

联合多级深度特征表示和有序加权距离融合的视频行人再识别方法

孙锐^{1,2}, 黄启恒^{1,2*}, 陆伟明^{1,2}, 高隽¹

¹合肥工业大学计算机与信息学院, 安徽 合肥 230009;

²工业安全与应急技术安徽省重点实验室, 安徽 合肥 230009

摘要 针对目前视频行人再识别中存在视角、光线变化, 背景干扰与遮挡, 行人外观与行为相似, 以及相同行人在不同模态特征下距离的差异性而导致的匹配不正确问题, 提出一种联合多级深度特征表示和有序加权距离融合的视频行人再识别方法。在行人特征表示阶段, 提出了行人多级深度特征表示网络, 该网络不仅能学习视频序列中行人的时空特征, 还能获取行人的全局外观特征和局部外观特征。在有序加权距离融合阶段, 将行人的特征表示输入到距离测度学习中, 分别计算行人在三类特征下的独立距离, 并将距离排序后, 根据距离的排名优化距离权重, 最后融合三类距离得到最终距离, 从而准确匹配行人。通过在公共数据集中的实验表明, 所提方法不仅能够提高视频行人再识别的识别率, 还具有丰富和完整的行人特征表示能力。

关键词 机器视觉; 视频行人再识别; 多级深度特征; 距离融合; 卷积神经网络; 循环神经网络; 有序加权

中图分类号 TP391.41

文献标识码 A

doi: 10.3788/AOS201939.0915006

Video-Based Person Re-Identification via Combined Multi-Level Deep Feature Representation and Ordered Weighted Distance Fusion

Sun Rui^{1,2}, Huang Qiheng^{1,2*}, Lu Weiming^{1,2}, Gao Jun¹

¹ School of Computer and Information, Hefei University of Technology,

Hefei, Anhui 230009, China;

² Anhui Provincial Key Laboratory of Industry Safety and Emergency Technology, Hefei, Anhui 230009, China

Abstract Video-based person re-identification problems are caused by perspective changes, lighting variations, background clutter, occlusion, appearance similarity, motion similarity, and mismatch resulting from the distance difference of same person with different modal features. This study proposes a video-based person re-identification method that combines multi-level deep feature representation and ordered weighted distance fusion. During the stage of person feature representation, the multi-level deep feature representation network proposed herein not only learns the space-time features of the persons in video sequences but also acquires the persons' global and local appearance features. In the stage of the ordered weighted distance fusion, the feature representations of persons are firstly input into distance metric learning, and the independent distances of persons under three types of features are calculated. The fusion algorithm then sorts the distances to optimize distance weights according to distance ranking. Finally, to accurately match a person, the algorithm fuses the three types of distances to obtain the final distance. Experimental results compared with the results of related methods in public datasets show that the proposed method not only improves the recognition rate of video-based person re-identification but also possesses abundant and integral ability for person feature representation.

Key words machine vision; video-based person re-identification; multi-level deep feature; distance fusion; convolutional neural network; recurrent neural network; ordered weighted

OCIS codes 150.1135; 100.4995; 100.4999; 100.4996

收稿日期: 2019-01-14; **修回日期:** 2019-04-12; **录用日期:** 2019-05-31

基金项目: 国家自然科学基金面上项目(61471154)、安徽省科技攻关强警项目(1704d0802181)、中央高校基本科研业务费专项资金资助项目(JZ2018YYPY0287)

* **E-mail:** jchqh123@163.com

1 引 言

行人再识别研究的是在非重叠视域的不同摄像机下识别匹配同一行人的问题,部分文献中也称其为行人检索^[1]。近些年来,行人再识别逐渐成为计算机视觉领域的研究热点。目前,行人再识别的研究主要集中于利用单帧静止图像获取行人外观特征,解决多摄像机视角下的单视行人匹配的问题,但单帧静止图像中可能存在的背景干扰、遮挡或外观相似及变化的问题易使行人特征表示受限。因此,本研究利用视频序列中隐含的时间运动关联信息和多样的外观信息来改善行人特征表示的能力。

虽然视频行人再识别相比于静止图像行人再识别在特征表示方面具有特有的优势,但是视频行人再识别也面临着许多挑战。首先,不同行人如果具有相似外观,便会出现行人不同但特征表示相似的问题。其次,虽然借助视频序列中行人运动关联的时空特征可改善性能,但也可能面临不同行人之间运动姿态相似的困难。最后,由于不同模态的外观特征与时空特征的简单级联可能会降低某些区分性特征的重要性,因此相同行人在不同模态特征下的距离有明显差异,特征的简单融合会使行人再识别准确性产生误差。因此,在本研究的算法中,如何构建行人完整丰富的特征表示能力和如何处理不同模态特征之间的相互影响成为算法的关键。一方面,针对视频行人再识别中不同行人外观相似和运动行为相似的问题,已有研究者提出采用卷积神经网络(CNN)和循环神经网络(RNN)相结合的方法来获取时空特征^[2-3]。这类算法主要是将CNN获取的空间外观特征向量作为RNN的输入,从而获得行人的时空特征表示。虽然这类研究工作有效地提高了视频行人再识别的识别率,但是其算法级联式的特征网络未充分获取行人的外观特征信息。当不同行人之间衣物和行为相似时,准确匹配相同行人仍然是一项难题。另一方面,现有的视频行人再识别研究工作采用的策略是将各类特征简单级联,从而忽略不同模态的特征融合所产生信息丢失的影响^[4]。同时,相同行人在不同特征下距离的差异性也会导致匹配不准确。因此,为了解决上述困难,本文采用多级深度特征表示的思想^[5]学习行人丰富的全局、局部外观特征和稳定的时空特征。同时,受到文献^[6-7]的启发,本文设计一种有序加权的距离测度融合算法,对行人的全局外观特征、局部外观特征和时空特征分别进行独立距离测度学习,再计算行

人之间在三类特征下的独立距离。最后根据三类距离值的排序,优化距离融合时的权值,从而达到最佳的距离学习效果,并有效减缓相同行人在不同特征下由距离差异导致的匹配不准确问题。将本文算法应用于i-LIDS和PRID_2011两个视频数据集中,多次实验的结果表明,本文联合多级深度特征表示和有序加权距离融合的视频行人再识别算法提高了行人识别的正确率。

2 行人再识别

2.1 行人再识别研究现状

目前,多数行人再识别算法所采用的策略可分为两部分:提取区分性特征信息^[8-10]和学习稳定的距离测度^[11-14]。

在行人再识别领域的研究中,获取行人的特征信息主要是提取行人图像或视频序列中的外观特征,利用该特征信息匹配行人,如图1所示。本研究认为对于行人再识别中可利用的行人特征信息可从以下两个角度分类。从特征结构的角度,可分为低层特征和中高层特征,低层特征包括颜色^[15]、纹理^[16]、轮廓^[17]等,而在近几年获取图像的中高层特征信息成为特征提取的热点,中高层特征主要指语义特征^[18]。从特征信息的角度,可分为可视特征和隐含特征,可视特征在行人图像中指的是行人的外观特征信息和部分生物行为特征信息,而隐含特征包括视频序列中的时空特征、行人图像中的语义特征和部分隐含的生物特征等。同时,行人再识别中特征提取的方式也可以分类为手动设计和自动学习。手动设计主要用于提取行人的颜色、纹理等多种特征,而自动学习提取特征主要是应用深度学习技术自动获得行人的外观特征^[19-20]。

目前,可以将行人再识别的距离测度学习部分分为学习特征距离矩阵和映射特征子空间的两类测度学习方法。测度学习算法被广泛应用于人脸识别、行人再识别、场景分类和其他的视觉领域中。测度学习的基本思想是从特征空间到具有某些优点的距离空间中寻找映射函数,例如来自同一个人的特征向量之间的距离比来自不同行人的特征向量之间的距离更接近。典型的测度学习方法主要包括马氏测度学习算法(KISSME)^[21]、边缘最近邻算法(LMNN)^[22]、局部自适应决定函数算法(LADF)^[12]和语义属性相同匹配算法^[23]。此外,上述行人再识别算法主要是有监督方法,且已有研究人员集中对无监督方式进行研究。例如,Ye等^[24]结合最优全

