

基于距离中心化与投影向量学习的行人重识别

丁宗元¹ 王洪元¹ 陈付华² 倪彤光¹

¹(常州大学信息科学与工程学院 江苏常州 213164)

²(西自由大学自然科学与数学系 美国西弗吉尼亚州西自由市 26074)

(dzyha2011@foxmail.com)

Person Re-Identification Based on Distance Centralization and Projection Vectors Learning

Ding Zongyuan¹, Wang Hongyuan¹, Chen Fuhua², and Ni Tongguang¹

¹(School of Information Science and Engineering, Changzhou University, Changzhou, Jiangsu 213164)

²(Department of Nature Science and Mathematics, West Liberty University, West Liberty, West Virginia, USA 26074)

Abstract Existing projection-based person re-identification methods usually suffer from long time training, high dimension of projection matrix, and low matching rate. In addition, the intra-class samples may be much less than the inter-class samples when a training data set is built. To solve these problems, this paper proposes a distance-centralization based algorithm for similarity metric learning. When a training data set is to be built, the feature values of a same target person are centralized and the inter-class distances are built by these centralized values, while the intra-class distances are still directly built from original samples. As a result, the number of intra-class samples and the number of inter-class samples can be much closer, which reduces the risk of overfitting because of class imbalance. In addition, during learning projection matrix, the resulted projection vectors can be approximately orthogonal by using a strategy of updating training data sets. In this way, the proposed method can significantly reduce both the computational complexity and the storage space. Finally, the conjugate gradient method is used in the projection vector learning. The advantage of this method is its quadratic convergence, which can promote the convergence. Experimental results show that the proposed algorithm has higher efficiency. The matching rate can be significantly improved, and the time of training is much shorter than most of existing algorithms of person re-identification.

Key words person re-identification; distance centralization; metric learning; projection vectors; conjugate gradient method

摘要 现有的基于投影的行人重识别方法具有训练时间长、投影矩阵维数高、识别率低等问题。此外在建立训练集时,还会出现类内样本数目远少于类间样本数目的情况。针对这些问题,提出了基于距离中心化的相似性度量算法。在构建训练集时,将同一组目标群体特征值中心化,利用中心特征值来构建类间距离,而类内距离保持不变。这样使得类内类间样本数目接近,可以很好地缓解类别不平衡所带来的过拟合风险。另外在学习投影矩阵时,利用训练集更新策略,学习若干组投影向量,使得到的投影向量近

收稿日期:2017-01-12;修回日期:2017-05-16

基金项目:国家自然科学基金项目(61572085,61502058)

This work was supported by the National Natural Science Foundation of China (61572085, 61502058).

通信作者:王洪元(hywang@cczu.edu.cn)

似正交,这样既可以有效减少运算复杂度和存储复杂度,又可以使得学习到的投影向量能够通过简单的相乘近似得到原来的投影矩阵.最后,在学习投影向量时采用共轭梯度法,该方法具有二次收敛性,能够快速收敛到目标精度.实验结果表明:提出的算法具有较高的效率,在不同数据集上的识别率都有明显的提升,训练时间也比其他常用的行人重识别算法要短.

关键词 行人重识别;距离中心化;度量学习;投影向量;共轭梯度法

中图法分类号 TP391.41

目前,越来越多的摄像系统被广泛地布置到公共场所,进行24小时不间断的监控,产生了大量的视频数据,使得主要依靠人工监控和人工鉴别的传统视频监控系统不仅耗费大量的人力,而且效率非常低下.因此,对视频数据进行自动化处理和分析对提高视频监控的效率有极大的帮助.视频监控中,当一个行人被位于公共场所的某个摄像机所捕捉到,即该行人的一帧或多帧图像被获取后,利用现有摄像机网络去发现该目标行人下一次出现的场所的过程称为行人重识别.行人重识别研究近几年获得了较大的进展.以往研究工作主要是利用投影矩阵的方式,将特征投影到共同子空间,以获得更好的判别性.一些方法对于行人数据的光照条件变化、拍摄角度的不同等一些列问题具有很好的鲁棒性.

近些年在行人重识别问题上的研究涌现出了不少有效的算法.这些方法主要分为2种类型,基于特征的行人重识别^[1-10]和基于度量学习的行人重识别方法^[11-18].而在基于度量学习的行人重识别方法中,主要以学习出“好的”度量为主要目的.其主要思想是利用机器学习的方法,学习出距离测度和分类器,使得类内距离尽量小,类间距离尽量大.该方法对特征选择的要求较低,但具有训练时间长、投影矩阵维数大、容易过拟合等一些问题.

2007年,Weinberger等人提出大间隔最近邻分类^[19](large margin nearest neighbor, LMNN)距离测度学习算法.该方法采用三元组的方式,并加入对不相似对的约束,学习使得尽可能多的三元组样本满足一定条件的最优矩阵 M .2008年,Weinberger又给出了LMNN的快速解法^[20],采用激活策略的方式来避免计算所有样本的梯度.2010年,Dikmen在LMNN基础上提出改进的算法^[21](LMNN-R),并用于行人重识别问题.2007年,Davis等人提出基于信息论方法^[22](information theoretic metric learning, ITML)的度量学习算法,将马氏度量矩阵 M 映射到一个高斯模型,利用信息散度来度量不同马氏矩阵 M 的相似性.2011年,Zheng等人提出基于

概率相对距离比较^[23](probabilistic relative distance comparison, PRDC)的度量学习算法.作者采用的是logistic误差函数,因此最终的目标函数是一个平滑的凸优化问题.为了避免过拟合,作者对矩阵 M 加入了额外约束.Kostinger等人提出的基于简单而且直接策略^[24](keep it simple and straight MEtric, KISSME)的度量学习算法,通过直接的方式来对模型的参数进行估计出度量矩阵 M ,该方法能够用于大尺度数据的学习,但该方法在训练数据不足时,估计出来的参数不准确.2013年,Pedagadi将局部线性判别分析^[25](local fisher discrimination analysis, LFDA)方法用于行人重识别问题,其对特征进行了主成分分析(PCA)的方式,提取不同特征的主要成分.2015年,Chen等人提出了建立相似对来进行度量学习^[26],利用给定的相似矩阵进行度量学习,但该方法需要事先给定数据集中的相似矩阵,因此当数据集规模很大时,该方法会显得非常耗时.2016年,Chen等人提出了非对称特征投影模型^[27],该方法可以针对相机属性参数的不同学习出不同的投影矩阵,可以有效地提高识别的效率,但该方法需要根据摄像场景的数目分别学习出数目与之对应的投影矩阵,这必然增加训练时间与存储复杂度.

