

基于贝叶斯模型的相机间人群目标识别

邓颖娜 朱虹 刘薇

(西安理工大学自动化与信息工程学院, 西安 710048)

摘要 准确获取相互遮挡粘连目标的位置特征,是在视野有重叠区域条件下进行相机间目标识别的关键。提出首先构造人体模型,利用贝叶斯模型将粘连目标的分割问题转换为求解最大后验概率问题,然后依据获得的目标轴线特征,在不同的相机间按照最小距离原则进行相同目标的匹配识别。结果表明,利用人体模型进行人群分割的抗干扰能力强,目标识别的准确率较高。

关键词 贝叶斯模型 目标识别 人群分割 人体模型

中图法分类号: TP394.1 文献标识码: A 文章编号: 1006-8961(2009)09-1750-06

Bayesian Human Recognition Across Multiple Cameras in Crowded Situations

DENG Ying-na, ZHU Hong, LIU Wei

(College of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048)

Abstract Getting exact human position under occlusion is a key problem to object recognition across multiple cameras with overlapped views. The problem of human segmentation was converted to maximize the posteriori estimation by constructing a human model and a Bayesian model. And then the same objects were matched in different views on the least distance principal by taking the human axis as a feature. Experiments show promising results on human segmentation and recognition in crowded situations and the accuracy rate is high.

Keywords Bayesian model, object recognition, crowd segmentation, human model

1 引言

对不同相机间的目标进行准确识别,是实现多相机联合工作的关键问题。遮挡粘连目标的识别,是其中的难点问题。依据相机间视野关系可以分为有重叠区域和无重叠区域两种类型,本文主要针对有重叠区域的情况进行目标识别。

不同视野内的相同目标匹配识别的方法一般分为基于区域和基于点特征两大类。基于区域特征的方法,一般选取目标的颜色等外观特征进行目标识别^[1]。由于颜色受外界光照条件以及相机的影响较大,通常联合目标的外观和位置等时空信息综合

进行目标识别^[2-5]。当目标之间发生遮挡时,目标的外观信息会发生变化,或者无法准确提取到目标的外观信息。Zhou 等人^[4]通过路径分析,对遮挡目标进行处理,该方法对目标之间发生简单短时间的遮挡有效,当目标之间发生前后遮挡时则无能为力。Nam 等人^[5]在目标发生遮挡或粘连时,则将相互遮挡的目标视为同一个目标,导致目标跟踪的精度不高。

基于点特征进行目标识别的基本思想为,选取代表目标位置的点作为特征点,通过比较特征点的位置关系进行目标识别。Guler 和 Everts 等人^[6-7]通过相机标定,获取目标的3维世界坐标位置信息作为特征。相机标定技术比较复杂,而且在很多实际

收稿日期:2008-10-05; 改回日期:2009-06-29

第一作者简介:邓颖娜(1980~),女,西安理工大学控制理论与控制工程专业在读博士研究生。主要研究方向为多相机联合视频监控。E-mail: dengyingna@126.com

场景无法实现。Khan 和 Calderara 等人^[8-9]则在目标所处位置位于同一平面的假设下,建立两个相机地面之间的单应变换关系,并依此获得不同视野内目标在像平面内的位置关系。上述方法一般都选取目标的脚底点作为特征点进行目标识别,当目标被遮挡或目标分割不完整时,都有可能无法准确找到目标的脚底点。针对该问题,Hu 等人^[10]提出一种利用目标轴线交点确定目标脚底点的方法。其将目标分为单个目标和粘连的多个目标,对粘连目标,该文通过目标区域的垂直投影实现对粘连目标的分割。这对于目标之间发生简单遮挡时是有效的,但是当目标之间发生严重遮挡,如前后遮挡时,目标投影的峰谷性不明显,无法实现对目标的准确分割。

针对上述问题,本文首先通过头肩模型与目标边缘信息的匹配获取目标的头顶候选点,然后构造人体模型,利用贝叶斯模型将分割问题转换为求解最大后验概率问题,寻找模型的最佳组合实现对目标分割,并获取目标轴线。最后将轴线作为特征进行相机间相同目标的匹配识别^[10]。利用贝叶斯模型对目标进行分割,对目标前后遮挡时仍可实现对目标的检测,而且无需对目标进行区分即可获得每个目标的轴线。实验结果表明,该方法对人群目标具有较强的适应性,目标识别的准确率较高。

