

基于压缩感知的在线多示例学习目标追踪

韩亚颖, 王元全

(天津理工大学计算机与通信工程学院, 天津, 300384)

摘要 近年来提出的多示例学习算法在一定程度上能够克服模板漂移问题。然而,在线学习需要获取足够多的有用数据才能达到稳定的追踪效果,但是这却增加了算法的复杂度。为了解决这一问题,在压缩感知理论的基础上,运用随机观测的方法对多尺度图像特征进行降维,提取的这些低维特征中包含大量的有用信息。因此,我们提出的算法是先利用压缩感知理论提取目标特征之后,再使用在线多示例学习算法分类器对这些特征进行分类从而实现目标的稳定跟踪。通过对不同的图像序列进行实验,结果表明基于压缩感知的在线多示例学习算法对实时的目标追踪有很好的适应性。

关键词 目标追踪;多示例学习;压缩感知

DOI 10.3969/j.issn.1009-3516.2014.05.018

中图分类号 TN82 **文献标志码** A **文章编号** 1009-3516(2014)05-0071-05

Visual Tracking with Multiple Instance Learning Based on Compressive Sensing

HAN Ya-ying, WANG Yuan-quan

(Tianjin University of Technology, Tianjin 300384 China)

Abstract: Visual tracking is one of the most popular research topics in the domain of computer vision. It is a challenging task to develop an effective and efficient tracking algorithm because of template drift problems. To alleviate the drift, the multiple instance learning (MIL) method has been applied to target tracking. However, there must be a sufficient amount of useful data for online MIL to learn at the outset, which actually increases the computational complexity. In this paper, an effective tracking algorithm is proposed which uses an online MIL based on the compressed appearance model to accomplish object tracking. In order to decrease the computational complexity and obtain sufficient data for online learning adaptive appearance model, Features are extracted by non-adaptive random projections of the multi-scale image feature space based on compressive sensing theories. The experimental results on various videos show that the proposed method has a satisfactory performance in real-time object tracking.

Key words: visual tracking; multiple instance learning; compressive sensing

视觉追踪是近年来新兴的一个研究方向,融合了计算机视觉、模式识别、人工智能等学科的技术,具有广泛的应用前景和深远的科学研究意义。视觉

追踪算法尽管很多,但是由于追踪过程中目标特征的变化如部分遮挡、目标的变形以及光照的强弱改变等实现可靠的目标追踪依然是一个非常艰巨的任

收稿日期:2013-09-24

基金项目:天津市自然科学基金资助项目(11JCZDJC15600)

作者简介:韩亚颖(1989-),女,山东菏泽人,硕士生,主要从事图像处理研究.E-mail:lchphy1@126.com

引用格式:韩亚颖,王元全. 基于压缩感知的在线多示例学习目标追踪[J]. 空军工程大学学报:自然科学版,2014,15(5):71-75. HAN Yaying, WANG Yuanquan. Visual tracking with multiple instance learning based on compressive sensing[J]. Journal of air force engineering university: natural science edition, 2014, 15(5): 71-75.

务。因此,如何有效的表达目标特征模型是视觉追踪系统中最重要的部分,也吸引了越来越多人的关注。视觉追踪算法大致上可以分为两类:产生模型和区别模型。

产生模型通过学习目标的特征表示模型来解决追踪过程中目标特征发生变化的问题。Lee 等人提出在线学习算法进行增量学习目标表达模型^[1]。David 等人提出一种线性子空间模型进行在线学习目标特征模型^[2],用粒子滤波实现目标的追踪。近年来,稀疏表达方法成功地被运用在目标追踪过程中^[3-5],处理目标遮挡以及光照变化的情况具有很强的鲁棒性。

