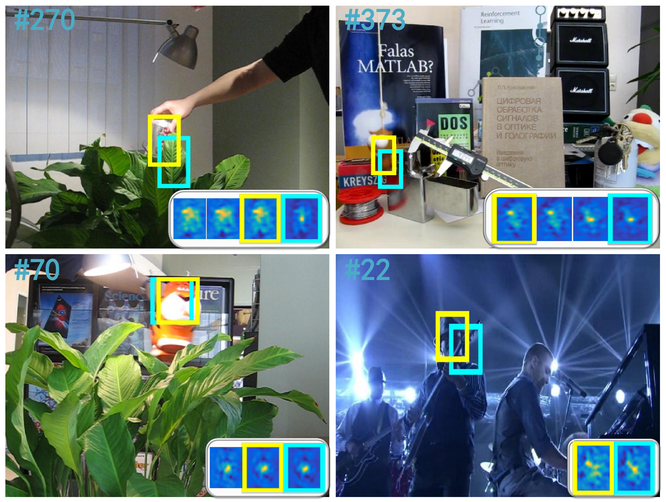
Robust visual tracking by exploiting the historical tracker snapshots

Jiatong Li[1,2], Zhibin Hong[1], Baojun Zhao[2]
[1] Faculty of Engineering and Information Technology, University of Technology, Sydney
[2] School of Information and Electronics, Beijing Institute of Technology

Motivation

➡ **Online Tracking Failure Factors:**
- Illumination changes
- heavy occlusions
- abrupt motions

➡ **Historical Tracker Snapshots:**

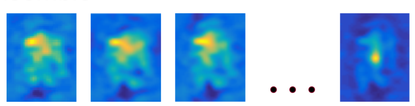


- The historical tracker snapshots are capable of recognizing the true target when come across challenging scenarios.
- Establish a tracker ensemble.
- Handle the tracking failures by exploiting the trajectory consistency between the current tracker and its historical tracker snapshots.

SME Tracker

➡ **Multi-Expert Selection**

✦ Tracker Ensemble



$$\mathbf{E} := \{ \quad \mathcal{T}_{t_1} \quad \mathcal{T}_{t_2} \quad \mathcal{T}_{t_3} \quad \cdots \quad \mathcal{T}_T \quad \}$$

✦ Select criteria: Expert Accumulative Score

$$\mathcal{T}^* = arg\max_{\mathcal{T}\in\mathbf{E}} \sum_{t\in[T-\Delta,T]} \mathcal{S}_{\mathcal{T}}^t,$$

- Expert Score

$$\mathcal{S}_{\mathcal{T}} = L_{\mathcal{T}} - \eta H_{\mathcal{T}}(Y|X,Z)$$

Trajectory consistency score

$$C_{\mathcal{T}}^t = \frac{1}{N-1} \sum_{(\mathcal{T}_i\in\mathbf{E})\bigcap(\mathcal{T}_i\neq\mathcal{T})} \mathcal{C}_{\mathcal{T}:\mathcal{T}_i}^t$$

where

$$\mathcal{C}_{\mathcal{T}_1:\mathcal{T}_2}^t = \exp(-\frac{\|\vec{x}_{\mathcal{T}_1}(t) - \vec{x}_{\mathcal{T}_2}(t)\|^2}{\sigma^2})$$

Entropy Regulization

$$H_{\mathcal{T}}(Y|X,Z) = -\sum_{Y\in Z} P(Y|X;\theta_{\mathcal{T}}) \log P_{\mathcal{T}}(Y|X;\theta_{\mathcal{T}})$$

possible label $Z = \{Y^1, Y^2, ..., Y^n\}$; label $Y^j = (l_1^j, l_2^j, ..., l_n^j)_{l_i\in\{1,-1\}}$; candidate set $X = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$

where

$$P(Y|X;\theta_{\mathcal{T}}) = \prod_i P(l_i|\mathbf{x}_i;\theta_{\mathcal{T}})$$

➡ **Base Tracker**

✦ Translation Estimation — Kernelized Correlation Filter (KCF).

$$f(\mathbf{z}) = \mathcal{F}^{-1}(\hat{\mathbf{k}}^{\mathbf{xz}} \odot \hat{\boldsymbol{\alpha}})$$

✦ Scale Estimation — two dimensional scale space pyramid (only conducted to $\mathcal{T}_T$ for computation efficiency).