2 获取目标头顶候选点

由于目标之间可能存在前后遮挡,目标头部在目标轮廓和区域投影中不能体现,所以利用目标的边缘信息进行人体头部检测。如图1(b)所示,首先通过背景差分法获得运动目标区域,然后在目标区域内进行头顶匹配。

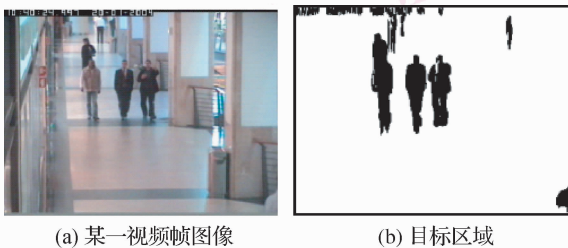


图1 相互遮挡时的目标区域
Fig. 1 Object region under occlusion

人体在运动过程中,其头肩的位置关系相对固定,所以利用标准头肩模型与目标边缘信息匹配的方法获得头顶候选点。首先对头肩模型设置检测

线,检测线的方向为模型轮廓的法线方向,如图2(b)所示,然后沿检测线对模型与目标边缘信息进行匹配。

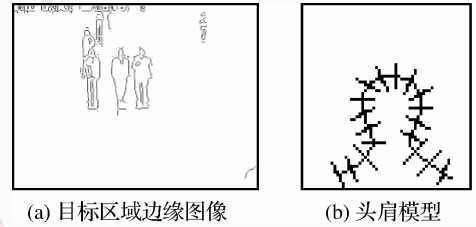


图2 目标边缘图和头肩模型
Fig. 2 Object edge and head shoulder model

对目标区域内某一边缘点 γ ,其与头肩模型的相似度为

$$P(\gamma) = \sum_{\nu} G(\nu, \nu_0, \delta) \mathbf{n}_0(\nu) \cdot \mathbf{n}_m(\nu) / N \quad (1)$$

其中, ν 为目标边缘图像中沿模型检测线方向距离模型轮廓最近的点, ν_0 为模型检测线与模型轮廓的交点, $G(\nu, \nu_0, \delta)$ 为高斯模型, $\mathbf{n}_0(\nu)$ 为目标边缘图像在点 ν 的法线方向, $\mathbf{n}_m(\nu)$ 为点 ν 所在检测线的方向, N 为模型中检测线的个数。

获得图像中每个边缘点与头肩模型的相似度后,通过设定阈值 T 获得头顶候选点,如图3(a)所示为 $T=0.45$ 时获得的头顶候选点,其中图3(b)为关于头顶候选点的局部图像。图3(b)中确定出的候选点共计有131个,而实际的目标只有4个,因此,在后续的处理中需要剔除其中126个误检测点,确定真实头顶点的位置。

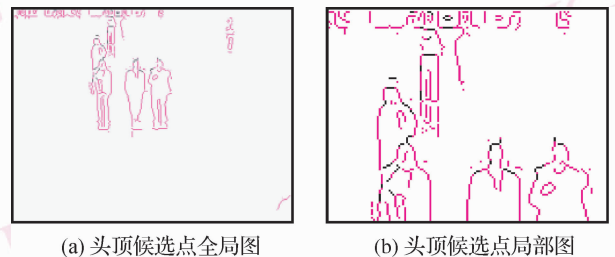


图3 头顶候选点
Fig. 3 Head candidates

3 利用贝叶斯模型获取目标

获得目标头顶候选点后,要实现对粘连目标分割,获得每个目标的位置,首先需要去除候选点中的非头顶点。通过设计人体模型,将粘连目标的分割

问题转换为求解最大后验概率的问题。假设检测到的目标整体区域为 I , 通过估计 θ^* 使其后验概率最大, 即

$$\theta^* = \arg \max_{\theta \in \Theta} P(\theta | I) \quad (2)$$

其中, $\theta = \{n, \{M_1, M_2, \dots, M_n\}\}$, n 表示目标个数, M 表示每个目标模型的相应参数。

依据贝叶斯理论, 可知:

$$P(\theta | I) \propto P(I | \theta) P(\theta) \quad (3)$$

假设目标的先验概率都相等, 这样, 只需要求解使 $P(I | \theta)$ 最大的 θ^* 即可。

3.1 目标姿态模型

利用椭圆分别代表头、躯干和腿来粗略地模拟人体的姿态, 如图 4 所示。模型中各部分比例依据人体实际比例粗略确定, 头、躯干和腿的长轴比例分别为 1:4:4, 短轴比例为 1:2.5:1.5。若已知目标的身高, 即可确定目标身体各部位的尺寸。对于每一个目标模型, 其参数由目标的头顶位置和身高组成, 其表达式为 $M = \{h, height\}$ 。