区别模型通过分类器进行在线训练和学习的方法来处理目标特征在追踪过程中发生变化的情况,进而从背景中分离出目标。一些有效的分类器如 SVM 和 Boosting 广泛地运用在目标追踪算法中。Adaboost 即自适应 Boosting 算法^[6-7],利用不断增强学习的思想,这样将逐渐产生的弱分类器进行加权联合,训练时以“更关注被错分的样本”为依据进行调整权重,这样从整体上才能达到不断增强学习的效果。Avidan 提出 Ensemble Tracking 方法^[8],利用 Ensemble 集合的思想,通过 Adaboost 方式利用这个强分类器计算下一帧置信图,通过 mean shift 算法来找到目标位置。Tian 等人提出了一种线性 SVM 追踪算法^[9],使得在追踪过程中很好的利用先前的信息来达到更加优越的追踪效果。Grabner 和 Bischof 提出的 Online Adaboost 算法^[10-11]首先利用很少的样本进行训练,然后利用新样本进行不断的更新,再从特征集中不断选择最好的鉴别特征并且保持这种特征进行目标追踪。在线 SVM 分类器用 co-tracking 框架^[12]训练不同的特征进行目标追踪。SemiBoost 算法^[13]根据检测算法进行跟踪并计算未标签的样本与先前样本的相似性。Online semi-supervised 算法^[14]通过第一帧标签的样本来训练分类器进而减轻了样本漂移的问题。尽管这些追踪算法在很多视频序列中能够达到很好的追踪效果,但是要在第一帧能够提取准确的目标位置仍然是一个很困难的任务。因此,Babenko 等人提出了一种在线多示例学习算法^[15-16]。该算法通过构造一个在线学习的多示例分类器作为检测器,不需要采取大量的样本进行离线的训练,只需要手动在第一帧选取目标,就可以自动生成正负样本,并在以后的帧图像序列中,根据跟踪到的目标位置进行自动更新分类器,从而有效的实现了目标的稳定跟踪。但在实际应用中由于背景复杂、光线强弱、摄像机抖动等原因目标通常会出现一些特征的

变化,所以多示例学习算法在解决这些复杂条件下的追踪时仍然存在很多问题。实时的压缩追踪算法^[17]基于特征压缩进行目标追踪,将特征用一个非常简单的随机测量矩阵进行压缩,是目标追踪算法中一个显著的改进。

为了实现更加鲁棒地追踪效果,本文将压缩感知理论与在线多示例学习算法进行巧妙的结合,提出了一种简单有效的追踪算法,首先利用符合压缩感知 RIP 条件^[18]的稀疏的测量矩阵对多尺度图像特征进行降维,所得到的低维压缩子空间可以很好的保留高维图像特征空间的信息,然后采用在线多示例学习算法对降维后的特征进行分类实现比较满意的追踪效果。

1 基于压缩感知理论的在线多示例学习算法

本文提出了利用压缩感知理论提取目标与背景的特征,然后再用在线多示例分类器进行分类的追踪算法。通过稀疏测量矩阵对提取前景目标和背景的特征进行降维,作为在线学习更新分类器的正样本和负样本,然后使用多示例分类器去分类下一帧图像的目标。

1.1 基于压缩感知的目标表达

压缩感知被广泛用于许多研究领域,包括信号处理和目标追踪。文献^[17]提出在压缩感知理论的基础上利用符合压缩感知 RIP 条件的随机预测矩阵对多尺度图像特征进行降维,用降维后的特征表示目标模型。

$\mathbf{R} \in \mathbb{R}^{n \times m}$ 表示测量矩阵, n 、 m 分别表示 \mathbf{R} 的行数与列数,可以将一个高维图像空间的 $x \in \mathbb{R}^m$ 变换到一个低维的空间 $v \in \mathbb{R}^n$,其数学表达式为:

$$v = \mathbf{R}x \quad (1)$$

式中 $n \ll m$ 。如果矩阵 \mathbf{R} 满足 RIP 条件我们将能够以很小的错误概率从 v 中重构出 x ,而且保证 x (重构信号)保存了原来的大部分信息。

随机高斯矩阵是一个比较典型的满足 RIP 条件测量矩阵,矩阵元素满足 $N(0, 1)$ 分布,但是如果高维空间很大,其运算及存储消耗还是非常大的,这里采用一个非常稀疏的随机矩阵作为测量矩阵 \mathbf{R} :

$$r_{ij} = \sqrt{s} \times \begin{cases} +1 & \text{with probability } \frac{1}{2s} \\ 0 & \text{with probability } 1 - \frac{1}{s} \\ -1 & \text{with probability } \frac{1}{2s} \end{cases} \quad (2)$$

经证明 s 取 2 或 3 时, \mathbf{R} 满足 RIP 条件。从式 (1) 中我们可以知道, 用测量矩阵 \mathbf{R} 的一行内的一个非零元素感知 x 中的一个元素等价于将一个方形窗口滤波器与输入图像某一固定位置进行灰度卷积。为了实现尺度不变性, 对每一个样本 $z \in \mathbb{R}^{w \times h}$ 与多尺度的矩形滤波器 $\{h_{1,1}, \dots, h_{w,h}\}$ 进行卷积, 对每一个尺度的矩形滤波器定义为:

$$h_{i,j}(x,y) = \begin{cases} 1, & 1 \leq x \leq i, 1 \leq y \leq j \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

