

基于在线判别分布域特征选择的鲁棒跟踪算法

郭强^{1,2}, 吴成东¹, 赵迎春²
(1. 东北大学 信息科学与工程学院, 辽宁 沈阳 110819; 2. 中国刑事警察学院 图书馆, 辽宁 沈阳 110854)

摘 要: 针对基于检测目标跟踪中的特征描述子 Haar-like 表征能力不强和易引入错误训练样本导致目标漂移的问题,提出了一种利用分布域描述算子进行示例层级的在线判别特征选择跟踪算法. 首先,用软直方图方法快速近似得到分布域特征,并利用此描述算子取代 Haar-like 特征有效表示目标的外观信息. 然后,基于示例级样本的先验信息进行有监督学习,利用在线判别特征选择算法选择最佳的分布域层特征以减少漂移现象发生. 实验利用多场景视频标准测试库及新的评价指标进行验证,结果表明本文算法性能优于对比算法.

关 键 词: 视觉跟踪;分布域描述算子;特征选择;软直方图;监督学习

中图分类号: TP 391 **文献标志码:** A **文章编号:** 1005-3026(2017)03-0305-05

Robust Visual Tracking with Distribution Fields Feature Selection Based on Online Discrimination

GUO Qiang^{1,2}, WU Cheng-dong¹, ZHAO Ying-chun²
(1. School of Information Science & Engineering, Northeastern University, Shenyang 110819, China; 2. Library, National Police University of China, Shenyang 110854, China. Corresponding author: GUO Qiang, E-mail: royingchina@163.com)

Abstract: The Haar-like features used in MIL (multiple instance learning) trackers are not efficient to represent the appearances of the targets, and the noise samples are prone to be involved for classifier training phase, then drift in targets may happen. To solve these problems, an online discriminative feature selection (ODFS) tracking algorithm based on distribution fields (DFs) descriptors at instance level was proposed. Firstly, soft histogram method is manipulated to fastly approximate DFs, and the Haar-like features are replaced with the layers of DFs, which are adopted to represent appearance information. Then, supervised learning with prior information of instance labels is conducted; the ODFS algorithm is used to select the most optimal discrimination layer features, which can handle drift more effectively. The proposed tracking method are tested in benchmark dataset of a large variety of scenarios and under new evaluation indexes. Experimental results show the effectiveness of the algorithm.

Key words: visual tracking; distribution fields descriptors; feature selection; soft histogram; supervised learning

视觉目标跟踪是计算机视觉领域的研究热点,广泛应用于场景监控、人机交换等方面. 众多学者在该领域展开了研究工作,已经取得了一些成果,较新的流行算法如 MIL (multiple instance learning) 跟踪^[1]、DFs (distribution fields) 跟踪^[2]等. 在线学习跟踪算法属于判别式跟踪方法,在此基础上发展出了当前流行的基于检测的跟踪算法. 文献[1]首次将多示例学习应用在目标跟踪上,得到了较好的跟踪效果. 但 MIL 存在着一些不足,文献[1,3]中用 Haar-like 特征描述

外观并不精准且选取的特征具有较少的辨别信息. 其次, MIL 使用包层级的概率获得弱分类器而没有直接采用示例层级, 影响效率, 因而应提高选取弱分类器的泛化能力. 文献[3]融合正样本集的先验信息提出一种新的在线判别特征选择算法, 比 MIL 特征选取更高效.

近来, Laura 等^[2]提出基于分布域描述子 (DFs) 的跟踪算法, 能很好地表征图像外观. 然而, 该算法属于匹配式跟踪, 没有利用背景信息, 尤其是在目标和相邻背景的 DFs 中不同层之间存在相似特性时易发生漂移. 因此, 选取最具有辨别性的 DFs 层特征有利于优化跟踪结果. 此外, DFs 是基于平滑局部直方图计算, 用群编码可提高 DFs 计算效率.

