

DOI: 10.11992/tis.201607010

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20170626.1739.010.html>

目标再确认中的优化扩散距离相似性度量

曹伟, 韩华, 王裕明, 孙宪坤

(上海工程技术大学 电子电气工程学院, 上海 201620)

摘要: 跨摄像头目标再确认是多摄像头监控领域中一个亟需解决的难点问题, 如何获得准确率更高的距离度量算法成为解决该难点的关键。为此本文在提取优秀的多特征基础上, 建立了一种无需训练, 适应更多场景的度量算法: 优化扩散距离相似性度量, 用于跨摄像头目标再确认。在高维扩散距离的基础上, 加入高斯金字塔图像采样和加权性优化处理, 用于提高相似空间向量的辨别力, 并提高计算效率。通过对高维扩散距离的二次优化建立起最终的相似性度量函数。最后对 VIPeR 和 ETHZ 数据库中的图片进行多次目标再确认实验, 排名第一的图片的正确匹配率达到了 50.5%。实验结果表明本文算法取得了较好的匹配结果。

关键词: 优化扩散距离; 相似性度量; 多特征融合; 目标再确认

中图分类号: TP391.9 **文献标志码:** A **文章编号:** 1673-4785(2018)02-0269-12

中文引用格式: 曹伟, 韩华, 王裕明, 等. 目标再确认中的优化扩散距离相似性度量[J]. 智能系统学报, 2018, 13(2): 269-280.

英文引用格式: CAO Wei, HAN Hua, WANG Yuming, et al. Target re-identification based on optimized diffusion distance[J]. CAAI transactions on intelligent systems, 2018, 13(2): 269-280.

Target re-identification based on optimized diffusion distance

CAO Wei, HAN Hua, WANG Yuming, SUN Xiankun

(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: Target re-identification via cross-camera is a challenging problem in the field of multi-camera surveillance. How to get a more accurate distance measurement algorithm is the key to solve this difficult problem. So, in this paper, we established a new measurement algorithm without training based on extracting more excellent features to do target re-identification, which is optimized diffusion distance. On the basis of high-dimensional diffusion distance, Gaussian pyramid image sampling and weight optimization are added to improve the discrimination of similar space vectors and increase computational efficiency. The final similarity measure function is established by the second optimization of the high-dimensional diffusion distance. At last, we do numerous target re-identification experiments based on databases VIPeR and ETHZ. The matching rate of rank first image can reach 50.5%. The experimental results show that the algorithm proposed in this paper has good performance.

Keywords: optimized diffusion distance; similarity measure; multi-feature fusion; target re-identification

跨摄像头目标再确认是目前多摄像机智能监控领域的一个重点研究方向^[1], 如何在不同的光照条件、不同的行人姿态以及不同尺度甚至遮挡等复杂环境下实现不同监控摄像头下的同一目标的再确

认, 是现在智能监控领域研究的热点。能否在复杂的环境中快速地匹配到同一目标的关键点在于能否找到目标在时间、空间上的不变特性以及选择合适的匹配算法。

目前最先进的算法对 VIPeR(viewpoint invariant pedestrian recognition) 数据库的视频目标进行再确认的正确率为 40%, 处于比较低的水平^[2]。根据上文所述, 目标再确认主要的研究方法分为两类:

收稿日期: 2016-07-13. 网络出版日期: 2017-06-26.

基金项目: 国家自然科学基金项目(61305014); 国家留学基金管理委员会项目(201508310033); 上海市教育委员会和上海市教育发展基金会“晨光计划”(13CG60); 上海高校青年教师培养资助计划(ZZGJD13006); 上海工程技术大学人才行动计划(2017RC112015, nhrc-2015-11).

通信作者: 韩华. E-mail: 2070967@mail.dhu.edu.cn.

目标外观特征类和度量算法类。而目标外观特征类的研究又可以归纳为3类:1)依据目标本身的整体特征进行描述,如文献[3-9]通过建立整体相关特征颜色直方图进行目标再确认;2)基于目标本身的局部特征进行目标匹配,如文献[10-13]通过训练大量样本得出的数据作为匹配函数,依据目标本身的颜色特征、纹理特征进行目标的再确认;3)通过整体和局部多特征融合来匹配目标,如文献[14-18]通过人体本身的局部LBP(local binary patterns)特征、SIFT(scale-invariant feature transform)特征以及整体的颜色特征进行描述,来完成目标再确认。度量算法类的研究方向相对特征类方法而言,对目标外观特征的要求较低,更多地依赖选择和训练优秀的度量函数。比如,ERSVM方法^[19]采用基于相对距离排序的SVM分类器集成多个特征向量权重;文献[20]采用多核SVM学习的方法融合颜色和视觉单词直方图;文献[21]采用直方图交叉核;文献[22]采用夹角余弦法等相似度量函数进行目标匹配。

以上两类方法在目标再确认过程中,度量算法类对特征的选择要求较低,通常能取得更好的再确认效果。但其训练的时间和空间复杂度较高,且场景和图库变化时需要重新训练。目标外观特征类方法则需要人为地寻找和设计较好的特征,该方法的性能依赖于特征的优劣程度。本文受两类方法优点的启发,在特征类方法的思路下设计和构造优秀的特征;同时在度量算法类方法的思路下,引入物理学中的热扩散原理作为度量函数^[23],并最终建立基于优化的扩散距离的度量函数完成目标再确认。

基于上文所述,引入度量算法的基础来源于扩散距离。扩散距离是物理学中用于描述热力场中存在的一种扩散现象的,国外学者Ling等^[23]曾提出了一种非线性基于扩散现象和距离判断方式二者结合的扩散距离,它是模拟热扩散和记录随机扩散遍历的时间计算两点之间的距离。这使得扩散距离具有可以很好地描述目标间相似性的特性,同时,传统度量算法存在诸多缺点,比如:对图像形变比较敏感;缺少考虑像素间的空间关系;较小的形变就能使距离计算变化较大,从而产生误匹配等,而扩散距离的引入克服了以上不足。基于扩散距离的众多优点,因此本文引入扩散距离,并对其进行优化,构造出优秀的相似性度量函数,以完成对行人目标的再确认。

1 多特征描述

目标再确认过程中,描述每个目标的外观相当于从目标身上找到可辨识性的特征。在描述目标

特征时,直方图是目前应用最广泛的外观描述器。但是,大部分基于直方图的方法,其不足之处是缺少特征分布的空间信息。比如,基于直方图的方法不能分辨出一个穿着白色衬衫和黑色裤子的行人与另一个穿着黑色衬衫和白色裤子的行人是否是不同的。

因此,文中提出了一个新颖而简单的方法:基于两级区域的直方图提取方法。首先,提出的第一级区域为:行人整个身体区域(H_{global})颜色直方图;提出的二级区域为:行人的上半身区域(H_{tor})和腿部区域(H_{leg})如图1所示。图中(a)、(d)为一级区域身体直方图,图(b)、(e)为二级区域身躯直方图,图(c)、(f)为二级区域腿部直方图。

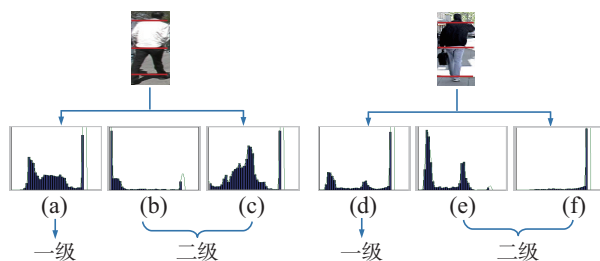


图1 两级区域直方图

Fig. 1 Two levels of area histogram

图1(a)、(d)一级所提取的 H_{global} 直方图,(d)直方图中的3个峰值在(a)直方图中可以找到相似的对应,因此单纯靠全局直方图很难发现目标是两个不同的人;通过对图1中的两个原始图像分别提取身躯段和腿部段的 H_{tor} 和 H_{leg} 直方图,对应于二级,如图1左图中的(b)、(c)和图1右图中的(e)、(f)所示,对比图1(b)和图1(e),可发现两者身躯直方图明显不同,同时对比图1(c)和图1(f)也能发现两者的腿部直方图也是明显不同的,可以通过局部直方图的差异,发现两目标并不是同一个人。通过上述方法可以有效地提取颜色的空间信息,并区分出图1两个行人的不同。

