



文章编号:1001-9081(2018)06-1751-04

DOI:10.11772/j.issn.1001-9081.2017112735

## 通道稳定性加权补充学习的实时视觉跟踪算法

樊佳庆<sup>1</sup>, 宋慧慧<sup>2</sup>, 张开华<sup>1\*</sup>

(1. 江苏省大数据分析技术重点实验室(南京信息工程大学), 南京 210044;  
2. 南京信息工程大学 大气环境与装备技术协同创新中心, 南京 210044)  
(\*通信作者电子邮箱 zkhua@gmail.com)

**摘要:**为解决补充学习(Staple)跟踪算法在平面内旋转、部分遮挡时存在的跟踪失败问题,提出了一种通过通道稳定性加权的补充学习(CSSStaple)跟踪算法。首先,使用标准相关滤波分类器检测出每层通道的响应值;然后,计算获得每层通道的稳定性权重,并乘到每层权重上,获得相关滤波响应;最后,通过融合颜色补充学习器的响应,得到最终的响应结果,响应中的最大值的位置即为跟踪结果。将所提算法与层和空间可靠性判别相关滤波(CSR-DCF)跟踪、对冲深度跟踪(HDT)、核化相关滤波(KCF)跟踪和Staple等跟踪算法进行了对比实验。实验结果表明,所提算法在成功率上表现最优,在OTB50和OTB100上比Staple分别高出2.5个百分点和0.9个百分点,验证了所提算法对目标在平面内旋转和部分遮挡时的有效性。

**关键词:**相关滤波;视觉跟踪;颜色直方图;补充学习;通道稳定性加权

**中图分类号:** TP391.4    **文献标志码:**A

## Real-time visual tracking algorithm via channel stability weighted complementary learning

FAN Jiaqing<sup>1</sup>, SONG Huihui<sup>2</sup>, ZHANG Kaihua<sup>1\*</sup>

(1. Jiangsu Key Laboratory of Big Data Analysis Technology  
(Nanjing University of Information Science and Technology), Nanjing Jiangsu 210044, China;  
2. Collaborative Innovation Center of Atmospheric Environment and Equipment Technology,  
Nanjing University of Information Science and Technology, Nanjing Jiangsu 210044, China)

**Abstract:** In order to solve the problem of tracking failure of the Sum of template and pixel-wise learners (Staple) tracking algorithm for in-plane rotation and partial occlusion, a simple and effective Channel Stability-weighted Staple (CSSStaple) tracking algorithm was proposed. Firstly, a standard correlation filter classifier was employed to detect the response value of each channel. Then, the stability weight of each channel was calculated and multiplied to the weight of each layer to obtain correlation filtering response. Finally, by integrating the response of the color complementary learner, the final response result was obtained, and the location of the maximum value in the response was the tracking result. The proposed algorithm was compared with several state-of-the-art tracking algorithms including Channel and Spatial Reliability Discriminative Correlation Filter (CSR-DCF) tracking, Hedged Deep Tracking (HDT), Kernelized Correlation Filter (KCF) Tracking and Staple. The experimental results show that, the proposed algorithm performs best in the success rate, it is 2.5 percentage points higher and 0.9 percentage points higher than Staple on OTB50 and OTB100 respectively, which proves the effectiveness of the proposed algorithm for target in-plane rotation and partial occlusion.

**Key words:** correlation filter; visual tracking; color histogram; complementary learning; channel stability weighting

## 0 引言

视觉跟踪是计算机视觉领域一个重要的基础问题。本文只考虑单目标跟踪,即第1帧中的跟踪目标单一且在接下来的帧中跟踪该目标。尽管人们对跟踪已研究多年,但是由于不同的干扰因素,如遮挡、光照、快速移动、运动模糊姿势变化等都会对跟踪产生较大影响,所以它仍是一个非常具有挑战性的计算机视觉任务。

近年来,相关滤波类<sup>[1-2]</sup>跟踪方法因其能利用快速傅里叶变换进行相关滤波运算,从而大大提高跟踪速度和精度,故

引起了广泛关注。在相关滤波的基础上涌现出了大量简单、有效的实时跟踪算法<sup>[3-9]</sup>。

Bolme等<sup>[1]</sup>首先把相关滤波引入视觉目标跟踪领域,提出了速度极快的最小误差输出平方和(Minimum Output Sum of Squared Error, MOSSE)跟踪算法。接着,Henriques等<sup>[2]</sup>提出循环结构核检测跟踪(Circulant Structure of tracking-by-detection with Kernels, CSK)算法,它通过对单幅图像的密集采样获得循环样本,最终实现较好的跟踪结果。为了进一步提升CSK的性能, Henriques等<sup>[4]</sup>接着提出了核化相关滤波(Kernelized Correlation Filter, KCF)算法,利用简单的方向梯