针对上述分析, 本文首先采用群编码理论近似计算 DFs 特征, 提高了文献[2]中 DFs 平滑直方图的计算效率. 其次, 采用 ODFS 算法从 DFs 特征中选取最有判别性的 DFs 层特征建立示例级分类器, 改进了 MIL 中从样本包获取弱分类器的方法. 最后, 采用自适应样本更新策略增加算法鲁棒性.

1 相关工作

1.1 基于平均移位柱状图和群编码的 DFs 近似法

分布域 DFs^[2]是由一组特征的概率分布构成, 通过量化图像的特征信息构成的一个复杂数据结构. DFs 是维度为 $2 + D$ 的矩阵, 前两维是图像的宽度和高度, D 是特征空间维度. 建立 DFs 有两个步骤: 图像扩展和平滑. 首先, 1 幅图像被分成若干层, 每层包含一组值, 或相近的灰度值, 将图像扩展成 DFs. 然后, 对扩展后的 DFs 进行高斯滤波平滑, 可降低目标模型对光照变化和背景噪声的敏感.

$$\text{df}_s = \text{df}(k) * h_{(k, \sigma_s)}(x; y), \quad (1)$$

$$\text{df}_{ss} = \text{df}_s(i, j) * h_{(\sigma_f)}(z). \quad (2)$$

其中: df_s 为平滑后的分布域; $\text{df}(k)$ 为 k 所对应的分布域; $h_{(k, \sigma_s)}(x; y)$ 是标准差为 σ_s 的二维高斯核函数; $*$ 是卷积运算. 通常对图像平滑后会导致图像空间信息丢失, 破坏了原图信息, 采用高斯核函数对分布域的空间域和特征域分别进行平滑处理, 保持了每个像素的原有信息. h 函数表示标准差为 σ_f 一维高斯核函数; df_{ss} 是计算得到的分布域描述子.

上述 DFs 计算过程融合了卷积后的池化和

加权平均两个处理环节, 因而并没有明确得到 DFs 的准确统计特征. 在空间平滑阶段, 每个像素点位置的概率分布是加权直方图即式(1)的结果. 在特征域平滑阶段, 式(2)结果为概率分布直方图的平均值. 因而, DFs 的统计特性服从平均移位柱状图 (averaged shifted histogram, ASH)^[4], 从而属于 Scott 等^[5]提出的 ASH 概率密度估计理论范畴. ASH 利用渐近性质进行概率密度估计优于普通直方图^[6], 前者密度曲线更光滑且不过度依赖于边界点选择.

ASH 通过将每 m 个变化的子区间上的估计值进行平滑作为 $f(x)$ 的估计值, 将每个长度为 h 的区间进行划分, 令 $\delta = h/m$, $t_k = k\delta$, 得分段子区间 $I_k = [t_{k-1}, t_k]$, r_k 为在每个小区间内的样本点的个数, 则针对加权 ASH, 每个区间上概率分布为

$$\hat{f}_{\text{ASH}}(x; m) = \frac{1}{m} \sum_{i=1}^{m-1} w_m(i) r_{k+i}. \quad (3)$$

其中 w_m 为权重函数, 式(3)另一种常见形式如下:

$$\hat{f}_{\text{ASH}}(x; m) = \frac{1}{m} \sum_{i=1}^{m-1} \left(1 - \frac{|i|}{m}\right) r_{k+i}. \quad (4)$$

当 $w_m(i)$ 为高斯核函数时, 式(3)等同于计算 DFs 特征的池化操作. 对式(4)求极限后得^[7]

$$\lim_{m \rightarrow \infty} \hat{f}_{\text{ASH}}(x; m) = \frac{1}{nh} \sum_{j=1}^n B\left(\frac{x - x_j}{h}\right). \quad (5)$$

当 m 趋近于无穷大即 $B(\cdot/h)$ 由 $\lim_{m \rightarrow \infty} w_m(\cdot)$ 得到时, 在统计学期望值上, 计算 DFs 特征即转为核密度估计 KDE. 而群编码或信道表示是近似 KDE 的一种有效方法^[8], 因而可有效计算 DFs.