式中 i 和 j 分别为矩形滤波器的宽和高, 多尺度图像特征向量 $\mathbf{x} = (x_1, \dots, x_m)^T \in \mathbb{R}^m$ 。因此, 通过稀疏矩阵 \mathbf{R} 将 \mathbf{x} 投影到低维空间得到 \mathbf{v} , 此时低维向量 \mathbf{v} 中包含原高维向量 \mathbf{x} 的大部分信息, 可以通过 \mathbf{v} 高概率地重构出 \mathbf{x} 。

1.2 多示例学习分类算法对特征进行分类

由压缩感知的描述可以知道, 低维子空间可以很好的保留高维图像空间的信息, 所以我们用稀疏测量矩阵提取目标和背景的特征, 作为在线学习更新分类器的正负样本。因此我们将降维后的样本分为正负两个包, 当包中至少有一个示例为正时, 则该包为正包, 反之为负包。对多尺度图像特征向量 $\mathbf{x} = (x_1, x_2, \dots, x_m)^T \in \mathbb{R}^m$ 压缩后变为低维特征向量 $\mathbf{v} = (v_1, v_2, \dots, v_n)^T \in \mathbb{R}^n$ 。由于多示例学习的样本由不同的包组成, $V_i = \{v_{i1}, v_{i2}, \dots, v_{ij}\}$ 代表样本的包, v_{ij} 为包中的事例。提取正负样本分别放入正负包中, 每一个事例 v_{ij} 的标签为 $y_i \in \{0, 1\}$, 而包的标签为 $y_i = \max(\{y_{ij}\})$ 。

基于特征压缩的在线多示例学习算法利用梯度下降 boosting 框架^[7]通过最大化样本包的可能性函数对分类器进行训练:

$$L(H|\{V_i\}) = \sum_{i=1}^L (y_i \log(p(y_{ij}=1|V_i)) + (1-y_i) \log(1-p(y_{ij}=1|V_i))) \quad (4)$$

可能性函数是根据包而不是示例进行定义的, 因为示例的标签在训练中是不知道的, 然而我们的目标是通过训练分类器来估计 $p(y|\mathbf{v})$ 。因此我们需要根据示例表达出包的可能性函数 $p(y_i|V_i)$ 。在文献[15~16]中用 Noisy-OR (NOR) 模型来表示包为正的的概率:

$$p(y_i|V_i) = 1 - \prod_j (1 - p(y_{ij}=1|V_i)) \quad (5)$$

对于每一个示例, 示例为正的的概率:

$$p(y_{ij}=1|V_i) = \sigma(H_{ij} + h_m(v_{ij})) \quad (6)$$

在实际的目标追踪中, 应该满足实时性要求, 所以训练样本要采用在线学习的训练方法。采用增强学习的框架, 进行优化目标函数 J , 从而顺序选取弱分类器:

$$(h_k, \alpha_k) = \operatorname{argmax}_{h \in H, \alpha} J(H_{k-1} + \alpha h) \quad (7)$$

式中: H_{k-1} 为前面 $k-1$ 弱分类器组成的强分类器, H 为所有可能弱分类器的集合。目标函数 J 对整个训练数据集进行的计算。

当前帧训练数据集 $\{(V_1, y_1), (V_1, y_2), \dots, (V_n, y_n)\}$ 其中 $V_i = \{v_{i1}, v_{i2}, \dots, v_{ij}\}$ 用来更新分类器。算法中共有 M 个候选弱分类器, 通过最小化损失函数 $L(H|\{V_i, y_i\}_{i=1,2,\dots,N})$ 从中选取由 K 个最好的弱分类器。

$$h_k = \operatorname{argmax}_{h \in \{h_1, h_2, \dots, h_m\}} L(H_{k-1} + h) \quad (8)$$

将选择出的 K 个弱分类器进行组成强分类器。

$$H(\mathbf{v}) = \sum_{k=1}^K h_k(\mathbf{v}) \quad (9)$$

用多示例分类器对所有的图像样本进行估计 $p(y|\mathbf{v})$, 通过 $l(\mathbf{v}^*) = l(\operatorname{arg max}_{\mathbf{v}} p(y=1|\mathbf{v}))$ 更新目标位置。

