

基于二次相似度函数学习的行人再识别

杜宇宁 艾海舟

(清华大学计算机科学与技术系 北京 100084)

摘 要 行人再识别是一个有着非常重要现实意义的研究问题,它可以应用于刑事侦查、在公共场所中寻找丢失的小孩、个人相册管理以及电子商务等领域.同时由于光照、视角、人的姿态以及背景的变化,同一个人的表观在不同的监控视频中往往变化很大,解决行人再识别问题也非常有挑战性.在设计行人再识别算法时,给定行人图像的特征,考虑到不同的特征分量具有不同的区分能力,学习合适的相似度度量非常重要.度量学习是一类学习相似度度量的主流算法,这些算法通过学习一个马氏距离相似度函数(Mahalanobis Similarity Function, MSF)来估计一对行人图像的相似度.然而 MSF 只与特征差分空间有关,忽略了一对图像中每个个体的表观特征,对于同一个人在不同场景中很大的表观变化的捕捉能力有限.为了加强相似度函数与每个个体的表观特征的联系,该文提出通过学习一个二次相似度函数(Quadratic Similarity Function, QSF),来估计一对行人图像的相似度. QSF 是 MSF 的泛化形式,不但描述了一对行人图像的互相关关系,而且关联了一对行人图像的自相关关系,可以更好地捕捉同一个人在不同监控视频中很大的表观变化.为了学习 QSF,该文分别从分类和排序的角度出发,设计两种不同的优化目标,提出了两种不同的学习 QSF 的算法.由行人再识别的公共数据集 VIPeR 和 CUHK 的实验表明,这两种不同的算法都可以学习到有效的 QSF,识别性能优于已有的行人再识别算法.

关键词 视频监控;行人再识别;行人分析;度量学习;二次相似度函数学习;机器学习

中图法分类号 TP391 **DOI 号** 10.11897/SP.J.1016.2016.01639

Learning Quadratic Similarity Function for Pedestrian Re-Identification

DU Yu-Ning AI Hai-Zhou

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084)

Abstract Pedestrian re-identification is a valuable problem that has enormous potential for practical applications like criminals monitoring and investigation, lost children search in the public area, personal photo album management, e-business, to name a few. Meanwhile, it is also very challenging since the appearance of the same person between camera views changes dramatically caused by illumination, viewpoint, pose, and background variations. Given a certain pedestrian feature space, because of the different discriminative ability of different feature components, learning a proper similarity is important for pedestrian re-identification. A dominant algorithm to learn a similarity is the metric learning that learns a Mahalanobis Similarity Function (MSF) to estimate the similarity of a pair of pedestrians. However, because MSF only projects a pair of pedestrians into the feature difference space and ignores the appearance of each individual, it is inadequate for modeling complex relationships of appearance variations from different cameras. In this paper, we propose to learn a Quadratic Similarity Function (QSF) that greatly strengthens the modeling ability of the similarity function. QSF is a generalization of MSF, and it not only represents the cross correlation relationship of a pair of pedestrians, but also describes the auto

correlation relationship. So QSF is better to tackle large appearance variations of the same pedestrians than MSF. To learn a QSF, inspired from the classification perspective and the ranking perspective, we design two algorithms with different optimizing object functions. Experiments on the VIPeR dataset and the CUHK campus dataset show that both algorithms are effective to learn a QSF and their performance outperforms the previous algorithms.

Keywords visual surveillance; pedestrian re-identification; pedestrian analysis; metric learning; quadratic similarity function learning; machine learning

1 引 言

行人再识别是指在已有的多个空间位置上不交叠的监控视频中,根据目标行人图像,检索目标行人是否出现在这些视频中.该技术可以广泛应用到多个领域,例如刑事侦查、个人相册管理、在公共区域寻找丢失的老人和小孩、电子商务等.同时,由于光照、视角、人的姿态以及背景的变化,在不同的监控视频中,同一个人的表观变化往往很大,因此解决行人再识别问题非常有挑战性.正因为此问题广泛的应用前景和面临的困难,近年来有大量的研究工作关注此问题^[1-15].为了便于叙述,如果一对行人图像来源于同一个人,本文简称为相关行人图像对,否则简称为不相关行人图像对.

文献[16]综述了近年来关于行人再识别问题的研究进展.为了匹配同一个人在不同摄像机视场中的图像,一个典型的行人再识别算法由两步组成:(1)抽取待查询图像和行人图像库中图像的特征;(2)利用上述行人图像的特征,根据相似度度量准则,计算行人图像之间的相似度,从而从行人图像库中检索出目标行人.先前的行人再识别算法根据这两步的侧重点不同大致分为两类:无监督的算法^[1-4]和有监督的算法^[5-14].

无监督的算法侧重于图像特征的提取,然后利用标准的相似度度量(例如 1 范数、2 范数、巴氏距离等)度量一对行人图像的相似度.其基本想法是手动设计具有稳定性和区分性的特征,以适应同一个人在不同场景下的表观变化.文献[1]利用行人图像的对称性,提出对称的局部累积特征来处理姿态、视角和光照的变化.为了处理姿态的变化,文献[2]从人体的每个部件区域抽取特征.各个部件区域的定位则采用人体姿态估计中经典算法,图案结构算法(pictorial structure).文献[3]同样使用图案结构算法,通过图匹配计算相似度.文献[4]借助一个参考

的行人图像库来学习每个行人的显著性特征,进而处理姿态和视角变换的问题.因为在不同的实际场景中,同一个人的表观变化很大,因此手动设计鲁棒的特征往往很困难.而提取这些手动设计的特征时间代价也很大.此外,当给定行人图像的特征,考虑到不同的特征分量具有不同的区分能力,而标准的相似度度量等同对待所有的特征分量,因此标准的相似度度量并不是最优的相似度度量.

有监督的算法侧重于从简单的行人图像特征中,利用机器学习算法(例如 AdaBoost、支持向量机、度量学习等)学习相似度度量,从数据中挖掘有区分性的信息.相对于无监督的算法,即使使用简单的行人图像特征,这些算法也可以捕获同一个人在不同场景中很大的表观变化,并且估计一对行人图像相似度的时间代价很低.此外,这些算法也有助于进一步提高通过无监督的算法得到的行人图像特征的性能.因此,本文主要关注当给定行人图像特征时,如何更好地学习一对行人图像的相似度度量.

文献[5]利用 Adaboost 算法从一个特征池中学习最优的特征组合,进而定义一对行人图像之间的相似度.上述特征池由通过密集采样生成的水平条带的颜色和纹理特征组成.文献[6]则利用支持向量机,结合预先定义的具有一定区分性的特征向量学习相似度度量.上述方法也采用颜色和纹理特征,但与文献[5]不同,特征来源于均分的 6 个水平条带.近年来基于度量学习的算法受到普遍重视^[7-12],这些算法通过学习一个马氏距离相似度函数(Mahalanobis Similarity Function, MSF)来度量一对行人图像的相似度,使相关行人图像对获得尽可能高的相似度,而不相关行人图像对获得尽可能低的相似度.然而,以上所有算法学习的相似度函数只与特征差分空间有关,忽略了每个个体的表观特征,因此对于同一个人在不同场景中的很大的表观变化的捕捉能力有限.与以上算法不同,文献[13]将图像的空间分成不同的区域,对于每个区域学习各自的相似

度函数. 文献[14]首先估计每个图像的显著性区域, 然后学习与显著性区域相关的相似度函数. 这两种算法都验证了学习与每个个体相关的相似度函数有助于提高识别性能.

基于上述分析, 本文提出利用一个二次相似度函数 (Quadratic Similarity Function, QSF) 来估计一对行人图像的相似度, 进而捕捉不同摄像机视场下同一个人较大的表观变化. QSF 与 MSF 有两方面的不同: (1) 一对行人图像包含两种二阶关系: 自相关关系和互相关关系. 在 MSF 中, 描述这两种关系的自相关矩阵和互相关矩阵互为负矩阵 (详见第 2 节), 因此 MSF 只与特征差分空间有关. 而在 QSF 中, 上述两个矩阵完全不同, 因此 QSF 与每个个体的表观特征有关; (2) MSF 仅仅描述了一对行人图像的二阶关系. 但是 QSF 不仅描述了上述关系, 而且描述了一对行人图像的一阶关系, 加强了相似度函数与每个个体的表观特征的相关性. 因此, QSF 是局部自适应的, 可以更好地捕捉同一个人不同的场景中的很大的表观变化.

在已有的度量学习算法中, 学习相似度函数的优化策略大致分为两类: 一类是类似于文献[8-10, 12], 从分类的角度出发, 将相关行人图像对和不相关行人图像对看成是两个不同的类别, 优化目标是使任

意的相关行人图像对的相似度高于任意的不相关行人图像对的相似度, 如图 1(a) 所示. 图中 A1 和 A2 表示同一个人的不同图像, B1 和 B2, C1 和 C2 类似. 相关图像对 A1 和 A2 的相似度, 不但要高于与 A 有关联的不相关图像对, 例如 A1 和 B1, 而且需要高于与 A 没有关联的不相关图像对, 例如 C1 和 B2. 另一类是类似于文献[7], 从排序的角度出发, 给定一个待考察的行人图像, 只考虑与此图像有关联的相关行人图像对和不相关行人图像对, 使它们的差异尽可能大, 而不关注与此行人图像无关联的行人图像对, 如图 1(b) 所示. 相关图像对 A1 和 A2 的相似度, 只要求高于与 A 有关联的不相关图像对, 例如 A1 和 B2, 不再要求高于与 A 没有关联的不相关图像对, 例如 C1 和 B2. 两种策略各有利弊. 基于分类的策略考虑的行人图像对之间的关系比基于排序的策略复杂, 但是前者比后者更容易受到噪声影响. 为了学习 QSF, 本文分别从分类和排序的角度出发, 设计了两种不同的新的优化策略. 在行人再识别的公共数据集 VIPeR^[15] 和 CUHK^[12] 的实验表明, 这两种不同的优化策略都可以学习到有效的二次相似度函数, 识别性能优于已有的行人再识别算法. 而这两种优化策略的识别性能相当.

