多源数据行人重识别研究综述

叶钰[1, 2, 3] 王正[1, 4] 梁超[1, 2, 3] 韩镇[1, 2, 3] 陈军[1, 2, 3] 胡瑞敏[1, 2, 3]

摘 要 行人重识别是近年来计算机视觉领域的热点问题, 经过多年的发展, 基于可见光图像的一般行人重识别技术已经趋近成熟. 然而, 目前的研究多基于一个相对理想的假设, 即行人图像都是在光照充足的条件下拍摄的高分辨率图像. 因此虽然大多数的研究都能取得较为满意的效果, 但在实际环境中并不适用. 多源数据行人重识别即利用多种行人信息进行行人匹配的问题. 除了需要解决一般行人重识别所面临的问题外, 多源数据行人重识别技术还需要解决不同类型行人信息与一般行人图片相互匹配时的差异问题, 如低分辨率图像、红外图像、深度图像、文本信息和素描图像等. 因此, 与一般行人重识别方法相比, 多源数据行人重识别研究更具实用性, 同时也更具有挑战性. 本文首先介绍了一般行人重识别的发展现状和所面临的问题, 然后比较了多源数据行人重识别与一般行人重识别的区别, 并根据不同数据类型总结了 5 类多源数据行人重识别问题, 分别从方法、数据集两个方面对现有工作做了归纳和分析. 与一般行人重识别技术相比, 多源数据行人重识别的优点是可以充分利用各类数据学习跨模态和类型的特征转换. 最后, 本文讨论了多源数据行人重识别未来的发展.

关键词 多源数据行人重识别, 跨模态, 度量学习, 特征模型, 统一模态

引用格式 叶钰, 王正, 梁超, 韩镇, 陈军, 胡瑞敏. 多源数据行人重识别研究综述. 自动化学报, 2020, **46**(9): 1869−1884

DOI 10.16383/j.aas.c190278

A Survey on Multi-source Person Re-identification

YE Yu[1, 2, 3] WANG Zheng[1, 4] LIANG Chao[1, 2, 3] HAN Zhen[1, 2, 3] CHEN Jun[1, 2, 3] HU Rui-Min[1, 2, 3]

Abstract Person re-identification (Re-ID) has been a popular and well-investigated topic in computer vision community. However, current researches have a relatively ideal assumption that person images are captured under a sufficient light condition and with high-resolution. Although most researches can achieve very exciting performances, they are not suitable for practical applications. Since practical conditions are a little complicated, and there are multiple sources to represent persons' appearance. In this paper, we focus on the multi-source person Re-ID, which refers to the problem of using multiple sources of data for person re-identification. Compared with general person Re-ID methods, multi-source person Re-ID researches are more practical, yet more challenging in reality. We need to face challenges caused by domain gap among different data sources, such as low-resolution images, infrared images, depth images, text information and sketch images. In this paper, we start with a brief introduction of general person Re-ID. The differences between general and multi-source person Re-ID are then compared. Five types of multi-source person Re-ID are further analyzed and summarized. From these discussions, it will become evident that several advantages exist in multi-source person Re-ID over general person Re-ID methods, as the former can make full use of data sources to learn cross-modality feature transformation. Finally, the future trends of multi-source person Re-ID are discussed.

Key words Multi-source person re-identification (Re-ID), cross-modality, metric learning, feature model, modality unifying

Citation Ye Yu, Wang Zheng, Liang Chao, Han Zhen, Chen Jun, Hu Rui-Min. A survey on multi-source person re-identification. *Acta Automatica Sinica*, 2020, **46**(9): 1869−1884

随着视频监控系统在城市中的广泛应用, 利用摄像机拍摄的画面判断出现在不同图像中的行人是否是同一个人, 并通过摄像机生成轨迹预测他们行为的技术已经广泛应用于智能视频监控、安保、刑侦等领域, 在日常调查中发挥着越来越重要的作用.

收稿日期 2019-04-01 录用日期 2019-10-17
Manuscript received April 1, 2019; accepted October 17, 2019
国家重点研发计划项目 (2017YFC0803700), 国家自然科学基金青年项目 (61801335, 61876135), 湖北省自然科学基金群体项目 (2018CFA024, 2019CFB472, 2018AAA062) 资助
Supported by National Key Research and Development Program of China (2017YFC0803700), National Natural Science Foundation of China (61801335, 61876135), and Natural Science Foundation of Hubei Province (2018CFA024, 2019CFB472, 2018AAA062)
本文责任编委 赖剑煌
Recommended by Associate Editor LAI Jian-Huang
1. 武汉大学计算机学院国家多媒体软件工程技术研究中心 武汉 430072, 中国 2. 武汉大学多媒体网络通信工程湖北省重点实验室 武汉 430072, 中国 3. 地球空间信息技术协同创新中心 武汉 430072, 中国 4. 日本国立信息学研究所 东京 1638001, 日本

1. National Engineering Research Center for Multimedia Software, School of Computer Science, Wuhan University, Wuhan 430072, China 2. Hubei Key Laboratory of Multimedia and Network Communication Engineering, Wuhan University, Wuhan 430072, China 3. Collaborative Innovation Center of Geospatial Technology, Wuhan 430072, China 4. National Institute of Informatics, Tokyo 1638001, Japan

这种运用计算机视觉和机器学习等方法判断某个摄像机中的特定行人是否出现在其他摄像机中的技术称为行人重识别 (Person re-identification, Re-ID), 如图 1 所示. 行人重识别不仅具有非常迫切的应用需求, 还具有非常重要的研究价值, 近年来, 行人重识别引起了学术界和工业界的广泛关注, 是计算机视觉领域的一个研究热点. 经过 10 多年的发展, 国内外相继提出了大量行人重识别模型, 在限定的仿真条件下已经取得了非常高的准确率[1−3], 在 Market-1501数据集上达到了 94.0 %, 在 CUHK03 数据集上则为 96.1 %, 这一准确率甚至超过了人类视觉的能力.

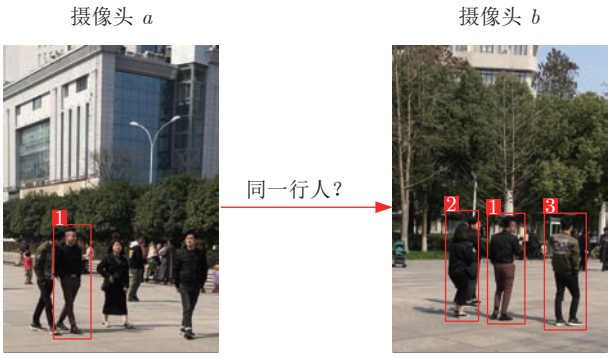


图 1 行人重识别示意图

Fig. 1 An example illustrating person re-identification

行人重识别研究是基于监控视频检索, 具有其特殊性, 在实际城市视频监控中, 行人对象的画面质量较差、分辨率较低, 而且还存在明显的视角、光照变化[4]. 因此, 相对于通用图像检索, 行人重识别仍面临以下问题: 1) 在不同监控摄像机中, 行人与摄像机的距离不同, 导致不同摄像机视域下的行人图像分辨率、光照和视角不同, 同一个行人对象的不同图像视觉特征会产生明显的变化; 2) 同一行人在不同摄像机视域拍摄的画面中受背景和其他因素导致的遮挡程度不同, 大量的行人遮挡问题导致完整的行人图像比较少; 3) 由于受行人姿势及摄像机角度变化影响, 在不同监控摄像机中, 不同行人图像之间的视觉特征差异可能比较小. 此外, 一些特定的问题也没有受到足够的重视, 比如大规模快速检索问题、数据不足问题、实际环境中人员信息情况复杂跨多模态问题等[3], 这使得行人重识别问题比一般基于实例的图像/视频检索更加困难.

现有的行人重识别工作多使用一般可见光摄像机所获取的同一类型数据, 然而实际生活中摄像机采集到的图像质量参差不齐, 仅利用可见光摄像机采集的图像取得的识别效果可能并不尽人意, 往往还需要结合其他类型的数据信息才能取得良好的效果, 如图像数据与视频数据[5]、可见光图像数据与其他图像数据、图像数据与文本数据等. 如果将同一数据特性下的行人重识别问题认为是一般的行人重识别, 则与之相对应, 我们总结了使用多种数据进行行人重识别的方法, 称之为多源数据行人重识别 (图 2). 由于数据来源和数据类型并不一致, 其成像原理和图像质量也不一致, 因此多源数据行人重识别除了需要克服一般行人重识别面临的问题外, 还需要着重解决跨模态的特征匹配这一关键难题.

在实际的行人识别过程中可用信息来源较多, 但鉴于数据获取和利用的难易程度, 本文所说的多源数据行人重识别主要考虑以下几种交叉类型/模态的行人重识别问题: 1) 使用不同的相机规格和设置, 如高分辨率与低分辨率图像; 2) 使用不同的拍摄设备, 如可见光与红外摄像机, 可见光与深度传感器; 3) 根据历史文档记录或对行人的描述获得的文本信息; 4) 由专家或者数字传感器自动获得的图像, 如刑侦系统使用的素描与数字照片.

低分辨率: 在当前社会的安全环境考虑下, 将低分辨率行人图像与高分辨率行人图像进行匹配是一个热门挑战. 而受到环境、成像条件等多方面因素的影响, 实际视频侦查中得到的行人图像分辨率多变, 且分辨率往往较低. 在这种数据特性更复杂的情况下, 传统的基于单一高分辨率的行人重识别方法辨识能力显著降低.

红外: 红外 (Infra-red, IR) 图像是由红外设备而非可见光设备拍摄的. 红外设备可以在可见光不可控的环境下建立受控的拍摄条件, 但红外设备的成像原理和方式与可见光设备完全不同, 多源数据行人重识别的挑战在于将红外图像与可见光图像进行匹配.

