

DOI: 10.19650/j.cnki.cjsi.J2210655

基于深度测量的行人体态特征提取与再识别方法*

刘明洋, 万九卿

(北京航空航天大学自动化科学与电气工程学院 北京 100191)

摘要:行人再识别是视觉监控系统的核心问题之一,然而传统基于彩色图像的特征提取方法难以用于极暗光照条件下的行人再识别。本文提出一种基于深度测量的行人体态特征提取方法,由于深度测量独立于光照条件,因此所提方法可以在极暗光照条件下对行人目标进行有效识别。由深度数据经过分割和滤波生成人体点云,将观测点云与初始人体模型进行配准,基于配准后的点云对人体模型的体态参数和姿态参数进行联合估计,计算体态特征向量的欧式距离实现行人再识别。在公开数据集和实验室自采数据集上进行验证,计算 Rank-n、累计匹配曲线、平均精度均值等性能指标,其中在 Single shot 评估模式下 BIWI 数据集的 Rank-1 可达到 70.71%、Rank-5 可达到 92.32%,结果表明本文所提算法可有效提高再识别精度。

关键词:行人再识别;极暗光照;深度测量;人体模型

中图分类号: TP391.41 TH789 **文献标识码:** A **国家标准学科分类代码:** 520.2050

Person shape feature extraction and reidentification based on depth measurement

Liu Mingyang, Wan Jiuqing

(School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China)

Abstract: Person re-identification is a fundamental problem in the smart video surveillance system. However, the traditional RGB-based feature extraction method cannot be used in dark environment. A new method for person shape feature extraction using depth measurement is proposed in this article. The depth data are independent from lighting condition. Therefore, the proposed method can be used for person re-id in the dark. Specifically, the point cloud of person is generated from depth data after segmentation and filtering. Then, the point cloud is registered to the initial human body model. The shape and pose parameters of the body model are estimated jointly based on the registered point cloud. Finally, the re-id is achieved by calculating the Euclidean distance in the vector space of shape parameters. The author applies this method on public and self-collected datasets in the laboratory to calculate performance indicators, including Rank-n, cumulative matching curve, and mean average precision, etc. Among the indicators, the Rank-1 of BIWI datasets in Single shot evaluation mode reaches 70.71% and the Rank-5 of BIWI datasets is up to 92.32%, which indicate that the proposed algorithm can effectively improve the re-recognition accuracy.

Keywords: pedestrian re-identification; extremely dark lighting; depth measurement; human body model

0 引 言

行人再识别^[1]技术用于判断图像或者视频中是否存在特定的行人,是智能视觉监控系统的核心问题之一。行人再识别可以基于如下特征:衣着特征^[2-3],人脸特征^[4-5],步态特征^[6-7]和体态特征^[8-12]。目前绝大多数特

征提取算法需要根据彩色图像进行提取,而获取有效的彩色图像需要适度的光照条件。然而在很多情况下,比如夜间监控或隐私保护的情形,需要在极暗光照条件下进行行人再识别。

在极暗光照下,深度传感器,比如深度相机或激光雷达传感器,仍可获得物体与相机之间距离的可靠测量^[13]。从深度数据中提取衣着和人脸特征较为困难,而

步态特征的有效性依赖于行人的行走状态。因此,本文将研究如何从深度数据提取人体体态特征,从而实现基于体态特征的行人再识别。人体体态是指人身体的形状特征,体态特征是人的固有属性,独立于行人运动状态的变化。已有研究表明,依据体态特征可以进行有效的行人识别,体态识别可以作为其他行人再识别方法的替代或必要的补充^[14]。

然而,目前在行人再识别领域关于体态特征提取的研究工作大多数集中在基于彩色图像^[8-12,15]或彩色与深度图像的组合^[16-18],无法应用于极暗光照场合。在没有彩色图像信息的条件下,从纯深度数据中获取行人体态特征是一个更为困难的问题,主要原因在于:1)行人的深度数据及由此生成的点云随观测角度的变化发生显著变化;2)行人的深度数据及由此生成的点云随行人动作的变化发生显著变化;3)由深度数据生成的点云是不完整的,对应于部分人体表面;4)深度数据存在大量噪声。由于上述原因,目前基于纯深度数据进行体态识别的研究十分稀少,且结果并不令人满意。在文献[8,11]中,作者从点云中提取局部特征描述人体体态,然而局部特征难以表征人体的全局特性。在文献[16,19]中,作者使用简单的人体尺寸测量进行体态识别,没有充分利用深度数据中的体态信息。

针对上述问题,本文提出一种基于纯深度数据的行人体态参数估计方法,所得体态参数给出了行人身体形状的三维全局描述,且独立于光照、视角、行人动作、以及衣着颜色和纹理的变化,可以有效提升极暗光照条件下行人再识别的准确率。同时,本文方法不需要训练过程,避免了基于机器学习的方法^[8-12,15-18]中常见的样本稀缺、过拟合或领域迁移等问题。本文方法基于蒙皮多人线性模型(skinned multi-person linear, SMPL)^[20],该模型实现了从人体体态参数和姿态参数到人体三维顶点的映射。

本文将体态特征提取问题转化为基于不完整点云观测的SMPL模型参数估计问题,通过体态和姿态参数的联合估计实现了两者的解耦。根据估计所得的体态参数向量,通过计算向量间的欧式距离完成行人再识别。本文主要贡献总结如下:

1) 提出一种极暗光照下的行人再识别方法,该方法将人体体态参数作为行人特征实现基于体态的行人再识别;

2) 给出了基于纯深度数据的体态估计与识别过程,包括:数据预处理(人体分割、点云生成和点云滤波),点云配准(粗配准和精配准),参数估计和行人再识别;

3) 在自采数据集和公开数据集上验证了本文方法的有效性,结果表明本文方法的性能明显优于当前的方法。

1 相关工作

1.1 基于体态的行人再识别

基于彩色图像,研究者提出了若干基于深度学习模型用于提取体态特征,主要用于解决行人换装条件下的再识别问题。例如,文献[12]中的体态提取网络,文献[21]中的自编码器模型,文献[22]中的对抗特征解耦网络,以及文献[23]中的三维体态网络模型。这些模型在不同程度上实现了体态特征与其他外观特征的解耦。文献[15]提出一种从彩色图像提取行人轮廓的方法,轮廓在一定程度上反映了行人的体态特征。也有一些方法从彩色图像和深度图像的组合中提取体态特征。例如,文献[24-25]从彩色图像中提取衣着局部描述子,并将其与从深度图像中提取的行人尺寸测量相融合进行行人再识别。然而,上述传统基于彩色图像的方法需要全部或部分依赖于彩色图像,往往难以取得理想效果,无法应用于极暗光照条件,如图1所示。