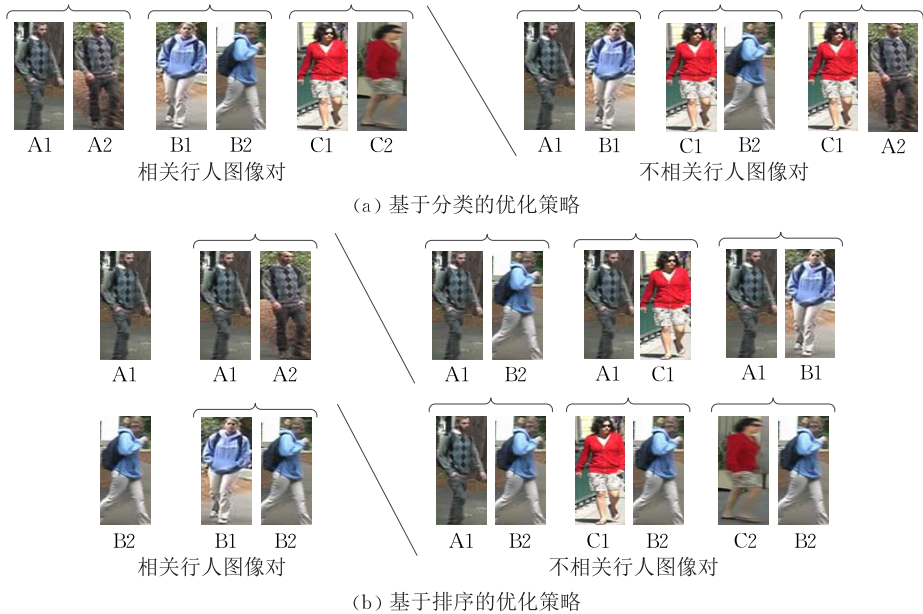


图 1 两种不同的学习相似度函数的策略说明

综上所述, 本文的贡献概括为两个方面: (1) 将行人再识别问题形式化为一个局部自适应的二次相似度函数 (Quadratic Similarity Function, QSF) 学习问题; (2) 分别从分类和排序的角度出发, 提出了

两种不同的优化策略学习二次相似度函数. 本文第 2 节介绍如何学习二次相似度函数; 第 3 节介绍基于 QSF 学习的行人再识别系统; 第 4 节介绍本文算法在公共数据集上的实验; 第 5 节总结全文.

2 二次相似度函数的学习

在给定的特征空间中学习样本之间的相似度是机器学习和数据挖掘领域中一个非常重要的基本问题. 许多算法需要一个合适的相似度度量来反映样本之间的重要关系, 进而提高算法性能, 例如 K 均值算法、最近邻分类器以及基于核函数的算法(例如支持向量机)等. 因此, 相似度度量的学习受到研究者的普遍关注, 其中包括 MLCC^[17]、ITML^[18] 以及 LMNN^[19] 等. 受这些算法的启发, 近年来, 行人再识别问题的研究者也尝试设计合适的相似度学习算法^[7-12], 更好地估计一对行人图像的相似度. 当给定一对可能来源于非交叠摄像机视场的行人图像(行人图像的特征向量为 $\mathbf{X}$ 和 $\mathbf{Y}$, $\mathbf{X}, \mathbf{Y} \in R^d$, d 是特征维度), 这些算法通过学习一个马氏距离相似度函数(Mahalanobis Similarity Function, MSF) $f_M(\mathbf{X}, \mathbf{Y})$ 来度量一对行人图像的相似度, 使相关行人图像对获得尽可能高的相似度, 而不相关行人图像对获得尽可能低的相似度.

MSF 形式如式(1)所示:

$$f_M(\mathbf{X}, \mathbf{Y}) = -(\mathbf{X} - \mathbf{Y})^T \mathbf{B}(\mathbf{X} - \mathbf{Y}) \quad (1)$$

其中 $\mathbf{B}$ 是一个半正定矩阵, 用于保证相似度函数的对称性.

正如引言所述, MSF 只与特征差分空间有关, 忽略了每个个体的表观特征, 因此对于同一个人在不同场景中的很大的表观变化的捕捉能力有限. 因此, 本文通过学习一个二次相似度函数(Quadratic Similarity Function, QSF) $f_Q(\mathbf{X}, \mathbf{Y})$ 来估计一对行人图像的相似度. QSF 是 MSF 的泛化形式, 而已有的工作却很少关注应用 QSF 来度量样本之间的相似度.

QSF 的一个复杂形式如式(2)所示:

$$f_Q(\mathbf{X}, \mathbf{Y}) = \mathbf{X}^T \mathbf{A}_1 \mathbf{X} + \mathbf{Y}^T \mathbf{A}_2 \mathbf{Y} + \mathbf{X}^T \mathbf{B}_1 \mathbf{Y} + \mathbf{Y}^T \mathbf{B}_2 \mathbf{X} + \mathbf{W}_1^T \mathbf{X} + \mathbf{W}_2^T \mathbf{Y} + e \quad (2)$$

其中 $\mathbf{A}_1, \mathbf{A}_2, \mathbf{B}_1, \mathbf{B}_2 \in R^{d \times d}$, $\mathbf{W}_1, \mathbf{W}_2 \in R^d$, $e \in R$. 考虑到一对行人图像对相似度的对称性, 即 $f_Q(\mathbf{X}, \mathbf{Y}) = f_Q(\mathbf{Y}, \mathbf{X})$, 令 $\mathbf{A}_1 = \mathbf{A}_2$, $\mathbf{B}_1 = \mathbf{B}_2$, $\mathbf{W}_1 = \mathbf{W}_2$, 忽略脚标, 则得到 QSF 的一般形式, 如式(3).

$$f_Q(\mathbf{X}, \mathbf{Y}) = \mathbf{X}^T \mathbf{A} \mathbf{X} + \mathbf{Y}^T \mathbf{A} \mathbf{Y} + \mathbf{X}^T \mathbf{B} \mathbf{Y} + \mathbf{Y}^T \mathbf{B} \mathbf{X} + \mathbf{W}^T \mathbf{X} + \mathbf{W}^T \mathbf{Y} + e \quad (3)$$

在 QSF 中, $\mathbf{W}^T \mathbf{X} + \mathbf{W}^T \mathbf{Y}$ 描述了一对行人图像的一阶关系, $\mathbf{X}^T \mathbf{A} \mathbf{X} + \mathbf{Y}^T \mathbf{A} \mathbf{Y} + \mathbf{X}^T \mathbf{B} \mathbf{Y} + \mathbf{Y}^T \mathbf{B} \mathbf{X}$ 描述了一对行人图像的二阶关系, e 是偏移量. 在二阶关系

中, $\mathbf{X}^T \mathbf{A} \mathbf{X} + \mathbf{Y}^T \mathbf{A} \mathbf{Y}$ 描述了一对行人图像的自相关关系, $\mathbf{X}^T \mathbf{B} \mathbf{Y} + \mathbf{Y}^T \mathbf{B} \mathbf{X}$ 描述了一对行人图像的互相关关系. 因为式(3)已经保证了相似度的对称性, 所以在 QSF 中, $\mathbf{A}$ 和 $\mathbf{B}$ 并不需要是半正定矩阵.

从式(3)可以看出, QSF 是 MSF 的泛化形式. 当 $\mathbf{A} = -\mathbf{B}$, $\mathbf{W} = 0$, $e = 0$, 并且 $\mathbf{B}$ 是一个半正定矩阵时, 可以从 QSF 得到 MSF 的形式. $\mathbf{A} = -\mathbf{B}$ 表示 MSF 的自相关矩阵和互相关矩阵互为负矩阵. $\mathbf{W} = 0$ 表示 MSF 忽略了一对行人图像的一阶关系. 因此, MSF 仅与一对行人图像的特征差分空间有关, 忽略了一对行人图像的每个个体的表观特征. 而 QSF 是局部自适应的, 捕捉用于行人再识别的区分性信息的能力更强.

为了说明上述观点, 作为举例说明, 可以考虑如下的一维空间的二分类情形. 其中一个类的样本来源于 $[-2, -0.3]$ 的均匀分布, 另一个类的样本来源于 $[0.3, 2]$ 的均匀分布. 当给定一个样本 $x = -0.5$, 图 2 显示了采用不同的相似度函数, x 和另一个样本 y 的相似度的分布, 其中 $c(x) = c(y)$ 表示 x 和 y 属于同一类, 用实线表示, $c(x) \neq c(y)$ 表示 x 和 y 属于不同类, 用虚线表示. 当相似度函数采用 MSF, 即 $f_M(x, y) = -\sigma(x - y)^2$, 如图 2(a) 所示, 同类样本和不同类样本之间的相似度的范围有重叠. MSF 并不能分开上述两类样本. 因为 σ 并不影响两类样本之间相似度的相对关系, 在图 2 中, $\sigma = 1$. 当相似度函数中的自相关矩阵和互相关矩阵不再互为负矩阵, 增强互相关关系的影响, 即 $f_{Q_1}(x, y) = -(x - y)^2 + 5xy$, 如图 2(b) 所示, 同类样本和不同类样本之间的相似度的范围存在间隔, 上述两类样本可以被分开. 当进一步增加一阶关系, 即 $f_Q(x, y) = -(x - y)^2 + 5xy - 15(x + y)$, 如图 2(c) 所示, 同类样本和不同类样本之间的相似度的范围的间隔将增大, 上述两类样本可以更好地被分开. 以上示例说明 QSF 比 MSF 更加鲁棒.

