

基于局部最大概率特征和映射模型学习的行人再识别^①

胡正平^② 张敏姣 李淑芳 任大伟

(燕山大学信息科学与工程学院 秦皇岛 066004)

摘 要 进行了行人再识别研究。针对行人自身差异和相机视域差异的存在导致再识别率低的问题,提出了一种基于局部最大概率特征和映射模型学习的行人再识别算法。该算法首先对行人图像提取局部最大概率特征,克服光照变化并提取图像的完整信息;然后学习交叉映射模型,利用学习好的模型进行行人特征变换,从而消除不同摄像机拍摄区域的特征差异;最后进行距离度量和排序。实验表明,该算法合理有效,能够获得较为完整的判别性特征表示,成功地提高了行人再识别的匹配精度。

关键词 行人再识别, 映射模型学习, 局部最大概率特征

0 引 言

随着视频监控设备的广泛应用,对行人再识别 (person re-identification) 技术的研究,显示出了其在社会公共安全上的重要现实意义。行人再识别技术有两个核心任务,图像描述和距离度量,目前已有的行人再识别算法大致可分为基于图像描述的方法和基于距离度量学习的方法。在图像描述中,最早人们使用手动操作提取行人描述符,其中行人描述最常用的是颜色特征。由 Zhao^[1] 等人提出 32 维的 LAB 颜色直方图,以及 128 维的尺度不变特征变换 (scale invariant feature transform, SIFT) 描述符,通过步长为 5 个像素的滑动子窗口在样本中密集采样获得 10 × 10 的结构块,从每个结构块中提取特征。文献[2]对每个局部结构块提取 11 维颜色名称描述符,并且通过一个词袋 (bag-of-words, BoW) 模型,将描述符集合成为一个全局向量。还有基于属性特征的行人描述符,属性特征是中层视觉特征,例如文献[3]提出将来自不同摄像机的同一个行人的二进制语义属性嵌入到一个连续的低秩属性空间。手动

提取特征后,合适的距离度量方法对行人成功匹配至关重要,最常用的方式是基于马氏距离学习距离度量函数,通过线性缩放和特征空间旋转该方程式能够涵盖欧式距离,Weinberger^[4] 等基于此方式一些经典的度量学习方法引入,提出了大间隔最近邻居 (large margin nearest neighbor, LMNN) 分类算法,该算法给邻近目标 (匹配对) 设定一个阈值,并且惩罚不同类之间的近距离,其属于有监督的局部距离度量学习模型。除使用马氏距离方法以外,一些研究者还使用其他的学习算法,例如支持向量机 (support vector machine, SVM)。文献[5]对每个训练身份学习一个特定的 SVM,并将每个测试图像通过其视觉特征映射成一个权重向量。另外一些研究从学习具有区分性的子空间角度展开。Liao^[6] 等人提出通过交叉视图数据,学习投影到一个低维的子空间上,以相同的方式解决线性判别分析 (linear discriminant analysis, LDA),在学习的子空间中,通过使用简单而且直接的量度学习 (keep it simple and straight metric learning, KISSME) 算法学习一个距离函数。为达到降维目的, Pedagadi^[7] 等人结合无监督的主成分分析 (principal component analysis,

① 国家自然科学基金 (61071199) 和河北省自然科学基金 (F2016203422) 资助项目。
② 男,1970 年生,博士,教授;研究方向:模式识别;联系人,E-mail: hzp@ysu.edu.cn
(收稿日期:2017-09-11)

PCA)和有监督的局部 Fisher 判别分析进行度量学习,从而保留了局部邻域结构。特征提取之后特征维度一般很高,因此 KISSME 之前需要对高维特征进行降维,而文献[8]提出了成对约束分量分析(pairwise constrained component analysis, PCCA)算法,用以学习一个线性映射函数,进而能够直接在高维数据上操作。文献[9]进一步将目前两个子空间投影方法进行改进,即正则化的成对约束分量分析(regularized pairwise constrained component analysis, rPCCA)^[8]和核局部费舍尔判别分析(kernel local Fisher discriminant analysis, kLFDA)^[7]。

