

基于在线学习判别性外观模型的多目标跟踪算法

侯建华 边群星 项俊

(中南民族大学 电子信息工程学院, 武汉 430074)

摘要 在基于检测的跟踪框架下,设计了一种在线学习的判别性外观模型并应用于多目标跟踪.对检测器输出的相邻帧间检测响应做保守关联,生成短小可靠的轨迹片;利用目标轨迹时空域约束条件,从轨迹片中提取训练样本及特征;采用 Adaboost 算法在线生成目标外观的判别性模型,计算轨迹片之间的外观相似度;最后采用匈牙利算法,经过多次迭代得到每个目标的完整轨迹.对实验结果做了定量和定性分析,结果表明:所设计的算法提高了跟踪精度,在复杂场景下能够较好地完成多目标跟踪任务.

关键词 多目标跟踪; 轨迹片; 外观模型; Adaboost 算法

中图分类号 TP391.41 文献标识码 A 文章编号 1672-4321(2017)01-0081-06

Multi-Object Tracking Algorithm Based on Online Learned Discriminative Appearance Model

Hou Jianhua, Bian Qunxing, Xiang Jun

(College of Electronic Information Engineering, South-Central University for Nationalities, Wuhan 430074)

Abstract An online learned discriminative appearance models is designed for multi-object tracking (MOT) under the tracking by detection framework. Firstly, by low level association strategy, short yet reliable tracklets are generated among detection responses which are produced by a pre-trained detector. Then, the training samples are collected by using spatial-temporal constraints of target trajectory, and features are extracted for representation of the appearance model. We adopt online Adaboost algorithm to train the discriminative appearance model, by which the appearance similarity between tracklets can be calculated. Finally, the Hungarian algorithm is used via several iterations to obtain the complete trajectory for each target. Experiments are conducted quantitatively and qualitatively. The results show that the proposed method has improved the tracking accuracy, and has satisfactory performance in complex scene.

Keywords multi-object tracking; tracklets; appearance model; Adaboost algorithm

近十年来,随着目标检测技术快速发展,基于检测的跟踪方法^[1-6]已成为当前多目标跟踪领域研究热点;该方法将多目标跟踪问题转化为数据关联问题,利用目标检测响应所包含的外观、运动、位置等信息设计关联模型,以此度量迹片间的连接概率(即相似度),实现轨迹片最优关联. Huang 等人^[6]提出的多级关联是一种比较流行的方法,该方法先通过初级关联得到可靠而短小的轨迹片,再通过逐级关联方式,计算轨迹片间的连接概率,应用 Hungarian 算法^[8]得到最优关联解,逐步将短小轨迹片关联为长的目标轨迹.但文献^[6]使用基于颜色直

方图的外观相似度,其外观模型简单,对位置相近、外观相似的易混淆目标区分性较差,在跟踪过程中容易产生目标轨迹间隙、及身份标识混淆.

针对上述问题,本文提出了一种基于在线学习判别性外观模型的多目标跟踪算法:利用目标轨迹时空域约束条件,从短小轨迹片中采集训练样本,分别提取颜色及纹理特征;采用 Adaboost 在线学习算法^[7]训练分类器,构建目标判别性外观模型,计算轨迹片之间的外观相似度;利用 Hungarian 算法^[8]对轨迹片多次迭代关联,最终得到可靠的目标完整轨迹.实验结果表明在拥挤、遮挡频繁发生的背景复

收稿日期 2017-01-16

作者简介 侯建华(1964-)男,教授,博士,研究方向:计算机视觉、模式识别, E-mail: hou8781@126.com

基金项目 国家自然科学基金资助项目(61671484);中南民族大学中央高校基本科研业务费专项资金(CZW15013)

杂中,该算法可以较好地实现准确稳定的目标跟踪.

1 训练样本采集

为提高算法的有效性,采用滑动窗^[5]处理模式,在每一个滑动窗内,从短小轨迹片中提取目标正负样本.

1.1 短小轨迹片

设 $R = \{r_i\}$ 代表检测器输出的检测响应集,其中第 i 个响应为 $r_i = (x_i, y_i, s_i, t_i)$, (x_i, y_i) 、 s_i 、 t_i 分别表示目标的位置、大小、所在帧序号. 定义相邻两帧间两个响应 r_i 与 r_j 的连接概率如下:

$$P_{score}(r_j | r_i) = \begin{cases} A_{appr}(r_j | r_i) & \text{if } t_j - t_i = 1 \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

其中 $A_{appr}(r_j | r_i)$ 表示两个检测响应的外观亲密度模型(即连接概率). 设第 k 帧、 $k + 1$ 帧分别有 m 、 n 个响应,按照(1)式构造 $m \times n$ 的连接概率矩阵 P .