信息群编码是一种由生物学概念转换而来的数据表达方式, 数据值 x 可通过非线性转换而编码为一个信道向量 \mathbf{c} . 更多理论内容见文献[8].

$$\mathbf{c} = [c_1, \dots, c_N] = [B(x - \xi_1), \dots, B(x - \xi_N)]. \quad (6)$$

其中: B 是局部对称非负的核函数; ξ_n , $n \in [1, N]$ 是呈均匀分布的各信道中心. 带宽参数即各信道中心是整数间隔 h , 软直方图系数定义为

$$c_k = \frac{1}{nh} \sum_{j=1}^n B(x_j/h - k), k \in \mathbf{N}. \quad (7)$$

式(7)参照式(5)可知, 群编码与 KDE 相近, 但用各核函数代替了各矩形柱划分样本数据. 该法对新加入的样本 x_j 只需对总样本计算一次求和操作, 而 KDE 需要移动核函数到每个样本点所在的位置, 对每个样本进行求和计算, 因此群编码效率更高. 式(7)中 $B(x)$ 为二次 B 样条曲线.

将式(2)改用群编码的式(7)计算,需要对前者带宽参数 h 同后者的参数 σ_{eff} 进行转换, σ_{eff} 是子区间块与特征域平滑高斯核 σ_f 卷积. 因只考虑这一个参数,选择同一个有效方差的联合核函数就可得到最佳近似效果. 假定矩形块宽度 $w = 16$,

$$\sigma_{\text{eff}} = \sqrt{\int_{-w/2}^{w/2} \frac{1}{w} (\sigma_f^2 + t^2) dt} = \sqrt{\sigma_f^2 + \frac{w^2}{12}}. \quad (8)$$

参数 $h = 2\sigma_{\text{eff}}$, 群编码由 $\lceil 256/h \rceil + 2$ 个信道构成.

1.2 基于检测跟踪的判别式模型

本文将一个目标样本 x 的特征向量表示为 $f(x) = (f_1(x), \dots, f_K(x))^T$, $h(x)$ 为判断样本类别的强分类器,是由若干个弱分类器 $\varphi_k(x)$ 组成,本文朴素贝叶斯分类器判别公式如下:

$$c(x) = P(y=1|x) = \sigma(h_K(x)). \quad (9)$$

假定每个特征是独立同分布的. $\sigma(\cdot)$ 是 sigmoid 函数, $y \in \{0, 1\}$ 是二值变量,表示样本正负类别. 本文用一组平滑后分布域的层特征表示样本,即 $f_k = \text{df}(x) = (\text{df}_x(k), \dots, \text{df}_x(K))^T$. 设 t 时刻样本 x 位置为 $l_t(x)$, $t-1$ 时刻目标位置是 $l_{t-1}(x^*) \in \mathbf{R}^2$, x^* 是对应的样本. 基于检测的跟踪算法首先在 t 时刻 $l_{t-1}(x^*)$ 目标位置附近采集一组样本集 $\mathbf{X}^\delta = \{x | \|l_t(x) - l_{t-1}(x^*)\| < \delta\}$, 搜索半径 δ , 然后对每个样本提取相应特征,计算所有满足条件的样本为正样本的概率 $P(y=1|x)$. 使用分类器对每个特征向量进行分类,通过计算最大分类响应 $x^* = \arg \max_x (c(x))$ 来找到跟踪位置. 确定 t 时刻的目标位置 $l_t(x^*)$ 后,据此选取新的正负训练样本对分类器更新. $\mathbf{X}^+ = \{x | \|l_t(x) - l_{t-1}(x^*)\| < \alpha\}$ 是新的正样本集, $\mathbf{X}^- = \{x | \beta < \|l_t(x) - l_{t-1}(x^*)\| < \gamma\}$ 是新的负样本集,其中 $\alpha < \beta < \gamma$; 标记为正样本的集合 $\{x_1^+, x_2^+, \dots, x_n^+\}$, 对应的分布域特征为 $\{\text{df}_1^+, \text{df}_2^+, \dots, \text{df}_n^+\}$. MIL^[1] 训练样本建立由弱分类器构成特征池 Φ , 再从池中选出 K 个弱分类器使分类置信度最大,并用贝叶斯分类器实现强分类器.

