

文章编号:1004-4213(2011)05-0758-6

基于支持向量机增量学习和 LPBoost 的人体目标再识别算法

许允喜^{1,2}, 蒋云良¹, 陈方^{1,3}

(1 湖州师范学院 信息与工程学院, 浙江 湖州 313000)

(2 浙江大学 信息与电子工程系, 杭州 310027)

(3 南京航空航天大学 自动化学院, 南京 210016)

摘 要:摄像机间目标关联是无重叠视域多摄像机目标持续跟踪的关键. 提出了一种只利用人体目标外观, 完全不依赖于空时关系的人体目标再识别算法, 利用识别结果直接进行跨摄像机间人体目标关联, 而不依赖于目标的捕获时间和路径限制. 对跟踪视频前景图像序列提取互补性视觉单词树直方图和全局颜色直方图二种特征, 采用支持向量机增量学习在线训练二种特征的人体外外观辨别模型, 再利用多类线性规划增强算法对二种特征的支持向量机模型进行在线自适应融合. 实验结果表明, 本文算法具有较强的在线学习能力, 能增量式表达人体目标辨别性外观模型, 特征融合后的模型区别性更强, 有效地降低多方面条件变化的影响, 获得了高识别率, 且能够实现快速实时实现, 相对于现有方法有了明显提升.

关键词:视频监控; 支持向量机增量学习; 局部描述子; 人目标再识别; 单词树; 线性规划增强

中图分类号: TP391.41

文献标识码: A

doi:10.3788/gzxb20114005.0758

0 引言

近年来由于单摄像机有限的视域无法满足广域视频监控的要求, 无重叠视域多摄像机目标跟踪成为广域视频监控领域的一个新兴的研究热点^[1-3]. 要实现目标的连续跟踪, 主要问题是确认不同位置的摄像机在不同时刻发现的目标是否是同一个目标, 即摄像机间的目标关联问题. 相对于单摄像机目标跟踪, 跨摄像机间的目标关联极具挑战性, 这是因为: 1) 不同传感器特性、光照条件、视角使同一目标的外观在不同摄像机间差异很大; 2) 摄像机间跟踪目标时空信息的可靠性低以及开阔的盲区极大地提高了跨摄像机间跟踪目标关联的难度.

目前, 绝大多数无重叠视域多摄像机视频跟踪算法都利用摄像机间的空时关系来简化跟踪目标的交接^[2-5]. 然而, 在一般情况下, 这种方法有一定的限制: 依赖于目标的捕获时间和路径限制. 最近, 一些研究人员提出了只利用人体目标外观, 完全不依赖于空时关系的人体目标再识别算法^[6-10], 这种方法

具有更广泛的应用, 不依赖于目标的捕获时间和路径限制.

用视觉单词包(Bag Of Visual Words, BoW)来表征图像内容在图像检索、图像分类等领域取得了良好的应用效果^[11-12]. 由于跨摄像机间目标外观的变化性强, 这种局部描述符能够适应粗糙的目标定位、部分遮挡以及光照变化. 文献[9]采用尺度不变特征转换(Scale-Invariant Feature Transform, SIFT)描述符的视觉单词树和 Learning++、MT 增量学习方法在线训练得到支持向量机(Support Vector Machine, SVM)辨别模型, 达到了实时性能, 在线训练 SVM 辨别模型的模式识别方法比基于相似性度量方法的识别率更高. 实验结果也表明: 视觉单词树方法比颜色直方图方法的识别率更高. 但文中基于 Learning++、MT 增量学习的 SVM 在线训练方法性能不佳, 需进一步提高.