采用文献^[6]提出的双阈值关联策略,将相邻帧间的检测响应做保守关联,生成短小可靠的轨迹片(tracklet). 双阈值策略如下:如果 $P_{score}(r_j | r_i)$ 大于门限 θ_1 ,并且比矩阵 P 的第 i 行、第 j 列中其他元素都大于门限 θ_2 ,则认为检测响应 r_i 与 r_j 是相互关联的,即属于同一目标. 公式化描述为:

$$\begin{cases} P_{score}(r_j | r_i) > \theta_1 \text{ and } \forall r_k \in R - \{r_i, r_j\} \\ \min [P_{score}(r_j | r_i) - P_{score}(r_j - r_k), \\ P_{score}(r_j | r_i) - P_{score}(r_k - r_j)] > \theta_2 \end{cases}, \quad (2)$$

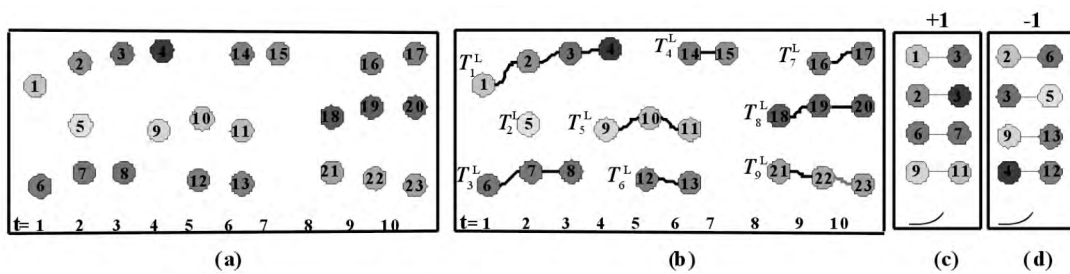


图1 训练样本采集示意图

Fig. 1 The diagram of obtaining training samples

2 目标外观描述及特征表达

2.1 目标外观描述

为了构建鲁棒的外观模型,采用颜色和纹理特征,提取标准的颜色直方图、HOG特征、协方差矩阵,分别描述目标外观的颜色、形状、纹理信息.

其中 θ_1 、 θ_2 为关联阈值. 实验证明,通过合理选择阈值(本文实验中双阈值分别选为 $\theta_1 = 0.82$ 、 $\theta_2 = 0.18$),可得到可靠的短小轨迹片集合 $\mathcal{T}^L = \{T_k^L\}$. 其中 k 代表小轨迹片编号,上标 L (Low level) 代表保守初级关联.

1.2 训练样本提取策略

根据目标轨迹在时空域分布特性,从轨迹片集合 $\mathcal{T}^L = \{T_k^L\}$ 中提取训练样本. 目标轨迹的时空域约束条件如下:(1)同一条轨迹片的响应均属于同一目标;(2)时间重叠的两条轨迹片,其响应属于不同的目标(同一目标不可能同时属于两条运动轨迹). 在样本采集前,本文先计算每个响应的可见率^[9],可见率越大则表明被遮挡的可能性越小;采集训练样本时优先考虑可见率较大的响应,以保证样本信息的完整性.

训练样本由一对响应组成,来自同一条轨迹片的两个响应构成正样本,来自不同目标轨迹片的两个响应构成负样本. 图1给出了训练样本采集示意图:(a)一个时间滑动窗口中的所有检测响应;(b)采用双阈值策略得到的短小轨迹片;(c)正样本生成;(d)负样本生成. 例如,图1(b) T_1^L 中所有的检测响应都为同一目标,可以选择 r_1 和 r_2 作为正样本; T_1^L 和 T_3^L 在时间上重叠,从中各选一个检测响应可作为负样本,如 r_2 和 r_6 .

(1) 颜色直方图. 采用 RGB 颜色空间,将每个通道设置为 8Bins,将三个通道直方图串联起来,组成一个 24 维颜色直方图向量 f_{RGB} .

(2) HOG 特征. 根据文献 [10],把目标区域划分为 4 个单元,每单元分 8 个方向统计,得到 32 维 HOG 特征向量 f_{HOG} .