深度学习近几年在计算机视觉领域得到了广泛的研究和应用,从 2014 年开始不少研究者提出了基于深度学习进行行人再识别的算法。Ahmed 等在文献[10]中提出了一种深层卷积神经网络(CNN)结构,该网络结构可以自动学习特征和距离度量函数,并且学习过程是同时进行的。然而基于深度学习方法,内部原理也隐含包含了特征提取和距离函数学习两个关键步骤,由于参数量大,需要大的训练数据库以及相当长的训练时间,对于不同的数据库需要学习不同的网络结构参数,移植性较差,这些都是深度模型存在的问题,并且训练时涉及到的调参过程是需要专业理论作为指导的,具有较大的主观性问题。总之,这一切都需要大的数据量作为研究基础。

可以看出,目前大多数行人再识别算法难以解决图像中的光照变化从而提取完整的判别性特征,针对这个问题本文研究了基于局部最大概率特征与映射模型学习的行人再识别方法。首先对行人图像提取局部最大概率特征^[6],从而克服光照变化提取图像完整信息,然后学习映射模型,利用学习好的模型进行行人特征变换,从而消除不同摄像机拍摄区域行人的特征差异,最后进行距离度量和排序。

1 基于局部最大概率特征和映射模型学习的行人再识别

1.1 局部最大概率特征提取

颜色是描述行人图像的重要视觉特征,不同摄

像机之间的光照条件差异较大,且每个摄像机的设置也不尽相同,因此在不同摄像机视域内对同一行人的颜色感知可能差别较大。针对此问题,首先应用 Retinex 算法对图像进行预处理^[11]。其中 Retinex 是人类调节视觉颜色和亮度的感知模型,其可以生成与人类眼睛所视的场景保持一致的彩色图像。然后对 Retinex 预处理后的行人图像进行 HSV 颜色直方图特征提取。

在不同摄像机下的行人通常以不同的角度出现,为解决此问题,这里将一个行人图像竖直方向平均分成 6 个水平条,然后在每个水平条上单独计算直方图。并使用滑动子窗口来描述行人图像的局部细节,子窗口的步长为 5,这样就可以得到行人图像局部图像块,在得到的每一个局部图像块中进行 HSV 和 SILTP 直方图特征提取。为解决角度变化问题,将同一水平位置上的所有的局部图像块提取的特征进行如图 1 所示的局部最大化操作,生成局部最大概率特征直方图。

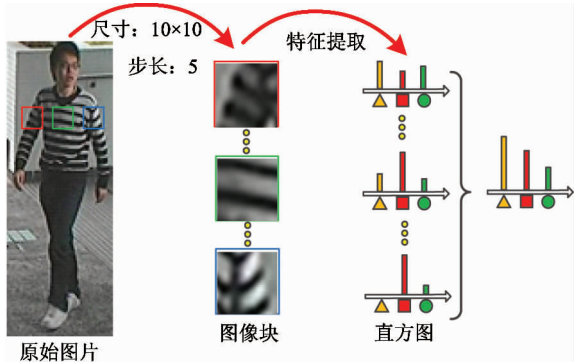


图 1 局部最大概率特征提取

考虑到各尺度之间的信息,需要构建三个尺度金字塔来表示,首先通过两次局部平均池化操作对行人原图像进行下采样,然后重复上述特征提取步骤,将得到的所有特征级联,同时对直方图特征值进行对数变换从而抑制直方图特征中较大的值,最后统一将 HSV 直方图特征和 SILTP 直方图特征归一化为单位长度,得到局部最大概率特征。如此过程后得到的局部最大概率特征能较好地处理光照变化和角度变化。

1.2 映射模型学习

1.2.1 不同视域特征差异

交叉视图映射模型用来寻找一个潜在的公共空间,在这个空间中,不同摄像机视域下的同一行人的特征更加相似,与此同时不同行人的特征差异更大。图2所示为交叉视图映射模型,假设 $\mathbf{X}^k = [\mathbf{x}_1^k, \mathbf{x}_2^k, \mathbf{x}_3^k, \dots, \mathbf{x}_{n^k}^k] \in \mathbb{R}^{d \times n^k}$ 表示从第 k 个视域内捕获的行人图像提取的特征矩阵, d 表示这个特征维数, n^k 是第 k 个视域内的样本个数,则数学模型如下式所