2.1 基于分类的 QSF 学习算法

本文将基于分类的 QSF 学习策略简称为 QSFC. QSFC 利用一个由相关行人图像对组成的集合 S 和一个由不相关行人图像对组成的集合 D , 学习相似度函数. 学习目标是在新的度量准则下, 相关行人图像对处于正无穷大, 不相关行人图像对处于负无穷大. 为了实现上述目标, 首先利用一个逻辑回归函数将二次函数相似度的范围规范到 $[0, 1]$, 式(4)为转换函数.

$$C_c(\mathbf{X}, \mathbf{Y}) = \frac{1}{1 + e^{-f_Q(\mathbf{X}, \mathbf{Y})}} \quad (4)$$

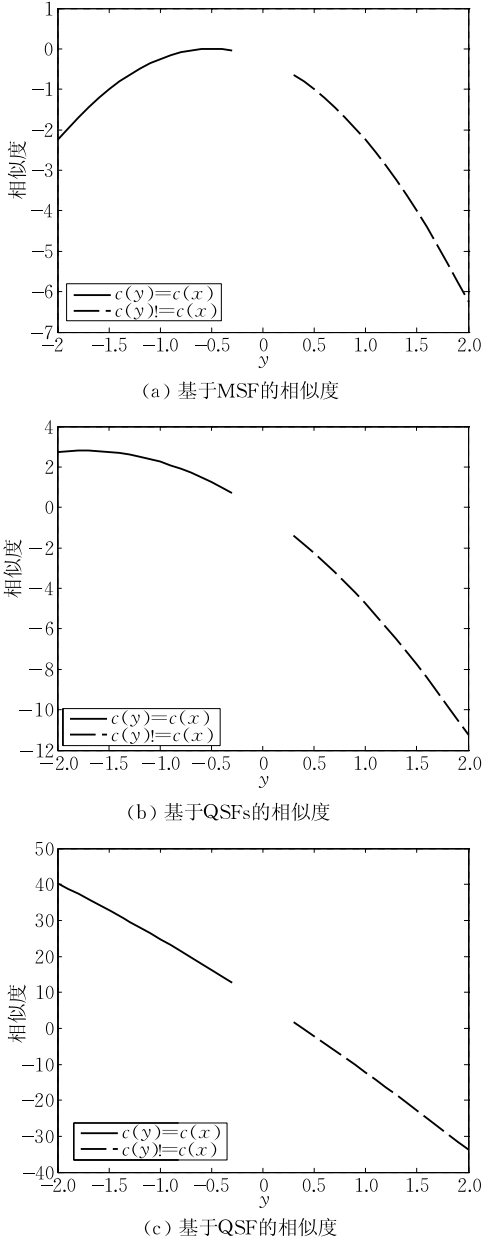


图2 利用QSF和MSF度量相似度的对比

然后通过最小化式(5)学到二次相似度函数的参数 $\theta = \{\mathbf{A}, \mathbf{B}, \mathbf{W}, e\}$.

$$J_c(\theta) = - \left[\frac{1}{|S|} J_c^S(\theta) + \frac{1}{|D|} J_c^D(\theta) \right] + \frac{\beta_1}{2} (\|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2) + \frac{\beta_2}{2} \|\mathbf{W}\|_F^2 \quad (5)$$

其中 $|S|$ 表示集合 S 中元素的数目, $\|\mathbf{A}\|_F$ 表示矩阵 $\mathbf{A}$ 的 Frobenius 范数, β_1 和 β_2 是为了防止过拟合的权重衰减参数, $J_c^S(\theta)$ 和 $J_c^D(\theta)$ 衡量由相似度函数估计相关图像对和不相关图像对的相似度的能力, 其定义如下:

$$J_c^S(\theta) = \sum_{(\mathbf{X}, \mathbf{Y}) \in S} \log[C_c(\mathbf{X}, \mathbf{Y})] \quad (6)$$

$$J_c^D(\theta) = \sum_{(\mathbf{X}, \mathbf{Y}) \in D} \log[1 - C_c(\mathbf{X}, \mathbf{Y})] \quad (7)$$

对于相关行人图像对 $\mathbf{X}$ 和 $\mathbf{Y}$, 优化式(5)可以使 $C_c(\mathbf{X}, \mathbf{Y})$ 接近于 1, $\mathbf{X}$ 和 $\mathbf{Y}$ 的相似度被压缩到正无穷大. 对于不相关行人图像对 $\mathbf{X}$ 和 $\mathbf{Y}$, 优化式(5)可以使 $C_c(\mathbf{X}, \mathbf{Y})$ 接近于 0, $\mathbf{X}$ 和 $\mathbf{Y}$ 的相似度被压缩到负无穷大. 因为目标函数 $J_c(\theta)$ 与逻辑回归函数类似, 本文采用一种拟牛顿迭代优化算法 L-BFGS 算法^[20]优化 $J_c(\theta)$. 优化算法的初始参数为欧氏距离相似度函数, 即 $\mathbf{A}_0 = -\mathbf{I}$, $\mathbf{B}_0 = \mathbf{I}$, $\mathbf{W}_0 = 0$, $e_0 = 0$. 式(8)~式(11)给了每一轮迭代中参数的更新量, 推导见附录 1.

$$\begin{aligned} \nabla_{\mathbf{A}} J_c(\theta) = & - \left\{ \frac{1}{|S|} \sum_{(\mathbf{X}, \mathbf{Y}) \in S} [1 - C_c(\mathbf{X}, \mathbf{Y})] \cdot \right. \\ & (\mathbf{X}\mathbf{X}^T + \mathbf{Y}\mathbf{Y}^T) - \frac{1}{|D|} \sum_{(\mathbf{X}, \mathbf{Y}) \in D} C_c(\mathbf{X}, \mathbf{Y}) \cdot \\ & \left. (\mathbf{X}\mathbf{X}^T + \mathbf{Y}\mathbf{Y}^T) \right\} + \beta_1 \mathbf{A} \end{aligned} \quad (8)$$

$$\begin{aligned} \nabla_{\mathbf{B}} J_c(\theta) = & - \left\{ \frac{1}{|S|} \sum_{(\mathbf{X}, \mathbf{Y}) \in S} [1 - C_c(\mathbf{X}, \mathbf{Y})] \cdot \right. \\ & (\mathbf{X}\mathbf{Y}^T + \mathbf{Y}\mathbf{X}^T) - \frac{1}{|D|} \sum_{(\mathbf{X}, \mathbf{Y}) \in D} C_c(\mathbf{X}, \mathbf{Y}) \cdot \\ & \left. (\mathbf{X}\mathbf{Y}^T + \mathbf{Y}\mathbf{X}^T) \right\} + \beta_1 \mathbf{B} \end{aligned} \quad (9)$$

$$\begin{aligned} \nabla_{\mathbf{W}} J_c(\theta) = & - \left\{ \frac{1}{|S|} \sum_{(\mathbf{X}, \mathbf{Y}) \in S} [1 - C_c(\mathbf{X}, \mathbf{Y})] \cdot \right. \\ & (\mathbf{X} + \mathbf{Y}) - \frac{1}{|D|} \sum_{(\mathbf{X}, \mathbf{Y}) \in D} C_c(\mathbf{X}, \mathbf{Y}) \cdot \\ & \left. (\mathbf{X} + \mathbf{Y}) \right\} + \beta_2 \mathbf{W} \end{aligned} \quad (10)$$

$$\begin{aligned} \nabla_e J_c(\theta) = & - \left\{ \frac{1}{|S|} \sum_{(\mathbf{X}, \mathbf{Y}) \in S} [1 - C_c(\mathbf{X}, \mathbf{Y})] - \right. \\ & \left. \frac{1}{|D|} \sum_{(\mathbf{X}, \mathbf{Y}) \in D} C_c(\mathbf{X}, \mathbf{Y}) \right\} \end{aligned} \quad (11)$$

算法 1 总结了基于分类的二次相似度函数学习算法(QSFC 算法).

算法 1. QSFC 算法.

输入: 由相关行人图像对组成的集合 S 和由不相关行人图像对组成的集合 D , 权重衰减参数 β_1 和 β_2

输出: QSF 的参数 $\mathbf{A}_T, \mathbf{B}_T, \mathbf{W}_T, e_T$

1. 初始化 $\mathbf{A}_0 = -\mathbf{I}, \mathbf{B}_0 = \mathbf{I}, \mathbf{W}_0 = 0, e_0 = 0$.
2. 根据式(8)~式(11), 计算参数更新量 $\nabla_{\mathbf{A}} J_c(\theta), \nabla_{\mathbf{B}} J_c(\theta), \nabla_{\mathbf{W}} J_c(\theta), \nabla_e J_c(\theta)$.
3. 利用 L-BFGS 算法计算参数更新率 ϵ .
4. 根据参数更新量和更新率更新参数, 即 $\mathbf{A}_{t+1} = \mathbf{A}_t - \epsilon \nabla_{\mathbf{A}} J_c(\theta), \mathbf{B}_{t+1} = \mathbf{B}_t - \epsilon \nabla_{\mathbf{B}} J_c(\theta), \mathbf{W}_{t+1} = \mathbf{W}_t - \epsilon \nabla_{\mathbf{W}} J_c(\theta),$

$$e_{t+1} = e_t - \epsilon \nabla_e J_c(\theta).$$

5. 重复步 2 到步 4, 直到收敛. 此时的参数为 $\mathbf{A}_T, \mathbf{B}_T, \mathbf{W}_T, e_T$.