为解决上述行人重识别算法的种种不足,本文提出基于样本距离中心化的相似性度量算法.首先,常用的基于距离学习的算法在构建训练集时,存在反例样本数目远多于正例数目的情况.每个样本在构建反例时需要与所有不同样本的特征向量求特征距离,而在此过程中会产生大量冗余的反例特征距离,大大地增加了训练的时间复杂度,而LMNN算法中利用构建三元组的方法会丢失一些重要的训练特征距离.因此在求样本的反例特征距离时,对同一样本的不同特征向量进行中心化.这样可以很好地缓解类别不平衡带来的过拟合问题.另外,对于同一个样本具有较多的类内样本时,本文采用局部距离中心化的方法,这样可以很好地保留一些重要信息.

其次,常用的基于投影矩阵学习的方法中投影

矩阵维数较高,带来了较大的运算和存储复杂度.本文对投影矩阵进行特征值分解,将其分解成低秩的投影矩阵.因此在训练时,与其他常用的学习整个投影矩阵不同的是,利用本文所提出的迭代优化策略,对样本特征的距离向量进行更新,获得新的样本分布,每次只需要利用更新后的训练集学习一组新的投影向量,在满足目标精度时停止更新,具有较好的降维效果,可以有效地降低运算和存储复杂度,并且该策略能够近似地保证矩阵特征值分解后的向量之间应该保持正交的特性.

最后,针对机器学习中常用的基于梯度下降的优化方法收敛速度慢、运算量大等问题,本文在学习投影向量时采用共轭梯度法的方式,该方法只需要计算一次初始梯度,而且对于二次函数具有二次终止性,可以很快地收敛到目标精度.

本文提出的针对行人重识别常见问题的解决方法主要包括 3 个部分:

- 1) 提出(局部)样本距离中心化的方法来构建训练集;
- 2) 提出新的迭代策略,只需更新一组近似正交的投影向量;
- 3) 利用共轭梯度法来更新投影向量.

1 目标函数的建立

本文将行人重识别问题转化为度量学习问题,假设利用特征向量来代表每个行人数据,将第 i 个行人特征向量记为 $\mathbf{x}_i \in \mathbb{R}^n$,其中 n 为特征维数.因此可以构造训练集为 $X = \{(\mathbf{x}_i, y_i)\}_{i=1}^m$,其中 y_i 为第 i 个行人的标签, m 为所有训练集中所包含的行人数据集数目.对于任意 2 个样本数据集 $(\mathbf{x}_a, \mathbf{x}_b)$ 之间的距离记作 $dis(\mathbf{x}_a, \mathbf{x}_b)$,假设样本 $(\mathbf{x}_i, \mathbf{x}_j)$ 代表同一个行人的数据集(类内样本),样本 $(\mathbf{x}_i, \mathbf{x}_k)$ 代表不同行人的数据集(类间样本),则根据类内距离小于类间距离的原则,有 $dis(\mathbf{x}_i, \mathbf{x}_j) < dis(\mathbf{x}_i, \mathbf{x}_k)$. 记 $S = \{S_t = (\mathbf{d}_t^{\text{pos}}, \mathbf{d}_t^{\text{neg}})\}_{t=1}^m$,其中 $\mathbf{d}_t^{\text{pos}}$ 代表第 t 个样本与其他样本之间的类内距离, $\mathbf{d}_t^{\text{neg}}$ 代表第 t 个样本与其他样本之间的类间距离.为了同时满足这 2 个条件,可以极小化

$$r(\mathbf{d}_t^{\text{pos}}, \mathbf{d}_t^{\text{neg}}) = \mathbf{d}_t^{\text{pos}} - \mathbf{d}_t^{\text{neg}}. \quad (1)$$

式(1)中的函数无界,因此在迭代时无法保证收敛,不利于之后的优化工作,所以,将其转化为常用的 Sigmoid 函数,使其具有连续性:

$$C(\mathbf{d}_t^{\text{pos}}, \mathbf{d}_t^{\text{neg}}) = (1 + \exp(\mathbf{d}_t^{\text{pos}} - \mathbf{d}_t^{\text{neg}}))^{-1}. \quad (2)$$

将式(2)中的所有 t 进行连乘并取对数,然后再取负数,将式(2)转化为求和问题,即可使所有差值向量满足以上约束:

$$f = -\ln\left(\prod_{S_t} C(\mathbf{d}_t^{\text{pos}}, \mathbf{d}_t^{\text{neg}})\right) = \sum_{S_t} \ln(1 + \exp(\mathbf{d}_t^{\text{pos}} - \mathbf{d}_t^{\text{neg}})). \quad (3)$$

在这里,极小化式(1)即等价于极大化式(2),而极大化式(2)又等价于极小化式(3).而由于马氏距离中的马氏矩阵具有很好的投影性质和可学习性,因此这里的距离函数取马氏距离:

$$dis(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{M} (\mathbf{x}_i - \mathbf{x}_j), \quad (4)$$

度量学习就是对矩阵 \mathbf{M} 进行学习,由于 \mathbf{M} 是半正定对称矩阵,在这里,将 \mathbf{M} 进行特征值分解,将其对角化,总能找到一组正交基 \mathbf{P} ,使得 $\mathbf{M} = \mathbf{P}\mathbf{P}^T$,其正交基数目可以小于原矩阵 \mathbf{M} 的秩,这样可以通过学习获得一个降维矩阵 $\mathbf{P} \in \mathbb{R}^{n \times d}$,该降维矩阵的每一列即可以作为每一组更新后特征空间的投影向量,其中 d 为降维后的正交基数目.于是:

$$dis(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{M} (\mathbf{x}_i - \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{P}\mathbf{P}^T (\mathbf{x}_i - \mathbf{x}_j) = \|\mathbf{P}^T (\mathbf{x}_i - \mathbf{x}_j)\|^2, \quad (5)$$