示:

$$d(\{\mathbf{x}_i^p, p\}, \{\mathbf{x}_j^q, q\}) = \|\mathbf{U}^{pT} \mathbf{x}_i^p - \mathbf{U}^{qT} \mathbf{x}_j^q\|_2 \quad (1)$$

其中 p 和 q 表示两个不同摄像机拍摄的视域,通常情况下 $\mathbf{U}^p \neq \mathbf{U}^q$ 。从本质上讲,构建的模型是通过学习矩阵 $\mathbf{U}^p$ 和 $\mathbf{U}^q$ 得到,此模型称为交叉视图映射模型(图2)。

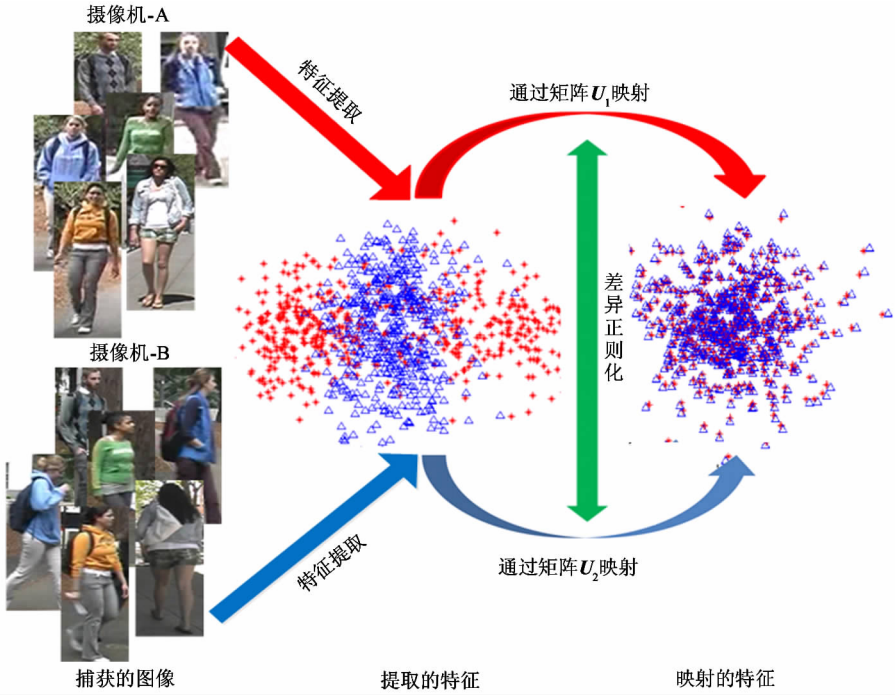


图2 交叉视图映射模型示意图

1.2.2 视图特性变换降低差异

应用映射模型对每一个摄像机视域学习特征转换,其中 $\mathbf{U}^p = [\mathbf{u}_1^p, \mathbf{u}_2^p, \dots, \mathbf{u}_C^p]$ 表示视域 p 的映射矩阵, $p = 1, 2, \dots, N$; C 表示映射空间的维数。其主要目的是学习映射矩阵 $\mathbf{U}^p$, 从而将特征 $\mathbf{X}^p$ 嵌入到具有判别力的共同欧式空间中,在这个空间中正样本对距离更近、负样本对距离更远。然后学习一个潜在的共同空间能够表示交叉视图(cross-view)样本对和内部视图(intra-view)样本对的关系模型,此关系模型由交叉视图模型和内部视图模型组成:

$$f = f_{\text{cross}} + \eta f_{\text{intra}} \quad (2)$$

η 表示控制内部视图模型 f_{intra} 的权重,其为正整数。交叉视图模型 f_{cross} 和内部视图模型 f_{intra} 由以下等式决定:

$$f_{\text{cross}} = \sum_{p=1}^{N-1} \sum_{q=p+1}^N \sum_{i=1}^{n^p} \sum_{j=1}^{n^q} \mathbf{W}_{ij}^{p,q} \|\mathbf{U}^{pT} \mathbf{x}_i^p - \mathbf{U}^{qT} \mathbf{x}_j^q\|_2^2 \quad (3)$$

$$f_{\text{intra}} = \sum_{p=1}^N \sum_{i=1}^{n^p} \sum_{j=1}^{n^p} \mathbf{W}_{ij}^{p,p} \|\mathbf{U}^{pT} \mathbf{x}_i^p - \mathbf{U}^{pT} \mathbf{x}_j^p\|_2^2 \quad (4)$$

