The Large margin tracking method ith circulant feature maps

创建时间： *2017/9/23 18:37*
作者： *XY.Wang*

the large margin tracking method with circulant featur

利用循环移位样本的特征图(**circulant feature maps**)与一个向量的内积，作为是目标的可能性；

目标函数的约束条件：使得目标中心处的patch是目标的可能性大于（w,h）处是目标的可能性超过一个距离。**意为large margin！！**

output pairs (x, y) with $F : X \times Y \rightarrow \mathbb{R}$ from which we can acquire a prediction by maximizing F over the response variable for a specific given input x. Then the general form of the function f can be denoted as

衡量是目标的可能性

$$f(\mathrm{x}; \mathrm{w}) = \arg\max_{\mathrm{y} \in Y} F(\mathrm{x}, \mathrm{y}; \mathrm{w}) \quad (1)$$

where we assume F to be a linear function, $F(\mathrm{x}, \mathrm{y}; \mathrm{w}) = \langle \mathrm{w}, \Psi(\mathrm{x}, \mathrm{y}) \rangle$ and w denotes the parameter vector which can be learned from the soft-margin support vector machine learning over structured outputs. F can also be extended to nonlinear situation which will be discussed in the next section. We penalize margin violations by a quadratic term, leading to the following optimization problem:

$$\min_{\mathbf{w}} \frac{1}{2}\|\mathrm{w}\|^2 + C \sum_{w=1}^{W-1} \sum_{h=1}^{H-1} \xi_{w,h}^2$$

松弛因子

$$\text{s.t.} \forall w, \forall h, \forall \mathrm{y}_{w,h} \in Y \backslash \mathrm{y}_{0,0} :$$

$$F(\mathrm{x}, \mathrm{y}_{0,0}; \mathrm{w}) - F(\mathrm{x}, \mathrm{y}_{w,h}; \mathrm{w}) \geqslant \sqrt{\Delta(\mathrm{y}_{0,0}, \mathrm{y}_{w,h})} - \xi_{w,h}$$

意义：寻找w的解，使得两个patches之间是目标的可能性的差距>=二者的高斯权重差的

where $\mathrm{y}_{0,0}$ denotes the observed output with no cyclic transform and $\xi_{w,h}$ is the slack variable which penalizes the margin violations. The regularization parameter $C > 0$ controls the trade-off between training error minimization and margin maximization. $\Delta(\mathrm{y}_{0,0}, \mathrm{y}_{w,h})$ quantifies the loss associated with a prediction $\mathrm{y}_{w,h}$ when the true output value is $\mathrm{y}_{0,0}$. We define the loss function as

$$\Delta(\mathrm{y}_{0,0}, \mathrm{y}_{w,h}) = m(\mathrm{y}_{0,0}) - m(\mathrm{y}_{w,h}) \quad (3)$$

a fast optimization algorithm that builds up a bridge b problem formulation and the well-known correlation fi

给定了需要优化的目标函数和约束条件，利用一种与CF结合的快速在线优化算法，**将求解过程转化到频域，加速了优化过程。**

w和z其中一个固定的话，优化过程就变得很简单了，一步即可求出来，其中w是在频域计算的。然后交替更新，优化。

（1）w固定，z的求解方式：

There are two variables w and z to be solved in Eq.7. Whenever one of them is known, the subproblem on the other has a closed form solution. Thus similar to [34], we introduce the alternating optimization algorithm to solve the model efficiently by iterating between the following two steps.

