计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2019.20180740
56(11): 2424-2437, 2019

基于跨视角判别词典嵌入的行人再识别

陆 萍[1,2] 董虎胜[1] 钟 珊[3] 龚声蓉[3]
[1](苏州经贸学院信息技术学院 江苏苏州 215009)
[2](浙江大学计算机科学与技术学院 杭州 310027)
[3](常熟理工学院 江苏常熟 215500)
(plu2015@QQ.com)

Person Re-identification by Cross-View Discriminative Dictionary Learning with Metric Embedding

Lu Ping[1,2], Dong Husheng[1], Zhong Shan[3], and Gong Shengrong[3]
[1](*School of Information Technology, Suzhou Institute of Trade and Commerce, Suzhou, Jiangsu* 215009)
[2](*School of Computer Science and Technology, Zhejiang University, Hangzhou* 310027)
[3](*Changshu Institute of Technology, Changshu, Jiangsu* 215500)

Abstract The task of person re-identification is to associate individuals who have been observed over disjoint camera views. Due to its value in applications of video surveillance, person re-identification has drawn great attention from computer vision and machine learning communities. To address this problem, current literature mainly focuses on extracting discriminative features or learning distance metrics from pedestrian images. However, the representation power of learned features or metrics might be limited, because a person's appearance usually undergoes large variations in different camera views, and many passers-by may take similar visual appearances in public spaces. In order to overcome these challenges and improve the person re-identification accuracies, we propose an effective re-identification method called cross-view discriminative dictionary learning with metric embedding. Different from traditional dictionary learning or metric learning approaches, the cross-view dictionary and distance metric are jointly learned in our model, thus their strengths can be combined. The proposed model not only captures the intrinsic relationships of representation coefficients, but also explores the distance constraints in different camera views. As a result, the re-identification can be performed with much more powerful representations in a discriminative subspace. To address the bias brought by unbalanced training samples in the metric learning phase, an automatic weighting strategy of training pairs is introduced. We devise an efficient optimization algorithm to solve the proposed model, in which the representation coefficients, dictionary, and metric are optimized alternately. Experimental results on three public benchmark datasets including VIPeR, GRID, and 3DPeS, show that the proposed method achieves remarkable performance compared with existing approaches as well as published results.

收稿日期:2018-11-01;**修回日期**:2019-05-08
基金项目:国家自然科学基金项目(61170124,61272258,61702055);江苏省自然科学基金项目(BK20151260);江苏省高等院校国内高级访问学者计划项目(2018GRFX052);江苏省高校青蓝工程骨干教师培养对象(2019年)
This work was supported by the National Natural Science Foundation of China (61170124, 61272258, 61702055), the Natural Science Foundation of Jiangsu Province of China (BK20151260), the Senior Visiting Scholars Program of Jiangsu Province (2018GRFX052), and the Backbone Teachers of Qinglan Project of Jiangsu Province (2019).
通信作者:董虎胜(hsdong2012@gmail.com)

Key words　person re-identification; feature representation; dictionary learning; distance metric; weight assignment

摘　要　行人再识别是指在具有不重叠视域的摄像机监控网络中根据行人外观进行身份关联的任务.由于在视频监控系统中具有广泛的应用前景,受到了计算机视觉与机器学习领域的广泛关注.当前的行人再识别研究主要关注从行人图像中提取判别性的特征描述子或学习距离度量.然而不同摄像机视角下行人的外观常常存在很大差异,同一摄像机下还会有行人外观相近的情况,这使得特征描述子或距离度量的表达能力受到了很大的影响.为了增强它们的表达能力并提升行人再识别的准确率,提出了一种基于跨视角判别性词典嵌入的行人再识别算法.在该算法中不仅学习了跨视角的词典还同时联合学习了一个距离度量矩阵,从而将两者的优势结合起来.该算法模型有效地挖掘了不同视角下词典表达的内在联系与距离约束,从而能够使用学习到的表达能力更强的特征在嵌入子空间中进行行人再识别.为了避免不均衡训练样本带来的度量矩阵偏差问题,在度量矩阵的学习中还引入了自适应的权重分配策略.在模型优化上,采用了高效的交替优化方法来求解词典与距离度量等模型参数.在 VIPeR,GRID,3DPeS 等数据集上的实验结果表明本文算法取得了非常优秀的行人再识别性能.

关键词　行人再识别;特征表达;词典学习;距离度量;权重分配

中图法分类号　TP394

在具有不重叠视域的摄像机监控网络中,根据行人表观信息进行跨摄像机身份关联的工作也被称为行人再识别[1],它是实现对特定目标的检索[2]、持续跟踪[3]和行为分析等智能视频监控应用的一项关键技术.由于受到光照、视角、姿态与遮挡等因素的影响,同一行人在不同摄像机拍摄的画面中可能会呈现出很大的外观差异,这给行人再识别带来了相当大的困难.由于在智能视频监控中具有广阔的应用前景,行人再识别引起了计算机视觉与机器学习领域广泛的关注并开展了大量的研究[4-6].

目前对行人再识别的研究可分为传统方法与基于深度学习的方法两大类.其中深度学习方法需要有大量标注的训练数据,因此在大型数据集上通常能够取得比较优秀的性能[7-8].但在较小的数据集上,深度学习模型极易发生过拟合问题,在性能上仍弱于传统的方法.本文工作主要关注小数据集上的行人再识别问题,且归属于传统方法类别.应用传统方法的行人再识别工作主要从特征描述子设计与度量学习算法两个方面来开展.

为了从行人图像中获取具有判别性的表观信息,研究人员设计了一系列用于行人图像匹配的特征描述子,如局部最大出现特征[6](local maximal occurrence, LOMO)、显著颜色名称[9](salient color names, SCN)、条状加权直方图[10](weighted histograms of overlapping stripes, WHOS)等,它们有力地促进了行人再识别研究的进展.但是由于不同摄像机下行人外观常常会存在很大的差异,同一摄像机下还会有行人外观相近的情况,以及特征描述子在语意上的模糊性等原因,使得特征描述子的表达能力受到了一定的限制.

直接在原始特征表达空间中进行行人再识别的准确率通常都比较低,通过学习度量矩阵将它们投影到更具判别性的子空间中通常能够带来比较显著的性能提升[11].度量学习旨在从训练数据中学习到某一特定的投影空间,使得具有相同标签的行人图像在该嵌入子空间中距离被收缩,而具有不同标签的图像之间的距离被拉大[12-13].尽管度量学习方法能够获得更为优秀的匹配效果,它们仍然会受到特征表达能力的影响.

针对行人外观描述子与距离度量表达能力受限的问题,本文提出了一种基于跨视角判别词典嵌入(cross-view discriminative dictionary learning with metric embedding, CDDM)的行人再识别匹配模型.在该模型中通过学习跨视角的判别词典将原始特征表达为过完备基(over-complete basis)的组合系数向量,从而获得比原始特征描述子更为鲁棒的表达.但与文献[14-15]等仅学习词典表达的方法不同,本文方法还利用了训练样本及标签中蕴含的距离约束信息,在学习判别词典的同时联合学习了一个度量矩阵来进行子空间嵌入,这样就可以在更具判别性的子空间中进行行人相似度的匹配.针对不同摄像机下行人图像正负样本对数量严重不均衡引起的度量偏差问题,本文还设计了样本对自适应权重分配策略.在 VIPeR,GRID,3DPeS 数据集上的实验结果验证了本文算法的有效性.

1 相关工作

在行人再识别的研究工作中,特征设计受到关注相对较早.为了抑制各种引起行人外观变化的因素,在行人再识别特征描述子的设计中大多使用了颜色、纹理与形状等信息.在 Liao 等人[6]设计的 LOMO 描述子中,从滑动窗口中提取了联合 HSV 直方图和尺度不变局部三值模式(scale invariant local ternary pattern, SILTP),并运用最大池化(max pooling)操作来增强描述子的抗视角变化能力.Matsukawa 等人[16]使用层次化的高斯模型来表达图像的颜色信息,设计了高斯化高斯(Gaussian of Gaussian, GOG)描述子.Yang 等人[9]从像素概率分布的角度提出了显著颜色名称 SCN 特征.Zhao 等人[17]通过学习最具有判别性的中层滤波器特征来表达行人图像外观.Ma 等人[18]设计了使用协方差描述的生物启发特征(bio-inspired features, BIF).

在获得行人图像的特征描述子之后,度量学习能够利用训练数据的标签信息,根据特定的距离约束来学习获得更有效的距离计算模型,取得更高的行人再识别准确率.Mignon 等人[19]设计了成对约束元件分析(pairwise constrained component analysis, PCCA)算法从高维样本中学习投影子空间;Liao 等人[20]提出了对训练样本采用不对称加权策略的度量学习方法.Zheng 等人[21]提出了概率相对距离比较模型(probabilistic relative distance comparison, PRDC),You 等人[22]在引入更严格的最近负样本约束后设计了"顶推"(top push)学习模型.利用贝叶斯准则,Köestinger 等人[23]提出了具有闭合形式解的简单直接度量(keep it simple and straightforward metric, KISSME)学习方法.Liao 等人[6]对 KISSME 加以改进后提出了联合学习度量矩阵与投影子空间的跨视角二次判别分析(cross-view quadratic discriminant analysis, XQDA)方法.

