

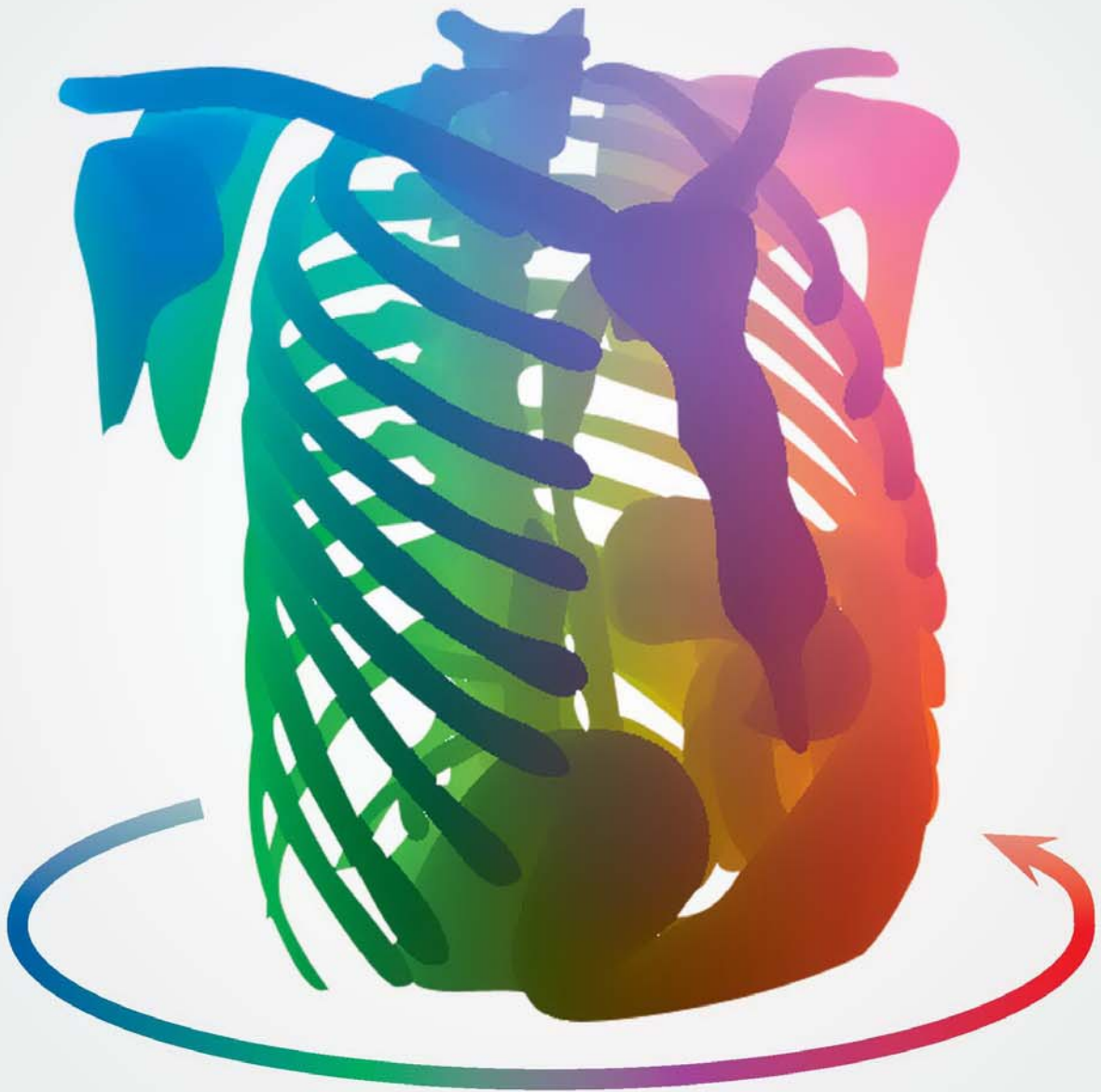


主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2015
02
VOL.20

ISSN1006-8961
CN11-3758/TB





第20卷第2期（总第226期）

2015年2月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院

主办单位 中国科学院遥感与数字地球研究所

中国图象图形学学会

北京应用物理与计算数学研究所

主 编 李小文

编辑出版《中国图象图形学报》编辑出版委员会

邮政信箱 北京9718信箱

邮 编 100101

电子信箱 jig@irsa.ac.cn

电 话 010-64807995

网 址 www.cjig.cn

广告经营许可证 京朝工商广字第0361号

总 发 行 北京报刊发行局

订 购 全国各地邮局

海外发行 中国国际图书贸易集团有限公司

（邮政信箱：北京399信箱 邮编：100048）

印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences

Sponsored by Institute of Remote Sensing and Digital Earth, CAS

China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief LI Xiaowen

Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics

P.O.Box 9718, Beijing, P.R.China

Zip code 100101

E-mail jig@radi.ac.cn

Telephone 010-64807995

Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals

Domestic All Local Post Offices in China

Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China))

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

ISSN 1006-8961

CODEN ZTTXFZ

国外发行代号 M1406

国内邮发代号 82-831

国内定价 50.00元



多尺度活动网格在云场景仿真中的应用(第0202页)



自然环境视频中基于显著鲁棒轨迹的行为识别(第0245页)



虚拟手抓持力觉生成算法真实性的评价(第0280页)

图像处理和编码

引入神经网络中间神经元的快速小波图像压缩

张海涛, 张永霖 0159

拟正态分布扩散的图像平滑

周先春, 汪美玲, 周林锋 0169

可选特征的快速分形图像编码

袁宗文, 鲁业频, 杨汉生 0177

图像分析和识别

融合多尺度码本的全局编码图像分类

董振宇, 赵杰煜, 祝军 0183

计算机图形学

结合K-D树和Shell的快速动态等值面光线跟踪法

罗月童, 石放放, 张伟, 朱会国 0193

多尺度活动网格在云场景仿真中的应用

范晓磊, 张立民, 张建廷, 徐涛 0202

第十届和谐人机环境联合会议专栏

低资源语言的无监督语音关键词检测技术综述

杨鹏, 谢磊, 张艳宁 0211

快速离散Curvelet变换域的图像融合

杨勇, 董松, 黄淑英 0219

高效视频编码中变换跳过模式的快速选择

王宁, 张永飞, 樊锐 0229

主动学习的多标签图像在线分类

徐美香, 孙福明, 李豪杰 0237

自然环境视频中基于显著鲁棒轨迹的行为识别

易云, 王瀚漓 0245

特征变换和数据集分块的行人比对

沈洋, 林巍晓, 殷惠清 0254

移动手持环境下触摸式目标选择技术对比

辛义忠, 李岩, 袁伟强, 郑谦 0264

虚拟环境的用户意图捕获

程成, 赵东坡, 卢保安 0271

虚拟手抓持力觉生成算法真实性的评价

杨文珍, 许艳, 颜传武, 吴新丽, 张昊, 孟闯 0280

交互式乐器演奏的六自由度力觉渲染方法

童号, 史有姣, 王党校, 张玉茹 0288

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Multi-scale moving grids and its application in clouds scene simulation(P0202)