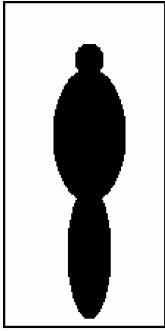


图 4 人体模型

Fig. 4 Human model

3.2 概率模型

目标的后验概率 $P(I | \theta)$ 与目标的边缘信息和所包含的区域有关系, 即

$$P(I | \theta) = P(I^{Edge} | \theta) P(I^{Region} | \theta) \quad (4)$$

3.2.1 边缘概率

目标关于边缘信息的概率, 可表示为

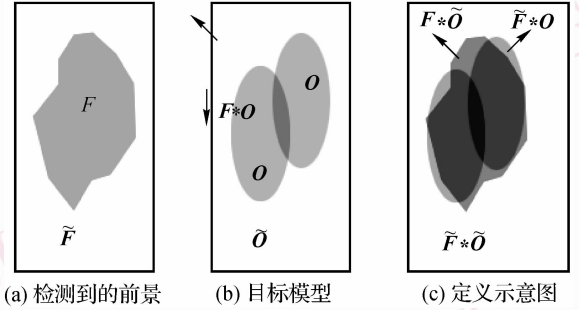
$$P(I^{Edge} | \theta) = \prod_{i=1}^n P(h_i) \quad (5)$$

其中, $P(h_i)$ 表示每个头顶候选点所对应头部区域与所建立的头肩模型的相似度, n 为目标个数。

3.2.2 区域概率

检测到的前景区域, 可能是由多个目标之间相互遮挡而形成, 所以需要建立多个目标模型与前景图像的联合相似度。对于检测到的前景区域, 如

图 5(a) 所示, 将其标记为 F , 对于给定解 θ , 可以得到其模型区域, 将其标记为 O , 如图 5(b) 所示, 模型与前景区域匹配示意图如图 5(c) 所示。



(a) 检测到的前景

(b) 目标模型

(c) 定义示意图

图 5 区域定义示意图

Fig. 5 Region definition

为了建立前景区域与模型区域的相似度, 分别对区域做如下定义: p_{11} 表示像素位于前景也位于目标模型内 ($F * O$) 的概率, N_{11} 表示该区域内对应的像素个数; p_{10} 表示像素位于前景但不位于目标模型内 ($F * \tilde{O}$) 的概率, N_{10} 表示该区域内对应的像素个数; p_{01} 表示像素不位于前景但位于目标模型内 ($\tilde{F} * O$) 的概率, N_{01} 表示该区域内对应的像素个数; p_{00} 表示像素不位于前景也不位于目标内 ($\tilde{F} * \tilde{O}$) 的概率, N_{00} 表示该区域内对应的像素个数。其中 $p_{11} + p_{10} = 1, p_{01} + p_{00} = 1$ 。由此建立模型与前景的相似度:

$$\begin{aligned} P(I^{Region} | \theta) &= \prod_{i \in I} P(i | \theta) \\ &= \prod_{i \in F * O} P(i | \theta) \prod_{i \in F * \tilde{O}} P(i | \theta) \times \\ &\quad \prod_{i \in \tilde{F} * O} P(i | \theta) \prod_{i \in \tilde{F} * \tilde{O}} P(i | \theta) \\ &= p_{11}^{N_{11}} p_{10}^{N_{10}} p_{01}^{N_{01}} p_{00}^{N_{00}} \\ &= (1 - p_{10})^{I^{F1} - N_{10}} p_{10}^{N_{10}} p_{01}^{N_{01}} (1 - p_{01})^{I^{\tilde{F}1} - N_{01}} \\ &= \alpha e^{-(\lambda_{10} N_{10} + \lambda_{01} N_{01})} \end{aligned} \quad (6)$$