2.2 基于排序的 QSF 学习算法

本文将基于排序的 QSF 学习策略简称为 QSFR. 给定一张行人图像 $\mathbf{X}$, 包含 $\mathbf{X}$ 的一个相关的行人图像对 $(\mathbf{X}, \mathbf{Y})$ 和一个不相关的行人图像对 $(\mathbf{X}, \mathbf{Z})$ 形成一个三元组 $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$. QSFR 利用由三元组组成的集合 O , 学习一个相似度函数区分一对行人图像. 学习的目标是使每个三元组中涉及的不相关行人图像对和相关行人图像对的相似度的差异尽可能的大. 为了实现上述目标, 首先利用一个逻辑回归函数将相似度的差规范到 $[0, 1]$, 式(12)为转换函数.

$$C_r(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \frac{1}{1 + e^{-(f_Q(\mathbf{X}, \mathbf{Y}) - f_Q(\mathbf{X}, \mathbf{Z}))}} \quad (12)$$

然后通过最小化式(13)学到二次相似度函数的参数 $\theta = \{\mathbf{A}, \mathbf{B}, \mathbf{W}, e\}$.

$$J_r(\theta) = -\frac{1}{|O|} \sum_{(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in O} \log(C_r(\mathbf{X}, \mathbf{Y}, \mathbf{Z})) + \frac{\beta_1}{2} (\|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2) + \frac{\beta_2}{2} \|\mathbf{W}\|_F^2 \quad (13)$$

其中 β_1 和 β_2 是为了防止过拟合的权重衰减参数.

优化式(13)可以使 $C_r(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ 接近于 1, 每个三元组中相关行人图像对和不相关行人图像对的差压缩到正无穷大. 同样 QSFR 采用 L-BFGS 算法^[20] 优化 $J_r(\theta)$. 优化算法的初始参数仍采用欧氏距离相似度函数. 式(14)~(17)给出了 QSFR 的每一轮迭代中参数的更新量, 推导见附录 2.

$$\nabla_{\mathbf{A}} J_r(\theta) = -\frac{1}{|O|} \sum_{(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in O} [1 - C_r(\mathbf{X}, \mathbf{Y}, \mathbf{Z})] \cdot (\mathbf{Y}\mathbf{Y}^T - \mathbf{Z}\mathbf{Z}^T) + \beta_1 \mathbf{A} \quad (14)$$

$$\nabla_{\mathbf{B}} J_r(\theta) = -\frac{1}{|O|} \sum_{(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in O} [1 - C_r(\mathbf{X}, \mathbf{Y}, \mathbf{Z})] \cdot (\mathbf{X}\mathbf{Y}^T + \mathbf{Y}\mathbf{X}^T - \mathbf{X}\mathbf{Z}^T - \mathbf{Z}\mathbf{X}^T) + \beta_1 \mathbf{B} \quad (15)$$

$$\nabla_{\mathbf{W}} J_r(\theta) = -\frac{1}{|O|} \sum_{(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in O} [1 - C_r(\mathbf{X}, \mathbf{Y}, \mathbf{Z})] \cdot (\mathbf{Y} - \mathbf{Z}) + \beta_2 \mathbf{W} \quad (16)$$

$$\nabla_e J_r(\theta) = 0 \quad (17)$$

算法 2 总结了基于排序的二次相似度函数学习算法(QSFR 算法).

算法 2. QSFR 算法.

输入: 三元组集合 O , 权重衰减参数 β_1 和 β_2

输出: QSF 的参数 $\mathbf{A}_T, \mathbf{B}_T, \mathbf{W}_T, e_T$

1. 初始化 $\mathbf{A}_0 = -\mathbf{I}, \mathbf{B}_0 = \mathbf{I}, \mathbf{W}_0 = 0, e_0 = 0$.

2. 根据式(14)~式(17), 计算参数更新量 $\nabla_{\mathbf{A}} J_r(\theta)$, $\nabla_{\mathbf{B}} J_r(\theta)$, $\nabla_{\mathbf{W}} J_r(\theta)$.

3. 利用 L-BFGS 算法计算参数更新率 ϵ .

4. 根据参数更新量和更新率更新参数, 即 $\mathbf{A}_{t+1} = \mathbf{A}_t - \epsilon \nabla_{\mathbf{A}} J_r(\theta)$, $\mathbf{B}_{t+1} = \mathbf{B}_t - \epsilon \nabla_{\mathbf{B}} J_r(\theta)$, $\mathbf{W}_{t+1} = \mathbf{W}_t - \epsilon \nabla_{\mathbf{W}} J_r(\theta)$, $e_{t+1} = 0$.

5. 重复步 2 到步 4, 直到收敛. 此时的参数为 $\mathbf{A}_T, \mathbf{B}_T, \mathbf{W}_T, e_T$.

3 基于 QSF 学习的行人再识别系统

基于 QSF 学习的行人再识别系统包括 3 个阶段: (1) 行人图像特征的提取; (2) QSF 学习; (3) 目标行人图像查询. 整个系统如图 3 所示. 而 QSF 则是根据 QSFC 或 QSFR 算法, 利用训练数据集, 离线学习得到.

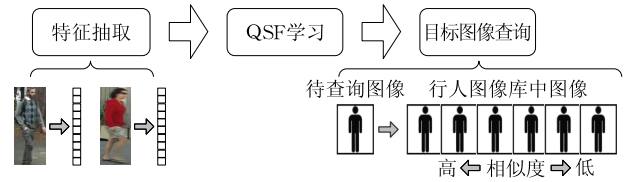


图 3 基于 QSF 学习的行人再识别系统框架

3.1 行人图像的特征描述

颜色和纹理是先行人再识别算法常用的两种特征. 后续试验表明, 只用颜色特征, 不用纹理特征, 也可以学到一个很好的相似度函数. 图 4 给出了提取一张行人图像的特征描述的示例. 本文从 4 种颜色空间提取特征, 即 HSV, 规范化的 RGB(NRGB), CIE Lab 和 YCbCr. 提取特征时, 将一张行人图像均分为 6 个水平条带, 对于每种颜色空间中每个通道, 在每个水平条带上抽取 16 维的颜色直方图特征. 上述所有特征级联形成行人图像的特征描述, 该特征的维度为 1152 ($6 \times 4 \times 3 \times 16$).

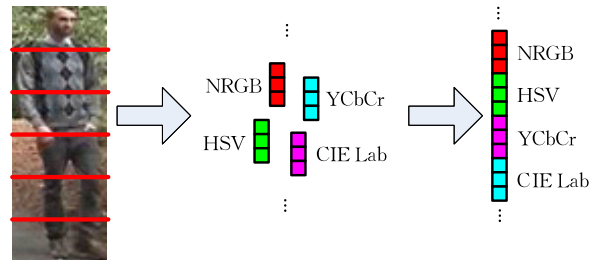


图 4 行人图像特征的提取

本文使用的特征在行人目标主体未能填充行人图像时也有一定的适用性. 不同场景中行人姿态的变化主要集中在水平方向上, 而在垂直方向上并不很明显. 本文使用的特征对于水平方向的移动具有一定的不变性. 此外, 也可以结合行人检测器和人体

姿态估计对行人图像进行校准,使行人目标主体较好填充行人图像。

为了说明本文提出的通过学习一个二次相似度函数,来估计一对行人图像的相似度的有效性,本文在实验中采用了上述比较简单的行人颜色直方图特征。后续实验表明,虽然采用的特征简单,但本文算法性能优于已有的行人再识别算法。如果采用更加复杂的特征来描述行人图像,本文算法的识别性能会进一步提升。

3.2 QSF 学习

为了提高特征的有效性,便于 QSF 的学习,实验中首先利用 PCA 降维,将行人图像的特征维数降到 150。对于 VIPeR 数据集^[15],保持大约 97.6% 的能量。对于 CUHK 数据集^[12],保持大约 98.4% 的能量。在训练阶段,根据 QSFC 或 QSFR 算法,离线学习一个 QSF。当估计出 QSF 的参数 A, B, W, e ,则根据式(3)评估一对行人图像 X 和 Y 的相似度。

3.3 目标行人图像识别

行人再识别的研究目标是在已有的多个空间位置上不交叠的监控视频中,根据目标行人图像,检索该行人是否出现在这些视频中。本文假设通过行人检测技术,已经将不交叠的监控视频转换为行人图像库。利用学习到的较优的相似度函数,度量目标

行人图像和行人图像库中所有图像的相似度。行人图像库中相似度高的图像为潜在的目标行人图像。

4 实 验

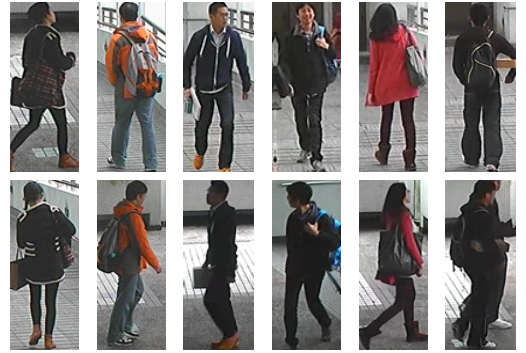
本节首先介绍实验细节,然后分析 QSF 模型,验证 QSF 的一阶关系和二阶关系的有效性。最后,比较本文算法和已有的行人再识别算法。