Action recognition from unconstrained videos via salient and robust trajectory (P0245)



Validation the Authenticity of Virtual fingers Grasping Haptic Algorithm(P0280)

Image Processing and Coding

- Fast wavelet image compression introducing neural network of relay neuron
Zhang Haitao, Zhang Yonglin 0159
- Image smoothing algorithm based on matching normal distribution diffusion
Zhou Xianchun, Wang Meiling, Zhou Linfeng 0169
- Optional features fast fractal image coding algorithm
Yuan Zongwen, Lu Yepin, Yang Hansheng 0177

Image Analysis and Recognition

- Image classification based on global coding combined with multi-scale codebook
Dong Zhenyu, Zhao Jieyu, Zhu Jun 0183

Computer Graphics

- Fast isosurface ray tracing method by combining K-D tree and Shell
Luo Yuetong, Shi Fangfang, Zhang Wei, Zhu Huiguo 0193
- Multi-scale moving grids and its application in clouds scene simulation
Fan Xiaolei, Zhang Limin, Zhang Jianting, Xu Tao 0202

Column of HHME'2014

- Survey on unsupervised spoken term detection for low-resource languages
Yang Peng, Xie Lei, Zhang Yanning 0211
- Image fusion based on fast discrete Curvelet transform
Yang Yong, Tong Song, Huang Shuying 0219
- Fast transform skip mode decision for high efficiency video coding
Wang Ning, Zhang Yongfei, Fan Rui 0229
- Online multi-label image classification with active learning
Xu Meixiang, Sun Fuming, Li haojie 0237
- Action recognition from unconstrained videos via salient and robust trajectory
Yi Yun, Wang Hanli 0245
- Pedestrian re-identification based on feature transformation and dataset classification
Shen Yang, Lin Weiyao, Yin Huiqing 0254
- Comparison of target selection techniques in mobile handheld environment
Xin Yizhong, Li Yan, Yuan Weiqiang, Zheng Qian 0264
- Intent understanding for virtual environments
Cheng Cheng, Zhao Dongpo, Lu Baoan 0271
- Validation the authenticity of virtual fingers grasping haptic algorithm
Yang Wenzhen, Xu Yan, Yan Chuanwu, Wu Xinli, Zhang Hao, Meng Chuang 0280
- Interactive simulation of music instrument playing by a six degree-of-freedom haptic rendering approach
Tong Hao, Shi Youjiao, Wang Dangxiao, Zhang Yuru 0288

中图法分类号: TP37 文献标识码: A 文章编号: 1006-8961(2015)02-0254-10

论文引用格式: Shen Y, Lin W Y, Yin H Q. Pedestrian re-identification based on feature transformation and dataset classification[J]. Journal of Image and Graphics, 2015, 20(2): 0254-0263. [沈洋, 林巍峤, 殷惠清. 特征变换和数据集分块的行人比对[J]. 中国图象图形学报, 2015, 20(2): 0254-0263.][DOI:10.11834/jig.20150212]

特征变换和数据集分块的行人比对

沈洋¹, 林巍峤¹, 殷惠清²

1. 上海交通大学电子信息与电气工程学院, 上海 200240; 2. 数字电视国家工程研究中心, 上海 200240

摘要: **目的** 随着监控摄像头的增多, 基于多摄像头的智能分析显得很重要。基于此, 提出一种新的基于特征变换和数据集分块的行人比对方法。**方法** 首先提出一种新的基于变换矩阵来消除特征差异的算法。这种算法能够在高维空间中, 通过变换矩阵, 让某特征向量逼近另一特征向量, 从而消除了同一行人特征间的差异。与此同时, 还提出一种新的将行人数据集中特征分块的算法, 使每个分块中的行人特征具有相似的性质, 从而能够共用某个变换矩阵, 从而能更好地消除同一行人在不同镜头下的特征差异。**结果** 基于 VIPeR 数据集和复杂街道场景数据集设计了行人比对实验。实验结果表明, 本文的比对方法具有较高的比对准确率, VIPeR 数据集 (50% 训练, 50% 检测) 的 Rank1 的比对准确率为 22%。同时本文设计了特征变换和数据集分块这 2 个模块的对照实验。实验结果表明, 特征变换和数据集分块模块对结果都有提升的效果。**结论** 本文新的行人比对方法通过恰当的特征变换让同一行人在多镜头下的特征互相接近。实验结果表明该方法能够较好地多镜头下匹配行人。

关键词: 行人比对; 数据集分块; 特征变换; 聚类

Pedestrian re-identification based on feature transformation and dataset classification

Shen Yang¹, Lin Weiyao¹, Yin Huiqing²

1. School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China;

2. National Engineering Research Center of Digital Television, Shanghai 200240, China

Abstract: Objective In recent years, a number of surveillance cameras have been placed. The number of surveillance videos that need to be observed and analyzed is rising rapidly because of increased traffic road monitoring and indoor surveillance needs. Single cameras are insufficient to meet the need of certain monitoring tasks because of the significant increase in the number of outdoor cameras. The multi-camera-based intelligent analysis of human behavior is increasingly valuable in the field of video surveillance, and multi-camera-based pedestrian re-identification is a significant research direction. This paper proposes a new method of pedestrian re-identification based on feature transformation and dataset classification.