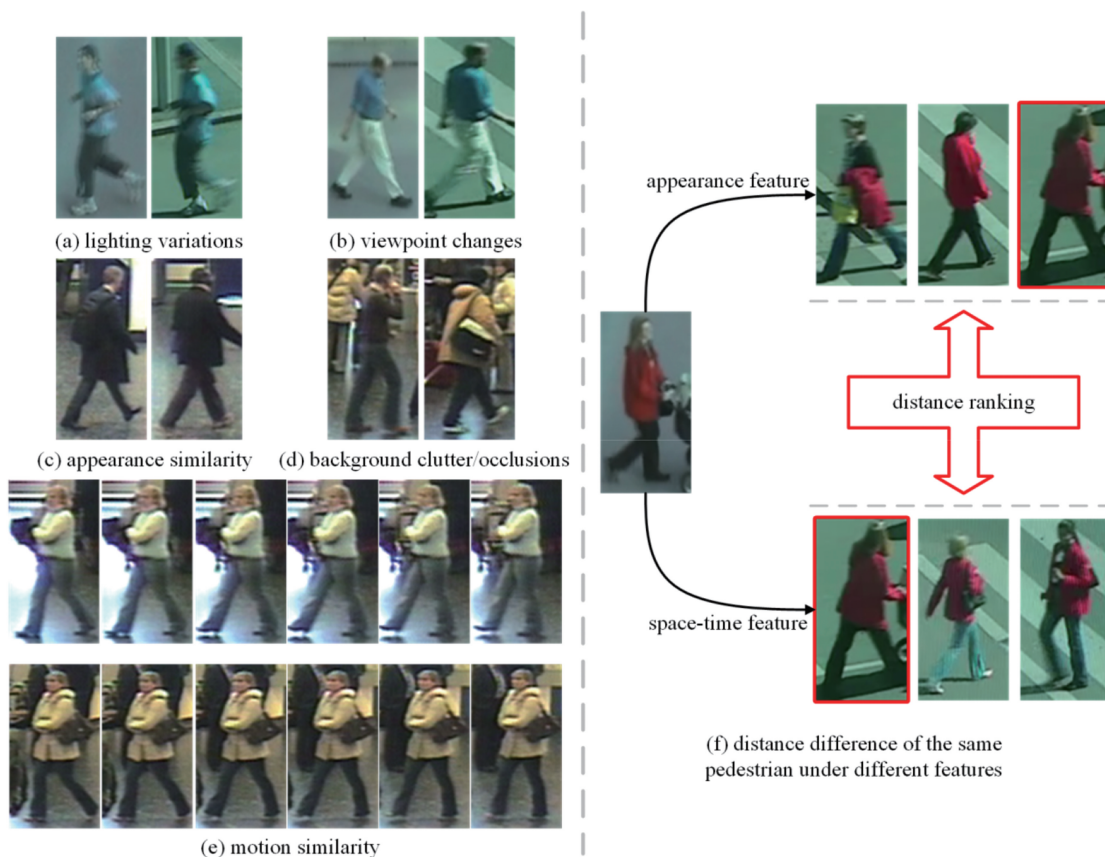


图 1 视频行人再识别中的挑战与困难

Fig. 1 Challenges and difficulties in video-based person re-identification

局匹配和摄像机之间关系的优势,采用图匹配无监督技术进行准确的标签估计。Lü 等^[25]提出了一种无监督的增量学习算法(TFusion),其通过迁移学习来辅助学习目标域中行人的时空模式之间的距离映射。

2.2 行人再识别的深度学习应用

随着深度学习在计算机视觉和模式识别领域的广泛应用,已经有部分文献采用深度学习网络方法处理行人再识别中的挑战和问题。Cheng 等^[26]提出了一种改进的深度学习网络,其采用成对的图像作为其输入,并输出表示两个输入图像是否为同一个人的相似度得分。Yi 等^[27]构建了一个 Siamese 神经网络来学习成对行人图像相似性,并且还使用身体部位来训练其 CNN 模型。Li 等^[28]提出了一种利用新的滤波器配对神经网络(FPNN)的方法,它们使用图像块匹配层来匹配视图中局部图像块的滤波器响应以及其他卷积,联合处理未对准、光照和几何变换、遮挡与背景模糊等问题,并采用最大池化层来处理身体部位的位移。

近些年,采用深度学习网络学习行人丰富的和稳定的特征的研究已经取得突破性进展,越来越多

的研究者提出不同的特征表示网络。例如, Yan 等^[29]采用长短期记忆网络(LSTM)学习并聚合视频序列的时空特征。而 Matsukawa 等^[18]是将预训练改进的 CNN 模型用于提取行人的语义特征,进而应用于行人匹配中。Wu 等^[30]将 CNN 与门循环单元(GRU)网络结合来获取行人的时空特征表示。Dai 等^[31]设计了一个能同时提取连续图像帧中全局特征与局部特征的时间残差学习模块,提高了视频行人再识别的识别率。而本研究针对级联式网络的特征融合问题,提出用行人多级特征表示网络来学习视频中行人的全局及局部外观特征信息和时空特征信息。

3 本文算法架构

所提出的多级深度特征表示网络与有序加权距离测度融合相联合的视频行人再识别方法如图 2 所示。该算法能学习视频序列中行人的全局外观特征、局部外观特征与时空特征,同时能有序加权融合三类特征下的独立距离测度。在行人多级深度特征表示网络中,视频序列经过由卷积层与全连接层组

成的全局外观特征流和由卷积层、Slice层、Concat层与全连接层组成的局部外观特征流之后,分别学习行人的全局外观特征向量 $f'_g(s)$ 和局部外观特征向量 $f'_l(s)$, s 为输入的图像数据。同时,在时空特征流中,部分 CNN 获得的特征向量 $f'_r(s)$ 经 RNN

学习后,可得到行人的时空特征向量 $O'(s)$ 。最后,采用平均时间池化方法将表示该行人的全局外观特征信息、局部外观特征信息与时空特征信息分别聚合为单个特征向量 $f_g(s)$ 、 $f_l(s)$ 和 $O(s)$,并以此联合描述表示该行人。

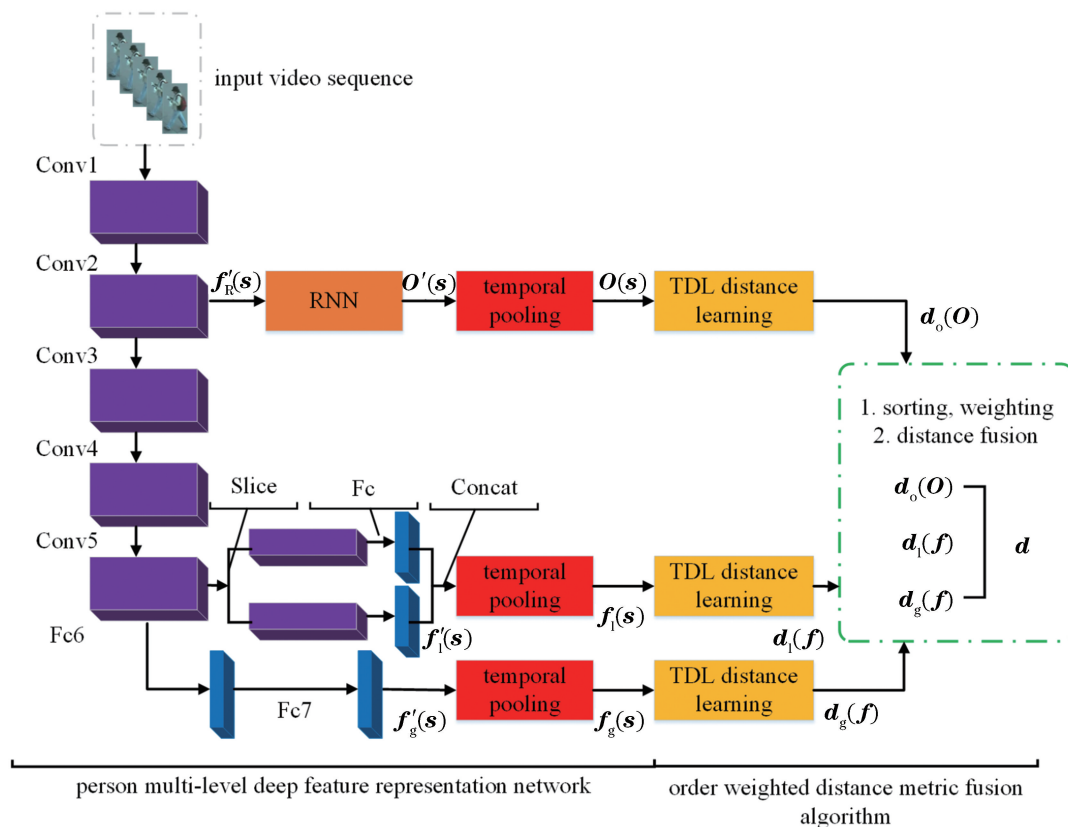


图 2 联合多级深度特征表示和有序加权距离融合的视频行人再识别算法架构

Fig. 2 Architecture of video-based person re-identification algorithm combining multi-level depth feature representation and ordered weighted distance fusion