深度图像: 深度图像 (Depth image) 也称为距离图像 (Range image), 是指将从摄像机到场景中各点的距离 (深度) 作为像素值的图像, 它直接反映了被拍摄物体可见表面的几何形状, 在行人衣着发生改变或照明条件较差时常使用深度图像进行行人身份识别.

文本: 大多数视频监控系统都依赖于在不同摄像机视域下拍摄的视频. 事实上, 在调查过程中, 除了监控视频外, 调查人员还手工标注了一些注记, 这些注记虽然信息不完整, 但准确性高, 有助于识别行人身份. 文本-视觉匹配旨在测量文本描述与图像之间的相似性.

素描: 素描行人重识别是根据手工或软件绘制

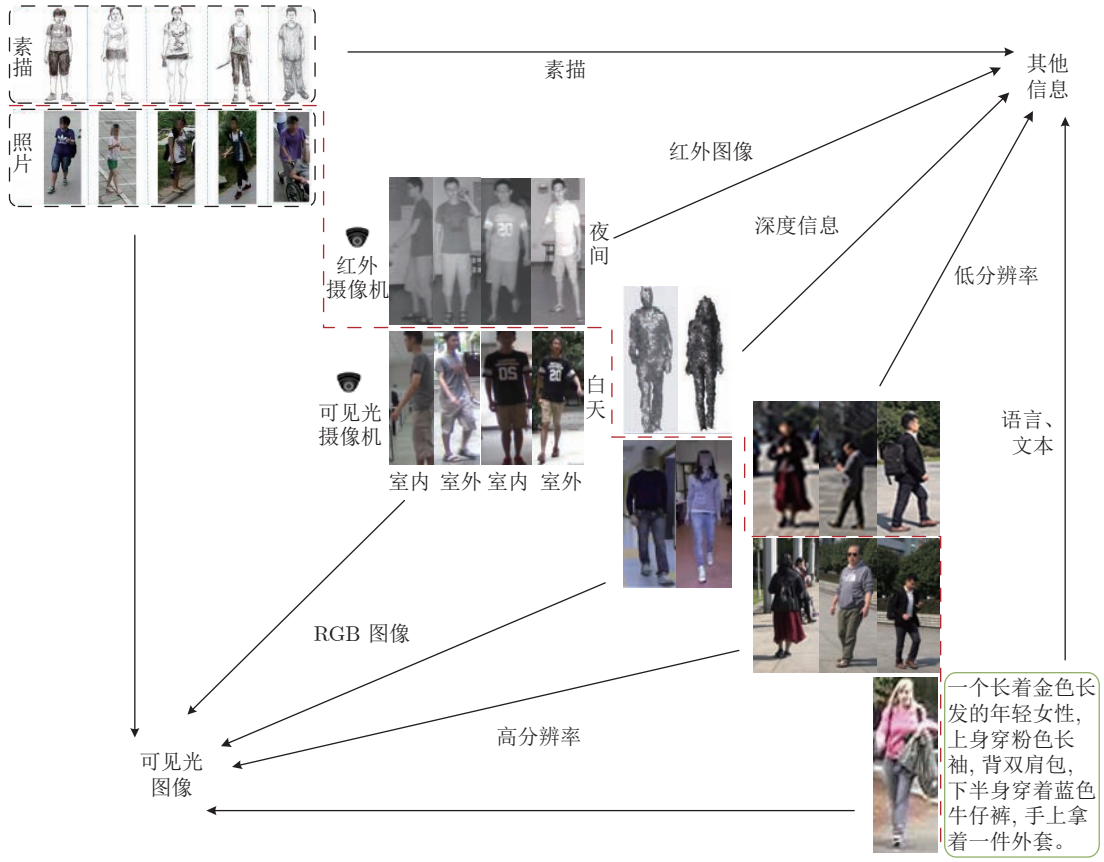


图 2　多源数据行人重识别类型

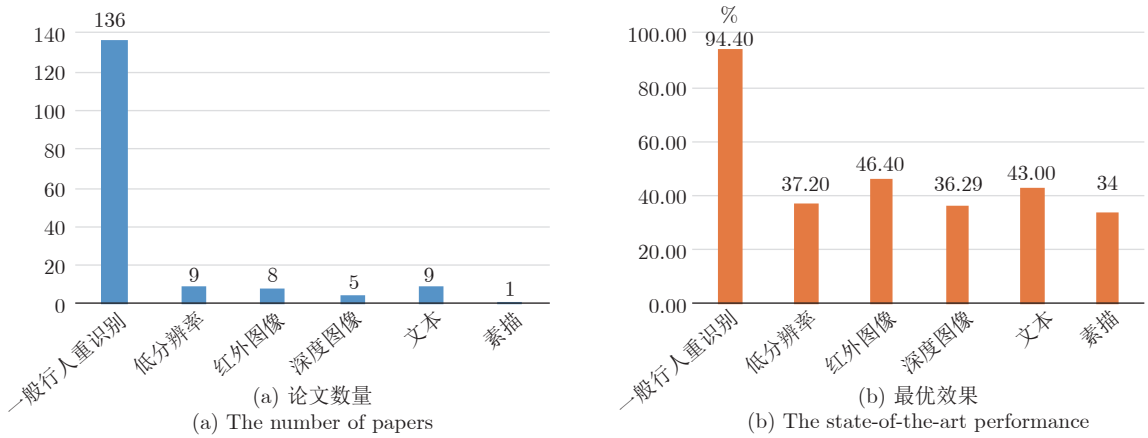
Fig. 2　Scope of multi-source data person re-identification studied in this survey

的行人全身素描图像, 与照片数据库中的行人图像进行匹配的过程. 在无法获取目标人物照片的情况下, 素描行人重识别能根据专业人员绘制的素描图像自动搜索所有监控图像, 迅速缩小目标人物的范围, 具有重要的现实意义.

1　多源数据行人重识别的基本特征

传统的行人重识别方法从特征提取和度量学习两个方向进行研究, 2014 年, Li 等[6] 率先使用深度学习方法进行行人重识别研究, 此后越来越多的研究者尝试将深度学习方法与行人重识别研究进行结合. 行人重识别的基本程序如下: 1) 根据行人特征提取方法从检索图片/视频库中提取特征; 2) 针对提取的特征利用相似性判别模型进行训练, 获得能够描述和区分不同行人的特征表达向量, 度量计算特征表达向量之间的相似性; 3) 根据相似性大小对图像进行排序, 将相似度最高的图像作为最终的识别结果. 近年来, 一般行人重识别技术在公共行人识别数据集上获得了很高的精确度, 但这些方法大多是基于一个关键的假设, 即所有人的图像都是在白天用可见光相机拍摄的, 且具有统一和足够高的分辨率. 而在实际应用过程中, 总是存在各种分辨率和尺度 (包括低分辨率和小尺度) 的图像; 在照明条件较差的夜间或者室内通道等环境下, 则常常利用红外设备或深度传感器而不是可见光设备进行拍摄; 此外, 刑侦人员通常还需要依靠证人的描述和素描图像来检索系统中的人物图像. 在这些情况下, 数据本质有很大的变化, 一般的行人重识别模型在匀质条件下的设计将失去其有效性.

多源数据行人重识别则针对每类数据使用一个特定于该类型的网络来提取或构造特定信息并映射到同一个表达空间, 然后, 利用共享网络在共享表达空间中生成特征, 这个通用的重识别网络通过中心损失、三重损失等损失函数进行训练并与普通网络相连, 实现跨数据类型的行人重识别. 然而, 当对近 6 年 ICCV、CVPR、AAAI 等顶级国际会议关于一般行人重识别和多源数据行人重识别的论文数量汇总后发现 (图 3(a)), 一般行人重识别问题是当前研究的热点方向, 而针对跨数据类型的行人重识别研究屈指可数, 我们对 5 种多源数据行人重识别方



(a) 论文数量 (b) 最优效果

(a) The number of papers (b) The state-of-the-art performance

图 3 一般行人重识别与多源数据行人重识别论文数量和最优效果对比

Fig. 3 The state-of-the-art performance and number of papers between general Re-ID and multi-source data Re-ID

法取得的最好效果与一般行人重识别的最优效果进行对比发现 (图 3(b)), 由于多源数据行人重识别涉及不同类型的信息, 加之行人本身诸多因素的影响, 导致其特征提取和匹配难度大, 准确率远低于一般行人重识别, 如素描行人重识别目前最高的准确率仅为 34 %, 红外图像行人重识别最高的识别率为 46.4 %, 不到一般行人重识别准确率的一半. 与一般行人重识别问题相比, 多源数据行人重识别在非均匀条件下的研究虽然更加实际, 但也是一个更具挑战性的问题 (表 1).

注 1. 图 3 (b) 分别选取了一般行人重识别和 5 种多源行人重识别在所有行人数据集上识别率最高的结果, 其中一般行人重识别识别率在 Market-1501 数据集上获得, 低分辨率行人重识别识别率在 MLR-VIPeR 数据集上获得, 红外图像行人重识别识别率在 RegDB 数据集上获得, 深度图像行人重识别识别率在 BIWI RGBD-ID 数据集上获得, 文本行人重识别识别率在 iLIDS-VID 数据集上获得, 素描行人重识别识别率在 SKETCH Re-ID 数据集上获得.

2 多源数据行人重识别的算法框架

目前国内外学者已经对多源数据行人重识别进行了一些初步研究, 本节将介绍基于 5 种不同数据类型的行人重识别学习框架和算法.

2.1 基于低分辨率的行人重识别

城市视频监控系统的成本高昂, 通常只在主要街道上布设高分辨率摄像机, 因此现实生活中由公共监控摄像机拍摄的行人图像仍多为低分辨率图像, 且不同摄像机拍摄的图像尺度不一, 导致分辨率不匹配的问题, 对行人重识别工作产生十分不利的影响. 现有的重识别方法要么选择忽略这个问题, 要么直接进行简单的图像缩放或将所有行人图像标准化为统一的尺寸, 使低分辨率图像中的行人信息损失严重, 不能真正有效地解决低分辨率行人重识别问题.