Method In this method, a new pedestrian re-identification algorithm is proposed based on the transformation matrix that can massively eliminate the differences of the features. This algorithm can be used in high-dimensional space, in which a vector

收稿日期: 2014-07-25; 修回日期: 2014-09-22

基金项目: 国家自然科学基金项目 (61471235); 国家重点基础研究发展计划 (973 计划) 项目 (2013CB329603); 国家“十二五”科技支撑计划项目 (2012BAH07B01); 上海市浦江人才项目 (12PJ1404300)

第一作者简介: 沈洋 (1991—), 男, 现为上海交通大学电子与通信专业硕士研究生, 主要研究行人检测与行人比对。E-mail: shenyang1715@163.com

通信作者: 林巍峤, wylin@sjtu.edu.cn

Supported by: National Natural Science Foundation of China (61471235); National Basic Research Program of China (2013CB329603); Key Projects in the National Science & Technology Pillar Program during the Twelfth Five-year Plan Period (2012BAH07B01)

approaches another vector through the transformation matrix. Therefore, this algorithm is capable of eliminating the feature differences of the same pedestrian. This paper also proposes a new algorithm based on the pedestrian feature classification of datasets; the pedestrian features of each class have similar properties. Thus, these features can share the same transformation matrix, and the algorithm can eliminate the feature differences of a pedestrian under multi-camera. In detail, blocks are built by clustering based on block features. The transformation matrix in each block is trained based on the multi-channel features under different cameras. Each block has a corresponding transformation matrix. The correct block for each testing pedestrian can be determined by comparing that block to that of the training pedestrians. Thus, the transformation matrix of the chosen block is the appropriate matrix. The chosen transformation matrix then eliminates the differences in the features of the testing pedestrian under multi-camera. **Result** Experimental results prove that the proposed method can improve the accuracy of finding the same pedestrian under multi-camera. Specifically, the Rank1 matching rate of the method in VIPeR dataset (50% training, 50% testing) is 22%, which is better than the results of other existing methods. The feature transformation module and the dataset classification module can match the same pedestrian under multi-camera. The matching rate of pedestrian can be reduced significantly if one of the two modules is eliminated. Moreover, the experiments are designed to verify the robustness characteristics of the chosen feature. The experimental results based on real street scene dataset show that the chosen feature is more outstanding and robust than some of the existing features.

Conclusion This paper proposes a new method of pedestrian re-identification based on feature transformation and dataset classification. An appropriate transformation matrix can eliminate the influences of illumination and pedestrian poses under multi-camera. Thus, the method can make the features of the same pedestrian close to each other. The method can also significantly improve the accuracy of matching pedestrians under multi-camera.

Key words: pedestrian re-identification; dataset classification; feature transformation; cluster

0 引言

现今的监控系统一般使用单一摄像头进行监控。每个监控摄像头仅负责自己视角所在的区域,各个摄像头之间可能有内容的重复,但相互之间没有沟通交互。在某些情况下,用户所需监控的范围往往远远超过1、2个摄像头所能覆盖的视角,如在警方办案的过程中往往需要通过各处的摄像头连续多天追踪某一罪犯的行踪,又如交警有时需要通过多个交通监控摄像头来锁定肇事的汽车。这些任务都涉及多摄像头下的监控技术,需要在多摄像头的镜头和场景状况下寻找出某一特定目标。

为了提高不同场景下的目标比对正确率,本文在总结现有不同镜头下行人比对方法的基础上,提出一种基于变换矩阵和数据集分类的行人比对算法。该算法通过利用变换矩阵将某一场景的特征空间映射到另一场景的特征空间,缩小了同一目标在不同场景下的特征差距。同时,提出了一种基于特定特征将行人数据集分块的算法。该算法基于特定的特征,将行人数据集分块,并通过建立目标函数和多次迭代等方法达到最佳的分类效果。能使每类中的行人数据特征具有相似的性质,能够共用某个变换矩阵。最终可以让不同场景下的同一目标的特征差距尽可能地小,而不同目标的特征差距尽可能地大。本文算法的整体框架如图1所示。

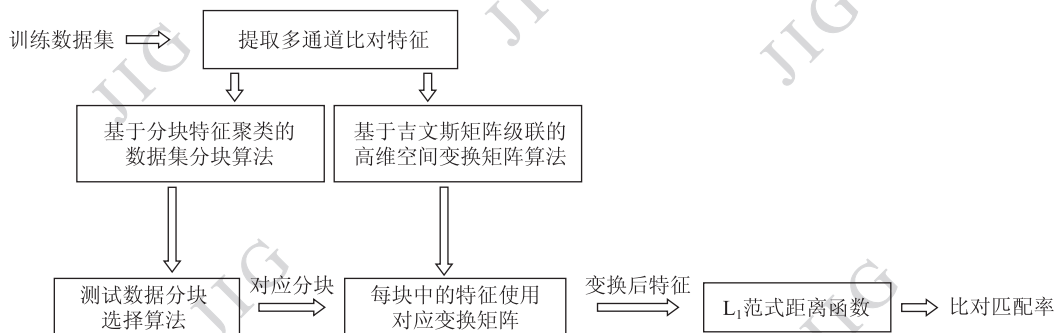


图1 基于变换矩阵和数据集分类的行人比对算法框架图

Fig. 1 Framework of pedestrian re-identification based on the feature transformation and dataset classification

1 相关工作

由于相机镜头的品质变化和不同场景的环境变化,行人在不同的相机镜头下被提取的特征往往有着极大的差异。在对于目标提取特征时,我们主要追求特征的稳定性和区分度。在选择了优秀的行人特征以后,需要进行距离计算,从而得到相应的分数。主流的距离计算方式有 L_1 范式^[1], 巴氏距离^[2], aggregating similarity scores 等等。目前国内外主流的行人再识别方法主要从如下 3 个研究方向进行突破:利用更丰富更易区分的人体细节、选择更恰当的距离函数、提取出更具稳定性统计或局部特征。

Wang 等人^[3,4]利用更丰富更易区分的人体细节来增强特征,主要是利用脸部和姿态的信息来增强特征。由于不同行人的面部是有显著差异的,利用人脸信息的方法^[3]提取出的特征具有很强的区分度。受到目前的交通监控摄像头的品质限制。利用人体细节信息的方法并不是主流的方法。