在有序加权距离融合算法中,将所得到的全局外观特征 $f_g(s)$ 、局部外观特征 $f_l(s)$ 和时空特征 $O(s)$ 分别输入最优距离差异的距离测度学习算法中,并计算行人在三类特征下独立的距离 $d_g(f)$ 、 $d_l(f)$ 和 $d_o(O)$ 。根据每一行人对三类距离值的排序来优化距离融合时的权值,距离融合后得到最终距离 d 并完成行人再识别。

3.1 行人多级深度特征表示网络

3.1.1 全局与局部的外观特征表示-CNN

如图 2 中的行人多级深度特征表示网络,通过 CNN 和时间池化层,获得视频序列中行人丰富的全局和局部外观特征信息。一般情况下,CNN 由卷积层、池化层、非线性激活函数和全连接层组成,本研究改进的 CNN 中增加了随机失活(Dropout)层和归一化(Normalization)层。如表 1 所示,全局外观特征流与局部外观特征流的网络结构主要由 5 层卷

积层、3 层池化层、Slice 层、Concat 层以及全连接层组成。在全局外观特征流中,在第一层池化层和第二层池化层之后都同时增加了归一化层,并且在全连接层 Fc6 后增加随机失活层。增加这两类网络层的目的是抑制和均衡视频序列中行人某些突出特征的表现,减少过拟合,从而获得稳定的外观特征信息。在局部外观特征流,本研究利用 Slice 层将最后一级卷积操作输出的行人特征图均分成上下两部分,即均分行人上半身与下半身。再者,利用全连接层分别将分割后的两个特征图形成局部特征映射。最后,采用 Concat 层合并行人上半身和下半身的局部特征,从而形成更好的行人上半身与下半身的局部外观信息。此外,该网络中池化层所采用的是最大池化(Max-pooling),非线性激活函数所采用的是修正线性单元(ReLU)激活函数。

表 1 全局与局部的外观深度特征网络结构参数

Table 1 Structure parameters of global and local appearance depth feature network

Network layer/Parameter	
Conv1/Filters(96,11×11),Stride 4	
Max pooling1/Filters(3×3),Stride 2	
Conv2/Filters(256,5×5),Stride 1,pad 2	
Max pooling2/Filters(3×3),Stride 2	
Conv3/Filters(384,3×3),Stride 1,pad 1	
Conv4/Filters(384,3×3),Stride 1,pad 1	
Conv5/Filters(256,3×3),Stride 1,pad 1	
Max pooling5/Filters(3×3),Stride 2	
Slice layer	
Fc6 layer	Fc6_Lh layer Fc6_Ll layer
Concat layer	
Fc7 layer	

给定长度为 L 的视频序列 $\mathbf{s} = \{\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \dots, \mathbf{s}^{(L)}\}$ 输入该 CNN 后,将会输出特征向量。该过程表示为

$$\mathbf{f}'_k(\mathbf{s}) = \text{conv}(\mathbf{s}^{(t)}), 1 \leq t \leq L, \quad (1)$$

式中:该特征向量 $\mathbf{f}'_k(\mathbf{s})$ 是经任意卷积层后输出的特征向量图, $K \in \{g, l, o, R\}$ 为每一类特征向量所对应的下标表示,在(1)式中 K 的取值为 g 或 l ,而 $\mathbf{f}'_g(\mathbf{s})$ 、 $\mathbf{f}'_l(\mathbf{s})$ 即为本研究所需的行人全局外观特征表示和局部外观特征表示,并将其作为平均时间池化层的输入; $\mathbf{s}^{(t)}$ 表示视频序列中 t 时刻的原始输入图像, $\text{conv}()$ 表达式所表示的是经任意层的卷积,即经池化,即经全连接层、或归一化层后非线性运算后的结果。此外,将经部分 CNN(2 层卷积层)输出的特征向量 $\mathbf{f}'_R(\mathbf{s})$ 作为 RNN 的输入。

3.1.2 时空特征表示-RNN

由于视频序列中包含某个时间段内的行人图像,所以采用 RNN 提取时空特征能充分利用视频中行人的运动信息,从而改善视频行人再识别的性能。因 RNN 具有递归处理和记忆历史信息的特点,故通常将其用于处理时间序列和空间序列中关联性较强的信息。在视频行人再识别中,视频序列的运动信息和时态信息的关联性是普遍存在的。采用 RNN 将视频中行人的关联信息转化表示为结构化的循环依赖关系,即是本研究使用 RNN 获取行人时空的特征信息。如图 3 所示,本研究的 RNN 是“环形”递归连接,图中的横向“环形”连接具有存储记忆功能,将当前时刻的信息作为下一时刻的部

分输入信息。该 RNN 的输入层是将 3.1.1 节中的部分 CNN 预处理后输出的特征向量作为输入。将第二层卷积层预处理后的输出向量作为时空特征流输入的主要原因是两层卷积层所处理的行人特征图具有高维的空间信息映射。如果采用更深层的卷积特征图将会丢失更多的全局空间特征信息,此时再学习时空特征将会产生信息的偏差。

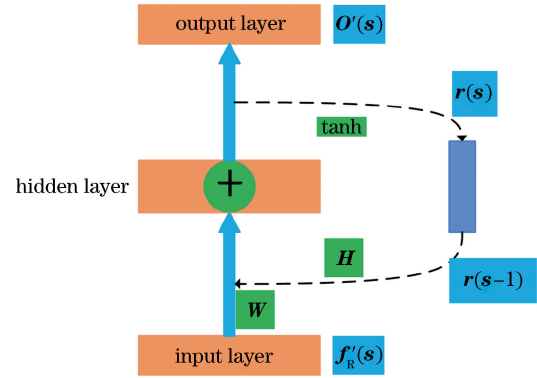


图 3 RNN 结构图

Fig. 3 Structure of RNN

具体地看,输入特征向量 $\mathbf{f}'_R(\mathbf{s})$ 经 RNN 后得到输出向量 $\mathbf{O}'(\mathbf{s})$,计算公式为

$$\mathbf{O}'(\mathbf{s}) = \mathbf{W}\mathbf{f}'_R(\mathbf{s}) + \mathbf{H}\mathbf{r}(\mathbf{s}-1), \quad (2)$$

$$\mathbf{r}(\mathbf{s}) = \tanh[\mathbf{O}'(\mathbf{s})], \quad (3)$$

式中: \mathbf{s} 表示的是视频序列 $\mathbf{s}^{(L)}$ 中 t 时刻的行人图像 $\mathbf{s}^{(t)}$; 输入向量 $\mathbf{f}'_R(\mathbf{s})$ 包含 t 时刻递归神经网络之前的图像信息; 系数矩阵 \mathbf{W} 和 \mathbf{H} 是将 $\mathbf{f}'_R(\mathbf{s})$ 映射到低维特征向量空间; $\mathbf{r}(\mathbf{s})$ 是 $\tanh[\]$ 非线性函数的输出; 输出向量 $\mathbf{O}'(\mathbf{s})$ 表示的是 t 时刻 RNN 的特征信息,即是行人的时空特征表示。

3.1.3 聚合序列多级特征-时间池化层

虽然本研究的多级深度特征表示网络可以分别提取行人的外观特征向量和时空特征向量,但是所获得的行人信息仍然存在某些局限性,该缺陷主要体现在两方面。第一,如果同一行人完整的视频序列经网络后输出的两类序列特征信息不经过任何处理,便会抑制外观特征信息和时空特征信息的作用。因为每一视频序列中具有分辨性的信息可能出现在任何一帧图像中,而并不能完全确定行人的区分性信息一定出现在视频序列的末尾。第二,因每一视频序列的时间尺度不同,故标准递归神经网络提取不同时间尺度下视频序列的时空特征是不明确的。针对上述两方面的缺陷,本研究在 CNN 和 RNN 的输出向量后增加时间池化层解决此问题。该时间池化层不仅能聚合所有时刻的信息和避免当前时刻与