(3) 纹理特征. 采用基于协方差的纹理特征描

述^[11], 该方法对于图像纹理和目标分类有很好的性能. 描述符 C 与协方差矩阵对应:

$$C = \frac{1}{n-1} \sum_{k=1}^n (z_k - \mu)(z_k - \mu)^T, \quad (3)$$

$$z_k = \left[\frac{\partial I}{\partial x} \quad \frac{\partial I}{\partial y} \quad \frac{\partial^2 I}{\partial x^2} \quad \frac{\partial^2 I}{\partial y^2} \quad \frac{\partial^2 I}{\partial xy} \right]^T, \quad (4)$$

其中 μ 、 n 分别表示检测响应区域灰度均值、像素个数 z_k 表示第 k 个像素的一阶导和二阶导.

综合以上三种特征, 单个检测响应 r_i (即目标) 的外观描述子表示为 $\{f_{RGB_i}, f_{HOG_i}, C_i\}$;

对于轨迹片 T_k^l , 选取可见率最大的前 5 个检测响应, 将各自的三种特征串联, 得到轨迹片外观描述子如下:

$$A_k = \{ \{f_{RGB_k}^l\} \quad \{f_{HOG_k}^l\} \quad \{C_k^l\} \}, \quad (5)$$

其中上标 l 表示轨迹片 T_k^l 的第 l 个检测响应索引 $l = 1, 2, \dots, 5$.

2.2 特征表达

首先考察检测响应 r_i, r_j 间的相似性向量. 对于颜色直方图和 HOG 特征, 采用欧式距离进行, 表示为 $\rho(f_{RGB_i}, f_{RGB_j}), \rho(f_{HOG_i}, f_{HOG_j})$. 对于纹理特征, 采用文献 [11] 提出的相似性距离:

$$\sigma(C_i, C_j) = \sqrt{\sum_{k=1}^5 \ln^2 \lambda_k(C_i, C_j)}, \quad (6)$$

其中 $\lambda_k(C_i, C_j)$ 是广义特征值.

检测响应 r_i, r_j 的相似性向量表示为:

$$h(r_i, r_j) = [\rho(f_{RGB_i}, f_{RGB_j}), \rho(f_{HOG_i}, f_{HOG_j}), \sigma(C_i, C_j)]. \quad (7)$$

相应地, 根据 (6) 式, 定义轨迹片 T_i^l 和 T_j^l 的特征向量如下:

$$h(A_i, A_j) = [\rho(f_{RGB_i}^1, f_{RGB_j}^1), \dots, \rho(f_{RGB_i}^5, f_{RGB_j}^5), \rho(f_{HOG_i}^1, f_{HOG_j}^1), \dots, \rho(f_{HOG_i}^5, f_{HOG_j}^5), \sigma(C_i^1, C_j^1), \dots, \sigma(C_i^5, C_j^5)]. \quad (8)$$

为方便描述, 以下将 $h(r_i, r_j)$ 记为 $h(x_i)$, 其中 x_i 代表样本 (即一对检测响应). 样本特征向量构成的集合记为 $F = \{h(x_i)\}$.

3 在线学习

本文采用 Adaboost 在线学习算法训练一个强分类器 (即判别性外观模型), 利用该分类器计算任意两个目标轨迹片的外观相似性 $H(A_i, A_j)$. $H(A_i, A_j)$ 越大, 表示轨迹片 T_i^l 和 T_j^l 属于同一目标的可能性就越大.

Adaboost 算法进行 T 轮迭代. 在第 t 轮迭代中,

输入为样本特征向量集 $F = \{h(x_i)\}$, 计算损失函数 $\sum_{i=1}^N w_i^t \exp(-\alpha_k y_i h_i(x_i))$, 并使之最小化, 得到该轮迭代对应的弱分类器 $h_i(\cdot)$ 、加权系数 α_i , 再对各样本权重做更新 (加大错分样本的权重). 通过 T 轮迭代最终得到强分类器如下:

$$H(A_i, A_j) = \sum_{i=1}^T \alpha_i h_i(A_i, A_j). \quad (9)$$

基于 Adaboost 算法的外观模型学习过程见算法 1. 其中 (x_i, y_i) 为样本集第 i 个样本 $y_i \in \{+1, -1\}$, $+1$ 代表来自同一目标检测响应对组成的正样本, -1 代表来自不同目标检测响应对组成的负样本; N 为特征库 F 中特征的数量, T 为选取的弱分类器个数, 实验中 T 取 40.