无论是在传统方法还是深度学习方法中, 度量学习在行人重识别中都是一种非常有效的模型匹配技术. 2015 年, Li 等[7] 首次提出了一种针对低分辨率行人重新识别问题的原则性解决方法, 他们设计

表 1 一般行人重识别与多源数据行人重识别的对比

Table 1 Comparison of general Re-ID and multi-source data Re-ID

	一般行人重识别	多源数据行人重识别
定义	给定一个监控行人图像, 检索跨设备下的该行人图像的技术	给定一个监控行人的跨类型或模态信息/图像, 检索跨设备跨模态下的该行人图像的技术
数据类型	单一类型的图像	多类型的图像/视频、文本、语言、素描等数据信息
方法	针对输入图像提取稳定、鲁棒且能描述和区分不同行人的特征信息, 计算特征相似性, 根据相似性大小排序	使用特定于类型/域的网络提取该类型/域的特征信息, 通过共享网络生成特征, 使用合适的损失函数进行训练并与普通网络相连确保重识别工作的有效性
数据集	单一的可见光图像、二分类属性数据集	多种图像、多种信息、多属性数据集
解决重点和难点	低分辨率、视角和姿势变化、光照变化、遮挡和视觉模糊性问题	模态变化以及一般行人重识别需要克服的问题

了一种新的联合多尺度判别框架 JUDEA (Joint multi-scale discriminant component analysis). 该框架的关键组成部分是一种用于低维子空间中跨尺度图像对齐的特征分布差异准则 HCMD (Heterogeneous class mean discrepancy), 最小化这一准则能够统一同一个行人对象在不同分辨率下的相似性判别信息. 通过这种跨尺度的图像统一过程, 可以实现正常分辨率行人图像和低分辨率行人图像判别信息的共享, 将行人图像在多个尺度上进行匹配. Wang 等[8] 发现改变图像尺度距离时可以区分同一个人或不同的人的图像对在不同尺度下产生的尺度距离函数 (Scale-distance function, SDF), 在此基础上提出了一个通过学习计算尺度距离函数进行重识别的多低分辨率行人重识别方法 SALR-REID (Scale-adaptive low resolution person re-identification). 除了度量学习外, 还有一些方法利用字典学习匹配模型. Jing 等[9] 使用一种半耦合低秩判别字典学习方法 SLD^{2}L (Semi-coupled low-rank discriminant dictionary learning) 从一对高−低分辨率图像特征中学习一对字典和一个映射函数, 将低分辨率图像特征转换为高分辨率特征. 为保证经字典学习和映射转换后的特征具有良好的判别能力, 还设计了一个用于半耦合字典学习的判别项, 使得转换得到的高分辨率特征与同一个人在高分辨率图库中的特征更相似而区别于其他人的特征. 此外, 为了保证字典对能更好的描述高分辨率图像和低分辨率图像的潜在特征子空间, 还引入了有监督低秩学习. 在此基础上, Jing 等[10] 还提出了一种多视图 SLD^{2}L 方法, 命名为 MVSLD^{2}L, 通过学习不同类型特征的不同映射, 将低分辨率图像的特征更有效地转化为高分辨率特征. 2018 年, Li 等[11] 提出了一种半耦合投影字典学习框架 DSPDL (Discriminative semi-coupled projective dictionary learning), 采用有效的投影技术, 与字典共同学习映射函数. 通过引入映射函数, 放松相同身份跨视图图像编码之间严格的对应关系, 从而使字典具有更大的泛化能力, 最大限度的提高特征表示能力. 同时, 框架中还设计了一种具有鲁棒性的无参数正则化器, 能显著提高学习字典的判别能力, 从而有效区分正确的行人对和错误的行人对. 该方法与现有最优方法在三个公共数据集上进行比较都具有显著优势, 在 VIPeR 数据集上 DSPDL 比基于特征学习的最佳方法 CMAE[12] 的匹配率高 11%. 比基于度量学习的最佳方法 MLAPG[13] 提高了 6 % 左右, 与现有的基于字典的学习方法相比也有显著提高.

超分辨率技术可以从低分辨率图像中重建出相应的高分辨率图像, 是获取低分辨率图像信息的一种有效手段. 2018 年, Jiao 等[14] 开发了一种超分辨率和行人身份识别联合学习的新方法 SING (Super-resolution and identity joint learning), 该方法设计了一个混合深层卷积神经网络有效连接超分辨率网络和身份识别模型, 通过增强低分辨率图像中有利于身份识别的高频外表信息来提高图像超分辨率和行人重识别的整合容量, 从而解决分辨率不匹配导致的信息量差异问题, 并通过一种混合非对称卷积神经网络联合身份识别损失和超分辨率重构损失函数来优化网络结构.Wang 等[15] 提出了一种统一的级联超分辨率框架 CSRGAN (Cascaded super-resolution generative adversarial network), 将多个 SR-GAN 串联起来对低分辨率图像使用尺度自适应超分辨率技术, 提高了尺度自适应超分辨率模块与身份识别模块的集成兼容性, 从而提高了超分辨率过程中高−低分辨率图像对的相似性. 此外, 还在高−低分辨率图像对内和对间分别创新性引入行人共性损失和行人个性损失, 使生成的高分辨率图像看起来更像人, 同时行人图像更容易被识别.

此外, 一些针对低分辨率行人重识别问题的其他方法也陆续被提出. 2018 年, Wang 等[16] 在基础的残差网络上 (ResNet50)[17] 进行了改进, 该模型在网络的低层 (高分辨率) 和高层 (多语义) 上构建融合嵌入, 着重进行资源约束下的行人重识别, 有效平衡了计算准确性和计算量. 清华大学的 Zhuang 等[18] 提出了一种深度对偶学习框架, 并首次提出了对比中心损失法 (Contrastive center loss, CCL), 可以不受分辨率差异的干扰从不同分辨率图像中学习, 这种框架普适性高, 在此基础上, 普通的行人重识别网络效果也能得到显著提升.

现有的低分辨率行人重识别工作主要基于超分辨率和特征空间投影转换技术, 如何提高识别输入图像有效特征的准确性并尽可能少地引入与行人重识别无关或不利的视觉结果是提高低分辨率行人重识别的关键. 而 SING、CSRGAN 等一系列具有尺度自适应能力的方法能有效利用低分辨率图像中区分行人的高频信息, 联合优化图像超分辨率和重识别问题, 为解决低分辨率行人重识别中不同分辨率图像信息差异问题带来了新思路.

2.2 基于红外数据的行人重识别

在实际生活中, 可视摄像机可能无法拍摄到所有的外观信息, 尤其是在条件较差的室内环境或光

照不足的夜间. 红外摄像机的优势在于它不依赖于人体对可见光的反射, 因此, 在低照度环境下, 红外摄像机拍摄的图像可用于行人的再识别.

早在 2010 年 Kai 等[19] 就提出了一种仅依赖局部图像特征的红外图像行人身份识别方法, 将重识别与行人检测和跟踪方法相结合. 其中用于行人检测的通用外观码本可作为重识别的索引结构, 而在跟踪过程中收集到的局部特征则用于生成行人结构元素, 从而得到有效的匹配模型. Møgelmose 等[20] 提出了一种可见光图像、深度信息和热红外图像数据相结合的行人重识别方法, 该方法使用可见光数据对身体不同区域的颜色信息进行建模, 然后使用深度数据计算人体的软生物特征 (胸廓宽度和关节长度等), 再使用热红外数据提取局部结构信息[21], 最后将三种信息组合构成一个联合分类器, 根据组合规则将首次出现的符合人物作为匹配结果.

Wu 等[22] 通过分析单流网络、双流网络和非对称的全连接网络三种常用的跨模态网络结构对可见光−红外图像行人重识别问题的有效性后, 发现双流结构和全连接结构在特殊情况下都可以用单流网络结构表示, 由此提出了一种采用深度零填充方法的单流网络. 相对于另外两种网络结构来说, 采用深度零填充方法训练的单流网络不但一样具有针对模态的结构和参数, 还具有更强的灵活性, 可以自动学习网络的隐式结构, 对红外行人图像进行有效的重识别. 此外, Wu 还创建了一个红外图像行人重识别的基准数据集 SYSU-MM01. Ye 等[23] 考虑到模态间和模态内的特征变化, 在 2018 年提出了一种具有双向约束高阶损失的双流网络学习可识别特征的表示, 该网络不需要额外的度量学习步骤就可以直接进行端到端的特征学习. 在此基础上, Ye 等[24] 结合特征损失和对比损失对双流网络进行了改进, 提出了一种分层跨模态学习方法 HCML (Hierarchical cross-modality metric learning), 改进后的网络通过学习可见光图像和红外图像两种跨模态不变 (共享) 的特征表示, 可以同时处理跨模态差异和跨视图的变化以及类内的行人模态变化. Dai 等[25] 针对可见光图像和红外图像对同一行人识别信息不足的问题, 设计了一种跨模态生成对抗网络, 该网络包括一个深度卷积神经网络作为学习图像特征表示的生成器和一个模态分类器作为鉴别器, 从两种不同图像中学习身份识别特征的表示. 结合识别损失和跨模态损失最大化类间跨模态相似性的同时最小化类内模糊性, 其工作在 SYSU-MM01 数据集上的累积匹配特征曲线 (Cumulative match characteristic curve, CMC) 和平均精度 (Mean average precision, mAP) 比 Wu 等[22] 的工作分别提高了 12.17% 和 11.85%.

2019 年的 CVPR 中, Wang 等[26] 首次利用 GAN 网络将可见光图像和红外图像两种类型图像合成多光谱图像用于红外条件下的行人重识别, 提出了一种双级差异减少方法 D^2RL (Dual-level discrepancy reduction learning scheme). 该方法由一个图像级差减子网络 φ_M 和一个特征级差减子网络 φ_A 组成, φ_M 通过将不同模态的图像投影至统一的图像空间来最小化模态差异, 并为 φ_A 提供更多可能的图像组合, φ_A 则用于消除外观差异并使 φ_M 生成更可靠的多光谱图像, 二者相辅相成, 以端到端的方式进行联合优化.