之前时刻的信息偏差,还能获取视频序列中的长尺度特征信息,RNN的短尺度时间特征和时间池化层的中尺度时间特征,从而建立完整的多时间尺度信息。

假设长度为 L 的行人视频序列经 CNN 和 RNN 后得到的行人的全局外观特征、局部外观特征和时空特征分别为 $\mathbf{f}'_g(\mathbf{s}) = \{\mathbf{f}'_g(\mathbf{s}^{(1)}), \mathbf{f}'_g(\mathbf{s}^{(2)}), \dots, \mathbf{f}'_g(\mathbf{s}^{(L)})\}$ 、 $\mathbf{f}'_l(\mathbf{s}) = \{\mathbf{f}'_l(\mathbf{s}^{(1)}), \mathbf{f}'_l(\mathbf{s}^{(2)}), \dots, \mathbf{f}'_l(\mathbf{s}^{(L)})\}$ 和 $\mathbf{O}'(\mathbf{s}) = \{\mathbf{O}'(\mathbf{s}^{(1)}), \mathbf{O}'(\mathbf{s}^{(2)}), \dots, \mathbf{O}'(\mathbf{s}^{(L)})\}$,再将三类特征向量分别输入到平均时间池化层中,该层主要把表示序列的特征向量输入并将其分别聚合转化为单个全局外观特征向量、局部外观特征向量和时空特征向量,并用此输出的特征信息表示该视频序列中的行人。该过程如(4)式、(5)式和(6)式所示:

$$\mathbf{f}_g(\mathbf{s}) = \frac{1}{L} \sum_{t=1}^L \mathbf{f}'_g(\mathbf{s}^{(t)}), \quad (4)$$

$$\mathbf{f}_l(\mathbf{s}) = \frac{1}{L} \sum_{t=1}^L \mathbf{f}'_l(\mathbf{s}^{(t)}), \quad (5)$$

$$\mathbf{O}(\mathbf{s}) = \frac{1}{L} \sum_{t=1}^L \mathbf{O}'(\mathbf{s}^{(t)}). \quad (6)$$

3.2 有序加权距离测度融合算法

3.2.1 距离测度学习

基于上述获得的行人的全局外观特征表示、局部外观特征表示和时空特征表示,采用文献[32]中的距离测度学习算法,分别计算出行人之间三类特征下的独立距离。下面将具体简述该距离测度学习算法模型建立与优化,如图4所示。

视频行人再识别中的理想结果是对于目标行人的视频序列,在视频序列集中匹配的最优排序总是正确的。这意味着每一目标行人与视频序列集中相同行人之间的距离总是比任一目标行人与视频序列集中不相同行人之间的距离小,即是最小类间距离总是大于最大类内距离。因此,该距离测度学习模型能增大类间距离,减小类内距离。

参数表示:将行人样本的全局外观特征、局部外观特征和时空特征分别表示为集合 $\mathbf{v}_{gm} = \{\mathbf{f}_{gm}(\mathbf{s}), x_m\}$ 、集合 $\mathbf{v}_{lm} = \{\mathbf{f}_{lm}(\mathbf{s}), x_m\}$ 和集合 $\mathbf{v}_{om} = \{\mathbf{O}_m(\mathbf{s}), x_m\}$,其中 $\mathbf{f}_{gm}(\mathbf{s})$ 、 $\mathbf{f}_{lm}(\mathbf{s})$ 和 $\mathbf{O}_m(\mathbf{s})$ 分别表示视频序列中行人 x_m 的全局外观特征、局部外观特征和时空特征, x_m 为行人的标签, m 表示第 m 个行人。同时,行人样本集也可表示为 $\{\mathbf{v}_{K1}, \mathbf{v}_{K2}, \dots, \mathbf{v}_{Km}\}_{m=1}^T$ 。再者,样本集中任意一对行人之间的距离表达式为

$$d_{K(m,w,p)} = \begin{cases} d_{K\Diamond}(\mathbf{v}_{Km}, \mathbf{v}_{Kp}), & x_m \neq x_p \\ d_{K\nabla}(\mathbf{v}_{Km}, \mathbf{v}_{Kw}), & x_m = x_w \end{cases}, \quad (7)$$

式中: $d_{K\Diamond}$ 表示类间距离, $d_{K\nabla}$ 表示类内距离; w, p 分别表示第 w 个行人和第 p 个行人。距离的定义与计算采用的是马氏距离,即

$$d_K(\mathbf{v}_{Km}, \mathbf{v}_{Kw}) = (\mathbf{v}_{Km} - \mathbf{v}_{Kw})^T \mathbf{M} (\mathbf{v}_{Km} - \mathbf{v}_{Kw}), \quad (8)$$

式中 \mathbf{M} 为半正定矩阵。最后,将行人之间的所有距离表示为

$$D_K = \begin{cases} D_{K\Diamond} = \{d_{K\Diamond 1}, d_{K\Diamond 2}, \dots, d_{K\Diamond a}\}_{a=z} \\ D_{K\nabla} = \{d_{K\nabla 1}, d_{K\nabla 2}, \dots, d_{K\nabla b}\}_{b=y} \end{cases}, \quad (9)$$

式中: $D_{K\Diamond}$ 表示类间距离集合, $D_{K\nabla}$ 表示类内距离集合, $a=z$ 和 $b=y$ 分别表示类间距离集合与类内距离集合中的数量。

模型建立及优化:首先,为实现相对距离比较,在类间距离集合中寻找最小距离表示为 $\min D_{K\Diamond}$,并期望最小类间距离总是大于每一类内距离,该思想反映为

$$\min D_{K\Diamond} > D_{K\nabla} + \epsilon, \quad (10)$$

式中 ϵ 是自由参数,设定 $\epsilon=1$ 。同时,既要保证最大化最小类间距离与类内距离的差值,又要减小类内距离 $\min D_{K\nabla}$,并在此基础上融入相对距离比较的思想,故将目标函数定义为

$$F(D_K) = \beta \min_y \sum_y D_{K\nabla} + (1-\beta) \cdot \max_{y,z} \sum_{y,z} \{D_{K\nabla} + \epsilon - \min D_{K\Diamond}, 0\}, \quad (11)$$

式中, β 为权值平衡参数,用于均衡该式中前后两项,其范围为 $\beta \in [0, 1]$ 。最后,采用随机梯度下降法优化此目标函数。优化过程中的梯度计算公式为

$$G_q = \beta \sum_y D_{K\nabla} + (1-\beta) \cdot \sum_{y,z} \{D_{K\nabla} - D_{K\Diamond}\}, \quad (12)$$

式中 q 表示第 q 次梯度运算。

3.2.2 有序加权的距离融合

为了解决相同行人在不同特征下距离的差异性而导致匹配错误的问题,设计了一种有序加权的距离融合算法,利用不同的距离权值减缓和消除不同特征在相同行人下的差异性,使得再识别行人时的匹配更加准确。此距离融合的策略即是由于相同行人的不同特征具有不同的置信值,最高置信值下的特征对应的是最小距离值,所以该特征应获得最高权值的奖励。受到文献[6, 33]启发,本研究依据该融合策略设计了有序加权的距离融合算法。如图5所示,根据所得到的行人之间在三类特征下的独立距离 $d_g(\mathbf{f})$ 、 $d_l(\mathbf{f})$ 和 $d_o(\mathbf{O})$,将每一行人对的三类距离排序,再根据距离的排名优化计算距离的权值。

Algorithm: distance metric learning

Input: Global appearance feature vector $f_g(s)$, local appearance feature vector $f_l(s)$ and space-time feature vector $O(s)$.

Output: Compute the independent distances $d_g(f)$, $d_l(f)$ and $d_o(O)$ between persons under three types of features.

Step 1: The initial processing of three types of feature vectors $f_g(s)$, $f_l(s)$ and $O(s)$ are expressed as v_{gm} , v_{lm} and v_{om} .

Step 2: Compute distance between any pair of persons as $d_K(v_{Km}, v_{Kw})$, $K \in \{g, l, o, R\}$ by Eq. (7) and Eq. (8).

Step 3: Represent the set of inter-class distances $D_{K\phi}$ and the intra-class distance $D_{K\psi}$ between all persons by Eq. (9).

Step 4: Find the smallest inter-class distance $\min D_{K\phi}$, construct the relationship model between the minimum

inter-class distance and the intra-class distance: $\max\{D_{K\psi} + \varepsilon - \min D_{K\phi}, 0\}$ by Eq. (10).

Step 5: Achieve reduced intra-class distance: $\min D_{K\psi}$.

Step 6: Construct the objective function of distance learning:

$$F(D_K) = \beta \min_y \sum_y D_{K\psi} + (1 - \beta) \max_{y,z} \sum_{y,z} \{D_{K\psi} + \varepsilon - \min D_{K\phi}, 0\}.$$

Step 7: Initial transformation of objective function: $V_K = (v_{Km} - v_{Kw})^T (v_{Km} - v_{Kw}) \longrightarrow D_K = \text{tr}(MV_K)$

$$\longrightarrow F(M) = \beta \min \sum \text{tr}(MV_{K\psi}) + (1 - \beta) \max \sum \{\text{tr}(MV_{K\psi}) + \varepsilon - \min \text{tr}(MV_{K\phi}), 0\}.$$

Step 8: Optimizing the objective function by stochastic gradient descent:

8-1 Initialization parameter: the metric matrix $M_0 := I$, the gradient $G_q = \beta \sum V_{K\psi}$, the counter $q := 0$.

