JIG
JOURNAL OF IMAGE AND GRAPHICS
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

2020
07
VOL.25
ISSN1006-8961
CN11-3758/TB

俯视深度头肩序列识别 P1393

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第25卷第7期（总第291期）
2020年7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
中文核心期刊

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 顾行发
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮　　编 100190
电子信箱 jig@radi.ac.cn
电　　话 010-58887035
网　　址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218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 Gu Xingfa
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@radi.ac.cn
Telephone 010-58887035
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

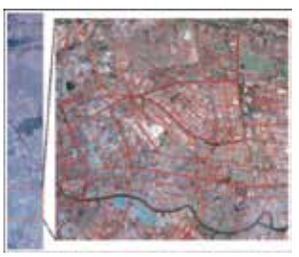
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

2020年7月 第25卷 第7期（总第291期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

目次

非真实感绘制技术研究现状与展望(第1283页)

图卷积网络下牙齿种子点自动选取(第1481页)



迁移学习下高分快视数据道路快速提取(第1501页)

Volume 25, Number 7
Published July 16, 2020

CONTENTS
JOURNAL OF IMAGE AND GRAPHICS

Research status and progress of non-photorealistic rendering technology(P1283)

Automatic selection of tooth seed point by graph convolutional network(P1481)

Rapid road extraction from quick view imagery of high-resolution satellites combined with transfer learning(P1501)

中图法分类号:TP391　　**文献标识码**:A　　**文章编号**:1006-8961(2020)07-1393-15

论文引用格式:Wang X N, Liu C H, Qi G Q and Zhang S Q. 2020. Person re-identification based on top-view depth head and shoulder sequence. Journal of Image and Graphics, 25(07):1393-1407(王新年,刘春华,齐国清,张世强. 2020. 俯视深度头肩序列行人再识别. 中国图象图形学报, 25(07):1393-1407)[DOI:10.11834/jig.190608]

俯视深度头肩序列行人再识别

王新年[1],刘春华[1],齐国清[1],张世强[1,2]

1. 大连海事大学, 大连　116026; 2. 华录智达科技有限公司, 大连　116023

摘　要:**目的**　行人再识别是指在一个或者多个相机拍摄的图像或视频中实现行人匹配的技术,广泛用于图像检索、智能安保等领域。按照相机种类和拍摄视角的不同,行人再识别算法可主要分为基于侧视角彩色相机的行人再识别算法和基于俯视角深度相机的行人再识别算法。在侧视角彩色相机场景中,行人身体的大部分表观信息可见;而在俯视角深度相机场景中,仅行人头部和肩部的结构信息可见。现有的多数算法主要针对侧视角彩色相机场景,只有少数算法可以直接应用于俯视角深度相机场景中,尤其是低分辨率场景,如公交车的车载飞行时间(time of flight,TOF)相机拍摄的视频。因此针对俯视角深度相机场景,本文提出了一种基于俯视深度头肩序列的行人再识别算法,以期提高低分辨率场景下的行人再识别精度。**方法**　对俯视深度头肩序列进行头部区域检测和卡尔曼滤波器跟踪,获取行人的头部图像序列,构建头部深度能量图组(head depth energy map group,HeDEMaG),并据此提取深度特征、面积特征、投影特征、傅里叶描述子和方向梯度直方图(histogram of oriented gradient,HOG)特征。计算行人之间头部深度能量图组的各特征之间的相似度,再利用经过模型学习所获得的权重系数对各特征相似度进行加权融合,从而得到相似度总分,将最大相似度对应的行人标签作为识别结果,实现行人再识别。**结果**　本文算法在公开的室内单人场景 TVPR(top view person re-identification)数据集、自建的室内多人场景 TDPI-L(top-view depth based person identification for laboratory scenarios)数据集和公交车实际场景 TDPI-B(top-view depth based person identification for bus scenarios)数据集上进行了测试,使用首位匹配率(rank-1)、前 5 位匹配率(rank-5)、宏 F1 值(macro-F1)、累计匹配曲线(cumulative match characteristic, CMC)和平均耗时等 5 个指标来衡量算法性能。其中,rank-1、rank-5 和 macro-F1 分别达到 61%、68% 和 67% 以上,相比于典型算法至少提高了 11%。**结论**　本文构建了表达行人结构与行为特征的头部深度能量图组,实现了适合低分辨率行人的多特征表达;提出了基于权重学习的相似度融合,提高了识别精度,在室内单人、室内多人和公交车实际场景数据集中均取得了较好的效果。

关键词:深度相机;俯视深度头肩序列;头部深度能量图组;相似度权重学习;行人再识别

Person re-identification based on top-view depth head and shoulder sequence

Wang Xinnian[1], Liu Chunhua[1], Qi Guoqing[1], Zhang Shiqiang[1,2]

1. Dalian Maritime University, Dalian　116026, China; 2. Hualuzhida Technology Co., Ltd., Dalian　116023, China

Abstract: **Objective** Person reidentification is an important task in video surveillance systems with a goal to establish the correspondence among images or videos of a person taken from different cameras at different times. In accordance with camera types, person re-identification algorithms can be divided into RGB camera-based and depth camera-based ones. RGB

收稿日期:2019-12-02;**修回日期**:2020-01-20;**预印本日期**:2020-01-27

基金项目:大连市科技创新基金项目(2019J12GX036)

camera-based algorithms are generally based on the appearance characteristics of clothes, such as color and texture. Their performances are greatly affected by external conditions, such as illumination variations. On the contrary, depth camera-based algorithms are minimally affected by lighting conditions. Person re-identification algorithms can also be divided into side view-oriented and vertical view-oriented algorithms according to camera-shooting angle. Most body parts can be seen in side-view scenarios, whereas only the plan view of head and shoulders can be seen in vertical-view scenarios. Most existing algorithms are for side-view RGB scenarios, and only a few of them can be directly applied to top-view depth scenarios. For example, they have poor performance in the case of bus-mounted low-resolution depth cameras. Our focus is on person re-identification on depth head and shoulder sequences. **Method** The proposed person re-identification algorithm consists of four modules, namely, head region detection, head depth energy map group (HeDEMaG) construction, HeDEMaG-based multifeature representation and similarity computation, and learning-based score-level fusion and person re-identification. First, the head region detection module is to detect each head region in every frame. The pixel value in a depth image represents the distance between an object and the camera plane. The range that the height of a person distributes is used to roughly segment the candidate head regions. A frame-averaging model is proposed to compute the distance between floor and the camera plane for determining the height of each person with respect to floor. The person's height can be computed by subtracting floor values from the raw frame. The circularity ratio of a head region is used to remove nonhead regions from the candidate regions because the shape of a real head region is similar to a circle. Second, the HeDEMaG construction module is to describe the structural and behavioral characteristics of a walking person's head. Kalman filter and Hungarian matching method are used to track multiple persons' heads in each frame. In the walking process, the head direction may change with time. A principal component analysis (PCA) based method is used to normalize the direction of a person's head regions. Each person's normalized head image sequence is uniformly divided into R_t groups in time order to capture the structural and behavioral characteristics of a person's head in local and overall time periods. The average map of each group is called the head depth energy map, and the set of the head depth energy maps is named as HeDEMaG. Third, the HeDEMaG-based multifeature representation and similarity computation module is to extract features and compute the similarity between the probe and gallery set. The depth, area, projection maps in two directions, Fourier descriptor, and histogram of oriented gradient (HOG) feature of each head depth energy map in HeDEMaG are proposed to represent a person. The similarity on depth is defined as the ratio of the depth difference to the maximum difference between the probe and gallery set. The similarity on area is defined as the ratio of the area difference to the maximum difference between the probe and gallery set. The similarities on projections, Fourier descriptor, and HOG are computed by their correlation coefficients. Fourth, the learning-based similarity score-level fusion and person re-identification module is to identify persons according to the similarity score that is defined as a weighted version of the above-mentioned five similarity values. The fusing weights are learned from the training set by minimizing the cost function that measures the error rate of recognition. In the experiments, we use the label of the top one image in the ranked list as the predicted label of the probe. **Result** Experiments are conducted on a public top view person re-identification (TVPR) dataset and two self-built datasets to verify the effectiveness of the proposed algorithm. TVPR consists of videos recorded indoors using a vertical RGB-D camera, and only one person's walking behavior is recorded. We establish two datasets, namely, top-view depth based person identification for laboratory scenarios (TDPI-L) and top-view depth based person identification for bus scenarios (TDPI-B), to verify the performance on multiple persons and real-world scenarios. TDPI-L is composed of videos captured indoors by depth cameras, and more than two persons' walking is recorded in each frame. TDPI-B consists of sequences recorded by bus-mounted low-resolution time of flight (TOF) cameras. Five measures, namely, rank-1, rank-5, macro-F1, cumulative match characteristic (CMC) and average time are used to evaluate the proposed algorithm. The rank-1, rank-5, and macro-F1 of the proposed algorithm are above 61%, 68%, and 67%, respectively, which are at least 11% higher than those of the state-of-the-art algorithms. The ablation studies and the effects of tracking algorithms and parameters on the performance are also discussed. **Conclusion** The proposed algorithm is to identify persons in head and shoulder sequences captured by depth cameras from top views. HeDEMaG is proposed to represent the structural and behavioral characteristics of persons. A learning-based fusing weight-computing method is proposed to avoid parameter fine tuning and improve the recognition accuracy. Experimental results show that proposed algorithm outperforms the state-of-the-art algorithms on public available

indoor videos and real-world low-resolution bus-mounted videos.