人体目标本身具有很多稳定不变的特征,这些不变特征在应对因光照变化或细微的肢体动作变化时具有很好的鲁棒性。因此,为了更好利用这些优势特征来克服由于光照变化或其他因素所带来的干扰,本文在下文中会从整体特征和局部特征两方面来介绍本文所提取的这些优势多特征,实验表明,使用这些多特征中的任何一类进行匹配时,识别率都不太理想,但结合下文中的多特征进行目标再确认时,可以很大程度上提高目标的识别率。

1.1 整体特征

整体特征使用基于一级区域(图1中Level 1)的RGB颜色直方图进行粗识别。在提取RGB颜色直方图时,本文将R、G、B这3个通道分别进行量化。

而量化过程中存在两个极端问题: 粗量化会造成大量的可辨别性信息被放弃, 细量化会生成稀疏直方图。本文在 3 个通道使用相同的量化级数 32bins。通过大量实验验证了本文所使用的 32bins 具有较好的可辨别性和准确性。通过图 1 可以很清楚地体现出来。

1.2 局部特征

局部特征相对整体特征而言更加注重局部小范围特征的不变性, 采用更细腻的特征描述子或者具有较强鲁棒性的特征来进行匹配。观察大部分行人目标的 H_{tor} 和 H_{leg} 特征发现: 最显著的局部差异特征仍是颜色特征, 实验过程中发现使用颜色特征应对摄像机镜头旋转变化的差异具有很好的辨别性。所以在下文中的局部特征表述过程中, 基于上文中二级区域 H_{tor} 和 H_{leg} 的基础上, 为了使人体目标取得更佳的匹配效果使用 GrabCut^[24]背景去除法, 将干扰背景(头部也当作背景)进行去除, 同时将 RGB 颜色通道模型和加权 HSV 颜色通道模型作为主要的局部特征; 而另一方面考虑到局部人体目标的表象和形状能够被梯度或边缘的方向密度分布很好地描述, 方向梯度直方图 (histogram of oriented gradient, HOG) 特征可以通过计算和统计图像局部区域的梯度方向直方图来构成特征, 并且能够对图像的几何和光学形变都能保持很好的不变性, 所以, 将结构性 HOG 描述子作为次要的局部特征。

1.2.1 局部 RGB 颜色特征

观察数据库中行人目标发现: 去除头部的影响后, 由于目标行人上身所穿的衣服和下身所穿的裤子在颜色上往往会表现出明显的差异, 所以本文在二级区域 H_{tor} 和 H_{leg} 基础上使用局部 RGB 颜色直方图匹配。通过观察图 2(b): 掩膜背景图 (图中干扰背景和头部已经去除) 和测试样本库里的图片 (国际通用标准库 VIPeR 中的样本图片大小: 48×128) 分析得到: 头部和身躯段的平均分界线在第 20 行, 身躯段和腿部段的平均分界线在 62 行。为了减少背景对匹配结果的干扰, 如图 2(c) 所示, 将身躯段去除左右边界各 4 列, 即使用整张图像的第 20~60 行, 第 6~44 列作为身躯段窗口; 同时, 在腿部段部分去除最下面的 8 行, 以及左右各 14 列, 使用整张图像的第 66~122 行, 第 14~36 列, 作为腿部段窗口。在量化过程中, 同整体特征量化数相同, 提取每个通道对其量化后的直方图统计如图 2。

图 2 中, 分段处理和图 1 中的两级区域一致, 将图像中的人物按照不对称原则分为头部、身躯和腿部 3 个水平段, 并将图像中的人物的身躯和腿部按照对称原则分为两个垂直段, 并且去掉权重比较小的头部像素点, 得到的实际效果如图 2(c) 所示。图 2(d) 代表实验过程中所绘制的图 2(b) 被去除头

部和背景干扰之后, 身躯和腿部的 R 、 G 、 B 分量直方图。其中, 图 2(d) 中的横轴, 0~31 代表身躯段和腿部段上的 R 颜色通道统计直方图 (其中 0~15 代表身躯段, 16~31 代表腿部段); 32~63 代表身躯段和腿部段上的 G 颜色通道统计直方图; 64~95 代表身躯段和腿部段上的 B 颜色通道统计直方图。

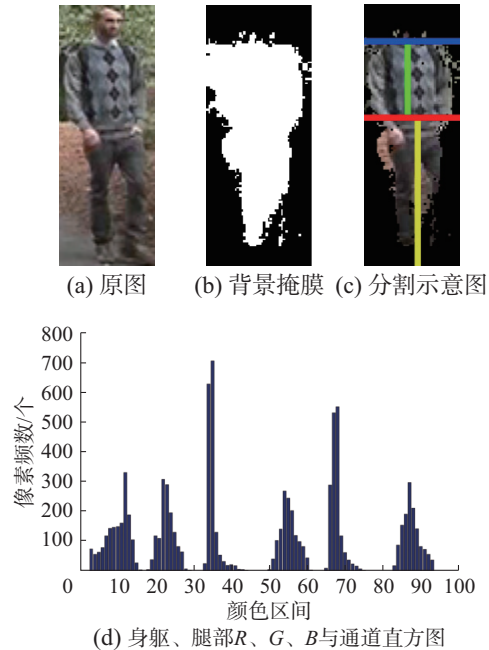


图 2 局部 RGB 颜色直方图

Fig. 2 Local RGB color histogram

局部 RGB 颜色直方图充分利用了空间分布信息, 可以有效区分出上文所提到的上身下身穿衣颜色的不同。在局部特征描述方面的应用中, 可以较好地实现应对图像旋转和平移的改变所带来的影响, 并且还可借助归一化技术, 来达到不受图像尺度的改变所带来的影响, 充分利用了局部颜色特征的优势。

1.2.2 加权 HSV 特征

根据人眼视觉特性, 颜色信息中 HSV 特征相比较 RGB 特征, 在应对光照变化方面, 可以很大程度上忽略图像亮度变化所带来的影响; 在二级区域中对 HSV 模型使用高斯分布的加权处理, 可以在提取空间信息基础上更好地对特征权重化处理, 所以本文选择 HSV 颜色模型作为局部特征之一来使用。

考虑到由于人体目标特征分布比例的差异性, 有些显著特征在人体身上的某些部位所占比例很重, 而有些特征则在人体身上的某些部位所占比例很小。通过观察图 3(b) 发现: 身躯和腿部中间部分的颜色特征比较明显, 对图像匹配时的影响会比较重, 所以相应地给予较大的权重分配, 越远离垂直坐标的颜色特征对匹配性能的提高影响较小, 甚至会降低匹配性能, 所以相应地给予远离垂直坐标的