收稿日期:2017-11-20;修回日期:2018-02-24;录用日期:2018-02-27。

基金项目:国家自然科学基金资助项目(41501377);江苏省自然科学基金资助项目(BK20150906, BK20170040)。

作者简介:樊佳庆(1994—),男,江苏南通人,硕士研究生,主要研究方向:目标跟踪;宋慧慧(1986—),女,山东聊城人,教授,博士,主要研究方向:遥感图像处理;张开华(1983—),男,山东日照人,教授,博士,CCF会员,主要研究方向:图像分割、目标跟踪。



度直方图(Histogram of Oriented Gradients, HOG)特征取得了不错的跟踪效果并且达到了远超实时的跟踪速度。

在此基础上,Bertinetto等<sup>[6]</sup>提出了实时补充学习(Sum of template and pixel-wise learners, Staple)跟踪,利用颜色直方图作为全局特征,再结合局部方向梯度直方图特征HOG构造相关滤波跟踪器,取得了较好的实时跟踪效果。但是,Staple方法并没有考虑每层响应的稳定性,即每层响应的权重是一致的,这样就使得噪声很大的响应给跟踪结果带来较大的负面影响。比如,在周围有类似物体干扰的情况下,目标就很容易跟丢。

针对上述问题,本文在实时补充学习(Staple)跟踪的基础上加入了响应稳定性权重,提出了通道稳定性加权的Staple(Channel Stability-weighted Staple, CSStaple)跟踪算法。如图1所示,本文首先使用标准相关滤波获得每层通道响应,然后,利用每个通道响应的峰值计算出稳定性权重,接着,把这些权重乘上对应的通道响应,最终和颜色直方图响应相结合得出跟踪结果。

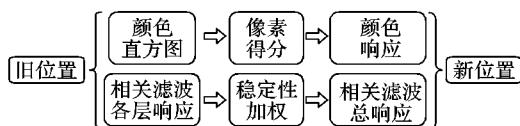


图1 CSStaple 算法原理图

Fig. 1 Schematic diagram of CSStaple algorithm

## 1 通道稳定性加权的跟踪算法

本文所提跟踪算法由加入通道稳定性权重的相关滤波算法模块和颜色直方图补充学习模块组成,最后将两者的响应结果加权线性组合,得出最终的跟踪结果。

### 1.1 通道稳定性权重估计

本文首先通过传统相关滤波算法得出多通道跟踪结果 $r_{\text{channel}}$ ,尺寸是 $m \times n \times c$ (其中: $m$ 是响应的高度, $n$ 是响应的宽度, $c$ 是响应的层数),然后通过峰值检测算法检测出每一层的前若干个峰值向量:

$$\mathbf{p}^i = [p_1^i, p_2^i, \dots, p_n^i]^T \quad (1)$$

其中 $i \in \{1, 2, \dots, c\}$ 是该层的序号。接着通过式(2)求出该层的权重:

$$w^i = \max\{(1 - p_2^i/p_1^i), 0.5\} \quad (2)$$

式中: $p_1^i, p_2^i$ 分别是该层检测出的峰值向量中的最大值和第二大的值。本文的权重度量方法是基于第二大值和最大值的比例即 $p_2^i/p_1^i$ ,当目标背景混乱或光照变化强烈时, $p_2^i/p_1^i$ 则变大,该权重则惩罚该层响应。然而存在一种特殊情况,当目标附近出现多个类似物体时,目标周围也会出现多个较大峰值,但是这些峰值刻画了目标及周围的上下文关系,对跟踪结果有积极作用。为了防止对该类响应的惩罚过重,本文把最低权重设置在0.5,这样避免了对形状相似物体的过度惩罚。

当该层响应噪声较大时,因为高噪声的响应具有较低的判别能力,所以应该得到较低权重,这样也就削弱了该层响应在最终响应中的影响力。相反地,当该层响应的波动较小时,本文认为这是一个比较理想的响应层,所以赋予较大权重,该层响应在最终响应中的影响也较大。