Key words: depth camera; top view depth head and shoulder sequence; head depth energy map group (HeDEMaG); similarity fusion weights learning; person re-identification

0 引　言

行人再识别是指在一个或者多个相机拍摄的图像或者视频中实现行人匹配的技术，以判断某个相机中的某位行人是否再次出现，广泛应用于图像检索、智能监控和智能安保等领域，在社会管理、突发事件重构等方面具有广阔的应用前景。

按照相机种类的不同，行人再识别算法可分为基于彩色相机的行人再识别算法和基于深度相机的行人再识别算法。基于彩色相机的行人再识别算法一般是根据行人衣服的颜色（Liao 等，2015；Zheng 等，2017；Kim 等，2017；蒋建国 等，2019）和纹理（Matsukawa 等，2016；Nguyen 等，2018，2019）等外观特征进行研究。Liao 等人（2015）提出了局部最大概率（local maximal occurrence，LOMO）特征，将色调饱和度明度空间颜色直方图和尺度不变三值模式纹理直方图（scale invariant local ternary pattern，SILTP）（Liao 等，2010）在水平方向进行最大池化，计算简单且表达能力强。Matsukawa 等人（2016）提出了多层高斯（Gaussian of Gaussian，GOG）特征，使用高斯分布对图像的局部区域建模，较好地描述了纹理特征。但是该类算法受外界环境的影响较大，无法解决不同行人穿着相似的问题。

深度图像的像素值表征着场景中某一点到相机平面的距离。与彩色图像相比，深度图像具有受外界环境干扰较小的优势。因此，基于深度相机的行人再识别算法能够解决不同行人穿着相似的问题。基于深度相机的行人再识别算法按照提取特征种类的不同可分为人工设计法（Han 和 Bhanu，2006；Sivapalan 等，2011；Hofmann 等，2012；Paolanti 等，2018；Imani 和 Soltanizadeh，2019）、骨架法（Barbosa 等，2012；Munaro 等，2014c）和点云法（Munaro 等，2014a，2014b；Wu 等，2017）。Han 和 Bhanu（2006）提出了步态能量图（gait energy image，GEI），通过将一系列的行人图像进行平均，从而获得步态信息实现行人再识别。Sivapalan 等人（2011）在 GEI 的基础上进行了扩展，利用深度信息将 2D 特征变为 3D 特征，提出了步态能量体积（gait energy volume，GEV），增强了识别性能。Hofmann 等人（2012）根据提取方向梯度直方图（histogram of oriented gradient，HOG）特征的思想，提出了深度梯度直方能量图（depth gradient histogram energy image，DGHEI），得到了较好的识别结果。Paolanti 等人（2018）根据俯视场景中行人的特点，将在深度图像中提取的头肩周长、宽度和面积等作为俯视角深度（top view depth，TVD）特征，与彩色图像中利用 HSV 直方图提取的俯视角颜色直方图（top view color histogram，TVH）特征结合，从而形成俯视角深度与颜色直方图（top view depth and color histogram，TVDH）特征，加强了行人特征的表达能力。Imani 和 Soltanizadeh（2019）将伽柏（Gabor）特征分别与局部二值模式（local binary pattern，LBP）、局部导数模式（local derivative pattern，LDP）组合形成伽柏局部二值模式（Gabor local binary pattern，GLBP）、伽柏局部导数模式（Gabor local derivative pattern，GLDP）特征，提高了识别性能。Barbosa 等人（2012）根据行人的骨架结构，选取人体各关节点之间的距离作为特征。Munaro 等人（2014b）将不同视图下的点云进行融合，利用点云匹配的方法实现行人再识别。

按照相机拍摄角度的不同，行人再识别算法可分为基于侧视角的行人再识别算法和基于俯视角的行人再识别算法，如图 1 所示。现有多数算法主要是针对侧视角彩色相机场景进行研究，只有少数算法（Paolanti 等，2018）能直接用于俯视角深度相机场景，尤其是低分辨率场景。例如在公交车中，为有效降低遮挡和光照影响，常将低分辨率的车载飞行时间（time of flight，TOF）相机安装到高处进行垂直式拍摄。

在垂直式深度相机拍摄的俯视深度头肩序列中，行人的身体特征大幅度减少，头部特征较为突出，利用骨架特征或者点云特征的方法不再适用。但其优点是行人间的遮挡现象大幅度降低，利于多人的区分与跟踪。

针对上述情况，本文以公交车等实际监控系统

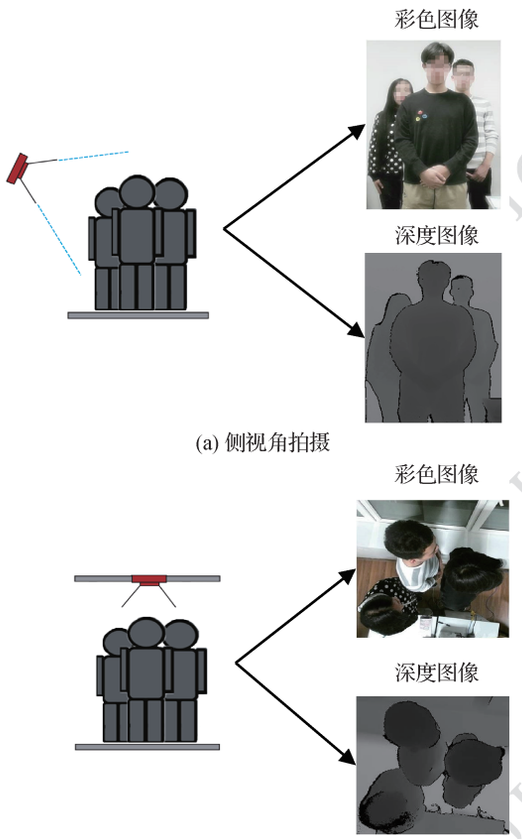


图 1　两种视角下拍摄的行人彩色图和深度图示意

Fig. 1　Illustration of the depth and RGB images of persons in two shooting views((a) side view; (b) top view)

为背景,以垂直式安装的深度相机作为视频源,提出了一种面向俯视深度头肩序列的行人再识别算法,以期提高低分辨率场景下的行人再识别精度。具体而言,本文主要工作有:

1)提出了用以表征结构特征和行为特征的头部深度能量图组,即将头部图像矫正并按照时间顺序分组计算头部深度能量图;

2)提出了基于头部深度能量图组的行人表达特征:深度特征、面积特征、投影特征、傅里叶描述子和 HOG 特征;

3)提出了基于权重学习的相似度融合与行人再识别算法。根据构建的模型学习权重系数,将头部深度能量图组的上述行人表达特征的相似度进行融合得到相似度总分,从而将最大相似度对应的行人作为识别结果,实现行人再识别。

1　算法流程

本文算法包括 4 个部分:头部区域检测、头部深度能量图组的构建、基于头部深度能量图组的行人多特征表达与相似度计算、基于权重学习的相似度融合与行人再识别,如图 2 所示。

首先对俯视深度头肩序列中的行人头部区域进

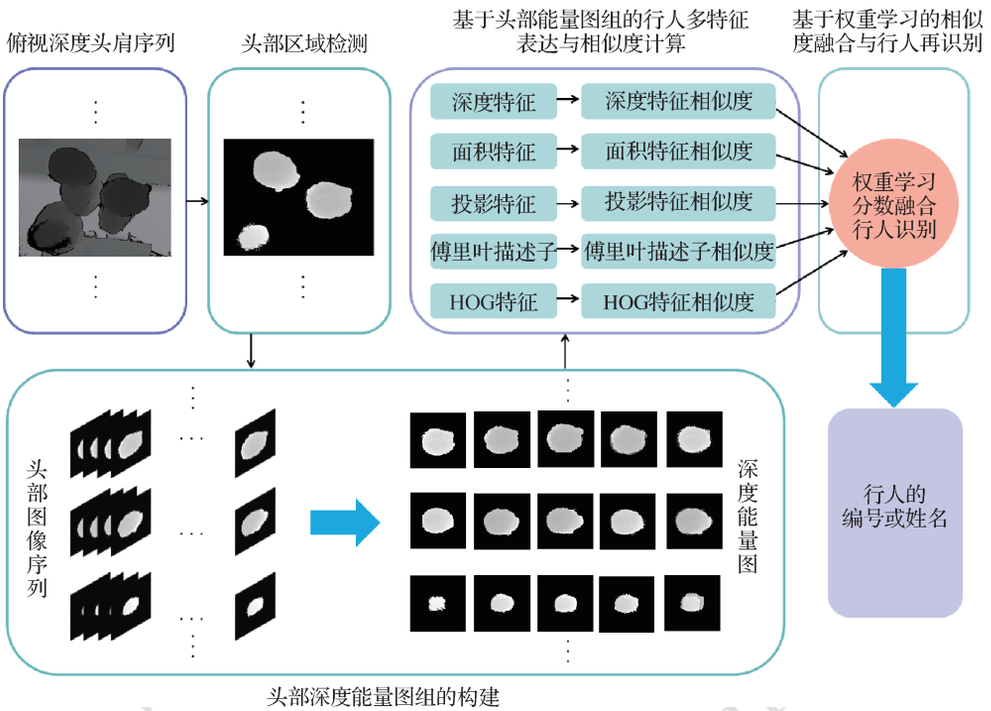


图 2　本文算法流程图

Fig. 2　Flow chart of the proposed algorithm

行检测。然后对行人进行跟踪，得到每位行人的头部图像序列并分组构建头部深度能量图。再提取头部深度能量图组的深度特征、面积特征、投影特征、傅里叶描述子和 HOG 特征，并计算行人间头部深度能量组对应特征之间的相似度。最后利用模型学习所获得的权重系数对多特征相似度进行加权融合，从而得到相似度总分，将最大相似度对应的行人作为结果，实现行人再识别。