算法 1: 外观模型学习

1) Input: 正样本集: $S^+ = (x_i, y_i = 1)$, 负样本集: $S^- = (x_i, y_i = -1)$;

2) 初始化样本权重: $w_i = \frac{1}{2 |S^+|}$ if $(x_i \in S^+)$; $w_i = \frac{1}{2 |S^-|}$ if $(x_i \in S^-)$;

3) For $t = 1$ to T do:
For $n = 1$ to N do:

$$r = \sum_i w_i y_i h_n(x_i);$$

$$\alpha_n = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right);$$

End for

$$\text{Select } n^* = \arg \min_n \sum_i w_i \exp[-\alpha_n y_i h_n(x_i)]$$

Set $\alpha_t = \alpha_{n^*}; h_t = h_{n^*}$.

样本权重更新: $w_i = w_i \cdot \exp[-\alpha_i y_i h_i(x_i)]$ 归一化 w_i

End for

4) 输出强分类器: $H(x) = \sum_{i=1}^T \alpha_i h_i(x)$.

此外, 考虑到目标外观的变化, 在每个时间滑动窗内重新获取新的样本, 在线更新强分类器 H (即外观判别性模型).

综合以上三节内容, 图 2 给出了一个时间滑动窗内的多目标跟踪算法框图.

1) 输入视频序列, 检测器输出检测响应 $R = \{r_i\}$, 经初级关联得到可靠的短小轨迹片集合 $\mathcal{S}^L = \{T_k^l\}$;

2) 利用时空域约束条件, 从 \mathcal{S}^L 集中生成训练样本, 提取样本和轨迹片特征;

- 3) 训练样本送至 Adaboost 分类器 在线学习强分类器 H 得到判别性外观模型;
- 4) 强分类器 H 计算两两轨迹片之间的相似性 $H(A_i, A_j)$ 利用 Hungarian 算法实现轨迹片关联; 多

次重复该步骤 得到该时间滑动窗内的目标轨迹. 通过移动时间滑动窗, 重复上述步骤, 最终得到视频序列中多目标的完整轨迹 $\mathcal{T} = \{T_k\}$.

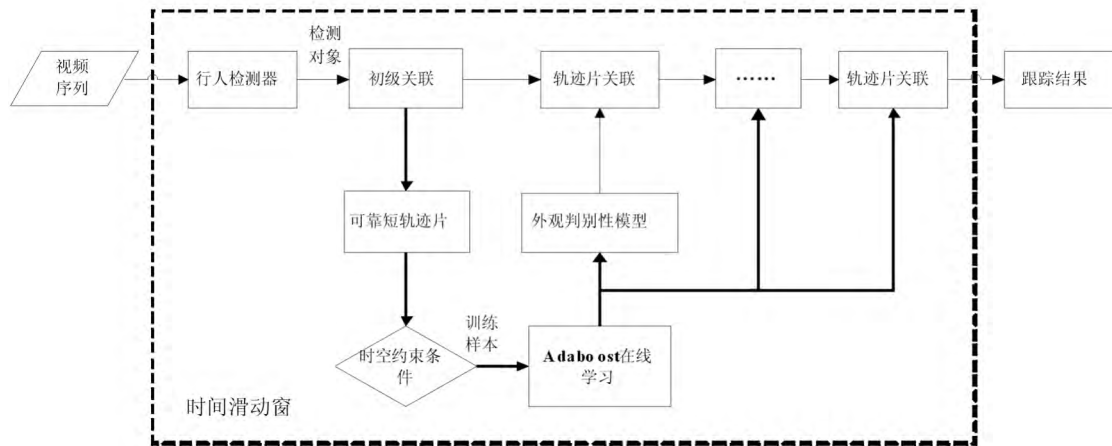


图2 时间滑动窗内多目标跟踪算法系统框架

Fig.2 The system framework of multi-target tracking algorithm in time sliding window

4 实验结果与分析

实验采用 TUD 数据库(包含 TUD-Stadtmitte、TUD-Crossing 两个视频序列)和 PET2009 数据库中的 S2L1 视频序列,跟踪目标为行人.