4.1 实验细节

VIPeR^[15]和 CUHK^[12]是行人再识别广泛使用的公共数据集,本文将在这两个数据集上评测本文算法。VIPeR 数据集中每个行人有 2 张图像,这两张图像来源于 2 个不同的摄像机场景,总共由 632 个行人组成,所有图像的分辨率都被规范到 48×128 。CUHK 数据集由 971 个行人组成,每个行人有 4 张图像,同样来源于 2 个不同的摄像机视场,每个视场有两张图像,所有图像的分辨率都被规范到 60×160 。图 5 显示了这两个数据集中一些图像,其中同一列中的行人图像为同一个人。因为这两个数据集中同一个人的表现由于姿态、视角、光照和背景的变化而变化很大,所以非常适合用于研究行人再识别问题。



(a) VIPeR数据集



(b) CUHK数据集

图 5 公共数据集行人图像对举例

为了评价行人再识别算法性能,累计匹配特性(Cumulative Match Characteristic, CMC)曲线在先前的算法中广泛被采用。CMC 曲线描述的是在一个给定的行人图像库中检索目标行人图像,在前 r 个搜索结果中识别出目标行人的比例。第 1 匹配率($r=1$)表征了算法真正识别能力,非常重要。但是,在实际应用中,因为可以借助人眼判断,从检索的前 r 个结果中发现目标行人,所以当 r 取值很小时的匹配率也值得去关注。因此,行人再识别问题也可以看成是一个信息检索问题。除了 CMC 曲线,本文引

入规范化折扣累计增益(NDCG),信息检索中常用的评测指标来评测算法性能。NDCG 可以根据 CMC 曲线,利用式(18)来计算。本文实验中评测算法性能时采用 NDCG@15。

$$\text{NDCG}@k = \frac{\sum_{i=1}^k H(i) \cdot G(i)}{\sum_{i=1}^k H(i)} \quad (18)$$

其中 $G(i)$ 是 CMC 曲线中第 i 匹配率,

$$H(i) = \begin{cases} 1, & i = 1 \\ 1/\log_2(i), & 2 \leq i \leq k \end{cases}$$

本文采用行人再识别的标准评测框架评测本文算法性能^[5,12,15]. 对于一个评测集中所有行人,随机选择一半作为训练集,余下的作为测试集. 测试集由行人图像库和查询集组成. 对于测试集中的每个行人,随机选择一个场景的行人图像并入行人图像库,另一个场景的行人图像并入查询集. 通过衡量在行人图像库中检索目标行人的能力来评测一个行人再识别算法的性能. 与此同时,从 VIPeR 数据集中随机选择 100 个行人作为训练集,余下的用于测试,形成一个新的实验场景用于评估基于学习的行人再识别算法的过拟合程度. 上述每组实验的过程重复 10 次,每组的实验结果为相关的 10 次实验的平均值. 因为本文组织实验是按照行人再识别的标准评测框架,已有算法的性能直接从相应文章的图表中获得.

4.2 QSF 模型分析

为了验证 QSF 的一阶关系和二阶关系的有效性,本文在两个公共数据集上,采用本文提出的两种优化策略分别学习 MSF、QSF 和 QSFs(只含有二阶关系的 QSF). 其中基于分类的策略的算法简称为 MSFC、QSFC、QSFCs;基于排序的策略的算法简称为 MSFR、QSFR、QSFRs. 表 1 和表 2 显示了比较结果. 在 VIPeR 数据集上,QSFCs 的检索性能比 MSFC 大约高 2.1%,QSFRs 的检索性能比 MSFR 大约高 2.5%,其中检索性能指 NDCG@15. 在 CUHK 数据集上,QSFCs 的检索性能比 MSFC 大约高 2.8%,QSFRs 的检索性能比 MSFR 大约高 3.2%. 由此可见,利用两个完全不同的矩阵描述自相关关系和互相关关系,可以使相似度函数局部自适应,有益于行人再识别. 此外,在 VIPeR 数据集上,QSFC 的检索性能比 QSFCs 大约高 2.8%,QSFR 的检索性能比 QSFRs 大约高 1.6%. 在 CUHK 数据集上,QSFC 的检索性能比 QSFCs 大约高 5.5%,QSFR 的检索性能比 QSFRs 大约高 4.3%. 在相似度函数中引入一对图像的一阶关系,可以进一步使相似度函数局部自适应,提高性能.

表 1 在 VIPeR 数据集上 QSF 和 MSF 的比较

算法	第 1 匹配率	第 5 匹配率	第 10 匹配率	第 15 匹配率	检索 性能
QSFC	0.301	0.628	0.767	0.839	0.605
QSFCs	0.259	0.604	0.744	0.842	0.577
MSFC	0.245	0.580	0.725	0.808	0.556
QSFR	0.303	0.614	0.761	0.831	0.595
QSFRs	0.271	0.603	0.745	0.821	0.579
MSFR	0.241	0.573	0.727	0.807	0.554

表 2 在 CUHK 数据集上 QSF 和 MSF 的比较

算法	第 1 匹配率	第 5 匹配率	第 10 匹配率	第 15 匹配率	检索 性能
QSFC	0.270	0.526	0.648	0.712	0.512
QSFCs	0.235	0.469	0.581	0.654	0.457
MSFC	0.208	0.434	0.553	0.621	0.429
QSFR	0.252	0.500	0.623	0.693	0.489
QSFRs	0.224	0.452	0.574	0.638	0.446
MSFR	0.203	0.422	0.531	0.597	0.414

防止过拟合的权重衰减参数 β_1 和 β_2 是本文提出的两种学习 QSF 的算法的主要参数. 本文从 VIPeR 数据集中随机产生一轮测试集规模为 316 的实验,通过交叉验证决定上述参数. 这轮实验与参与评测算法性能的 10 轮实验不同. 图 6 描述了 QSFC 和 QSFR 的检索性能随 β_1 和 β_2 的变化. 在本文的实验中,对于 QSFC, $\beta_1 = 2.5 \times 10^{-6}$, $\beta_2 = 0.1$; 对于 QSFR, $\beta_1 = 1.0 \times 10^{-6}$, $\beta_2 = 0.5$.

4.3 与先前的行人再识别算法对比

表 3 将本文的算法 QSFC 和 QSFR 与已有的无监督的行人再识别算法在 VIPeR 数据集(测试集规模为 316)上进行对比,其中包括 SDALF^[1]、CPS^[2]、MCTCS^[3]、eSDC^[4]. QSFC 的检索性能比 SDALF 大约高 21.6%,比 CPS 大约高 16.3%,比 MCTCS 大约高 13.9%,比 eSDC 大约高 10.7%. QSFR 的检索性能比 SDALF 大约高 20.6%,比 CPS 大约高 15.3%,比 MCTCS 大约高 12.9%,比 eSDC 大约高 9.7%. 这些表明,即使使用简单的颜色直方图特征,从数据中学习有区分性的信息,要优于只是单纯的通过手动设计鲁棒性的特征的算法.

表 3 在 VIPeR 数据集,与无监督的行人再识别算法比较,其中测试集规模为 316

算法	第 1 匹配率	第 5 匹配率	第 10 匹配率	第 15 匹配率	检索 性能
QSFC	0.301	0.628	0.767	0.839	0.605
QSFR	0.303	0.614	0.761	0.831	0.595
SDALF ^[1]	0.200	0.390	0.490	0.580	0.389
CPS ^[2]	0.220	0.450	0.570	0.650	0.442
MCTCS ^[3]	0.230	0.480	0.400	0.690	0.466
eSDC ^[4]	0.267	0.507	0.624	0.700	0.498

表 4 将本文的算法 QSFC 和 QSFR 与已有的有监督的行人再识别算法在 VIPeR 数据集(测试集规模为 316)上进行对比,其中包括 LF^[11]、LAFT^[13]、RPLM^[10]、SalMatch^[14]. 因为这些算法明显优于文献[5-9,12]中的有监督的行人再识别算法,因此本文只与这些算法进行对比. QSFC 的检索性能比 LF 大约高 10.1%,比 LAFT 大约高 7.1%,比 RPLM