2　头部区域检测

2.1　头部区域粗筛选

深度图像中的像素值表示物体到相机平面之间的距离。如果事先知道待检测物体到地面的高度范围，就可以确定深度图像中待检测物体的深度值范围。因此利用地面模型和行人的高度阈值进行分割，初步确定头部候选区域。地面模型 $G(x,y)$ 的建立过程为

$$G(x,y)=\frac{\sum_{l=1}^{L} g_l(x,y)}{L} \tag{1}$$

式中，$\boldsymbol{g}_l$ 表示第 l 帧地面深度图像，L 表示建立地面模型 $\boldsymbol{G}$ 的总帧数。地面深度图像是指在没有行人的情况下拍摄的图像。

设俯视深度头肩图像为 $\boldsymbol{J}_0$，则粗分割后的头肩区域候选图像 $\boldsymbol{J}_1$ 为

$$J_1(x,y)=\begin{cases} J_0(x,y) & G(x,y)-H_{\max}\leqslant J_0(x,y)\leqslant G(x,y)-H_{\min} \\ 0 & \text{其他} \end{cases} \tag{2}$$

式中，$H_{\min}$ 和 $H_{\max}$ 分别表示深度图像中行人肩膀高度的最小值和最大值，由多次实验确定。处理后的头肩区域候选图像一般存在着噪声和空洞，所以采用中值滤波和形态学方法去除干扰区域。

根据头部区域是极值区域的特点，确定 $J_1(x,y)$ 中的头部候选区域。在头肩区域候选图像 $J_1(x,y)$ 中，对每一个深度值大于零的区域采用最大极值稳定区域法（maximally stable extremal regions，MSER）进行再次分区，若分区数目小于 2，则该区域不可能既有头部和肩部区域（头部和肩部有明显的深度差），不予考虑；否则将各个分区按照由大到小面积排序，选择前 3 个分区中均值最小的分区作为头部候选区域。

2.2　头部区域精定位

利用头部区域是类圆形的特点进一步确定头部区域。设头肩区域候选图像 $J_1(x,y)$ 中有 E 个头部候选区域，对每个候选区域 $\boldsymbol{U}_e(e=1,2,\cdots,E)$ 计算类圆形判定参数 C_{U_e}，根据类圆形判定参数判断该区域 $\boldsymbol{U}_e$ 是否为行人头部区域 $\boldsymbol{V}_f(f=1,2,\cdots,F)$。$F$ 表示精确定位后的头部区域个数，其小于或者等于头部候选区域个数 E。若 S_{U_e} 表示候选区域 $\boldsymbol{U}_e$ 的面积，a_{U_e} 表示 $\boldsymbol{U}_e$ 的直径，则类圆形判定参数 C_{U_e} 为

$$C_{U_e}=\frac{S_{U_e}}{\pi\times\left(\frac{a_{U_e}}{2}\right)^2} \tag{3}$$

式中，C_{U_e} 的值越接近于 1，表明该区域越接近圆形。

3　头部深度能量图组的构建

当地面高度和行走方向变化时，深度视频中行人的头部区域也会发生相应变化。例如在上台阶时，行人头部与深度相机之间的距离减小，头部区域的深度值也会随之减小。使用单帧图像进行特征提取时，易受到噪声的影响，具有偶然性。然而使用头部序列的平均图像进行特征提取时，忽略了头部区域的变化。因为头部区域的变化比较缓慢，所以在较短的时间内可认为头部区域是基本不变的。综合上述分析，本文提出了构建头部深度能量图组，以表征头部区域的结构特征和行为特征，具体过程如图 3 所示。

首先对俯视深度头肩序列进行头部区域检测，确定图像中行人头部区域的位置，然后构建头部深度能量图组。头部深度能量图组的构建过程为：先使用卡尔曼滤波器获取每位行人的头部图像，再将头部区域进行矫正，最后将矫正后的图像按照时间顺序形成序列，以分组构建头部深度能量图。

3.1　头部图像序列的获取

头部图像序列的获取由 5 个部分组成：建立卡尔曼滤波器模型、当前帧行人位置预测、匹配行人、更新头部图像序列和更新卡尔曼滤波器，具体过程为：

1）使用卡尔曼滤波器对深度视频中的每位行人建模跟踪，建模过程为

状态方程：　$\boldsymbol{X}_k=\boldsymbol{A}\boldsymbol{X}_{k-1}+\boldsymbol{W}_{k-1}$　　(4)

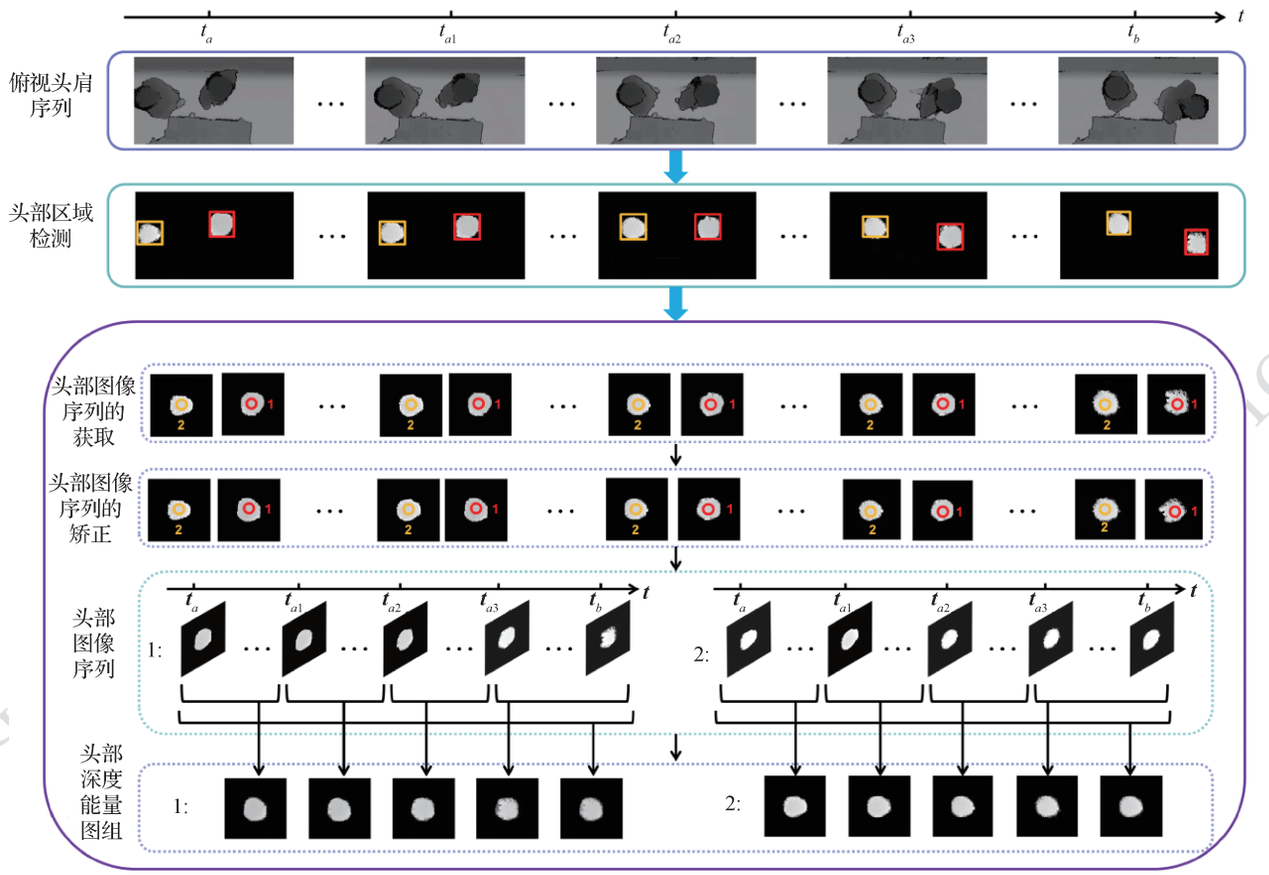


图 3　头部深度能量图组(HeDEMaG)的构建过程

Fig. 3　Constructing process of the head depth energy map group(HeDEMaG)

观测方程：　$\boldsymbol{Z}_k = \boldsymbol{H}\boldsymbol{X}_k + \boldsymbol{V}_k$　　(5)