### 1.2 相关滤波分类器

本文利用标准的相关滤波框架,训练出一个岭回归分类器。目标是找到一个函数 $f(\mathbf{z}) = \mathbf{w}^T \mathbf{z}$ ,使得在循环样本 $\{\mathbf{x}_i\}$ 上的检测结果和回归目标 $\{y_i\}$ 之间的最小平方误差最小,即:

$$\min_{\mathbf{w}} \left\{ \sum_i (f(\mathbf{x}_i) - y_i)^2 + \lambda \|\mathbf{w}\|_2^2 \right\} \quad (3)$$

进一步地,利用核技巧<sup>[10]</sup>,本文直接得出式(3)的闭式解:

$$\hat{\boldsymbol{\alpha}} = F^{-1}(\hat{\mathbf{y}} / (\hat{\mathbf{k}}^{\text{xx}} + \lambda)) \quad (4)$$

式中: $\mathbf{k}^{\text{xx}}$ 是 $\mathbf{x}$ 和它自己的核相关;“ $\hat{\cdot}$ ”表示离散傅里叶变换;而 $F^{-1}$ 表示离散傅里叶逆变换。

本文采用一种在线更新的策略来更新学到的参数 $\boldsymbol{\alpha}^t$ ,通过:

$$\boldsymbol{\alpha}^t = (1 - \eta_{\text{cf}}) \boldsymbol{\alpha}^{t-1} + \eta_{\text{cf}} \hat{\boldsymbol{\alpha}}^t \quad (5)$$

式中: $\eta_{\text{cf}}$ 是相关滤波分类器的学习率; $\hat{\boldsymbol{\alpha}}^t$ 利用当前 $t$ 帧的跟踪结果通过式(4)计算得到。最终,当输入新一帧即 $t+1$ 帧图片 $\mathbf{z}^{t+1}$ 时,它每层的检测响应结果即:

$$r_{\text{channel}} = F^{-1}(\hat{\boldsymbol{\alpha}}^t \odot \hat{\mathbf{k}}^{\text{xz}^{t+1}}) \quad (6)$$

每层响应与式(2)中每层的稳定性权重 $w^i$ 相乘并相加之后,便可得到层通道稳定性加权后的相关滤波响应:

$$r_{\text{ch}} = \sum_i r_{\text{channel}}^i \times w^i \quad (7)$$

### 1.3 颜色直方图分类器

为了找到一个较好的颜色分类器,本文使用一种特殊形式的特征表示,对于每个RGB像素 $\mathbf{u} \in \mathbb{R}^3$ ,其特征表示为:

$$\psi[\mathbf{u}] = \mathbf{e}_{k[\mathbf{u}]} \quad (8)$$

其中 $\mathbf{e}_i = [\underbrace{0, 0, \dots, 0}_{i-1}, 1, 0, \dots, 0]^T$ 表示第 $i$ 个元素为1,其余为0。这里的序号 $k[\mathbf{u}]$ 通过式(9)得到:

$$k[\mathbf{u}] = \text{floor}(\mathbf{u}/(255/bins)) + 1 \quad (9)$$

其中 $bins$ 是直方图柱的数量。之后,本文在目标区域 $\Omega_o \subset \mathbb{R}^2$ 和背景区域 $\Omega_b \subset \mathbb{R}^2$ 上使用一个线性回归目标函数:

$$\begin{aligned} \varepsilon(\boldsymbol{\beta}) &= \frac{1}{|\Omega_o|} \sum_{\mathbf{u} \in \Omega_o} (\boldsymbol{\beta}^T \psi[\mathbf{u}] - 1)^2 + \\ &\quad \frac{1}{|\Omega_b|} \sum_{\mathbf{u} \in \Omega_b} (\boldsymbol{\beta}^T \psi[\mathbf{u}] - 0)^2 \end{aligned} \quad (10)$$

式中 $\boldsymbol{\beta}$ 是系数向量。

再把特征表示式(7)代入目标函数,得到:

$$\varepsilon(\boldsymbol{\beta}) = \sum_{i=1}^n \left[ \frac{N^i(\Omega_o)}{|\Omega_o|} (\beta^i - 1)^2 + \frac{N^i(\Omega_b)}{|\Omega_b|} (\beta^i - 0)^2 \right] \quad (11)$$