采用文献 [3, 12] 定义的标准来评估本文跟踪算法的性能,包含多个评估指标,如表 1 所示。“↑”表示其值越大、对应跟踪效果越好,“↓”表示其值越小、对应跟踪效果越好)。

为了保证算法评价的公平性,检测结果(即跟踪算法的输入)均来自文献 [14, 15],其中 TUD 数据库采用 DPM 检测器,PET2009 数据库采用 HOG + SVM 检测器.参与比较的算法包括文献 [13, 14] 中的两种方法.

表1 多目标跟踪性能指标

Tab.1 Multi-target tracking performance index

性能指标	含义
Recall(↑)	正确跟踪的检测数目/总的真实检测数目
Precision(↑)	正确跟踪的检测数目/总的跟踪结果检测数目
FAF(↓)	每帧中平均的检测虚警数
GT	视频序列 ground truth 包含的目标总数
MT(↑)	有超过 80% 被正确跟踪的轨迹数
ML(↓)	有低于 20% 被正确跟踪的轨迹数
PT	部分正确跟踪的轨迹数,即 $1 - MT - ML$
Frag(↓)	目标 ground truth 对应轨迹在跟踪中断开的次数
IDS(↓)	跟踪中,目标 id 的改变次数

表 2、表 3 分别给出了在 TUD 数据库、PET2009 数据库上的实验结果.实验中的 3 个视频序列均具有背景复杂、目标间频繁发生遮挡的特点.由表 2、表 3 可知,本文算法在多个性能指标上都优于文献 [13, 14] 算法,例如在 FAF、Frag、IDS 指标上,本文算法的优势明显.

表2 在 TUD 数据库上的实验结果比较

Tab.2 Comparison results on TUD dataset

视频序列	跟踪方法	Recall%(↑)	Precision%(↑)	FAF(↓)	GT	MT(↑)	PT	ML(↓)	Frag(↓)	IDS(↓)
TUD-Stadtmitte	文献 [13]	61.8	83.6	0.78	10	3	7	0	26	30
	文献 [14]	68.3	84.4	0.50	10	1	8	1	8	12
	本文算法	99.4	80.7	0.20	10	10	0	0	5	0
TUD-Crossing	文献 [13]	36.8	51.6	1.73	28	2	15	11	67	23
	文献 [14]	34.5	65.0	0.93	31	2	9	20	43	33
	本文算法	98.6	95.6	0.04	13	9	1	0	6	1

表 3 在 PET2009 数据库上的实验结果比较
Tab. 3 Comparison results on PET2009 dataset

跟踪方法	Recall% (↑)	Precision% (↑)	FAF (↓)	GT	MT (↑)	PT	ML (↓)	Frag (↓)	IDS (↓)
文献[13]	95.9	78.7	#	22	22	0	0	8	10
文献[14]	81.4	76.1	#	19	19	0	0	21	15
本文	84.5	92.4	#	19	16	2	1	8	7

图 3 给出了本文算法在 TUD-Stadtmitte 上的跟踪效果图. 其中第 7024 帧中目标 4、5(图中向上箭头)并肩行走, 在第 7051 帧目标 6、7(图中向下箭头)相反方向朝目标 4、5 走来, 被目标 4、5 严重遮

挡, 本文算法可处理此遮挡问题, 实现对目标的正确跟踪; 另外 4 号目标从第 7024 帧到 7068 帧行走过程中, 背景变化较大, 本文算法也可以很好地克服.



图 3 TUD-Stadtmitte 跟踪的结果
Fig. 3 Tracking results on TUD-Stadtmitte

图 4 是本文算法在 TUD-Crossing 上的跟踪效果图. 该视频序列中由车辆运动导致背景干扰严重. 第 38 帧目标 2、5、6、9 至左向右行走(图中向上箭头), 其中目标 5、9 并肩行走, 被从右往左迎面行走的目

标 8(图中向下箭头)依次遮挡, 到第 85 帧遮挡结束; 目标身份没有发生混淆(见第 85 帧 5、6、9、8), 始终保持对目标的稳定跟踪.



图 4 TUD-Crossing 跟踪的结果
Fig. 4 Tracking results on TUD-Crossing

图 5 是本文算法在 PET2009 中 S2L1 视频序列上的跟踪效果图. 摄像机由上向下拍摄, 视频场景中目标之间身体形态和外观相似、相互遮挡, 且场景中心存在路灯杆对行人的遮挡发生. 在第 104 帧中目标 1、7 和 10(图中向上箭头)之间发生严重遮挡, 目标 4 和 5 并肩行走(图中向上箭头); 到 137 帧时目标 1 和 7, 4 和 5 均被准确跟踪; 在 528 帧到 566 帧中, 目标 28 和 29 并肩行走、相互发生遮挡(图中向下箭头), 本文算法始终保持对其准确稳定的跟踪. 图 5 表明本文算法对位置相近、外观相似的目标具有较好的鉴别能力.