式中，$\boldsymbol{X}_k$ 是 k 时刻的行人的运动状态，$\boldsymbol{X}_{k-1}$ 是 $k-1$ 时刻的行人的运动状态，$\boldsymbol{A}$ 是运动状态转移矩阵，$\boldsymbol{H}$ 是观测矩阵，$\boldsymbol{W}_{k-1}$ 和 $\boldsymbol{V}_k$ 为互不相关的高斯白噪声，$\boldsymbol{Z}_k$ 是 k 时刻观测到的头部区域的中心位置。

深度视频中相邻两帧图像之间的时间间隔 Δt 非常短，所以可以近似地认为相邻两帧图像间行人是匀速运动的，则运动状态向量 $\boldsymbol{X}_k$ 由 4 个分量组成

$$\boldsymbol{X}_k = [x_k, y_k, v_{xk}, v_{yk}]^{\mathrm{T}} \tag{6}$$

式中，x_k 与 y_k 分别表示 k 时刻头部区域的中心位置在 x 轴和 y 轴上的坐标；v_{xk} 与 v_{yk} 分别表示 k 时刻行人在 x 轴和 y 轴上的运动速度。

设观测向量 $\boldsymbol{Z}_k = [x_k, y_k]^{\mathrm{T}}$，则运动状态转移矩阵 $\boldsymbol{A}$ 与观测矩阵 $\boldsymbol{H}$ 分别为

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{7}$$

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{8}$$

高斯白噪声 $\boldsymbol{W}_{k-1}$ 和 $\boldsymbol{V}_k$ 的协方差矩阵 $\boldsymbol{Q}$ 和 $\boldsymbol{R}$ 分别为

$$\boldsymbol{Q} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{9}$$

$$\boldsymbol{R} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{10}$$

2)根据运动模型预测行人在当前时刻的状态和位置，预测过程为

$$\hat{\boldsymbol{X}}_k = \boldsymbol{A}\hat{\boldsymbol{X}}_{k-1} \tag{11}$$

$$\hat{\boldsymbol{Z}}_k = \boldsymbol{H}\hat{\boldsymbol{X}}_k \tag{12}$$

式中，$\hat{\boldsymbol{X}}_{k-1}$ 表示 $k-1$ 时刻行人运动状态的最优估计值，$\hat{\boldsymbol{X}}_k$ 表示 k 时刻行人运动状态的预测值，$\hat{\boldsymbol{Z}}_k$ 表示 k 时刻行人位置的预测值。

3)利用头部区域检测获得当前帧中的行人位置，计算观测位置和预测位置之间的欧氏距离，通过

匈牙利匹配法进行数据关联(Bewley 等,2016),确认每位行人在当前帧的观测位置 $\boldsymbol{Z}_k$。

若当前跟踪到了 B 位行人,设第 $b(b=1,2,\cdots,B)$ 位行人在 k 时刻的预测坐标为 $\hat{\boldsymbol{Z}}_{bk}=[\hat{x}_{bk},\hat{y}_{bk}]^{\mathrm{T}}$,检测到的某个头部区域 $\boldsymbol{V}_f(f=1,2,\cdots,F)$ 的中心坐标为 (x_f,y_f),则它们之间的欧氏距离为

$$d_{fb}=\sqrt{(x_f-\hat{x}_{bk})^2+(y_f-\hat{y}_{bk})^2} \tag{13}$$

计算当前帧头部区域检测到的 F 个中心坐标与 B 位行人在当前帧的预测坐标之间的欧氏距离矩阵 $\boldsymbol{D}$。若观测位置 f 与行人 b 的预测位置之间的欧氏距离小于距离阈值 d_m,则认为观测位置 f 有可能是行人 b 在当前帧的位置,即 $m_{fb}=1$,否则为不可能,即 $m_{fb}=0$。将欧氏距离矩阵 $\boldsymbol{D}$ 中的所有值与 d_m 进行比较获得可能性矩阵 $\boldsymbol{M}_{\mathrm{p}}$,即对于 $\boldsymbol{M}_{\mathrm{p}}$ 中的每个点 m_{fb} 有

$$m_{fb}=\begin{cases}1 & d_{fb}<d_m\\ 0 & \text{其他}\end{cases} \tag{14}$$

则可能性矩阵为

$$\boldsymbol{M}_{\mathrm{p}}=\begin{bmatrix} m_{11} & \cdots & m_{1b} & \cdots & m_{1B}\\ & & \vdots & & \\ m_{f1} & \cdots & m_{fb} & \cdots & m_{fB}\\ & & \vdots & & \\ m_{F1} & \cdots & m_{Fb} & \cdots & m_{FB}\end{bmatrix} \tag{15}$$

对矩阵 $\boldsymbol{M}_{\mathrm{p}}$ 采用匈牙利匹配法将 F 个观测位置与 B 个预测位置对应的行人轨迹相关联,从而确定每位行人在当前帧的观测位置 $\boldsymbol{Z}_k$。

4)获取观测位置 $\boldsymbol{Z}_k$ 处的头部区域图像,然后将其以时间顺序关联到头部图像序列中,更新每位行人的头部图像序列。

5)更新卡尔曼滤波器,以便进行下一帧的跟踪。更新过程为

$$\hat{\boldsymbol{X}}_k=\hat{\boldsymbol{X}}_k+\boldsymbol{K}_k(\boldsymbol{Z}_k-\hat{\boldsymbol{Z}}_k) \tag{16}$$

式中,$\boldsymbol{K}_k$ 表示 k 时刻的卡尔曼增益矩阵。

3.2　头部图像序列的矫正

在行走的过程中,头部方向可能会发生变化,这个变化在一定范围内是随机的,因此需要对头部图像序列中的每幅图像进行方向矫正。图 4 展示了某位行人的头部区域方向 θ 随时间 t 变化的情况。

本文选择主成分分析(principal component analysis,PCA)算法进行头部区域的方向矫正,过程为:

1)将头部区域图像进行二值化;

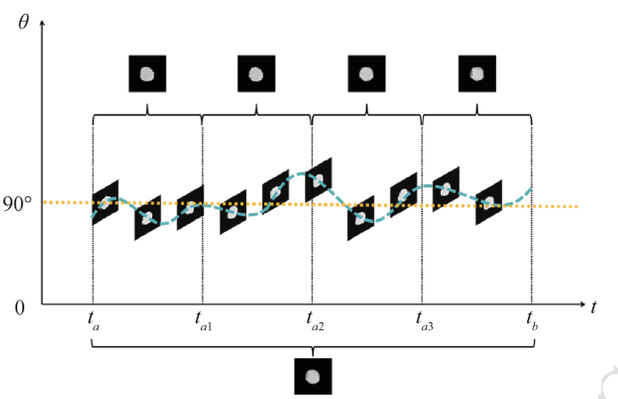


图 4　矫正前头部序列图

Fig. 4　Head sequence before normalizing directions

2)设二值化的头部区域图像中共有 N_{w} 个白色像素,获取 N_{w} 个白色像素的坐标值,按照自左向右自上到下的顺序形成点集 $\boldsymbol{P}_{\mathrm{w}}$,$\boldsymbol{P}_{\mathrm{w}}\in\mathbf{R}^{2\times N_{\mathrm{w}}}$;

3)使用 PCA 算法计算点集 $\boldsymbol{P}_{\mathrm{w}}$ 的特征向量矩阵 $\boldsymbol{E}_{\mathrm{w}}=\begin{bmatrix}e_{11} & e_{12}\\ e_{21} & e_{22}\end{bmatrix}$;

4)根据矩阵 $\boldsymbol{E}_{\mathrm{w}}$ 中的 e_{11} 计算旋转角度

$$\alpha=\arccos(e_{11})\times\frac{180}{\pi} \tag{17}$$

5)利用计算出的旋转角度 α 对头部区域图像进行方向矫正。

3.3　构建头部深度能量图组

在运动的过程中,行人的头部区域会受到地面高度和行走路线的影响,使得不同时间段头部区域的形状、大小和方向随之变化。因此本文提出了构建头部图像序列的深度能量图组,以表征不同时间段头部区域的结构特征和行为特征。

若在 $t_a\sim t_b$ 时间段内跟踪到某位行人的头部区域,则划分的 R_t 个时间组为:先将 $t_a\sim t_b$ 平均分为 R_t 个时间组,再将总时间段 $t_a\sim t_b$ 作为第 R_t 组时间,如图 4 所示。若 R_t 为 1,则只有 1 个时间组,即 $t_a\sim t_b$。设 t 时刻行人的头部图像为 $\boldsymbol{D}_t$,则行人的第 $r(r=1,2,\cdots,R_t)$ 幅头部深度能量图为

$$DEI_r(x,y)=\begin{cases}\dfrac{1}{\Delta t_0}\displaystyle\sum_{t=t_a+(r-1)\Delta t_0}^{t_a+r\Delta t_0-1}D_t(x,y) & r<R_t\\ \dfrac{1}{t_b-t_a+1}\displaystyle\sum_{t=t_a}^{t_b}D_t(x,y) & r=R_t\end{cases} \tag{18}$$