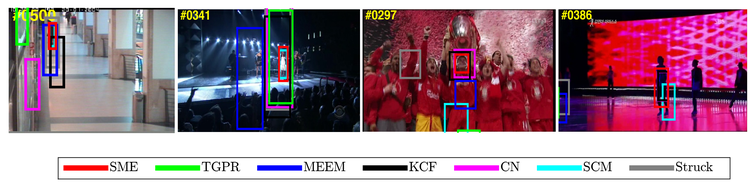
Algorithm 1: SME Tracker

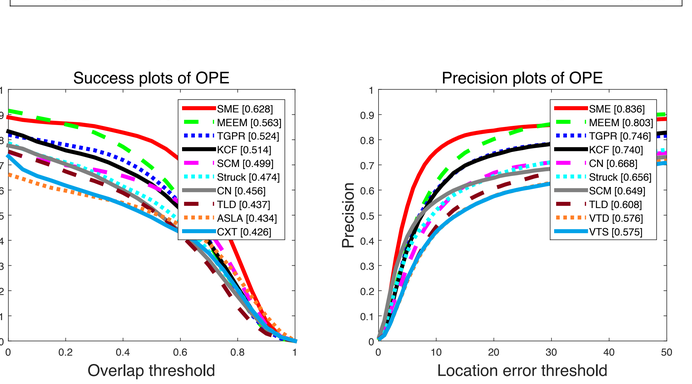
input : Initial target bounding box $\mathbf{x}_1$
output: The estimated target state $\mathbf{x}_t = (\hat{x}_t, \hat{y}_t, \hat{s}_t)$

$\mathbf{E} \leftarrow \mathcal{T}_1$
repeat
  Get the target candidate set X by $\mathbf{E}$;
  for $\mathcal{T} \in \boldsymbol{E}$ **do**
    if $\mathcal{T} = \mathcal{T}_t$ **then**
      Build the target pyramid at $(\hat{x}_{t-1}, \hat{y}_{t-1})$;
      Get the response pyramid, estimate the target position $(x_{\mathcal{T}_t}.y_{\mathcal{T}_t})$ and scale $\hat{s}_t$;
    else
      Get the response map and estimate the target position $(x_{\mathcal{T}}, y_{\mathcal{T}})$;
    Compute the expert score $\mathcal{S}_{\mathcal{T}}^t$;
  if *expert disagreement is reported* **then**
    Select $\mathcal{T}^* \in \mathbf{E}$ according to Expert Accumulative Score;
    $\mathbf{x}_t = (x_{\mathcal{T}_t^*}, y_{\mathcal{T}_t^*}, \hat{s}_t)$;
    $\mathcal{T}_t \leftarrow \mathcal{T}^*$;
  else
    $\mathbf{x}_t = (x_{\mathcal{T}_t}, y_{\mathcal{T}_t}, \hat{s}_t)$;
  if $\mod(t, \Omega) == 0$ **then**
    $\mathbf{E} \leftarrow \mathbf{E} \cup \mathcal{T}_t$;
    discard the oldest snapshot when $|\mathbf{E}| > N_{\mathbf{E}}$;
  Update $\mathcal{T}_t$;
until *Last frame of video sequences*;

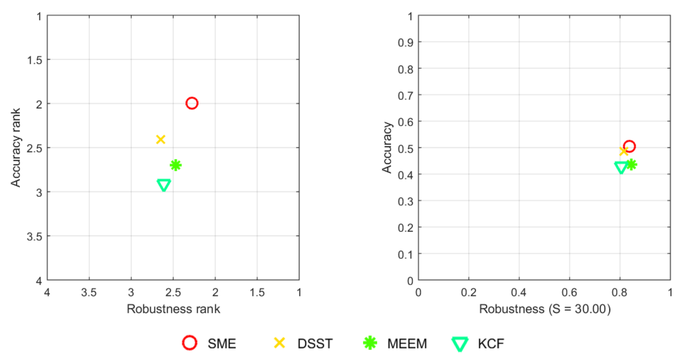
Experiments

- **Visual Tracker Benchmark:**
  – 50 sequences and 33 trackers.





- **VOT 2015 Dataset:**
  – 60 sequences and 3 top performance trackers.



References

– J. Zhang, M. Shugao, and S. Stan. "MEEM: Robust tracking via multiple experts using entropy minimization." In ECCV 2014.

– J. F. Henriques, et al. "High-speed tracking with kernelized correlation filters." TPAMI 2015.