8-2 While (not converged) do

8-2-1 Find the smallest intra-class distance: $\min D_{K\phi}$.

8-2-2 Compute the distance between persons by Eq. (8), and construct the set of distance.

8-2-3 Compute gradient G_q by Eq. (12).

8-2-4 Compute $M_q := M_q - \lambda G_q$.

8-2-5 $q := q + 1$.

End

8-3 Return matrix M .

Step 9: Compute the independent distances $d_g(f)$, $d_l(f)$ and $d_o(O)$ between persons under three types of features by Eq. (8) and matrix M .

图 4 本文所采用的距离测度算法计算行人之间距离的流程

Fig. 4 Flowchart of distance metric learning algorithm used for computing distance between persons

最后将三类距离加权融合成最终距离 d 。

$d_g(f)$ 、 $d_l(f)$ 和 $d_o(O)$ 作为有序加权距离融合算法

具体地说,将每一行人三类特征下的独立距离

的输入,将输入的距离由小到大排序表示为集合

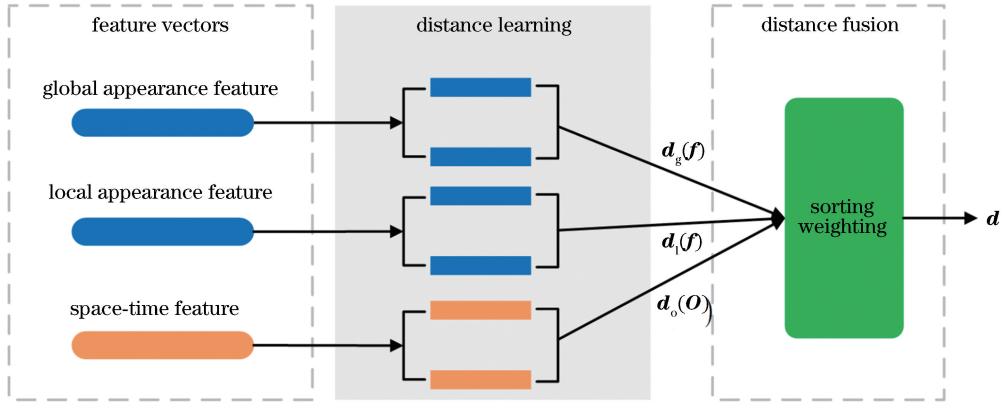


图 5 有序加权距离融合算法图

Fig. 5 Algorithm of ordered weighted distance fusion

$\{[d_g(\mathbf{f}), u_g], [d_l(\mathbf{f}), u_l], [d_o(\mathbf{O}), u_o]\}_{x_m, x_n}$, $1 \leq u_K \leq 3, u_K \in \mathbf{N}_+$, 其中 $K \in \{g, l, o, R\}$, 而 u_K 表示所对应距离的排名, 即最小距离对应 $u_K = 1$ 。最后, 根据排名加权融合距离, 此过程表示为

$$\mathbf{d} = \alpha_g \mathbf{d}_g(\mathbf{f}) + \alpha_l \mathbf{d}_l(\mathbf{f}) + \alpha_o \mathbf{d}_o(\mathbf{O}), \quad (13)$$

$$\alpha_K = \frac{\prod_{K'} u_K - u_K}{\sum_K u_K}, K \in \{g, l, o, R\}, \quad (14)$$

式中: α_K 表示的是每一类特征的特征权值; \mathbf{d} 表示行人之间融合的最终距离。

4 实验介绍

4.1 实验数据集介绍

行人再识别研究中常用的实验数据集如表 2 所示。下面将主要介绍本研究实验所采用的 i-LIDS^[34] 数据集和 PRID-2011^[35] 数据集。

i-LIDS 数据集包括某机场 119 位行人的 476 张图片(每位行人平均 4 张图片), 是常用的视频(多帧)图像序列的行人再识别实验数据集。i-LIDS 数据集的视频数据由两个摄像机视角 1 和 2 组成, 其中包括 300 位行人, 600 段视频序列, 每位行人视频序列的长度是在 23 frame 到 192 frame 图像之间, 该数据集不仅存在行人视角、姿态和光照方面的

变化, 还存在较严重的部分遮挡现象。

PRID 数据集包含两个数据库, 分别是 PRID-2011 和 PRID-450S。其中, PRID-2011 数据集由两个摄像机视角 A 和 B 组成, 其中 A 视角下包括 385 位行人, B 视角下包括 749 位行人, 每个行人视频序列的长度在 5 frame 到 675 frame 图像之间。特别注意的是, PRID-2011 数据集中, A、B 视角下存在 200 位行人是相同的。这些视频序列存在视角、光照和背景的差异。

4.2 实验设计及评价准则

实验分别将 PRID-2011 和 i-LIDS 两个数据集随机均分为两部分: 训练数据集和测试数据集。在 i-LIDS 数据集中, 训练集和测试集中的行人分别为 150, 在 PRID-2011 数据集中, 训练集和测试集中的行人分别为 89。实验中设定视频序列的长度 L 为 8。当数据集中视频序列长度小于 8 时, 则将弃用该视频序列行人。实验平台采用的是搭载显卡 GTX 1070, 内存 16G 的计算机。实验采用的框架和软件是深度学习框架 Caffe^[36] 和 Matlab 软件。在行人多级深度特征表示网络训练阶段, 选择三个 Softmax 函数作为损失函数联合预训练行人多级深度特征表示网络, 并获得特征模型。训练网络时参数设定如下: 基础学习率为 0.0001, 学习率下降策略

表 2 行人再识别常用公共数据集

Table 2 Public datasets for person re-identification

Experimental dataset	Characteristic	Object
PRID-2011	Viewpoint, lighting, and background changes	Outdoor person
i-LIDS	Viewpoint and lighting changes, occlusions	Airport person
VIPeR	Viewpoint, lighting, and background changes	School person
CAVIAR4REID	Lighting changes and occlusions	Shopping person
GRID	Viewpoint and lighting changes, low resolution	Subway person

为“均值下降(step)策略”,批大小为 50,最大迭代为 10000 次,网络训练中所采用的优化方法为随机梯度下降法(SGD),而所采用的预训练模型是在 Imagnet 数据集训练好的 CaffeNet 模型。在距离测度学习阶段,利用获得的特征向量分别训练优化独立的距离测度矩阵。在算法的测试(行人再识别)阶段,实验选取一个摄像机视角下的视频序列作为视频集,从另一个摄像机视角中挑选某个人行人视频作为目标行人视频。并将视频序列输入到已完成训练的模型中提取行人的全局外观特征、局部外观特征和时空特征,再通过距离测度学习分别计算行人之间在三类特征下的独立距离。最后利用有序加权距离融合策略计算最终距离,并通过匹配判断是否为相同行人。实验重复 10 次,并记录平均结果,从而获得可靠的实验结果。

实验评价行人再识别算法优劣的准则是累计匹配特征曲线(CMC)和排序- n (Rank- n)识别率。CMC 是描述和计算行人再识别算法性能表现的评价依据,Rank- n 识别率(单位为%)是指在 Rank- n 的列表中,目标视频序列与视频序列集中匹配的准确率。

4.3 与基于静止图像的相关方法的对比实验

将所提算法与基于静止图像的相关行人再识别方法在 i-LIDS 和 PRID-2011 两个数据集进行比较,结果如表 3、表 4 所示,所选择的相关传统经典的基于静止图像行人再识别的方法包括以下 4 种:GRDL 算法^[37]、DVDL 算法^[38]、SRID 算法^[39]、Saliency 算法^[40],表中的加粗数据表示为最高的 Rank- n 识别率。在 i-LIDS 数据集中,综合来看,所提算法的 Rank- n 识别率都高于其余 4 种算法的识别率;具体来看,所提算法的 Rank-1 识别率比 GRDL 算法高出 30%,比 DVDL 算法高出 29.8%。在 PRID-2011 数据集中,所提算法的 Rank-1 识别率与 DVDL 算法高出 31.2%,比 GRDL 算法高出 30.2%,而与其余算法相比,所提算法的识别率都优于其余两类算法。分析认为所提算法相比传统算法识别率较高的原因是描述行人的多级深度特征相比传统手工特征具有较强的辨识性,从而算法中行人的表征能力得到增强。图 6 给出所提算法与其余 4 类算法的 CMC 比较结果如图中所示,黑色虚线即表示所提算法在 i-LIDS 和 PRID-2011 实验数据集中的 CMC。