2 本文跟踪算法

2.1 在线判别分布域特征选择

因为构成 DFs 的层分布域 $\text{df}(k)$ 的数量远小于 Haar-like 特征数量,本算法将其作为目标特征向量. 每个弱分类器由 DFs 中的层特征构成, MIL 从特征池中穷尽选出弱分类器算法效率不高. 本

文用在线判别特征选择算法实现,训练分类器时不再用 MIL 跟踪算法中基于包的训练集进行,弱分类器选取直接通过优化示例层级概率实现.

分类器工作的原理是由式(9)计算得到样本 x 为真实目标的置信图,最大置信度对应的点即为跟踪位置. 假定样本空间可以划分为正样本区域 $R^+ = \{x, y=1\}$ 和负样本区域 $R^- = \{x, y=0\}$, 用双方区域内样本平均置信度差值作为双方分界线:

$$E_{\text{margin}} = \frac{1}{|R^+|} \int_{x \in R^+} c(x) dx - \frac{1}{|R^-|} \int_{x \in R^-} c(x) dx. \quad (10)$$

其中 $|R^+|$ 和 $|R^-|$ 是正负集合的基数. 在训练集中 N 个正样本集 $R^+ = \{x_i\}_{i=0}^{N-1}$, L 个负样本集合 $R^- = \{x_i\}_{i=N}^{N+L-1}$, 将积分用加法表示,式(10)转换为

$$E_{\text{margin}} \approx \frac{1}{N} \left(\sum_{i=0}^{N-1} \sigma \left(\sum_{i=1}^M \varphi_k(x_i) \right) - \sum_{i=N}^{N+L-1} \sigma \left(\sum_{i=1}^M \varphi_k(x_i) \right) \right). \quad (11)$$

式中样本 x_i 由一个特征向量 $(\text{df}_x(1), \dots, \text{df}_x(K))^T$ 表示,得到特征池 $\Phi = \{\varphi_m\}_{m=1}^M$. 本算法只要从 Φ 中选出 K 个弱分类器,保证最大化正样本集的平均置信度并抑制负样本集的平均置信度,即:

$\{\varphi_1, \dots, \varphi_k\} = \arg \max_{\varphi_1, \dots, \varphi_k \in \Phi} E_{\text{margin}}(\varphi_1, \dots, \varphi_k)$. 本文从 Φ 中顺序选择弱分类器最大化 E_{margin} , 结合式(11)得

$$\varphi_k = \arg \max \left\{ \left(\sum_{i=0}^{N-1} \sigma(h_{k-1}(x_i) + \varphi(x_i)) \right) - \left(\sum_{i=N}^{N+L-1} \sigma(h_{k-1}(x_i) + \varphi(x_i)) \right) \right\}. \quad (12)$$

利用 ODFS 准则选择弱分类器,推导细节见文献[3].

2.2 增强的在线判别分布域特征选择跟踪算法

增强的在线判别分布域特征选择跟踪算法如下:

输入 $t+1$ 时刻视频图像.

1) 采集一组样本集,用式(7)群编码系数替换式(2)特征池化,近似得到分布域特征集 $\{\text{df}_k(x)\}_{k=1}^K$;

2) $h_k(x)$ 代入式(9)对每个特征向量 $\text{df}(x)$ 进行分类,通过计算最大分类器响应确定跟踪位置 $l_t(x^*)$, 其中 $x^* = \arg \max_{x \in \mathbf{X}^\alpha} (c(x))$;

3) 利用 1.2 节采样策略获得 2 组样本集;

4) 由 ODFS 算法提取 DFs 层特征,更新 DFs 模型.