另外,在对小样本进行训练时,仍然会有过拟合的情况发生,因此为了进一步缓解过拟合的风险,同时使学习到的投影矩阵具有一定的稀疏性,本文引入正则化项 $r\|\mathbf{P}\|^2$,其中 r 是正则化因子,则目标函数为

$$f = \sum_{S_t} \ln(1 + \exp(\|\mathbf{P}^T \mathbf{d}_t^{\text{pos}}\|^2 - \|\mathbf{P}^T \mathbf{d}_t^{\text{neg}}\|^2)) + r\|\mathbf{P}\|^2. \quad (6)$$

2 样本距离中心化与迭代策略

2.1 样本距离中心化

考虑到在建立正反例时,由于每一个样本与其他样本之间的正例特征距离数目(即为同一个行人不同数据集的特征距离数目,如图 1 中的实线所示)远少于反例特征距离数目(即为不同行人的不同特征距离数目,如图 1 中的粗虚线所示),这就带来了类别不平衡的问题,于是在度量学习时会出现反例样本训练过拟合或者正例样本欠拟合的情况,可能造成学习出来的度量矩阵并不能提高识别率.因此,对样本的特征向量进行适当的调整,尽量缓解类别严重不平衡的情况显得非常重要.本文主要利用类别不平衡问题常用的 3 个解决方法之一——欠采样的方式,将距离中心化,在求反例样本之间的特征

距离数目时,将每一组正例样本的特征值平均化(如图1中黑点所示),用特征中心值距离来代替原先的反例特征距离数目,这样可以减少较多的较为相近的反例特征距离数目(如图1中点虚线所示),同时也有利于缓解过拟合的风险.即原本训练特征集为 $\{S_l = (\mathbf{d}_l^{\text{pos}}, \mathbf{d}_l^{\text{neg}}); t=1, 2, \dots, m\}$, 距离中心化后的特征训练集为 $\{S_l = (\mathbf{d}_l^{\text{pos}}, \bar{\mathbf{d}}_l^{\text{neg}}); t=1, 2, \dots, m\}$, 其中 $\mathbf{d}_l^{\text{pos}} = \text{dis}(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{x}_i, \mathbf{x}_j$ 为同一行人的不同图像的特征向量,而 $\bar{\mathbf{d}}_l^{\text{neg}} = \text{dis}(\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_k)$, $\bar{\mathbf{x}}_i, \bar{\mathbf{x}}_k$ 为不同行人图像的平均特征向量,即 $\bar{\mathbf{x}}_i = \frac{1}{n_i} \sum_{i=1}^{n_i} \mathbf{x}_i, \bar{\mathbf{x}}_k = \frac{1}{n_k} \sum_{k=1}^{n_k} \mathbf{x}_k$, n_i, n_k 是目标样本集 i, k 中同一个行人图像的数目.

当同一目标群体里含有较多的样本时,本文针对其提出了局部距离中心化的思想,即将每一个目标群体中的较多样本分为若干个集合,对每个集合依次利用以上的距离中心化的方法.而求正例样本特征距离仍利用原先的策略来求解,这样在一定程度上能够缓解类别不平衡的问题,而且不会损失太多样本.

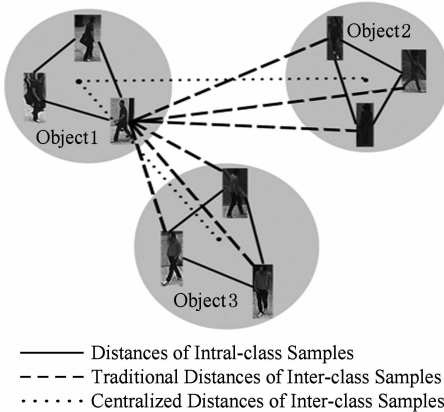


Fig. 1 Samples distance centralization (Dataset 3DPeS)
 图1 样本距离中心化(3DPeS数据集)

2.2 投影向量学习算法

采用距离中心化策略后,可以获得数量较为平衡的正反例样本特征分布,以这些特征分布,利用特定的优化算法对目标函数进行极小化搜索,以此来学习在这些特征分布下的投影向量,在未达到总体目标精度之前,再利用迭代策略(即式(7))来更新得到一组新的特征分布,然后再学习一组新的投影向量,直到达到总体目标精度为止.通常在机器学习中的较多的优化算法是梯度下降法,梯度下降法是

用负梯度方向为搜索方向的,而梯度下降法有越接近目标值步长越小、前进越慢等缺点,且直线搜索时可能会产生一些问题.共轭梯度法具有存储量小、步收敛性、稳定性高的特点,而且不需要任何外来参数,因此本文提出了基于共轭梯度法的投影向量学习方法.

本文采用共轭梯度法.首先需要给定初始搜索点 \mathbf{p}_l^0 和共轭梯度法的优化误差 ϵ_g , 通过 \mathbf{p}_l^0 计算得到目标函数的梯度,再根据 PRP 公式计算共轭方向,然后再利用一维精确搜索来确定搜索步长,如此优化直到收敛.优化到第 k 步时的目标函数为

$$f(\mathbf{p}_l^k) = \sum_{\mathbf{d}_l \in S^l} \ln(1 + \exp(\|(\mathbf{p}_l^k)^T \mathbf{d}_l^{\text{pos}}\|^2 - \|(\mathbf{p}_l^k)^T \mathbf{d}_l^{\text{neg}}\|^2)) + r \|\mathbf{p}_l^k\|^2, \quad (7)$$