1.3 弱分类器的构造

特征 f_k 和 4 个参数 $(\mu_1, \sigma_1, \mu_0, \sigma_0)$ 的在线估计表示的对比数来构成弱分类器。

$$h_k(\mathbf{v}) = \log \left[\frac{p(f_k(\mathbf{v})|y=1)}{p(f_k(\mathbf{v})|y=0)} \right] \quad (10)$$

弱分类器 $h_k(\mathbf{v})$ 的条件概率服从高斯函数分布:

$$p_k(f_k(\mathbf{v})|y=0) \sim N(\mu_0, \sigma_0), p_k(f_k(\mathbf{v})|y=1) \sim N(\mu_1, \sigma_1)。更新规则:$$

$$\mu_1 \leftarrow \lambda \mu_1 + (1-\lambda) \bar{\mu} \quad (11)$$

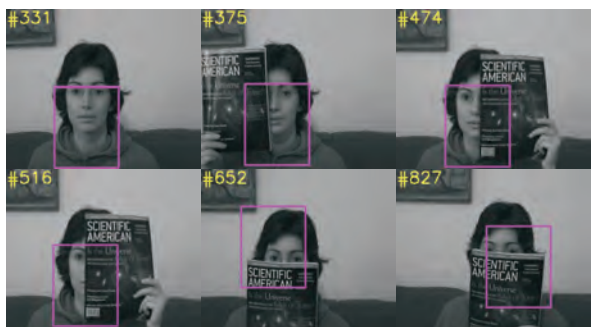
$$\sigma_1 \leftarrow \sqrt{\lambda(\sigma_1)^2 + (1-\lambda) \frac{1}{N} \sum_{j=0,1}^{N-1} (f_k(\mathbf{v}) - \mu)^2 + \lambda(1-\lambda)(\mu_1 - \mu)^2} \quad (12)$$

式中 λ 为学习率, μ_0, σ_0 更新规则与 μ_1, σ_1 类似。

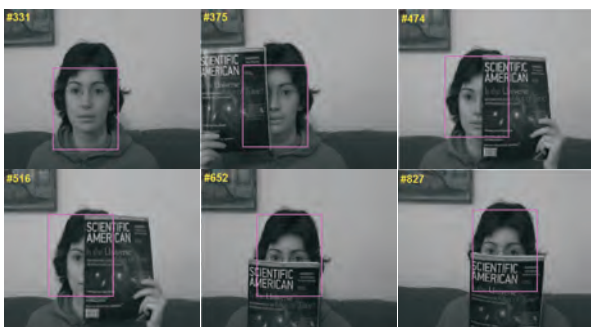
2 实验结果

为了验证基于压缩感知的在线多示例学习算法的有效性和鲁棒性, 分别使用该算法对不同的图像序列进行了实验。选取进行实验的图像序列均涉及到了目标的部分遮挡, 光照、姿态的变化以及一些固有的畸变。图 1~图 3 展现视频序列的追踪结果。

图 1 表示一个脸部遮挡的图像序列。这个实验主要为了验证算法在处理目标遮挡时的性能。图中分别说明了目标部分遮挡时的追踪情况。可以看出, 当书遮挡人脸的时候, 提取的目标特征发生变化, 而在线多示例学习算法不能很好地处理目标遮挡的情况, 导致追踪框发生了明显的偏离现象。从图 1(b) 中可以看到提出的算法能够很好地解决这种特征的变化, 不会因为目标遮挡而发生追踪不准确的情况, 在整个视频序列中追踪效果都比较好。



(a) 在线多示例学习算法



(b) 基于压缩感知的在线多示例学习算法

图1 脸部遮挡图像集(帧数 331,375,474,516,652,827)

Fig.1 Occluded face dataset (frame 331,375,474,516,652,827)

图 2 为摇滚歌唱图像序列。这个实验主要为了验证算法在处理背景杂乱、光照变化严重及目标迅速发生移动的情况的追踪性能。在 60 帧时光照发生突然的变化,从图 2(a)可以看出目标几乎完全丢失。由于目标的头部运动非常的迅速,图 2(a)追踪的目标虽然没有丢失但是偏移现象比较严重,而图 2(b)中目标追踪框几乎没有发生任何偏移。所以得到结论,在目标快速运动和光照发生强烈变化的情况下本算法能很好地适应这种变化,达到令人满意的追踪效果。