其中, α 是与 θ 无关的常数, $\lambda_{10}, \lambda_{01}$ 是与 p_{10}, p_{01} 相关的系数。

3.3 求解概率模型

3.3.1 估计模型身高

由于目标距离相机远近的不同, 其在图像中所呈现的高度也有所不同。为估计目标身高, 假设场景中所有目标的脚底与头顶分别处于同一平面内, 则对于某一目标头顶点 h , 其脚底点 f 的位置可以通过下式估计得到:

$$f = H \cdot h \quad (7)$$

其中, H 为从目标头顶所在平面到脚底所在平面的

单应变换矩阵^[11],可以通过最小二乘法估计得到。

3.3.2 求解 θ^*

设可以获得 N_s 个头顶候选点,实际场景中的目标个数 $n \leq N_s$,采用迭代循环的方法寻找使后验概率 $P(I|\theta)$ 最大的解 θ^* ,具体步骤如下:

(1) 初始化,令

$$\theta^* = \{N_s, \{M_1, M_2, \dots, M_{N_s}\}\}$$

$$M_i = \{h_i, height_i\}$$

此时目标的联合概率为 $P(I|\theta^*)$ 。

(2) 遍历 N_s 个头顶候选点,依次去掉其中的一个,也就是在每次迭代中,去掉当前目标 $t(t=1, \dots, N_s)$,令

$$\theta_{cur} = \theta^* - \theta_t$$

$$\theta_t = \{1, \{M_t\}\}$$

重新计算当前的联合概率 $P(I|\theta_{cur})$,如果 $P(I|\theta_{cur}) > P(I|\theta^*)$,则令 $\theta^* = \theta_{cur}$,否则,保持 θ^* 不变, $t = t + 1$,返回步骤 2。

获得使后验概率最大的解 θ^* ,也就获得了每个目标的头顶点。依据前面建立脚底平面与头顶平面的单应关系,可以估计出目标的脚底点,从而获得每个目标的轴线。图 6 所示为对图 3 进行处理后得到的检测目标。

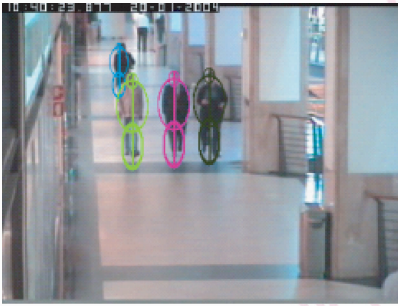


图 6 目标检测结果

Fig. 6 Detected objects

4 相机间的目标识别

4.1 相机间目标相似性度量

对相机间目标进行识别之前,首先需要要对目标的轴线特征进行相似性度量。

假设 L_s^i 表示在第 i 个视野内第 s 个目标的轴线, X_s^i 表示该目标对应的脚底点。 L_k^j 表示第 j 个视野内第 k 个目标的轴线,其脚底点表示为 X_k^j , L_s^j 表示 L_s^i 变换到第 j 个视野内得到的轴线, Q_{sk}^{ij} 表示 L_s^i

与 L_k^j 的交点。同理, Q_{sk}^{ij} 表示将 L_k^j 转换到第 i 个视野的轴线与 L_s^i 的交点。两条轴线属于同一目标的相似性与 Q_{sk}^{ij} 与 X_s^i 的距离以及 Q_{sk}^{ij} 和 X_k^j 的距离有关,如下式所示,距离越小,说明两个目标越相似。

$$R(L_s^i, L_k^j) = 2\pi\delta \exp\left(-\frac{1}{2}D_{sk}^{ij}\right)$$

$$D_{sk}^{ij} = (X_s^i - Q_{sk}^{ij})(X_s^i - Q_{sk}^{ij})^T + (X_k^j - Q_{sk}^{ij})(X_k^j - Q_{sk}^{ij})^T \quad (8)$$

4.2 相机间的目标匹配

假设在相机 i 的视野内有 M_s 个目标,分别表示为 L^1, L^2, \dots, L^{M_s} 。相机 j 的视野内有 N_s 个目标,分别表示为 L^1, L^2, \dots, L^{N_s} 。这里进行匹配的思想为使两个视野间互相匹配目标之间的距离之和最小。具体步骤如下:

(1) 构造所有目标的可能匹配对序列 ω ,计算这些匹配对所对应的距离。

(2) 对匹配对序列中的每一个匹配对 $\{m, n\}$,判断其距离是否满足 $D_{mn}^{ij} < D_T$,其中 D_T 为设定的距离阈值。如果不满足,则在 ω 中将该匹配对去掉,这样在匹配对序列中,只包含满足距离阈值条件的匹配对。