点以较小的权重分配。根据人体不同部位所对应的分段坐标对 HSV 直方图进行高斯分布的加权^[16]处理。

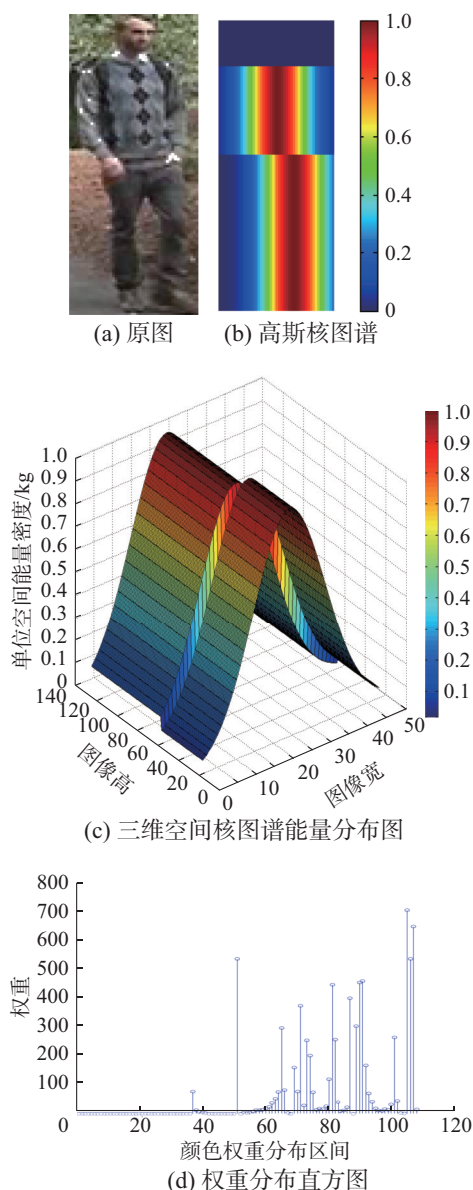


图3 人体加权 HSV 处理图

Fig. 3 Body weighted HSV processing figure

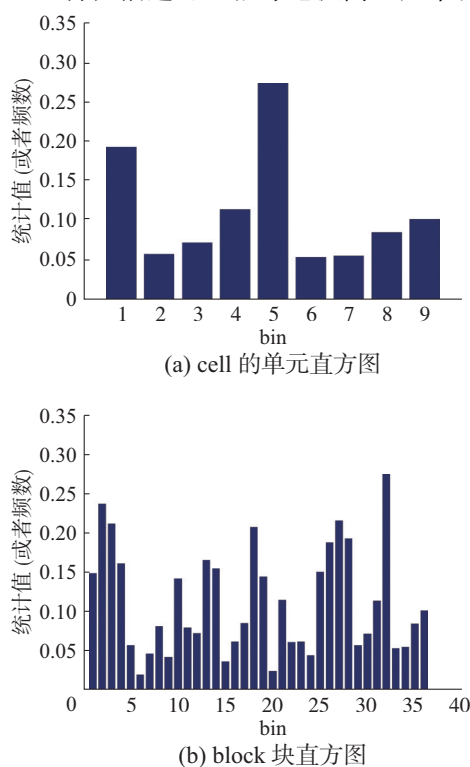
如图3所示,为了更形象地表现人体身上特征权重的分布,文中提取人体不同位置对应的高斯分布图谱3(b)和三维空间中的核能量图谱3(c)。考虑到光照的影响会对匹配精度造成很大的影响,因而为了抑制光照变化,将 HSV 中的亮度值(V)分配较小区间段。文中的H、S、V通道分别分成16、16、4段,进而得到一个加权的108段(其中0~31段为头部段,已被忽略计算;32~71段为身躯段;72~107为腿部段)HSV直方图,其相应的实验效果如图3(d)。

1.2.3 HOG 特征

HOG 是目标检测常用的特征描述子。目标局部的表象和形状能够被梯度或边缘的方向的密度分布很好地描述^[24]。具体的实现方法是:首先将图像

分成小的连通区域,我们把它叫细胞单元^[25]。然后采集细胞单元中各像素点梯度或边缘的方向直方图。最后把这些直方图组合起来就可以构成特征描述器。为了提高性能,还可以把这些局部直方图在图像的更大的范围内(也可称为区间或 block)进行对比度归一化,所采用的方法是:先计算各直方图在这个区间(block)中的密度,然后根据这个密度对区间中的各个细胞单元做归一化。本文在实际操作中使用的参数为:窗口大小统一为 48×128 , block 块大小为 16×16 ,每个 block 划分为4个 cell,并且 block 每次滑动8个像素(即一个 cell 的宽),以及梯度方向划分为9个区间,最后在 $0 \sim 360^\circ$ 范围内统计,将 cell 的梯度方向 360° 分成9个方向块,也就是每 $360^\circ/9=40^\circ$ 分到一个方向,方向大小按像素边缘强度加权,对 cell 内每个像素用梯度方向在直方图中进行加权投影(映射到固定的角度范围),其相应的实验效果如图4(c)所示。实验过程中,根据上文相应参数对图4(a)进行 HOG 特征提取,提取过程中的实验效果如图4所示。

从图中可以清晰的看出:本文将 cell 的梯度方向以 40° 为一个单位,分成一个方向;图4(d)代表实验过程中提取的归一化后 HOG 3 维特征曲面图。由于 HOG 方法是在图像的局部细胞单元上操作的,所以它对图像几何和光学的形变能够保持很好的不变性,归一化后能对光照变化和阴影获得更好的效果。它的优点不仅包括对图像几何和光学形变方面能保持良好的不变性,还包括归一化后允许行人有细微的肢体动作,且不影响检测效果,多特征融合 HOG 特征描述可以很好地提高匹配率。



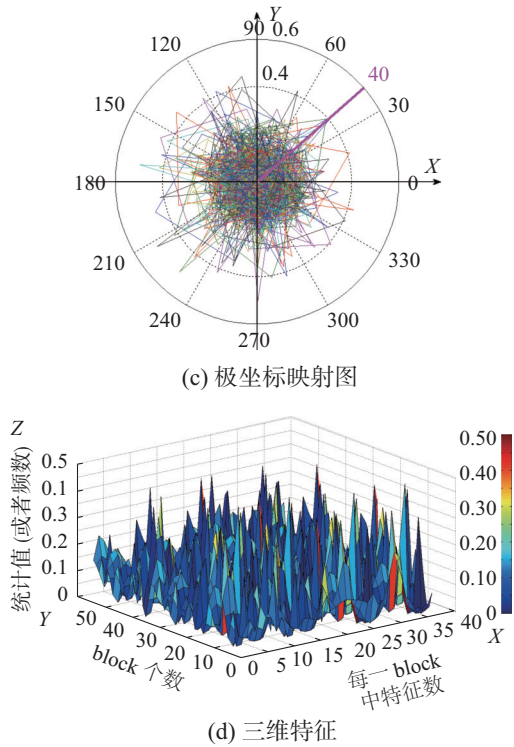


图4 HOG特征描述子实验效果图

Fig. 4 HOG feature descriptor experimental result

2 相似性度量函数

在相似性度量函数的建立方面,传统度量算法大多基于欧氏距离或者余弦距离等进行优化。然而,欧式距离和余弦距离本身存在诸多缺点,比如:欧式距离不能从高维特征向量恢复到具有低维的几何机构,对图像的形变比较敏感和缺少像素间的空间关系;余弦距离存在位置判定上差异性太大以及对数值的不敏感导致结果的众多误差等缺点。考虑到传统相似性度量算法的不足,本文引进物理学温度场扩散距离^[26]的概念来建立相似性度量函数。扩散距离的本质来源于物理学中的温度场,它不同于欧式距离中的距离长短决定相似度大小的标准。扩散距离中的相似度取决于温度扩散路径的多少以及热扩散的快慢。本文所使用的基于优化扩散距离相似性度量函数用于衡量个体在空间上存在的距离,距离越远,说明个体间的差异越大,反之,差异越小。优化扩散距离在一定程度上弥补了欧式和余弦距离的不足的同时,在匹配性能上显示了较优的表现^[26]。其原理如下:

首先,假设 $\text{hist}_1(x)$ 和 $\text{hist}_2(x)$ 为两个一维分布,其差异可以表示成:

$$d(x) = \text{hist}_1(x) - \text{hist}_2(x) \quad (1)$$

如果存在一个独立的温度场 $T(x, t)$ 中,在时间 $t=0$ 时,有 $T(x, 0) = d(x)$ 。那么温度场 $T(x, t)$ 的计算公式为

$$\begin{cases} T(x, t) = T_0(x) \cdot \phi(x) \\ T_0(x) = T(x, 0) = d(x) \\ \phi(x, t) = \frac{1}{\sqrt{2\pi t}} \exp^{-x^2/2t} \end{cases} \quad (2)$$