交替的优化算法：两个参数交替进行优化计算。

Update z. Given w, the subproblem on z becomes:

$$\min_{\mathbf{z}} \left\| \mathrm{z} - \left(\mathrm{w}^T \Phi_0 - \mathrm{w}^T \Phi - \Upsilon \right) \right\|_2^2, \text{s.t. } \mathrm{z} \geqslant 0 \quad (9)$$

Then the closed form solution of z is:

$$\mathrm{z} = \max\left\{\mathrm{w}^T\Phi_0 - \mathrm{w}^T\Phi - \Upsilon, 0\right\} \qquad (10)$$

w和z其中一个固定的话，优化过程就变得很简单了，
一步即可求出来。然后交替更新，优化。

（2）z固定，w转化到频域进行求解，加速了运算（计算公式如下图）

（3）非线性模型的求解：是需要求解参数alpha，由公式可计算出w（计算公式如下图）

Update w. Given z, the subproblem on w becomes:

$$\min_{\mathrm{w}} \frac{1}{2}\|\mathrm{w}\|^2 + C\left\|\mathrm{w}^T\Phi - \left(\mathrm{w}^T\Phi_0 - \Upsilon - \mathrm{z}\right)\right\|_2^2 \qquad (11)$$

哪里有用到CF理论？

In order to employ the correlation filter theory, we define $\mathrm{u}_0 = \mathrm{w}^T\Phi_0$ which stands for a plane whose height is the highest peak of $\mathrm{w}^T\Phi$ in the last iteration. Then the closed form solution of w is:

将循环样本的计算转换到了频域求解？
就是滤波器理论？

$$\hat{\mathrm{w}} = \frac{\hat{\Psi}^*\left(\mathrm{x}, \mathrm{y}_0\right) \circ \hat{\mathrm{u}}^T}{\hat{\Psi}^*\left(\mathrm{x}, \mathrm{y}_0\right) \circ \hat{\Psi}\left(\mathrm{x}, \mathrm{y}_0\right) + \frac{1}{2C}} \qquad (12)$$

where $\mathrm{u} = \mathrm{u}_0 - \Upsilon - \mathrm{z}$ and $\frac{\bullet}{\bullet}$ denotes the element-wise division.

Nonlinear extension. The proposed linear model can be extended to a nonlinear model by the kernel trick $K_{ij} = \left\langle \varphi\left(\Psi\left(\mathrm{x}, \mathrm{y}_i\right)\right), \varphi\left(\Psi\left(\mathrm{x}, \mathrm{y}_j\right)\right)\right\rangle$ where $\varphi(\bullet)$ indicates the implicit use of a high-dimensional feature space. The solution w can be represented as $\mathrm{w} = \sum_{w=0}^{W-1} \alpha_w \varphi\left(\Psi\left(\mathrm{x}, \mathrm{y}_w\right)\right)$. The optimization now is rewritten as

$$\begin{aligned} &\min_{\alpha} \alpha^T \mathcal{F}^{-1}\left(\hat{\mathrm{k}}^{\Psi_0\Psi_0} \circ \hat{\alpha}\right) \\ &+ C\left\|\mathcal{F}^{-1}\left(\hat{\mathrm{k}}^{\Psi_0\Psi_0} \circ \hat{\alpha}\right) - \left(\mathrm{u}_0 - \Upsilon - \mathrm{z}\right)^T\right\|_2^2 \\ &\text{s.t. } \mathrm{z} \geqslant 0 \end{aligned} \qquad (13)$$

where $\Psi_0 = \Psi\left(\mathrm{x}, \mathrm{y}_0\right)$ and $\hat{\mathrm{k}}^{\Psi_0\Psi_0}$ denotes the DFT of the first row of the circulant kernel matrix K whose elements are K_{ij}. The closed form of the subproblem on α is

$$\hat{\alpha} = \frac{\hat{\mathrm{u}}^T}{\hat{\mathrm{k}}^{\Psi_0\Psi_0} + \frac{1}{2C}} \qquad (14)$$

where $\frac{\bullet}{\bullet}$ denotes the element-wise division.