基于距离学习的行人比对算法的目标是找到更优秀的距离函数。Zheng 等人^[5]利用机器学习和迭代修正目标函数参数的方法,逐步地训练出优秀的距离函数。Schultz 等人^[6-7]计算距离函数时采用更高阶的函数。Gray 等人^[1]则通过 boosting 的变体方法,对特征各个部分的权重比例进行调整,将在当前场景中辨识度高的部分权重增大,以此来提升比匹配率。在通过统计信息和局部信息提取特征时,我们都利用行人图片中的各种有用信息来获得更具稳定度的行人特征。不同的是, Mikolajczyk 等人^[8-9]提出局部信息的方法通常使用特征点提取算子,对行人提取出特征区块,而后针对提取特征区块进行逐一的比对。而利用统计信息的方法针对整个行人外表提取特征,Wang 等人^[10]总结了该特征主要包含了颜色通道信息和轮廓信息,然后进行比对。Orwell 等人^[11-14]在特征中加入 YCrCb, LAB 等颜色通道,颜色通道的信息对于镜头角度和行人姿态的变化不敏感。Felzenszwalb 等人^[15-16]将轮廓信息加入到特征当中,例如 HOG 特征等。这种方法一般使用直方图来表征行人特征,从而降低了维度,又具备一定的旋转不变性。

如上的主流方法致力于寻找更恰当的特征和更优秀的距离函数,但没有从行人特征针对相机镜头变换的角度考虑,通过两相机镜头的特征空间映射,来弥补特征旋转不变性上的不足。本文提出一种更合理的基于变换矩阵和数据集分块的行人比对算法,能在使用较为简单的距离函数的情况下,正确地区分同一行人。

2 本文算法

2.1 基于吉文斯旋转的变换矩阵构造算法概述

行人的特征在 2 个相机镜头下,由于镜头角度和光线强度的不同,往往有很大的差异。很难找到 1 个适用于所有场景的特征。

针对以上问题,提出一种基于吉文斯旋转的变换矩阵构造算法,能够消除同一行人特征在不同镜头下的差异。第 1 节提到同一行人在不同相机镜头下会表现出差异很大的特征,这可以理解为不同的相机镜头拥有不同的行人特征空间。我们的思路是:将不同相机的行人特征映射到 1 个特征空间,在这个特征空间中同一行人的特征向量要尽可能地接近,不同行人的特征向量需要尽可能有差异。为克服以上缺点,使用变换矩阵的方法(图 2)直接将 1 个相机的特征空间映射到另 1 个相机的特征空间,如此一来消除行人在不同相机镜头下的变化,从而提高行人比对的准确率。

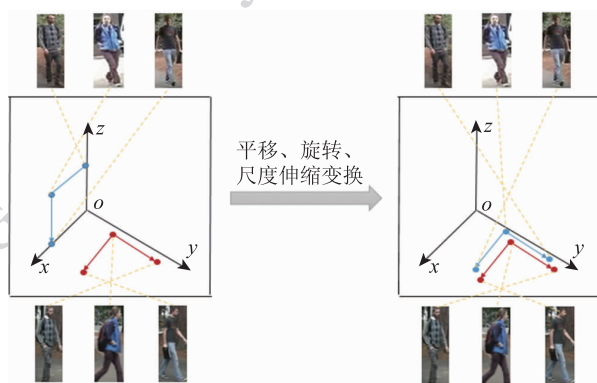


图 2 行人特征在低维空间的变换示意图

Fig. 2 Feature transformation in low-dimensional space

2.2 行人数据集聚类分块算法概述

2.1 节提到使用变换矩阵的方法直接将 1 个相机的特征空间映射到另 1 个相机的特征空间,以此来消除行人在不同相机镜头下的变化。由于

行人特征数据的多样性, 统一的映射方式(统一的变换矩阵)并不能适用于数据集中的所有行人特征, 所以我们需要将行人数据集分块。每小块中包含不定数目的行人特征, 要使这些特征可以共享同1个映射方式, 也即可以共享同一组平移旋转矩阵。

在对行人数据集进行分块时, 首先要在行人特征的基础上再进行组合, 提取出分块特征。在组合行人特征构成分块特征时, 加入了特征相互位置关系信息和特征点本身的信息, 然后我们需要相互位置关系和特征本身信息接近的特征分在同一块中。

3 实 现

3.1 基于吉文斯旋转的变换矩阵构造算法实现

3.1.1 低维空间的平移旋转变换

重点考虑旋转变换3维情况下可以将旋转变换矩阵看成是3个基本旋转矩阵的复合。3个基本旋转则是分别绕着 x, y, z 这3轴的旋转, 分别称为 pitch, yaw 和 roll 旋转。

3维旋转矩阵 M 可以用3个基本旋转矩阵 R 的连乘表示, 即

$$M = R_x(\theta_x)R_y(\theta_y)R_z(\theta_z) \quad (1)$$

3.1.2 基于吉文斯旋转矩阵的高维空间变换矩阵

上升到更高维的情况, 3维旋转变换的方法并不能直接使用。为有效解决该问题, 提出级联吉文斯旋转矩阵的方法来求解近似旋转变换矩阵。

其中吉文斯旋转(Givens rotation)是某2个坐标轴所确定的平面中的旋转。当固定了2个坐标轴之后, 可以在这2个坐标轴所确定的平面上计算出两向量的旋转角度。4维的吉文斯旋转矩阵为

$$G(i, k, \theta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & c & -s & 0 \\ 0 & s & c & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

吉文斯旋转矩阵中的非零元素共有6个。除了矩阵对角线的2个值为1的端点, 出现在 i 行 i 列和 k 行 k 列的4个交叉点上, 其中的 i 和 k 是构建矩阵之前所确定的2个维度, θ 是旋转角。对于每1

个吉文斯旋转矩阵, 虽然维度很高, 但是矩阵非常稀疏, 主要的运算集中在 $c = \cos(\theta)$ 和 $s = \sin(\theta)$ 所在的 2×2 部分, 所以单个吉文斯旋转矩阵的运算的复杂度并不高。

对于高维空间的变换, 可以通过遍历地选择高维空间的2个维度, 进行吉文斯旋转, 最后将所有吉文斯旋转矩阵级联起来, 近似代替高维变换矩阵。

假设特征空间的维度为 N 维, 在选择2个维度时, 当确定1个维度, 则有 $n-1$ 维度可以和它组合, 遍历到整个 N 维可以得出共有 $n(n-1)/2$ 种组合。也即共有 $n(n-1)/2$ 个吉文斯矩阵, 通过级联这些吉文斯旋转矩阵, 得到近似的旋转变换矩阵。