从训练数据中学习判别性词典能够将原始特征表达为更鲁棒的组合系数向量,实现对原始特征的变换[24].在文献[25]中,Liu 等人通过学习跨视角的半监督耦合词典来匹配行人图像.Prates 等人[26]通过学习核化的跨视角词典,使用协同表达向量来对行人图像进行匹配.Zhang 等人[27]为每个行人学习了支持向量机(support vector machine, SVM)的判别向量,并进一步创建最小二乘半耦合词典.Srikrishna 等人[14]通过对相互关联的稀疏编码施加判别约束来解决行人图像因视角变化引起的差异.Kodirov 等人[28]通过引入 L_1 范数的拉普拉斯图正则项来进行无监督的行人再识别.

与上述工作不同,本文方法采用了联合学习度量矩阵与判别词典的策略.在学习模型中充分挖掘了不同视角下词典表达的内在联系与距离约束,把度量学习与词典学习的优势结合起来进行行人再识别.

2 跨视角判别词典嵌入

2.1 词典学习

设 $\boldsymbol{X}\in\mathbb{R}^{d\times n}$ 为含有 n 个训练样本的特征矩阵,其每列 $\boldsymbol{x}^i\in\mathbb{R}^d$ 为从第 i 张图像中提取的外观特征.词典学习的主要目的是从训练数据中学习获得一个由过完备基组成的词典 $\boldsymbol{D}=(\boldsymbol{d}^1,\boldsymbol{d}^2,\cdots,\boldsymbol{d}^m)\in\mathbb{R}^{d\times m}$,使用该词典能够将原始 d 维特征空间中的样本投影到由 $\boldsymbol{D}$ 的各列张成的一个 m 维子空间 $\mathcal{S}=span\{\boldsymbol{d}^1,\boldsymbol{d}^2,\cdots,\boldsymbol{d}^m\}$ 中.这样就能够以比较低的维度表达出原始信息,使得学习任务简化,模型的复杂度得以降低.学习词典 $\boldsymbol{D}$ 的模型表示为

$$\begin{aligned}&\arg\min_{\boldsymbol{D},\boldsymbol{Z}}\|\boldsymbol{X}-\boldsymbol{D}\boldsymbol{Z}\|_{\mathrm{F}}^2+\lambda_1\Omega(\boldsymbol{Z}),\\&\text{s.t. }\|\boldsymbol{d}^i\|_2^2\leqslant 1,\forall i,\end{aligned}\tag{1}$$

其中,$\|\cdot\|_{\mathrm{F}}$ 为矩阵的 Frobenius 范数;$\boldsymbol{Z}=(\boldsymbol{z}^1,\boldsymbol{z}^2,\cdots,\boldsymbol{z}^n)\in\mathbb{R}^{m\times n}$ 被称为系数矩阵,也就是各训练样本在新的维空间中的表达;$\boldsymbol{d}^i$ 指代词典 $\boldsymbol{D}$ 的第 i 列,式中对它们施加单位长度约束旨在使获得的词典具有更好的紧凑性;$\Omega(\boldsymbol{Z})$ 为对 $\boldsymbol{Z}$ 的正则项,常用的正则函数有 $\Omega(\boldsymbol{Z})=\sum_{i=1}^{n}\|\boldsymbol{z}^i\|_1$ 或 $\Omega(\boldsymbol{Z})=\|\boldsymbol{Z}\|_{\mathrm{F}}^2$,前者能够获得稀疏的表达向量但运算代价相对较高,后者求解较为容易但不具有稀疏性.由于行人再识别使用的特征描述子维度远高于训练样本数,使用稀疏表达难以捕捉到具有巨大差异的跨摄像机特征向量数据的内在相关性,因此在本文中选用了 $\Omega(\boldsymbol{Z})=\|\boldsymbol{Z}\|_{\mathrm{F}}^2$ 正则项.由于 $\boldsymbol{Z}$ 为 $\boldsymbol{X}$ 在新特征空间中的投影,因此可以使用学习到的词典 $\boldsymbol{D}$ 实现对 $\boldsymbol{X}$ 的重建,也就是 $\boldsymbol{X}\approx\boldsymbol{D}\boldsymbol{Z}$.

2.2 跨视角判别词典嵌入模型

在行人再识别中,需要对不同摄像机下捕捉到的行人图像进行相似度匹配.但采用式(1)学习到的词典无法捕捉不同视角下数据的内在结构,针对该问题,在本文方法中为每个摄像机视角分别学习了词典表达.设 $\boldsymbol{X}_{\mathrm{p}}\in\mathbb{R}^{d\times n}$ 与 $\boldsymbol{X}_{\mathrm{g}}\in\mathbb{R}^{d\times n}$ 分别为训练集

中检测集(probe set)与匹配集(gallery set)的特征矩阵;$\boldsymbol{Y}\in\mathbb{R}^{n\times n}$为它们之间的匹配标签矩阵;$\boldsymbol{D}\in\mathbb{R}^{d\times m}$为对应的判别词典;可以建立的跨视角判别词典学习模型为

$$\begin{aligned}\arg\min_{\boldsymbol{D},\boldsymbol{Z}_{\mathrm{p}},\boldsymbol{Z}_{\mathrm{g}}}&\|\boldsymbol{X}_{\mathrm{p}}-\boldsymbol{D}\boldsymbol{Z}_{\mathrm{p}}\|_{\mathrm{F}}^{2}+\|\boldsymbol{X}_{\mathrm{g}}-\boldsymbol{D}\boldsymbol{Z}_{\mathrm{g}}\|_{\mathrm{F}}^{2}+\\&\lambda_1\|\boldsymbol{Z}_{\mathrm{p}}\|_2^2+\lambda_1\|\boldsymbol{Z}_{\mathrm{g}}\|_2^2,\\&\text{s.t.}\ \ \|\boldsymbol{d}^i\|_2^2\leqslant 1,\forall i.\end{aligned}\tag{2}$$

其中,λ_1为调节系数;$\boldsymbol{Z}_{\mathrm{p}}\in\mathbb{R}^{m\times n}$和$\boldsymbol{Z}_{\mathrm{g}}\in\mathbb{R}^{m\times n}$分别指代$\boldsymbol{X}_{\mathrm{p}}$与$\boldsymbol{X}_{\mathrm{g}}$在使用词典$\boldsymbol{D}$表达时的组合系数向量,也就是变换后的特征表达.式(2)的前2项表达了学习词典对原始特征数据的重建误差,后2项为正则项,用来抑制模型的过拟合风险.

尽管式(2)能够描述跨视角行人图像数据的内在结构,但是对训练数据与标签中蕴含的距离约束信息却未能有效利用.在行人再识别中,我们希望不同摄像机视角下正确匹配图像(正样本对)之间距离应尽可能的小,而错误匹配图像(负样本对)间的距离要尽可能的大,从而在正、负样本之间建立起一个距离间隔.这样就可以在给定某一检索图像后,达到将正确匹配图像从所有待匹配图像中识别出来的目标.为此,本文引入的约束损失函数为

$$\psi_{\boldsymbol{M}}(\boldsymbol{z}_{\mathrm{p}}^i,\boldsymbol{z}_{\mathrm{g}}^j)=[y_{ij}(d_{\boldsymbol{M}}(\boldsymbol{z}_{\mathrm{p}}^i,\boldsymbol{z}_{\mathrm{g}}^j)-\mu)]_+,\tag{3}$$

其中,$[\cdot]_+$为铰链损失(hinge loss)函数,即$[x]_+=\max(0,x)$;$\boldsymbol{z}_{\mathrm{p}}^i$与$\boldsymbol{z}_{\mathrm{g}}^j$分别指代$\boldsymbol{Z}_{\mathrm{p}}$和$\boldsymbol{Z}_{\mathrm{g}}$的第$i,j$列;$y_{ij}$取自匹配标签矩阵$\boldsymbol{Y}$,若$\boldsymbol{z}_{\mathrm{p}}^i$与$\boldsymbol{z}_{\mathrm{g}}^j$正确匹配则$y_{ij}=1$,否则$y_{ij}=-1$;$\mu$为一个正的常数,用作为判断阈值;$d_{\boldsymbol{M}}(\boldsymbol{z}_{\mathrm{p}}^i,\boldsymbol{z}_{\mathrm{g}}^j)$为标准的马氏距离函数,定义为

$$d_{\boldsymbol{M}}(\boldsymbol{z}_{\mathrm{p}}^i,\boldsymbol{z}_{\mathrm{g}}^j)=(\boldsymbol{z}_{\mathrm{p}}^i-\boldsymbol{z}_{\mathrm{g}}^j)^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{z}_{\mathrm{p}}^i-\boldsymbol{z}_{\mathrm{g}}^j),\tag{4}$$

其中$\boldsymbol{M}$为待求解的距离度量矩阵,其半正定性($\boldsymbol{M}\geq 0$)保证了$d_{\boldsymbol{M}}$能够满足距离所需的三角不等式与非负性.对$\boldsymbol{M}$可进一步作Cholesky分解得$\boldsymbol{M}=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}$,因此式(3)等价于:

$$\psi_{\boldsymbol{W}}(\boldsymbol{z}_{\mathrm{p}}^i,\boldsymbol{z}_{\mathrm{g}}^j)=[y_{ij}(\|\boldsymbol{W}(\boldsymbol{z}_{\mathrm{p}}^i-\boldsymbol{z}_{\mathrm{g}}^j)\|_2^2-\mu)]_+.\tag{5}$$

在行人再识别中,由于不同摄像机下错误匹配行人图像的数量远多于正确匹配图像,这会使得学习到的度量矩阵倾向于将所有行人图像对判定为错误匹配,引起度量偏差问题[20].为了解决该问题,可以采用从训练样本邻域学习度量矩阵的方案[29],通过减少容易识别的负样本对在模型中的贡献度来抑制数据不平衡问题.由此可以把整个训练集上的损失函数表达为

$$\mathcal{L}_{\boldsymbol{W}}=\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_{ij}\psi_{\boldsymbol{W}}(\boldsymbol{z}_{\mathrm{p}}^i,\boldsymbol{z}_{\mathrm{g}}^j),\tag{6}$$

其中β_{ij}为样本对$(\boldsymbol{z}_{\mathrm{p}}^i,\boldsymbol{z}_{\mathrm{g}}^j)$的贡献权重,通过设置$\beta_{ij}$的取值即可实现从样本邻域学习的目标.