在上述的两个模型中, $\mathbf{W}_{ij}^{p,q}$ 表示视域 p 与视域 q 之间的每一对样本的权重, $\mathbf{U}^p$ 为视域 p 的映射矩阵, $\mathbf{W}_{ij}^{p,q}$ 定义如下:

$$\mathbf{W}_{ij}^{p,q} = \begin{cases} \frac{1}{n_{\text{pos}}^{p,q}} \mathbf{A}_{ij}^{p,q}, & (\mathbf{x}_i^p, \mathbf{x}_j^q) \in \xi^{p,q} \\ -\gamma \frac{1}{n_{\text{neg}}^{p,q}}, & \text{其他} \end{cases} \quad (5)$$

$\mathbf{A}_{ij}^{p,q}$ 可以设置为局部加权项,或简单设置为1。 $n_{\text{pos}}^{p,q}$ 和 $n_{\text{neg}}^{p,q}$ 各自表示为区域 p 与区域 q 之间的正样

本对和负样本对的个数, γ 是一个标量。因为正样本对的个数远远小于负样本对的个数, 所以使用 $\frac{1}{n_{\text{pos}}^{p,q}}$ 和 $\frac{1}{n_{\text{neg}}^{p,q}}$ 分别对交叉视图模型和内部视图模型进行标准化, 这样 γ 可以控制交叉视图模型和内部视图模型的权重, 最小化目标函数 f 将减小类内不同, 同时增大类间不同。当 $p \neq q$ 时, $\mathbf{W}^{p,q}$ 描述交叉视图关系, 当 $p = q$ 时, $\mathbf{W}^{p,q}$ 表示内部视图关系。

为避免平凡解, 即 $\mathbf{U}^k = \mathbf{0}$, 其 $k = 1, 2, \dots, N$, 加上约束条件并制定以下优化问题:

$$\begin{aligned} \min_{\mathbf{U}^1, \mathbf{U}^2, \dots, \mathbf{U}^N} & \sum_{p=1}^{N-1} \sum_{q=p+1}^N \sum_{i=1}^{n^p} \sum_{j=1}^{n^q} \mathbf{W}_{ij}^{p,q} \| \mathbf{U}^{pT} \mathbf{x}_i^p - \mathbf{U}^{qT} \mathbf{x}_j^q \|^2 \\ & + \sum_{p=1}^N \sum_{i=1}^{n^p} \sum_{j=1}^{n^p} \mathbf{W}_{ij}^{p,p} \| \mathbf{U}^{pT} \mathbf{x}_i^p - \mathbf{U}^{pT} \mathbf{x}_j^p \|^2 \\ \text{s. t. } & \mathbf{U}^{kT} \mathbf{M}^k \mathbf{U}^k = \mathbf{I}, k = 1, 2, \dots, N \end{aligned} \quad (6)$$

其中 $\mathbf{M}^k = \mathbf{X}^k \mathbf{X}^{kT} + \mu \mathbf{I}$, $\mathbf{I}$ 表示单位矩阵, 从而可避免协方差矩阵的奇异性。

1.2.3 核扩展

利用一致性正则化方法学习线性映射矩阵来进行特征变换, 而在实际应用中给定的数据常常是非线性关系, 这里通过核扩展方法来解决。第 k 个视域中隐含的高维子空间基可表示为: $\tilde{\mathbf{X}} \boldsymbol{\alpha}^k$, 其中 $\tilde{\mathbf{X}}$ 表示所有训练数据的高维特征矩阵。因此, 映射特征可以表示为

$$h^k(\tilde{\mathbf{x}}^k) = \boldsymbol{\alpha}^{kT} \tilde{\mathbf{X}}^T \tilde{\mathbf{x}}^k = \boldsymbol{\alpha}^{kT} \mathbf{k}(X, \mathbf{x}^k) \quad (7)$$