多峰目标检测

新的一帧到来时，通过公式(1)向量参数与特征图做内积得到的响应图来预测目标位置，

tion f can be denoted as

衡量是目标的可能性

$$f\left(\mathrm{x}; \mathrm{w}\right) = \arg\max_{\mathrm{y} \in Y} F\left(\mathrm{x}, \mathrm{y}; \mathrm{w}\right) \qquad (1)$$

x是以上一帧目标位置为中心，与目标区域成比例的一个iamge patch；

是目标可能性的函数F定义为两个向量的内积：

$F(x,y;w) = \langle w, \Psi(x,y) \rangle$

由前面的内容可知，求解参数w的过程可以通过FFT来加速。

单峰目标检测(只找最高峰)

所有的循环样本的响应图放在一起，就变成了：

单峰 $$F(\mathrm{s},\mathrm{y};\mathrm{w}) = \mathcal{F}^{-1}\left(\hat{\Psi}^{*}_{\mathrm{s0}} \circ \hat{\mathrm{w}}\right) = \mathcal{F}^{-1}\left(\hat{\mathrm{k}}^{\Psi_{\mathbf{x}0}\Psi_{\mathrm{s}0}} \circ \hat{\alpha}\right) \quad (15)$$

寻找响应图的最高峰，作为预测的目标位置。（s是前面的x）

解释：s的循环移位样本的特征与w做内积运算==原始image patch s的联合特征图与w做循环卷积（移位量为0时）->转换到频域，就是二者的element-wise pr

缺陷：然而这种检测方法是单峰检测，如果有相似物体或背景干扰存在的话，它们对应的响应值可能很接近目标甚至比目标更高，因而响应图中的最高峰可能是把最高峰当做目标，而是将所有的峰都考虑进去（即**多峰检测**），从中寻找代表目标位置的那个峰。

多峰目标检测（从多个峰里再找最高分）

在响应图的基础上，乘上一个与响应图大小相同的0/1矩阵B（B的元素在局部最大值处的值设为1，其余为0）。则P(s)中的非零值就表示响应图中的多峰。——其实

文中这样描述：

Consequently, a multimodal target detection method is proposed to improve localization precision further. For the unimodal detection response map $F(\mathrm{s},\mathrm{y};\mathrm{w})$, the multiple peaks are computed by

多峰 $$P(\mathrm{s}) = F(\mathrm{s},\mathrm{y};\mathrm{w}) \circ \mathrm{B} \quad (16)$$

where B is a binary matrix with the same size as $F(\mathrm{s},\mathrm{y};\mathrm{w})$, which identifies the locations of local maxima in $F(\mathrm{s},\mathrm{y};\mathrm{w})$. The elements at the locations of local maxima in B are set to 1, while others are set to 0. All non-zero elements in $P(\mathrm{s})$ indicate multiple peaks in the response map of s. P(s)只保留了响应图中的局部最大值（多峰）。

多峰响应公式：

多峰 $$P(\mathrm{s}) = F(\mathrm{s},\mathrm{y};\mathrm{w}) \circ \mathrm{B} \quad (16)$$

多峰检测方法：当多峰和最高峰的比例超过预设的阈值时，就以这些峰所在的位置为中心，分别提取patch作为s，然后用公式15寻找每个峰所在区域的最大值置。下图中间是检测到的多峰，最右是多峰检测后的最终目标位置。