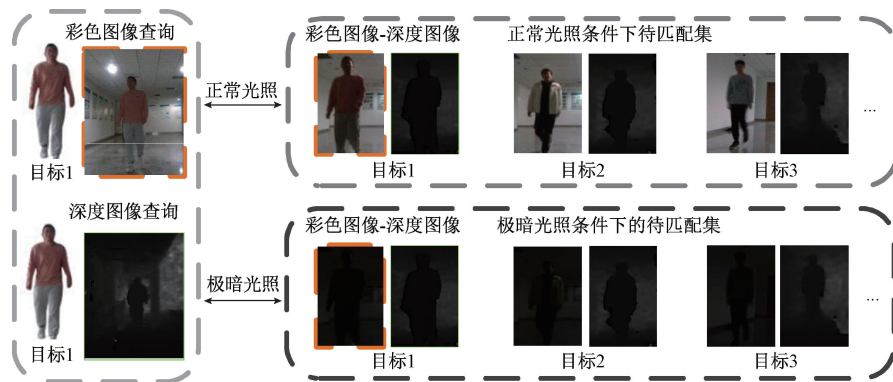


图1 正常光照与极暗光照下的行人再识别对比

Fig.1 Person re-identification in the light and in the dark

极暗光照情况下,只有少数工作基于纯深度数据提取体态特征进行行人再识别。在文献[16]中,作者计算

了人体关节间的欧式距离以及人体表面上的测地线距离,并以此作为体态特征进行再识别。文献[19]中,作

者计算对齐后点云之间的距离并据此进行行人再识别,但作者忽略了由于局部点云的刚体变换造成的人体形变问题。文献[11]中,作者提出一种再励学习框架,利用点云视频数据提取人体形态和运动的局部信息,但该方法需要点云视频数据训练模型。文献[8]中,作者从点云数据中提取局部描述子作为行人体态特征进行再识别,但全局体态信息没有得到充分利用。

与上述方法不同,本文方法利用 SMPL 模型,给出了行人体态的紧凑的三维全局描述,并据此进行行人再识别,避免了现有方法中简单利用尺寸测量^[16,19]或局部描述^[8,11]的缺陷。

1.2 基于模型的体态参数估计

针对人体建模问题,研究者已经提出了多种人体模型^[20,26-29],当前主流的模型包括线性混合蒙皮(linear blending skinning, LBS)^[26],姿势空间变形(pose space deformation, PSD)^[27],人体体态动作模型(shape completion and animation of people, SCAPE)^[28],SMPL模型^[20]和稀疏训练的铰接式人体模型(sparse trained articulated human body regressor, STAR)^[29],其中 SMPL 模型^[20]的应用最为广泛。已有许多工作致力于基于彩色图像^[30-33]或视频^[34-35]的 SMPL 模型参数估计。

最近也出现了一些纯深度数据的 SMPL 模型参数估计方法。在文献[36]中,作者提出一种深度回归网络框架,将点云深度特征映射到人体骨架,再由人体骨架映射到 SMPL 模型参数。然而,该深度回归网络需要大量的训练样本进行有监督学习,训练样本由具有一一对应关

系的点云和人体三维网格顶点集合组成。类似的工作包括^[37-38],作者分别采用了回归网络^[37]和自编码器^[38]架构。

然而,目前提出的基于纯深度数据的 SMPL 模型参数估计方法依赖于有监督的训练过程,因而存在样本稀缺、过拟合以及领域迁移等问题。同时,现有方法针对的应用场景主要是时装行业、健康监测、虚拟现实和增强现实领域,所关注的重点在于三维体重构的精度,而本文的目标是寻求一种人体形态的描述以便进行可靠的行人再识别。因此,本文没有采用基于深度学习的方法,而是通过将 SMPL 模型拟合到配准后的点云获取体态参数的最优估计。

2 方法

2.1 整体流程

本文所提的行人再识别方法总体架构如图2所示,主要包含4个步骤:1)预处理:从相机采集到的深度图像 D 中将人体分割,接着生成对应于部分人体表面的三维点云 P 并对点云滤波以去除噪声干扰;2)配准:通过寻求恰当的刚体变换 T 实现点云 P 与 SMPL 模型 M 对齐,以便于建立点云顶点与模型顶点之间的对应关系;3)参数估计:通过最小化点云 P 与模型 M 顶点间距离,寻求模型体态参数 β 和姿态参数 θ 的最优估计;4)行人再识别:将体态参数 β 作为人体体态特征,计算体态特征间欧式距离,实现行人再识别。

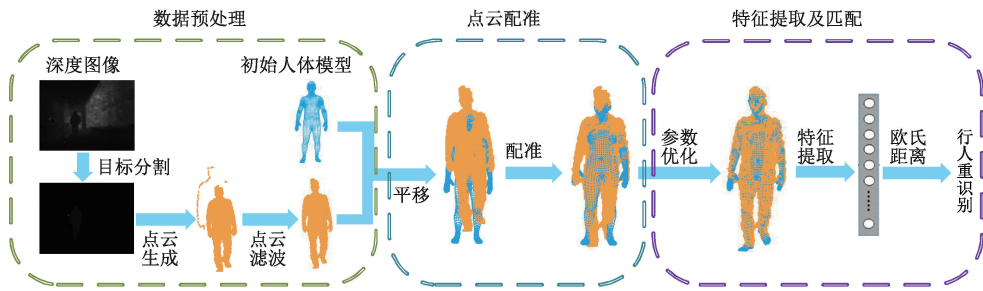


图2 基于深度测量的行人再识别整体框架

Fig. 2 The framework of pedestrian re-identification based on depth measurement

2.2 SMPL 人体模型

本文使用 SMPL 模型^[20]作为人体模型,该模型可以有效地表征不同姿态下各种体型的人体。SMPL 模型实现了从人体体态参数 β 和人体姿态参数 θ 到人体三维网格顶点地映射,映射关系如式(1)所示。

$$M(\beta, \theta; \Phi) = W(T_p(\beta, \theta; \bar{T}, S), J(\beta; \bar{T}, S), \theta, \omega) \quad (1)$$