式中 $N^i(\Omega_a) = \{ \mathbf{u} \in \Omega_a \mid k[\mathbf{u}] = i \}, a \in \{o, b\}$ 。令 $\frac{\partial \varepsilon(\boldsymbol{\beta})}{\partial \beta^i} = 0$ ,得到目标函数的闭式解:

$$\bar{\beta}^i = \frac{N^i(\Omega_o)}{N^i(\Omega_o) + \frac{1}{|\Omega_b|} N^i(\Omega_b)} \quad (12)$$

为了自适应目标的表观变化,类似相关滤波分类器,本文使用一种简单的在线更新策略来更新分类器系数:

$$\boldsymbol{\beta}^t = (1 - \eta_{\text{ch}}) \boldsymbol{\beta}^{t-1} + \eta_{\text{ch}} \bar{\boldsymbol{\beta}}^t \quad (13)$$



式中:  $\beta'$  是利用当前  $t$  帧的跟踪结果计算出来的分类器系数;  $\eta_{ch}$  是颜色直方图分类器的学习率。

在检测阶段,  $t + 1$  帧时, 输入图片之后, 就能得到它在像素点  $u$  处的颜色直方图响应, 即:

$$r_{ch} = \beta'^T \psi[u] \quad (14)$$

#### 1.4 融合相关滤波响应和颜色响应

得到稳定性加权的相关滤波响应  $r_{cf}$  和颜色直方图响应  $r_{ch}$  之后, 本文采用一种线性加权的结合方式得到最终的响应:

$$r = \eta r_{ch} + (1 - \eta) r_{cf} \quad (15)$$

其中:  $\eta$  是融合因子;  $r$  中的最大值的位置就是跟踪结果。

## 2 实验与分析

在这部分中, 首先介绍了实现的细节, 然后详细分析了本文设计的跟踪器 CSStaple 与当前先进的跟踪器在 OTB50<sup>[11]</sup>、OTB100<sup>[12]</sup> 测试集上的性能对比实验结果。

### 2.1 实验设置

在相关滤波部分, 本文使用简单的 HOG 特征并设置它的单元尺寸为  $4 \times 4$ , 设置学习率  $\eta_{cf} = 0.01$ 。在颜色直方图方面, 直接使用 RGB 特征, 颜色直方图柱子数量设定为 32, 颜色分类器的学习率为 0.04。另外, 设置固定区域大小为  $150 \times 150$ , 融合因子  $\eta = 0.5$ 。

### 2.2 OTB50 上的评估

在 OTB50 的 50 个视频上, 首先将本文的跟踪器 CSStaple 与 5 个先进的跟踪器进行一次通过型(One Pass Evaluation, OPE)成功率对比实验, 并对实验结果进行分析。之后, 又分析了基于属性的成功率性能对比结果。

#### 2.2.1 与先进跟踪器对比

在 OTB50 上, 本文选取了 5 个先进的跟踪器进行对比实验, 包括: 层和空间可靠性判别相关滤波( Channel and Spatial Reliability Discriminative Correlation Filter, CSR-DCF) 跟踪<sup>[13]</sup>、对冲深度跟踪(Hedged Deep Tracking, HDT)<sup>[14]</sup>、核化相关滤波(KCF) 跟踪<sup>[4]</sup>、对偶线性结构化 SVM 跟踪(Dual Linear Structured SVM Tracker, DLSSVM)<sup>[15]</sup>、补充学习(Staple) 跟踪<sup>[1]</sup>。总的对比实验结果如图 2 所示, 使用的是成功率这一指标。其中本文提出的 CSStaple 跟踪算法最优, 分别超过基准算法 Staple、核化相关滤波(KCF) 跟踪算法 2.5 个百分点和 10.4 个百分点。

#### 2.2.2 基于属性分析的对比

本文在 OTB50 上进行了属性分析的对比实验。所有的视频被分为 11 种不同的属性, 即: 光照变化、尺度变化、遮挡、形变、运动模糊、平面内旋转、快速移动、平面外旋转、脱离视线、背景混乱以及低像素。平面内旋转、遮挡这两种属性下的成功率如图 3 所示。在平面内旋转属性下, HDT 取得了 58.0% 的曲线下面积(Area Under Curve, AUC) 得分, 本文的跟踪器比它高出 1.8 个百分点。本文的跟踪器在遮挡属性中也取得了最佳的表现, 在图像背景混乱或部分遮挡等强干扰下, 原基准跟踪器 Staple 表现不好, 跟踪成功率较低。而本文的 CSStaple 加入了通道稳定性权重之后, 自动减小响应十分