颜色和局部描述符这二种特征具有一定的互补性, 因此, 对这二种特征进行融合建模可以有效提高识别率. 本文采用视觉单词树直方图和全局颜色直

基金项目:国家自然科学基金(No. 60872057)和浙江省自然科学基金(No. Y1101237, No. R1090244, No. Y1080212)资助

第一作者:许允喜(1978—), 男, 讲师, 博士研究生, 主要研究方向为智能视频监控、图像处理, 三维视觉计算等. Email: xuyunxi@hutc.zj.cn

收稿日期:2010-11-17; **修回日期:**2011-01-06

方图构建跟踪目标的人体外观模型. 在模型在线训练方面, 采用在增量训练过程中仅保留支持向量的 SVM 增量学习进行 SVM 在线训练, 获得了比 Learning++、MT 增量学习方法更高的训练和识别性能; 在特征融合方面, 对全局颜色直方图 SVM 模型和视觉单词树直方图 SVM 模型利用多类 LPBoost 算法进行特征融合, 获得了比单一特征更高的识别性能.

1 人体目标特征提取方法

本文使用二种互补特征建立人体目标观测模型: 一种是颜色直方图特征, 为全局特征; 另一种是 SIFT 描述符, 为局部特征. 全局颜色直方图可以直接作为 SVM 训练和识别特征. 因为一幅图像有许多 SIFT 局部描述符, 需使用视觉单词包方法形成直方图以用作 SVM 训练和识别特征. 本文采用分层结构的视觉单词树方法^[9,14], 与传统平面结构的视觉单词包方法相比具有计算负荷小、单词个数不受约束、搜索速度快等优越性.

视觉单词树方法将所有训练图像中获得的 SIFT 局部描述子进行分层 k 均值聚类量化为树形码书. 树形码书可采用一般图像库进行离线训练以满足视频跟踪的实时需求. 树形码书形成后, 每幅图像就可以采用最近邻技术寻找每个特征局部描述子对应的码书中的视觉单词表示, 据此可计算图像的特征表示: 视觉单词树直方图.

1.1 局部特征提取和 SIFT 局部描述符

使用特征提取技术从图像中获得多个图像块, 这些特征提取技术可以直接使用规则网格, 也可以使用各种兴趣点检测子, 还可采用随机取样等提取方法. 本文采用随机取样方法. SIFT 描述符^[13]广泛应用于图像检索和分类领域, 是目前公认的各项综合性能较好的外观特征描述子, 它具有尺度和仿射不变性, 适合各种复杂条件下的特征描述. SIFT 描述子的计算方法是, 首先进行高斯滤波以获得尺度不变性, 再计算关键点周围区域的梯度大小及方向, 将周围区域均分成 4×4 个子区域, 累计每个子区域的梯度方向直方图. 梯度方向分为 8 个方向, 则描述子长度为 128 维.

1.2 基于分层 k 均值聚类的视觉单词树构建

基于分层 k 均值聚类的视觉单词树^[14]定义了一种分层的量化策略, 在训练阶段提取样本图像集的特征描述向量进行无监督训练以生成视觉单词树. 不同于传统聚类, 类别数 k 被定义为分支因子, 即每个树节点的分支数. 首先, 对训练图像的特征集进行 k 均值聚类, 将样本特征划分成 k 个分支, 对每

个分支递归地执行 k 均值聚类得到下一层的 k 个更细分支, 直到约定的最大层数 L . 在识别阶段, 对新图像的每一特征向量只需自上而下逐层与 k 个候选类中心进行比较, 以选择最近邻的类. 在每一层只需执行 k 次点乘运算, 使得整个过程有 kL 次点乘的复杂度. 用分层的树型结构定义视觉单词树, 不同于传统视觉单词包的平面结构, 可以设计近似于最近邻搜索的算法更有效搜寻最佳匹配单词. 同时在平面结构的单词包中, 整个图像检索算法的计算代价随单词包维数增加而剧增, 而在分层的单词树中, 算法的计算代价只与叶子数呈对数增长, 在 L 层 k 支树中可以表示的单词个数为

$$M = \sum_{i=1}^L k^i = \frac{k^{L+1} - k}{k - 1} - 1 \quad (1)$$

不同的树节点具有不同的重要性, 需要为视觉单词树的每个树节点分配一个权重 w_i . 本文采用文本检索中典型的逆向文档频率加权法确定节点权重^[9,14]

$$w_i = \ln \frac{N}{N_i} \quad (2)$$

式中, N 是数据库包含的总图像数, N_i 是至少有一个特征路径通过节点 i 的图像数.