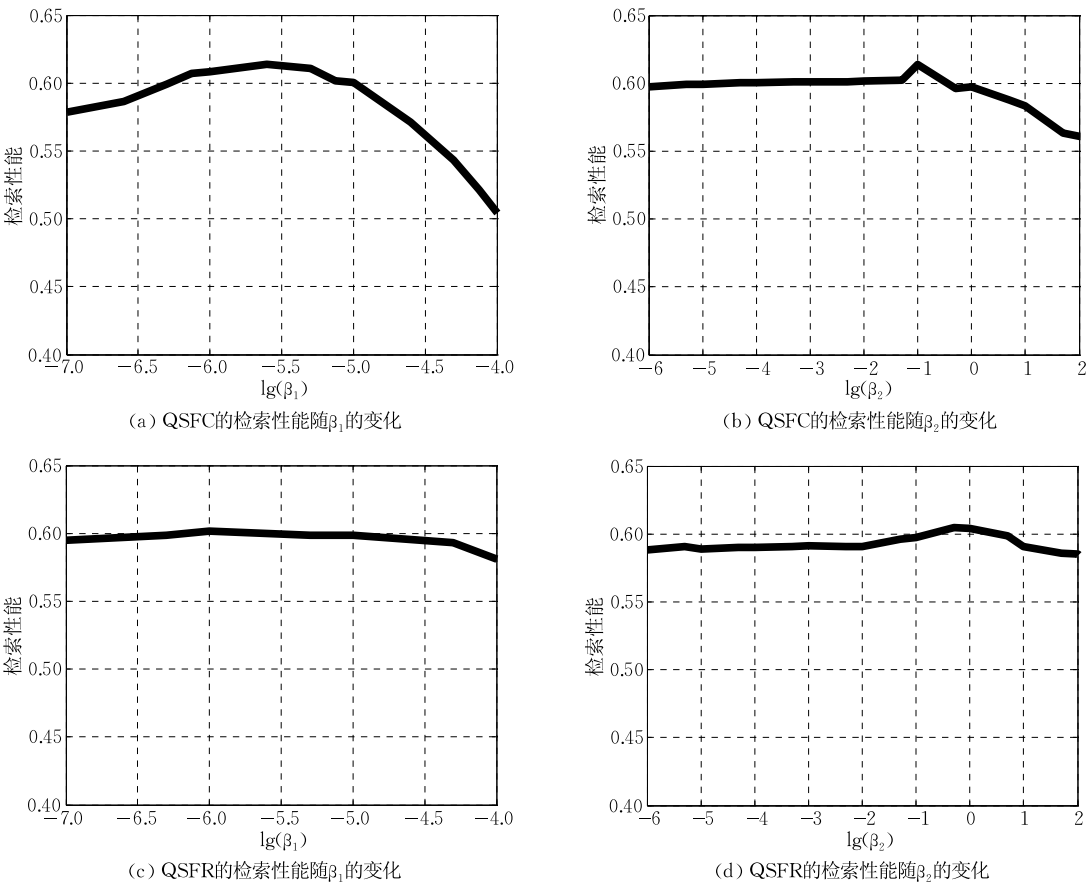


图 6 QSFC 和 QSFR 的检索性能随 β_1 和 β_2 的变化曲线

大约高 7.4%，比 SalMatch 大约高 8.2%。QSFR 的检索性能比 LF 大约高 9.1%，比 LAFT 大约高 6.1%，比 RPLM 大约高 6.4%，比 SalMatch 大约高 7.2%。QSFC 和 QSFR 优于基于 MSF 的算法 LF 和 RPLM，可见 QSF 与 MSF 相比，更易描述不同摄像机视场下同一个人的有区分性的信息。此外，QSFC 和 QSFR 也优于基于与个体相关的相似度函数学习的算法 LAFT 和 SalMatch。可见利用 QSF 挖掘相似度函数与个体表现之间的关系更加有效。

表 4 在 VIPeR 数据集，与有监督的行人再识别算法比较，其中测试集规模为 316

算法	第 1 匹配率	第 5 匹配率	第 10 匹配率	第 15 匹配率	检索性能
QSFC	0.301	0.628	0.767	0.839	0.605
QSFR	0.303	0.614	0.761	0.831	0.595
LF ^[11]	0.242	0.500	0.671	0.750	0.504
LAFT ^[13]	0.296	0.530	0.693	0.785	0.534
RPLM ^[10]	0.270	0.540	0.690	0.770	0.531
SalMatch ^[14]	0.302	0.525	0.660	0.740	0.523

表 5 将本文的算法 QSFC 和 QSFR 与已有的有监督的行人再识别算法在训练数据较少时进行比较，其中包括 PRDC^[7]、PCCA^[9]、LAFT^[13] 和

RPLM^[10]。QSFC 的检索性能比 PRDC 大约高 13.1%，比 PCCA 大约高 12.6%，比 LAFT 大约高 6.6%，比 RPLM 大约高 11.3%。QSFR 的检索性能比 PRDC 大约高 12.1%，比 PCCA 大约高 11.6%，比 LAFT 大约高 5.6%，比 RPLM 大约高 10.3%。这些表明，当训练数据较少时，QSFC 和 QSFR 缓解了基于学习的算法的过拟合问题。

表 5 在 VIPeR 数据集，与有监督的行人再识别算法比较，其中测试集规模为 532

算法	第 1 匹配率	第 5 匹配率	第 10 匹配率	第 15 匹配率	检索性能
QSFC	0.138	0.378	0.522	0.617	0.378
QSFR	0.137	0.364	0.509	0.598	0.368
PRDC ^[7]	0.091	0.242	0.344	0.420	0.247
PCCA ^[9]	0.092	0.249	0.374	0.440	0.252
LAFT ^[13]	0.129	0.303	0.427	0.520	0.312
RPLM ^[10]	0.110	0.250	0.280	0.460	0.265

表 6 将本文的算法 QSFC 和 QSFR 与 SalMatch^[14] 在 CUHK 数据集上进行对比。因为文献[14]中表明 SalMatch 在这个数据集上优于已有的算法，因此本文只与 SalMatch 进行对比。虽然 QSFC 和 QSFR 的第 1 匹配率低于 SalMatch，但是 QSFC 和 QSFR

自第 5 匹配率以后,明显高于 SalMatch. QSFC 的检索性能比 SalMatch 大约高 5.2%,QSFR 的检索性能比 SalMatch 大约高 2.9%. SalMatch 是基于与个体相关的相似度函数学习的算法,实验结果再次说

表 6 在 CUHK 数据集,与已有的行人再识别算法比较

算法	第 1 匹配率	第 5 匹配率	第 10 匹配率	第 15 匹配率	检索性能
QSFC	0.270	0.526	0.648	0.712	0.512
QSFR	0.252	0.500	0.623	0.693	0.489
SalMatch ^[14]	0.285	0.460	0.550	0.635	0.460

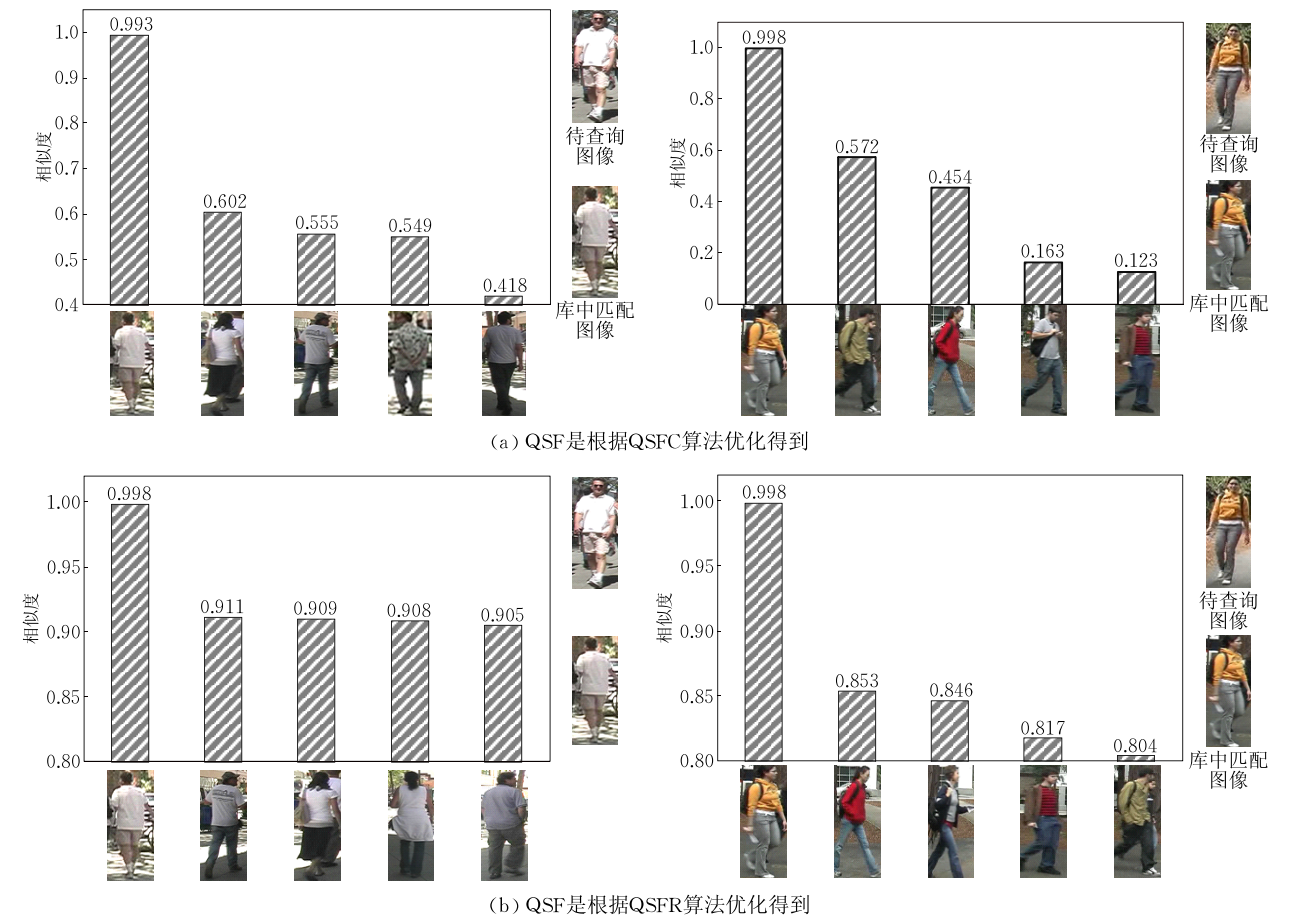


图 7 利用 QSF 从行人图像库中查询目标行人的示例

5 结 论

本文将行人再识别问题形式化为一个二次相似度函数学习问题. QSF 是马氏距离相似度函数的泛化形式. 与 MSF 相比,QSF 重新设计了一对行人图像的二阶关系,并且引入一对行人图像的一阶关系,从而可以更好地捕捉同一个人在不同摄像机视场下较大的表观变化. 为了学习 QSF,本文分别从分类和排序的角度出发,设计了两种不同的新的优化策略. 由行人再识别的公共数据集 VIPeR 和 CUHK 的实验表明,这两种不同的优化策略可以学习到有

效的二次相似度函数,识别性能优于已有的行人再识别算法.