式中,$\boldsymbol{DEI}_r$ 表示第 r 幅头部深度能量图,Δt_0 表示

$t_a \sim t_b$ 均匀划分的时间间隔。

将 R_t 个时间组的头部图像序列均进行上述计算，使得每位行人均有 R_t 幅头部深度能量图，从而表达了局部和整体时间段中行人头部区域的结构特征和行为特征。

4 基于头部深度能量图组的行人多特征表达与相似度计算

根据深度视频中行人的特点，提取了深度特征、面积特征、投影特征、傅里叶描述子和 HOG 特征等 5 种不同角度的特征，以更好地表达低分辨率视频中的行人特征。再计算待识别行人与候选者之间的各特征相似度，以衡量不同角度下的行人相似度。

4.1 计算深度特征及相似度

深度值表示行人到深度相机所处平面的距离。当行人越高时，距离相机越近，深度值就越小。因此，深度值可以反映行人的身高信息。为了去除噪声的干扰，提取头部深度能量图中出现次数最多的深度值来表征行人的身高信息。具体过程为：先计算头部深度能量图的直方图，再将直方图中最大概率值对应的深度值作为特征。

设第 $q(q = 1,2,\cdots,N)$ 位待识别行人与第 $m(m = 1,2,\cdots,M)$ 位候选者的第 r 幅头部深度能量图的深度特征分别为 $d_q(r)$ 和 $d_m(r)$，则深度特征相似度 $S_{\mathrm{d}}(r,q,m)$ 为

$$S_{\mathrm{d}}(r,q,m) = 1 - \frac{|d_q(r) - d_m(r)|}{d_{qM}(r)} \tag{19}$$

式中，$d_{qM}(r)$ 为待识别行人与所有候选者的第 r 幅头部深度能量图之间的深度特征之差的最大值。

4.2 计算面积特征及相似度

不同行人的头部区域的大小和面积不同，因此可以利用面积特征计算相似度。

设第 q 位待识别行人与第 m 位候选者的第 r 幅头部深度能量图的面积值分别为 $a_q(r)$ 和 $a_m(r)$，则面积特征相似度 $S_{\mathrm{a}}(r,q,m)$ 为

$$S_{\mathrm{a}}(r,q,m) = 1 - \frac{|a_q(r) - a_m(r)|}{a_{qM}(r)} \tag{20}$$

式中，$a_{qM}(r)$ 为待识别行人与所有候选者的第 r 幅头部深度能量图之间的面积特征之差的最大值。

4.3 计算投影特征及相似度

投影特征可以反映头部深度能量图在水平方向和垂直方向上的结构特征。设行人的第 r 幅头部深度能量图为 $\boldsymbol{DEI}_r$，其长度和宽度分别为 d_x 和 d_y，则 x 和 y 方向的投影为

$$P_{xr}(x) = \sum_{y=1}^{d_y} DEI_r(x,y) \tag{21}$$

$$P_{yr}(y) = \sum_{x=1}^{d_x} DEI_r(x,y) \tag{22}$$

将 x 方向和 y 方向的投影进行串联以表示投影特征 $P_r(z)$，记 $d_z = d_x + d_y$，则 $P_r(z) \in \mathbf{R}^{1\times d_z}$。采用相关系数来计算头部深度能量图投影特征之间的相似度。

设第 q 位待识别行人与第 m 位候选者的第 r 幅头部深度能量图的投影特征分别是 $P_{qr}(z)$ 和 $P_{mr}(z)$，对应的平均值分别是 $\overline{P_{qr}}$ 和 $\overline{P_{mr}}$，则投影特征相似度

$$S_{\mathrm{p}}(r,q,m) = \frac{\sum_{z=1}^{d_z}(P_{qr}(z) - \overline{P_{qr}})(P_{mr}(z) - \overline{P_{mr}})}{\sqrt{\sum_{z=1}^{d_z}(P_{qr}(z) - \overline{P_{qr}})^2 \sum_{z}(P_{mr}(z) - \overline{P_{mr}})^2}} \tag{23}$$

式中，相关系数 $S_{\mathrm{p}}(r,q,m)$ 的值处于 0 ～ 1 之间。根据相关系数的特点可知，$S_{\mathrm{p}}(r,q,m)$ 越接近于 1，两幅头部深度能量图越相似。

4.4 计算傅里叶描述子及相似度

傅里叶描述子是一种形状描述子，可以用低频分量近似地表达头部深度能量图的轮廓特征。提取傅里叶描述子的过程为：

1）将头部深度能量图进行二值化；

2）使用 Canny 算子提取二值化的头部深度能量图的外轮廓边缘点；

3）将 N_{c} 个边缘点的坐标值按照自左向右自上到下的顺序形成点集 $\boldsymbol{E}_{\mathrm{c}}$，$\boldsymbol{E}_{\mathrm{c}} \in \mathbf{R}^{2\times N_{\mathrm{c}}}$；

4）利用点集 $\boldsymbol{E}_{\mathrm{c}}$ 构建复数点集 $\boldsymbol{F}_{\mathrm{c}}$，$\boldsymbol{F}_{\mathrm{c}} \in \mathbf{C}^{1\times N_{\mathrm{c}}}$，$\mathbf{C}$ 为复数集。若 $\boldsymbol{E}_{\mathrm{c}}$ 中的某一个坐标点为 (e_x,e_y)，则复数点集 $\boldsymbol{F}_{\mathrm{c}}$ 中的对应点值为 $e_x + \mathrm{j}e_y$，j 是虚数单位；

5）对点集 $\boldsymbol{F}_{\mathrm{c}}$ 进行傅里叶变换，使用前 24 个分量来描述头部深度能量图的轮廓特征。

与式(23)类似，采用相关系数来计算第 q 位待识别行人与第 m 位候选者之间的第 r 幅头部深度能量图的傅里叶描述子相似度 $S_{\mathrm{f}}(r,q,m)$。

4.5　计算 HOG 特征及相似度

HOG 特征通过计算和统计头部深度能量图局部区域的梯度方向直方图来构成特征，以表达行人头部区域的结构特征和纹理特征。其中，细胞单元(cell)大小为 8×8 像素，一个图像块(block)由 4 个细胞单元(cell)组成。

与式(23)类似，采用相关系数来计算第 q 位待识别行人与第 m 位候选者之间的第 r 幅头部深度能量图的 HOG 特征相似度 $S_{\mathrm{h}}(r,q,m)$ 。

5　基于权重学习的相似度融合与行人再识别

为了提高识别精度，本文提出一种基于权重学习的相似度融合与行人再识别方案。通过对深度特征相似度、面积特征相似度、投影特征相似度、傅里叶描述子相似度和 HOG 特征相似度等 5 种特征相似度分配不同权重实现加权融合，从而将最大相似度对应的行人作为识别结果，实现行人再识别。具体过程为：

若第 $q(q=1,2,\cdots,N)$ 位待识别行人存在着 M 位候选者，则该行人与第 $m(m=1,2,\cdots,M)$ 位候选者第 $r(r=1,2,\cdots,R_t)$ 幅头部深度能量图之间的相似度组合 $\boldsymbol{S}_r(q,m)$ 为

$$\boldsymbol{S}_r(q,m)=[S_{\mathrm{d}}(r,q,m),S_{\mathrm{a}}(r,q,m),S_{\mathrm{p}}(r,q,m),S_{\mathrm{f}}(r,q,m),S_{\mathrm{h}}(r,q,m)]^{\mathrm{T}} \tag{24}$$

设多种特征相似度的权重向量为 $\boldsymbol{w}=[w_{\mathrm{d}},w_{\mathrm{a}},w_{\mathrm{p}},w_{\mathrm{f}},w_{\mathrm{h}}]^{\mathrm{T}}$ 。则行人再识别结果 $f_q(\boldsymbol{w})$ 为

$$f_q(\boldsymbol{w})=\arg\max_{m}\sum_{r=1}^{R_t}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_r(q,m) \tag{25}$$

式中，$f_q(\boldsymbol{w})$ 是第 q 位待识别行人的 R_t 组头部深度能量图的多特征相似度融合的最大值对应的行人编号，即为行人再识别的结果。

为了确定多种特征的相似度融合系数，使用训练集进行权重学习。多特征相似度权重向量的解

$$\boldsymbol{w}^*=\arg\min_{\boldsymbol{w}}\sum_{q=1}^{N}L_q(\boldsymbol{w}) \tag{26}$$

式中，$L_q(\boldsymbol{w})$ 表示第 q 位待识别行人的损失函数。为了求解 $\boldsymbol{w}^*$，借鉴铰链损失函数思想，构建对应的损失函数

$$L_q(\boldsymbol{w})=\sum_{m\neq y_q}\max\Big(0,\sum_{r=1}^{R_t}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_r(q,m)-\sum_{r=1}^{R_t}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_r(q,y_q)\Big) \tag{27}$$

则求导结果为

$$\nabla_w L_q=\sum_{m\neq y_q}1\Big(\sum_{r=1}^{R_t}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_r(q,m)-\sum_{r=1}^{R_t}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_r(q,y_q)>0\Big)\times\Big(\sum_{r=1}^{R_t}\boldsymbol{S}_r(q,m)-\sum_{r=1}^{R_t}\boldsymbol{S}_r(q,y_q)\Big) \tag{28}$$