其中, $\Phi = (\bar{T}, S, \omega)$ 为给定的模型参数,在 β 和 θ 的估计过程中保持不变。函数 W 为标准的线性混合蒙皮函

数,输入为零姿态下修正后的人体模板 T_p ,关节位置 J ,人体姿态 θ 和混合权重参数 ω ,输出 6 890 个顶点的三维坐标。文献[20]对大量零姿态人体网格进行主元素分析,得到人体均值模板 \bar{T} ,以及偏差向量矩阵 $S = [S_1, \dots, S_{|\beta|}]$,该矩阵的每一列对应于一个主元方向。体态参数向量 $\beta \in R^{10}$ 由前 10 个主元向量的组合系数构成,刻画了特定人体的体态特征。姿态参数向量 $\theta \in R^{72}$ 包含描述人体 23 个关节相对旋转关系的 3×23 个轴角参数和描述根节点方向的 3 个轴角参数。

SMPL模型通过6890个网格顶点的三维坐标来表征人体形状,这些顶点均匀分布在整个人体表面。由于人类体型的约束,这6890个顶点位置并非相互独立。反映到SMPL中,就体现为6890个顶点的位置可以由体态和姿态共82(10+72)个参数来控制。由于人体形状的约束,可以通过对人体部分表面形状的观测估计SMPL模型参数,从而获取人体整体形状的描述。

2.3 深度数据预处理

1) 人体分割

原始的深度图像中包含了人体和背景,本文采用背景差分法将对应于人体部分的像素分割出来。如图3所示,图3(a)表示原始深度图,图3(b)表示背景深度图,图3(c)是将前两者相减而得,图3(d)为图3(c)中的非零像素区域,对应于人体区域掩模,图3(e)为分割出的人体区域,可以看到图3(e)中只包含了人体深度像素。

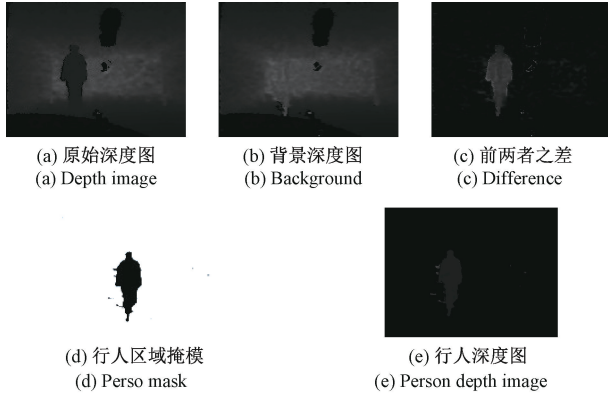


图3 深度图像人体分割

Fig. 3 Person segmentation in depth image

2) 点云生成

将世界坐标系选取为相机坐标系,根据相机内参将二维深度图像转换为三维点云。对深度图中的每个像素 (u, v) ,可按式(2)计算其在点云中的三维坐标 (x_w, y_w, z_w) ,如图4(a)所示。坐标定义在世界坐标系。

$$\begin{cases} x_w = z_c \times (u - u_0) \times \frac{dx}{f} \\ y_w = z_c \times (v - v_0) \times \frac{dy}{f} \\ z_w = z_c \end{cases} \quad (2)$$

其中, $z_c(u, v)$ 表示像素 (u, v) 的深度值, f 表示相机焦距, (u_0, v_0) 表示中心像素点坐标。

3) 滤波

经过人体分割、点云生成后,我们得到一个初始人体点云。但由于分割误差和深度相机空间分辨率的限制,使得在人体轮廓边缘处部分背景像素可能被误判为人体

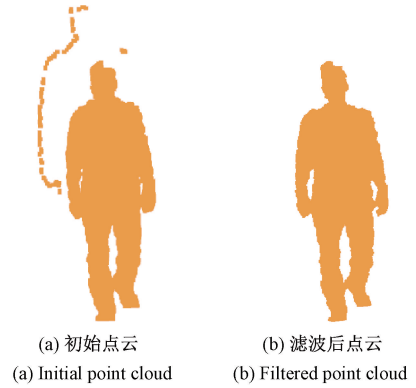


图4 滤波效果

Fig. 4 Filtering effect

区域像素,导致生成的人体点云中出現孤立的条状野值,如图4(a)所示。本文采用半径滤波方法对其进行滤除,基本思想如下:对于点云中的每个点,通过指定其一定半径范围内存在的最少点数阈值进行滤波。具体来说,指定每点在一定半径范围内至少要有足够多数量的近邻点,符合阈值要求的点将被保留,不符合阈值要求的点将被剔除。以图5为例各形状的点均表示点云中的点,此时进行基于半径距离的点云滤波,若指定半径内最少临近点阈值为1,那方形的点会被滤除;若指定半径内最少临近点阈值为2,那方形和六边形的点都将被滤除。通过选取适当的半径和阈值,可有效滤除远离人体的孤立条状野值,本文实验中,取半径为0.05 m,阈值点数为20,滤波后点云如图4(b)所示。

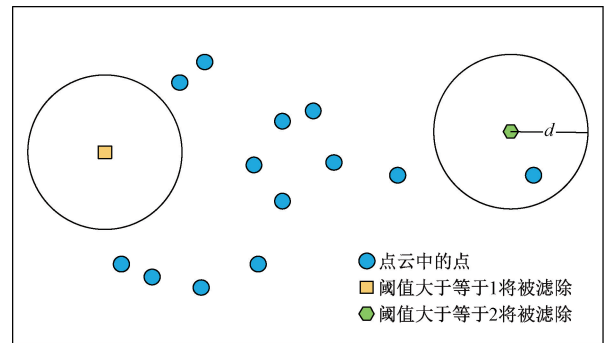


图5 基于半径的点云滤波

Fig. 5 Radius-based point cloud filtering

2.4 点云配准

SMPL模型 M 原点与相机坐标系原点重合,无法直接利用观测点云 P 估计SMPL模型的参数。因此,需要寻求一个刚体变换 T 完成模型 M 与点云 P 的配准。配准分为粗配准和精配准两个步骤。