相比旋转变换而言, 平移变换和3维空间的情况没有太多的区分, 将向量的起始点移动到目标点, 向量中的剩余的点随着起始点同时移动即可。

对数据集中的行人提取特征后, 每个行人可表示为1个特征向量, 也即可以用上述方法得到变换矩阵来消除同一行人在不同镜头下的特征差异。

取典型情况, 利用局部线性嵌入(LLE)算法将高维特征点(200维)压缩到3维, 并作图像展示基于吉文斯矩阵的高维空间特征变换的效果, 如图3所示。图3中“*”“o”两符号分别表示两相机镜头下的行人特征点, 1个行人用1种颜色表示。

针对数据集中的 m 个行人构造变换矩阵时, 需要全面考虑 m 个行人。我们分别取在2个镜头下行人特征的平均值作为特征来构建高维变换矩阵。经过实验证实, 当 m 较小(小于6), 且在2个镜头下, m 个行人的特征相互之间都比较相似时, 构建出的变换矩阵可以适用于这 m 个行人。

3.2 行人数据集聚类分块算法

3.2.1 行人数据集的行人特征和分块特征

之所以要进行数据集的分块, 是为了让测试数据的行人特征找到最合适的变换矩阵, 从而尽可能消除同一行人在不同镜头下的特征差异。

从3.1节可知, 需要用1个特征来表征2个镜头下行人的特征之间的相似性, 称此特征为分块特征。

分块特征由行人特征组合变换得到, 分块特征接近的行人, 可以共用1个变换矩阵。为了使

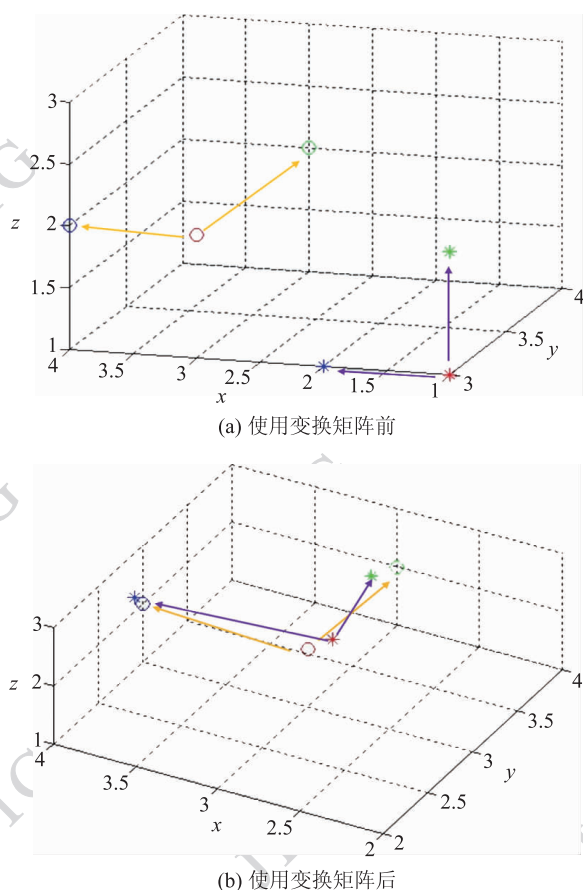


图3 高维空间特征变换的低维展示

Fig. 3 Result of high-dimension feature transformation showed in low-dimension space

行人特征具有区分度,该行人特征中包含了8个颜色通道(RGB, HS 和 YCbCr),8个被不同参数的Gabor滤波器滤波的亮度通道,各个通道如图4所示。

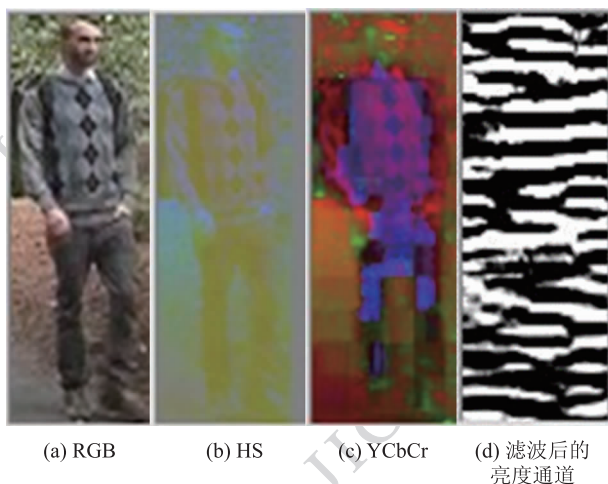


图4 行人特征各个通道示意图

Fig. 4 Pedestrian features of each channel

对行人按照水平方向均匀分成6块,对于每块都用直方图来统计信息,直方图的横轴取值范围依然是 $[0, 255]$ 。由于这样得出的行人特征维数过高,对每个直方图的横轴进行降采样,将直方图的横轴均分成16块。这样就得到 $1536(16 \times 6 \times 16)$ 维度的行人特征。

下面详细描述如何通过行人特征组合变换,最终得到3072维度的分块特征 f 。

由于分块特征表征了多个行人特征之间的相似性(2个镜头下)。首先保证在单一镜头下,行人的特征之间比较相似,故将单一镜头下行人特征本身加入分块特征。分块特征中的此部分记为 x 。

行人特征 x 仅能表示行人在单一镜头下特征的相似度,而分块特征表征2个镜头下特征的相似度。故需要在分块特征中加入同一行人在2个镜头下的2个特征的相互关系。设某行人在镜头 x 下的特征点为 x ,在镜头 y 下的特征点为 y ,则 $y - x$ 表示为1个行人在2个相机镜头下的特征差向量。如果数据集中 m 个行人的特征差向量 $y - x$ 之间很接近,这表明 m 个行人在2个镜头下的特征点相互位置关系很类似,若这些行人在单个镜头下特征本身也很接近,则说明在2个镜头下这 m 个行人的特征都很相似,也即这 m 个行人可以公用1个旋转变换矩阵。也即通过找到某组特征差向量 $y - x$ 类似的行人,就找到了可以共用1个旋转矩阵的一组行人。故将特征差向量 $y - x$ 加入分块特征。

