An Integrated Framework for Pedestrian Tracking

Taihong Xiao, Jinwen Ma
Department of Information Science, School of Mathematical Science and LMAM, Peking University
Why Pedestrian Tracking?
General Object Tracking

VOT Challenges
Pedestrian Tracking

1. Various gestures, appearances and poses
2. Distraction from similar person
3. Complete occlusion
The ILFPT Model Overview

- **Detection Frame?**
- **Faster R-CNN Sampling**
- **Local Sampling**
- **Proposals**
- **Predicted location**
- **Previous location**
- **Predicted location**
- **Pretrained Faster R-CNN**
- **Pretrained Re-id Net**
- **Deep Re-id and Haar Features**
- **Bayesian Classifier**
- **Online Update**
- **Final location**

Frame $t$
Target Prediction

Assume that we are tracking on frame $t$, the speed of the target in the current frame can be estimated by

$$
\begin{align*}
    v_x^t &= \rho v_x^{t-1} + (1 - \rho)\Delta x^{t-1}, \\
    v_y^t &= \rho v_y^{t-1} + (1 - \rho)\Delta y^{t-1}
\end{align*}
$$

where $v_x^{t-1}$ and $v_y^{t-1}$ are respectively the horizontal and vertical speed in frame $t - 1$, $\rho \in [0, 1]$ is the momentum factor that controls the weights of the previous speed and

$$
\begin{align*}
    \Delta x^{t-1} &= x^{t-1} - x^{t-2}, \\
    \Delta y^{t-1} &= y^{t-1} - y^{t-2}
\end{align*}
$$

Therefore, the center of tracking window in frame $t$ can be predicted as

$$
\begin{align*}
    x^t &= x^{t-1} + v_x^t \\
    y^t &= y^{t-1} + v_y^t
\end{align*}
$$
Two Sampling Techniques

Local sampling

Faster RCNN sampling
Detection Frames & Non-detection Frames

1. Faster RCNN sampling is activated in detection frames.
2. Faster RCNN sampling helps adapt the bounding box size.
3. Switching scheme improves speed.

Periodical Frames

Detection Frames

Feed-back Frames

Non-detection Frames
Online Learning Model

A positive sample set $S^+$ and a negative sample set $S^-$ are initialized for storing new pedestrian patterns and updating online model. For each candidate sample $c$, we obtain a compressed low-dimensional feature $v = (v_1, v_2, \ldots, v_n)^T \in \mathbb{R}^n$.

$$R(v) = \log \frac{p(y=1|v)}{p(y=0|v)} = \sum_{i=1}^{n} \log \frac{p(v_i|y=1)}{p(v_i|y=0)}$$

The conditional probabilities are assumed to be Gaussian distributed with four parameters $(\mu^1_i, \sigma^1_i, \mu^0_i, \sigma^0_i)$

$$p(v_i|y = 1) = \mathcal{N}(\mu^1_i, \sigma^1_i), \quad p(v_i|y = 0) = \mathcal{N}(\mu^0_i, \sigma^0_i)$$
Online Learning Model

Update rules for parameters

\[
\sigma_i^j \leftarrow \sqrt{\lambda (\sigma_i^j)^2 + (1 - \lambda) (\bar{\sigma}_i^j)^2 + \lambda (1 - \lambda) (\mu_i^j - \bar{\mu}_i^j)^2}
\]

\[
\mu_i^j \leftarrow \lambda \mu_i^j + (1 - \lambda) \bar{\mu}_i^j
\]

\[
(i = 1, 2, \ldots, n; j = 0, 1)
\]

where \(\lambda \in (0, 1)\) is the inertial factor that controls the updating speed and

\[
\bar{\mu}_i^j = \frac{1}{|S_j|} \sum_{k \in S_j} v_i(k)
\]

\[
\bar{\sigma}_i^j = \sqrt{\frac{1}{|S_j|} \sum_{k \in S_j} (v_i(k) - \bar{\mu}_i^j)^2}
\]
Evaluation Methodology

Center Location Error (CLE)  Pascal VOC overlap ratio (VOR)
Experimental Results

Precision plots of OPE:
- Ours (0.646)
- CF2 (0.356)
- FCNT (0.354)
- KCF (0.345)
- MDNet (0.619)
- C-COT (0.595)
- SAMF (0.512)
- SCT (0.320)
- Staple (0.564)
- Struck (0.322)
- TLD (0.222)

Success plots of OPE:
- Ours (0.657)
- CF2 (0.481)
- FCNT (0.449)
- KCF (0.462)
- MDNet (0.671)
- C-COT (0.638)
- SAMF (0.583)
- SCT (0.447)
- Staple (0.624)
- Struck (0.473)
- TLD (0.305)
Effectiveness of Deep Re-id Feature

Precision plots of OPE

Success plots of OPE

Precision Curve

Success Curve
An Integrated Framework for Pedestrian Tracking

Taihong Xiao, Jinwen Ma
Peking University

https://www.youtube.com/watch?v=HQIi0Z9b4Pw
Thank you!