参考实验表 3、表 4 中的数据 and 图 6 中的 CMC,与基于静止图像的相关方法对比,发现所提算法体

现基于视频的行人再识别方法中视频的特殊优势,在一定程度上增强了视频行人再识别中行人的多级特征表示能力,丰富的行人多级特征信息能提高行人再识别的准确率。

表 3 与基于静止图像的相关方法在 i-LIDS 数据集比较结果

Method	Matching rate of i-LIDS dataset (image-based methods) / %			
	Rank-1	Rank-5	Rank-10	Rank-20
Proposed	55.7	84.7	95.3	96.7
GRDL	25.7	49.9	63.2	77.6
DVDL	25.9	48.2	57.3	68.9
SRID	24.9	44.5	55.6	66.2
Saliency	10.2	24.8	35.5	52.9

表 4 与基于静止图像的相关方法在 PRID-2011 数据集比较结果

Method	Matching rate of PRID-2011 dataset (image-based methods) / %			
	Rank-1	Rank-5	Rank-10	Rank-20
Proposed	71.8	90.4	94.1	97.5
GRDL	41.6	76.4	84.9	89.9
DVDL	40.6	69.7	77.8	85.6
SRID	35.1	59.4	69.8	79.7
Saliency	25.8	43.6	52.6	62.0

4.4 与基于视频的相关方法对比实验

为了评价所提算法与相关的基于视频行人再识别方法的有效性,实验选取 6 种相关的视频行人再识别的方法,分别如下:TDL^[32]、DGM + IDE^[24](IDE 表示距离测度算法)、VR^[34]、PAMM^[41]、UNKISS^[42]、ISR^[43]。如表 5、表 6 所示,表中加粗数据表示最高的 Rank- n 识别率。从表中具体来看,在 i-LIDS 数据集中,所提算法的 Rank-1 识别率优于大部分算法,仅低于 TDL 算法 0.6%。在 PRID-2011 数据集中,所提算法的 Rank-1 识别率都高于相对比的所有算法,并高于 TDL 算法 15.1%。图 7 给出所提算法与 6 种对比算法在 i-LIDS 和 PRID-2011 两个数据集上的 CMC,虚线表示所提算法。分析认为所提算法相比于 6 种视频行人再识别传统算法呈现较好结果的原因是丰富的外观特征表示给算法的特征表示模型带来了稳定性与全面性。表 5、表 6 和图 7 的结果验证了所提算法表现性能

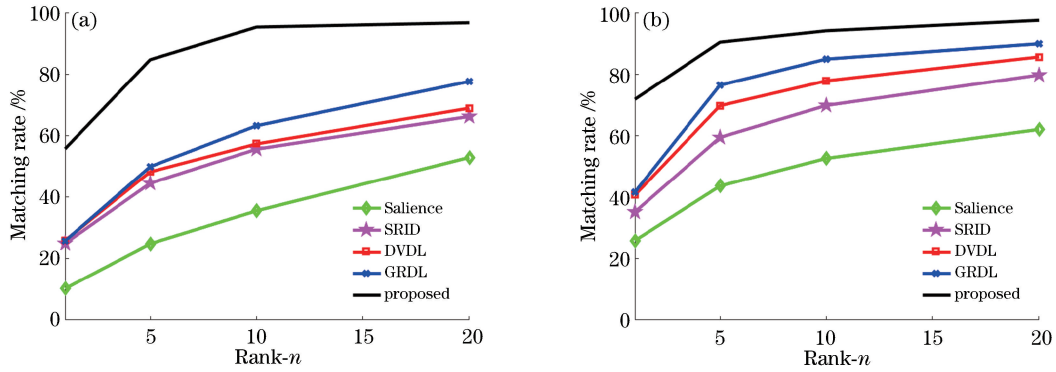


图 6 与基于静止图像行人再识别方法在 i-LIDS 数据集和 PRID-2011 数据集相比较的 CMC。

(a) i-LIDS 数据集结果; (b) PRID-2011 数据集结果

Fig. 6 CMCs of comparison with still-image-based person re-identification methods on i-LIDS and PRID-2011 datasets.

(a) i-LIDS dataset; (b) PRID-2011 dataset

表 5 与基于视频行人再识别的相关方法在 i-LIDS 数据集比较结果

Table 5 Results of comparison with video-based person re-identification methods on i-LIDS dataset

Method	Matching rate of i-LIDS dataset (video-based methods) / %			
	Rank-1	Rank-5	Rank-10	Rank-20
Proposed	55.7	84.7	95.3	96.7
TDL	56.3	87.6	95.6	98.3
DGM+IDE	37.2	62.6	73.4	80.8
VR	23.3	42.4	55.3	68.4
PAMM	30.3	56.3	70.3	82.7
UNKISS	35.9	63.3	74.9	83.4
ISR	11.6	22.1	27.4	36.7

表 6 与基于视频行人再识别的相关方法在 PRID-2011 数据集比较结果

Table 6 Results of comparison with video-based person re-identification methods on PRID 2011 dataset

Method	Matching rate of PRID-2011 dataset (video-based methods) / %			
	Rank-1	Rank-5	Rank-10	Rank-20
Proposed	71.8	90.4	94.1	97.5
TDL	56.7	80.0	87.6	93.6
DGM+IDE	61.6	89.0	94.8	98.2
VR	28.9	55.3	65.5	82.8
PAMM	45.0	72.0	85.0	92.5
UNKISS	58.1	81.9	89.6	96.0
ISR	17.6	35.8	43.0	54.4

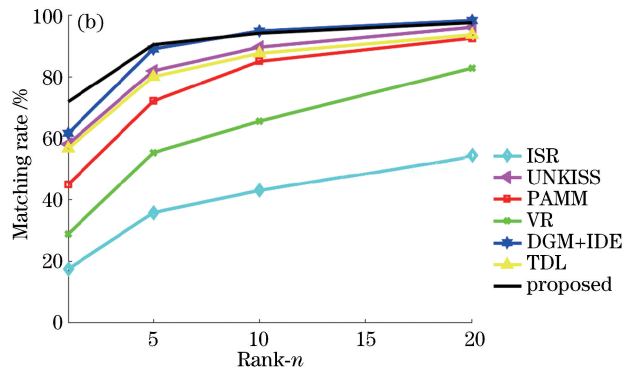
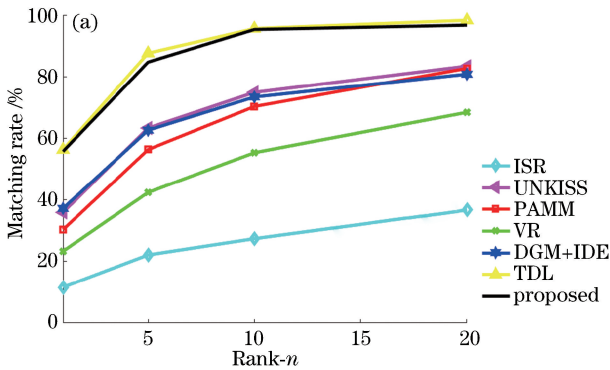


图 7 与基于视频行人再识别方法在 i-LIDS 数据集和 PRID-2011 数据集相比较的 CMC。

(a) i-LIDS 数据集; (b) PRID-2011 数据集

Fig. 7 CMCs of comparison with video-based person re-identification methods on i-LIDS and PRID-2011 datasets.