最终,得到3072维度的分块特征

$$f = w(y - x) + x \quad (3)$$

式中, w 用于调整两者之间的权重关系,在本文的实验中,行人特征本身 x 在分块特征中有更大的权重, w 取0.2。

3.2.2 行人数据集的分块方式

通过对训练数据的行人特征进行组合变换得到分块特征之后,基于分块特征进行聚类,得到不同镜头下的分块。然后基于每个分块中的行人特征按照3.1节的方法对每个分块构造变换矩阵,也即每个分块对应1个变换矩阵。

基于分块特征运用 k -平均算法将训练行人数据集划分成 k 个块。其中距离函数通常使用 L_2 范式的平方。由于在设置初始点时是随机地选取 k 个点作为每块的初始中心。由于初始值是随机选取的,比对的结果因此会有一定影响。

对 k 均值算法的每块中的初始中心选择方法, 进行改良。理想的聚类结果是: 1 个块中的分块特征很相似, 而不同块之间特征差异很大。我们通过多次迭代计算的方式, 比较最后得到的每个簇的中心值, 找到各个簇的中心值相差最远, 中心值两两之间最不相似的初始值。这组迭代产生初始值是由相异聚类中的对象的不相似特性所得到的, 与随机产生初始值相比, 选用这组初始值显然会获得更好的聚类效果。

由于训练数据的分块特征中行人特征本身 x 占有很大的权重, 且测试数据不具有不同镜头下行人的对应关系, 无法获得同一行人在不同镜头下的特征差向量 $y - x$, 故选用测试数据的行人特征。基于测试数据的行人特征, 用最近邻居法将测试数据对应到合适的分块。在最后比对的步骤中, 基于特征差向量 $y - x$ 排除一些明显的错误分块。

在最近邻居法中, 首先需要先确定 k 值, 所谓 k 值既是最近 k 邻居算法中 k 的大小, 含义是对于 1 个待分类的数据点, 需要为其寻找邻居数目。其次, 选择距离函数表达, 本文选择的是欧氏距离。然后

将测试数据的行人特征和训练数据集中行人特征进行比较, 计算距离, 找出距离测试数据的最近 k 个训练数据, 并统计这 k 个样本分别所属的分块, 测试数据则对应到累计最多的分块中, 也即测试数据找到了合适的变换矩阵。在本文的实验中, k 取值为 6。

最后, 对测试数据中某行人在镜头 x 下的特征进行变换后, 和镜头 y 下的行人特征比对。理想情况下, 特征最接近的是相同的行人。

基于镜头 x 下变换后的特征, 锁定了镜头 y 下的候选行人特征 y 之后, 计算该行人的特征差向量 $y - x$, 若该行人的特征差向量与选定分块中的训练数据的平均特征差向量的相似度低于阈值, 则重新选择对应的分块。

数据集分块流程如图 5 所示。最后将数据点归类到数量最大的邻近类别中。这样我们就给测试数据集中的某个行人找到对应的分块, 也即该行人特征使用此分块的变换矩阵后。在整个数据集的行人特征都找到对应的分块, 并使用了相应的变换矩阵之后, 使用简单的 L_1 范式作为距离函数也能得到较高的准确率。

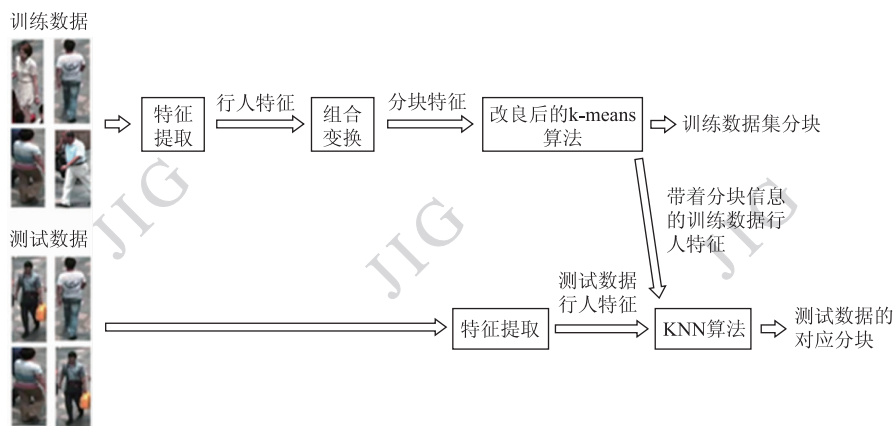


图5 数据集分块部分的整体流程

Fig. 5 Framework of dataset clustering

4 实验结果

为了比较本文中实现的基于数据集分块及变换矩阵的行人比对算法和其他行人比对算法的优劣, 选用行人比对最常用的 VIPeR 数据集, 该数据集包含在 2 个相机镜头下获取到的 632 个行人组。同时还构建了 1 个基于复杂街道场景的数据集, 该数据

集包含 200 个行人组。

用 CMC(累计匹配特性)曲线来表征行人比对效果。

设数据集中共有 L 个行人, $X = \{x_i | 0 < i < L + 1\}$ 是镜头 x 下的行人数据集, $Y = \{y_i | 0 < i < L + 1\}$ 是镜头 y 下的行人数据集。 $\{x_i, y_i\}$ 表示第 i 个行人在两监控摄像头下的行人对。匹配率为

$$R_a = \frac{\sum_{1 \leq i \leq N} \varepsilon(r - \text{rank}(y_i))}{N} \times 100\% \quad (4)$$

式中, r 表示 CMC 曲线横轴某点, $rank$ 函数获取该行人的相似度排序, ε 表示阶跃函数。

同时基于 VIPeR 数据集设计了对照实验, 实验结果证实特征变换模块和数据集分块模块存在的必要性, 2 个模块对结果都有提升的效果。

4.1 VIPeR 数据集

由于两镜头下的图像具有不同的光线, 行人角度和姿势, VIPeR 数据集在行人比对领域是最常用也较具挑战性的数据集, 该数据集的例子如图 6 所示。将该数据集中的图像归一化到统一的大小, 并省略了提取前景的操作, 也即在比对时行人图像中包含着少量背景, 这也增加了比对的难度。