(3) 在匹配对序列 ω 中,可能包含前一场内某一目标与另一相机视野内多个目标满足阈值条件的情况。所以,存在多种可能的匹配模式。所有可能匹配的模式集合表示为 W ,在某一匹配模式中,匹配对的个数为 l ,可以表示为 $W = \{W_k = (L_{k_1}^i, L_{k_1}^j), (L_{k_2}^i, L_{k_2}^j), \dots, (L_{k_l}^i, L_{k_l}^j)\}$,其中 k 为匹配模式的索引号。

(4) 在匹配模式中,寻找使距离之和最小的模式 λ ,使得:

$$\lambda = \arg \min_k \left(\sum_{w=1}^l (D_{(k_w, k'_w)}^{i,j}) \right) \quad (9)$$

(5) 对得到的匹配对,在相应相机视野内进行标记。

5 实验结果分析

为验证本文方法的有效性,分别对 CAVIAR 数据库和 PETS2001 数据库中的视频进行实验。CAVIAR 视频为室内拍摄,场景中的运动目标为行人。PETS2001 视频为在室外拍摄,运动目标包括行人和汽车等非人体目标。其中每组视频均由两个相

机视野组成,分别标注为 C_1 和 C_2 。利用 CAVIAR 视频验证本文方法对相互遮挡的行人目标检测识别的有效性,用 PETS2001 视频验证本文方法对非人体目标抗干扰的鲁棒性。

5.1 目标检测实验结果

5.1.1 CAVIAR 视频实验结果

CAVIAR 视频的 C_1 视图中,目标之间的遮挡情况较轻,一般只有比较简单的左右遮挡,偶尔存在前后遮挡。在 C_2 视图中,场景中包含的目标个数较多,遮挡情况也比较复杂,同时在目标的前后遮挡和左右遮挡。图 7 为对 CAVIAR 视频进行目标检测与识别的效果图。

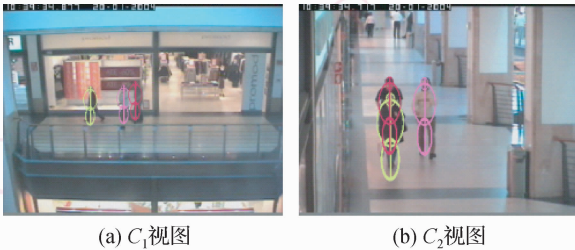


图 7 CAVIAR 视频(第 1398 帧)的目标识别效果

Fig. 7 Object recognition result of CAVIAR video(1398th)

表 1 和表 2 所示为对 C_1 和 C_2 相机视野分别进行目标检测的对比结果。从表 1 可以看出,当目标之间遮挡较轻时,本文方法与垂直投影法的差异不大,但是当场景中目标个数较多且遮挡情况复杂时,本文方法明显优于目标投影的方法,如表 2 所示。

表 1 CAVIAR C_1 视频的目标检测结果

Tab. 1 Object detection results of CAVIAR C_1 video

	垂直投影法	本文方法
有效目标个数	245	245
准确检测目标	203	214
漏检目标	42	31
误检目标	24	22
准确率(%)	82.86	87.35
漏检率(%)	17.14	12.65
误检率(%)	9.80	9.00

表 2 CAVIAR C_2 视频的目标检测结果

Tab. 2 Object detection results of CAVIAR C_2 video

	垂直投影法	本文方法
有效目标个数	506	506
准确检测目标	208	431
漏检目标	298	75
误检目标	61	52
准确率(%)	41.11	85.18
漏检率(%)	58.89	14.82
误检率(%)	12.06	10.28

5.1.2 PETS2001 视频实验结果

PETS2001 的视频监视场景中存在汽车遮挡人体目标的情况。选取其中一段视频进行实验,如图 8 所示为采用本文方法对该场景进行目标识别效果图。表 3 所示为对 PETS2001 视频进行目标检测的对比结果,可以看出,本文方法对汽车等非人体目标具有较强的抗干扰能力。

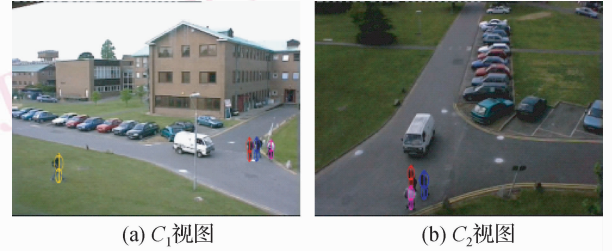


图 8 PETS2001 视频(第 861 帧)的目标识别效果

Fig. 8 Object recognition result of PETS2001 video(861th)