(a) i-LIDS dataset; (b) PRID-2011 dataset

在 i-LIDS 和 PRID-2011 两个数据集中具有一定的有效性。

4.5 与基于深度学习的相关方法对比实验

本节实验是将所提算法与相关的基于深度学习

行人再识别方法进行比较,目的是评价所提出联合多级深度特征表示和有序加权距离融合的视频行人再识别方法的性能表现。如表 7、表 8 所示,此次实验选取的 6 种基于深度学习行人再识别的方法如

表 7 与基于深度学习行人再识别的相关方法在 i-LIDS 数据集比较结果

Table 7 Results of comparison with deep-learning-based person re-identification methods on i-LIDS dataset

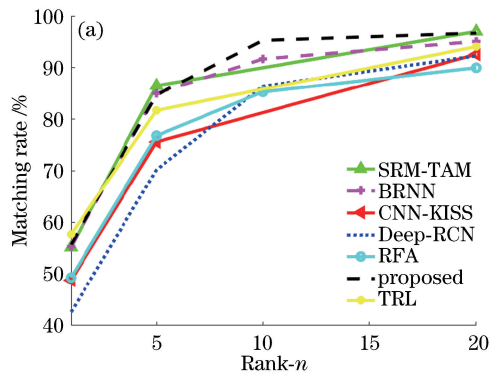
Method	Matching rate of i-LIDS dataset (deep learning methods) / %			
	Rank-1	Rank-5	Rank-10	Rank-20
Proposed	55.7	84.7	95.3	96.7
RFA	49.3	76.8	85.3	90.0
Deep-RCN	42.6	70.2	86.4	92.3
CNN-KISS	48.8	75.6	—	92.6
BRNN	55.3	85.0	91.7	95.1
SRM-TAM	55.2	86.5	—	97.0
TRL	57.7	81.7	—	94.1

表 8 与基于深度学习行人再识别的相关方法在 PRID-2011 数据集比较结果

Table 8 Results of comparison with deep-learning-based person re-identification methods on PRID 2011 dataset

Method	Matching rate of PRID-2011 dataset (deep learning methods) / %			
	Rank-1	Rank-5	Rank-10	Rank-20
Proposed	71.8	90.4	94.1	97.5
RFA	58.2	85.8	93.4	97.9
Deep-RCN	49.8	77.4	90.7	94.6
CNN-KISS	69.9	90.6	—	98.2
BRNN	72.8	92.0	95.1	97.6
SRM-TAM	79.4	94.4	—	99.3
TRL	87.8	97.4	99.3	99.3

下: RFA^[29]、Deep-RCN^[30]、CNN-KISS^[44]、BRNN^[3]、SRM-TAM^[45]、TRL^[31],表中加粗字体表示最高的识别率。从表 7、表 8 可以看出,在 i-LIDS



数据集中,所提算法的 Rank-1 识别率领先 BRNN 算法仅 0.4%,与 SRM-TAM 算法相比仅大于 0.5%,与 TRL 算法相比小于 2.0%,而与其余三种算法相比,所提算法的优势显著。在 PRID-2011 数据集中,所提算法的 Rank-n 识别率与 BRNN 算法、SRM-TAM 算法、TRL 算法相比具有一定的差距,但都优于其余三类算法。分析认为所提算法的识别率与 BRNN 算法、SRM-TAM 算法、TRL 算法在 PRID-2011 数据集中相比存在差距的可能原因是 BRNN 算法、SRM-TAM 算法与 TRL 算法在应对视角变化(PRID-2011 数据集基本特点)时具有优异的表现性能,而所提算法集中于应对背景干扰与遮挡(i-LIDS 数据集基本特点)的问题。同时,这三类算法是端到端的学习模型,在训练优化时相比本研究的阶段式算法确有优势。图 8 给出所提算法与 6 种基于深度学习行人再识别方法的 CMC 结果,其中黑色曲线表示的仍然是所提算法的 CMC。综合来看此次对比实验的结果,本研究中的行人多级深度特征表示网络发挥了学习提取特征的优势,具有一定的有效性和稳健性。

4.6 关键因素的验证实验

距离融合的影响:本次实验是为了验证有序加权的距离融合算法能缓解相同行人在不同特征下距离差异性导致的错误匹配问题,并能有效提高视频行人再识别结果。因此,将所提算法分别与独立的局部外观特征(LAF-1)、时空特征(ST-1)和全局外观特征(GAF-1)在 i-LIDS 和 PRID-2011 两个数据集中进行实验比较,如表 9、表 10 中第 1~4 行所示。由以上两个数据集的实验数据可以看出,所提算法的 Rank-n 识别率显著高于对独立特征分别进行距离学习后的识别率,这表明采用有序加权的距

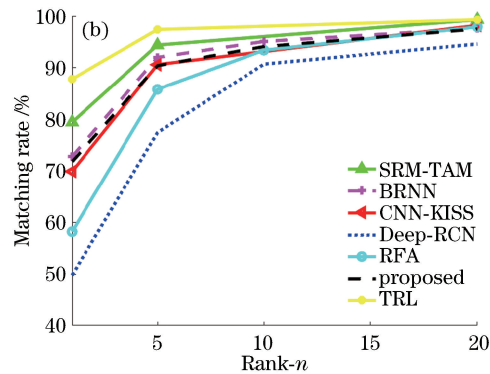


图 8 与基于深度学习行人再识别方法在 i-LIDS 数据集和 PRID-2011 数据集相比较的 CMC。

(a) i-LIDS 数据集;(b) PRID-2011 数据集

Fig. 8 CMCs of comparison with deep-learning-based person re-identification methods on i-LIDS and PRID-2011 dataset.

(a) i-LIDS dataset; (b) PRID-2011 dataset

离融合算法可以改善视频行人再识别的识别率。此外,局部外观特征(LAF-1)即将局部外观特征向量应用于距离测度学习算法后直接匹配行人,计算Rank- n 识别率,全局外观特征(GAF-1)和时空特征(ST-1)同理。

表 9 关键因素在 i-LIDS 数据集的比较结果

Table 9 Results of comparison with key factors on i-LIDS dataset

Method	Matching rate of i-LIDS dataset / %			
	Rank-1	Rank-5	Rank-10	Rank-20
Proposed	55.7	84.7	95.3	96.7
LAF-1	5.3	14.0	21.3	32.7
ST-1	38.7	71.3	82.3	88.0
GAF-1	46.0	74.0	79.3	89.3
LAF+ST	24.1	52.0	70.6	83.4
GAF+ST	31.5	61.4	71.3	84.0

表 10 关键因素在 PRID-2011 数据集的比较结果

Table 10 Results of comparison with key factors on PRID 2011 dataset

Method	Matching rate of PRID-2011 dataset / %			
	Rank-1	Rank-5	Rank-10	Rank-20
Proposed	71.8	90.4	94.1	97.5
LAF-1	12.4	30.3	41.6	61.8
SF-1	33.7	65.2	75.3	87.6
GAF-1	40.4	64.0	79.8	91.0
LAF+ST	28.1	53.9	68.5	80.9
GAF+ST	42.7	67.4	81.8	89.9

多级特征表示的影响:此次实验是为了验证多级特征表示网络中所学习的行人多级特征信息能有效提高视频行人再识别的正确率。因此,将所提算法分别与局部外观特征及时空特征组合(LAF+ST)、全局外观特征及时空特征组合(GAF+ST)在 i-LIDS 和 PRID-2011 两个数据集中进行验证比较,如表 9、表 10 中第 1、第 5、第 6 行。由在两个数据集的实验数据看出,所提算法的 Rank- n 识别率显著高于 LAF+ST 特征组合和 GAF+ST 特征组合的结果,验证了本研究的多级特征表示网络具有丰富的特征表示能力。其中,GAF+ST 特征组合(即不分割的行人外观特征)的识别率低于所提算法(即分割后的行人外观特征)的识别率,由此能够验证局部外观特征有益于构建完整的行人特征表示。

参数 β 的影响:本研究的距离测度学习算法中权值参数 β 的改变将会影响实验结果,在 PRID-2011 数据集中实验时,发现参数设定为 $\beta=0.15$ 使算法模型在实验中获得较好的结果。而在 i-LIDS 数据集中实验时,将参数设定为 $\beta=0.13$ 能获得较高的识别率。在两个数据中,当参数设定为 $\beta=1$ 时,类内距离的变化最小,都将会导致过拟合现象的出现。

5 结 论

提出联合多级深度特征表示和有序加权距离融合的视频行人再识别方法学习视频中行人的全局外观特征、局部外观特征和时空特征,再计算行人之间在三类特征下的独立距离并有序加权融合为最终距离,从而能准确再识别行人。其中,所提算法中提出的行人多级深度特征表示网络和距离融合算法能有效解决视频序列中行人外观、行为相似和相同行人在不同特征下距离的差异性而导致匹配不正确的问题。同时,实验结果也表明所提算法在与前沿的基于深度学习行人再识别方法比较时,依然具有一定的识别差距,分析认为这是由于本研究的行人多级深度特征表示所提取的某些背景干扰和外观相似的行人特征不具有区分性,并且本研究所采用数据集的样本数量有限而引起的。因此,进一步优化改进算法和增加训练样本数据的数量将是本课题组今后工作的重点。

参 考 文 献

- [1] Bauml M, Tapaswi M, Schumann A, *et al.* Contextual constraints for person retrieval in camera networks [C] // 2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance, September 18-21, 2012, Beijing, China. New York: IEEE, 2012: 221-227.
- [2] McLaughlin N, del Rincon J M, Miller P. Recurrent convolutional network for video-based person re-identification [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 1325-1334.
- [3] Zhang W, Yu X D, He X. Learning bidirectional temporal cues for video-based person re-identification [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(10): 2768-2776.
- [4] Tao D P, Guo Y N, Yu B S, *et al.* Deep multi-view feature learning for person re-identification [J]. IEEE