为了合理地分配β_{ij}的权重值,对每个变换后的特征表达$\boldsymbol{z}_{\mathrm{p}}^i$,首先计算新特征空间中待匹配集$\{\boldsymbol{z}_{\mathrm{g}}^j\}_{j=1}^n$内所有样本与其之间的距离,然后将$\{\boldsymbol{z}_{\mathrm{g}}^j\}_{j=1}^n$划分为3个组:

$$\begin{aligned}\mathcal{X}_{\mathrm{h}}&=\{\boldsymbol{z}_{\mathrm{g}}^j\,|\,\mathcal{R}_i(\boldsymbol{z}_{\mathrm{g}}^j)<\mathcal{R}_i(\boldsymbol{z}_{\mathrm{g}}^+)\},\\\mathcal{X}_{\mathrm{m}}&=\{\boldsymbol{z}_{\mathrm{g}}^j\,|\,\mathcal{R}_i(\boldsymbol{z}_{\mathrm{g}}^+)<\mathcal{R}_i(\boldsymbol{z}_{\mathrm{g}}^j)\leqslant n/2\},\\\mathcal{X}_{\mathrm{e}}&=\{\boldsymbol{z}_{\mathrm{g}}^j\,|\,\mathcal{R}_i(\boldsymbol{z}_{\mathrm{g}}^+)>n/2\},\end{aligned}\tag{7}$$

其中:$\mathcal{R}_i(\boldsymbol{z}_{\mathrm{g}}^j)$指的是$\boldsymbol{z}_{\mathrm{p}}^i$的排序列表中$\boldsymbol{z}_{\mathrm{g}}^j$的排序位置(rank);$\mathcal{R}_i(\boldsymbol{z}_{\mathrm{g}}^+)$指代与$\boldsymbol{z}_{\mathrm{p}}^i$正确匹配的图像$\boldsymbol{z}_{\mathrm{g}}^+$所在的位置;$\mathcal{X}_{\mathrm{h}},\mathcal{X}_{\mathrm{m}},\mathcal{X}_{\mathrm{e}}$分别指代$\boldsymbol{z}_{\mathrm{p}}^i$的困难匹配集(hard set)、中等匹配集(medium set)与容易匹配集(easy set).

图1给出了$\mathcal{X}_{\mathrm{h}},\mathcal{X}_{\mathrm{m}},\mathcal{X}_{\mathrm{e}}$的划分图示,由图1可知$\mathcal{X}_{\mathrm{h}}$包含了那些排在正确匹配图像$\boldsymbol{z}_{\mathrm{g}}^+$之前的负样本,它们通常与检索图像具有接近的外观,但却具有不同的标签,因此是训练时需要着重处理的对象.

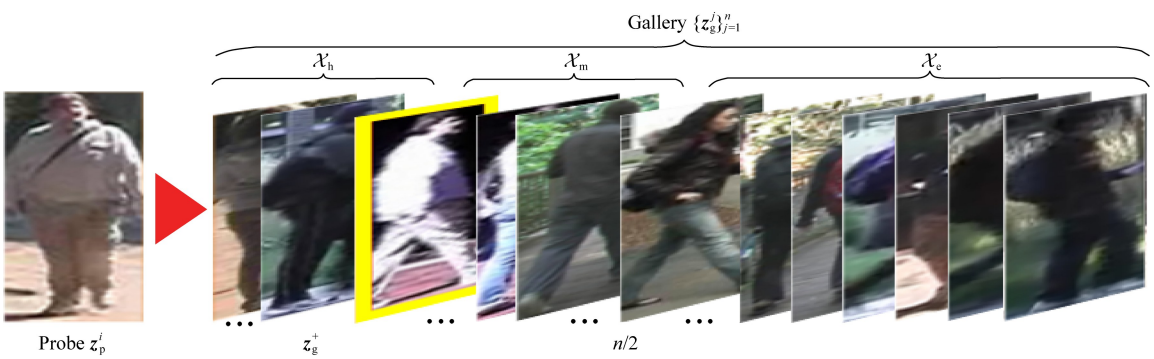


Fig. 1　Partition of the hard/medium/easy sets

图1　困难/中等/容易匹配集划分示意

中等匹配集 $\mathcal{X}_m$ 指代在前 1/2 排序列表中排在 z_g^+ 之后的那些负样本，它们对模型训练的贡献一般，只需要赋予比较低的权重即可.容易匹配集 $\mathcal{X}_e$ 包含的都是非常容易区分的错误匹配图像，对训练模型没有帮助，权重可以置 0.在根据式(7)划分 $\mathcal{X}_h$，$\mathcal{X}_m$，$\mathcal{X}_e$ 时，若 z_g^+ 位于后 1/2 排序列表时，可以忽略 $\mathcal{X}_m$，将 z_g^+ 前的样本划分为 $\mathcal{X}_h$，而后面的样本划分为 $\mathcal{X}_e$.

根据分析，本文采用的训练样本对自适应加权方案为：若 $y_{ij}=1$ 即为正确匹配时，取 $\beta_{ij}=1/N^+$，这里 N^+ 为训练集中正样本对的数量；若 $y_{ij}=-1$，β_{ij} 取值为

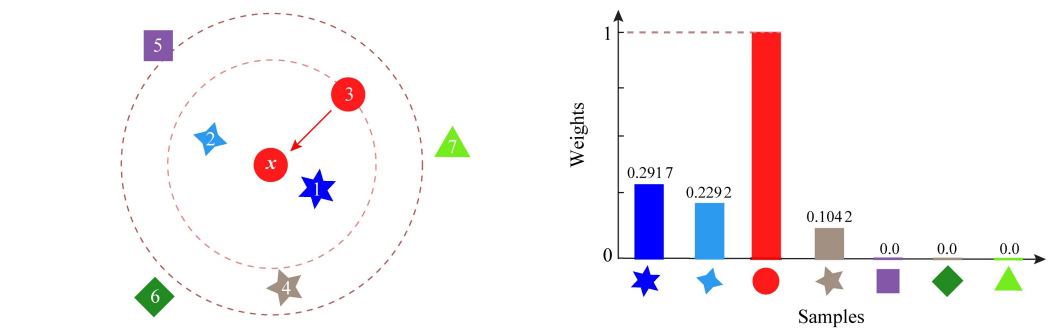
$$\beta_{ij}=\begin{cases}\dfrac{1}{N^-}+\dfrac{3}{4N^-}\left(\dfrac{\mathcal{R}_i(z_g^j)-\mathcal{R}_i(z_g^+)}{1-\mathcal{R}_i(z_g^+)}\right), & z_g^j\in\mathcal{X}_h,\\ \dfrac{1}{N^-}+\dfrac{3}{4N^-}\left(\dfrac{\mathcal{R}_i(z_g^j)-\mathcal{R}_i(z_g^+)}{1-\mathcal{R}_i(z_g^+)}\right), & z_g^j\in\mathcal{X}_m,\\ 0, & z_g^j\in\mathcal{X}_e,\end{cases} \tag{8}$$

其中 N^- 为训练集中负样本对的数量.从式(8)可知对于每个检索图像 z_p^i，与其接近且难以区分的负样本会被赋予较大的权重，而那些容易被识别出的负样本将会被赋予较低的权重或直接丢弃.图 2 给出了按式(7)与式(8)进行自适应权值分配的示例，图 2 中“x”指代检索样本，“3”为正确匹配样本，其余为错误匹配样本，数字标记了它们的排序次序.从图 2 中可知错误匹配样本的权值得到了有效的抑制.在将此方案应用于行人再识别中的距离度量学习时，正负样本对的贡献度能够基本上得到均衡，从而避免度量偏差问题.

根据式(2)与式(6)，可以将本文提出的联合学习跨视角判别词典与度量嵌入的模型表达为

$$\begin{aligned}\arg\min_{\boldsymbol{D},\boldsymbol{Z}_p,\boldsymbol{Z}_g,\boldsymbol{W}}&\|\boldsymbol{X}_p-\boldsymbol{D}\boldsymbol{Z}_p\|_F^2+\|\boldsymbol{X}_g-\boldsymbol{D}\boldsymbol{Z}_g\|_F^2+\\&\lambda_0\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_{ij}\psi_{\boldsymbol{W}}(z_p^i,z_g^j)+\lambda_1\|\boldsymbol{Z}_g\|_F^2+\\&\lambda_1\|\boldsymbol{Z}_p\|_F^2+\lambda_2\|\boldsymbol{W}\|_F^2,\\&\text{s.t.}\ \ \|\boldsymbol{d}^i\|_2^2\leqslant 1,\ \forall\, i.\end{aligned} \tag{9}$$

其中 $\boldsymbol{Z}_p=(z_p^1,\cdots,z_p^i,\cdots,z_p^n)$，$\boldsymbol{Z}_g=(z_g^1,\cdots,z_g^j,\cdots,z_g^n)$ 即 z_p^i 与 z_g^j 分别为 $\boldsymbol{Z}_p$ 与 $\boldsymbol{Z}_g$ 的第 i，j 列.