表 3 PETS2001 视频的目标检测结果

Tab. 3 Object detection results of PETS2001 video

	垂直投影法	本文方法
有效目标个数	420	420
准确检测目标	286	375
漏检目标	134	45
误检目标	72	65
准确率(%)	68.10	89.29
漏检率(%)	31.90	10.71
误检率(%)	17.14	15.48

5.2 目标识别实验结果

为统计本文方法对相机间人群目标识别的准确率,做如下定义:假设 C_1 视野中有 M_s 个目标, C_2 视野中有 N_s 个目标,目标识别的准确率如下式:

$$r = 1 - \frac{N_f}{(M_s + N_s)} \quad (10)$$

其中, N_f 表示两个视野间错误识别的目标个数。如表 4 所示为目标识别的准确率对比结果,可以看出,利用本文方法对人群目标进行分割后,目标识别的准确率有明显提高。

表 4 目标识别准确率

Tab. 4 Object recognition rate 单位:%

实验视频	垂直投影法	本文方法
CAVIAR	49.3	77.2
PETS2001	62.6	82.1

6 结 论

目标的位置信息对于有重叠区域的相机间目标识别,是相对比较稳定的信息。与目标位置相关的目标轴线特征对进行目标识别具有较高的鲁棒性。然而目标之间的相互遮挡粘连,给准确提取目标的位置信息带来很大干扰。针对该问题,本文采用头肩模型寻找目标的头顶候选点,然后依据贝叶斯模型,将最终确定目标的分割问题转换为求解最大后验概率问题。对 CAVIAR 和 PETS2001 中的视频分别进行了实验,视频分别包括目标之间存在简单遮挡和复杂遮挡以及非人体目标遮挡的情况。实验结果表明,本文方法对复杂人群粘连分割的适应性较强,目标识别准确率较高。

参考文献 (References)

- 1 Price M, Nicolls F, Jager G. POD: Colour Detection of People and Objects for Multi-camera Video Tracking [EB/OL]. <http://www.dip.ee.uct.ac.za/~mathew/publish/prasa2004.pdf>, 2009-05-25.
- 2 Mittal A, Davis L. Unified multi-camera detection and tracking using region-matching [A]. In: Proceedings of IEEE Workshop on Multi-Object Tracking [C], Vancouver, Canada, 2001:3-10.
- 3 Kang J, Cohen I, Medioni G. Tracking objects from multiple stationary and moving cameras [A]. In: Proceedings of the IEE Intelligent Distributed Surveillance Systems [C], London, Briantan, 2004: 31-35.
- 4 Zhou Q, Aggarwal J K. Object tracking in an outdoor environment using fusion of features and cameras [J]. Image and Vision Computing, 2006, **24**(11):1244-1255.
- 5 Nam Y, Ryu J, Choi Y, *et al.* Learning spatio-temporal topology of a multi-camera network by tracking multiple people [J]. International Journal of Signal Processing, 2007, **4**(4):254-259.
- 6 Guler S, Griffith J M, Pushee I A. Tracking and handoff between multiple perspective camera views [A]. In: Proceedings of the 32nd Applied Imagery Pattern Recognition Workshop [C], Washington, DC, USA, 2003:275-281.
- 7 Everts I, Sebe N, Jones G A. Cooperative object tracking with multiple PTZ cameras [A]. In: Proceedings of the 14th International Conference on Image Analysis and Processing [C], Modena, Italy, 2007:323-330.
- 8 Khan S, Shah M. Consistent labeling of tracked objects in multiple cameras with overlapping fields of view [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, **25**(25):1355-1360.
- 9 Calderara S, Prati A, Vezzani R, *et al.* Consistent Labeling for Multi-Camera Object Tracking [M]. Lecture Notes in Computer Science, Heidelberg: Springer, 2005, **3617**:1206-1214.
- 10 Hu W, Hu M, Zhou X. Principal axis-based correspondence between multiple cameras for people tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, **28**(4): 663-671.
- 11 Zhe L, Davis L S, Doermann D, *et al.* Hierarchical part-template matching for human detection and segmentation [A]. In: Proceedings of 11th International Conference on Computer Vision [C], Rio de Janeiro, Brazi, 2007:1-8.