则此时的目标函数的梯度 \mathbf{g}_l 为

$$\frac{\partial f}{\partial \mathbf{p}_l} = \sum_{\mathbf{d}_l \in S^l} \frac{2\exp(\|(\mathbf{p}_l^k)^T \mathbf{d}_l^{\text{pos}}\|^2 - \|(\mathbf{p}_l^k)^T \mathbf{d}_l^{\text{neg}}\|^2)}{1 + \exp(\|(\mathbf{p}_l^k)^T \mathbf{d}_l^{\text{pos}}\|^2 - \|(\mathbf{p}_l^k)^T \mathbf{d}_l^{\text{neg}}\|^2)} \times (\mathbf{d}_l^{\text{pos}} (\mathbf{d}_l^{\text{pos}})^T - \mathbf{d}_l^{\text{neg}} (\mathbf{d}_l^{\text{neg}})^T) \mathbf{p}_l^k + 2r \mathbf{p}_l^k. \quad (8)$$

而第 k 步迭代后的投影向量为

$$\mathbf{p}_l^{k+1} = \mathbf{p}_l^k + \alpha_k \mathbf{q}_k, \quad (9)$$

其中,步长 α_k 是 $f(\mathbf{p}_l^k + \alpha \mathbf{q}_k)$ 通过一维精确搜索求得, \mathbf{q}_k 为第 k 步投影向量的搜索方向:

$$\mathbf{q}_k = -\mathbf{g}_k + \beta_{k-1} \mathbf{q}_{k-1}, \quad (10)$$

$$\beta_{k-1} = \begin{cases} \frac{\mathbf{g}_k^T (\mathbf{g}_k - \mathbf{g}_{k-1})}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}}, & k > 1, \\ 0, & k = 1. \end{cases} \quad (11)$$

式(11)为 PRP 公式,当 $|f(\mathbf{p}_l^k) - f(\mathbf{p}_l^{k-1})| < \epsilon_g$ 时,停止迭代,即已经达到目标精度,此时的 \mathbf{p}_l^k 作为第 l 步迭代得到的投影向量 \mathbf{p}_l .

这里设第 l 步迭代的投影向量 \mathbf{p}_l^0 为

$$\mathbf{p}_l^0 = \frac{1}{\|S_{\text{pos}}^l\|} \sum_{S_{\text{pos}}^l} \mathbf{d}_i^{\text{pos}} - \frac{1}{\|S_{\text{neg}}^l\|} \sum_{S_{\text{neg}}^l} \mathbf{d}_i^{\text{neg}}, \quad (12)$$

其中, $\|S_{\text{pos}}^l\|$ 表示第 l 次更新后的训练集中的正例数目, $\|S_{\text{neg}}^l\|$ 亦如此,这里的 S 为第 l 步迭代更新后的新的训练集.

2.3 特征空间的迭代更新策略

本文通过迭代更新得到新的特征距离集合(即 S),可以得到一组新的特征分布,通过新的特征分布来学习得到一组近似正交的投影向量 \mathbf{p}_i . 这样,在满足一定精度要求的前提下可以大大减少运算复杂度,用少量的列向量 \mathbf{p}_i 来构造具有很好判别性的度量矩阵 \mathbf{M} ,通过这种方式可以减少数据冗余度,即降噪的效果.

假设在 l 次迭代后,已经通过学习得到了一组

投影向量 $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_l$, 与文献[28]使用的策略类似, 本文在其基础上通过增加扰动项, 利用下面的迭代策略来更新 $\mathbf{d}_i^{s,l+1}$, 其中 $s \in \{\text{pos}, \text{neg}\}, t \in \{1, 2, \dots, |S|\}$.

$$\mathbf{d}_i^{s,l+1} = \mathbf{d}_i^{s,l} - \frac{\mathbf{p}_l \mathbf{p}_l^T}{\|\mathbf{p}_l + \mathbf{u}\|^2} \mathbf{d}_i^{s,l}. \quad (13)$$

假设初始 $\mathbf{p}_0 = 0$, 则当 $l > 0$ 时, 使用式(13)来更新特征距离集合, 而当 $l = 0$ 时, 即直接利用初始构造的特征距离集合(即 $\mathbf{d}_i^s, s \in \{\text{pos}, \text{neg}\}, t \in \{1, 2, \dots, |S|\}$)来学习投影向量 \mathbf{p}_1 . 根据式(9)、式(12)可知, \mathbf{p}_l 在 $\{\mathbf{d}_i^{\text{pos},l}\} \cup \{\mathbf{d}_i^{\text{neg},l}\}$ 的生成子空间中, 即 $\mathbf{p}_l \in \text{span}\{\mathbf{d}_i^{s,l}\}$, 其中 $s \in \{\text{pos}, \text{neg}\}$, 其中 $i \in \{1, 2, \dots, |S|\}$ 而由式(13)可知, $\mathbf{p}_j^T \mathbf{d}_i^{s,j+1} \approx 0, j = 1, 2, \dots, l-1$, 而 $\mathbf{p}_l \in \text{span}\{\mathbf{d}_i^{s,l}\}$, 并且有 $\text{span}\{\mathbf{d}_i^{s,l}\} \subseteq \text{span}\{\mathbf{d}_i^{s,l-1}\} \subseteq \dots \subseteq \text{span}\{\mathbf{d}_i^{s,0}\}$. 因此, \mathbf{p}_l 与 $\mathbf{p}_j, j = 1, 2, \dots, l-1$ 近似正交. 由于每次迭代得到的投影向量对应一个投影空间, 而这些投影空间的关系并不是完全孤立的, 因此本文在式(13)中加入数值较小的扰动项 \mathbf{u} , 使得每一次更新迭代后的投影向量近似正交, 保留每个投影空间的一些联系, 使得投影更具有实际意义.

2.4 具体操作步骤

本文在提取得到行人图像的特征之后, 利用2.1节所提出的局部距离中心化的方法构建训练集, 然后在该训练集上利用2.2节与2.3节的方法来优化投影向量, 最后将这些投影向量拼接成投影矩阵并输出.