式中，$1(\cdot)$ 是指示函数，若内部表达式为真，则结果为 1，否则为 0。所以权重向量 $\boldsymbol{w}$ 的更新为

$$\boldsymbol{w}=\boldsymbol{w}-\eta\sum_{q=1}^{N}\nabla_w L_q \tag{29}$$

η 是学习速率，取 0.001。

6　实验与分析

6.1　数据集

为了验证算法的有效性，采用 1 个公开的 TVPR(top view person re-identification)数据集(Paolanti 等，2018)和自建的 2 个数据集实验。

TVPR 是使用垂直式 RGB-D 相机在室内录制的数据集，包含 100 位行人的 23 个视频序列，训练集和测试集中的行人运动方向相反。空间分辨率为 640×480 像素，帧率为 30 帧/s，总长度大约为 2 000 s，同一时间内仅有一个人存在。相机的安装高度为 4 m，并可以覆盖 4.43 m×3.31 m 的区域。图 5 展示了 TVPR 数据集中的部分片段。

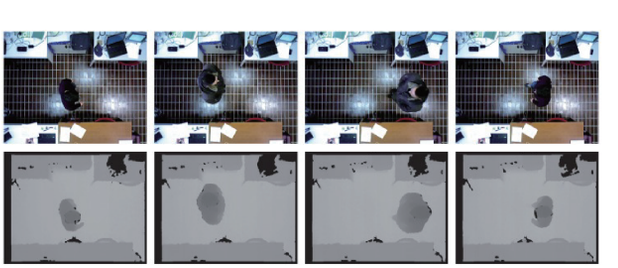
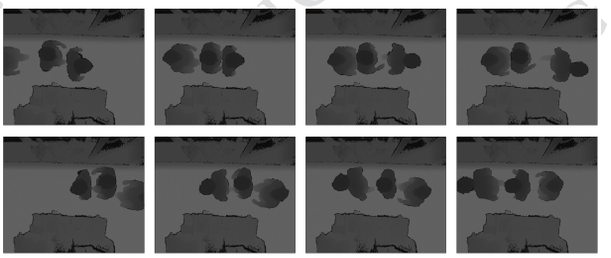


图 5　TVPR 数据集示例图

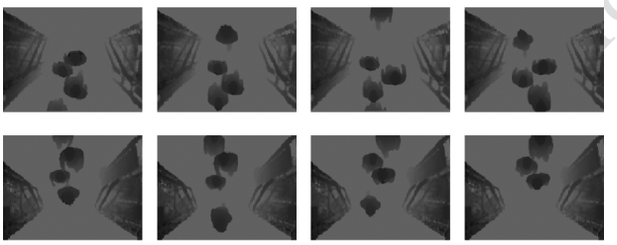
Fig. 5　Samples from TVPR dataset

TVPR 数据集是理想情况下的室内单人场景数据集。为了验证多人场景和实际场景中算法的性能，建立了两个数据集，一个是室内多人场景数据集(top-view depth based person identification for laboratory scenarios，TDPI-L)，另一个是公交车实际场景数

据集(top-view depth based person identification for bus scenarios, TDPI-B),如图 6 所示。实验采用深度相机,安装高度均为 2.5 m。



(a) TDPI-L



(b) TDPI-B

图 6 自建数据集部分片段示例图

Fig. 6 Samples from self-built dataset

((a) TDPI-L; (b) TDPI-B)

TDPI-L 是使用 Kinect v2 相机在实验室中录制的俯视角数据集,包含 31 位行人的 20 个视频序列,训练集和测试集中的行人运动方向相反。空间分辨率为 512×424 像素,帧率为 30 帧/s,每个视频长度在 20 ~ 60 s 之间,同一时间内至少有 2 个人存在。

TDPI-B 是使用两个 TOF 相机分别在公交车的进出口高处垂直式录制的数据集,包含 71 位行人的 50 个视频序列。训练集和测试集分别由不同的 TOF 相机录制,运动方向相反。空间分辨率为 160×120 像素,帧率为 30 帧/s,每个视频长度在 30 ~ 90 s 之间,同一时间内至少有 2 个人存在。

TDPI-B 数据集场景复杂,帧分辨率远低于 TVPR 和 TDPI-L 数据集中的帧分辨率,挑战性也更大。

6.2 评价指标

首位匹配率(rank-1)是指待查询的行人与在候选库中利用相似度最大策略匹配到的行人属于同一个人的概率,计算为

$$r_1 = \frac{\sum_{q=1}^{N} P_q}{N} \tag{30}$$

式中,N 为待查询行人的总数,P_q 表示查询的第 q 位行人是否匹配成功,成功时 $P_q = 1$,失败时 $P_q = 0$。类似地,可以考察前 k 位匹配率。前 k 位匹配率是指待查询的行人能在候选库中检索到的前 k 位行人中找到的概率。常用的前 k 位匹配率有 rank-1, rank-3, rank-5 以及 rank-10 等。

累计匹配曲线(cumulative match characteristic, CMC)也是一种常用的评价指标,横坐标为秩 k,纵坐标为第 k 匹配率 rank-k,曲线上的每个点表示不超过 k 个检索结果中找到正确结果的概率。CMC 曲线的起点与走势能够体现算法的优劣,当将两个算法的 CMC 曲线对比时,曲线整体越高则算法性能越好。

宏 F1 值(macro-F1)是用来衡量多分类模型精确度的一种指标,通过将每个类的 F1 值进行平均得到,兼顾了多分类模型的精确率和召回率。其中,F1 值是每个类的精确率和召回率的调和平均数。macro-F1 的值越大,算法的整体性能越好。

平均耗时是指使用多个测试数据进行实验的平均时间,不包含训练集的特征提取和参数学习时间。

6.3 算法性能

为了验证本文算法的有效性,将本文算法与俯视角深度与颜色直方图特征(top view depth and color histogram, TVDH)(Paolanti 等, 2018)、步态能量图(gait energy image, GEI)(Han 和 Bhanu, 2006)、步态能量体积(gait energy volume, GEV)(Sivapalan 等, 2011)、深度梯度直方能量图(depth gradient histogram energy image, DGHEI)(Hofmann 等, 2012)、Gabor 局部二值模式(Gabor local binary pattern, GLBP)、Gabor 局部导数模式(Gabor local derivative pattern, GLDP)(Imani 和 Soltanizadeh, 2019)在 TVPR 数据集、TDPI-L 数据集和 TDPI-B 数据集上进行实验。所有实验均在 MATLAB R2016a 平台上完成,机器配置为 3.10 GHz Intel(R) Pentium(R) CPU G3240, 8 GB RAM,结果如表 1 和图 7 所示。由于 TVDH 算法没有提供源代码和给出部分指标的具体数值,所以在表 1 中缺失数据。从表 1 和图 7 可以看出:本文算法在识别性能上与其他算法相比有很强的优势,但在平均耗时上比最快的 GEI 算法平均慢 37 ms。

表 1　不同行人再识别算法比较

Table 1　Performance of different person re-identification algorithms

数据集	指标	TVDH	GEI	GEV	DGHEI	GLBP	GLDP	本文算法
TVPR	rank-1	–	0. 21	0. 35	0. 51	0. 42	0. 53	**0. 75**
	rank-5	–	0. 38	0. 59	0. 69	0. 64	0. 68	**0. 86**
	macro-F1	0. 83	0. 36	0. 55	0. 68	0. 58	0. 65	**0. 84**
	平均耗时/ms	–	**120**	126	138	142	168	157
TDPI-L	rank-1	–	0. 21	0. 31	0. 46	0. 43	0. 49	**0. 69**
	rank-5	–	0. 28	0. 64	0. 67	0. 58	0. 69	**0. 79**
	macro-F1	–	0. 24	0. 59	0. 65	0. 54	0. 67	**0. 76**
	平均耗时/ms	–	**127**	132	146	157	172	165
TDPI-B	rank-1	–	0. 19	0. 27	0. 44	0. 35	0. 46	**0. 61**
	rank-5	–	0. 25	0. 43	0. 57	0. 46	0. 62	**0. 68**
	macro-F1	–	0. 23	0. 41	0. 52	0. 42	0. 60	**0. 67**
	平均耗时/ms	–	**131**	139	147	159	178	167

注：加粗字体为每行最优值；“–”表示相关文献中没有给出该指标值。

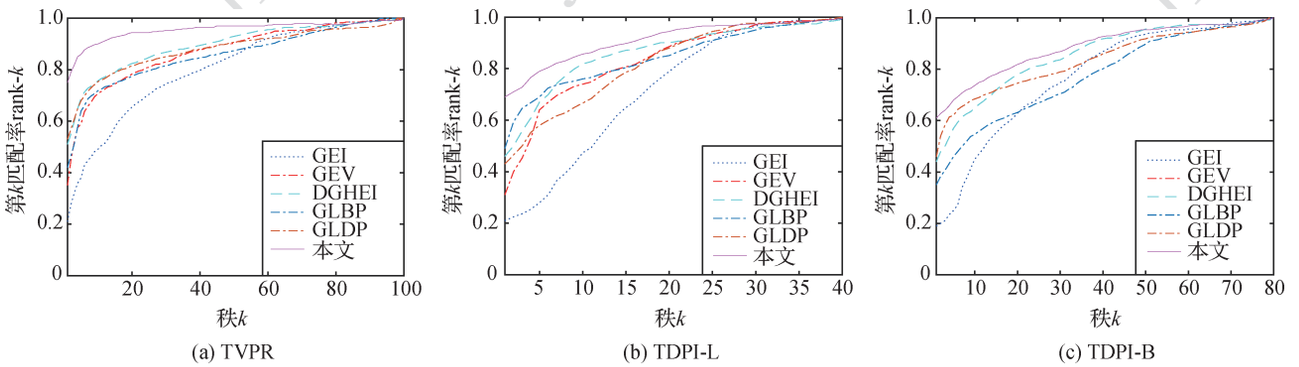


图 7　各算法在 3 个数据集上的 CMC 曲线

Fig. 7　CMC curves of the proposed algorithm and other algorithms on three datasets((a) TVPR;(b) TDPI-L;(c)TDPI-B)