随着摄影技术的发展及成本的降低, 红外捕捉装置已经逐渐成为日常监控摄像机的一部分, 结合可见光图片进行行人匹配的潜力巨大, 引起了人们对基于红外图像的行人重识别问题越来越大的兴趣. 目前红外图像行人重识别研究主要使用特征空间投影转换等方法解决跨模态特征匹配的问题, 但由于红外数据跨模态识别的独特之处在于照明类型的变化, 与完全依赖机器学习或基于不变特征提取的方法相比, 基于物理知识的跨模态光度标准化建模或许更有效.

2.3 基于深度图像的行人重识别

相对于可见光图像的视觉特征无法在照明较差的环境下识别的特性, 深度图像与红外图像一样不受光照条件影响, 其特有深度信息在极暗的光照条件下仍然保持不变. 此外, 深度图像还包含了行人的身体形状和骨架特征等信息, 这些信息不受行人的服装变化影响. 因此, 在对不同时间段的同一行人进行匹配时, 基于深度图像的行人重识别效果显著.

2012 年, Barbosa 等[27] 首次基于深度图像进行了行人重识别研究. 为解决重识别过程中行人服装变化的问题, Barbosa 等提取了一组 3D 软生物识别特征代替视觉外观特征, 此外, 他们还收集了一个深度信息行人数据集 PAVIS. 之后, Møgelmose 等[20] 提出了一个结合可见光、深度和热数据的联合分类器, 首次将深度数据与可见光图像、热红外数据三种信息结合起来用于行人重识别领域. Munaro 等[28] 使用基于点云跟踪的自由移动人群重建 3D 模型, 利用 3D 模型实现目标行人匹配, 并收集了一个包含 50 个不同行人生物特征的深度数据

集 BIWI RGBD-ID. Haque 等[29] 提出了一种基于注意力的行人重识别模型, 该模型通过卷积神经网络和递归神经网络组合而成, 在没有 RGB 信息的情况下通过提取人体形状和运动动力学的 4D 时空特征来识别指示行人身份的判别区域, 从而识别不同行人身份. Wu 等[30] 提出了一种局部旋转深度形状不变描述符来描述行人的体型, 然后通过基于核的隐式特征转移将深度特征与 RGB 视觉特征相结合. 2018 年, Hafner 等[31] 受 Gupta 等[32] 监督迁移工作的启发, 提出了一种跨模态蒸馏的迁移学习方法, 首先训练神经网络进行单模态特征识别, 然后利用深度信息与可见光图像的内在关系, 成功将该模态中学习到的特征转移到另一模态, 实现深度信息和 RGB 信息两个模态特征之间的相互转换.

总的来说, 当光照较差或者行人衣着发生改变的情况下, 利用深度信息及其与可见光图像之间的关联能获得更好的识别效果. 然而当拍摄视角发生变化时, 所获得的深度图像中人体形状和骨架信息并不能被有效区分, 且由于深度信息随着行人和相机之间的距离增大而迅速减少, 实际生活中深度相机多用于室内, 而很少布设在室外场景, 因此在实际应用中, 基于深度图像的行人识别问题仍未得到充分研究.

2.4 基于文本数据的行人重识别

大多数视频监控调查系统都依赖于在不同摄像机视域下拍摄的视频图像. 事实上, 在调查过程中, 除了图像/视频资料外, 还有调查人员手工标记的一些注记和来自他人的口头描述的语义信息, 这些标记虽然不完整, 但准确性高, 基于文本数据的行人重识别即利用这些信息进行大型图像数据库搜索匹配, 在行人视频监控应用中有着重要的作用, 而基于属性匹配的行人重识别根据用户对检索对象的描述所获得的属性标签在行人数据库中准确快速地将某个符合描述的目标行人标识出来, 是文本行人重识别的一个重要研究方向.

近年来学者们提出了一系列基于属性匹配的行人重识别方法, 如使用特定的属性配置或支持向量机方法等提取细微属性并从大量的监控视频数据中搜索匹配该配置文件的图片[33−34]. 2015 年, Shi 等[35] 提出了一种可以从具有强/弱注释的数据或混合数据中进行属性学习的框架, 该框架具有强大的自适应能力, 对监督与非监督行人重识别都有十分显著的效果. Wang 等[36] 发现手工标记的注释虽然信息不全但是准确性很高, 对行人重识别工作具有重要意义, 并由此提出一个多步骤的融合算法. 该算法首先利用视觉特征和标记预测完整精确的属性向量, 然后基于统计属性和显著性特征构建一种优势显著性匹配模型用于测量属性向量之间的距离, 最后, 利用视觉特征和属性向量对所有图像进行相似性大小排序. 此外, 也相继提出了一系列基于深度学习的方法, 2017 年, Lin 等[37] 为了提高行人重识别网络的整体精度, 提出了一种属性−行人联合识别网络 APR (Attribute-person recognition), 该网络包含一个身份识别卷积神经网络和一个属性分类模型, 通过身份识别进行属性预测, 同时又集成属性学习用于改进识别网络. Su 等[38] 提出了一种新的半监督深层属性学习算法 SSDAL (Semi-supervised deep attribute learning), 该网络分为三个不同的阶段: 首先在有标签的行人属性识别网络上进行监督训练以得到初始的属性识别网络; 鉴于同一个人的属性识别结果更类似, 故使用三重损失函数提高初始网络的识别能力; 然后用微调后的网络预测一部分无标签数据, 将这部分无标签数据和原始有标签的数据一起用于微调属性识别网络; 最后利用属性之间的差距进行最终的再识别. 该方法不需要对目标数据集进行进一步的训练, 但属性检测鲁棒性仍然很强. 2018 年, Su 等[39] 提出了一种基于低秩属性嵌入的多任务学习方法 MTL-LORAE (Multi-task learning with low rank attribute embedding), 将不同相机间的行人再识别视为相关的多任务, 在多任务学习框架中使用低层可视化特征和中层属性特征作为行人身份特征. 在此基础上还引入了低秩属性嵌入, 利用每对属性间的相关性将原始二值属性映射到连续空间中, 提高特征描述的准确性.

但文本信息和目击者的描述往往是一段自然的句子描述而并不是离散的属性, 基于属性匹配的方法并不能完全适用, 在实际应用中有较大的限制, 因此, Ye 等[40] 提出一种基于对偶的度量学习方法, 通过将不完整的文本描述转换为属性向量, 采用基于线性稀疏重构的方式补全属性向量, 其效果显著, 也是首个真正意义上基于文本信息的行人重识别方法. Li 等[41] 针对文本−视觉匹配问题提出了一种基于身份感知的两阶段框架, 该框架由卷积神经网络 (Convolutional neural networks, CNN) 和长短期记忆 (Long short term memory, LSTM) 两个深度神经网络组成. 第 1 阶段网络引入跨模态交叉熵损失 (Cross-modal cross-entropy loss, CMCE), 利用特征学习中的标识级注释学习表示图像和文本的可识别特征, 在减小交叉熵损失的同时最小化文本特征和图像特征之间的距离, 同时也是第 2 阶段网络

训练的起始点; 第 2 阶段网络通过构建一种具有潜在共注意机制的线性译码器 LSTM 来共同学习潜在空间注意力和语义注意力, 并自动对齐不同的单词和图像区域, 最大限度减少句子结构变化造成的影响. 此外, Li 等[42] 还收集了一个包括不同来源的个人样本和详细的自然语言注释的大型个人资料数据集, 称为 "CUHK 个人资料集" (CUHK-pedes), 并提出了一种基于门控神经注意机制的递归神经网络 GNA-RNN (Recurrent neural network with gated neural attention mechanism), 可以根据查询对象的文本描述利用搜索算法对数据库中所有样本进行排序, 检索与描述最相关的样本.

以往基于文本的行人重识别工作多被看成是行人属性重识别问题, 但由于行人属性的低维度特性导致识别结果往往不如人意. 将基于文本的行人重识别任务当做独立的一类跨模态识别问题后, 通过学习两个模态间共有的具有判别力的子空间能大大缩小模态间的差异性, 可以学习到判别能力更强的特征, 避免了直接预测属性导致匹配误差过大的问题.

2.5 基于素描数据的行人重识别

行人重识别旨在匹配查询照片与图像数据库中的人物. 但现实情况中通常较难取得目标人物的照片, 多数情况下只有专家根据目击证人的描述绘制的人物素描, 人物素描对有关人员的执法行动具有重要的意义. 早期的一些人脸识别研究意识到了这问题并取得了一些成果[43−44], 但与传统的面部素描人脸识别不同的是, 行人重识别领域的素描不仅仅局限于脸部, 而是对全身的素描. 此外, 素描行人重识别与文本行人重识别工作的源数据虽然都包括来自目击证人的描述信息, 但二者仍存在较大的区别: 文本行人重识别更偏向于利用手工注记等语义信息, 这些信息虽然准确, 但往往不够完整和细致; 素描行人重识别则需要素描专家或软件根据更全面细致的描述生成行人素描图像, 视觉上更加直观. 同时, 素描专家和目击证人还可以根据检索到的相似照片交互式地改进素描图像, 进一步提高匹配精度.

素描是一种抽象的描述, 与照片是两个不同的范畴, 加之相机的视角、人的姿势和拍摄范围中的遮挡导致的行人在照片中的不确定性, 利用素描进行行人重识别十分具有挑战性. Pang 等[45] 意识到这一问题并率先进行了研究, 他们提出的素描深度对抗学习框架通过过滤低级特征和保留高级语义特征来共同学习素描和照片中的身份特征、轮廓和纹理等跨模态不变特征, 实现了素描人物图像和一般行人图像的匹配, 并提出了一个包含 200 人的素描–照片跨领域数据集, 弥补了素描行人数据集的缺失. 此外, 该方法在 CUFSF 数据集[44] 和 QMUL-shoe 数据集[46] 两个素描照片数据集上也表现出显著的性能. 但 Pang 的方法丢失了部分有利于进行行人身份判别的模态特有信息, 没有联合优化素描和行人照片特征表达学习的优点. 另外, 由于使用的数据集规模较小, 素描与照片相似性较大, 因此并没有有效反应真实识别情况, 基于素描的行人重识别研究虽然具有重要的现实意义, 但相关研究仍有待重视.