1) 粗配准

首先计算观测点云 P 的均值, 计算公式如式(3)所示。

$$\bar{P} = \frac{1}{|P|} \sum_{p \in P} p \quad (3)$$

其中, $p \in R^3$ 为点云中点的坐标, $|P|$ 为点云中点的数目。粗配准后的点云表示为式(4)。

$$P' = \{p' : p' = p - \bar{P}, p \in P\} \quad (4)$$

粗配准效果如图 6 所示, 可以看到, 粗配准实现了点云平移, 可以显著减小人体模型和观测点云之间的距离, 为精配准提供了一个可靠的初始解。

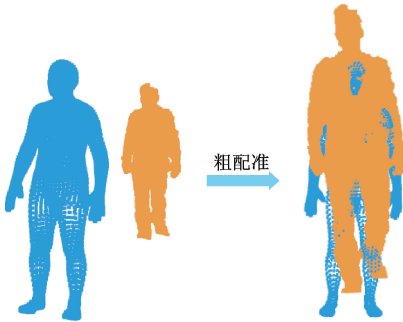


图 6 粗配准效果
Fig. 6 Coarse registration effect

2) 精配准

利用最近点迭代算法 (iterative closest point, ICP) 算法^[39]实现人体模型 M 和平移后的观测点云 P' 的精配准, 即计算刚体变换 $T = (R, t)$, 包括旋转矩阵 R 和平移矢量 t 。精配准算法由以下步骤组成:

(1) 根据给定的初始刚体变换 $T = (R, t)$, 变换后的模型 M' 的计算公式如式(5) 所示。

$$M'(R, t) = \{m' : m' = Rm + t, m \in M\} \quad (5)$$

(2) 计算模型 M' 和点云 P' 之间的距离。由于模型 M' 覆盖了整个人体表面, 而点云 P' 只包含人体部分表面, 因此对于 P' 中的每一个点, 我们遍历 M' 寻找与之最近的点并计算二者的欧式距离。模型 M' 和点云 P' 之间距离 d 计算公式如式(6) 所示。

$$d(M', P') = \frac{1}{|P'|} \sum_{p' \in P', m' \in M'} \min \|p' - m'\|_2^2 \quad (6)$$

(3) 计算最优刚体变换 $T^* = (R^*, t^*)$, 最小化观测点云 P' 与变换后模型 M' 之间的距离, 表示为式(7) 所示。

$$(R^*, t^*) = \underset{R, t}{\operatorname{argmin}} d(M'(R, t), P') \quad (7)$$

(4) 如果距离 $d(M', P')$ 小于给定阈值, 结束迭代; 否则, 重复步骤(1) ~ (3)。

精配准效果如图 7 所示, 图 7(a) 为粗配准后的初始结果, 图 7(b) 为精配准 20 轮迭代的效果, 图 7(c) 为精配准 40 轮迭代的效果, 图 7(d) 为精配准 60 轮迭代的效

果, 可以看出随着迭代次数的增加, 点云之间逐渐靠近贴合, 最终实现了较高的配准精度。

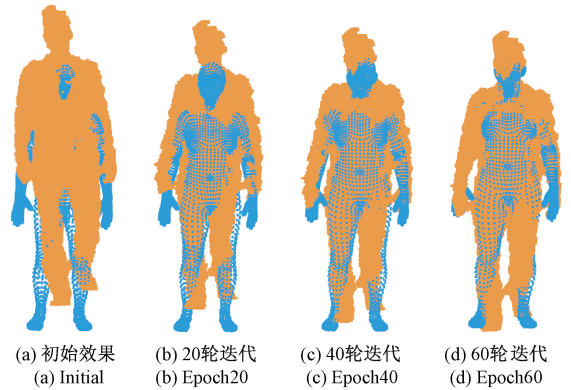


图 7 精配准效果
Fig. 7 Fine registration effect

2.5 参数估计

配准过程确定了人体模型与观测点云之间的刚体变换, 但由于人体形状会随着人体的体型和姿态 (即关节角度) 的变化而变化, 刚体变换并不能实现模型与点云的精准贴合。利用 SMPL 模型, 通过改变人体体态参数 β 和姿态参数 θ 调整模型的形状 $M''(\beta, \theta)$, 使之与观测点云 P' 之间的距离 $d(M'', P')$ 最小, 即可获得人体体态和姿态参数的最优估计如式(8) 所示。

$$(\beta^*, \theta^*) = \underset{\beta, \theta}{\operatorname{argmin}} d(M''(\beta, \theta), P') + h(\theta) \quad (8)$$

其中, $d(M'', P')$ 按式(6) 计算, $h(\theta)$ 对应于各关节角度的先验分布。

利用梯度下降法求解上述优化问题, 优化过程如图 8 所示, 可以看到, 随着迭代轮次的增加, 人体模型和观测点云逐渐贴合, 表明人体体态和姿态参数被越来越准确的估计。

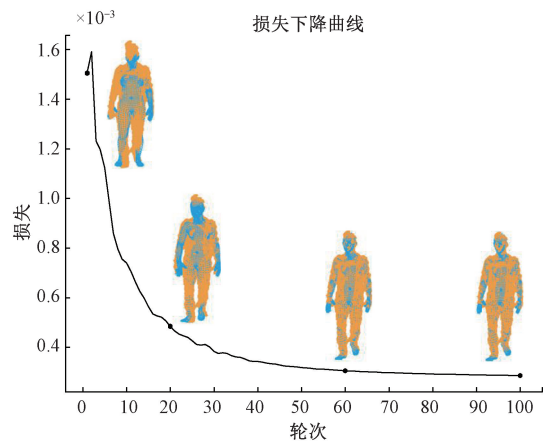


图 8 参数估计过程
Fig. 8 Parameter estimation process

2.6 基于体态的再识别

参数估计得到了人体体态参数 β 和姿态参数 θ 的估计,其中姿态参数随人体关节运动而改变,不能用于行人再识别。而体态参数是特定行人的固有属性,不随其动作发生改变。因此,将体态参数 β 作为行人特征用于再识别。