(a) Distribution of samples in 2-dimensional space and their ranks (b) The weights of each example

Fig. 2 Illustration of the adaptive weight assignment in 2-dimensional space

图 2 二维空间中样本自适应权重分配示例

2.3 模型优化

在式(9)所示的模型中需要同时优化 $\boldsymbol{D}$，$\boldsymbol{Z}_p$，$\boldsymbol{Z}_g$，$\boldsymbol{W}$ 这 4 个相互耦合的参数，模型并非关于所有参数联合凸，因此无法对它们同时进行优化.但该模型中各项均为二次项或 max 函数，在固定其他参数仅优化某一变量时为凸模型，故本文采用交替优化的方法来求解各模型参数.

1) 更新 $\boldsymbol{Z}_p$

在固定 $\boldsymbol{D}$，$\boldsymbol{Z}_g$，$\boldsymbol{W}$，仅对 $\boldsymbol{Z}_p$ 进行优化时，由于式(9)需要根据 z_p^i 与 z_g^j 来计算距离约束损失函数，所以这里采用逐列求取 z_p^i 的方法来优化 $\boldsymbol{Z}_p$.此时，仅优化 z_p^i 的目标函数为

$$\begin{aligned}z_p^*=\arg\min_{z_p}&\|x_p-\boldsymbol{D}z_p\|_2^2+\lambda_1\|z_p\|_2^2+\\&\lambda_0\sum_{j=1}^{n}\beta_{ij}\psi_{\boldsymbol{W}}(z_p,z_g^j).\end{aligned} \tag{10}$$

为简化表达，式(10)中略去了 x_p^i 与 z_p^i 的上标 i.对式(10)求解可获得 z_p 闭合形式的解表达式为

$$\begin{aligned}z_p^*=&(\boldsymbol{D}^T\boldsymbol{D}+\lambda_1\boldsymbol{I}+\lambda_0\boldsymbol{W}^T\boldsymbol{W}\sum_{j=1}^{n}\beta_{ij}\delta_{ij})^{-1}\times\\&(\boldsymbol{D}^Tx_p+\lambda_0\boldsymbol{W}^T\boldsymbol{W}\sum_{j=1}^{n}\beta_{ij}\delta_{ij}z_g^j),\end{aligned} \tag{11}$$

其中，$\boldsymbol{I}$ 为单位矩阵；$\delta_{ij}=\mathbb{I}(\psi_{\boldsymbol{W}}(z_p,z_g^j))$ 为示性函数，若 $\psi_{\boldsymbol{W}}(z_p,z_g^j)>0$ 则定义 $\delta_{ij}=y_{ij}$，否则 $\delta_{ij}=0$.

2) 更新 $\boldsymbol{Z}_g$

与优化 $\boldsymbol{Z}_p$ 类似，在对式(9)固定 $\boldsymbol{D}$，$\boldsymbol{Z}_p$，$\boldsymbol{W}$，对

$\boldsymbol{Z}_{\mathrm{g}}$ 进行优化时也需要采取逐列优化 $\boldsymbol{z}_{\mathrm{g}}$ 的方式,最终可以获得的解表达式为

$$\boldsymbol{z}_{\mathrm{g}}^{*}=(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}+\lambda_1\boldsymbol{I}+\lambda_0\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\sum_{i=1}^{n}\beta_{ij}\delta_{ij})^{-1}\times(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{x}_{\mathrm{g}}+\lambda_0\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\sum_{i=1}^{n}\beta_{ij}\delta_{ij}\boldsymbol{z}_{\mathrm{p}}^{i}),\tag{12}$$

其中,$\delta_{ij}=\mathbb{I}(\psi_{\boldsymbol{W}}(\boldsymbol{z}_{\mathrm{p}}^{i},\boldsymbol{z}_{\mathrm{g}}))$.

3) 更新 $\boldsymbol{D}$

在固定 $\boldsymbol{Z}_{\mathrm{p}},\boldsymbol{Z}_{\mathrm{g}},\boldsymbol{W}$ 对式(9)仅考虑 $\boldsymbol{D}$ 的优化时,等价于二次规划问题:

$$\arg\min_{\boldsymbol{D}}\|\boldsymbol{X}_{\mathrm{p}}-\boldsymbol{D}\boldsymbol{Z}_{\mathrm{p}}\|_{\mathrm{F}}^{2}+\|\boldsymbol{X}_{\mathrm{g}}-\boldsymbol{D}\boldsymbol{Z}_{\mathrm{g}}\|_{\mathrm{F}}^{2}\quad\text{s.t. }\|\boldsymbol{d}^{i}\|_{2}^{2}\leqslant 1,\ \forall i.\tag{13}$$

为简化求解,这里令 $\boldsymbol{X}=(\boldsymbol{X}_{\mathrm{p}},\boldsymbol{X}_{\mathrm{g}})$ 表示检索集特征矩阵与匹配集特征矩阵的拼合矩阵;类似地,令 $\boldsymbol{Z}=(\boldsymbol{Z}_{\mathrm{p}},\boldsymbol{Z}_{\mathrm{g}})$ 为学习到的系数矩阵的拼合.对式(13)应用拉格朗日对偶方法可以解得:

$$\boldsymbol{D}=\boldsymbol{X}\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}+\boldsymbol{\Lambda}^{*})^{-1},\tag{14}$$

其中,$\boldsymbol{\Lambda}^{*}$ 为由最优对偶变量组成的一个对角矩阵.在实际运算时 $\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}+\boldsymbol{\Lambda}^{*}$ 可能会出现奇异的情况,此时可以进行适当的正则平滑或取伪逆.

4) 更新 $\boldsymbol{W}$

在固定 $\boldsymbol{Z}_{\mathrm{p}},\boldsymbol{Z}_{\mathrm{g}},\boldsymbol{D}$ 时,式(6)关于 $\boldsymbol{W}$ 的优化目标等价于:

$$\min_{\boldsymbol{W}}\Gamma(\boldsymbol{W})=\frac{\lambda_0}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_{ij}\psi_{\boldsymbol{W}}(\boldsymbol{z}_{\mathrm{p}}^{i},\boldsymbol{z}_{\mathrm{g}}^{j})+\frac{\lambda_2}{2}\|\boldsymbol{W}\|_{\mathrm{F}}^{2}.\tag{15}$$

对式(15)计算关于 $\boldsymbol{W}$ 的导数:

$$\frac{\partial\Gamma(\boldsymbol{W})}{\partial\boldsymbol{W}}=\lambda_0\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_{ij}\delta_{ij}\boldsymbol{W}\boldsymbol{C}_{ij}+\lambda_2\boldsymbol{W},\tag{16}$$

其中,$\boldsymbol{C}_{ij}=(\boldsymbol{z}_{\mathrm{p}}^{i}-\boldsymbol{z}_{\mathrm{g}}^{j})(\boldsymbol{z}_{\mathrm{p}}^{i}-\boldsymbol{z}_{\mathrm{g}}^{j})^{\mathrm{T}}$.

但是如果直接采用式(16)计算 $\frac{\partial\Gamma(\boldsymbol{W})}{\partial\boldsymbol{W}}$ 需要计算大量外积矩阵 $\boldsymbol{C}_{ij}$,必然会带来巨大的运算开销.为了降低运算代价和提高计算性能,可以进一步将式(16)表达为矩阵运算形式:

$$\frac{\partial\Gamma(\boldsymbol{W})}{\partial\boldsymbol{W}}=\lambda_0\boldsymbol{W}(\boldsymbol{Z}_{\mathrm{p}}\boldsymbol{R}\boldsymbol{Z}_{\mathrm{p}}^{\mathrm{T}}-\boldsymbol{Z}_{\mathrm{p}}\boldsymbol{H}\boldsymbol{Z}_{\mathrm{g}}^{\mathrm{T}}-\boldsymbol{Z}_{\mathrm{g}}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{Z}_{\mathrm{p}}^{\mathrm{T}}+\boldsymbol{Z}_{\mathrm{g}}\boldsymbol{R}\boldsymbol{Z}_{\mathrm{g}}^{\mathrm{T}})+\lambda_2\boldsymbol{W},\tag{17}$$

其中,$\boldsymbol{R}=diag(\sum_{i}\beta_{ij}\delta_{ij},\ \forall j)$ 是一个对角矩阵,它的主对角元素是以 $\beta_{ij}\delta_{ij}$ 为元素的矩阵的行和;$\boldsymbol{H}=diag(\sum_{j}\beta_{ij}\delta_{ij},\ \forall i)$ 是以 $\beta_{ij}\delta_{ij}$ 为元素的矩阵列和组成的对角阵.显然,采用式(17)计算梯度能够显著降低运算量.在获得 $\frac{\partial\Gamma(\boldsymbol{W})}{\partial\boldsymbol{W}}$ 后,可以采用梯度下降方法更新 $\boldsymbol{W}$,在第 t 步迭代中的计算式为 $\boldsymbol{W}^{t+1}=\boldsymbol{W}^{t}-\eta\frac{\partial\Gamma(\boldsymbol{W})}{\partial\boldsymbol{W}}$,$\eta$ 为更新步长.

最终,本文提出的联合学习跨视角判别词典与嵌入矩阵的算法模型可以被描述为算法 1 所示的流程框架,本文将其称为跨视角判别词典嵌入(cross-view discriminative dictionary learning with metric embedding, CDDM)算法.

算法 1. 跨视角判别词典嵌入(CDDM)算法.