1.3 视觉单词树直方图

使用视觉单词树对每幅图像提取的所有 SIFT 描述符量化, 形成视觉单词树直方图. 则目标 c 在 t 时刻前景分割图像的局部描述符外观特征表达为

$$\text{App}(I_t^c) = \{x_1, x_2, \dots, x_M\} \quad (3)$$

M 为由式(1)定义的词汇树单词总数. x_i 为加权的单词直方图

$$x_i = n_i w_i \quad (4)$$

n_i 为包含视觉单词 i 的描述符数量. w_i 为式(2)定义的权重. 文献[9]对不同的 k 和 L 参量设置进行了实验比较, 结果表明 $k=10$ 和 $L=4$ 效果较好, 本文采用一样的参量设置, 则视觉单词树直方图向量的维数 $M=11\ 110$. 该直方图向量是稀疏的, 绝大多数元素都为 0, 只有上百个元素不为 0, 可以采用稀疏存储方法节省空间. 另外, 利用稀疏结构 SVM 的点积计算也可以很快完成.

1.4 颜色直方图特征

颜色特征由于具有很好的区分性, 在人体目标跟踪中被广泛应用. 本文使用 RGB 颜色空间. 每个颜色通道计算 32 维的直方图. 整个前景人体图像全局颜色直方图特征共有 $96(3 \times 32)$ 维.

2 SVM 增量学习

SVM 的学习方法是基于小样本的学习方法, 它

实现了结构风险最小化,避免了在学习过程中存在的过学习现象。

在视频跟踪中,人体前景图像序列具有动态性,不断有已有目标新的样本或新目标出现.人体外观模型必须适应这一情况,持续的更新模型.需要给已在某一学习样本集中学习过的 SVM 增添新的学习样本,以进一步提高分类准确度;同时希望 SVM 在新的学习中能够充分利用历史的学习结果,从而显著地减少后继的学习时间.本文利用 SVM 增量学习方法,在很小的时间空间代价下实现新样本的学习,达到实时视频跟踪的要求。

本文采用线性核的支持向量机,其分类函数为

$$f(x) = \text{sgn}\left\{\sum_{i=1}^n \alpha_i^* y_i(x_i x) + b^*\right\} \quad (5)$$

式中 α_i^* 不为零的样本,即为支持向量,可以由二次规划方法计算. b^* 为分类阈值.支持向量机为二值分类算法,而本文是多目标识别,所以采用一对多组合(one against all)的多类 SVM.

支持向量可完全描述整个数据样本集的分类特征,而支持向量集只是数据样本集的一小部分.分类仅由在超平面边缘上的少数支持向量决定.鉴于此,本文在增量训练过程中仅保留支持向量,采用文献[15]中基于超越边界的增量学习方法,并扩展到多类情况.设 t 时刻的分类模型为 SVM $_t^c$, $c=1, \dots, C$.当已有目标新的样本数据出现时,检验新增样本是否超越 SVM $_t^c$ 定义的边界,即 $y_i f_t^c(x_i) \leq 1$. 满足条件的样本保留,不满足的则丢弃.当满足条件的样本达到一定数目时,则新样本联合支持向量重新训练得到新模型和新的支持向量.当新目标出现且样本达到一定数目时,则进行训练获得新目标的 SVM 模型和支持向量。

本文对跟踪目标采用了简单的时间窗管理策略,即对长时间未出现的跟踪目标在人体外观模型库中清除,相应的支持向量也随之清除.采用这种退出机制,有三个方面作用:1) 限制人体外观模型库的不断增加,较少人体目标再识别时间;2) 减少模型的训练时间,避免了支持向量的不断增加;3) 只采用当前时间窗的人体目标训练样本,增强了模型的辨别能力。