与 TVDH、GEI、GEV、DGHEI、GLBP 和 GLDP 算法相比，本文算法构建了用以表征结构特征和行为特征的头部深度能量图组，并在此基础上提取多角度特征计算相似度，将其以损失函数最小的目标进行融合，从而有效地利用融合的多特征相似度实现行人再识别。因此，相比现有的基于深度相机的行人再识别算法，本文算法在俯视深度头肩序列上具有一定优势，在低分辨率的公交车实际场景 TDPI-B 数据集中也能取得较好的识别结果。

6.4　跟踪对算法性能的影响

TDPI-L 数据集和 TDPI-B 数据集中同一时间内至少有 2 个人存在，所以为了分析跟踪对算法性能的影响，分别使用核相关滤波器(kernelized correlation filters，KCF)(Henriques 等，2015)、基于回归网络的通用目标跟踪(generic object tracking using regression networks，GOTURN)(Held 等，2016)、马尔可夫决策过程(Markov decision process，MDP)(Xiang 等，2015)和卡尔曼等算法进行跟踪，使用多特征相似度加权融合算法进行识别，实验结果如表 2 所示。

从实验结果中可以看出，本文选择使用的卡尔曼算法的效果最好，MDP 次之，KCF 和 GOTURN 的效果较差。深度视频中行人与行人之间、行人与背

表 2　跟踪对算法性能的影响

Table 2　Tracking effects on the proposed algorithm performance

数据集	指标	KCF	GOTURN	MDP	卡尔曼
TDPI-L	rank-1	0.23	0.36	0.49	**0.69**
	rank-5	0.33	0.41	0.62	**0.79**
	macro-F1	0.32	0.38	0.57	**0.76**
TDPI-B	rank-1	0.20	0.28	0.46	**0.61**
	rank-5	0.24	0.37	0.53	**0.68**
	macro-F1	0.23	0.34	0.51	**0.67**

注：加粗字体为每行最优值。

景之间的区分度较小，使得 KCF、GOTURN 和 MDP 等跟踪算法的适用性较差。然而，卡尔曼算法是利用位置信息更新模型，适用性良好。因此本文选择卡尔曼算法进行跟踪，以有效获得头部图像序列，从而构建表达结构特征和行为特征的头部深度能量图组。

6.5　损失函数对算法性能的影响

为了分析损失函数对本文算法的影响，在 TVPR 数据集、TDPI-L 数据集和 TDPI-B 数据集上分别进行了 0-1 损失函数、交叉熵损失函数和本文的损失函数的性能比较，结果如表 3 所示。从实验结果中可以看出，本文使用的损失函数的效果最好，交叉熵损失函数次之，0-1 损失的性能稍差。

表 3　各损失函数对算法性能的影响

Table 3　Effect of loss functions on the performance of the proposed algorithm

数据集	指标	0-1 损失	交叉熵损失	本文算法
TVPR	rank-1	0.69	0.73	**0.75**
	rank-5	0.78	0.82	**0.86**
	macro-F1	0.76	0.81	**0.84**
TDPI-L	rank-1	0.60	0.67	**0.69**
	rank-5	0.74	0.78	**0.79**
	macro-F1	0.71	0.75	**0.76**
TDPI-B	rank-1	0.53	0.59	**0.61**
	rank-5	0.61	0.67	**0.68**
	macro-F1	0.59	0.65	**0.67**

注：加粗字体为每行最优值。

6.6　参数对算法性能的影响

6.6.1　头部深度能量图的组数 R_t 对算法性能的影响

为了验证头部深度能量图的组数 R_t 对本文算法的影响，在 TVPR 数据集、TDPI-L 数据集和 TDPI-B 数据集上分别进行 R_t 取不同值的测试，观察 rank-1 的变化情况，结果如图 8 所示。

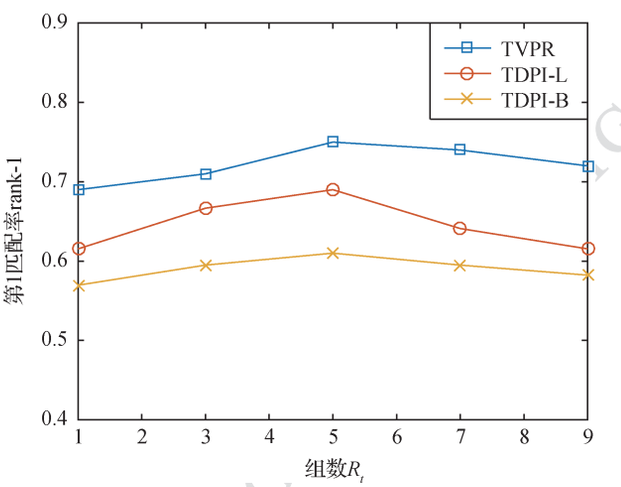


图 8　R_t 值对算法性能的影响

Fig. 8　The effect of the parameter R_t on the performance of the proposed algorithm

从实验结果中可以看出，头部深度能量图的组数 R_t 为 1 时，效果较差。R_t 取 1 时会将所有的行人头部图像进行叠加，使得行人头部图像的边缘特征大幅度减少，误差较大。但是，头部深度能量图的组数 R_t 过大时，rank-1 也会有一定下降。R_t 值过大时，会将行人头部图像序列分成太多组，计算量加倍，更易受到噪声干扰的影响。当头部深度能量图的组数 R_t 为 5 时，效果较好。

6.6.2　各特征相似度对算法性能的影响

首先分析各个特征相似度的作用，在 TVPR 数据集、TDPI-L 数据集和 TDPI-B 数据集上分别使用深度特征相似度 S_d、面积特征相似度 S_a、投影特征相似度 S_p、傅里叶描述子相似度 S_f 和 HOG 特征相似度 S_h 实现行人再识别，结果如表 4 所示。从实验结果中可以看出，HOG 特征相似度 S_h 的效果最好，傅里叶描述子相似度 S_f 次之，深度特征相似度 S_d、面积特征相似度 S_a 和投影特征相似度 S_p 的性能稍差。

6.6.3　相似度权重向量 $\boldsymbol{w}$ 对算法性能的影响

为了展示相似度权重向量对本文算法的影响，

表 4　各特征对算法性能的影响

Table 4　Effect of features on the performance of the proposed algorithm

数据集	指标	S_d	S_a	S_p	S_f	S_h	融合算法
TVPR	rank-1	0.28	0.39	0.31	0.46	0.55	**0.75**
	rank-5	0.46	0.45	0.51	0.62	0.69	**0.86**
	macro-F1	0.44	0.42	0.50	0.61	0.68	**0.84**
TDPI-L	rank-1	0.26	0.38	0.28	0.41	0.54	**0.69**
	rank-5	0.31	0.44	0.36	0.49	0.64	**0.79**
	macro-F1	0.30	0.42	0.34	0.48	0.61	**0.76**
TDPI-B	rank-1	0.22	0.25	0.32	0.39	0.47	**0.61**
	rank-5	0.27	0.34	0.38	0.43	0.54	**0.68**
	macro-F1	0.26	0.32	0.36	0.40	0.53	**0.67**

注：加粗字体为每行最优值。

在 TVPR 数据集、TDPI-L 数据集和 TDPI-B 数据集上分别使用不同的权重向量进行测试，结果如表 5 所示。

表 5　不同的权重向量对算法的影响

Table 5　Effect of fusion weights on the performance of the proposed algorithm

数据集	指标	组合 1	组合 2	组合 3	组合 4	组合 5
TVPR	rank-1	0.56	0.60	**0.75**	0.64	0.68
	rank-5	0.63	0.71	**0.86**	0.75	0.74
	macro-F1	0.60	0.68	**0.84**	0.72	0.73
TDPI-L	rank-1	0.54	0.59	0.64	**0.69**	0.51
	rank-5	0.56	0.67	0.72	**0.79**	0.62
	macro-F1	0.55	0.64	0.70	**0.76**	0.60
TDPI-B	rank-1	0.51	0.54	0.58	0.53	**0.61**
	rank-5	0.56	0.59	0.66	0.63	**0.68**
	macro-F1	0.53	0.58	0.64	0.60	**0.67**

注：加粗字体为每行最优值。

根据本文算法在 TVPR、TDPI-L 和 TDPI-B 的训练集上进行的实验，可知 $\boldsymbol{w}^* = [w_d, w_a, w_p, w_f, w_h]^T = [0.14, 0.11, 0.20, 0.32, 0.23]^T$ 时，TVPR 数据集上的损失值最小；$\boldsymbol{w}^* = [w_d, w_a, w_p, w_f, w_h]^T = [0.18, 0.22, 0.11, 0.17, 0.32]^T$ 时，TDPI-L 数据集上的损失值最小；$\boldsymbol{w}^* = [w_d, w_a, w_p, w_f, w_h]^T = [0.12, 0.13, 0.20, 0.22, 0.33]^T$ 时，TDPI-B 数据集上的损失值最小。