输出 跟踪结果 (x_t, y_t) .

上述算法中涉及的子算法 ODFS 算法如下:

输入 $\{x_i, y_i\}_{i=0}^{N+L-1}, y_i \in \{0, 1\}$, 初始化

$h_0(x_i)$.

- 1) 用 $\{x_i, y_i\}_{i=0}^{N+L-1}$ 更新特征池 $\Phi = \{\varphi_m\}_{m=1}^M$;
- 2) 更新平均弱分类器值 $\bar{\varphi}_m^+$ 和 $\bar{\varphi}_m^-$;
- 3) **for** $k=1$ to K do
- 4) 计算最速下降方向;
- 5) **for** $m=1$ to M do
- 6) 计算 E_{margin} ;
- 7) **end for**
- 8) $m^* = \arg \max E_{\text{margin}}$, 计算得 $h_k(x_i)$;
- 9) **end for**
- 输出 $h_K(x), c(x)$.

3 实验结果和分析

3.1 实验说明

为了验证本文算法的有效性,本文进行了 2 组实验,并采用了新的跟踪算法评价指标. 本实验平台基于 Matlab R2010b, CoreI7 3.4 GHz, 4 GB 内存. 本文所采用的测试标准库是由 Wu 等^[9]提出的国际上权威的跟踪算法评测库. 本文算法与基于信道表示的 DF 跟踪 (CBDF)^[8]、分布域跟踪 (DFMIL)^[2] 和基于分布域特征的在线多示例学习跟踪 (DFMIL)^[10] 进行比较. 将图像的像素光亮度特征划分 16 段, 特征池中候选特征数量 $M=16$, 远低于 MIL 跟踪器中 250 个 Haar-like 特征数量. 本算法从特征池中选择 5 个特征即 $K=5$, 而 MIL 方法需要选择 50 个特征. 对下一帧新目标位置搜索半径 $r_s=25$ 像素; 正样本半径 $r_p=4$; 负样本集圆环内径 $r_n=8$.

3.2 实验结果

跟踪算法性能的评价标准通常是成功率和准确率, 成功率定义为 $\text{score} = (A \cap B) / (A \cup B)$, 其中 A 是目标矩形框的真实位置, B 是算法的跟踪结果矩形框位置, \cap 和 \cup 分别表示交集和并集. 当 $\text{score} > t_0$, 则认为该跟踪方法在此帧图像跟踪成功, t_0 通常取 0.5. 但单一阈值 t_0 并不准确, 成功率指标^[9]由曲线下不同面积重叠阈值得到.

图 1 给出了两种不同的成功率曲线, 为了验证跟踪算法对测试序列初始化位置和时序的鲁棒性, 实验用两个评价指标: 时间鲁棒性成功率 (每个测试序列划分 20 个子序列, 分别测试) 和空间鲁棒性成功率 (初始位置附近扰动, 对不同标定

位置进行测试), 本文均具有更高的成功率.

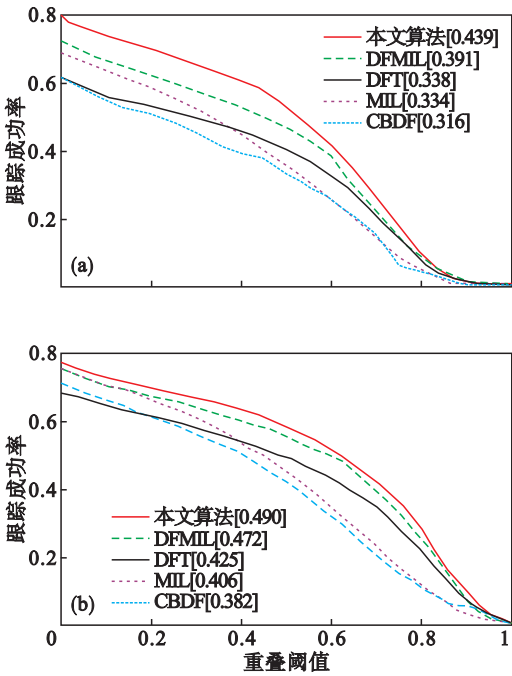


图 1 空间鲁棒性跟踪成功率和时间鲁棒性跟踪成功率曲线

Fig. 1 Spatial robustness evaluation and temporal robustness evaluation curves

(a)—空间鲁棒性; (b)—时间鲁棒性.