那么一维分布 $\text{hist}_1(x)$ 和 $\text{hist}_2(x)$ 的扩散距离定义为

$$\hat{K}(\text{hist}_1, \text{hist}_2) = \int_0^{\bar{t}} k(|T(x, t)|) dt \quad (3)$$

式中: \bar{t} 是一个正常数,且 $k(|T(x, t)|) = \int_{-\infty}^{+\infty} |T(x, t)| dx$ 。

由于一维扩散距离的应用没有多大实际意义,所以推广到高维且引入高斯金字塔对扩散距离进行优化,进一步提高了相似性度量函数的计算效率。

假设有两个 p 维直方图 $\text{Hist}_1(X)$ 和 $\text{Hist}_2(X)$,其中 X 为 p 维的实矩阵,由于高斯金字塔中图像的平滑子采样不会产生混淆现象,能有效的描述温度场 $T(x, t)$ 的离散化连续性扩散过程,所以本文使用基于成熟的高斯金字塔的交替距离定义高维扩散距离,即

$$K(\text{Hist}_1, \text{Hist}_2) = \sum_{l=0}^L k(|d_l(x)|) \quad (4)$$

式中:

$$d_0(X) = \text{Hist}_1(X) - \text{Hist}_2(X), l = 0 \quad (5)$$

$$d_l(X) = [d_{l-1}(X) * \phi(X, \sigma)] \downarrow_{2/3}, l = 1, 2, \dots, L \quad (6)$$

$$\phi(X, \sigma) = \frac{1}{(2\pi)^{p/2} \sigma} \exp^{-X^T X / 2\sigma^2} \quad (7)$$

式(6)中的 \downarrow 代表降采样,如果采样值偏小,有效信息会缺失;如果采样值偏大,会带来计算效率降低的问题,综合以上考虑,本文采样值大小设定为经验值 $\downarrow_{2/3}$,即取采样值的 $2/3$ 。 L 是高斯金字塔的层数, σ 是高斯函数 $\phi(X, \sigma)$ 的标准差,式(7)中 p 代表直方图的维数。然后,使用 L_1 范数标准对式(4)进行简化^[26]:

$$K(\text{Hist}_1, \text{Hist}_2) = \sum_{l=1}^L d_l(X) \quad (8)$$

从式(6)可以看出, d_l 中的向下采样操作在运算量上呈指数减少,同时使用采样值的 $2/3$,又能够最大限度保留样本值的信息;其次,高斯滤波卷积操作是线性运算,在计算效率上明显有优势。但是,优化后的扩散距离虽然在计算效率上提高不少,但是由式(4)和(5)可以知道,优化扩散距离在表示空间向量直方图 $\text{Hist}_1(X)$ 和 $\text{Hist}_2(X)$ 之间的累积差时,缺少对内部单个元素之间差异的表示。虽然使用式(8)可以直接应用到相似距离度量中,但其忽略了空间向量对应元素之间的相似性的影响,因而其精确度会存在误差。由于不同 $\text{Hist}(X)$ 中的每个元素代表不同的特征属性,比如 wHSV(weighted

HSV, 加权 HSV) 直方图中的第 z 个元素代表第 z 个区域内的像素点的总个数或者代表图像中的颜色值为 z 。所以, 进行相似性距离度量时, 既要考虑累积差异的影响, 也要考虑对应元素间相似性的影响。假设: $\text{Hist}_1(X) = [3, 7, 10]$, $\text{Hist}_2(X) = [1, 1, 5]$, $\text{Hist}_3(X) = [3, 6, 2]$, 为简化计算, 令 $\phi(X, \sigma) = 1$, 不使用降采样, 相应的 $K(\text{Hist}_1, \text{Hist}_2) = \sqrt{65} = K(\text{Hist}_1, \text{Hist}_3)$, 无法区分哪两个直方图更相似。通过观察 $\text{Hist}_1(X)$ 、 $\text{Hist}_2(X)$ 和 $\text{Hist}_3(X)$ 发现, $\text{Hist}_{13}(X)$ 相比 $\text{Hist}_{12}(X)$ 拥有更多相同或者相近的元素个数。从直观上, 只能定性的认为 $\text{Hist}_1(X)$ 与 $\text{Hist}_3(X)$ 应该比 $\text{Hist}_1(X)$ 与 $\text{Hist}_2(X)$ 更相似。

通过以上分析, 在计算空间向量直方图之间距离时, 不应该将内部元素归置一样的权值, 而应该对那些元素值间对应相同或者相近的给予更大的权值。所以, 本文在优化扩散距离基础上进行二次优化, 提出加权式扩散距离, 相应表达式为

首先, 将式 (5) 改写成

$$d_0(X) = \sqrt{\sum_{i=0}^n (\text{Hist}_{1i}(X) - \text{Hist}_{2i}(X))^2} \quad (9)$$

式中: $1i$ 和 $2i$ 分别代表直方图 1 中第 i 个元素和直方图 2 中第 i 个元素, $i \in \{0, n\}$; 式 (5) 和 (9) 代表类似的含义, 即两个直方图差异的距离等同于两个直方图中每一个对应元素差异的累积。那么计算时使用的公式即为式 (9)。

其次, 将式 (9) 改写成

$$\begin{cases} d_0(X) = \sqrt{\sum_{i=0}^n \frac{r_i}{R} (\text{Hist}_{1i}(X) - \text{Hist}_{2i}(X))^2} \\ r_i = \exp^{-|\text{Hist}_{1i} - \text{Hist}_{2i}|/\sigma}, \sigma = 1 \end{cases} \quad (10)$$

式中: R 为归一化因子, $R = \sum r_i$; r_i 为权值; σ 为调节因子。根据式 (10) 计算, 同样按照 $K(\text{Hist}_1, \text{Hist}_2) = 2.337$, $K(\text{Hist}_1, \text{Hist}_3) = 1.025$, 因而能从直观上, 定量的得出 $\text{Hist}_1(X)$ 与 $\text{Hist}_3(X)$ 应该比 $\text{Hist}_1(X)$ 与 $\text{Hist}_2(X)$ 更相似。由此可以看出, 本文提出的对优化扩散距离的二次优化, 对计算匹配距离时拥有更好的区分性, 精度进一步提高。

图 5 显示了不同时间段扩散距离的状态。图 5(a) 中显示同一个人不同视角下的直方图差异的扩散过程; 图 5(b) 中显示不同的人在不同视角下的直方图差异的扩散过程。通过两幅图的对比我们可以看到, (a) 图相比 (b) 图衰减更快, 即同类人直方图的扩散距离小于不同类人的。

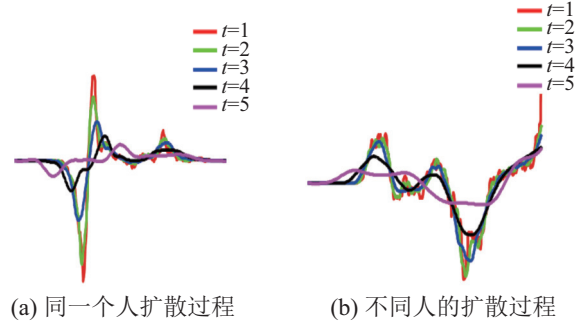


图 5 扩散距离匹配过程
Fig. 5 Diffusion distance matching process

3 目标再确认

3.1 归一化处理

单独使用上述 3 种特征描述中任何一种所建立的相似性度量函数都不能达到较好的匹配效果, 所以本文使用多特征融合进而建立起的相似性度量函数进行目标再确认, 对上述 3 种特征提取相应特征直方图, 然后分别计算相应的优化扩散距离, 为防止由于不同特征算出的扩散距离有较大差异而影响目标再确认结果, 对每类特征描述计算所得到的扩散距离都做归一化处理。