输入:训练集特征矩阵 $\boldsymbol{X}_{\mathrm{p}},\boldsymbol{X}_{\mathrm{g}}$,标签矩阵 $\boldsymbol{Y}$,参数 $\lambda_0,\lambda_1,\lambda_2$;

初始化:根据式(2)获得初始的 $\boldsymbol{D},\boldsymbol{Z}_{\mathrm{p}},\boldsymbol{Z}_{\mathrm{g}},\boldsymbol{W}=\boldsymbol{I},\mu=E[d_{\boldsymbol{I}}(\boldsymbol{z}_{\mathrm{p}},\boldsymbol{z}_{\mathrm{g}})]$;

① for $t=1,2,\cdots,T$ do

② 　根据式(4)(7)(8)计算 β_{ij};

③ 　根据式(11)更新 $\boldsymbol{Z}_{\mathrm{p}}$;

④ 　根据式(12)更新 $\boldsymbol{Z}_{\mathrm{g}}$;

⑤ 　根据式(14)更新 $\boldsymbol{D}$;

⑥ 　while 不收敛 do

⑦ 　　根据式(17)计算 $\frac{\partial\Gamma(\boldsymbol{W})}{\partial\boldsymbol{W}}$;

⑧ 　　$\boldsymbol{W}^{t+1}\leftarrow\boldsymbol{W}^{t}-\eta\frac{\partial\Gamma(\boldsymbol{W})}{\partial\boldsymbol{W}}$;

⑨ 　end while

⑩ end for

输出:$\boldsymbol{D}^{*}=\boldsymbol{D}^{t},\boldsymbol{Z}_{\mathrm{p}}^{*}=\boldsymbol{Z}_{\mathrm{p}}^{t},\boldsymbol{Z}_{\mathrm{g}}^{*}=\boldsymbol{Z}_{\mathrm{g}}^{t},\boldsymbol{W}^{*}=\boldsymbol{W}^{t}$.

2.4　在行人再识别中的应用

在完成跨视角判别词典嵌入算法模型的训练后,假设待测试的检索图像特征表达为 $\boldsymbol{x}_{\mathrm{pt}}$,待匹配图像集的特征为 $\boldsymbol{x}_{\mathrm{gt}}^{i}(i=1,2,\cdots,N)$,则实施行人再识别的过程为:

1) 对于每个匹配集特征 $\boldsymbol{x}_{\mathrm{gt}}^{i}$,根据获得的判别词典 $\boldsymbol{D}$ 计算的系数表达向量 $\boldsymbol{z}_{\mathrm{gt}}^{i}$ 为

$$\boldsymbol{z}_{\mathrm{gt}}^{i}=\arg\min_{\boldsymbol{z}}\|\boldsymbol{x}_{\mathrm{gt}}^{i}-\boldsymbol{D}\boldsymbol{z}\|_{2}^{2}+\lambda_1\|\boldsymbol{z}\|_{2}^{2}.\tag{18}$$

2) 采用类似过程 1)的方法根据式(18)获得 $\boldsymbol{x}_{\mathrm{pt}}$ 的系数表达 $\boldsymbol{z}_{\mathrm{pt}}$.

3) 对 $\boldsymbol{z}_{\mathrm{pt}}$ 与 $\boldsymbol{z}_{\mathrm{gt}}^{i}(i=1,2,\cdots,N)$ 根据 $\boldsymbol{d}(i)=\|\boldsymbol{W}(\boldsymbol{z}_{\mathrm{pt}}-\boldsymbol{z}_{\mathrm{gt}}^{i})\|_{2}$ 进行距离计算.

4) 对距离向量 $\boldsymbol{d}$ 排序,获得各匹配图像按距离升序排序的列表.

3　实　　验

本节对提出的跨视角判别词典嵌入算法 CDDM

在 VIPeR,GRID,3DPeS 这 3 个常用的行人识别数据集上进行了性能测试,并对实验结果进行了比较和分析.

3.1 实验设置

1) 数据集

实验使用 3 个数据集:

① VIPeR[30]是最早公开的专门用于检测行人再识别算法性能的数据集,在行人再识别的研究中应用最为广泛.该数据集中包含有从 2 个不重叠摄像机视角下拍摄的 632 个行人,每个行人在各摄像机下均只有 1 张图像,因此该数据集共有 1264 张图像.这些行人图像已经被统一为 128×48 的像素大小,他们在不同视角下的外观差异主要来自于强烈的光照变化、姿态与视角差异.

② GRID 数据集[3]由安装在地铁站中的 8 台摄像机拍摄获得,行人图像被组织到了检索集 Probe 与匹配集 Gallery 2 个目录下.其中有 250 个行人在 2 个目录下各有 1 张图片,Gallery 目录下还有 775 个行人在 Probe 下没有正确匹配的图像.由于存在干扰图像和强烈的光照视角变化,以及摄像机视角数多达 8 个,在 GRID 数据集上的行人再识别工作相当困难.

③ 3DPeS 数据集[31]中包含有从 8 个摄像机视角下拍摄的 192 个行人,每个行人的图像数为 2～26 张不等.由于 3DPeS 在采集时持续了数天中不同的时间段,因此该数据集中的图像存在强烈的光照变化,另外行人在不同摄像机下的姿态差异也比较大.

图 3 给出了从上述 3 个数据集中随机选取的部分行人图像示例,每一列的 2 张图像取自于同一行人在不同摄像机下的视频画面.

(a) VIPeR (b) GRID (c) 3DPeS

Fig. 3 Example images from VIPeR, GRID, and 3DPeS

图 3 VIPeR,GRID,3DPeS 数据集中部分行人图像

2) 特征提取

实验中采用了文献[32]中改进后的局部最大出现特征和使用深度残差网络[33](deep residual net, ResNet)提取的深度特征来表达行人图像.在文献[32]设计的特征描述子中融合了从密集网格提取的 LOMO[6]描述子与从图像前景两层水平条空间中提取的 LOMO 变体,其中使用的基本特征有联合 HSV 与 RGB 颜色直方图、局部三值模式(local ternary pattern, LTP)和显著颜色名称 SCN 特征.该描述子中从密集网格提取的特征能够比较好地捕捉图像的细节,从水平条中提取的特征能够更好地刻画图像的整体外观,两者的融合赋予了描述子"由粗到细"的行人外观表达能力.在使用深度残差网络提取图像特征时,使用了在 ImageNet 上训练好的 152 层的 ResNet-152 网络,提取的特征为 2048 维.

3) 参数设置

实验中模型的超参数通过交叉验证获得,具体设置为 $\lambda_0=1,\lambda_1=0.2,\lambda_2=0.1$.在使用梯度下降更新 $\boldsymbol{W}$ 时,学习率 η 的初始值设为 0.01;在迭代中若目标函数值下降则对 η 扩大 1.2 倍,否则对 η 乘上 0.9 的收缩因子.在选择词典基的数量时取 $m=200$,关于基数量的选择将在 3.4 节中作进一步的讨论.

4) 评价方案与指标

实验中对各数据集均采用了单张-单张(single-shot vs single-shot)的匹配测试方案,由于在 3DPeS 中每个行人的图像数不等,因此与文献[34]中的方法相同,对每个行人随机选择一张图像用于检索,剩余图像均作为匹配集.在评价指标上选择了在行人再识别研究中应用最为广泛的累积匹配特征(cumulative matching characteristic, CMC)曲线,它反映了在前个匹配集图像中发现正确匹配的概率.为了便于和文献公开的方法作性能对比,在表格中仅选择了 CMC 曲线部分排序位置(rank)上的匹配精度.为了获得更具有鲁棒性的实验结果,在每个数据集上都进行了 10 次随机的训练集/测试集划分,取它们的平均 CMC 作为最终实验数据.

3.2　与文献公开的结果对比

实验中首先把本文 CDDM 算法在各个数据集上取得的行人再识别结果与文献中公开的数值进行了对比.

在 VIPeR 数据集上进行行人再识别时采用了当前应用最为广泛的等量划分方案,数据集中 632 个行人被划分为 2 组,每组 316 个行人.其中一组作为训练集,另一组作为测试集.实验对比的方法包含有监督平滑流形[35](supervised smoothed manifold, SSM)方法、空间约束相似度学习[34](spatial constrained similarity learning on polynomial feature map, SCSP)算法、零空间 Foley-Sammon 变换[11](null Foley-Sammon transform, NFST)、度量组合[13](metric ensemble, ME)、摄像机相关性已知的特征扩增[36](camera correlation aware feature augmentation, CRAFT)、加权线性编码[37](weighted linear coding, WLC)、基于核化跨视角协同表达分类[26](kernel cross-view collaborative representation based classification, KX-CRC)、基于加速近邻梯度的度量学习[20](metric learning by accelerated proximal gradient, MLAPG)、XQDA[6]、GOG[16]、深度多层相似度[5](deep multi-level similarity, DMS)和 SpindleNet[7]等.

表 1 与图 4① 给出了 CDDM 算法及其他算法在 VIPeR 数据集上的行人再识别结果对比.从对比结果可以看出 CDDM 在性能上明显优于其他方法.特别是在 rank-1 上,CDDM 取得了 60.93%的正确匹配率,也是唯一达到 60%匹配率的方法.和此前 SpindelNet 取得的最优结果 53.80%相比,CDDM 比其高出了 7.13%,这充分展现了 CDDM 优异的性能.在其他的各个 rank 上,CDDM 也表现出显著的性能优势.在对比方法中,SpindelNet,CRAFT,DMS 都是基于深度学习模型的方法,但是在 VIPeR 数据集上由于样本相对较少,无法完全发挥它们的性能,虽然它们在 rank-1 上都达到 50%以上的匹配率,但整体性能仍相对较弱.在对比方法中 SSM,SCSP,NFST,MLAPG,XQDA 等均为度量学习算法,KX-CRC 与 WLC 为基于词典学习的方法,与它们相比,CDDM 联合学习了判别词典与度量矩阵,能够同时利用两者的优势,因此具有更强的匹配性能.