3 基于线性规划增强的 SVM 多特征在线自适应融合算法

线性规划增强(Linear Programming Boosting, LPBoost)^[16]是一种监督多分类器集成算法,它使不同类训练样本之间的软间隔最大化,与SVM训练

目标相同,是性能优越的 SVM 多分类融合算法.本文中 LPBoost 算法与 SVM 增量学习结合,利用 SVM 增量学习的支持向量作为训练样本,计算复杂度低,实现多特征高效融合。

3.1 LPBoost

在 LPBoost 算法中,最后的分类函数为

$$f(x) = \sum_{j=1}^J \beta_j h_j(x) \quad (6)$$

$h_j(x)$ 为弱分类器, β_j 是弱分类器的非负权重, $f(x)$ 为多个弱分类器的线性联合. LPBoost 算法就是利用线性规划技术得到弱分类器的最佳线性联合.在 LPBoost 算法中,所有分类器的权重在每次迭代中被整体调整.早期的 Boosting 算法,例如 AdaBoost 不具有这种特性,算法收敛速度较慢。

LPBoost 的基本线性规划表达为^[16]

$$\min_{\alpha, \xi, \rho} -\rho + D \sum_{n=1}^l \xi_n \text{ s.t. } \sum_{j=1}^J y_n \alpha_j h_j(x_n) + \xi_n \geq \rho, \sum_{j=1}^J \alpha_j = 1, \xi_n \geq 0, \alpha_j \geq 0, n=1, \dots, l \quad (7)$$

ξ 为松弛因子,允许错分样本的存在; D 是一个正常量,为惩罚因子, l 为训练样本总数.式(7)中的基本线性规划问题可以转化为对偶线性规划

$$\max_{\lambda, \gamma} \gamma \text{ s.t. } \sum y_n h_j(x_n) \lambda_n + \gamma \leq 0, 0 \leq \lambda_n \leq D, \sum_{i=1}^n \lambda_n = 1, n=1, \dots, l \quad (8)$$

式(8)中的线性规划问题可以用列生成技术高效解决,详细迭代算法参考文献[16]。

3.2 SVM 多特征融合算法

多类多特征支持向量机分类函数^{[17][18]}可表达为

$$f(x) = \arg \max_{c=1, \dots, C} \sum_{j=1}^J B_j^c h_{j,c}(x) \quad (9)$$

$h_{j,c}(x)$ 为多类支持向量机分类函数, B_j^c 为权重; J 为特征种类数目,本文中 $J=2$; C 为跟踪人体目标数,由于视频跟踪的动态性, C 的大小随时间变化.由第 2 节的 SVM 增量学习可以获得视觉单词树直方图特征的 SVM 分类函数 $h_{1,c}(x)$ 和颜色直方图特征的 SVM 分类函数 $h_{2,c}(x)$. 现在通过 LPBoost 算法优化参量 B_j^c . 多类 LPBoost 的优化目标函数为

$$\min_{\alpha, \xi, \rho} -\rho + D \sum_{n=1}^l \xi_n \text{ s.t. } \sum_{j=1}^J B_j^{y_i} h_{j,y_i}(x_n) - \sum_{j=1}^J B_j^{y_j} h_{j,y_j}(x_n) + \xi_n \geq \rho, \xi_n \geq 0, n=1, \dots, l, y_j \neq y_i \sum_{j=1}^J B_j^i = 1, m=1, \dots, J, B_j^i \geq 0, \quad (10)$$

这仍然是一个线性规划问题,可由上一节的 LPBoost 算法高效解决.目标函数的数据样本为第 2 节中二个特征人体外观模型的 SVM 支持向量.由于 SVM 支持向量数目一般很少,所以上述的线性

规划计算复杂度较低.