算法1. 投影矩阵学习算法

输入: 行人特征集 X , 正则化因子 r , 扰动因子 \mathbf{u} , 目标精度 ϵ_0 , 投影向量精度 ϵ_g ;

输出: 投影矩阵 $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_l]$;

- ① 用 X 构建训练集 S ;
- ② $\mathbf{P} = \mathbf{p}_0 = 0, l = 0$;
- ③ 用式(6)计算 f^0 ;
- ④ While True
- ⑤ $l = l + 1$;
- ⑥ 用式(13)更新 S^l ;
- ⑦ 用式(12)计算 \mathbf{p}_l^0 ;
- ⑧ $k = 0$;
- ⑨ 用式(7)计算 $f(\mathbf{p}_l^0)$;
- ⑩ While True
- ⑪ $k = k + 1$;
- ⑫ 用式(8), (10), (11)分别计算 $\mathbf{g}_l, \mathbf{q}_k, \beta_{k-1}$;
- ⑬ 用 \mathbf{q}_k 更新 \mathbf{p}_l^k ;
- ⑭ 用式(7)计算 $f(\mathbf{p}_l^k)$;

- ⑮ If $|f(\mathbf{p}_l^k) - f(\mathbf{p}_l^{k-1})| \leq \epsilon_g$
- ⑯ Break;
- ⑰ End If
- ⑱ End While
- ⑲ 令 $\mathbf{P} = [\mathbf{P}, \mathbf{p}_l]$;
- ⑳ 用式(6)计算 f^l ;
- ㉑ If $|f^l - f^{l-1}| \leq \epsilon_0$
- ㉒ Break;
- ㉓ End If
- ㉔ End While

算法1中各个符号及字母的含义与前文含义相同, 这里不再具体解释.

3 实验与分析

3.1 数据集与评价指标

本文采用3种常用公开数据集 VIPeR^[29], i-LIDs(MCTS)^[30], 3DPeS^[31], 如图2所示, 每组中上下2行代表同一个行人. 其中, VIPeR数据集分辨率被规范化为 128×64 像素, 由632个行人组成一共1024张行人图像, 每个行人都有2张不同的图像, 这2张图像是从不同摄像机不同角度拍摄得到的, 同一个人在不同摄像机下的图像姿势有较大的变化. i-LIDs数据集分辨率也是 128×64 像素, 该数据集是在机场入境大厅拍摄得到, 一共有119个行人组成的476个行人图像, 平均每个行人有4张行



(a) VIPeR

(b) i-LIDs



(c) 3DPeS

Fig. 2 Pedestrian image lists of public datasets

图2 公共数据集行人图像对

人图像,由于行人较多,因此行人图像中不免有些遮挡以及角度等问题. 3DPeS 数据集由 8 个不同的监控摄像机在不同时间拍摄得到的,该数据集的分辨率并不一样,由 204 个行人共 1012 张行人图像组成,由于该数据集有时间跨度,因此每个行人数据集光照变化大. 本文分别将 VIPeR 数据集、i-LIDs 数据集和 3DPeS 数据集中的 200 个行人、59 个行人和 134 个行人作为训练集,余下的用作测试集.

对于评价指标,本文采用行人重识别问题常用的累计匹配特性曲线(cumulative match characteristic, CMC)的方式. 累计匹配特性曲线的横坐标是排名(rank),纵坐标是匹配率(matching rate),排名为 r 时的匹配率表示排序后前 r 个行人中匹配正确的概率. 其中排名较靠前的匹配率大小更具有实际意义,因为在实际应用中,不管用什么方法来解决行人重识别问题,其结果都是返回若干张与该行人最相似的图像,这几张图像即排名靠前的几张图像,因此当排名靠前的匹配率较大时,能够在很大程度上使得返回的结果中一定有该行人的图像,从而提高行人重识别的效率.

为了测试本文算法的计算效率,本文的评价指标除了累计匹配特性曲线,还对本文算法与其他常用的基于度量学习的算法的时间花销进行对比,包括训练时间与测试时间.

3.2 特征表示

本文对行人图像分别提取 RGB, YCbCr, HSV, Lab, YIQ, Gabor 6 种特征,图 3 给出了提取一张行人图像的特征描述的示例. 其中前 5 种特征为颜色空间特征,提取的是直方图特征,即统计特征,RGB, YCbCr 分别提取全部 3 组颜色特征,而 HSV 特征只提取色调(H)、饱和度(S)特征,Lab 特征和 YIQ 去除像素的亮度特征(即 L 分量与 Y 分量),这些待提取的特征全部分为 16 维直方图统计特征. 而 Gabor 特征是一种纹理特征,根据不同波长、方向、相位偏移、空间纵横比、带宽等分别取 16 组不同的

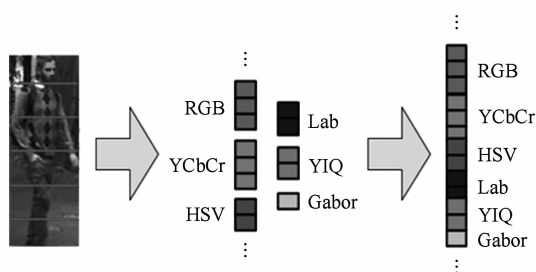


Fig. 3 Feature extraction of pedestrian images

图 3 行人图像的特征提取

Gabor 滤波器,而每个滤波器同样再提取 16 维直方图统计特征. 对于每一张行人图像,将其在水平方向平均分为 6 个水平条带. 因此每个水平条带中有 28 个特征通道,每个通道又被表示为 16 维直方向量,因此每幅图像在特征空间中被表示为 2 688 维特征向量.

3.3 对比分析

常用的基于度量学习的行人重识别的解决方法有 ITML^[24] (information theoretic metric learning), LMNN^[21] (large margin nearest neighbor), KISSME^[25] (keep it simple and straightforward metric), PRDC^[19] (probabilistic relative distance comparison), LFDA^[20] (local fisher discriminant analysis), CVDCA^[26] (cross-view discriminant component analysis), RMLLC^[27] (relevance metric learning method with listwise constraints) 等方法. 本节重点比较了这 7 种方法与本文所提出的方法,利用 CMC 曲线中的 Rank 1, Rank 5, Rank 10, Rank 20 所对应的 Matching Rate 来比较这 7 种方法与本文方法的优劣. 其中,ITML, LMNN 以及 KISSME 三种方法是基于学习高维的马氏度量矩阵的算法,这样不仅运算复杂度高,而且会带来存储量较大等问题;PRDC 方法与本文方法类似,其主要以概率相对距离比较为方法,学习出较好的马氏度量矩阵;LFDA 方法主要先利用主成分分析进行降维,并对不同的特征赋予不同的权重,其主要思想仍然是度量学习;CVDCA 方法是基于非对称投影的方式,综合考虑了不同摄像机的属性不同的情况来学习出不同的投影矩阵,但其只适用于训练集中只有 2 个摄像场景的情况;RMLLC 方法是基于相似对的方式,可以很好地减少相似但不同行人匹配出错的情况. 这几种方法都未考虑类别不平衡的情况,而且都是以学习投影矩阵为主要方法,所以这些方法仍然会有存储量大、耗时长等问题.