设 $\boldsymbol{w} = [0.20, 0.20, 0.20, 0.20, 0.20]^T$ 为组合 1，$\boldsymbol{w} = [0.30, 0.20, 0.10, 0.20, 0.20]^T$ 为组合 2，$\boldsymbol{w} = [0.14, 0.11, 0.20, 0.32, 0.23]^T$ 为组合 3，$\boldsymbol{w} = [0.18, 0.22, 0.11, 0.17, 0.32]^T$ 为组合 4，$\boldsymbol{w} = [0.12, 0.13, 0.20, 0.22, 0.33]^T$ 为组合 5。其中，组合 1 是将各特征相似度进行平均加权，组合 2 是手工设定的在 TVPR 数据集、TPI-L 数据集和 TDI-B 数据集中算法性能均较好的一组权重向量，组合 3 是 TVPR 数据集上的最优解，组合 4 是 TDPI-L 数据集上的最优解，组合 5 是 TDPI-B 数据集上的最优解。选取上述权重向量在 TVPR 数据集、TDPI-L 数据集和 TDPI-B 数据集上进行实验，实验结果如表 5 所示。

本文算法在 TVPR 数据集、TDPI-L 数据集和 TDPI-B 数据集的识别效果均优于融合前算法和其他 4 种组合算法，可见经过模型学习到的相似度权重向量能根据使损失函数最小的目标，将深度特征相似度、面积特征相似度、投影特征相似度、傅里叶描述子相似度和 HOG 特征相似度进行融合，证明了本文算法的有效性。

7　结　论

针对俯视深度头肩序列，本文提出了一种基于头部深度能量图组的多特征表达与权重学习的行人重识别算法，主要步骤包括：构建表达行人头部结构特征与行为特征的头部深度能量图组，提取适合低分辨率行人表达的深度、面积、投影、傅里叶描述子以及方向梯度直方图等特征，学习特征相似度融合权重系数。在公开的室内单人场景、自建的室内多人场景以及公交车车厢内实际场景 3 个数据集上的实验结果表明，本文算法均优于典型算法，尤其是在车载低分率场景（如公交车车厢内场景）的性能优势更为明显。

由于仅考虑了行人头部的结构和行为特征，本文算法性能易受头部装饰物变化的影响，因此在下一步研究中，拟引入俯视深度头肩序列中的非头部结构特征以及人走路的行为特征，通过多部位结构特征和行为特征联合表达的方式提高算法的适用性。

参考文献(References)

Barbosa I B, Cristani M, Del Bue A, Bazzani L and Murino V. 2012. Re-identification with RGB-D sensors//Proceedings of 2012 European Conference on Computer Vision. Florence: Springer: 433-442 [DOI: 10.1007/978-3-642-33863-2_43]

Bewley A, Ge Z Y, Ott L, Ramos F and Upcroft B. 2016. Simple online and realtime tracking//Proceedings of 2016 IEEE International Conference on Image Processing. Phoenix: IEEE: 3464-3468 [DOI: 10.1109/ICIP.2016.7533003]

Han J and Bhanu B. 2006. Individual recognition using gait energy image. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(2): 316-322 [DOI: 10.1109/TPAMI.2006.38]

Held D, Thrun S and Savarese S. 2016. Learning to track at 100 FPS with deep regression networks//Proceedings of 2016 European Conference on Computer Vision. Amsterdam: Springer: 749-765 [DOI: 10.1007/978-3-319-46448-0_45]

Henriques J F, Caseiro R, Martins P and Batista J. 2015. High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(3): 583-596 [DOI: 10.1109/TPAMI.2014.2345390]

Hofmann M, Bachmann S and Rigoll G. 2012. 2.5D gait biometrics using the depth gradient histogram energy image//Proceedings of the 5th IEEE International Conference on Biometrics: Theory, Applications and Systems. Arlington: IEEE: 399-403 [DOI: 10.1109/BTAS.2012.6374606]

Imani Z and Soltanizadeh H. 2019. Local binary pattern, local derivative pattern and skeleton features for RGB-D person re-identification. National Academy Science Letters, 42(3): 233-238 [DOI: 10.1007/s40009-018-0736-9]

Jiang J G, Yang N, Qi M B and Chen C Q. 2019. Person re-identification with region block segmentation and fusion. Journal of Image and Graphics, 24(04): 513-522 (蒋建国, 杨宁, 齐美彬, 陈翠群. 2019. 区域块分割与融合的行人再识别. 中国图象图形学报, 24(04): 513-522) [DOI: 10.11834/jig.180370]

Kim M, Jung J, Kim H and Paik J. 2017. Person Re-identification using color name descriptor-based sparse representation//Proceedings of the 7th IEEE Annual Computing and Communication Workshop and Conference. Las Vegas: IEEE: 1-4 [DOI: 10.1109/CCWC.2017.7868394]

Liao S C, Hu Y, Zhu X Y and Li S Z. 2015. Person re-identification by local maximal occurrence representation and metric learning//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE: 2197-2206 [DOI: 10.1109/CVPR.2015.7298832]

Liao S C, Zhao G Y, Kellokumpu V, Pietikäinen M and Li S Z. 2010. Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco: IEEE: 1301-1306 [DOI: 10.1109/CVPR.2010.5539817]

Matsukawa T, Okabe T, Suzuki E and Sato Y. 2016. Hierarchical Gaussian descriptor for person re-identification//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE: 1363-1372 [DOI: 10.1109/CVPR.2016.152]

Munaro M, Basso A, Fossati A, Van Gool L and Menegatti E. 2014a. 3D reconstruction of freely moving persons for re-identification with a depth sensor//Proceedings of 2014 IEEE International Conference on Robotics and Automation. Hong Kong, China: IEEE: 4512-4519 [DOI: 10.1109/ICRA.2014.6907518]

Munaro M, Fossati A, Basso A, Menegatti E and Van Gool L. 2014b. One-shot person re-identification with a consumer depth camera//Gong S G, Cristani M, Yan S C, Loy C C, eds. Person Re-Identification. London: Springer: 161-181 [DOI: 10.1007/978-1-4471-6296-4_8]

Munaro M, Ghidoni S, Dizmen D T and Menegatti E. 2014c. A feature-based approach to people re-identification using skeleton keypoints//Proceedings of 2014 IEEE International Conference on Robotics and Automation. Hong Kong, China: IEEE: 5644-5651 [DOI: 10.1109/ICRA.2014.6907689]

Nguyen T B, Nguyen H Q, Le T L, Pham T T T and Pham N N. 2019. A quantitative analysis of the effect of human detection and segmentation quality in person re-identification performance//Proceedings of 2019 International Conference on Multimedia Analysis and Pattern Recognition. Ho Chi Minh City: IEEE: 1-6 [DOI: 10.1109/MAPR.2019.8743532]

Nguyen T B, Tran D L, Le T L, Pham T T T and Doan H G. 2018. An effective implementation of Gaussian of Gaussian descriptor for person re-identification//Proceedings of the 5th NAFOSTED Conference on Information and Computer Science. Ho Chi Minh City: IEEE: 388-393 [DOI: 10.1109/NICS.2018.8606858]

Paolanti M, Romeo L, Liciotti D, Pietrini R, Cenci A, Frontoni E and Zingaretti P. 2018. Person re-identification with RGB-D camera in top-view configuration through multiple nearest neighbor classifiers and neighborhood component features selection. Sensors, 18(10): #3471 [DOI: 10.3390/s18103471]

Sivapalan S, Chen D, Denman S, Sridharan S and Fookes C. 2011. Gait energy volumes and frontal gait recognition using depth images//Proceedings of 2011 International Joint Conference on Biometrics. Washington: IEEE: 1-6 [DOI: 10.1109/IJCB.2011.6117504]

Wu A C, Zheng W S and Lai J H. 2017. Robust depth-based person re-identification. IEEE Transactions on Image Processing, 26(6): 2588-2603 [DOI: 10.1109/TIP.2017.2675201]

Xiang Y, Alahi A and Savarese S. 2015. Learning to track: online multi-object tracking by decision making//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 4705-4713 [DOI: 10.1109/ICCV.2015.534]

Zheng S T, Li X Y, Jiang Z Q and Guo X Q. 2017. LOMO3D descriptor for video-based person re-identification//Proceedings of 2017 IEEE Global Conference on Signal and Information Processing. Montreal: IEEE: 672-676 [DOI: 10.1109/GlobalSIP.2017.8309044]

作者简介

王新年，1973 年生，男，副教授，博士生导师，主要研究方向为数字图像处理、生物特征识别和模式识别。
E-mail：wxn@dlmu.edu.cn

张世强，通信作者，男，教授级高工，主要研究方向为智能交通系统。
E-mail：zhangsq@ehualu.com

刘春华，女，硕士研究生，主要研究方向为图像处理与计算机视觉。E-mail：liuchunhua1.edu@163.com

齐国清，男，教授，主要研究方向为雷达通信与图像信号处理。E-mail：qgq@dlmu.edu.cn