整个系统如如图 1. 多特征融合的系数随时间变化, 是在线自适应融合, 反映了视频跟踪的动态性. 多特征融合只在 SVM 增量学习过程中支持向量发生变化时才进行系数调整, 从而节省计算资源, 且调整前的系数作为 LPBoost 算法的迭代初始值, 保证算法的快速收敛.

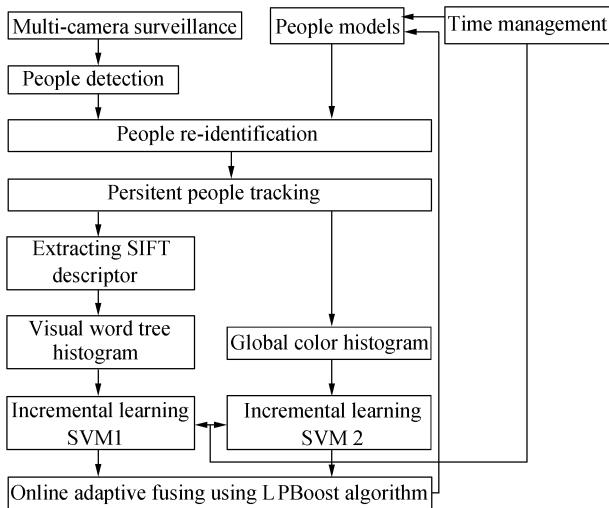


图 1 系统流程
Fig. 1 Flow chart of the system

4 实验结果与分析

4.1 数据集

本文采用 CAVIAR 数据集^[19] 评价算法的性能. 该数据集广泛用于视频跟踪和人体目标再识别算法的实验评价^[6,8-9]. 数据集由安装在购物中心走廊中不同角度的 2 个摄像机拍摄的 26 段视频监控片段组成, 如图 2. 数据集提供了基于 XML 的基准数据. 基准数据中包含视频序列中人体目标的前景分割图像: 粗糙的边界框. 本文从中提取 20 个不同人体目标的跟踪序列. 这些跟踪序列包含部分遮挡、严重遮挡、跨摄像机以及剧烈光照变化的图像, 实验使用的 20 个人体目标数据库如图 3.



图 2 在不同时间被不同摄像机捕获的同一目标
Fig. 2 Same visual object captured at different instants by different cameras



图 3 20 个人的目标数据库
Fig. 3 20 visual objects used in the dataset

视觉单词树的码书采用一般的样本图像集进行离线训练, 本文采用 PASCAL 人体目标数据库^[20].

4.2 增量学习能力

目标外观模型应该是自适应的. 在视频监控环境下, 目标外观模型随着目标被跟踪的视角和光照而变化, 所以在模型训练中能够反映这些变化. 最直接的方法为: 当出现新的样本数据时, 遗忘以前的模型, 新数据联合以前的数据一起训练得到新模型. 这种方法对于实时系统是不可行的, 因为其需要大的存储空间和计算资源, 然而, 这种方法可以提供增量学习方法的对比参考. 文献[9]采用了基于 Learning

++ 的增量学习方法. 本文与这二种方法进行了对比. Learning++ 算法中, 20 个人体目标跟踪序列数据库分为 3 个子集, 用于测试其增量学习能力. 和文献[9]一样, 所有算法都只使用局部描述符一种特征.

从表 1 的结果看, 本文的增量学习算法识别性能远远高于 Learning++ 算法, 和保留全部样本的学习算法相当. 本文采用的增量学习只保留支持向量, 有效地利用了 SVM 算法的特性, 模型辨别性强, 训练时间快. 而 Learning++ 算法只是对各批量数据的 SVM 模型进行集成学习, 模型辨别

性弱.