本文实验在 MATLAB 7. 11. 0 平台下完成,实验环境为 CPU Intel Core™ i5-4460T 1. 90 GHz,内存 8 GB. 在本文方法中,正则化因子 $r=1$ 、扰动因子 $u=10^{-3}$ 能够获得较好的结果. 首先利用本文方法与其他方法在不同的数据集上分别进行比较. 本文对 VIPeR, i-LIDs, 3DPeS 分别利用局部距离中心化,在求反例样本距离时,需中心化的同一样本的不同图像分别按 2 个行人、2 个行人、3 个行人作为一个集合分别进行局部距离中心化. 在数据集 VIPeR, i-LIDs, 3DPeS 上的实验结果分别如表 1~3 所示.

Table 1 Performance Comparison Among Different Methods on Dataset VIPeR ($P=432$)**表 1 各种方法在数据集 VIPeR 上的结果比较(测试集规模为 432)**

Methods	Rank 1	Rank 5	Rank 10	Rank 20	Training Time/s	Testing Time/s
Ours	0.2622	0.5662	0.7180	0.8232	475	3
CVDCA	0.1668	0.4100	0.5411	0.6773	1378	16
RMLLC	0.1522	0.3580	0.4737	0.5952	2113	15
LFDA	0.1145	0.2972	0.4098	0.5603	2214	20
PRDC	0.1036	0.2829	0.3762	0.5414	1205	12
KISSME	0.0721	0.2116	0.3068	0.4446	3002	18
ITML	0.0255	0.1052	0.1618	0.2390	8978	43
LMNN	0.0100	0.0367	0.0713	0.1291	3658	26

Table 2 Performance Comparison Among Different Methods on Dataset i-LIDs ($P=60$)**表 2 各种方法在数据集 i-LIDs 上的结果比较(测试集规模为 60)**

Methods	Rank 1	Rank 5	Rank 10	Rank 20	Training Time/s	Testing Time/s
Ours	0.4084	0.7735	0.8519	0.9056	177	2
CVDCA	0.3498	0.5943	0.7158	0.7910	1082	8
RMLLC	0.2622	0.5262	0.6480	0.7532	1503	10
LFDA	0.1036	0.2829	0.4462	0.6414	1789	11
PRDC	0.2183	0.3937	0.5112	0.6543	910	7
KISSME	0.1458	0.3414	0.4637	0.5996	2579	18
ITML	0.0721	0.1816	0.2968	0.4795	4688	20
LMNN	0.0445	0.1597	0.2280	0.3849	3236	25

Table 3 Performance Comparison Among Different Methods on Dataset 3DPeS ($P=70$)**表 3 各种方法在数据集 3DPeS 上的结果比较(测试集规模为 70)**

Methods	Rank 1	Rank 5	Rank 10	Rank 20	Training Time/s	Testing Time/s
Ours	0.3980	0.8660	0.9440	0.9920	419	4
CVDCA	0.3077	0.7000	0.8838	0.9640	1189	10
RMLLC	0.3020	0.6220	0.8180	0.9200	1602	12
LFDA	0.1731	0.4385	0.6615	0.8442	1836	16
PRDC	0.2945	0.5119	0.7121	0.8748	1371	13
KISSME	0.0800	0.3040	0.5376	0.8264	2836	19
ITML	0.0790	0.2700	0.4355	0.7674	5422	24
LMNN	0.0931	0.2236	0.3591	0.5381	3105	22

对于数据集 VIPeR,每个行人只有 2 张行人图像,从图 2(a)可以看出,该数据集给行人重识别造成困难的因素是拍摄角度等问题,而且每个行人的图像较少,常用的行人重识别方法在该数据集上的表现一直不是很好.但通过表 1 可以看出,本文方法在该数据集上的表现比其他常用的行人重识别方法至少高出 10 个百分点.除此之外,由于本文方法采

用了距离中心化的思想,明显减少了运算量,而且由于共轭梯度法具有快速收敛的特点,大大减少了运算复杂度,从表 1 的最后 2 列可以看出,本文方法的训练时间与测试时间明显比其他方法少.因此,本文算法在数据集 VIPeR 上的运算效率有了很大的提高.

而对于数据集 i-LIDs,通过图 2 可以看出,该数

据集中的每个行人由于有较多部分的遮挡,还有一些角度问题,这些问题给行人重识别问题带来了较大的困难,以往常用的方法在该数据集上的表现都很一般,而通过表 2 可以看出,本文方法比其他表现最好的方法仍然高出大约 6 个百分点.而且表 2 中最后 2 列也可以看出,本文方法的时间花销也比其他方法要少.此外,由于 i-LIDs 数据集属于小规模数据集,对于一般常用的算法会出现过拟合的问题,使得识别效果下降,但本文方法的识别效果仍然比其他方法要好,因此本文方法缓解了基于学习的算法的过拟合问题.

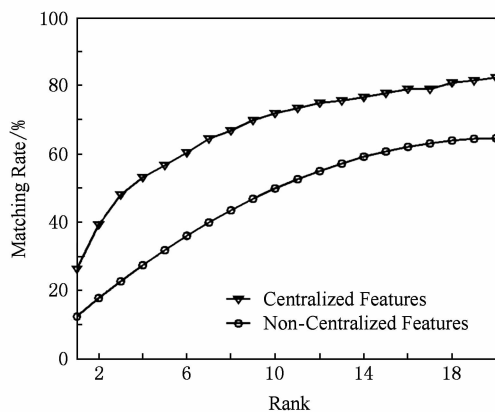
在数据集 3DPeS 上,其对行人重识别造成困难的因素是不同摄像机所摄取的时间差,因此同一行人的不同图像会有明显的光照区别.除此之外,该数据集对于同一行人的不同图像也有角度姿势的不同等问题.尽管如此,本文方法在该数据集上的表现仍然高于其他常用方法,比其他方法至少高 9 个百分点.算法的训练时间与测试时间也明显少于其他方法.因此本文方法在该数据集上的运算效率也有了较大的提高.