2.6 小结

对目前的多源数据行人重识别工作进行总结后, 我们认为现有的工作主要基于三种方法 (图 4): 1) 基于度量学习. 基于度量学习的多源数据行人重识别工作仅利用一般的度量学习方法来学习如何匹配属于不同类型或模态的行人特征. 2) 基于统一的特征模型. 这类方法侧重于学习两个类型/模态间具有判别力的潜在子空间, 通过将不同类型/模态的信息投影到同一子空间学习更具有判别性的特征模型. 3) 基于统一的模态. 基于统一类型/模态的方法通过各种方式将其中一种类型/模态信息转换成另一种类型/模态信息生成统一模态的特征模型.

从表 2 可以看出, 现有的多源数据行人重识别工作主要基于统一的特征模型和度量学习方法, 而统一模态的方法较少, 且主要集中在跨分辨率的行人重识别工作中. 对于跨文本和素描的行人重识别工作来说, 统一模态的方法目前仍较难实现, 但 Wang 等[26] 和 Hafner 等[31] 的工作首次实现了可见光图像与红外图像、可见光图像与深度图像特征之间的转换, 为将来统一模态的行人重识别工作提供了新的思路.

3 数据集

3.1 数据集介绍

目前已经公布了许多用于行人重识别的数据集, 如常用的 VIPeR[47]、CUHK01[48]、Market-1501[49] 和 iLIDS[50] 等, 但包含多源数据的跨模态行人数据集却很少, 我们总结了一些常用的一般行人重识别数据集和跨模态行人数据集的对比情况如表 3.

1) 低分辨率行人数据集

CAVIAR 数据集[57] 由里斯本一家室内购物中心的两个摄像机记录, 虽然数据集规模较小, 但是这两个摄像机的距离设置刚好一近一远, 它包含 72 个不同行人的 1 220 幅照片, 其中有 50 人同时

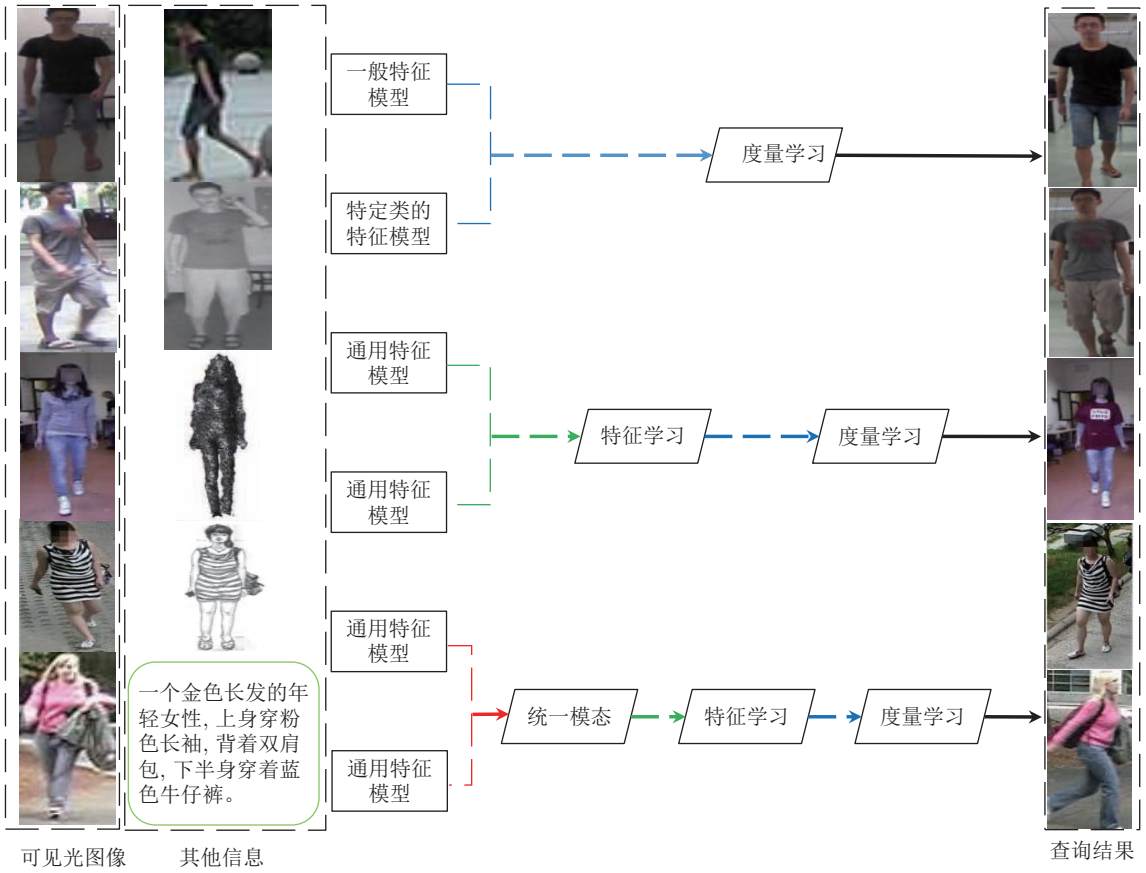


图 4　三类多源数据行人重识别方法描述

Fig. 4　Three types of methods for multi-source data re-ID

出现在两个摄像机中, 另外 22 人只有高分辨率图像 (由距离近的摄像机拍摄), 图像分辨率从 17 像素×39 像素到 72 像素×144 像素不等, 是首个适合低分辨率行人重识别的真实数据集.

2) 红外行人数据集

早在 2013 年, M Φ gelmose 等[20] 就收集了一个小规模红外行人数据集, 该数据集由 35 对 RGB−红外行人图像组成, 每幅图像均为 640 像素×480 像素, 除了包含 RGB 和红外数据之外, 还包含行人的深度数据, 这是行人重识别领域的首个跨三模态的数据集. 2017 年公开的 SYSU-MM01 数据集[22] 最大的特点是包含了由 4 个可见光摄像机和 2 个红外摄像机采集的两种图像, 采集环境也包括室内和室外两种情况, 该数据集包含 491 个不同身份行人的 287 628 幅可见光图像和 15 792 幅红外图像. RegDB 数据集[58] 于 2017 年 3 月份发布, 该数据集使用可见光和红外双摄像机同时拍摄, 不存在时间差. 为了模拟监控系统的正常工作状态, 双摄像机被安放在距地面垂直距离约 6 m 的建筑物顶部, 拍摄所有行人的自然走动状态. RegDB 数据集包括 412 人, 其中女性 254 人, 男性 158 人, 正视图包含 156 人, 后视图包含 256 人. 每人分别对应 10 幅可见光图像和 10 幅红外图像. 由于所有图像都是在人体运动时拍摄的, 所以每个人对应的 10 幅图像中人体姿势、拍摄距离和照明条件都存在差异, 但同一个人的 10 幅图像拍摄时的天气状况、相机视角和拍摄的视图 (前/后视图) 是一致的.

3) 深度图像行人数据集

PAVIS 数据集[27] 由四组不同的数据组成, 第 1 组 “协同小组” 分别记录了 79 个人的正面视图、缓慢行走、避开障碍和伸展手臂视图, 第 2 组 (行走) 和第 3 组 (行走 2) 记录了这 79 个人正常进入工作区域的正面视图和行走画面, 第 4 组 (后视) 记录了他们离开工作区域时的后视图画面. 每个人离摄像机至少 2 m, 由于拍摄时间不同, 因此每个人的服装也不一致, 除此之外, 该数据集还包含利用每个人的骨架和测地信息提取的 10 个软生物特征.

BIWI RGBD-ID 数据集[28] 由 Microsoft Kinect SDK 传感器收集的 50 个不同人员在不同的地点和时间的运动视频序列组成, 其中包括 RGB 图像 (1 280 像素×960 像素)、深度图像、行人分割图、骨骼数据以及地平面坐标. 此外, 还收集了其中

表 2 多源数据行人重识别工作中的代表性方法

Table 2 A summary of representational methods in multi-source data Re-ID

方法	模态	年份	会议/期刊	方法类别	数据集	度量学习	特征模型	统一模态
JUDEA[7]	高−低分辨率图像	2015	ICCV	度量学习	⑩⑪⑫	√	×	×
SLD^{2}L[9]		2015	CVPR	字典学习	⑪⑬⑭	×	√	×
SALR-REID[8]		2016	IJCAI	子空间学习	⑩⑮⑯	√	√	×
SING[14]		2018	AAAI	超分辨率	⑰⑱⑲	×	√	√
CSR-GAN[15]		2018	IJCAI	超分辨率	⑩⑮⑯	×	√	√
DSPDL[11]		2018	AAAI	字典学习	⑪⑭⑳	×	√	×
Zhuang[18]		2018	CVPR	深度对偶学习	㉑㉒㉓	√	×	√
Wu[22]	红外−可见光图像	2017	ICCV	深度零填充	㉔	×	√	×
TONE[24]		2018	AAAI	度量学习	㉕	√	√	×
Ye[23]		2018	IJCAI	特征学习	㉔㉕	√	√	×
cmGAN[25]		2018	IJCAI	特征嵌入	㉔	×	√	×
D^{2}RL[26]		2019	CVPR	图像生成	㉔㉕	×	√	√
Barbosa[27]	深度−可见光图像	2012	ECCV	度量学习	㉖	√	×	×
Wu[30]		2017	TIP	子空间学习	㉖㉗㉘	√	√	×
Hafner[31]		2018	CVPR	模态转移	㉗㉚	×	√	√
Ye[40]	文本−可见光图像	2015	ACM	度量学习	①④㉛	√	×	×
Shi[35]		2015	CVPR	属性识别	①⑤㉛	√	×	×
APR[37]		2017	CVPR	属性识别	⑦⑧	√	×	×
GNA-RNN[42]		2017	CVPR	密切关系学习	㉜	×	√	×
CNN-LSTM[41]		2017	ICCV	特征学习	㉜	×	√	×
MTL-LORAE[39]		2018	PAMI	特征学习	①③④⑨	√	√	×
Pang[45]	素描−可见光图像	2018	ACM MM	特征学习	㉝	×	√	×

28 人的静止和行走序列作为测试集, 这些视频以大约 8~10 帧/s 的速度拍摄, 每个对象的拍摄时间约为 1 min, 在拍摄行走视频中, 每个人面对传感器正面行走两次, 对角行走两次. 由于视频拍摄时间和地点不一致, 因此同一行人的服装也不相同.