表 1 采用不同增量学习策略的识别性能

Table 1 Recognition performance with different incremental learning method

Learning methods	Identification rate/(%)
Retain all samples	84.1
SVM incremental learning	83.7
Learning++ . MT	66.2

4.3 LPBoost 算法的特征融合性能

如表 2, 在使用 LPBoost 方法融合颜色和 SIFT 局部描述符二种特征后, 系统的识别率得到了较大的提高. 由单一特征得到的外观模型不够鲁棒. 这二种特征有一定的互补性, 融合后的模型区别性更强, 有效地降低多方面条件变化的影响. 实验中发现, 局部描述符对形变较大的人体目标识别性能较差, 而颜色特征对光照、局部遮挡不够鲁棒.

表 2 采用不同特征的识别性能

Table 2 Recognition performance with different features

Adopted feature	Identification rate/(%)
Color	76.2
SIFT local descriptor	83.7
Feature combination	90.1

4.4 算法实时性能分析

对于每帧图像, 算法时间主要耗在 SIFT 描述符提取和视觉单词数直方图计算上. 算法采用 C++ 语言实现, 在主频为 2.8 GHz, 内存为 4G 的 PC 机器上运行. 平均每帧视频前景图像的 SIFT 描述符提取时间约为 65ms, 视觉单词树直方图的计算时间约为 36 ms, 全局颜色直方图计算时间为 2 ms, 平均每秒约处理 9.7 帧. 为了加快处理速度, 算法还可以进一步优化, 如采用并行实现等. 和文献[9]相比, 本文算法只是多了 2 ms 的全局颜色直方图计算.

SVM 模型训练和 LPBoost 特征融合的计算效率很高, 一般在 60 ms 以内, 由于模型更新和特征融合不是在每帧视频图像上进行, 一般几十帧至几百帧更新一次, 不影响整个算法的实时性, 且这部分计算可另外独立完成. 由于本文采用只保留支持向量的 SVM 增量学习, 每次训练样本数很少, 训练速度明显快于 Learning++ . MT 算法.

5 结论

本文提出了基于外观的人体目标再识别算法. 算法完全不依赖于摄像机间的空时关系, 具有在线学习能力, 能增量式表达人体目标辨别性外观模型, 并采用 LPBoost 高效地融合了局部描述符和全局颜色直方图特征. 虽然算法获得了较高的识别效果,

但仍有一些地方值得改进和进一步研究, 如本文把单摄像机的目标检测和跟踪结果直接作为识别算法的输入, 如果识别算法可以和跟踪算法集成则可以提高整个视频跟踪的性能, 另外, 融合其他高效的特征也可以进一步提高识别性能.

参考文献

- [1] KONG Qing-jie, LIU Yun-cai. Non-Overlapping multi-camera object tracking for wide area visual surveillance[C]. In Proc Chinese Conference on Pattern Recognition, Beijing, China, 2007: 154-161.
孔庆杰, 刘允才. 面向广域视频监控的无重叠视域多摄像机目标跟踪[C]. 2007 年全国模式识别会议论文集, 北京, 2007: 154-161.
- [2] JAVED O, SHAFIQUE K, RASHEED Z, *et al.* Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views[J]. *Computer Vision and Image Understanding*, 2008, **109**(2): 146-162.
- [3] MEHMOOD M O, KHAWAJA A. Multi-camera based human tracking with non-overlapping fields of view[C]. Fifth International Conference on Image and Graphics, 2009: 313-318.
- [4] MOTAMED C, WALLART O. A temporal fusion strategy for cross-camera data association [J]. *Pattern Recognition Letters*, 2007, **28**(2): 233-245.
- [5] MONTCALM T, BOUFAMA B. Object inter-camera tracking with non-overlapping views: a new dynamic approach[C]. In Conference on Computer and Robot Vision, Windsor, ON, Canad, 2010: 354-361.
- [6] DEOLIVEIRA I O, DESOUZAPIO J L. People reidentification in a camera network [C]. Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, dasc, 2009: 461-466.
- [7] GHEISSARI N, SEBASTIAN T B, HARTLEY R. Person reidentification using spatiotemporal appearance[C]. In 2006 Conference on Computer Vision and Pattern Recognition, 2006: 1528-1535.
- [8] HAMDOUN O, MOUTARDE F, STANCIULESCU B, *et al.* Re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences [C]. 2nd ACM/IEEE International Conference on Distributed Smart Cameras, 2008: 1-6.
- [9] TEIXEIRA L F, CORTE-REAL L. Video object matching across multiple independent views using local descriptors and adaptive learning[J]. *Pattern Recognition Letters*, 2009, **30**(2): 157-67.
- [10] JEONG K, JAYNES C. Object matching in disjoint cameras using a color transfer approach [J]. *Machine Vision and Applications*, 2008, **19**(5): 443-455.
- [11] SIVIC J, ZISSERMAN A. Video google: a text retrieval approach to object matching in videos[C]. ICCV 2003, Nice, France, 2003: 11-17.
- [12] QUELHAS P, MONAY F, ODOBEZ J M, *et al.* A thousand words in a scene[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(9): 1575-1589.
- [13] LOWE D G. Distinctive Image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, **2**(60): 91-110.
- [14] NISTER D, STEWENIUS H. Scalable recognition with a vocabulary tree[C]. In 2006 Conference on Computer Vision and Pattern Recognition, 2006: 2161-2168.