在 GRID 数据集上,实验中将在 Probe 与 Gallery 目录下都有图像的 250 人均分为 2 组.其中一组作为训练集,另一组和 Gallery 目录下的 775 张干扰图像作为测试集.在该数据集上本文 CDDM 算法与样本独立的 SVM[27](sample specific SVM, SSSVM), NK3ML[38](nullspace kernel maximum margin metric learning)等其他文献中公开的结果对比如表 2 和图 5 所示.从表 2 可知,CDDM 再次取得了最优的结果.在 rank-1 上 CDDM 取得的正确匹配率达到了 28.20%,比此前最优的 NK3ML 和 SSM 高出了 1%,在其他 rank 上 CDDM 也取得了更为优秀的再识别性能.这说明 CDDM 能够较好地应对 GRID 数据集中复杂的视角变化与光照等干扰.

Table 1　Performance Comparison of CDDM with State-of-the-Art Algorithms on VIPeR

表 1　CDDM 与其他算法在 VIPeR 数据集上匹配率对比　%

Method	rank-1	rank-5	rank-10	rank-20	Reference
CDDM	60.93	86.68	93.89	98.35	Ours
SpindelNet	53.80	74.10	83.20	92.10	Ref[7]
SSM	53.73		91.49	96.08	Ref[35]
SCSP	53.54	82.59	91.49	96.65	Ref[34]
KX-CRC	51.40	81.20	89.70	95.60	Ref[26]
WLC	51.40	76.40	84.80		Ref[37]
NFST	51.17	82.09	90.51	95.52	Ref[11]
CRAFT	50.28	79.97	89.56	95.51	Ref[36]
DMS	50.10	73.10	84.35		Ref[5]
GOG	49.72	79.72	88.67	94.53	Ref[16]
ME	45.90	77.50	88.90	95.80	Ref[13]
MLAPG	40.73	69.96	82.34	92.37	Ref[20]
XQDA	40.00	68.13	80.51	91.08	Ref[6]

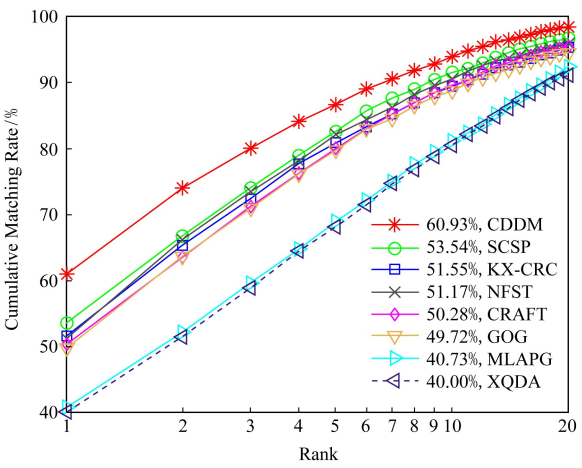


Fig. 4　CMC curves of different algorithms on VIPeR dataset

图 4　不同算法在 VIPeR 数据集上的 CMC 曲线

① 由于表 1 中部分方法未公开代码或 CMC,因此未能全部绘制.

在3DPeS数据集上实验时采用了与文献[34]相同的数据集分割方案,从该数据集随机选择96人作为训练集,剩余96人作为测试集.对于每个行人,随机选择一张图像来创建匹配集,剩余图像均用于检索.在该数据集上与本文CDDM算法进行对比的方法有核化局部Fisher线性判别[39](kernel local Fisher discriminant analysis, KLFDA)、深度排序大间隔度量学习[40](deep ranking by large adaptive margin learning, DRLAML)、域引导丢弃方法[41](domain guided dropout, DGD)、SpindelNet、SCSP和ME.表3列出了这些算法在rank1,5,10,20上取得的累积匹配正确率.

Table 2 Performance Comparison of CDDM with State-of-the-Art Algorithms on GRID

表2 CDDM与其他算法在GRID数据集上匹配率对比 %

Method	rank-1	rank-5	rank-10	rank-20	Reference
CDDM	28.20	52.40	64.00	74.10	Ours
NK3ML	27.20		60.96	71.04	Ref[38]
SSM	27.20		61.12	70.56	Ref[35]
KX-CRC	26.90	45.70	57.50	70.20	Ref[26]
CRAFT	26.00	50.60	62.50	73.30	Ref[36]
GOG	24.72	46.96	58.40	68.96	Ref[16]
SCSP	24.24	44.56	54.08	65.20	Ref[34]
SSSVM	22.40	40.40	51.28	61.20	Ref[35]
MLAPG	16.64	33.12	41.20	52.96	Ref[20]
XQDA	16.56	33.84	41.84	52.40	Ref[6]

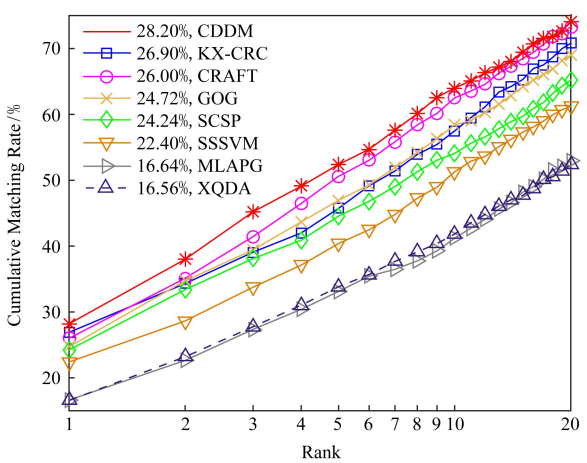


Fig. 5 CMC curves of different algorithms on GRID dataset

图5 不同算法在GRID数据集上的CMC曲线

从表3中的数据可以看出与其他方法相比,本文CDDM算法取得的匹配结果依然领先于其他方法.在rank-1上CDDM的匹配率为65.57%,比排在第2名的SpindelNet高出了3.47%,在其他rank上也均优于各对比方法.尽管基于深度学习方法的SpindelNet,DGD,DRLAML在该数据集上的识别性能比其他方法有所提升,但仍弱于本文CDDM算法.与SCSP,ME,KLFDA等度量学习方法相比,CDDM也具有明显的性能优势.

Table 3 Performance Comparison of CDDM with State-of-the-Art Algorithms on 3DPeS

表3 CDDM与其他算法在3DPeS数据集上匹配率对比 %

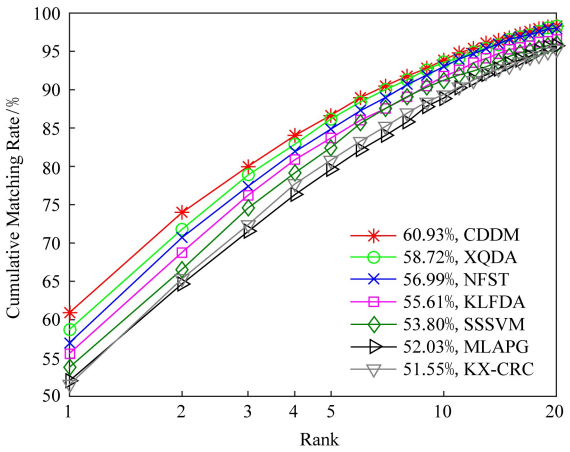
Method	rank-1	rank-5	rank-10	rank-20	Reference
CDDM	65.57	84.53	91.60	96.24	Ours
SpindelNet	62.10	83.40	90.50	95.70	Ref[7]
DRLAML	58.30	74.00		88.50	Ref[40]
SCSP	57.29	78.97	86.02	91.51	Ref[34]
DGD	55.20	76.40	84.90	91.90	Ref[41]
ME	53.33	76.79	84.95	92.78	Ref[13]
KLFDA	54.02	77.74	85.90	92.38	Ref[39]

3.3 采用相同特征描述子时算法性能比较

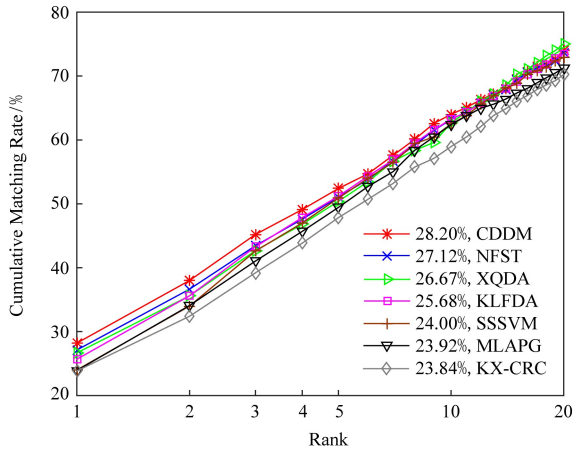
在3.2节的行人再识别结果数据对比中,尽管各算法模型均采用了相同的数据集划分方案,但是各模型的结构与使用的特征描述子各不相同,因此性能对比中必然存在一定的不公平性.特别是对于SpindelNet[7]等基于深度学习的方法,尽管已经取得比较优异的性能,但是受到数据集中样本数量较少的限制,它们的性能难以得到完全发挥.为了进一步对CDDM算法的性能进行分析,本节对CDDM与其他可获得源码的算法在采用相同特征时的再识别性能进行了测试.实验中对比的方法有SSSVM,MLAPG,XQDA,KLFDA,NFST,KX-CRC,其中SSSVM和KX-CRC为学习判别词典的方法,其余为度量学习方法.