由于 Microsoft Kinect SDK 提供的算法不能对人物进行非正面的骨骼估计, 因此 Munaro 等还收集了 IAS-Lab RGBD-ID 数据集[28], 该数据集共包含 11 个不同身份的行人, 由 OpenNI SDK 和 NITE 传感器收集的 33 个序列组成, 分为一个训练集和两个测试集, 其中训练集和测试集 A 中的人员所穿服装不同, 而训练集与测试集 B 中的人员服装一致但拍摄环境不一致.

4) 文本行人数据集

PETA 数据集[34] 由从 CUHK、VIPeR、PRID 等 10 个小规模数据集中挑选的 19 000 幅图像组成, 像素分辨率从 17 像素×39 像素到 169 像素×365 像素不等. 每幅图像都包含 61 个二值属性和 4 个多类属性. 其中二值属性包括人口统计学信息 (如性别和年龄范围)、外观 (如发型)、上下半身服装风格 (如休闲或正式) 和配饰等特征信息, 4 个多类属性中包含 11 种分别用于鞋类、头发、上半身服装和下半身服装基本颜色名[61]. 与一般行人属性数据集相比, PETA 数据集有三个显著的特征: a) 数据集更大. PETA 数据集有 API 和 VIPeR 数据集的 5 倍和 15 倍之大. b) 多样性强. 为了尽可能地使数据集更丰富, Deng 等[34] 特意选择从不同的场景和不同条件下采集的小规模数据集中挑选图像, 因此 PETA 数据集中的图像在照明条件、摄像机视角、图像分辨率、背景复杂性和室内/室外环境等各方面都具有很大的差异. c) 丰富的注释. 与现有的数据集相比, PETA 数据集包含更丰富的注释, 如 VIPeR, 只有 15 个二值属性, API[62] 有 11 个二值属性和 2 个多类属性, 而 PETA 数据集包括 61 个二值属性和 4 个多类属性, 特别是这些二值属性中还包括英国内政部和英国警方建议的在跟踪和刑事鉴定方面最有价值的 15 个属性. 另一个使用自然语言描述行人外观的大型语言数据集是 CUHK-PEDES[42], 包括 13 003 人的40 206 幅图像, 每幅图像由两名不相干的 AMT 工人用两句话进行描述, 其中包含丰富的词汇、短语、句子模式和结构, 该数据集共有 1 893 118 个单词, 其中包含 9 408

表 3 常用的一般行人重识别数据集与跨模态行人重识别数据集

Table 3 A summary of general Re-ID dataset and multi-source data Re-ID datase

类别	数据集名称	发布时间	数据集类型	人数	相机数量	数据集大小
一般行人数据集	①VIPeR[51]	2008	真实数据集	632	2	1 264幅 RGB 图像
	②3DPES[52]	2011		192	8	1 011 幅 RGB 图像
	③i-LIDS[50]	2009		119	2	476 幅 RGB 图像
	④PRID2011[53]	2011		934	2	1 134 幅 RGB 图像
	⑤CUHK01[48]	2012		971	2	3 884幅 RGB 图像
	⑥CUHK03[6]	2014		1 467	10	13 164幅 RGB 图像
	⑦Market-1501[54]	2015		1 501	6	32 217 幅 RGB 图像
	⑧DukeMT MC-REID[55]	2017		1 812	8	36 441 幅 RGB 图像
	⑨SAIVT-SoftBio[56]	2012		152	8	64 472 幅 RGB 图像
低分辨率行人数据集	⑩CAVIAR[57]	2011	真实数据集	72	2	720 幅高分辨率图像 500 幅低分辨率图像
	⑪LR-VIPeR[7, 9-11]	2015		632	2	1 264 幅 RGB 图像
	⑫LR-3DPES[7]	2015		192	8	1 011 幅 RGB 图像
	⑬LR-PRID2011[9, 15]	2015		100	2	200 幅 RGB 图像
	⑭LR-i-LDIS[9, 11]	2015		119	2	238 幅 RGB 图像
	⑮SALR-VIPeR[8, 15]	2016		632	2	1 264 幅 RGB 图像
	⑯SALR-PRID[8, 15]	2016		450	2	900 幅 RGB 图像
	⑰MLR-VIPeR[14]	2018	模拟数据集	632	2	1 264 幅 RGB 图像
	⑱MLR-SYSU[14]	2018		502	2	3 012 幅 RGB 图像
	⑲MLR-CUHK03[14]	2018		1 467	2	14 000 幅 RGB 图像
	⑳LR-CUHK01[11]	2018		971	2	1 942 幅 RGB 图像
	㉑LR-CUHK03[18]	2018		1 467	10	13 164 幅 RGB 图像
	㉒LR-Market-1501[18]	2018		1 501	6	32 217 幅 RGB 图像
	㉓LR-DukeMTMC-REID[18]	2018		1 812	8	36 441 幅 RGB 图像
红外行人数据集	㉔SYSU-MM01[22]	2017	真实数据集	491	6	287 628 幅 RGB 图像 15 792幅红外图像
	㉕RegDB[58]	2017		412	2	4 120 幅 RGB 图像 4 120 幅红外图像
深度图像行人数据集	㉖PAVIS[27]	2012	真实数据集	79	–	316 组视频序列
	㉗BIWI RGBD-ID[28]	2014		50	–	22 038 幅 RGB-D 图像
	㉘IAS-Lab RGBD-ID[28]	2014		11	–	33 个视频序列
	㉙Kinect REID[59]	2016		71	–	483 个视频序列
	㉚RobotPKU RGBD-ID[60]	2017		90	–	16 512 幅 RGB-D 图像
文本行人数据集	㉛PETA[34]	2014	真实数据集	8 705	–	19 000 幅图像 66 类文字标签
	㉜CUHK-PEDES[42]	2017		13 003	–	40 206 幅图像 80 412 个句子描述
素描行人数据集	㉝Sketch Re-ID[45]	2018	真实数据集	200	2	400 幅 RGB 图像 200 幅素描

个特有单词. 每个句子最少有 15 个单词, 最长的句子有 96 个单词, 平均单词长度为 23.5, 大多数句子有 20 到 40 个单词.

5) 素描行人数据集

Pang 等[45] 在 2018 年收集了一个素描行人数据集 Sketch Re-ID, 该数据集包含 200 人, 每个人对应 2 幅照片和 1 幅素描, 这两幅照片由两个交叉视域相机在白天拍摄, 而所有素描图像则由 5 位风格各不相同的专家共同绘出. 此外, 所有的照片和素描都有与之对应的 ID 标签, 同一个人的照片和

素描图像其 ID 是一致的, 这也是目前为止首个用于行人重识别的素描数据集. 表 4 列出了多源数据行人重识别问题中几种具有代表性的方法在常用的一般行人数据集和跨模态行人数据集上的识别结果.

3.2 小结

由于低分辨率行人数据集较少且规模较小, 因此大部分低分辨率行人重识别工作的做法仍然是使用 VIPeR、CAVIAR、CUHK-01 等基准行人数据集或模拟数据集, 这在很大程度上限制了跨分辨率行人重识别的发展. 而素描数据集 Sketch Re-ID 中的素描图像由专业人员严格按照行人照片进行描绘, 素描图像与真实照片相似度大, 直接消除了现实场景中由目击者口头描述带来的噪声和错误信息, 这与实际情况并不符合, 研究结果与实际应用效果有较大的出入. 总的来说, 目前用于多源数据行人重识别的跨模态数据集较少, 这些数据集多数只有几千幅甚至几百幅图片, 严重阻碍了多源数据行人重识别工作的发展, 构建大规模且贴合实际的真实跨模态多源数据集是当前多源数据行人重识别研究最重要的工作之一.