- [15] DOMENICONI C, GUNOPULOS D. Incremental support vector machine construction[C]. In Proc IEEE International Conference on Data Mining, San Jose, USA, 2001: 589-592.
- [16] DEMIRIZ A, BENNETT K P, TAYLOR J S. Linear programming boosting via column generation[J]. *Machine Learning*, 2002, **46**(1-3): 225-254.
- [17] GEHLER P, NOWOZIN S. On feature combination for multiclass object classification[C]. Proceedings of the Twelfth IEEE International Conference on Computer Vision, NJ, USA, 2009: 221-228.
- [18] SAFFARI A, GODEC M, POCK T, *et al.* Online multi-class LPBoost[C]. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010: 3570-3577.
- [19] EVERINGHAM M, GOOL L V, WILLIAMS C, *et al.* The PASCAL Visual Object Classes Challenge[DB/OL]. [2010-04-10]. <http://www.pascal-network.org/challenges/VOC/>
- [20] FISHER R, SANTOS-VICTOR J, CROWLEY J. CAVIAR: context aware vision using image-based active recognition [DB/OL]. [2010-04-10]. <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>.

Person Re-Identification Algorithm Based on Support Vector Machine Incremental Learning and Linear Programming Boosting

XU Yun-xi^{1,2}, JIANG Yun-liang¹, CHEN Fang^{1,3}

(1 School of Information & Engineering, Huzhou Teachers College, Huzhou, Zhejiang 31300, China)

(2 Department of information science & electronic engineering, Zhejiang University, Hangzhou 310027, China)

(3 College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: Object association between the cameras is a key of persistent object tracking in non-overlapping multi-cameras. A people re-identification algorithm was proposed only using people appearance completely independent on the space-time relations, object association across disjoint views was carried out by directly utilizing identification result, and this method does not depend on captured time of object and path restrictions. Complementary visual word tree histogram and global color histogram were extracted from the video image sequence, and Support Vector Machine(SVM) incremental learning was used to train online distinguishing people appearance models of two features. Finally, multi-class Linear Programming Boosting (LPBoost) algorithm was introduced into on-line adaptive fusion of two SVM models. The proposed method has strong online learning ability, and can incrementally represent discriminative people appearance model. The model after fusing two features is more discriminative and effectively reduces the influence of changes in various conditions. Experimental results show that the proposed method achieves high identification rate and rapid real-time implementation which are markedly improved compared to the existent methods.

Key words: Visual surveillance; Support Vector Machine(SVM) incremental learning; Local descriptor; Person re-identification; Vocabulary tree; Linear Programming Boosting(LPBoost)