3.2 总体相似性度量函数的建立

目标再确认中, 通常会有两组图像数据: 待识别目标 I_A 和候选目标 I_B 。本文利用两组图像数据库, 对上述 3 类特征得到的扩散距离都做相应的归一化处理之后, 使用式 (8) 结合式 (10) 来计算不同图像库 I_A 和 I_B 中的图像之间的匹配距离 $d(I_A, I_B)$ 。

$$\begin{aligned} d(I_A, I_B) = & \alpha \cdot d_g(\text{RGB}_g(I_A), \text{RGB}_g(I_B)) + \\ & \beta \cdot d_{lca}(\text{RGB}_{lca}(I_A), \text{RGB}_{lca}(I_B)) + \\ & \lambda \cdot d_{lca}(\text{wHSV}_{lca}(I_A), \text{wHSV}_{lca}(I_B)) + \\ & \mu \cdot d_{lca}(\text{HOG}_{lca}(I_A), \text{HOG}_{lca}(I_B)) \end{aligned} \quad (11)$$

式中: g 代表一级区域; lca 代表二级区域; $\alpha, \beta, \lambda, \mu$ 分别代表不同特征直方图的权重, $\alpha + \beta + \lambda + \mu = 1$ 。本文对于粗识别过程中的整体 RGB 特征直方图分配经验值为 $\alpha = 0.2$ 影响因素较大的局部 RGB 直方图给予较大的权值, 选择目前最佳的经验值 $\beta = 0.4$; 对于影响因素较弱的加权 HSV 直方图给予目前最佳的经验值 $\lambda = 0.3$; 对于影响因素最小的局部 HOG 直方图给予相对较弱的经验值 $\mu = 0.1$ 。总的评价准则函数即总的相似性度量函数为式 (11)。

4 实验验证

上文所述理论相比传统再确认算法在特征选择和度量算法上都进行了优化, 同时, 本文进行多组实验验证了本文所提出优化算法的可行性, 下文会分别从实验运行效果和实验数据统计来阐述本文的

有效性。

4.1 实验运行效果

从实验运行结果来看应用本文原理编程实现的匹配成功率较为理想, 能够很好地实现目标再确认的功能, 如图 6(a)、图 7 所示。下面显示的效果图中图 6(a) 和图 7 是目标再确认中显示正确的效果图, 图 6(a) 图显示了实验过程中成功再确认的案列, 而图 7 是实验完成后所有再确认成功的部分汇总。尽管本文算法可以很好地实现目标再确认, 但是, 仍存在少数失败的案列, 比如图 6(b) 所示。



图 6 匹配效果图
Fig. 6 Matching figures

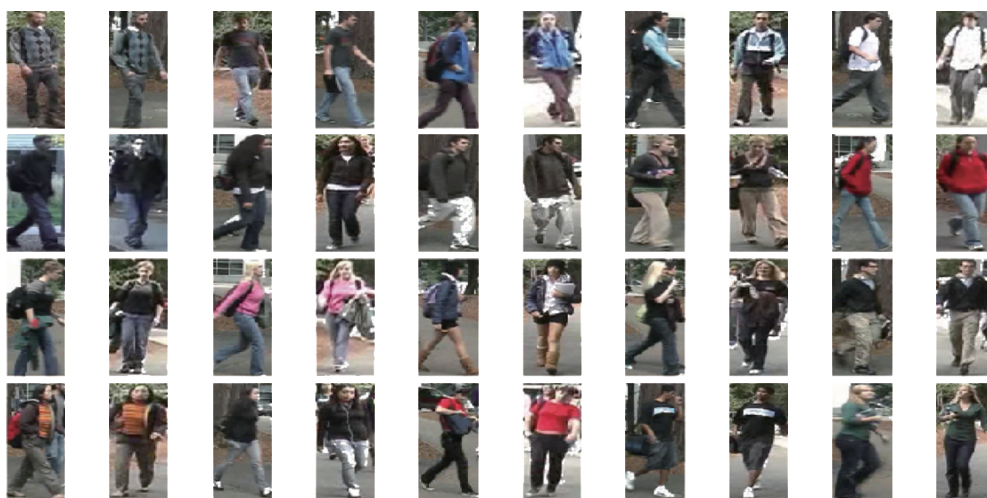


图 7 目标再确认成功对显示图
Fig. 7 Target to reconfirm success of figure

图 6(a) 是实验中匹配正确的效果图, 从图中可以看出两个摄像头中所呈现的目标的颜色特征有很大相似性, 在图 6(a) 中行人身上的局部颜色特征, 如上身身躯和下身身躯的颜色差异性很明显, 其他特征的干扰比较小, 因而在特征表达更具体, 更有辨别性。所以在算法中可以比较容易地将目标进行再确认。相比之下, 图 6(b) 中两个行人在进行躯体分割, 特征提取等过程后, 由于用于提取特征而分割出来的部分特征较为接近 (左右两图上身衣服和下身裤子的颜色分别比较接近), 因而导致出现了错误的再确认结果。

本文通过随机选取图片库中 45 对图片进行目标再确认实验, 将部分匹配成功对在 MATLAB 画布中显示在一起 (图 7), 可以更加直观地看出本文的算法达到了良好的再确认效果。具体性能指标可以通过下文实验数据统计评估结果进行说明。

4.2 实验数据统计评估

评价目标再确认的性能是否优良并没有统一的标准, 但目前在评估行人再确认性能方面主要采用的是累积匹配特性曲线 CMC^[27] (cumulative match characteristic)。CMC 曲线反映的是在目标图库前

n 个最相似候选目标内找到正确匹配图片的概率^[27]。实验中随机抽取 50 对行人目标进行测试, 实验的总次数是 5 次。最后给出平均 CMC 曲线 (横坐标为排名, 纵坐标为识别率)。假设曲线上有一点 (m, n) , 代表排名前 m 个的匹配目标中能够命中的概率是 n , 当然其中的 m 可以理解成所有匹配数目中具有最大相似度的前 m 个目标, n 自然是这 m 个目标中包含正确匹配结果的比率^[27], 实验运行第 1 次和 5 次实验运行之后生成 CMC 平均曲线效果图如图 8 中的 (a) 和 (b) 图所示。

图 8(a) 图中显示的是第 1 次实验完成后生成的 CMC 性能曲线, 其中图中的横轴代表排名等级, 竖轴是识别率, 从曲线上可以看出第 1 次实验至少前 20 位匹配的效率才能达到 100%, 此处的识别率也代表: 达到最高效率所需要匹配的数量越少, 说明该算法越好。图 8(b) 图中显示的 5 次实验完成后生成的平均 CMC 性能曲线, 从曲线上可以看出 5 次实验完成后求平均值生成的 CMC 曲线图中显示, 至少前 19 位匹配的效率才能达到 100%, 当然这里达到最高效率所需要匹配数量越少, 也说明该算法越好。图中第一次实验获得了 44% 的匹配率,