4 总结与展望

行人重识别的目标是实现对行人的跨视域定位和追踪, 是当今计算机视觉领域的关键技术, 具有重大的理论意义和应用前景, 而多源数据行人重识别更是行人重识别技术需要攻克的核心和难点问题, 具有重要的现实意义. 虽然目前关于行人重识别的研究层出不穷, 也取得了一定的研究成果, 但针对多源数据的跨模态和跨类型行人重识别研究尚处于初步探索阶段. 在对目前主要的多源数据行人重识别工作与数据集进行比较和总结后, 我们认为当前的多源数据行人重识别研究还需要着重解决以下问题:

1) 真实跨模态多源行人数据集较少, 规模较小. 虽然多源数据行人重识别能够更充分利用各类有效信息, 但由于跨模态和类型的非线性映射比低维空间的简单映射更加复杂, 因此深度学习过程中所需

表 4 几种多源数据行人重识别方法在常用的行人数据集上的识别结果

Table 4 Comparison of state-of-the-art methods on infra-red person re-identification dataset

	数据集	算法	年份	Rank1 (%)	Rank5 (%)	Rank10 (%)
低分辨率	VIPeR	SLD^{2}L[9]	2015	16.86	41.22	58.06
		MVSLD^{2}L[10]	2017	20.79	45.08	61.24
		DSPDL[11]	2018	28.51	61.08	76.11
	CAVIAR	JUDEA[7]	2015	22.12	59.56	80.48
		SLD^{2}L[9]	2015	18.40	44.80	61.20
		SING[14]	2018	33.50	72.70	89
红外	SYSU-MM01	Wu等[22]	2017	24.43	–	75.86
		Ye等[23]	2018	17.01	–	55.43
		CMGAN[25]	2018	37.00	–	80.94
	RegDB	Ye等[23]	2018	33.47	–	58.42
		TONE[24]	2018	16.87	–	34.03
深度图像	BIWI RGBD-ID	Wu等[30]	2017	39.38	72.13	–
		Hafner[31]	2018	36.29	77.77	94.44
	PAVIS	Wu等[30]	2017	71.74	88.46	–
		Ren等[63]	2017	76.70	87.50	96.10
素描	SKETCH Re-ID	Pang等[45]	2018	34	56.30	72.50
文本	VIPeR	Shi等[35]	2015	41.60	71.90	86.20
		SSDAL[38]	2016	43.50	71.80	81.50
		MTL-LORAE[39]	2018	42.30	42.30	81.6
	PRID	SSDAL[38]	2016	22.60	48.70	57.80
		MTL-LORAE[39]	2018	18	37.40	50.10
					Top1	Top10
文本	CUHK-PEDES	CNN-LSTM[41]	2017		25.94	60.48
		GNA-RNN[42]	2017		19.05	53.64

要的训练数据也更多, 但当前存在的跨模态多源数据集屈指可数, 这些数据集多数只有几千幅或几百幅图片, 可选择的余地非常有限, 很多低分辨率行人重识别工作甚至直接使用一般的数据集或模拟数据集, 这些问题大大限制了多源数据行人匹配的效果. 在今后的发展中, 收集规模更大属性更全的真实跨模态和跨类型多源行人数据集是研究者们亟需解决的重点问题.

2) 筛选有效的数据信息. 由于多源数据行人重识别的特殊性, 在行人匹配过程中提供了比一般行人重识别更多类型的数据和信息, 与此同时带来的信息冗余情况也更严重. 目前的多源数据行人重识别研究中, 如跨文本、素描等模态时仍有很多工作需要人工参与, 但人工参与过程是带有主观意识的, 针对同一任务不同的人可能会得到不同的信息, 这些信息通常又是不全面的, 因此, 在利用深度学习网络进行多源数据行人重识别工作时, 如何针对特定的数据类型设计并选择合适的网络过滤无效信息, 挖掘整合有效信息变得十分重要.

3) 基于统一模态的研究. 现有的多源数据行人重识别工作主要基于统一的特征模型和度量学习方法. 在低分辨率行人重识别中 Jiao 等[14–15, 18] 的工作使用超分辨率技术成功将低分辨率图像转换成高分辨率图像, 成功实现了跨分辨率行人重识别的模态统一, 但对于跨文本和素描的行人重识别问题来说, 目前还没有基于统一模态方法的研究成果, 而 Wang 等[26] 的工作通过使用 Cycle-GAN 首次实现了可见光图像与红外图像的模态统一, Hafner 等[31] 提出的跨模态蒸馏的迁移学习方法成功实现了深度信息和 RGB 信息两种模态特征之间的相互转换, 这些都为将来基于统一模态的多源数据行人重识别研究提供了新的思路.

4) 集成跨多类型数据行人重识别工作. 目前的多源数据行人重识别研究主要针对跨两种或三种类型和模态的行人匹配问题, 但事实上整合多类型和多模态信息进行特征提取不仅可以获得更多有效的身份识别信息, 而且更贴合实际应用情况. 因此, 在同一行人重识别过程中使用多种数据和信息进行行人匹配将是未来多源数据行人重识别研究的一个重要方向.

本文首先分别介绍了一般行人重识别和多源数据行人重识别方法及其区别, 然后总结了基于低分辨率、红外图像、深度图像、文本以及素描的 5 种不同类型数据行人重识别方法和数据集情况, 并分析和展望了当前多源数据行人重识别技术面临的挑战和未来的发展方向, 可以看出, 多源数据行人重识别具有重要的现实意义和巨大的发展空间.

References

1 Song Wan-Ru, Zhao Qing-Qing, Chen Chang-Hong, Gan Zong-Liang, Liu Feng. Survey on pedestrian re-identification research. *CAAI Transactions on Intelligent Systems*, 2017, **12**(6): 770−780
(宋婉茹, 赵晴晴, 陈昌红, 干宗良, 刘峰. 行人重识别研究综述. 智能系统学报, 2017, **12**(6): 770−780)

2 Li You-Jiao, Zhuo Li, Zhang Jing, Li Jia-Feng, Zhang Hui. A survey of person re-identification. *Acta Automatica Sinica*, 2018, **44**(9): 1554−1568
(李幼蛟, 卓力, 张菁, 李嘉锋, 张辉. 行人再识别技术综述. 自动化学报, 2018, **44**(9): 1554−1568)

3 Zheng Wei-Shi, Wu An-Cong. Asymmetric person re-identification: cross-view person tracking in a large camera network. *Scientia Sinica Informationis*, 2018, **48**(5): 545−563
(郑伟诗, 吴岸聪. 非对称行人重识别: 跨摄像机持续行人追踪. 中国科学: 信息科学, 2018, **48**(5): 545−563)

4 Wang Zheng. Person Re-identification in Complicated Conditions [Ph.D. dissertation], Wuhan University, China, 2017.
(王正. 条件复杂化行人重识别关键技术研究 [博士学位论文]. 武汉大学, 中国, 2017.)

5 Zhu X, Jing X Y, You X, Zuo W, Shan S, Zheng W S. Image to video person re-identification by learning heterogeneous dictionary pair with feature projection matrix. *IEEE Transactions on Information Forensics and Security*, 2018, **13**(3): 717−732

6 Li W, Zhao R, Xiao T, Wang X G. DeepReID: deep filter pairing neural network for person re-identification. In: Proceedings of the 27th IEEE International Conference of Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014. 152−159

7 Li X, Zheng W, Wang X, Xiang T, Gong S. Multi-scale learning for low-resolution person re-identification. In: Proceedings of the 28th IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 3765−3773

8 Wang Z, Hu R M, Yu Y, Jiang J J, Chao L, Wang J Q. Scale-adaptive low-resolution person re-identification via learning a discriminating surface. In: Proceedings of the 2016 International Joint Conference on Artificial Intelligence. New York, USA, 2016. 2669−2675

9 Jing X Y, Zhu X K, Wu F, You X G, Liu Q L, Yue D, et al. Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning. In: Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015. 695−704

10 Jing X Y, Zhu X K, Wu F, Hu R M, You X G, Wang Y H, et al. Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning. *IEEE Transactions Image Process*, 2017, **26**(3): 1363−1378

11 Li K, Ding Z M, Li S, Fu Y. Discriminative semi-coupled projective dictionary learning for low-resolution person re-identification. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence. Louisiana, USA: IEEE, 2018. 2331−2338

12 Wang S Y, Ding Z M, Fu Y. Coupled marginalized auto-encoders for cross-domain multi-view learning. In: Proceedings of the 2016 International Joint Conference on Artificial Intelligence. New York, USA, 2016. 2125−2131

13 Liao S C, Li S Z. Efficient psd constrained asymmetric metric learning for person re-identification. In: Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 3685−3693

14 Jiao J N, Zheng W S, Wu A C, Zhu X T, Gong S G. Deep low-resolution person re-identification. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence. Louisiana, USA: IEEE, 2018. 6967−6974

15 Wang Z, Ye M, Yang F, Bai X, Satoh S I. Cascaded SR-GAN for scale-adaptive low resolution person re-identification. In: Proceedings of the 2018 International Joint Conferences on Artificial Intelligence. Stockholm, Sweden, 2018. 3891−3897

16 Wang Y, Wang L Q, You Y R, Zou X, Chen V, Li S, et al. Resource aware person re-identification across multiple resolutions. In: Proceedings of the 31st IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 1−10

17 He K M, Zhang X Y, Ren S Q, Jian S. Deep residual learning for image recognition. In: Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 770−778

18 Zhuang Z J, Ai H Z, Chen L, Shang C. Cross-resolution person re-identification with deep antithetical learnin. In: Proceedings of the 31st IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 1−16

19 Kai J L, Arens M. Local feature based person reidentification in infrared image sequences. In: Proceedings of the 7th IEEE International Conference on Advanced Video and Signal Based Surveillance. Boston, USA: IEEE, 2010. 448−455

20 Møgelmose A, Bahnsen C, Moeslund T B, Clapes A, Escalera S. Tri-modal person re-identification with RGB, depth and thermal features. In: Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition Workshops. Portland, USA: IEEE, 2013. 301−307

21 Bay H, Ess A, Tuytelaars T, Gool L V. Speeded-up robust features. *Computer Vision and Image Understanding*, 2008, **110**(3): 346−359

22 Wu A C, Zheng W S, Yu H X, Gong S G, Lai J H. RGB-infrared cross-modality person re-identification. In: Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 5390−5399

23 Ye M, Wang Z, Lan X Y, Yuen P C. Visible thermal person re-identification via dual-constrained top-ranking. In: Proceedings of the 2018 International Joint Conferences on Artificial Intelligence. Stockholm, Sweden, 2018. 1092−1099

24 Ye M, Lan X Y, Li J W, Yuen P C. Hierarchical discriminative learning for visible thermal person re-identification. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence. Louisiana, USA: AAAI, 2018. 7501−7508

25 Dai P Y, Ji R R, Wang H B, Wu Q, Huang Y Y. Cross-modality person re-identification with generative adversarial training. In: Proceedings of the 2018 International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 2018. 677−683

26 Wang Z X, Wang Z, Zheng Y Q, Chuang Y-Y, Satoh S I. Learning to reduce dual-level discrepancy for infrared-visible person re-identification. In: Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, California, USA: IEEE, 2019. 618−626