混乱的通道的权重, 而赋予响应振荡较小的通道以较大权重, 使得最终加权之后的响应变得更为可靠, 所以本文的跟踪器能在这两种属性下表现相对更优。

—+— CSStaple(本文) [0.618]	--- DLSSVM(CVPR2016) [0.608]
-▲- Staple(CVPR2016) [0.593]	-◆- HDT(CVPR2016) [0.603]
-◆- KCF(PAMI2015) [0.514]	-▼- CSR-DCF(CVPR2017) [0.593]

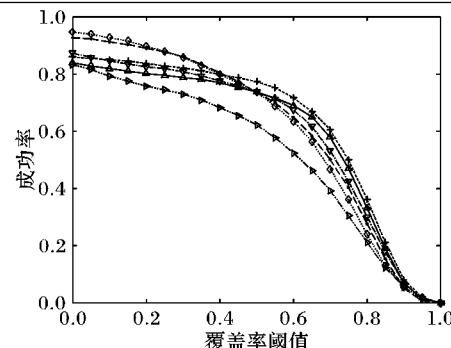
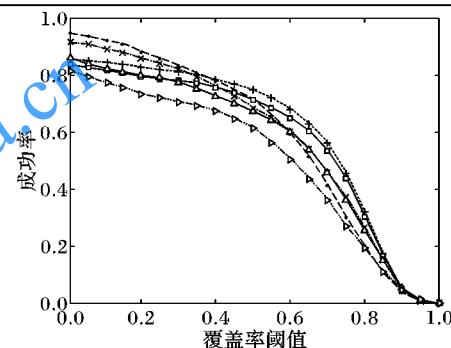


图 2 OTB50 上不同跟踪器对比

Fig. 2 Comparison of different trackers on OTB50

—+— CSStaple(本文) [0.598]	--- DLSSVM(CVPR2016) [0.580]
-▲- HDT(CVPR2016) [0.580]	-◆- Staple(CVPR2016) [0.576]
-◆- KCF(PAMI2015) [0.497]	-▼- CSR-DCF(CVPR2017) [0.554]



(a) 平面内旋转成功率

—+— CSStaple(本文) [0.613]	--- DLSSVM(CVPR2016) [0.592]
-▲- HDT(CVPR2016) [0.603]	-◆- Staple(CVPR2016) [0.585]
-◆- KCF(PAMI2015) [0.514]	-▼- CSR-DCF(CVPR2017) [0.571]

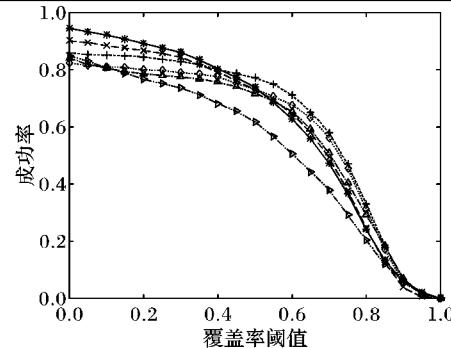


图 3 OTB50 上平面内旋转、遮挡两种属性下成功率

Fig. 3 Success rate under two attributes of in-plane rotation and occlusion on OTB50

### 2.3 OTB100 上的评估

在 OTB100 上本文选取了 5 个先进的跟踪器进行了成功率对比实验, 包括: 层和空间可靠性判别相关滤波(CSR-DCF) 跟踪<sup>[13]</sup>、长期相关跟踪(Long-term Correlation Tracking, LCT)<sup>[5]</sup>、对冲深度跟踪(HDT)<sup>[14]</sup>、核化相关滤波(KCF) 跟踪<sup>[4]</sup>、补充学习(Staple) 跟踪<sup>[6]</sup>。各跟踪器在 OTB100 上的成



功率如图4所示。

由图4可以看出,CSStaple在这100个视频上表现得最好,甚至超过了最新的跟踪算法比如Staple和CSR-DCF;在实时跟踪算法中,Staple的AUC得分为57.9%,LCT的AUC得分为56.2%,本文跟踪方法的AUC得分为58.8%,相比Staple和LCT分别提高了0.9个百分点和2.2个百分点。