5次平均实验后获得的匹配率为46%，相比现有提出的参考算法，本文算法的匹配率是比较理想的。

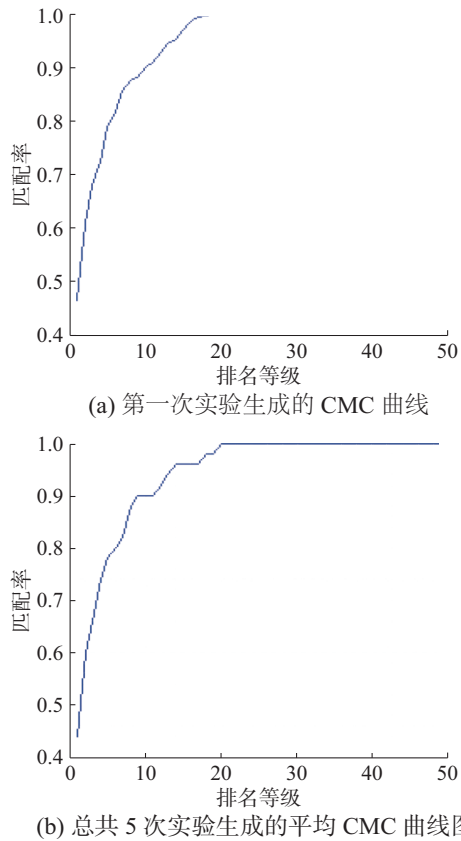


图 8 基于 VIPeR 数据集的 CMC 性能曲线

Fig. 8 CMC performance curve based on the data set of VIPeR

4.3 不同数据集实验对比分析

4.3.1 基于 VIPeR 数据库实验结果

本文在实现基于优化扩散距离算法的基础上使用多特征融合的方式进行目标再确认，实验过程中，本文为了突出多特征组合的优势，相应的做了多组实验：单一特征和多特征在同样距离度量算法的软硬件环境下进行了识别率对比，其相应实验效果如图 9(a) 所示；同时将本文算法和传统的目标再确认的其他优秀算法（夹角余弦度量算法和直方图交叉核度量算法）进行了性能方面的对比，其相应实验效果图如图 9(b) 所示。如图 9(a) 所示，通过 CMC 曲线图可以看出，使用本文中的多特征来识别时匹配率更高。从图中统计出 Rank=1, 3, 6, 8, 10 排名时所对应单一特征和多特征组合的识别率，可以从数据中观察出，本文中使用多特征组合所得识别率为 46.2%；仅使用 RGB 特征所得识别率为 28.2%；仅使用 HSV 特征所得识别率为 27.5%；仅使用 HOG 特征所得识别率为 4.5%；无论在哪个排名阶段，总体上，本文使用的多特征组合相对使用单一特征来进行识别时，多特征组合的识别率更高，更有优势。

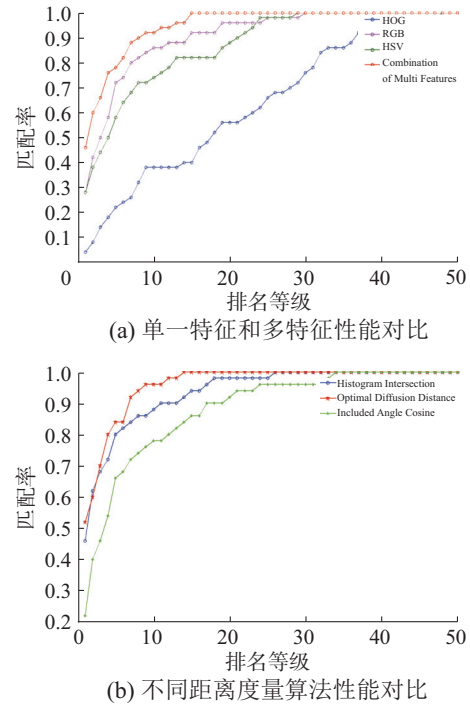


图 9 VIPeR 图库性能对比图

Fig. 9 The performance contrast figure based on the VIPeR

如图 9(b) 所示，通过 3 种算法的 CMC 曲线对比图可以很明确地看出，本文算法在相同条件下识别率更高。从图中统计出 Rank=1, 3, 6, 8, 10 排名时所对应不同算法的识别率，可以从数据中观察出，本文的识别率已经达到了 50.5%，而其他优秀算法中的直方图交叉核和夹角余弦算法的识别率分别为 47.3% 和 22.5%；无论在哪个排名阶段，总体上，优化扩散距离算法相比其他两种优秀算法在识别率方面都比较高。

为了体现本文算法在识别率上的优势，实验中，同时进行了不同优秀算法和本文算法的性能对比，对应性能表如下表 1 所示。

表 1 本文算法与其他特征类算法的性能比较

Table 1 Comparison between the proposed algorithm and other feature-class algorithms

方法	R=1	R=3	R=6	R=8	R=10	文献
SDALF	19.77	36.56	41.78	45.82	50.78	CVPR 2010 ^[16]
CPS	20.32	37.42	50.23	52.36	56.34	BMVC 2010 ^[28]
eLDFV	21.42	38.33	44.42	56.42	60.23	ECCV 2012 ^[29]
SCEAF	25.36	41.55	52.66	57.65	62.14	CVPR 2013 ^[10]
eSDC	26.77	44.24	53.38	58.66	63.56	CVPR 2013 ^[30]
HSCD	29.32	45.56	57.45	61.45	65.55	IECAS 2014 ^[31]
Diffusion Distance	50.5	71.22	84.12	94.23	96.24	本文

通过表 1 特征类算法的对比：SDALF (symmetry-driven accumulation of local features) 方法在

提取前景后分割人体目标为头部、躯干和腿部3部分的基础上进行wHSV、最大稳定颜色区域和重复纹理块组合识别,最后使用欧式距离进行测距,实验中的匹配率达到了19.77%,这个应用广泛的算法的优势在于较好特征的建立,但是在距离度量上还是使用传统的欧式距离,相比而言,本文在距离度量上的优势是显而易见的;CPS(custom pictorial structures)方法的优势在于能够根据人体结构检测出人体的各个部位进行识别,能够更好地提取前景特征,匹配率达到了20.32%;eLDFV(encoded local descriptors fisher vectors)方法结合SDALF使用Fisher向量编码图像的亮度和梯度信息,匹配率达到了21.42%,但是对于度量算法仍然需要改进;SCEFA(structural constraints enhanced feature accumulation)方法虽然使用基于Gabor纹理模式的关键点检测和双向匹配颜色区域特征的方法,匹配率达到了25.36%,其多特征的构建是最大的优点;eS-DC(effectiveness salience and dense correspondence)方法也是结合SDALF以及非监督的学习方式去寻找个体更好的特征,匹配率达到了很高的26.77%,但是欠缺考虑距离度量算法的改进;特征类中的HSCD(hybrid spatiogram and covariance descriptor)方法结合了空间直方图和区域协方差两种优秀的统计描述,并且从多个层次的统计区域提取多类互补性较好的统计向量,这个方法和同类相比是最好的,得到了29.32%的识别率。通过表2所展现的度量算法来看:ELF(ensemble of localized features)方法虽然AdaBoost分类器选取颜色和纹理特征,这种方法充分利用了分类器的选优特性,识别率仅仅为11.36%;ERSVM(Ensemble RankSVM)方法和PRDC(relative distance comparison)相类似,利用相对距离排序的SVM分类器集成多个特征向量权重,此方法在度量算法方面优势很明显,但是在特征方面仍需要改进;LMNN-R(large margin nearest neighbor with rejection)方法也使用相对距离比较的PRDC方法,使用了大边近邻分类器,在特征设计方面同样需要改进。

度量算法类中,性能表现最好的是HI(histogram intersection),直方图交叉核,该算法是一种基于隐式对应关系的内核函数,解决了无序、可变长度的矢量集合的判别分类的问题,而且是正定的,由于该算法的优势,应用到目标再确认领域,匹配率高达47.38%,因此,通过上文所述,同等条件下,度量算法类相比特征类在匹配率方面具有一定的优势。而查看表格,本文算法无论在特征类还是距离

度量类对比中,其50.5%的匹配率仍然具有一定的优势。通过以上优秀算法和本文算法在数据上对比可以看出,本文在基于VIPeR图像库的基础上,识别率方面性能相对比较好,这是因为本文特征设计上的优势和优秀度量算法引入的原因。

表2 本文算法与其他距离度量类算法的性能比较
Table 2 Comparison between the proposed algorithm and other distance metric-class algorithms