4.1.1 VIPeR 数据集的 CMC 曲线

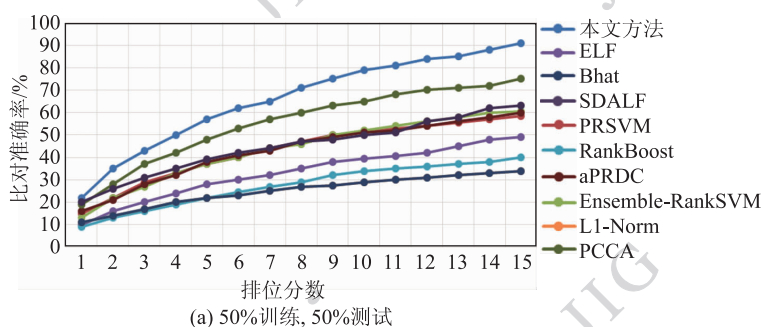
随机选取 VIPeR 数据集中的 50% (316 个行人) 作为训练数据, 剩下 50% (316 个行人) 作为测试数据, 并运用多种的比对方法得到相应的 CMC 曲线, 如图 7(a)。另外也随机选取 VIPeR 数据集中的

25% (158 个行人) 作为训练数据, 剩下 75% (474 个行人) 作为测试数据, 运用多种的比对方法得到相应的 CMC 曲线, 如图 7(b) 所示。

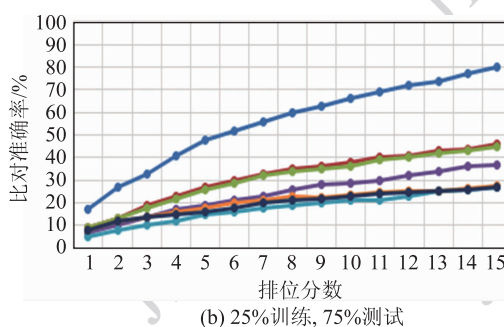


图 6 VIPeR 数据集图像案例

Fig. 6 Examples of VIPeR dataset



(a) 50%训练, 50%测试



(b) 25%训练, 75%测试

图 7 VIPeR 数据集的 CMC 曲线

Fig. 7 CMC curves of VIPeR dataset

在图 7 中, 将本文方法和下面的目标比对方法进行了比较: 1) Ensemble-RankSVM^[17], 2) PRSVM^[17], 3) ELF^[2], 4) RankBoost^[18], 5) Bhat^[17], 6) L1-Norm^[17], 7) aPRDC^[19], 8) SDAALF^[21], 9) PCCA^[20]。

本文行人比对算法主要是基于数据集分块和特征变换, 距离函数选择的是 L_1 范式。从图 7 中可以看出, 同样使用简单距离函数 (L_1 范式, Bhattacharyya 距离) 的普通比对方法的匹配效果很差, 也表明该数据集的挑战性。

同时也可以看出, 本文方法虽然使用了简单的距离函数, 但在匹配率上相比于其他比对方法有着较大的提升, 这得益于合理的行人特征变换。

特别地, 本文方法在训练数据较少的情况下, 比现有其他方法有着较大的提升。如图 7(b) 在训练和检测数据 (25%, 75%) 的情况下, Rank1 时本文方法匹配率是 17%, 而其他比对方法中最好的匹配率是 9%, 也即本文比对方法在 Rank1 中匹配率提升了 89%。

4.1.2 特征变换和数据集分块模块的对照实验

为了证实本文算法中特征变换模块和数据集分块模块存在的必要性, 我们设计了对照实验。在 VIPeR 数据集上训练和检测数据 (50%, 50%) 的情况下, 加入仅用特征变换模块时的行人比对率和仅用数据集分块模块时的行人比对率。

仅用特征变换模块,也即不存在分块对应,全部测试数据中的行人特征都用1个变换矩阵;仅用数据集分块,也即测试数据仅对应分块,不使用变换矩阵,将特征直接进行比较。实验结果如图8。

由图8可知,特征变换和数据集分块2个模块中缺少任何1个,都会明显地降低行人比对的准确率。这说明这2个模块都有必要存在。

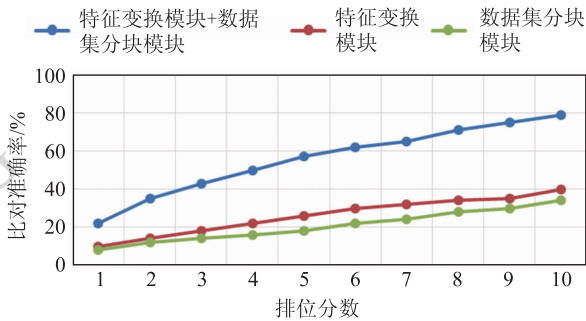


图8 特征变换和数据集分块模块的对照实验

Fig. 8 Control experiment of feature transformation and dataset classification

4.2 复杂街道场景数据集

为了进一步验证方法的有效性,构建1个基于复杂街道场景的数据集,并基于该数据集来验证我们选取的行人特征(8个颜色通道加8个经过不同参数的 Gabor filter 过滤的亮度通道的分块直方图统计特征)的优越性。该数据集由实际街道场景监控视频得到,从1组监控视频中取出200对行人,同时将本文的行人特征与现有的其他行人特征(HOG, CENTRIST, SIFT等)进行对比实验。

构建的数据集的实际场景以及得到的个体目标框见图9。需要指出的是,本数据集具有较大的挑战性,表现为:1)由于场景中目标较为拥挤,因此有些目标框中会出现被其他目标遮挡的情况,如图9(b);2)为了使实验更接近实际应用场景,我们采用自动的基于行人检测的方法^[22]代替手动方法实现视频中目标框的提取,这样提取出的目标框的精度和手动提取的相比也有一定的下降(如框得过大或过小等),这也进一步增加了比对的难度。

表1显示了本文方法和现有方法不同 Rank 下的比对率。同时,从表1中选取匹配效果最优秀的4个和本文方法的行人特征进行比较,做出 CMC 曲线,这4个特征分别是 HOG, HS, LSF 和 MFF,如图10所示。



(a) 街道场景



(b) 框出的行人

图9 复杂街道场景数据集案例

Fig. 9 Examples of street scene dataset

表1 实际街道场景数据集的比对结果

Table 1 Matching rate of street scene dataset

方法	排位分数			
	1	5	10	20
本文	62	86	98	100
HOG	39	54	70	84
CENTRIST	17	30	44	60
SIFT	25	36	50	72
LAB	22	50	68	79
LSF	32	55	71	80
HS	36	60	79	86
MFF	50	68	79	88