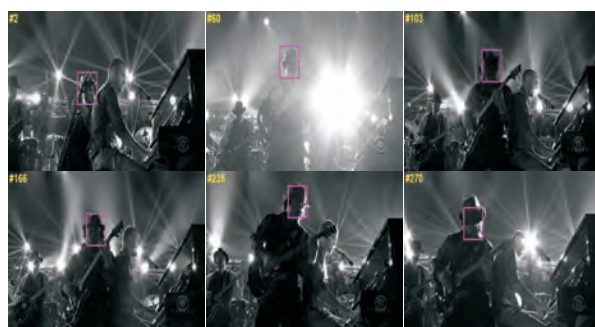
图 3 为鱼镜头头拍摄的图像,以路边的路灯作为追踪的目标。由于鱼镜头头具有球面透视模型,产生的图像存在固有的畸变,镜头周围畸变比较严重中间畸变比较小。图 3(a)中可以看出目标发生了严重的偏移,即使在图像中间没有形变的时候,追踪效果也不理想,目标几乎丢失。而图 3(b)中不论目标是否发生严重的变形追踪效果一直保持很好。这个实验充分验证了提出的算法在目标发生严重形变的情况下也能够达到很好的追踪效果,而在线多示例学习算法却不能很好地解决这种固有的形变。

图 4 分别为脸部遮挡和摇滚图像序列追踪性能的误差分析。图 4(a)前 230 帧 2 种算法目标追踪效果都比较好,但是在后面序列中在线多示例学习算法误差比较大,提出的算法达到令人满意的追踪效果。图 4(b)由于目标背景杂乱运动迅速,在线多示例学习算法不能很好解决特征变化,特别是在 60

帧时光照发生突然变化,目标几乎完全丢失,因此误差曲线发生了急剧地上升。然而提出的算法却能保持平衡。因此实验表明提出的算法在视觉追踪中性能优越。



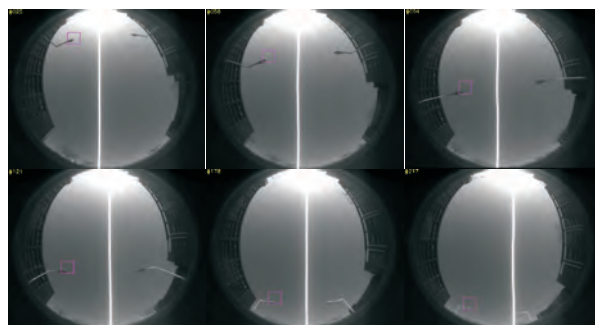
(a) 在线多示例学习算法



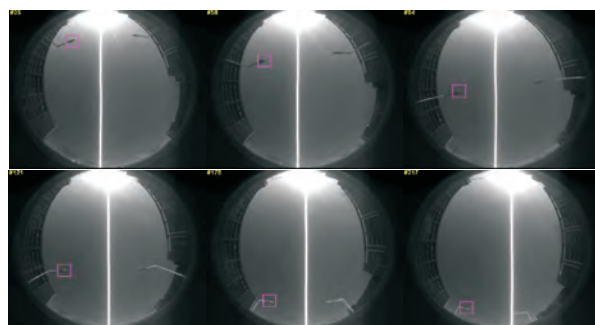
(b) 基于压缩感知的在线多示例学习算法

图2 摇滚图像集(帧数 2,60,103,166,235,270)

Fig.2 The shaking dataset (frame 2, 60, 103, 166, 235, 270)



(a) 在线多示例学习算法



(b) 基于压缩感知的在线多示例学习算法

图3 鱼眼图像集(帧数 25,58,94,121,178,217)

Fig.3 Fish eyes dataset (frame 25, 58, 94, 121, 178, 217)