5 结语

本文设计并实现了一种基于在线学习判别性外观模型的多目标跟踪算法, 该算法以时间滑动窗为基本关联单位, 首先通过初级关联生成短小可靠的轨迹片; 利用目标轨迹时空域约束条件从中提取训练样本及特征; 采用 Adaboost 算法在线生成目标外观的判别性模型, 计算轨迹片之间的外观相似度; 最后采用匈牙利算法、经过多次迭代得到每个目标的完整轨迹. 实验结果表明, 针对复杂场景中的背景干扰、目标遮挡、位置相近、外观相似的易混淆目标等,

该算法均具有较好的处理能力.然而,本文算法未考虑长时间多目标之间的遮挡,后期可以引入多个目

标间的遮挡处理机制,进一步提高跟踪性能.



图5 PET2009中S2L1视频序列跟踪的结果

Fig.5 Tracking results of S2L1 video sequence in PET2009.

参 考 文 献

- [1] Yang B, Nevatia R. Multi-target tracking by online learning a CRF model of appearance and motion patterns [J]. *International Journal of Computer Vision*, 2014, 2(107): 203-217.
- [2] Yang B, Nevatia R. Online learned discriminative part-based appearance models for multi-human tracking [C]// *ECCV. Proceedings of European Conference on Computer Vision*. Firenze: *ECCV*, 2012: 1-14.
- [3] Yang B, Nevatia R. Multi-target tracking by online learning of non-linear motion patterns and robust appearance models [C]// *CVPR. Proceedings of Computer Vision and Pattern Recognition*. Washington D C: *CVPR*, 2012: 1918-1925.
- [4] Yang B, Huang C, Nevatia R. Learning affinities and dependencies for multi-target tracking using a CRF model [C]// *CVPR. Proceedings of Computer Vision and Pattern Recognition*. Piscataway: *CVPR*, 2011: 1233-1240.
- [5] 黄奇,项俊,侯建华等.联合特征融合和判别性外观模型的多目标跟踪[J].*中国图象图形学报*, 2015, 20(9): 1188-1198.
- [6] Huang C, Wu B, Nevatia R. Robust object tracking by hierarchical association of detection responses [C]// *ECCV. Proceedings of European Conference on Computer Vision*. Marseille: *ECCV*, 2008: 788-801.
- [7] Schapire R E, Singer Y. Improved Boosting Algorithms Using Confidence-rated Predictions [J]. *Machine Learning*, 1999, 37(3): 297-336.
- [8] Kuhn H W. The Hungarian method for the assignment problem [J]. *Naval Research Logistics Quarterly*, 1955, 2(1): 83-97.
- [9] Anton M, Stefan R, Konrad S. Stefan R, Konrad Schindler. Continuous Energy Minimization for Multitarget Tracking [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(1): 58-72.
- [10] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]// *CVPR. Proceedings of Computer Vision and Pattern Recognition*. San Diego: *CVPR*, 2005: 886-893.
- [11] Tuzel O, Porikli F, Meer P. Region covariance: A fast descriptor for detection and classification [C]// *ECCV. Proceedings of European Conference on Computer Vision*. Graz: *ECCV*, 2006: 589-600.
- [12] Li Y, Huang C, Nevatia R. Learning to associate: HybridBoosted multi-target tracker for crowded scene [C]// *CVPR. Proceedings of Computer Vision and Pattern Recognition*. Piscataway: *CVPR*, 2009: 2953-2960.
- [13] Andriyenko A, Schindler K, Roth S. Discrete-continuous optimization for multi-target tracking [C]// *CVPR. Proceedings of Computer Vision and Pattern Recognition*. Washington D C: *CVPR*, 2012: 1926-1933.
- [14] Andriyenko A, Schindler K. Multi-target tracking by continuous energy minimization [C]// *CVPR. Proceedings of Computer Vision and Pattern Recognition*. Piscataway: *CVPR*, 2011: 1265-1272.
- [15] Yang B, Nevatia R. An online learned CRF model for multi-target tracking [C]// *CVPR. Proceedings of Computer Vision and Pattern Recognition*. Washington D C: *CVPR*, 2012: 2034-2041.