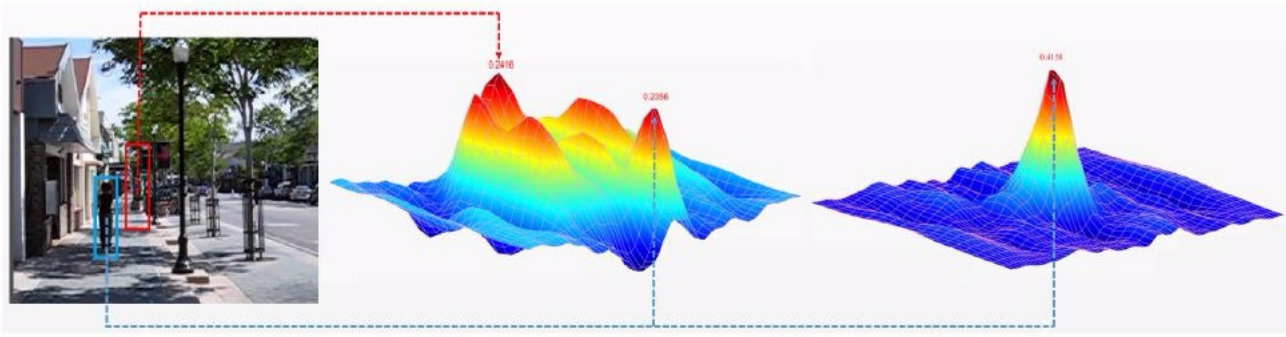
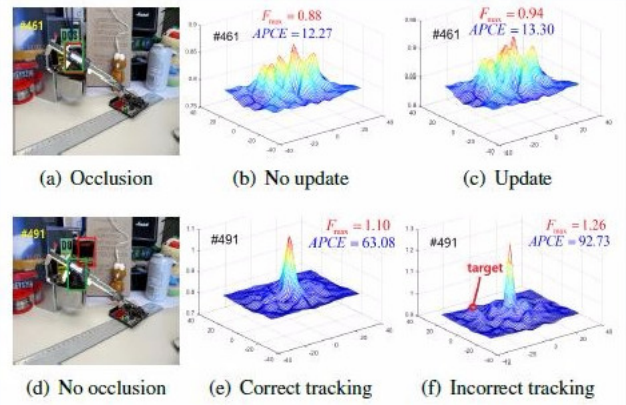


Figure 1. Illustration of multimodal target detection in sequence *human9* from OTB-15 [30]. The blue bounding box indicates the correct location of target, the red one is an incorrect detection. The response of the target is weaker than the background area within the red bounding box as shown in the middle. The unimodal detection will regard the highest peak as the target leading to false detection. The proposed multimodal target detection will redetect the areas centered at other peaks to find the maximum peak among these response maps as the right subfigure and locate the correct position of the target.

High-coinfidence update

大多数的跟踪器不考虑检测结果是否精确，简单粗暴地每帧都更新。实际上，当目标被遮挡或消失了，或者当前帧的检测结果根本就不准确时，再用这些结果去更

下图展示了目标被遮挡后继续每帧更新模型，造成后续某一帧跟踪结果错误的现象。（第一列的绿色框对应每帧更新，空色是high-coinfidence更新；第二列是
更新的跟踪结果，可以看到(f)图跟踪错了）



(a) Occlusion (b) No update (c) Update

(d) No occlusion (e) Correct tracking (f) Incorrect tracking

Figure 2. The first column are the shots of sequence *box* from OTB-15, where the red bounding boxes indicate the tracking results of LMCF with high-confidence update strategy and the green ones belong to the LMCF-NU which updates the tracking model in each frame. The response maps in the second column are corresponding to LMCF and the third column corresponding to LMCF-NU. The red annotation in the last subfigure points out the right position of the target in this response map.

本文：利用跟踪结果的反馈信息（跟踪结果是否足够可信），来决定是否有必要更新模型！

理想的响应图：只有一个sharp peak（锐利的峰），其他区域平滑下降。相关峰越尖锐，检测结果越精确。

非理想的响应图：有很多的波动。如果继续用不确定是否是目标的区域作为训练样本去更新模型，可能导致跟踪结果出错。

A. 定义两个条件：

Fmax = max F(s,y;w)：响应图中的最高相应分数

APCE：表示响应图的波动程度和检测结果的可信度。对于更尖锐的峰和少的噪音，响应图只有一个尖锐的峰和平滑下降的区域，APCE变得更大。当目标被遮

confidence feedback mechanism with two criteria. The first one is the maximum response score $F_{\max}$ of the response map $F(\mathrm{s}, \mathrm{y}; \mathrm{w})$ defined as