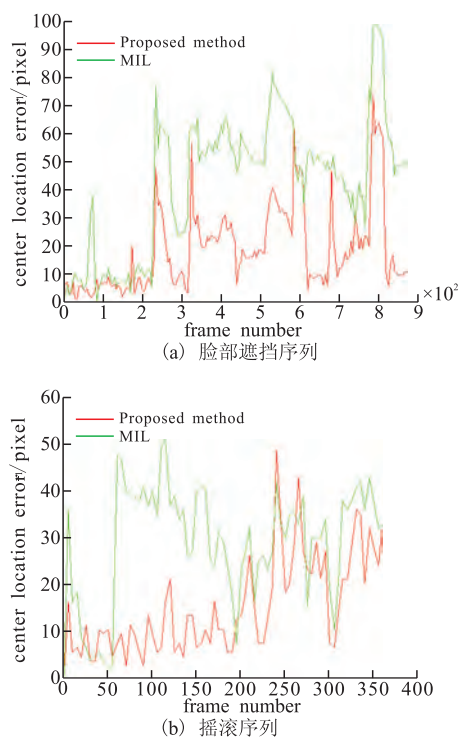


图 4 测试序列的中心误差

Fig.4 Error plots of test sequences

3 结语

针对追踪过程中由于特征变化而追踪失败的问题,提出了一种简单有效的追踪算法,即基于压缩感知的在线多示例学习算法。通过对压缩感知和在线多示例学习的深入研究,提出的方法解决了视觉追踪中挑战性问题。实验结果证明提出的算法不仅实现了目标的稳定追踪,而且满足实时性要求。由于稀疏理论的蓬勃发展,将稀疏知识成功地运用到视频追踪上仍然是很艰巨的任务。下一步我们将深入研究压缩感知理论的稀疏表示问题,并提出更加新颖的思想来解决视频追踪中难以解决的问题。

参考文献 (References):

[1] Lee K C, Kriegman D. Online learning of probabilistic appearance manifolds for video-based recognition and tracking[C]// IEEE computer society conference on computer vision and pattern recognition[S.l.]: IEEE press, 2005: 852-859.

[2] Ross David A, Lim Jongwoo, Lin Rwei-Sung, et al. Incremental learning for robust visual tracking[J]. International journal of computer vision, 2008, 77(1-3): 125-141.

[3] Wright J, Yang A Y, Ganesh G, et al. Robust face recognition via sparse representation[J]. IEEE transaction on pattern analysis and machine intelligent, 2009, 31(2): 210-227.

[4] Mei X, Ling H. Robust visual tracking using l1 minimization [C]// IEEE 12th international conference on computer vision.

Kyoto: IEEE press, 2009: 1436-1443.

- [5] Liu B Y, Yang L, Huang J Z, et al. Robust and fast collaborative tracking with two stage sparse optimization[C]// IEEE European conference on computer vision, Heraklion, Crete: [s.n.], 2010: 624-637.
- [6] Mason L, Baxter J, Bartlett P, et al. Boosting algorithms as gradient descent[C]// Advances in neural information processing systems. Denver, Colorado: NIPS conference, 1999: 512-518.
- [7] Friedman J H. Greedy function approximation: A gradient boosting machine[J]. The annals of statistics, 2001, 29(5): 1189-1232.
- [8] Avidan S. Ensemble tracking[C]// IEEE transactions on pattern analysis and machine, 2007, 29(2): 261-271.
- [9] Tian M, Zhang W, Liu F. On-line ensemble SVM for robust object tracking [C]// Asian conference on computer vision. Tokyo: [s.n.], 2007: 355-364.
- [10] Grabner H, Bischof H. Online boosting and vision[C]// IEEE international conference on computer vision and pattern recognition. Piscataway: IEEE press, 2006: 260.
- [11] Grabner H, Grabner M, Bischof H. Real-time tracking via on-line boosting[C]// Proc of British machine vision conference. [S.l.]: BMVA press, 2006: 47-56.
- [12] Tang F, Brennan S, Zhao Q, et al. Co-tracking using semi-supervised support vector machines[C]// IEEE 11th international conference on computer vision, Rio de Janeiro: IEEE press, 2007: 1-8.
- [13] Mallapragada P K, Jin R, Jain A K, et al. SemiBoost: boosting for semi-supervised learning[J]. IEEE transaction on pattern analysis and machine intelligent, 2009, 31(11): 2000-2014.
- [14] Grabner H, Leistner C, Bischof H. Semi-supervised on-line boosting for robust tracking[J]. ECCV 2008 computer vision, 2008, 5302: 234-247.
- [15] Babenko B, Yang M, Belongie B. Visual tracking with online multiple instance learning[C]// IEEE conference on computer vision and pattern recognition, Miami, FL: IEEE press, 2009: 983-990.
- [16] Babenko B, Yang M, Belongie S. Robust object tracking with online multiple instance learning [J]. IEEE transactions on pattern analysis and machine intelligence, 2011, 33(8): 1619-1632.
- [17] Zhang Kaihua, Zhang Lei, Yang Ming-Hsuan. Real-time compressive tracking[J]. European conference on computer vision, 2012, 7574: 864-877.
- [18] Candes E, Tao T. Decoding by linear programming[J]. IEEE trans on inform theory, 2005, 51(12): 4203-4215.

(编辑:徐楠楠)