- Transactions on Circuits and Systems for Video Technology, 2018, 28(10): 2657-2666.
- [5] Chu M D, Wu S, Gu Y F, *et al.* Rich features and precise localization with region proposal network for object detection[M]//Zhou J, Wang Y H, Sun Z A, *et al.* Biometric recognition. Lecture notes in computer science. Cham: Springer, 2017, 10568: 605-614.
- [6] Mirmahboub B, Mekhalfi M L, Murino V. Distance penalization and fusion for person re-identification [C]//2017 IEEE Winter Conference on Applications of Computer Vision (WACV), March 24-31, 2017, Santa Rosa, CA, USA. New York: IEEE, 2017: 1306-1314.
- [7] Xu Y, Lu Y W. Adaptive weighted fusion: a novel fusion approach for image classification [J]. Neurocomputing, 2015, 168: 566-574.
- [8] Kviatkovsky I, Adam A, Rivlin E. Color invariants for person reidentification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35 (7): 1622-1634.
- [9] Wu Z Y, Li Y, Radke R J. Viewpoint invariant human re-identification in camera networks using pose priors and subject-discriminative features [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(5): 1095-1108.
- [10] Wu S X, Chen Y C, Li X, *et al.* An enhanced deep feature representation for person re-identification [C]//2016 IEEE Winter Conference on Applications of Computer Vision (WACV), March 7-10, 2016, Lake Placid, NY, USA. New York: IEEE, 2016: 7477681.
- [11] Zhu X B, Che J. Person re-identification algorithm based on feature fusion and subspace learning [J]. Laser & Optoelectronics Progress, 2019, 56 (2): 021503.
朱小波, 车进. 基于特征融合与子空间学习的行人重识别算法[J]. 激光与光电子学进展, 2019, 56(2): 021503.
- [12] Li Z, Chang S Y, Liang F, *et al.* Learning locally-adaptive decision functions for person verification [C]//2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 23-28, 2013, Portland, OR, USA. New York: IEEE, 2013: 3610-3617.
- [13] Liao S C, Hu Y, Zhu X Y, *et al.* Person re-identification by local maximal occurrence representation and metric learning [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 2197-2206.
- [14] Xiong F, Gou M R, Camps O, *et al.* Person re-identification using kernel-based metric learning methods[M]//Fleet D, Pajdla T, Schiele B, *et al.* Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8695: 1-16.
- [15] Li W, Wang X G. Locally aligned feature transforms across views [C]//2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 23-28, 2013, Portland, OR, USA. New York: IEEE, 2013: 3594-3601.
- [16] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24 (7): 971-987.
- [17] Farenzena M, Bazzani L, Perina A, *et al.* Person re-identification by symmetry-driven accumulation of local features [C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA, USA. New York: IEEE, 2010: 2360-2367.
- [18] Matsukawa T, Suzuki E. Person re-identification using CNN features learned from combination of attributes[C]//2016 23rd International Conference on Pattern Recognition (ICPR), December 4-8, 2016, Cancun. New York: IEEE, 2016: 2428-2433.
- [19] Chen B, Zha Y F, Li Y Q, *et al.* Person re-identification based on convolutional neural network discriminative feature learning [J]. Acta Optica Sinica, 2018, 38(7): 0720001.
陈兵, 查宇飞, 李运强, 等. 基于卷积神经网络判别特征学习的行人重识别[J]. 光学学报, 2018, 38 (7): 0720001.
- [20] Xie L X, Tian Q, Zhang B. Simple techniques make sense: feature pooling and normalization for image classification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(7): 1251-1264.
- [21] Kostinger M, Hirzer M, Wohlhart P, *et al.* Large scale metric learning from equivalence constraints [C]//2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 16-21, 2012, Providence, RI. New York: IEEE, 2012: 2288-2295.
- [22] Weinberger K Q, Saul L K. Distance metric learning for large margin nearest neighbor classification [J]. The Journal of Machine Learning Research, 2009, 10: 207-244.
- [23] Khamis S, Kuo C H, Singh V K, *et al.* Joint learning for attribute-consistent person re-

- identification[M] // Agapito L, Bronstein M, Rother C. Computer vision-ECCV 2014 workshops. Lecture notes in computer science. Cham: Springer, 2015, 8927: 134-146.
- [24] Ye M, Li J W, Ma A J, *et al.* Dynamic graph co-matching for unsupervised video-based person re-identification [J]. IEEE Transactions on Image Processing, 2019, 28(6): 2976-2990.
- [25] Lü J, Chen W H, Li Q, *et al.* Unsupervised cross-dataset person re-identification by transfer learning of spatial-temporal patterns [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT. New York: IEEE, 2018: 7948-7956.
- [26] Cheng D S, Cristani M, Stoppa M, *et al.* Custom pictorial structures for re-identification [C] // The British Machine Vision Conference 2011, August 29-September 2, 2011, Dundee. [S.l.]: BMVC Press, 2011: 68.
- [27] Yi D, Lei Z, Liao S C, *et al.* Deep metric learning for person re-identification [C] // 2014 22nd International Conference on Pattern Recognition, August 24-28, 2014, Stockholm, Sweden. New York: IEEE, 2014: 34-39.
- [28] Li W, Zhao R, Xiao T, *et al.* DeepReID: deep filter pairing neural network for person re-identification [C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 23-28, 2014, Columbus, OH, USA. New York: IEEE, 2014: 152-159.
- [29] Yan Y C, Ni B B, Song Z C, *et al.* Person re-identification via recurrent feature aggregation[M] // Leibe B, Matas J, Sebe N, *et al.* Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9910: 701-716.
- [30] Wu L, Shen C H, van den Hengel A. Deep recurrent convolutional networks for video-based person re-identification: an end-to-end approach [J/OL]. (2016-06-12)[2018-12-30]. <https://arxiv.org/abs/1606.01609>.
- [31] Dai J, Zhang P P, Wang D, *et al.* Video person re-identification by temporal residual learning[J]. IEEE Transactions on Image Processing, 2019, 28(3): 1366-1377.
- [32] You J J, Wu A C, Li X, *et al.* Top-push video-based person re-identification[C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 1345-1353.
- [33] Yager R R. On ordered weighted averaging aggregation operators in multicriteria decisionmaking [J]. IEEE Transactions on Systems, Man, and Cybernetics, 1988, 18(1): 183-190.
- [34] Wang T Q, Gong S G, Zhu X T, *et al.* Person re-identification by video ranking[M] // Fleet D, Pajdla T, Schiele B, *et al.* Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8692: 688-703.
- [35] Hirzer M, Beleznai C, Roth P M, *et al.* Person re-identification by descriptive and discriminative classification [M] // Heyden A, Kahl F. Image analysis. Lecture notes in computer science. Berlin, Heidelberg: Springer, 2011, 6688: 91-102.
- [36] Jia Y Q, Shelhamer E, Donahue J, *et al.* Caffe: convolutional architecture for fast feature embedding [J/OL]. (2014-06-20)[2019-01-01]. <https://arxiv.org/abs/1408.5093>.
- [37] Kodirov E, Xiang T, Fu Z Y, *et al.* Person re-identification by unsupervised l_1 graph learning[M] // Leibe B, Matas J, Sebe N, *et al.* Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 178-195.
- [38] Karanam S, Li Y, Radke R J. Person re-identification with discriminatively trained viewpoint invariant dictionaries [C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 4516-4524.
- [39] Karanam S, Li Y, Radke R J. Sparse re-id: block sparsity for person re-identification[C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 33-40.
- [40] Zhao R, Ouyang W L, Wang X G. Unsupervised saliency learning for person re-identification [C] // 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 23-28, 2013, Portland, OR, USA. New York: IEEE, 2013: 3586-3593.
- [41] Cho Y J, Yoon K J. Improving person re-identification via pose-aware multi-shot matching [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 1354-1362.
- [42] Khan F M, Bremond F. Unsupervised data association for metric learning in the context of multi-shot person re-identification [C] // 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), August 23-26, 2016, Colorado Springs, CO, USA. New York:

- IEEE, 2016: 256-262.
- [43] Lisanti G, Masi I, Bagdanov A D, *et al.* Person re-identification by iterative re-weighted sparse ranking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(8): 1629-1642.
- [44] Zheng L, Bie Z, Sun Y F, *et al.* MARS: a video benchmark for large-scale person re-identification [M]//Leibe B, Matas J, Sebe N, *et al.* Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9910: 868-884.
- [45] Zhou Z, Huang Y, Wang W, *et al.* See the forest for the trees: joint spatial and temporal recurrent neural networks for video-based person re-identification [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI. New York: IEEE, 2017: 6776-6785.