方法	R=1	R=3	R=6	R=8	R=10	文献
ELF	11.36	23.15	35.58	39.96	43.65	CVPR 2010 ^[32]
ERSVM	14.38	29.32	40.98	47.89	50.69	BMVC 2010 ^[19]
PRDC	15.65	30.24	42.36	50.26	53.68	ECCV 2012 ^[33]
LMNN-R	21.22	37.25	53.69	58.98	64.48	CVPR 2013 ^[34]
COS	22.56	47.31	67.63	74.55	77.55	CEA 2011 ^[22]
HI	47.38	67.45	82.43	86.21	87.54	CVPR 2015 ^[21]
Diffusion Distance	50.5	71.22	84.12	94.23	96.24	本文

4.3.2 基于ETHZ数据库实验结果

ETHZ数据库是从运动的摄像机中获取的图像,ETHZ1包含83个目标的4857幅图像;ETHZ2包含35个目标的1936幅图像;ETHZ3包含28个行人的1762幅图像。本文使用这类数据库作为测试样本库,用来验证本文算法和余弦、直方图正交算法的性能对比。在数据库中,光照的改变和行人的遮挡相对严重,人体姿态方面变化比较小。根据获得的目标的帧数,可以将目标再确认分为以下3类:

- 1) 单帧对单帧(single vs single, SvsS)
- 2) 多帧对单帧(multiple vs single, MvsS)
- 3) 多帧对多帧(multiple vs multiple, MvsM)

本文分别测试了SvsS、MvsS、MvsM模式下的不同算法的目标再确认的性能,相应的CMC曲线图如图10所示。

在图10中,(a)、(d)分别代表ETHZ1数据集下MvsS和MvsM的性能表现;(b)、(e)分别代表ETHZ2数据集下MvsS和MvsM的性能表现;(c)、(f)分别代表ETHZ3数据集下MvsS和MvsM的性能表现。通过测试ETHZ数据库中的实验图像可以发现,在不同ETHZ集下,本文算法在匹配精度方面,一直保持领先。ETHZ1中,在MvsS和MvsM帧下,本文算法在Rank=1时都达到了87%左右;ETHZ2中,在MvsS和MvsM帧下,本文算法在Rank=1时分别达到了87.5%和93%左右;ETHZ3中,在MvsS和MvsM帧下,本文算法在Rank=1时分别达到了

94% 和 96% 左右。相比余弦算法和直方图正交算法, 本文算法能够得到更好的表现, 其原因在于: 由于多特征的使用, 可以应对摄像机镜头的不同、遮挡所造成的差异; 更重要的是在距离匹配过程中所应用的优化扩散距离, 可以有效应对光亮、阴影等造成的差异, 并且利用直方图的空间信息更加精确

其对应距离, 从而准确的进行目标再确认。这也是本文适合更多场景且无需监督的原因, 通过在 ETHZ 数据集上的表现, 可以看出本文算法的性能还是不错的, 观察匹配出的效果图发现, 本文所提出算法在应对姿态、视角、和光照变化等方面具有较好的鲁棒性。

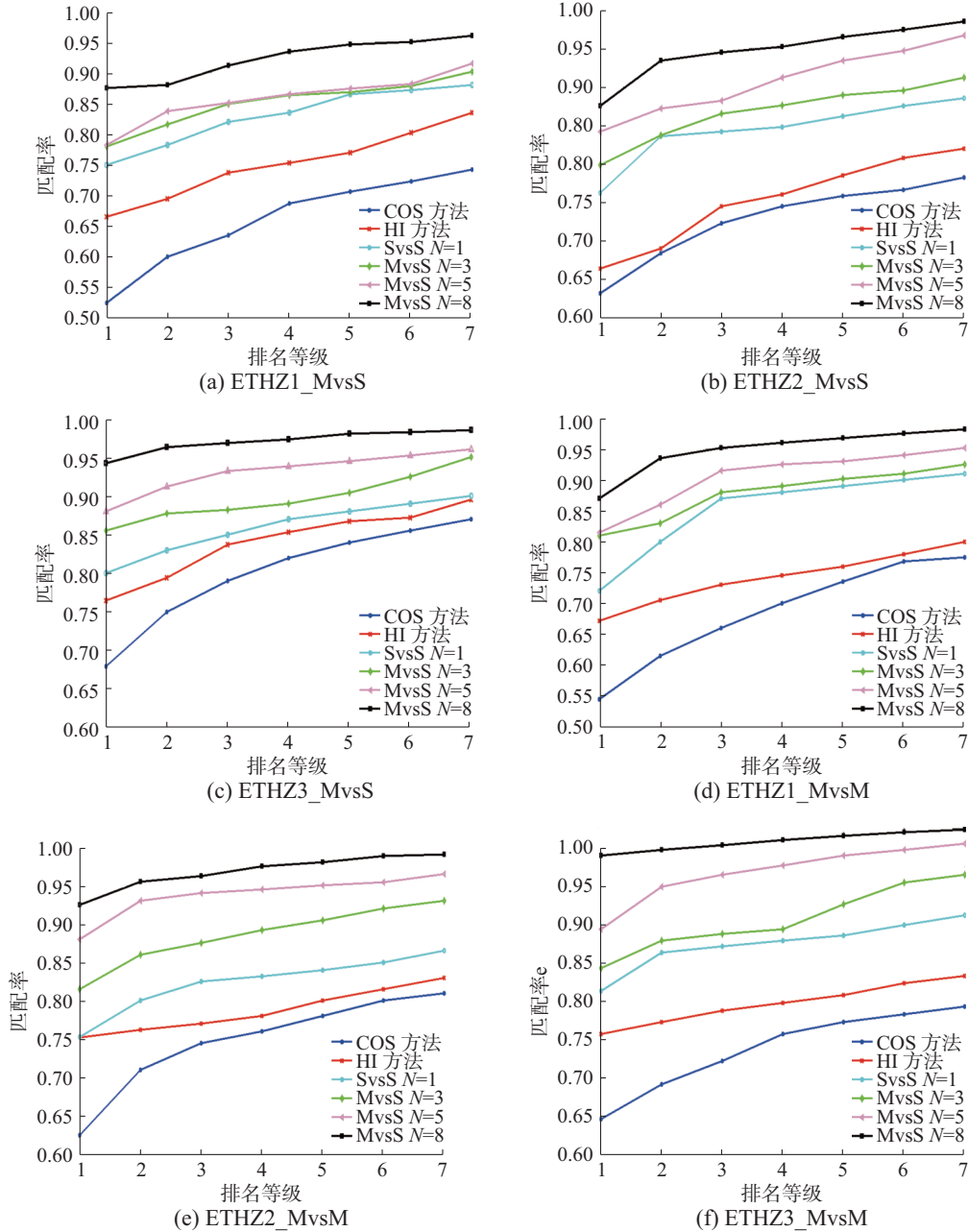


图 10 基于 ETHZ 数据集下的多帧实验

Fig. 10 The performance contrast figure based on the ETHZ

5 结束语

通过上文中的理论分析和实验验证, 本文最大的创新点在于引入物理学温度场的扩散距离的概念并进行优化, 来建立相似性度量函数。并为了提高优化扩散距离的计算效率, 进行了二次加权性优

化, 进而提高了计算的准确性。特征提取方面, 定义了两级区域直方图提取方案, 这种简易而有效的方法可以在特征基础上实现空间信息的添加, 从而提高了识别率。在两级区域方案基础上, 通过高斯加权 HSV 直方图, 整体、局部 RGB 直方图, 以及 HOG 特征描述子 3 类特征的综合使用, 并辅以适当

的加权处理,较好地实现了行人目标再确认。最后,本文优化算法与同类别的其他优秀算法以及传统目标再确认过程中所使用的有监督的训练学习型度量算法进行了对比。相比其他算法,本文的优点是:目标再确认准确率明显提高,同时不需要大量的训练样本、节省了训练时间、可以适应大部分的场景变化、节省内存开支等,在一定程度上体现了本文算法的优越性。但本文仍然还有许多需要改进的地方,比如如何在更加复杂的环境中可以更高效率的进行目标再确认,这也是本文下一步研究的重点。