行人再识别问题可描述如下:给定查询样本 β^p 以及候选样本集合 $\{\beta_1^c, \dots, \beta_k^c\}$,根据 β^p 与每个候选样本 β_i^c 的相似度获取一个候选样本集合的排序,候选样本排序越靠前,则与查询样本属于同一个行人的概率越高。本文将体态参数向量间的欧氏距离 $\|\beta^p - \beta_i^c\|_2^2$ 作为样本相似性度量,行人再识别算法性能指标计算详见 3.2 节。

3 实验结果

3.1 数据集

采用的数据集包括公开数据集 BIWI^[19] 和实验室自采数据集。

1) BIWI 数据集

BIWI 是一个用于行人再识别的深度数据集,数据集中出现的演员总数为 50。训练集为所有演员在实验室环境中采集的一段视频数据,测试集中包含其中 28 位演员在行走与站立两种状态下采集的视频数据。BIWI 数据集是在光照充足的实验室环境下采集,包括彩色图像和深度数据,本文只利用了其中的深度数据。

2) 自采数据集

为了进一步验证极暗光照情况下的算法性能,我们在极暗光照条件下分别在 5 个地点采集了 10 个演员的实测数据,共 100 段视频。每个行人在每个地点均有两种衣着的数据,用于测试换装的影响。采用的相机为 Intel RealSense D435i 深度相机,如图 9(a) 所示。该相机可同时采集彩色图像与深度图像,实验中设定彩色图像与深度图像空间分辨率均为 640×480 ,帧率为 30 FPS。深度测量数据采用 16 位无符号二进制数表示,单位为 mm,有效工作距离 10 m。该相机采集到的彩色图像和深度图像如图 9(b) 所示,可以看到,由于光照极暗,从彩色图像中难以对行人进行识别。数据采集在夜间关闭灯光的密闭空间内进行,环境照度约为 $0.01 \sim 0.03$ Lx。

3.2 评价指标

本文采用行人再识别文献中通用的评价指标,包括 Rank-n, 累计匹配曲线,平均精度均值@n。

1) Rank-n 指标

Rank-n 识别率即将所有的检索结果按相似度距离从小到大排序后,与查询目标距离最近的 n 个结果中有

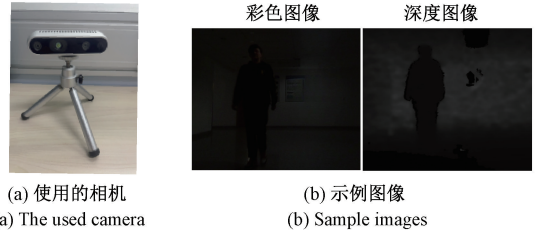


图 9 自采数据集使用的相机与示例图像

Fig. 9 The camera used in self-collected dataset and sample image

正确结果的概率,计算公式如式(9)所示。

$$Rank - n = \frac{\text{前 } n \text{ 个结果中正确结果的样本数量}}{\text{总样本数量}} \quad (9)$$

2) 累计匹配曲线

累计匹配曲线(cumulative matching characteristics, CMC)是生物特征识别领域的重要评价指标。该曲线的横坐标是指定的 Rank-n,纵坐标是各 Rank-n 识别率对应的百分比,可以有反映当前特征用于检索的速度和准确度,收敛到 100% 的速度越快,检索效果越好,特征越准确。

3) 平均精度均值

平均精度均值(mean average precision@n, mAP@n)针对待匹配集中有多个查询图像的情况,指标查询与查询图像最近匹配的 n 个样本,找到正确匹配的样本,计算此时准确率 P_i ,表示在当前已查询到的结果中正确结果所占百分比,计算公式如式(10)所示。

$$P_i = \frac{\text{当前正确查询结果数}}{\text{当前已查询的所有结果数}} \quad (10)$$

若匹配的 n 个样本中正确查询结果的数目为 m ,即得到 $P_1, P_2, \dots, P_i, \dots, P_m$, 则平均精度 (average precision, AP) 计算公式如式(11)所示,表示对本次查询中正确查询结果的准确率取平均。

$$AP = \frac{\sum_{i=1}^m P_i}{m} \quad (11)$$

假设第 j 次查询的结果为 AP_j , 查询集中查询样本数目为 N , 则 mAP 计算公式如式(12),表示对查询集中的所有查询样本的 AP 取平均。

$$mAP = \frac{\sum_{j=1}^N AP_j}{N} \quad (12)$$

需要注意的是再识别实验中有两种评估模式:Single shot 与 Multi shot。Single shot 是指候选集中仅包含每个行人目标的一个特征,而 Multi shot 是候选集中每个行人目标有多组特征的情况。本实验中的 Single shot 结果是待匹配集中每一类随机选取一个特征作为该类的表示,计算 10 次,取平均值作为最终结果。

3.3 实验结果对比

1) BIWI 数据集

对于 BIWI 数据集,以训练集中出现的 50 个人作为候选集,测试集中出现的 28 个人的行走和站立两种状态作为查询集,计算 Rank-n 指标如表 1 所示。

表 1 BIWI 数据集 Rank-n 结果
Table 1 BIWI dataset Rank-n results %

方法	Single-shot		Multi-shot	
	Rank-1	Rank-5	Rank-1	Rank-5
RGB	CESD ^[40]	-	44.66	68.84
	CCD ^[41]	10.71	35.28	13.08
	SD ^[19]	17.93	47.84	21.75
Depth	EDSKL ^[8]	27.50	64.25	34.67
	RAM ^[11]	30.10	-	45.30
RGBD	RTA ^[42]	25.40	50.00	-
	GE ^[6]	-	63.30	-
Depth	本文	70.71	92.32	71.42

由表 1 可知,在 BIWI 公开数据集上,本文所提方法的 Rank-n 指标明显优于其他现有方法。抑制衣着的形状特征 (cloth-elimination shape-distillation, CESD)^[40] 在彩色图像上进行行人体态参数测量,但该方法在极暗光照下使用时将受到一定限制。在基于深度图像的方法中,紧凑协方差描述符 (compact covariance descriptors, CCD)^[41] 与深度特征骨架特征方法 (eigen-depth feature and skeleton-based feature, EDSKL)^[8] 均基于三维点云从设计特征描述符的方向进行研究;骨架描述符 (skeleton descriptor, SD)^[19] 将所有的观测点云扭曲为一个标准姿