表 1 是 6 种跟踪器在部分标准测试序列的运行时间比较结果. 本文跟踪算法处理速度约为 12 帧/s, 略低于 CBDF 算法. ODFS 算法在选择一个特征后, 只需要更新分类器的梯度, 比 MIL 算法更有效率. 因为本文算法和 DFMIL 算法需要在每帧重新选择特征, 所以两种算法跟踪速度要低于 DFT. 但本文算法运行速度比 DFMIL 更快, 因采用了软直方图近似计算 DF 特征及快速的特征选择方法.

表 1 不同算法平均运行速率比较

Table 1 Comparisons of average tracking time in different algorithm (FPS) (帧/s)

跟踪方法	本文算法	DFT	DFMIL	ODFS	MIL	CBDF
平均运行速率	12	14	9	28	6	16

在图 2a Car4 序列中, 目标汽车经历光照的逐渐变化和小幅的尺度变化 (见第 187 帧), 在未出现阴影时, 4 种算法都能准确跟踪目标 (见第 50 帧), 本文算法利用 ODFS 选出最具有辨别性的分布域层特征以便更精确地跟踪目标. 当目标驶出桥下阴影瞬间经历了比较剧烈的光照变化 (见第 235 帧), 本文算法通过对所有正样本的弱分类器的输出求平均值来降低其他噪音的干扰, 其他算法均出现了漂移, DFMIL 跟踪失败 (见第

600 帧). 图 2b Matrix 序列中, 目标姿态持续变化, 并存在强环境光变化. MIL 和 DFT 从开始阶段出现偏移, 误差不断累积. DFMIL 和 DF 跟踪精度上都不如本文算法, 基于示例层级的有效样本模板更新发挥了作用. 同时, DFs 的池化技术对背景干扰鲁棒性更强(见第 50 帧). 图 2c David3

序列中, 目标经历部分遮挡、形变以及复杂背景干扰. 在前 100 帧, 几种算法都可以实现稳定跟踪, 当目标被树木短时遮挡后, 本文算法更准确定位目标, DFT 算法在短暂遮挡后出现停滞, 直到目标折返到原跟踪框误判停留区域后恢复跟踪.