$$F_{\max} = \max F(\mathrm{s}, \mathrm{y}; \mathrm{w}) \tag{17}$$

The second one is a novel criterion called average peak-to-correlation energy (APCE) measure which is defined as

$$APCE = \frac{|F_{\max} - F_{\min}|^2}{mean\left(\sum_{w,h}(F_{w,h} - F_{\min})^2\right)} \tag{18}$$

where $F_{\max}$, $F_{\min}$ and $F_{w,h}$ denote the maximum, minimum and the w-th row h-th column elements of $F(\mathrm{s}, \mathrm{y}; \mathrm{w})$. APCE indicates the fluctuated degree of response maps and the confidence level of the detected targets. For sharper peaks and fewer noise, i.e., the target apparently appearing in the detection scope, APCE will become larger and the response map will become smooth except for only one sharp peak. Otherwise, APCE will significantly decrease if the object is occluded or missing. 更新条件：

B. 更新条件(High-coinfidence)：

当 当前帧的Fmax和APCE与它们各自的历史均值的比值超过beta1和beta2时，就认为当前帧的跟踪结果是高可信的，这个时候才选择更新模型。

the object is occluded or missing. 更新条件：

When these two criteria $F_{\max}$ and APCE of the current frame are greater than their respective historical average values with certain ratios β_1, β_2, the tracking result in the current frame is considered to be high-confidence. Then the proposed tracking model will be updated online with a learning rate parameter η as

$$\begin{aligned}\hat{\alpha}^t &= (1-\eta)\,\hat{\alpha}^{t-1} + \eta\hat{\alpha} \\ \hat{\Psi}^t_{\mathbf{x}0} &= (1-\eta)\,\hat{\Psi}^{t-1}_{\mathbf{x}0} + \eta\hat{\Psi}_{\mathbf{x}0}\end{aligned} \tag{19}$$

C. 这样做的结果是

防止跟踪结果不准确时还去更新模型，造成模型堕落的问题。

本文的算法流程图

Algorithm 1 LMCF tracking algorithm

Input: Frames $\{\mathbf{I}_t\}_1^T$, initial target location $\mathbf{p}_1$, $\mathbf{z} = 0$, $\mathbf{u}_0 = ones\,(W, H)$

Output: Target locations of each frame $\{\mathbf{p}_t\}_2^T$.

1: **repeat**

检测阶段

2: Crop an image region s from $\mathbf{I}_t$ at the last location $\mathbf{p}_{t-1}$ and extract its joint feature map $\Psi\,(\mathbf{s}, \mathbf{y}_{0,0})$.

3: Detect the target location $\mathbf{p}_t$ with the multimodal detection via Eq.15 and Eq.16.

4: Estimate the scale of the target as [5].

5: Calculate $F_{\max}$ and APCE with Eq.17 and Eq.18.

决定是否更新

6: **if** $F_{\max}$ and APCE satisfy the update condition, **then**

7: Train the $\mathbf{u}_0$, z and $\hat{\mathbf{w}}$ $(\hat{\alpha})$ with Eq.10 and Eq.12 (14).

8: Update the tracking model with Eq.19.

9: Update the scale estimation model as [5] with η.

10: **end if**

11: **until** end of video sequence.

The la
to extr
layer i
Our tr
a PC \
k40 Gl

性能分析

只和相关滤波器的跟踪算法进行了比较

1.使用传统特征，fps可以达到85帧每秒；使用CNN特征，明显降低，变化成8f/s。但使用高层特征的精度更高。

使用传统特征，fps可以达到85帧每秒；使用CNN特征，明显降低，变化成8f/s。
但使用高层特征的精度更高。

Table 2. Characteristics and tracking results of LMCF, DeepLMCF, LMCF-Uni, LMCF-NU and LMCF-N2. The entries in red denote the best results and the ones in blue indicate the second best.

