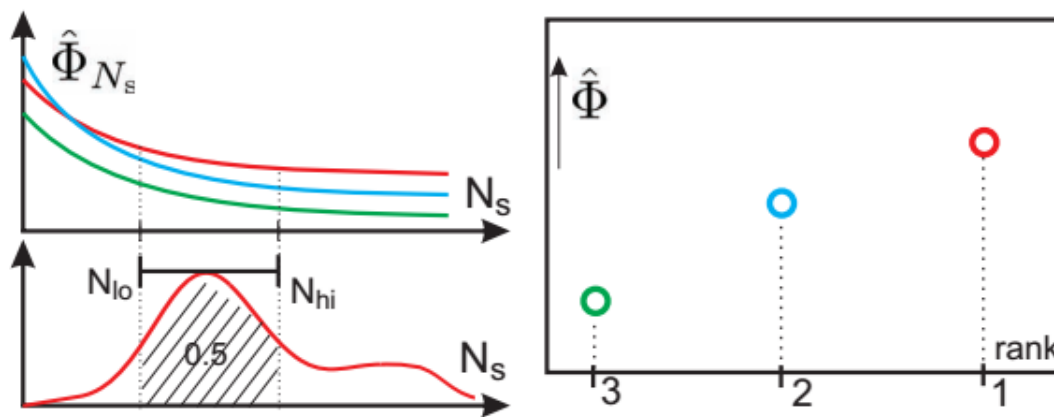
### Ranking

将tracker在不同属性序列上的表现按照accuracy (A) 和 robustness (R) 分别进行排名，再进行平均，得到该tracker的综合排名，依据这个综合排名的数字大小对tracker进行排序得出最后排名。这个排名叫做AR rank。



## 2 VOT(Visual Object Tracking)

EAO (Expected Average Overlap) : 因为之前将排名作为评价标准的体系仍然不够可靠, 所以VOT2015提出EAO, 即利用accuracy的原始数据 (raw data) 而非排名 (rank), 大致思路为对一定长度的序列计算得到该长度对应的overlap (重叠值, 和accuracy定义相同), 并且对每个长度都测一遍, 得到一条EAO曲线, 横坐标为序列长度, 纵坐标为EAO值





## 2 VOT(Visual Object Tracking)

EFO (Equivalent Filter Operations) : 速度是tracker很重要的一个评价指标, 由于编程语言、硬件平台的不同都会导致算法速度不同, 所以VOT2014提出EFO, 通过EFO, 使不同条件下测试的tracker得以在同一评价标准下进行比较



## 2 VOT(Visual Object Tracking)

(1) VOT (Visual Object Tracking) 评价指标综述:

[https://blog.csdn.net/dr\\_destiny/article/details/80108255](https://blog.csdn.net/dr_destiny/article/details/80108255)

(2) VOT2013-2016论文的重点内容:

[https://blog.csdn.net/dr\\_destiny/article/details/80115311](https://blog.csdn.net/dr_destiny/article/details/80115311)

(3) VOT2016和TraX的配置

[https://blog.csdn.net/dr\\_destiny/article/details/79997361](https://blog.csdn.net/dr_destiny/article/details/79997361)



电子科技大学  
University of Electronic Science and Technology of China

**谢谢聆听  
请指正**