其次,本文还在同样的度量学习方法上,将样本距离中心化后的识别效果与样本距离未中心化的识别效果进行了对比,2 种不同策略在 3 个数据集上的 CMC 曲线如图 4 所示,可以看出,由于样本距离中心化后能够很好地缓和过拟合的风险,将样本距离中心化后的行人重识别的效果明显要优于未中心化后的识别效果.可以看出,利用样本距离中心化后在排名靠前的匹配率明显高于样本距离未中心化的算法,虽然排名靠后的匹配率增长幅度逐渐降低,但由前面叙述可知,样本距离中心化后效果的提升更具有实际意义.而对于数据集中同一目标样本数量较多的数据集 3DPeS 而言,如果将同一行人的所有不同图像的特征值全部中心化后会损失较多的类间距离,因此会使得识别效果反而有所下降,所以利用局部距离中心化后效果提升的会比较明显,其与未利用局部距离中心化的效果对比如表 4 所示:

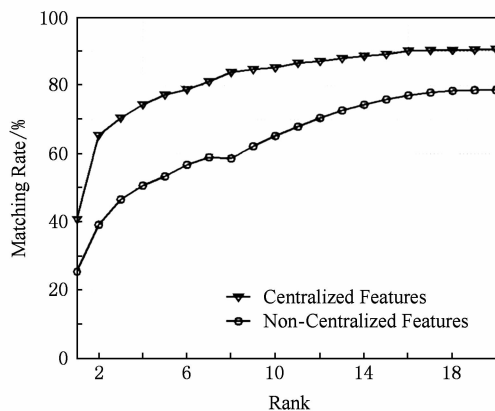
Table 4 Performance Comparison Between Local and Global Distance Centralization (3DPeS)

表 4 局部与全局距离中心化的效果对比 (3DPeS)

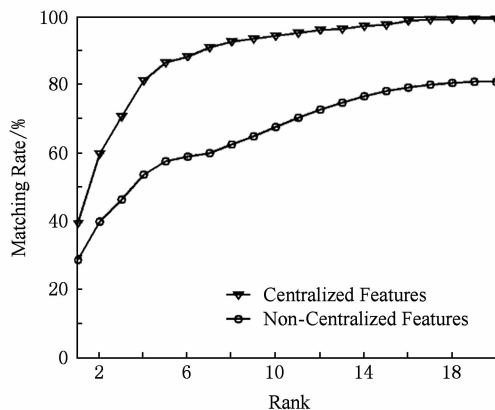
Methods	Rank 1	Rank 5	Rank 10	Rank 20
Local Distance Centralized	0.3980	0.8660	0.9440	0.9920
Global Distance Centralized	0.3565	0.7863	0.8605	0.9396



(a) VIPeR



(b) i-LIDs



(c) 3DPeS

Fig. 4 Performance comparison between feature distance centralization and non-centralization
图 4 特征中心化与未中心化的效果比较

4 结 论

针对以往的方法中未考虑样本类别不平衡问题,本文提出了基于样本距离中心化的,并使用迭代策略的度量学习的行人重识别方法.相比其他基于度量学习方法,本文方法采取的样本距离中心化能

够很好的去除较多的冗余的类间距离,使得类内距离与类间距离的数量趋于平衡,同时为了不使反例样本损失太多,本文在同样本因此在训练的时候能够很好的缓解因类别不平衡导致的过拟合的风险. 本文的采用的迭代策略能够很好地减少运算以及存储复杂度,能使每次优化得到的投影向量之间都是近似正交的. 而且利用特征值分解的方法与主成分分析(PCA)方法类似,除了能够起到降维的作用,还能达到去噪的效果. 此外,本文运用共轭梯度法来训练投影向量能够快速的收敛到目标精度,由于共轭梯度法的特有性质,使得训练时所需的运算与存储复杂度大大降低. 通过实验结果发现,本文提出的方法比其他基于度量学习的方法也更具有泛化性,而且训练时间也比其他基于度量学习的方法要短.