明了 QSF 可以更好地挖掘相似度函数与个体表观之间的关系.

图 7 给出一些利用 QSF 从行人图像库中查询目标行人的示例. 在每个示例中,最右边的两张行人图像显示了待查询图像和行人图像库中与其匹配的图像. 而 x 轴上的 5 张图则是根据 QSF 计算的相似度,从左到右依次为行人图像库中与待查询行人图像相似度最高的 5 张图像. 图中的相似度的值是利用式(4)进行规范化后的值.

效的二次相似度函数,识别性能优于已有的行人再识别算法.

参 考 文 献

[1] Farenzena M, Bazzani L, Perina A, et al. Person re-identification by symmetry-driven accumulation of local features// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA, 2010: 2360-2367

[2] Cheng Dong-Seon, Cristani M, Stoppa M, et al. Custom pictorial structures for re-identification//Proceedings of the British Machine Vision Conference. Dundee, British, 2011: 749-760

- [3] Xu Yuan-Lu, Lin Liang, Zheng Wei-Shi, et al. Human re-identification by matching compositional template with cluster sampling//Proceedings of the IEEE International Conference on Computer Vision. Sydney, Australia, 2013: 3152-3159
- [4] Zhao Rui, Ouyang Wan-Li, Wang Xiao-Gang. Unsupervised salience learning for person re-identification//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA, 2013: 3586-3593
- [5] Gray D, Tao Hai. Viewpoint invariant pedestrian recognition with an ensemble of localized features//Proceedings of the European Conference on Computer Vision. Marseille, France, 2008: 262-275
- [6] Prosser B, Zheng Wei-Shi, Gong Shao-Gang, et al. Person re-identification by support vector ranking//Proceedings of the British Machine Vision Conference. Aberystwyth, British, 2010: 231-242
- [7] Zheng Wei-Shi, Gong Shao-Gang, Xiang Tao. Person reidentification by probabilistic relative distance comparison //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA, 2011: 649-656
- [8] Kostinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 2288-2295
- [9] Mignon A, Jurie F. PCCA: A new approach for distance learning from sparse pairwise constraints//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 2666-2672
- [10] Hirzer M, Ranz P M, Kostinger M, et al. Relaxed pairwise learned metric for person reidentification//Proceedings of the European Conference on Computer Vision. Firenze, Italy, 2012: 780-793
- [11] Ma Bing-Peng, Su Yu, Jurie F. Local descriptors encoded by fisher vectors for person re-identification//Proceedings of the Workshops of European Conference on Computer Vision. Firenze, Italy, 2012: 413-422
- [12] Li Wei, Zhao Rui, Wang Xiao-Gang. Human reidentification with transferred metric learning//Proceedings of the Asian Conference on Computer Vision. Daejeon, Korea, 2012: 31-44
- [13] Li Wei, Wang Xiao-Gang. Locally aligned feature transforms across views//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA, 2013: 3594-3601
- [14] Zhao Rui, Ouyang Wan-Li, Wang Xiao-Gang. Person re-identification by salience matching//Proceedings of the IEEE International Conference on Computer Vision. Sydney, Australia, 2013: 2528-2535
- [15] Gray D, Brennan S, Tao Hai. Evaluating appearance models for recognition, reacquisition, and tracking//Proceedings of the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. Rio de Janeiro, USA, 2007: 41-47
- [16] Gong Shao-Gang, Cristani M, Yan Shui-Cheng, et al. Person Re-Identification. London: Springer Verlag, 2014
- [17] Globerson A, Roweis S. Metric learning by collapsing classes//Proceedings of the Conference on Neural Information Processing Systems. Vancouver, Canada, 2005: 451-458
- [18] Davis J, Kulis B, Jain P, et al. Information-theoretic metric learning//Proceedings of the International Conference on Machine Learning. Oregon, USA, 2007: 209-216
- [19] Weinberger K, Saul L. Distance metric learning for large margin nearest neighbor classification. The Journal of Machine Learning Research, 2009, 10(2): 207-244
- [20] Nocedal J, Wright S. Numerical Optimization(2nd). Germany: Springer Verlag, 2006

附录 1. QSFC 优化迭代中参数更新量的推导.

式(8)~(11)中 QSFC 每一轮优化迭代的参数的更新量 $\nabla_A J_c(\theta)$, $\nabla_B J_c(\theta)$, $\nabla_W J_c(\theta)$, $\nabla_e J_c(\theta)$ 的推导如下:

(1) $\nabla_A J_c(\theta)$ 的推导

$$\nabla_A J_c(\theta) = - \left[\frac{1}{|S|} \nabla_A J_c^S(\theta) + \frac{1}{|D|} \nabla_A J_c^D(\theta) \right] + \beta_1 \mathbf{A} \quad (1)$$

其中 $\nabla_A J_c^S(\theta)$ 和 $\nabla_A J_c^D(\theta)$ 分别是根据 $J_c^S(\theta)$ 和 $J_c^D(\theta)$ 计算的梯度, 推导过程如下:

$$\begin{aligned} \nabla_A J_c^S(\theta) &= \frac{\partial J_c^S(\theta)}{\partial \mathbf{A}} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}) \in S} \frac{C_c(\mathbf{X}, \mathbf{Y}) \cdot [1 - C_c(\mathbf{X}, \mathbf{Y})]}{C_c(\mathbf{X}, \mathbf{Y})} \cdot \frac{\partial f_q(\mathbf{X}, \mathbf{Y})}{\partial \mathbf{A}} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}) \in S} \{ [1 - C_c(\mathbf{X}, \mathbf{Y})] \cdot (\mathbf{X}\mathbf{X}^T + \mathbf{Y}\mathbf{Y}^T) \} \quad (2) \\ \nabla_A J_c^D(\theta) &= \frac{\partial J_c^D(\theta)}{\partial \mathbf{A}} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}) \in D} - \frac{C_c(\mathbf{X}, \mathbf{Y}) \cdot [1 - C_c(\mathbf{X}, \mathbf{Y})]}{[1 - C_c(\mathbf{X}, \mathbf{Y})]} \cdot \frac{\partial f_q(\mathbf{X}, \mathbf{Y})}{\partial \mathbf{A}} \end{aligned}$$

$$= \sum_{(\mathbf{X}, \mathbf{Y}) \in D} [-C_c(\mathbf{X}, \mathbf{Y}) \cdot (\mathbf{X}\mathbf{X}^T + \mathbf{Y}\mathbf{Y}^T)] \quad (3)$$

(2) $\nabla_B J_c(\theta)$ 的推导

$$\nabla_B J_c(\theta) = - \left[\frac{1}{|S|} \nabla_B J_c^S(\theta) + \frac{1}{|D|} \nabla_B J_c^D(\theta) \right] + \beta_1 \mathbf{B} \quad (4)$$

其中 $\nabla_B J_c^S(\theta)$ 和 $\nabla_B J_c^D(\theta)$ 分别是根据 $J_c^S(\theta)$ 和 $J_c^D(\theta)$ 计算的梯度, 推导过程如下:

$$\begin{aligned} \nabla_B J_c^S(\theta) &= \frac{\partial J_c^S(\theta)}{\partial \mathbf{B}} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}) \in S} \frac{C_c(\mathbf{X}, \mathbf{Y}) \cdot [1 - C_c(\mathbf{X}, \mathbf{Y})]}{C_c(\mathbf{X}, \mathbf{Y})} \cdot \frac{\partial f_q(\mathbf{X}, \mathbf{Y})}{\partial \mathbf{B}} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}) \in S} \{ [1 - C_c(\mathbf{X}, \mathbf{Y})] \cdot (\mathbf{X}\mathbf{Y}^T + \mathbf{Y}\mathbf{X}^T) \} \quad (5) \\ \nabla_B J_c^D(\theta) &= \frac{\partial J_c^D(\theta)}{\partial \mathbf{B}} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}) \in D} - \frac{C_c(\mathbf{X}, \mathbf{Y}) \cdot [1 - C_c(\mathbf{X}, \mathbf{Y})]}{[1 - C_c(\mathbf{X}, \mathbf{Y})]} \cdot \frac{\partial f_q(\mathbf{X}, \mathbf{Y})}{\partial \mathbf{B}} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}) \in D} - [C_c(\mathbf{X}, \mathbf{Y}) \cdot (\mathbf{X}\mathbf{Y}^T + \mathbf{Y}\mathbf{X}^T)] \quad (6) \end{aligned}$$

(3) $\nabla_{\mathbf{W}} J_c(\theta)$ 的推导

$$\nabla_{\mathbf{W}} J_c(\theta) = - \left[\frac{1}{|S|} \nabla_{\mathbf{W}} J_c^S(\theta) + \frac{1}{|D|} \nabla_{\mathbf{W}} J_c^D(\theta) \right] + \beta_2 \mathbf{W} \quad (7)$$