采用本文使用的特征描述子,在3个数据集上各算法取得的CMC曲线及rank-1匹配率如图6所示.从图6可以看出本文CDDM算法在3个数据集上均取得了优于其他算法的匹配性能.在VIPeR数据集上,CDDM的rank-1匹配率为60.93%,排在第2名的是XQDA,其正确匹配率为58.72%,比CDDM弱了2.21%.在GRID与3DPeS数据集上,排在第2名的方法分别是NFST和XQDA.与它们相比,CDDM分别具有1.08%和3.33%的rank-1性能优势.综合各方法在3个数据集上的再识别性能可以发现,在使用相同特征描述子时,尽管各方法在

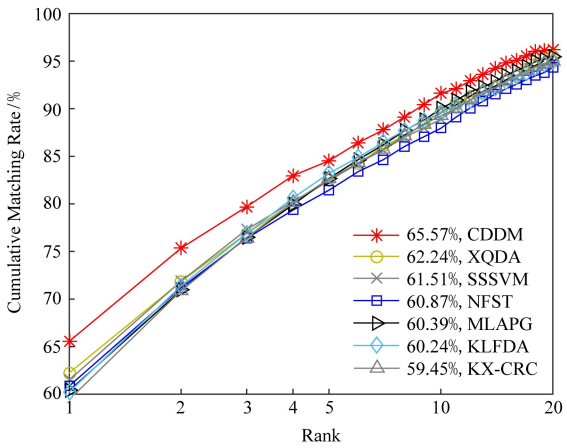
不同数据集上的性能会存在差异,但是本文 CDDM 由于同时学习了判别词典与度量矩阵,始终表现出最优的行人再识别性能.该实验充分说明了联合学习判别词典与度量矩阵所带来的优势.



(a) VIPeR



(b) GRID



(c) 3DPeS

Fig. 6　Performance comparison of CDDM with other algorithms using the same feature representation

图 6　采用相同特征描述子时 CDDM 与其他算法的性能对比

3.4　算法分析

在本文提出的 CDDM 算法中,学习的判别词典中基向量的数量、样本对权重的分配、使用的特征描述子等均会给算法的最终性能带来很大的影响,在本节实验中对它们分别进行了分析.

1) 词典基向量数量对算法性能的影响

图 7 给出了在 VIPeR,GRID,3DPeS 数据集上,采用本文 CDDM 算法进行行人再识别时不同的词典基向量数量对 rank-1 正确匹配率的影响.从图 7 可以看出,随着词典基向量数量的增长,各数据集上的 rank-1 匹配率均呈上升趋势;但在词典数达到 200 后,各匹配率基本上保持稳定.因此,本文选择了 200 作为词典基向量数.

2) 联合学习判别词典与距离度量的作用

在本文 CDDM 算法中联合学习了判别词典与度量矩阵,为了验证联合学习度量矩阵所带来的性能提升,实验中将算法 1 中的投影矩阵设置为单位矩阵进行了实验(下面标记为 CDDI),并与 CDDM 作了对比.表 4 给出了它们在不同数据集上的实验结果,从表 4 数据可知联合学习判别词典与度量矩阵时,CDDM 的匹配性能显著优于 CDDI.在 VIPeR,GRID,3DPeS 上,CDDM 的 rank-1 匹配率比 CDDI 分别高出了 7.13%,4.88%,5.15%,说明联合学习度量矩阵更有助于发现数据的内在结构,获得的投影子空间比使用欧氏距离具有更优的判别性.

3) 融合深度特征与手工特征带来的性能提升

本文实验中使用了手工设计的特征描述子(标记为 HCFeat)与 ResNet152 学习到的深度特征表达(标记为 DeepFeat),图 8 给出了它们在融合使用(标记为 ConFeat)与独立使用时获得的 CMC 曲线.

从图 8 可以发现 2 种特征融合后取得的匹配性能显著优于分开独立使用时的结果,本文认为这主要是因为它们捕获了具有互补性的图像低层外观与高层语意信息.

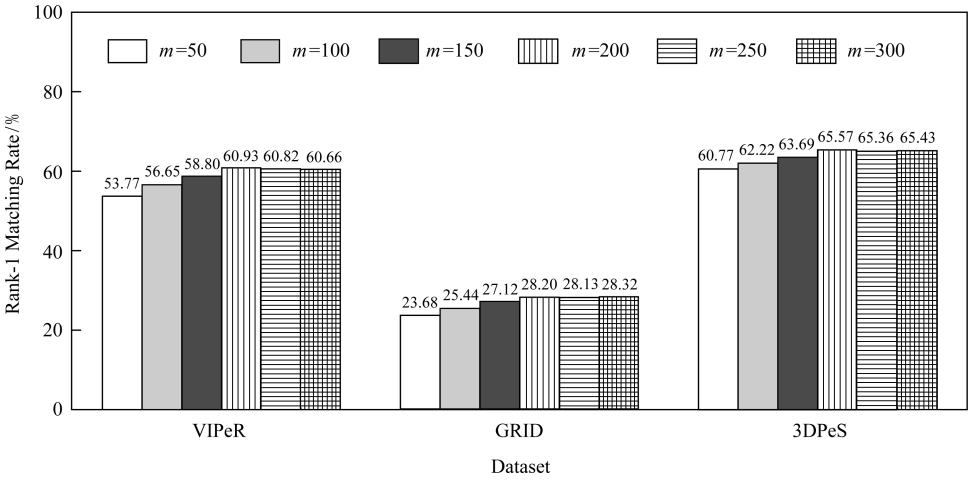


Fig. 7 Influence of the number of bases for dictionary learning on rank-1 matching rate

图 7 词典基向量数 m 对 rank-1 匹配率的影响

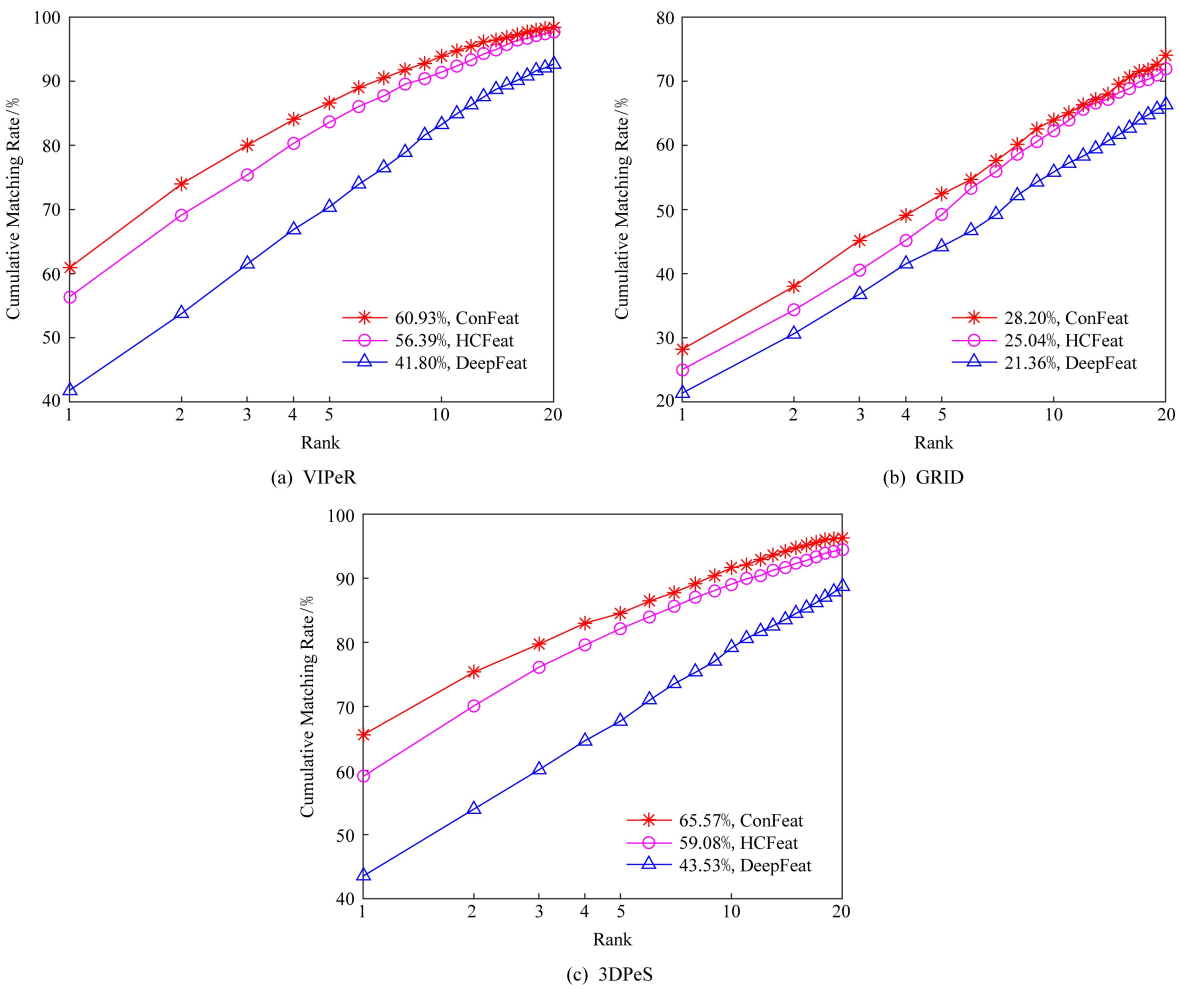


Fig. 8 Performance comparison of feature descriptors

图 8 特征描述子性能对比

Table 4　Matching Rate Comparison of CDDM with CDDI

表 4　CDDM 与 CDDI 匹配率对比　　%

Dataset	Method	rank-1	rank-5	rank-10	rank-20
VIPeR	CDDM	60.93	86.68	93.89	98.35
	CDDI	53.80	81.96	91.77	97.47
GRID	CDDM	28.20	52.40	64.00	74.10
	CDDI	23.32	47.68	61.12	70.80
3DPeS	CDDM	65.57	84.53	91.60	96.24
	CDDI	60.42	82.41	89.20	95.04

4) 样本对的权重分配对算法性能的影响

为了降低不均衡训练样本带来的度量偏差问题,本文采用了自适应的样本对权重分配策略.为了考查样本对权重分配对算法性能的影响,实验中对所有样本对在不考虑权重(设置式(8)中 $\beta_{ij}=1$)时的匹配性能与使用式(8)权重分配方案取得的结果进行性能对比.图 9 给出了这 2 种情况下在各数据集上的 rank-1 匹配率.从图 9 可以发现,使用了自动权重分配策略比不考虑权重分别带来了 7.07%,3.68%,8.66%的性能提升,说明本文权重分配策略对训练样本数量不平衡引起的度量偏差问题具有良好的抑制作用.