27 Barbosa I B, Cristani M, Bue A D, Bazzani L, Murino V. Re-identification with RGB-D sensors. In: Proceedings of the 12th International Conference on Computer Vision. Florence, Italy: ECCV, 2012. 433−442

28 Matteo M, Alberto B, Andrea F, Luc V G, Menegatti E. 3D reconstruction of freely moving persons for reidentification with a depth sensor. In: Proceedings of the 2014 IEEE International Conference on Robotics and Automation. Hong Kong, China: IEEE, 2014. 4512−4519

29 Haque A, Alahi A, Li F F. Recurrent attention models for depth-based person identification. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 1229−1238

30 Wu A C, Zheng W S, Lai J H. Robust depth-based person re-identification. *IEEE Transactions on Image Processing*, 2017: 2588−2603

31 Hafner F, Bhuiyan A, Kooij J F P, Granger E. A cross-modal distillation network for person re-identification in rgb-depth. In: Proceedings of the 31st IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 1−18

32 Gupta S, Hoffman J, Malik J. Cross modal distillation for supervision transfer. In: Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 2827−2836

33 Jason T, Jeanette B G, Daniel B, Michael C, Heather Z. Person attribute search for large-area video surveillance. In: Proceedings of the 2012 IEEE International Conference on Technologies for Homeland Security. Boston, USA: IEEE, 2012. 55−61

34 Deng Y B, Luo P, Loy C C, Tang X O. Pedestrian attribute recognition at far distance. In: Proceedings of the 22nd ACM International Conference on Multimedia. Orlando, USA: ACM MM, 2014. 789−792

35 Shi Z Y, Hospedales T M, Xiang T. Transferring a semantic representation for person re-identification and search. In: Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA: IEEE, 2015. 4184−4193

36 Wang Z, Hu R M, Yu Y, Liang C, Huang W X. Multi-level fusion for person re-identification with incomplete marks. In: Proceedings of the 23rd ACM International Conference on Multimedia. Brisbane, Australia: ACM MM, 2015. 1267−1270

37 Lin Y T, Liang Z, Zheng Z D, Yu W, Yi Y. Improving person re-identification by attribute and identity learning. In: Proceed-

ings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA: IEEE, 2017: 1−10

38 Su C, Zhang S L, Xing J L, Wen G, Qi T. Deep attributes driven multi-camera person re-identification. In: Proceedings of the 2016 European Conference on Computer Vision. Amsterdam, the Netherlands, 2016. 475−491

39 Su C, Yang F, Zhang S L, Tian Q, Davis L S, Gao W. Multi-task learning with low rank attribute embedding for multi-camera person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, **40**(5): 1167−1181

40 Ye M, Liang C, Wang Z, Leng Q M, Chen J, Liu J. Specific person retrieval via incomplete text description. In: Proceedings of the 5th ACM on International Conference on Multimedia Retrieval. Shanghai, China: ACM, 2015. 547−550

41 Li S, Xiao T, Li H S, Yang W, Wang X G. Identity-aware textual-visual matching with latent co-attention. In: Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 1890−1899

42 Li S, Xiao T, Li H S, Zhou B L, Yue D Y, Wang X G. Person search with natural language description. In: Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA: IEEE, 2017. 5187−5196

43 Galoogahi H K, Sim T. Face photo retrieval by sketch example. In: Proceedings of the 20th ACM International Conference on Multimedia. Nara, Japan: ACM, 2012. 949−952

44 Zhang W, Wang X G, Tang X O. Coupled information-theoretic encoding for face photo-sketch recognition. In: Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA: IEEE, 2011. 513−520

45 Pang L, Wang Y W, Song Y Z, Huang T J, Tian Y H. Cross-domain adversarial feature learning for sketch re-identification. In: Proceedings of the 2018 ACM Multimedia Conference on Multimedia Conferenc. Seoul, Korea: ACM, 2018. 609−617

46 Yu Q, Liu F, Song Y Z, Xiang T, Hospedales T M, Chen C L. Sketch me that shoe. In: Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 799−807

47 Gray D, Brennan S, Tao H. Evaluating appearance models for recognition, reacquisition, and tracking. In: Proceedings of the 10th International Workshop on Performance Evaluation for Tracking and Surveillance. Rio de Janeiro, Brazil: IEEE, 2007. 1−7

48 Li W, Zhao R, Wang X G. Human reidentification with transferred metric learning. In: Proceedings of the 2012 Asian Conference on Computer Vision. Daejeon, Korea, 2012. 31−44

49 Roth P M, Martin H, Köstinger M, Beleznai C, Bischof H. Mahalanobis distance learning for person re-identification. *Person Re-Identification*, 2014: 247−267

50 Zheng W S, Gong S G, Tao X. Associating groups of people. In: Proceedings of the 2009 British Machine Vision Conference. London, UK, 2009: 1−11

51 Gray D, Hai T. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Proceedings of the 10th European Conference on Computer Vision. Marseille, France, 2008. 262−275

52 Baltieri D, Vezzani R, Cucchiara R. 3Dpes: 3D people dataset for surveillance and forensics. In: Proceedings of the 2011 ACM Joint ACM Workshop on Human Gesture and Behavior Understanding. Scottsdale, USA: ACM, 2011. 59−64

53 Hirzer M, Beleznai C, Roth P M, Bischof H. Person re-identification by descriptive and discriminative classification. In: Proceedings of the 2011 Scandinavian Conference on Image Analysis. Ystad, Sweden, 2011. 91−102

54 Zheng L, Shen L Y, Tian L, Wang S J, Wang J D, Tian Q. Scalable person re-identification: A benchmark. In: Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 2380−7504

55 Zheng Z D, Zheng L, Yang Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: Proceedings of the 2017 IEEE International Conference on Computer Vision. Honolulu, USA: IEEE, 2017. 3774−3782

56 Bialkowski A, Denman S, Sridharan S, Fookes C, Lucey P. A database for person re-identification in multi-camera surveillance networks. In: Proceedings of the 2012 International Conference on Digital Image Computing Techniques and Applications. Fremantle, Australia, 2012. 1−8

57 Dong S C, Cristani M, Stoppa M, Bazzani L, Murino V. Custom pictorial structures for re-identification. In: Proceedings of the 2011 British Machine Vision Conference. Dundee, Scotland, 2011. 1−11

58 Nguyen D T, Hong H G, Kim K W, Park. K R. Person recognition system based on a combination of body images from visible light and thermal cameras. *Sensors*, 2017, **17**(3): 605−633

59 Pala F, Satta R, Fumera G, Roli F. Multimodal person reidentification using RGB-D cameras. *IEEE Transactions on Circuits and Systems for Video Technology*, 2016, **26**(4): 788−799

60 Hong L, Liang H, Ma L Q. Online RGB-D person re-identification based on metric model update. *CAAI Transactions on Intelligence Technology*, 2017, **2**(1): 48−55

61 Joost V D W, Cordelia S, Jakob V, Diane L. Learning color names for real-world applications. *IEEE Transactions on Image Processing*, 2009, **18**(7): 1512−1523

62 Zhu J Q, Liao S C, Lei Z, Yi D, Li S. Pedestrian attribute classification in surveillance: Database and evaluation. In: Proceedings of the 2013 IEEE International Conference on Computer Vision Workshops. Sydney, Australia: IEEE, 2013. 331−338

63 Ren L L, Lu J W, Feng J J, Zhou J. Multi-modal uniform deep learning for RGB-D person re-identification. *Pattern Recognition*, 2017, **72**: 446−457

叶 钰 武汉大学计算机学院国家多媒体软件工程技术研究中心博士研究生. 主要研究方向为图像处理, 计算机视觉.
E-mail: ms.yeyu@whu.edu.cn
(**YE Yu** Ph.D. candidate at the National Engineering Research Center for Multimedia Software, School of Computer Science, Wuhan University. Her research interest covers person re-identification and computer vision.)

王 正 日本国立信息学研究所学术振兴会外国人特别研究员. 2017 年获得武汉大学计算机学院国家多媒体软件工程技术研究中心博士学位. 主要研究方向为行人重识别和实例搜索. 本文通信作者.
E-mail: wangz@nii.ac.jp
(**WANG Zheng** JSPS Fellowship Researcher at National Institute of Informatics, Japan. He received his Ph.D. degree from the National Engineering Research Center for Multimedia Software, School of Computer Science, Wuhan University in 2017. His research interest covers person re-identification and instance search. Corresponding author of this paper.)

梁 超 武汉大学副教授. 2012 年获得中国科学院自动化研究所博士学位. 主要研究方向为多媒体内容分析和检索, 计算机视觉和模式识别.
E-mail: cliang@whu.edu.cn
(**LIANG Chao** Associate professor at Wuhan University. He received his Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences in 2012. His research interest covers multimedia content analysis and retrieval, computer vision, and pattern recognition.)

韩 镇 武汉大学副教授. 2009 年获得武汉大学博士学位. 主要研究方向为图像/视频压缩与处理, 计算机视觉和人工智能.
E-mail: hanzhen_2003@hotmail.com
(**HAN Zhen** Associate professor at Wuhan University. He received his Ph.D. degree in computer application technology from Wuhan University in 2009. His research interest covers image and video compressing and processing, computer vision, and artificial intelligence.)

陈 军 武汉大学教授. 主要研究方向为多媒体分析, 计算机视觉和安防应急信息处理.
E-mail: chenj@whu.edu.cn
(**CHEN Jun** Professor at Wuhan University. His research interest covers multimedia analysis, computer vision, and security emergency information processing.)

胡瑞敏 武汉大学教授. 主要研究方向为多媒体技术与大数据分析, 多媒体信号处理, 音视频处理, 模式识别, 人工智能.
E-mail: hrm1964@163.com
(**HU Rui-Min** Professor at Wuhan University. His research interest covers multimedia technology and big data analysis, multimedia signal processing, audio and video processing, pattern recognition, and artificial intelligence.)