态后进行识别;循环注意模型 (recurrent attention model, RAM)^[11] 通过卷积神经网络和递归神经网络提取时空特征进行再识别;但这些方法并没有利用人体模型的先验知识,因而识别率较低。基于彩色图像和深度数据的方法包括强化时间注意力模型 (reinforced temporal attention, RTA)^[42] 和步态编码器 (gait encoder, GE)^[6], RTA^[42] 认为从彩色图像上学习到的特征提取对于深度图像仍然适用,于是利用彩色图像预训练模型,在深度图像上进行迁移应用;GE^[6] 基于三维骨架将行人的步态进行编码;但这类方法仍然受到光照条件的限制。本文方法基于深度数据进行行人体态参数测量,利用模型体态特征进行行人再识别,对于极暗光照的情况具有较强鲁棒性。在所列方法中取得最佳 Rank-n 指标。

2) 自采数据集

对于自采数据集,按照行人衣着将 100 段视频平均分为两组,分别作为查询集和候选集。在自采数据集上,将本文方法与自行实现的基于外观特征的多粒度网络 (multiple granularity network, MGN)^[43]、基于体态特征的视频推理算法 (video inference for human body pose and shape estimation, VIBE)^[35]、基于三维点云 PointNet++ 算法^[44] 进行比对。图 10 给出了部分样本的重识别效果,给定查询样本,重识别算法返回一个候选集样本的排序,相似度高的样本排在前面。图中绿框表示正确匹配,红框表示错误匹配。为了展示清楚起见,我们将每张深度图对应的行人照片放在图 10 中,再识别算法仅使用深度图。可以看到,大部分正确匹配的候选样本排在靠前的位置。CMC 曲线对比如图 11 所示,其中图 11 (a) 为 Single-shot 模式 CMC 曲线,图 11 (b) 为 Multi-shot 模式 CMC 曲线,可以看到本文算法相对其他 3 种算法具有明显优势。

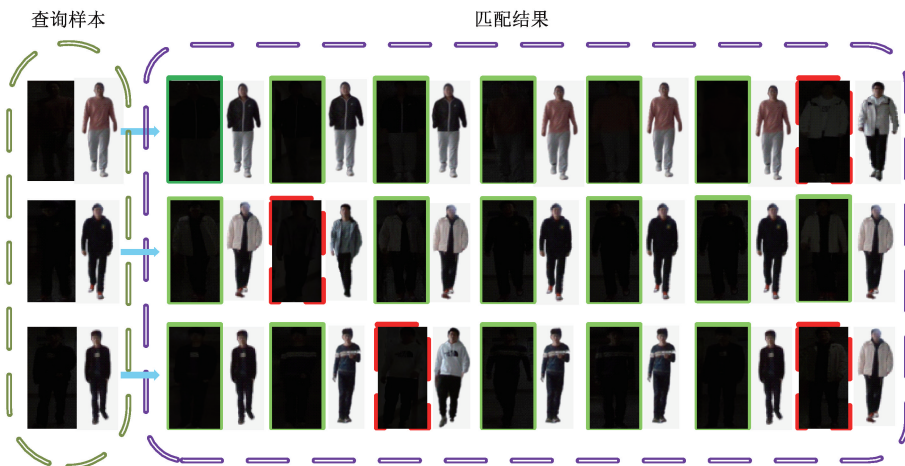


图 10 自采数据集行人再识别效果图(部分样本)

Fig. 10 Re-identification effect of self-collected dataset (partial samples)

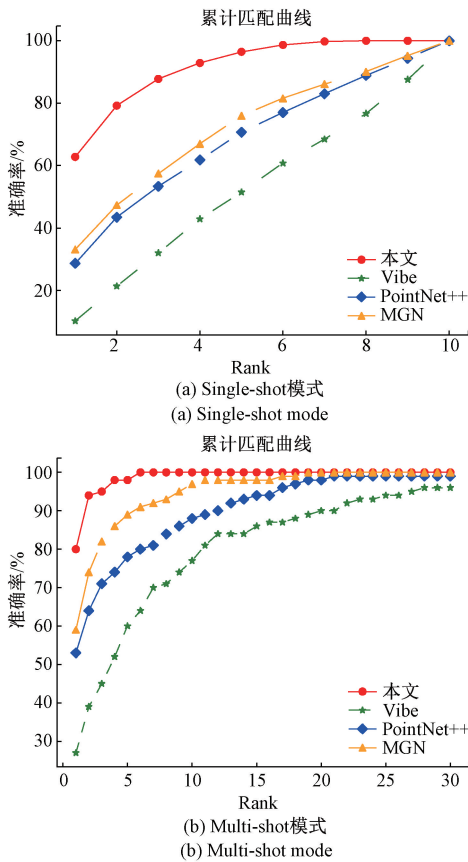


图 11 自采数据集 CMC 曲线

Fig. 11 CMC curves on the self-collected dataset

表 2 自采数据集实验结果

Table 2 Results of the self-collected dataset

方法	Single-shot			Multi-shot				
	Rank-1	Rank-5	Rank-9	Rank-1	Rank-5	Rank-10	mAP@5	mAP@10
MGN ^[43]	33.11	76.00	95.22	59.00	89.00	97.00	66.68	62.11
VIBE ^[35]	10.22	51.44	87.55	27.00	59.00	77.00	36.26	34.71
Pointnet++ ^[44]	28.66	70.66	94.33	53.00	78.00	88.00	59.61	54.32
本文	62.77	96.44	100.00	80.00	98.00	100.00	85.42	79.79

表 2 给出了不同算法的 Rank-n 与 mAP@n 指标。可以看到,在极暗光照的情况,我们所提方法在 Single-shot 与 Multi-shot 评估模式下的各项评价指标上相较于其余 3 种算法展现出明显优势,Multi-shot 模式下的 mAP@5 指标相比 MGN, VIBE, PointNet++ 分别提升约 18%, 49% 和 25%。基于二维图像的 VIBE 方法、MGN 方法会受到极暗光照的影响,无法完成精确的特征提取。本文方法与 Pointnet++ 方法依赖于点云数据,本文方法通过点云数据测量行人生物属性进行三维人体建模的方法进行再识别,在光照改变情况下表现出很强的鲁棒性。相比较于 Pointnet++ 学习全局点云特征的方法,我们方法所提特征对于行人目标更加具有判别力。