其中 $\nabla_{\mathbf{W}} J_c^S(\theta)$ 和 $\nabla_{\mathbf{W}} J_c^D(\theta)$ 分别是根据 $J_c^S(\theta)$ 和 $J_c^D(\theta)$ 计算的梯度, 推导过程如下:

$$\begin{aligned} \nabla_{\mathbf{W}} J_c^S(\theta) &= \frac{\partial J_c^S(\theta)}{\partial \mathbf{W}} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}) \in S} \frac{C_c(\mathbf{X}, \mathbf{Y}) \cdot [1 - C_c(\mathbf{X}, \mathbf{Y})]}{C_c(\mathbf{X}, \mathbf{Y})} \cdot \frac{\partial f_Q(\mathbf{X}, \mathbf{Y})}{\partial \mathbf{W}} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}) \in S} \{ [1 - C_c(\mathbf{X}, \mathbf{Y})] \cdot (\mathbf{X} + \mathbf{Y}) \} \quad (8) \end{aligned}$$

$$\begin{aligned} \nabla_{\mathbf{W}} J_c^D(\theta) &= \frac{\partial J_c^D(\theta)}{\partial \mathbf{W}} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}) \in D} - \frac{C_c(\mathbf{X}, \mathbf{Y}) \cdot [1 - C_c(\mathbf{X}, \mathbf{Y})]}{[1 - C_c(\mathbf{X}, \mathbf{Y})]} \cdot \frac{\partial f_Q(\mathbf{X}, \mathbf{Y})}{\partial \mathbf{W}} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}) \in D} - [C_c(\mathbf{X}, \mathbf{Y}) \cdot (\mathbf{X} + \mathbf{Y})] \quad (9) \end{aligned}$$

(4) $\nabla_e J_c(\theta)$ 的推导

$$\nabla_e J_c(\theta) = - \left[\frac{1}{|S|} \nabla_e J_c^S(\theta) + \frac{1}{|D|} \nabla_e J_c^D(\theta) \right] \quad (10)$$

其中 $\nabla_e J_c^S(\theta)$ 和 $\nabla_e J_c^D(\theta)$ 分别是根据 $J_c^S(\theta)$ 和 $J_c^D(\theta)$ 计算的梯度, 推导过程如下:

$$\begin{aligned} \nabla_e J_c^S(\theta) &= \frac{\partial J_c^S(\theta)}{\partial e} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}) \in S} \frac{C_c(\mathbf{X}, \mathbf{Y}) \cdot [1 - C_c(\mathbf{X}, \mathbf{Y})]}{C_c(\mathbf{X}, \mathbf{Y})} \cdot \frac{\partial f_Q(\mathbf{X}, \mathbf{Y})}{\partial e} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}) \in S} [1 - C_c(\mathbf{X}, \mathbf{Y})] \quad (11) \end{aligned}$$

$$\begin{aligned} \nabla_e J_c^D(\theta) &= \frac{\partial J_c^D(\theta)}{\partial e} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}) \in D} - \frac{C_c(\mathbf{X}, \mathbf{Y}) \cdot [1 - C_c(\mathbf{X}, \mathbf{Y})]}{[1 - C_c(\mathbf{X}, \mathbf{Y})]} \cdot \frac{\partial f_Q(\mathbf{X}, \mathbf{Y})}{\partial e} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}) \in D} - C_c(\mathbf{X}, \mathbf{Y}) \quad (12) \end{aligned}$$

附录 2. QSFRR 优化迭代中参数更新量的推导.

假设 $J_r^O(\theta) = \sum_{(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in O} \log(C_r(\mathbf{X}, \mathbf{Y}, \mathbf{Z}))$, 则式(14)~(17)

中 QSFRR 每一轮优化迭代的参数的更新量 $\nabla_{\mathbf{A}} J_r(\theta)$, $\nabla_{\mathbf{B}} J_r(\theta)$, $\nabla_{\mathbf{W}} J_r(\theta)$ 的推导如下:

(1) $\nabla_{\mathbf{A}} J_r(\theta)$ 的推导

$$\nabla_{\mathbf{A}} J_r(\theta) = - \frac{1}{|O|} \nabla_{\mathbf{A}} J_r^O(\theta) + \beta_1 \mathbf{A} \quad (1)$$

其中 $\nabla_{\mathbf{A}} J_r^O(\theta)$ 是根据 $J_r^O(\theta)$ 计算的梯度, 推导过程如下:

$$\begin{aligned} \nabla_{\mathbf{A}} J_r^O(\theta) &= \frac{\partial J_r^O(\theta)}{\partial \mathbf{A}} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in O} \left\{ \frac{C_r(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \cdot [1 - C_r(\mathbf{X}, \mathbf{Y}, \mathbf{Z})]}{C_r(\mathbf{X}, \mathbf{Y}, \mathbf{Z})} \cdot \left[\frac{\partial f_Q(\mathbf{X}, \mathbf{Y})}{\partial \mathbf{A}} - \frac{\partial f_Q(\mathbf{X}, \mathbf{Z})}{\partial \mathbf{A}} \right] \right\} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in O} \{ [1 - C_r(\mathbf{X}, \mathbf{Y}, \mathbf{Z})] \cdot [(\mathbf{X}\mathbf{X}^T + \mathbf{Y}\mathbf{Y}^T) - (\mathbf{X}\mathbf{X}^T + \mathbf{Z}\mathbf{Z}^T)] \} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in O} \{ [1 - C_r(\mathbf{X}, \mathbf{Y}, \mathbf{Z})] \cdot (\mathbf{Y}\mathbf{Y}^T - \mathbf{Z}\mathbf{Z}^T) \} \quad (2) \end{aligned}$$

(2) $\nabla_{\mathbf{B}} J_r(\theta)$ 的推导

$$\nabla_{\mathbf{B}} J_r(\theta) = - \frac{1}{|O|} \nabla_{\mathbf{B}} J_r^O(\theta) + \beta_1 \mathbf{B} \quad (3)$$

其中 $\nabla_{\mathbf{B}} J_r^O(\theta)$ 是根据 $J_r^O(\theta)$ 计算的梯度, 推导过程如下:

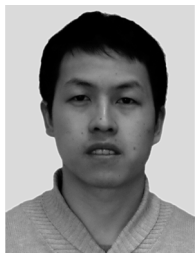
$$\nabla_{\mathbf{B}} J_r^O(\theta) = \frac{\partial J_r^O(\theta)}{\partial \mathbf{B}}$$

(3) $\nabla_{\mathbf{W}} J_r(\theta)$ 的推导

$$\nabla_{\mathbf{W}} J_r(\theta) = - \frac{1}{|O|} \nabla_{\mathbf{W}} J_r^O(\theta) + \beta_2 \mathbf{W} \quad (5)$$

其中 $\nabla_{\mathbf{W}} J_r^O(\theta)$ 是根据 $J_r^O(\theta)$ 计算的梯度, 推导过程如下:

$$\begin{aligned} \nabla_{\mathbf{W}} J_r^O(\theta) &= \frac{\partial J_r^O(\theta)}{\partial \mathbf{W}} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in O} \left\{ \frac{C_r(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \cdot [1 - C_r(\mathbf{X}, \mathbf{Y}, \mathbf{Z})]}{C_r(\mathbf{X}, \mathbf{Y}, \mathbf{Z})} \cdot \left[\frac{\partial f_Q(\mathbf{X}, \mathbf{Y})}{\partial \mathbf{W}} - \frac{\partial f_Q(\mathbf{X}, \mathbf{Z})}{\partial \mathbf{W}} \right] \right\} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in O} \{ [1 - C_r(\mathbf{X}, \mathbf{Y}, \mathbf{Z})] \cdot [(\mathbf{X} + \mathbf{Y}) - (\mathbf{X} + \mathbf{Z})] \} \\ &= \sum_{(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in O} \{ [1 - C_r(\mathbf{X}, \mathbf{Y}, \mathbf{Z})] \cdot (\mathbf{Y} - \mathbf{Z}) \} \quad (6) \end{aligned}$$



DU Yu-Ning, born in 1987, Ph. D. candidate. His main research interest is computer vision.

AI Hai-Zhou, born in 1964, professor, Ph. D. supervisor. His main research interest is computer vision.

Background

Pedestrian re-identification aims to identify the same person's images in an existing database come from non-overlapping camera views. It is a valuable problem and has enormous potential for practical applications like criminals monitoring and investigation, video index, lost person search, to name a few. It is also very challenging because the appearance of the same person between different cameras changes dramatically caused by pose, viewpoint and illumination variations. The appearance variations are large not only between the images of the different persons, but also between the ones of the same person.

In recent years, many researchers pay attention to this problem and propose lots of approaches. Those approaches can be roughly divided into two categories: unsupervised and supervised.

Unsupervised approaches aim to extract visual features that are distinctive and stable between different camera views by hand. Since there are large intra-class appearance variations under realistic conditions, it is difficult to design robust visual features worked well for all kinds of variations by hand. Moreover, those features extraction is often very time consuming.

The idea of supervised approaches is exploring the discriminative information from the data with machine learning algorithms, such as AdaBoost, SVM, Metric Learning and so on. They tend to have better performance than unsupervised ones and need less time to test an image. However, most of

supervised approaches project a pair of pedestrians into the feature difference space or the absolute feature difference space at first and it isn't an optimal projection for pedestrian re-identification since they neglect the appearance of each individual. Our approach is a supervised one. Different from previous work, we learn a quadratic similarity function to strengthen the connection between the similarity function and each individual appearance of a pair of pedestrians. So that the similarity function can deal better with large intra-class appearance variations. The performance of our approaches outperforms the previous ones.

This work is supported in part by the National Science Foundation of China under Grant No.61075026 (Visual Content Extraction and Analysis of Pedestrian Video), the National Basic Research Program of China under Grant No.2011CB302203 (Intelligent Processing of Internet Visual Media).

In the past, our lab explores some efficient approaches on the fundamental problems of visual content extraction and analysis of pedestrian video, such as detection, tracking, segmentation and pose estimation on pedestrian. This work is also a fundamental problem for pedestrian video analysis. In the future, we will build a prototype system of pedestrian retrieval in visual surveillance by jointing above fundamental approaches.