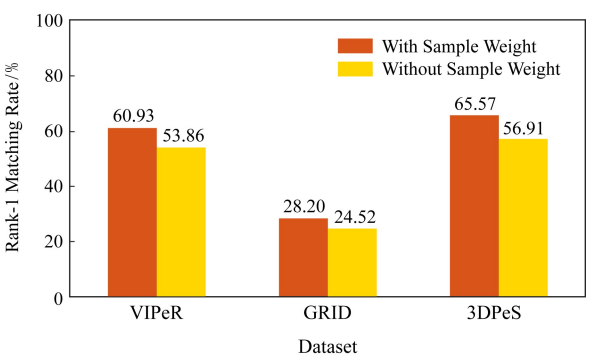


Fig. 9　Comparison of rank-1 matching rate

图 9　rank-1 匹配率对比

4　结束语

本文提出了一种跨视角判别词典嵌入的行人再识别算法,该算法中通过交替迭代优化的方式联合学习了跨视角的判别性词典和嵌入子空间,从而将词典表达与度量学习的优势结合了起来.为了降低在学习距离度量时由于正负样本对数量不均衡带来的度量偏差问题,在算法中还引入了对训练样本自适应赋予权重的策略.在 3 个广泛使用的行人再识别数据集上的实验结果表明,本文方法取得了优秀的跨视角行人再识别性能.由于当前的工作主要关注于小数集上的行人再识别,在后续的工作中将尝试基于深度学习模型学习判别词典,并应用到更接近现实场景的大型数据集上.

参　考　文　献

[1] Gong Shaogang, Cristani M, Yan Shuicheng, et al. Person Re-identification [M]. Berlin: Springer, 2014

[2] Huang Jipeng, Shi Yinghuan, Gao Yang. Multi-scale faster-RCNN algorithm for small object detection [J]. Journal of Computer Research and Development, 2019, 56(2): 319-327 (in Chinese)
(黄继鹏, 史颖欢, 高阳. 面向小目标的多尺度 Faster-RCNN 检测算法[J]. 计算机研究与发展, 2019, 56(2): 319-327)

[3] Chen Change Loy, Xiang Tao, Gong Shaogang. Multi-camera activity correlation analysis [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2009: 1988-1995

[4] Zheng Liang, Yang Yi, Hauptmann A G. Person re-identification: Past, present and future [J]. arXiv preprint, arXiv:1610.02984, 2016

[5] Guo Yiluan, Cheung N M. Efficient and deep person re-identification using multi-level similarity [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 2335-2344

[6] Liao Shengcai, Hu Yang, Zhu Xiangyu, et al. Person re-identification by local maximal occurrence representation and metric learning [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 2197-2206

[7] Zhao Haiyu, Tian Maoqing, Sun Shuyang, et al. Spindle Net: Person re-identification with human body region guided feature decomposition and fusion [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 1077-1085

[8] Zhong Zhun, Zheng Liang, Zheng Zhedong, et al. Camera style adaptation for person re-identification [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 5157-5166

[9] Yang Yang, Yang Jimei, Yan Junjie, et al. Salient color names for person re-identification [C] //Proc of European Conf on Computer Vision. Berlin: Springer, 2014: 536-551

[10] Lisanti G, Masi I, Bagdanov A D, et al. Person re-identification by iterative re-weighted sparse ranking [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(8): 1629-1642

[11] Zhang Li, Xiang Tao, Gong Shaogang. Learning a discriminative null space for person re-identification [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 1239-1248

[12] Wang Jing, Wang Zheng, Liang Chao, et al. Equidistance constrained metric learning for person re-identification [J]. Pattern Recognition, 2018, 74: 38-51

[13] Paisitkriangkrai S, Wu Lin, Shen Chunhua, et al. Structured learning of metric ensembles with application to person re-identification [J]. Computer Vision and Image Understanding, 2017, 156: 51-65

[14] Srikrishna K, Yang Li, Richard J R. Person re-identification with discriminatively trained viewpoint invariant dictionaries [C] //Proc of IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2015: 4516-4524

[15] Yang Yang, Lei Zhen, Zhang Shifeng, et al. Metric embedded discriminative vocabulary learning for high-level person representation [C] //Proc of the 30th AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI Press, 2016: 3648-3654

[16] Matsukawa T, Okabe T, Suzuki E, et al. Hierarchical gaussian descriptor for person re-identification [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 1363-1372

[17] Zhao Rui, Ouyang Wanli, Wang Xiaogang. Learning mid-level filters for person re-identification [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 144-151

[18] Ma Bingpeng, Su Yu, Jurie F. Covariance descriptor based on bio-inspired features for person re-identification and face verification [J]. Image and Vision Computing, 2014, 32(6): 379-390

[19] Mignon A, Jurie F. PCCA: A new approach for distance learning from sparse pairwise constraints [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2012: 2666-2672

[20] Liao Shengcai, Li Stan Z. Efficient PSD constrained asymmetric metric learning for person re-identification [C] //Proc of Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2015: 3685-3693

[21] Zheng Weishi, Gong Shaoang, Xiang Tao. Re-identification by relative distance comparison [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(3): 653-668

[22] You Jinjie, Wu Ancong, Li Xiang, et al. Top-Push video-based person re-identification [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 1345-1353

[23] Köestinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2012: 2288-2295

[24] Sumit S, Patel V M, Nasrabadi N M, et al. Joint sparse representation for robust multimodal biometrics recognition [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 36(1): 113-126

[25] Liu Xiao, Song Mingli, Tao Dacheng, et al. Semi-supervised coupled dictionary learning for person re-identification [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 3550-3557

[26] Prates R, Schwartz W R. Kernel cross-view collaborative representation based classification for person re-identification [J]. Journal of Visual Communication and Image Representation, 2019, 58: 304-315

[27] Zhang Ying, Li Baohua, Lu Huchuan, et al. Sample-specific SVM learning for person re-identification [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 1278-1287

[28] Kodirov E, Xiang Tao, Fu Zhenyong, et al. Person re-identification by unsupervised ℓ_1 graph learning [C] //Proc of European Conf on Computer Vision. Berlin: Springer, 2016: 178-195

[29] Lu Jiwen, Zhou Xiuzhuang, Yap-Pen T, et al. Neighborhood repulsed metric learning for kinship verification [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 36(2): 331-45

[30] Gray D, Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features [C] //Proc of European Conf on Computer Vision. Berlin: Springer, 2008: 262-275

[31] Baltieri D, Vezzani R, Cucchiara R. 3DPes: 3D people dataset for surveillance and forensics [C] //Proc of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding. New York: ACM, 2011: 59-64

[32] Dong Husheng, Lu Ping, Zhong Shan, et al. Person re-identification by enhanced local maximal occurrence representation and generalized similarity metric learning [J]. Neurocomputing, 2018, 307: 25-37

[33] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 770-778

[34] Chen Dapeng, Yuan Zejian, Chen Badong, et al. Similarity learning with spatial constraints for person re-identification [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 1268-1277

[35] Bai Song, Bai Xiang, Tian Qi. Scalable person re-identification on supervised smoothed manifold [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 2530-2539

[36] Chen Yingcong, Zhu Xiatian, Zheng Weishi, et al. Person re-identification by camera correlation aware feature augmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2018, 40(2): 392-408

[37] Yang Yang, Wen Longyin, Lyu Siwei, et al. Unsupervised learning of multi-level descriptors for person re-identification [C] //Proc of the 31st AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2017: 4306-4312

[38] Ali T M F, Chaudhuri S. Maximum margin metric learning over discriminative nullspace for person re-identification [C] //Proc of European Conf on Computer Vision. Berlin: Springer, 2018: 122-138

[39] Xiong Fei, Gou Mengran, Camps O, et al. Person re-identification using kernel-based metric learning methods [C] //Proc of European Conf on Computer Vision. Berlin: Springer, 2014: 1-16

[40] Wang Jiayun, Zhou Sanping, Wang Jinjun, et al. Deep ranking model by large adaptive margin learning for person re-identification [J]. Pattern Recognition, 2018, 74: 241-252

[41] Xiao Tong, Li Hongsheng, Ouyang Wanli, et al. Learning deep feature representations with domain guided dropout for person re-identification [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 1249-1258

Lu Ping, born in 1979. Master, associate professor. Her main research interests include digital image processing and pattern recognition.

Dong Husheng, born in 1981. PhD, lecturer. His main research interests include computer vision, image and video processing, and machine learning.

Zhong Shan, born in 1983. PhD, lecturer. Her main research interests include machine learning and deep learning.

Gong Shengrong, born in 1966. PhD, professor, and PhD supervisor in the School of Computer Science & Technology, Soochow University. His main research interests include image and video processing, pattern recognition, and computer vision.