采用 t 分布-随机邻近嵌入方法 (t-distributed Stochastic Neighbor Embedding, t-SNE) 对高维特征进行降维处理,并进行可视化表示,可视化结果图如图 12 所示。不同颜色对应数据集中的不同行人目标,可以看到 MGN 算法, VIBE 算法和 PointNet++ 算法所提取的特征向量在不同行人之间分布较分散,没有呈现出良好的聚类性;本文所提特征相同标签的特征聚合较为紧密,并与其他标签的数据有较大距离。

3) 算法运行时间

所提算法主要包括数据预处理、点云配准、参数估计、特征匹配 4 个环节,表 3 给出了每个环节所需要的时间。输入的深度图像大小为 640×480 pixels,计算机 CPU

为 Intel i5-10600KF,主频 4.1 GHz,内存 32 G,采用显卡加速,显卡为一块 NVIDIA GeForce GTX 1080Ti,显存为 11 G,算法用 Python 实现。

表 3 算法运行时间

Table 3 Running time of the algorithm

过程	预处理	点云配准	参数估计	特征匹配	合计
时间	0.10	3.13	0.67	0.03	3.93

由表 3 可以看出,完成一次体态特征提取与匹配所需时间约为 3.93 s,其中最为耗时的部分为点云精配准 ICP 过程,其运算复杂度为 $O(N \log N)$,其中 N 为点云中的点数。

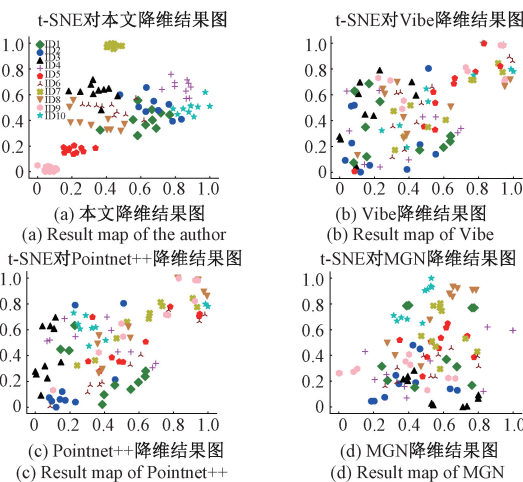


图 12 特征可视化效果图

Fig. 12 Feature visualization results

4 结 论

本文提出了一种能够在极暗光照下进行行人再识别的方法,该方法通过测量行人的三维生物信息进行再识别匹配,测量过程中仅依赖于深度数据,在极暗光照情况下算法表现出强鲁棒性。为了便于让模型学习到精确的三维信息,本文提出了基于深度数据的一系列分割、滤波、配准方法。最后,本文提出通过测量三维人体模型的参数作为行人再识别中的特征表征,在公开数据集和实测数据集上进行了实验,结果表明本文方法的性能优于当前的先进方法。

参考文献

- [1] YE M, SHEN J, LIN G, et al. Deep learning for person re-identification: A survey and outlook [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(6): 2872-2893.
- [2] MA X, ZHU X, GONG S, et al. Person re-identification by unsupervised video matching [J]. Pattern Recognition, 2017, 65: 197-210.
- [3] VARIOR R R, WANG G, LU J, et al. Learning invariant color features for person reidentification [J]. IEEE Transactions on Image Processing, 2016, 25(7): 3395-3410.
- [4] XIAO Y, CAO Z, WANG L, et al. Local phase quantization plus: A principled method for embedding local phase quantization into fisher vector for blurred image recognition [J]. Information Sciences, 2017, 420: 77-95.
- [5] 薛杉,朱虹,吴文欢.单样本的低分辨率单目标人脸识别算法 [J]. 仪器仪表学报, 2019, 40(3): 196-202.
- XUE SH, ZHU H, WU W H. Low-resolution single object face recognition algorithm with single sample [J]. Chinese Journal of Scientific Instrument, 2019, 40(3): 196-202.
- [6] RAO H, WANG S, HU X, et al. A self-supervised gait encoding approach with locality-awareness for 3D skeleton-based person re-identification [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(10): 6649-6666.
- [7] 赵东辉,杨俊友,白殿春,等.基于节点迭代模糊 Petri 网的非接触异常步态识别方法 [J]. 仪器仪表学报, 2019, 40(4): 255-264.
- ZHAO D H, YANG J Y, BAI D CH, et al. Non contact recognition method of abnormal gait based on node iteration type fuzzy Petri net [J]. Chinese Journal of Scientific Instrument, 2019, 40(4): 255-264.
- [8] WU A, ZHENG W S, LAI J H. Robust depth-based person re-identification [J]. IEEE Transactions on Image Processing, 2017, 26(6): 2588-2603.
- [9] PATRUNO C, MARANI R, CICIRELLI G, et al. People re-identification using skeleton standard posture and color descriptors from RGB-D data [J]. Pattern Recognition, 2019, 89: 77-90.
- [10] LIN K, WANG L, LIU Z. End-to-end human pose and mesh reconstruction with transformers [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 1954-1963.
- [11] HAQUE A, ALAHI A, LI F F. Recurrent attention models for depth-based person identification [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1229-1238.
- [12] HAN K, HUANG Y, GONG S, et al. 3D shape temporal aggregation for video-based clothing-change person re-identification [C]. Proceedings of the Asian Conference on Computer Vision, 2022: 2371-2387.
- [13] 孟浩,尹维考,李洪进,等.基于深度信息的指尖检测-跟踪-监督算法 [J]. 仪器仪表学报, 2019, 40(6): 171-180.
- MENG H, YIN W K, LI H J, et al. Fingertip detecting-tracking-supervising algorithm based on depth information [J]. Chinese Journal of Scientific Instrument, 2019, 40(6): 171-180.
- [14] MATZNER S, HEREDIA-LANGNER A, AMIDAN B, et al. Standoff human identification using body shape [C]. 2015 IEEE International Symposium on Technologies for Homeland Security, 2015: 1-6.
- [15] YANG Q, WU A, ZHENG W S. Person re-identification by contour sketch under moderate clothing change [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(6): 2029-2046.
- [16] BARBOSA I B, CRISTANI M, BUE A D, et al. Re-identification with RGB-D sensors [C]. European