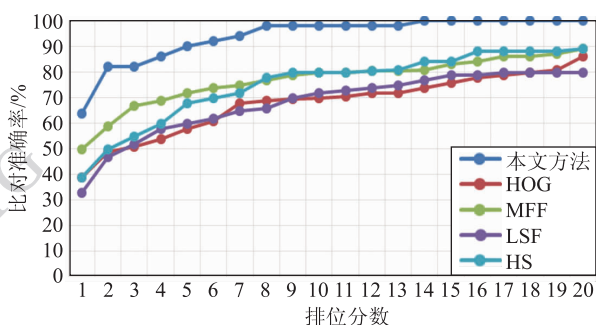


图 10 实际街道场景数据集的 CMC 曲线

Fig. 10 CMC curves of street scene dataset

从表 1 和图 10 可知,本文选用的行人特征在匹配效果上要优于其他行人特征,特别是随着 Rank 增加时,行人匹配率可以以很快的速度上升,当 Rank 等于 10 时,匹配率已经非常接近 100%。从整体上看,使用颜色通道的行人特征(LAB, HS)的比对算法在匹配率方面,要比使用轮廓信息的行人特征(SIFT, HOG, CENTRIST)的比对算法更高,特别是 Rank 大于 4 之后,匹配效果上有着明显的优势,故本文选用的行人特征中没有包含轮廓信息。同样也是使用多通道特征的 MFF,在匹配效果上要优于使用单一通道的特征,这说明只使用表征行人某方面属性的单一特征来做行人比对效果不理想,而用多通道融合的方法可以增强行人特征的区分度和稳定度,有助于提高匹配率。

5 结 论

为了提高不同场景下的目标比对准准确率,本文提出了一种基于数据集分块和特征空间变换的目标比对、再识别算法。该算法通过利用变换矩阵将 2 个相机镜头下的行人特征映射到 1 个特征空间,缩小了同一目标在不同场景下的特征差距。

同时,提出了一种基于特定特征将行人数据集的行人分类的算法。该算法基于特定的特征,将行人数据集分类,并通过建立目标函数和多次迭代等方法达到最佳的分类效果。能使每类中的行人数据特征具有相似的性质,能够共用某个变换矩阵,最终使在选择简单的距离函数的情况下,也达到较好的比对准准确率。

目前本文算法只能通过有监督学习来得到分块。在未来的工作中,将尝试运用无监督的方法来得到数据集的分块,使得算法更加鲁棒和全面。

参考文献 (References)

- [1] Wang X, Doretto G, Sebastian T, et al. Shape and appearance context modeling[C]//Proceedings of IEEE International Conference on Computer Vision. Washington DC: IEEE Computer Society, 2007: 1-8.
- [2] Gray D, Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features[C]//Proceedings of Computer Vision-ECCV. Berlin Heidelberg: Springer, 2008: 262-275.
- [3] Hsu R L, Abdel-Mottaleb M, Jain A K. Face detection in color images[J]. Pattern Analysis and Machine Intelligence, 2002, 24(5): 696-706.
- [4] Wang L, Tan T, Ning H, et al. Silhouette analysis-based gait recognition for human identification[J]. Pattern Analysis and Machine Intelligence, 2003, 25(12): 1505-1518.
- [5] Zheng W S, Gong S, Xiang T. Reidentification by relative distance comparison[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(3): 653-668.
- [6] Lee J E, Jin R, Jain A K. Rank-based distance metric learning: an application to image retrieval[C]//Proceedings of IEEE Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2008: 1-8.
- [7] Schultz M, Joachims T. Learning a distance metric from relative comparisons[J]. Advances in Neural Information Processing Systems, 2004, 41(1): 3-19.
- [8] Cohen W, Ravikumar P, Fienberg S. A comparison of string metrics for matching names and records[C]// Proceedings of KDD Workshop on Data Cleaning and Object Consolidation. New York: ACM Digital Library, 2003: 73-78.
- [9] Mikolajczyk K, Schmid C, Zisserman A. Human detection based on a probabilistic assembly of robust part detectors[C]//Proceedings of Computer Vision-ECCV. Berlin Heidelberg: Springer, 2004: 69-82.
- [10] Hamdoun O, Moutarde F, Stanculescu B, et al. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences[C]//Proceedings of IEEE International Conference on Distributed Smart Cameras. Washington DC: IEEE Computer Society, 2008: 1-6.
- [11] Wang X. Intelligent multi-camera video surveillance: a review[J]. Pattern recognition letters, 2013, 34(1): 3-19.
- [12] Orwell J, Remagnino P, Jones G A. Multi-camera colour tracking[C]// Proceedings of IEEE Visual Surveillance. Washington DC: IEEE Computer Society, 1999: 14-21.
- [13] Krumm J, Harris S, Meyers B, et al. Multi-camera multi-person tracking for easy living[C]//Proceedings of IEEE Visual Surveillance. Washington DC: IEEE Computer Society, 2000: 3-10.
- [14] Mittal A, Davis L S. M2tracker: a multi-view approach to segmenting and tracking people in a cluttered scene[J]. International

- al Journal of Computer Vision, 2003, 51(3): 189-203.
- [15] Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with discriminatively trained part-based models [J]. Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.
- [16] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]//Proceedings of IEEE Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2005: 886-893.
- [17] Prosser B, Zheng W S, Gong S, et al. Person re-identification by support vector ranking [C]//Proceedings of British Machine Vision Conference. Aberystwyth, Wales: Aberystwyth University, 2010: 1-11.
- [18] Freund Y, Iyer R, Schapire R E, et al. An efficient boosting algorithm for combining preferences [J]. The Journal of Machine Learning Research, 2003, 4(1): 933-969.
- [19] Liu C, Gong S, Loy C C, et al. Person re-identification: what features are important? [C]//Proceedings of Computer Vision-ECCV. Berlin Heidelberg: Springer, 2012: 391-401.
- [20] Mignon A, Jurie F. PCCA: a new approach for distance learning from sparse pairwise constraints [C]//Proceedings of IEEE Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2012: 2666-2672.
- [21] Farenzena M, Bazzani L, Perina A, et al. Person re-identification by symmetry-driven accumulation of local features [C]//Proceedings of IEEE Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2010: 2360-2367.
- [22] Wu J X, Liu N, Geyer C, et al. C4: a real-time object detection framework [J]. IEEE Transactions on Image Process, 2013, 22(10): 4096-4107.