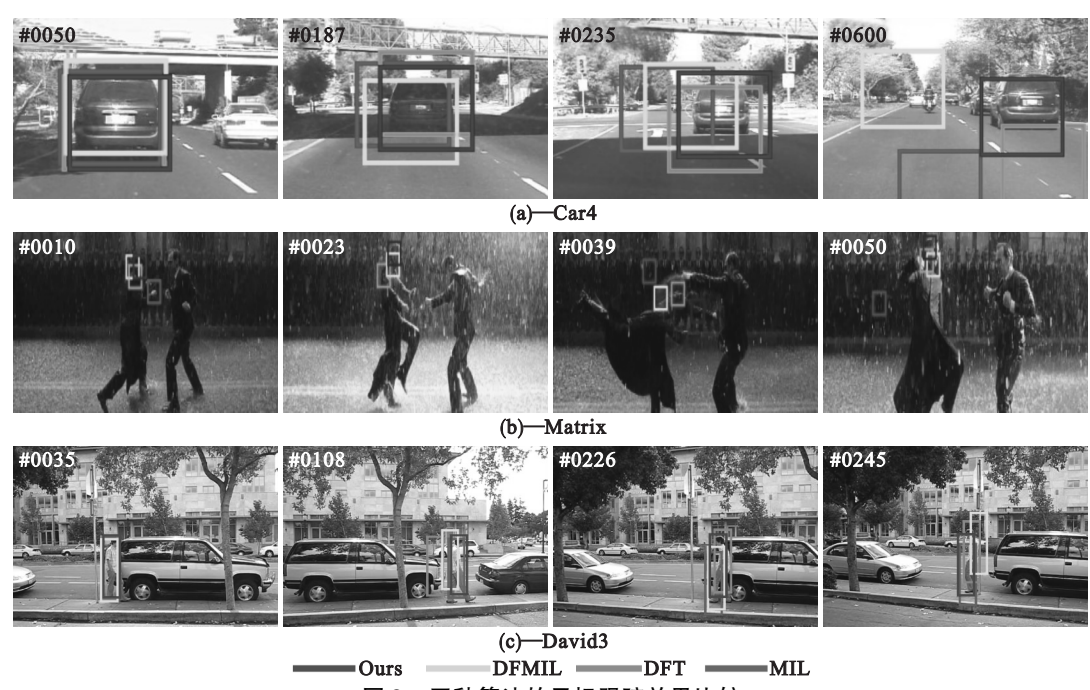


图 2 四种算法的目标跟踪效果比较

Fig. 2 Tracking results of the sequences with the four algorithms

综上所述, 本文算法的跟踪效果稳定, 能选出最有效的特征, 将样本的重要程度直接同分类器分数关联. 鲁棒的特征结构外观建模、更合理的示例级特征选择算法结合监督学习框架使得本文算法可有效防止漂移问题发生, 尤其是在光照变化和严重遮挡情况下.

4 结 论

本文提出一种新的高效跟踪算法, 结合产生式和判别式两种分类模型的优点, 利用在线判别特征选择方法, 同时将先验信息整合到监督学习算法中, 从而选出最具判别性的目标分布域层级特征. 文中采用软直方图来计算分布域特征, 提高了计算速度. 实验结果表明, 本文算法对光照变化、部分遮挡、形变等复杂环境具有较好鲁棒性, 兼具较快的运行速度.

参考文献:

[1] Babenko B, Yang M H, Belongie S. Visual tracking with online multiple instance learning [C]//Proceedings of the 2009 IEEE Computer Society Conference on Computer

Vision and Pattern Recognition. Miami, 2009; 983 – 990.

[2] Laura S L, Erik L M. Distribution fields for tracking [C]//Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Providence, 2012; 1910 – 1917.

[3] Zhang K, Zhang L, Yang M H. Real-time object tracking via online discriminative feature selection [J]. *IEEE Transaction on Image Processing*, 2013, 22 (12) : 4664 – 4677.

[4] Scott D W. Multivariate density estimation; theory, practice, and visualization [J]. *Journal of the American Statistical Association*, 1994, 425 (89) : 359 – 360.

[5] Scott D W. Averaged shifted histograms: effective nonparametric density estimators in several dimensions [J]. *The Annals of Statistics*, 1985, 13 (3) : 1024 – 1040.

[6] Felsberg M. Enhanced distribution field tracking using channel representations [C]// IEEE International Conference on Computer Vision Workshops. Sydney, 2013; 121 – 128.

[7] Jonsson E, Felsberg M. Reconstruction of probability density functions from channel representations [J]. *Image Analysis*, 2005, 3540 (1) : 491 – 500.

[8] Felsberg M. Adaptive filtering using channel representations [M]. London : Springer London, 2012; 31 – 48.

[9] Wu Y, Lim J, Yang M H. Online object tracking: a benchmark [C]//Proceeding of the 2013 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Portland, 2013; 2411 – 2418.

[10] Ning J F, Shi W, Yang S, et al. Visual tracking based on distribution fields and online weighted multiple instance learning [J]. *Image and Vision Computing*, 2013, 31 (11) : 853 – 863.