参考文献:

- [1] ZHAO R, OYANG W, WANG X. Person re-identification by saliency learning[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2017, 39(2): 356–370.
- [2] JIA J, RUAN Q, JIN Y. Geometric preserving local fisher discriminant analysis for person re-identification[J]. *Neuro-computing*, 2016, 205: 92–105.
- [3] UKITA N, MORIGUCHI Y, HAGITA N. People re-identification across non-overlapping cameras using group features[J]. *Computer vision and image understanding*, 2016, 144(C): 228–236.
- [4] MARTINEL N, MICHELONI C. Re-identify people in wide area camera network[C]//2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. RI, USA, 2012: 31–36.
- [5] HUA J I, PAN M L. Person re-identification based on global color context[C]//ACCV'10 Proceedings of the 2010 international conference on Computer vision. Queenstown, New Zealand, 2010, 6468(I): 205–215.
- [6] ZHAO T, NEVATIA R. Tracking multiple humans in complex situations[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2004, 26(9): 1208–1221.
- [7] LI P, WU H, CHEN Q. Person re-identification using color enhancing feature[C]//2015 3rd IAPR Asian Conference on Pattern Recognition. Kuala Lumpur, Malaysia, 2015: 86–90.
- [8] 李金, 胡文广. 基于颜色的快速人体跟踪及遮挡处理[J]. *智能系统学报*, 2010, 5(4): 353–359.
LI Jin, HU Wenguang. Tracking fast movement using colors while accommodating occlusion[J]. *CAAI transactions on intelligent systems*, 2010, 5(4): 353–359.
- [9] 刘宇, 向高林, 王伊冰. 一种改进的行人导航算法研究[J]. *重庆邮电大学学报:自然科学版*, 2016, 28(2): 233–238.
LIU Yu, XIANG Gaolin, WANG Yibing. An improved pedestrian navigation algorithm[J]. *Journal of Chongqing university of posts and telecommunication: natural science edition*, 2016, 28(2): 233–238.
- [10] HU Y, LIAO S, LEI Z. Exploring structural information and fusing multiple features for person re-identification[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Portland, USA, 2013: 794–799.
- [11] LIU P, GUO J M, CHAMNONGTHAI K, et al. Fusion of color histogram and lbp-based features for texture image retrieval and classification[J]. *Information sciences*, 2017, 1(390): 95–111.
- [12] CHANG C, LIN C. LIBSVM: a library for support vector machines[J]. *Journal ACM transactions on intelligent systems and technology*, 2011, 2(3): 389–396.
- [13] CAI Y, CHEN W, HUANG K, TAN T. Continuously tracking objects across multiple widely separated cameras[C]//ACCV'07 Proceedings of the 8th Asian conference on Computer vision. Tokyo, Japan, 2012(I): 843–852.
- [14] PATHYA B, NAINAN S. Performance evaluation of face recognition using LBP, PCA and SVM[J]. *International journal of advanced trends in computer science and engineering*, 2016, 3(4): 85–88.
- [15] 王彩玲, 詹松, 荆晓远. 基于图像显著特征的非重叠视域行人再识别[J]. *南京邮电大学学报:自然科学版*, 2016, 36(3): 106–111.
WANG Cailing, ZHAN Song, JING Xiaoyuan. Pedestrian re-identification based on salient features in non-overlapping areas[J]. *Journal of Nanjing university of posts and telecommunications: natural science edition*, 2016, 36(3): 106–111.
- [16] FARENZENA M, BAZZANI L. Person re-identification by symmetry-driven accumulation of local features[C]//2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA, 2010: 2360–2367.
- [17] 彭程, 刘帅师, 万川, 等. 基于局部纹理 ASM 模型的人脸表情识别[J]. *智能系统学报*, 2011, 6(3): 231–238.
PENG Cheng, LIU Shuaishi, WAN Chuan, et al. An active shape model for facial expression recognition based on a local texture model[J]. *CAAI transactions on intelligent systems*, 2011, 6(3): 231–238.
- [18] FENG X, PIETIKAINEN M, HADID A. Facial expression recognition with local binary patterns and linear programming[J]. *Pattern recognition and image analysis*, 2005, 15(2): 550–552.
- [19] PROSSER B, ZHENG W S, GONG S. Person re-identification by support vector ranking[C]//BMVC2010 British Machine Vision Conference. Aberystwyth, UK, 2010(42): 1–11.
- [20] FANG C. People re-identification based on online multiple kernel learning in video surveillance[J]. *Opto-electronic*

- engineering, 2012, 39(9): 65–71.
- [21] PANG Y, XIN-CHU S. Multiway histogram intersection for multi-target tracking[C]//2015 18th International Conference on Information Fusion. Washington, DC, USA, 2015: 1938–1945.
- [22] YIN J, ZHOU J, JIN Z. Principal component analysis and kernel principal component analysis based on cosine angle distance[J]. Computer engineering and applications, 2011, 47(3): 9–12.
- [23] HAIBIN L, OKADA K. Diffusion distance for histogram comparison[C]//Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC, USA, 2006, 1(1): 246–253.
- [24] ROTHER C, KOLMOGOROV V, BLAKE A. “GrabCut”: interactive foreground extraction using iterated graph cuts[J]. ACM transactions on graphics, 2004, 23(3): 307–312.
- [25] QING-JUN W, RU-BO Z. LPP-HOG: a new local image descriptor for fast human detection[C]//2008 IEEE International Symposium on Knowledge Acquisition and Modeling Workshop. Wuhan, China, 2008: 640–643.
- [26] JUN Y. HOCK-SOON S. fuzzy diffusion distance learning for cartoon similarity estimation[J]. Journal of computer science and technology: English edition, 2011, 26(2): 203–216.
- [27] MAZZON R, TAHIR S F, CAVALLARO A. Person re-identification in crowd[J]. Pattern recognition letters, 2012, 33(14): 1828–1837.
- [28] CHENG D S, CRISTANI M, STOPPA M. Custom pictorial structures for re-identification[C]//BMVC2010 British Machine Vision Conference. Dundee, UK, 2011, (68): 1–11.
- [29] MA B, SU Y, JURIE F. Local descriptors encoded by fisher vectors for person re-identification[C]//12th European Conference on Computer Vision Workshops. Florence, Italy, 2012, 7583: 413–422.
- [30] ZHAO R, OUYANG W, WANG X. Unsupervised saliency learning for person re-identification[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA, 2013, 9(4): 3586–3593.
- [31] ZENG M Y, WU Z M, CHANG T. Fusing appearance statistical features for person re-identification[J]. Journal of electronics and information technology, 2014, 36(8): 1844–1851.
- [32] GRAY D, TAO H. Viewpoint invariant pedestrian recognition with an ensemble of localized features[C]//ECCV'08 Proceedings of the 10th European Conference on Computer Vision. Marseille, France, 2008, (I): 262–275.
- [33] ZHENG W S, GONG S, XIANG T. Reidentification by relative distance comparison[J]. IEEE transactions on pattern analysis and machine intelligence, 2013, 35(3): 653–668.
- [34] DIKMEN M, AKBAS E, HUANG T S. Pedestrian recognition with a learned metric[C]//ACCV'10 Proceedings of the 10th Asian conference on Computer vision. Queenstown, New Zealand, 2010, 6495(IV): 501–512.

作者简介:



曹伟,男,1989年生,硕士研究生,主要研究方向为多目标跟踪、目标传递。



韩华,女,1983年生,副教授,主要研究方向为多目标跟踪、目标再确认、目标传递、智能计算。



王裕明,男,1962年生,教授,主要研究方向为计算机信息技术研究。