参 考 文 献

- [1] Chen Puqiang, Guo Lijun, Zhang Rong, et al. Patch matching with global spatial constraints for person re-identification [J]. *Journal of Computer Research and Development*, 2015, 52(3): 596-605 (in Chinese)
(陈普强, 郭立君, 张荣, 等. 基于全局空间约束匹配的目标人体识别[J]. *计算机研究与发展*, 2015, 52(3): 596-605)
- [2] Bazzani L, Cristani M, Perina A, et al. Multiple-shot person re-identification by chromatic and epitomic analyses [J]. *Pattern Recognition Letters*, 2012, 33(7): 898-903
- [3] Ma Bingpeng, Su Yu, Jurie F. Local descriptors encoded by fisher vectors for person re-identification [C] //Proc of the 12th European Conf on Computer Vision. Berlin: Springer, 2012, 7583: 413-422
- [4] Zeng Mingyong, Wu Zemin, Tian Chang, et al. Fusing appearance statistical features for person re-identification [J]. *Journal of Electronics & Information Technology*, 2014, 36(8): 1844-1851 (in Chinese)
(曾明勇, 吴泽民, 田畅, 等. 基于外观统计特征融合的人体目标再识别[J]. *电子与信息学报*, 2014, 36(8): 1844-1851)
- [5] Ma Bingpeng, Su Yu, Jurie F. Covariance descriptor based on bio-inspired features for person re-identification and face verification [J]. *Image & Vision Computing*, 2014, 32(6/7): 379-390
- [6] Kviatkovsky I, Adam A, Rivlin E. Color invariants for person reidentification [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2013, 35(7): 1622-1634
- [7] Fan Caixia, Zhu Hong, Lin Guangfeng, et al. Person re-identification based on multi-features [J]. *Journal of Image and Graphics*, 2013, 18(6): 711-717 (in Chinese)
(范彩霞, 朱虹, 蔺广逢, 等. 多特征融合的人体目标再识别[J]. *中国图象图形学报*, 2013, 18(6): 711-717)
- [8] Wu Ziyang, Li Yang, Radke R J. Viewpoint invariant human reidentification in camera networks using pose priors and subject discriminative features [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2014, 37(5): 1095-1108
- [9] Li Wei, Wang Xiaogang. Locally aligned feature transforms across views [C] //Proc of the 2013 IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2013. 4321-4328
- [10] Zhao Rui, Ouyang Wanli, Wang Xiaogang. Learning mid-level filters for person re-identification [C] //Proc of the 2014 IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2014. 144-151
- [11] Qi Meibin, Tan Shengshun, Wang Yunxia, et al. Multi-feature subspace and kernel learning for person re-identification [J]. *Acta Automatica Sinica*, 2016, 42(2): 299-308 (in Chinese)
(齐美彬, 檀胜顺, 王运侠, 等. 基于多特征子空间与核学习的行人再识别[J]. *自动化学报*, 2016, 42(2): 299-308)
- [12] Kostinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints [C] //Proc of the 2012 IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2012: 2288-2295
- [13] Li Zhen, Chang Shiyu, Liang Feng, et al. Learning locally adaptive decision functions for person verification [C] //Proc of the 2013 IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2013: 3610-3617
- [14] Du Yuning, Ai Haizhou. A statistical inference approach for person re-identification [J]. *Journal of Electronics and Information Technology*, 2014, 36(7): 1612-1618 (in Chinese)
(杜宇宁, 艾海舟. 基于统计推断的行人再识别算法[J]. *电子与信息学报*, 2014, 36(7): 1612-1618)
- [15] Ma Lianyang, Yang Xiaokang, Tao Dacheng. Person re-identification over camera networks using multi-task distance metric learning [J]. *IEEE Trans on Image Process*, 2014, 23(8): 3656-3670
- [16] Lisanti G, Masi I, Bagdanov A D, et al. Person re-identification by iterative re-weighted sparse ranking [J]. *IEEE Trans on Pattern Analysis & Machine Intelligence*, 2015, 37(8), 1629-1642
- [17] Nguyen H V, Bai L. Cosine similarity metric learning for face verification [C] //Proc of the 10th Asian Conference on Computer Vision (ACCV). Berlin: Springer, 2010: 709-720
- [18] Loy C C, Liu Chunxiao, Gong Shaogang. Person re-identification by manifold ranking [C] //Proc of the 20th IEEE Int Conf on Image Processing (ICIP). Piscataway, NJ: IEEE, 2013: 3567-3571
- [19] Weinberger K Q, Saul L K. Distance metric learning for large margin nearest neighbor classification [J]. *Journal of Machine Learning Research*, 2009, 10(2): 207-244

- [20] Weinberger K Q, Saul L K. Fast solvers and efficient implementations for distance metric learning [C] //Proc of the 25th Int Confe on Machine Learning. New York: ACM, 2008: 1160-1167
- [21] Dikmen M, Akbas E, Huang T S, et al. Pedestrian recognition with a learned metric [C] //Proc of the 10th Asian Conf on Computer Vision. Queenstown, Berlin: Springer, 2010, 6495: 501-512
- [22] Davis J V, Kulis B, Jain P, et al. Information theoretic metric learning [C] //Proc of the 24th Int Conf on Machine Learning. New York: ACM, 2007: 209-216
- [23] Zheng Weishi, Gong Shaogang, Xiang Tao. Person re-identification by probabilistic relative distance comparison [C] //Proc of the 2011 IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2011: 649-656
- [24] Kostinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints [C] //Proc of the 2012 Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2012: 2288-2295
- [25] Pedagadi S, Orwell J, Velastin S. Local fisher discriminant analysis for pedestrian reidentification [C] //Proc of the 2013 IEEE Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2013: 3318-3325
- [26] Chen Jiaxin, Zhang Zhaoxiang, Wang Yunhong. Relevance metric learning for person reidentification by exploiting listwise similarities [J]. IEEE Trans on Image Processing, 2015, 24(12): 4741-4755
- [27] Chen Yingcong, Zheng Weishi, Lai Jianhua, et al. An asymmetric distance model for cross-view feature mapping in person re-identification [J]. 2016 [2017-03-20]. <http://ieeexplore.ieee.org/document/7373616/>
- [28] Zheng Weishi, Gong Shaogang, Xiang Tao. Reidentification by relative distance comparison [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2013, 35(3): 653-668
- [29] Gray D, Tao Hai. Viewpoint invariant pedestrian recognition with an ensemble of localized features [C] //Proc of the 10th European Conf on Computer Vision. Berlin: Springer, 2008, 5302: 262-275
- [30] Zheng Weishi, Gong Shaogang, Xiang Tao. Associating groups of people [C] //Proc of the 20th British Machine Vision Conf (BMVC). Piscataway, NJ: IEEE, 2009: 1-11
- [31] Baltieri D, Vezzani R, Cucchiara R. 3DPeS: 3D people dataset for surveillance and forensics [C] //Proc of Joint ACM Workshop on Human Gesture and Behavior Understanding. New York: ACM, 2011: 59-64



Ding Zongyuan, born in 1991. Master candidate at Changzhou University. Member of CCF. His main research interests include image processing and pattern recognition.



Wang Hongyuan, born in 1960. Received his PhD degree from Nanjing University of Science and Technology in 2004. Professor and master supervisor at Changzhou University. His main research interests include image processing, artificial intelligence and pattern recognition.



Chen Fuhua, born in 1966. Received his PhD degree from Nanjing University of Science and Technology in 2003. Associate professor at Westliberty University. His main research interests include variation image segmentation and inverse problems.



Ni Tongguang, born in 1978. Received his PhD degree from Jiangnan University in 2015. Lecturer at Changzhou University. His main research interests include pattern recognition, intelligent computation and their application.