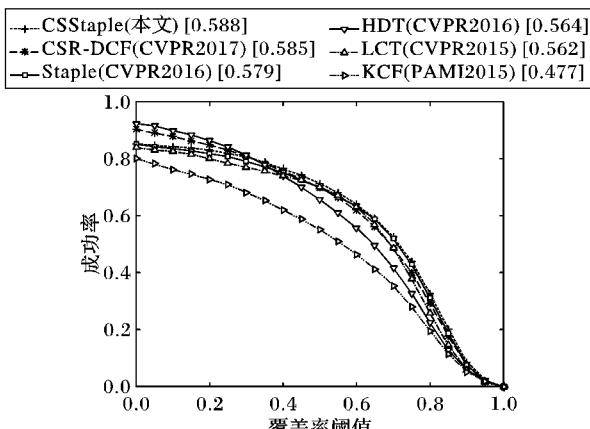


图4 OTB100 上不同跟踪器对比  
Fig. 4 Comparison of different trackers on OTB100

### 3 结语

本文在补充学习(Staple)跟踪器的基础上加入了通道稳定性权重估计,提出了通道稳定性加权的补充学习(CSStaple)跟踪器。改进后的跟踪器能较好地解决背景干扰、场景混乱等问题,因而在跟踪标准测试集OTB50和OTB100上取得了较高的结果,甚至优于一些基于深度学习的跟踪器。但是,本文方法对于强烈光照变化、低分辨率像素等其他常见问题处理得不太理想,后续将针对强烈光照变化、低分辨率像素等问题再进行进一步的研究。

### 参考文献 (References)

- [1] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters [C]// Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2010: 2544 – 2550.
- [2] HENRIQUES J F, CASEIRO R, MARTINS P, et al. Exploiting the circulant structure of tracking-by-detection with kernels [C]// Proceedings of the 2012 European Conference on Computer Vision, LNCS 7575. Berlin: Springer, 2012: 702 – 715.
- [3] ZHANG K H, ZHANG L, LIU Q S, et al. Fast visual tracking via dense spatio-temporal context learning [C]// Proceedings of the 2014 European Conference on Computer Vision, LNCS 8693. Cham: Springer, 2014: 127 – 141.
- [4] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583 – 596.
- [5] MA C, YANG X K, ZHANG C Y, et al. Long-term correlation tracking [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 5388 – 5396.
- [6] BERTINETTO L, VALMADRE J, GOLODETZ S, et al. Staple: complementary learners for real-time tracking [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 1401 – 1409.
- [7] MUELLER M, SMITH N, GHANEM B. Context-aware correlation filter tracking [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 1387 – 1395.
- [8] 朱明敏,胡茂海.基于相关滤波器的长时视觉目标跟踪方法[J].计算机应用,2017,37(5):1466 – 1470.(ZHU M M, HU M H. Long-term visual object tracking algorithm based on correlation filter [J]. Journal of Computer Applications, 2017, 37 (5): 1466 – 1470.)
- [9] 彭爽,彭晓明.基于高效多示例学习的目标跟踪[J].计算机应用,2015,35(2):466 – 469, 475. (PENG S, PENG X M. Object tracking with efficient multiple instance learning [J]. Journal of Computer Applications, 2015, 35(2): 466 – 469, 475.)
- [10] 李航.统计学习方法[M].北京:清华大学出版社,2012: 116. (LI H. Statistical Learning Method [M]. Beijing: Tsinghua University Press, 2012: 116.)
- [11] WU Y, LIM J, YANG M H. Online object tracking: a benchmark [C]// Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2013: 2411 – 2418.
- [12] WU Y, LIM J, YANG M H. Object tracking benchmark [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834 – 1848.
- [13] LUKEZIC A, VOJIR T, ZAJC L C, et al. Discriminative correlation filter with channel and spatial reliability [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 4847 – 4856.
- [14] QI Y K, ZHANG S P, QIN L, et al. Hedged deep tracking [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 4303 – 4311.
- [15] NING J F, YANG J M, JIANG S J, et al. Object tracking via dual linear structured SVM and explicit feature map [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 4266 – 4274.

This work is partially supported by the National Natural Science Foundation of China (41501377), the Natural Science Foundation of Jiangsu Province (BK20150906, BK20170040).

**FAN Jiaqing**, born in 1994, M. S. candidate. His research interests include object tracking.

**SONG Huihui**, born in 1986, Ph. D., professor. Her research interests include remote sensing image processing.

**ZHANG Kaihua**, born in 1983, Ph. D., professor. His research interests include image segmentation, object tracking.