Trackers	multimodal detection	high-confidence update	feature representations	OPE		TRE		SRE		mean FPS
				precision	success	precision	success	precision	success	
LMCF-N2	No	No	conventional	0.799	0.586	0.813	0.612	0.740	0.540	60.74
LMCF-Uni	No	Yes	conventional	0.809	0.606	0.815	0.616	0.757	0.549	61.38
LMCF-NU	Yes	No	conventional	0.813	0.605	0.820	0.619	0.750	0.545	46.45
LMCF	Yes	Yes	conventional	0.839	0.624	0.829	0.625	0.760	0.552	85.23
DeepLMCF	Yes	Yes	deep CNNs	0.892	0.643	0.877	0.649	0.850	0.596	8.11

2.对于遮挡、光线变化等各种情况的鲁棒性都表现的最好：（平均成功率最高）

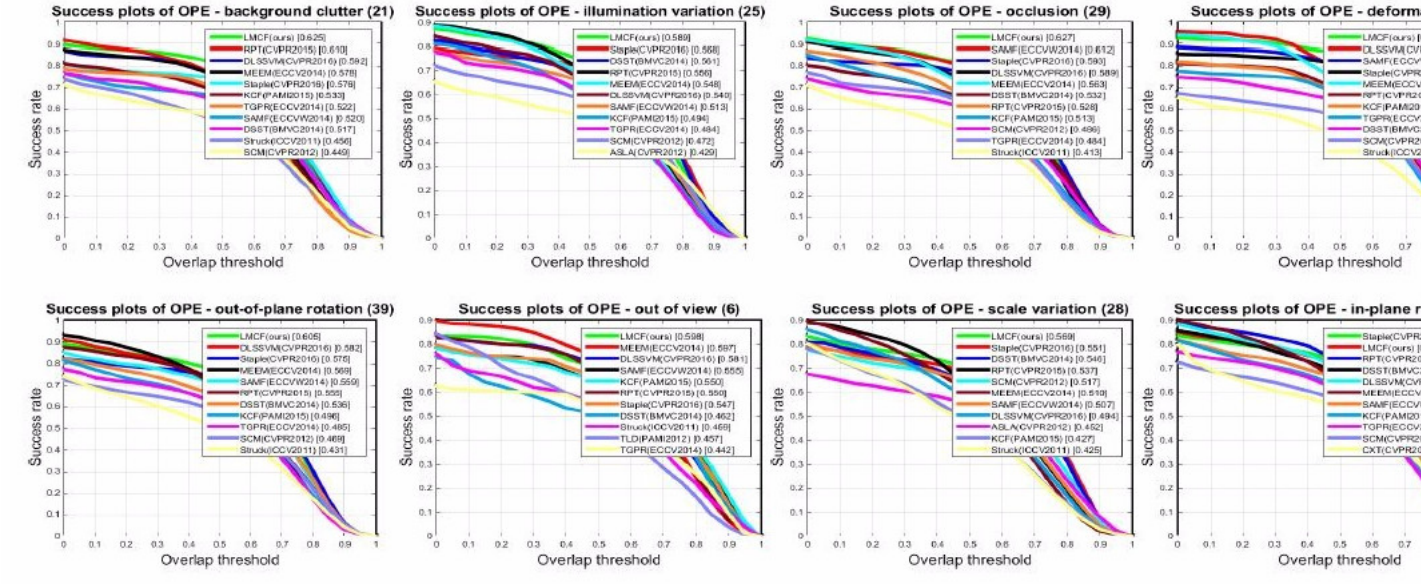


Figure 4. The success plots for 8 challenging attributes including background clutter, illumination variation, occlusion, deforma of-plane rotation, out-of-view, scale variation and in-plane rotation. The proposed LMCF performs best in almost all the attribute are best viewed on high-resolution displays.

3.DeepLMCF的精度仅次于C-COT