- Conference on Computer Vision, 2012: 433-442.
- [17] WU J, JIANG J, QI M, et al. An end-to-end heterogeneous restraint network for RGB-D cross-modal person re-identification [J]. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2022, 18(4): 1-22.
- [18] UDDIN M K, LAM A, FUKUDA H, et al. Fusion in dissimilarity space for RGB-D person re-identification[J]. *Array*, 2021, 12: 100089.
- [19] MUNARO M, FOSSATI A, BASSO A, et al. One-shot person re-identification with a consumer depth camera[M]. *Person Re-Identification*, Springer, London, 2014: 161-181.
- [20] LOPER M, MAHMOOD N, ROMERO J, et al. SMPL: A skinned multi-person linear model [J]. *ACM Transactions on Graphics*, 2015, 34(6): 1-16.
- [21] ZHANG Z, TRAN L, LIU F, et al. On learning disentangled representations for gait recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 44(1): 345-360.
- [22] XU W, LIU H, SHI W, et al. Adversarial feature disentanglement for long-term person re-identification [C]. *IJCAI*, 2021: 1201-1207.
- [23] CHEN J, JIANG X, WANG F, et al. Learning 3D shape feature for texture-insensitive person re-identification[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 8146-8155.
- [24] IMANI Z, SOLTANIZADEH H. Person reidentification using local pattern descriptors and anthropometric measures from videos of kinect sensor[J]. *IEEE Sensors Journal*, 2016, 16(16): 6227-6238.
- [25] PALA F, SATTI R, FUMERA G, et al. Multimodal person reidentification using RGB-D cameras[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2015, 26(4): 788-799.
- [26] MAGNENAT-THALMANN N, THALMANN D. Human body deformations using joint-dependent local operators and finite-element theory [M]. *Making Them Move: Mechanics, Control, and Animation of Articulated Figures*, 1991: 243-262.
- [27] LEWIS J P, CORDNER M, FONG N. Pose space deformation: A unified approach to shape interpolation and skeleton-driven deformation[C]. *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, 2000: 165-172.
- [28] ANGUELOV D, SRINIVASAN P, KOLLER D, et al. SCAPE: Shape completion and animation of people[J]. *ACM Transactions on Graphics*, 2005, 24(3): 408-416.
- [29] OSMAN A A A, BOLKART T, BLACK M J. STAR: Sparse trained articulated human body regressor [C]. *European Conference on Computer Vision*, 2020: 598-613.
- [30] BOGO F, KANAZAWA A, LASSNER C, et al. Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image[C]. *European Conference on Computer Vision*, 2016: 561-578.
- [31] KOLOTOUROS N, PAVLAKOS G, BLACK M J, et al. Learning to reconstruct 3D human pose and shape via model-fitting in the loop[C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 2252-2261.
- [32] BLACK A, BLACK M J, JACOBS D W, et al. End-to-end recovery of human shape and pose[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7122-7131.
- [33] PAVLAKOS G, ZHU L, ZHOU X, et al. Learning to estimate 3D human pose and shape from a single-color image [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 459-468.
- [34] PAVLLO D, FEICHTENHOFER C, GRANGIER D, et al. 3D human pose estimation in video with temporal convolutions and semi-supervised training [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 7753-7762.
- [35] KOCABAS M, ATHANASIOU N, BLACK M J. Vibe: Video inference for human body pose and shape estimation [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 5253-5263.
- [36] JIANG H, CAI J, ZHENG J. Skeleton-aware 3D human shape reconstruction from point clouds[C]. *Proceedings*

- of the IEEE/CVF International Conference on Computer Vision, 2019: 5431-5441.
- [37] WANG K, XIE J, ZHANG G, et al. Sequential 3D human pose and shape estimation from point clouds[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 7275-7284.
- [38] ZHOU B, FRANCO J S, BOGO F, et al. Reconstructing human body mesh from point clouds by adversarial gp network[C]. Proceedings of the Asian Conference on Computer Vision, 2020.
- [39] BESL P J, MCKAY N D. A method for registration of 3D shapes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992, 14(2): 239-256.
- [40] QIAN X, WANG W, ZHANG L, et al. Long-term cloth-changing person re-identification[C]. Proceedings of the Asian Conference on Computer Vision, 2020:71-88.
- [41] FEHR D, CHERIAN A, SIVALINGAM R, et al. Compact covariance descriptors in 3d point clouds for object recognition[C]. IEEE International Conference on Robotics and Automation, 2012: 1793-1798.
- [42] KARIANAKIS N, LIU Z, CHEN Y, et al. Reinforced temporal attention and split-rate transfer for depth-based person re-identification[C]. Proceedings of the European Conference on Computer Vision, 2018: 715-733.
- [43] WANG G, YUAN Y, CHEN X, et al. Learning discriminative features with multiple granularities for

person re-identification[C]. Proceedings of the 26th ACM International Conference on Multimedia, 2018: 274-282.

- [44] QI C R, YI L, SU H, et al. PointNet ++ deep hierarchical feature learning on point sets in a metric space[C]. Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 5105-5114.

作者简介



刘明洋, 2020年于济南大学获得学士学位, 现为北京航空航天大学硕士研究生, 主要研究方向为目标识别。

E-mail: lmy1998@buaa.edu.cn

Liu Mingyang received his B. Sc. degree from University of Jinan in 2020. He is currently a master student at Beihang University. His main research interest is object recognition.



万九卿(通信作者), 2003年于北京航空航天大学获得博士学位, 现为北京航空航天大学自动化科学与电气工程学院副教授, 主要研究方向为目标检测, 跟踪与识别。

E-mail: wanjiuqing@buaa.edu.cn

Wan Jiuqing (Corresponding author) received his Ph. D. degree from Beihang University in 2003. He is currently an associated professor at Beihang University. His main research interests include object detection, tracking and recognition.