$\mathbf{k}(X, \mathbf{x}) = [k(X_1, \mathbf{x}), \dots, k(X_n, \mathbf{x})]^T$, 其中 $k(\cdot, \cdot)$ 是核函数, $h^k(\cdot)$ 是第 k 视域的映射函数, n 是训练的样本个数。用 $\boldsymbol{\alpha}^p, \boldsymbol{\alpha}^q, \mathbf{k}(X, \mathbf{x}_i^p)$ 和 $\mathbf{k}(X, \mathbf{x}_j^q)$ 分别替换 $\mathbf{U}^p, \mathbf{U}^q, \mathbf{x}_i^p$ 和 $\mathbf{x}_j^q$, 并将式(7)带入到式(2), 得到核扩展后的损失函数。

利用复制内核希尔伯特空间的复制属性 $\langle k(\cdot, \mathbf{x}), k(\cdot, \mathbf{y}) \rangle = k(\mathbf{x}, \mathbf{y})$, 在隐含的高维空间里, 正则项则可以进一步表示为 $\sum_{k=1}^N \boldsymbol{\alpha}^{kT} \mathbf{K} \boldsymbol{\alpha}^k - \sum_{p=1}^{N-1} \sum_{q=p+1}^N \boldsymbol{\alpha}^{pT} \mathbf{K} \boldsymbol{\alpha}^q$, 其中 $\mathbf{K}$ 是由格拉姆矩阵 (Gram Matrices) 定义的, $\mathbf{K} = [k(X, X_1), k(X, X_2), \dots, k(X, X_n), \dots]$, 核扩展优化算法如下:

$$\min_{\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \dots, \boldsymbol{\alpha}^N} \sum_{p=1}^{N-1} \sum_{q=p+1}^N \sum_{i=1}^{n^p} \sum_{j=1}^{n^q} \mathbf{W}_{ij}^{p,q} \| \boldsymbol{\alpha}^{pT} \mathbf{k}(X, \mathbf{x}_i^p) - \boldsymbol{\alpha}^{qT} \mathbf{k}(X, \mathbf{x}_j^q) \|^2$$

$$\begin{aligned} & + \sum_{p=1}^N \sum_{i=1}^{n^p} \sum_{j=1}^{n^p} \mathbf{W}_{ij}^{p,p} \| \boldsymbol{\alpha}^{pT} \mathbf{k}(X, \mathbf{x}_i^p) - \boldsymbol{\alpha}^{pT} \mathbf{k}(X, \mathbf{x}_j^p) \|^2 \\ & + \text{tr}(\lambda \sum_{k=1}^N \boldsymbol{\alpha}^{kT} \mathbf{K} \boldsymbol{\alpha}^k - 2\lambda' \sum_{p=1}^{N-1} \sum_{q=p+1}^N \boldsymbol{\alpha}^{pT} \mathbf{K} \boldsymbol{\alpha}^q) \\ \text{s. t. } & \boldsymbol{\alpha}^{kT} \mathbf{M}^k \boldsymbol{\alpha}^k = \mathbf{I}; k = 1, 2, \dots, N \end{aligned} \quad (8)$$

在上述公式中 $\mathbf{M}^k = \mathbf{K}^k \mathbf{K}^{kT} + \lambda \mathbf{K}$, $\mathbf{K}^k = [k(X, \mathbf{x}_1^k), k(X, \mathbf{x}_2^k), \dots, k(X, \mathbf{x}_{n_k}^k)]$ 。

1.2.4 优化求解

线性情况和非线性情况的优化求解过程是相似的, 以线性情况为例进行优化求解。待优化的目标函数可以用以下形式来表示:

$$\begin{aligned} f = \text{tr} & \left(\sum_{p=1}^{N-1} \sum_{q=p+1}^N \mathbf{U}^{pT} \mathbf{H}^{p,q} \mathbf{U}^p + \mathbf{U}^{qT} \mathbf{H}^{q,p} \mathbf{U}^q \right. \\ & \left. - 2\mathbf{U}^{pT} \mathbf{R}^{p,q} \mathbf{U}^q + \lambda \sum_{k=1}^N \mathbf{U}^{kT} \mathbf{U}^k \right) \end{aligned} \quad (9)$$

其中 $\mathbf{H}^{p,q} = \mathbf{X}^p (\mathbf{D}^{p,q} + \eta \mathbf{D}^{p,p} - \eta \mathbf{W}^{p,p}) \mathbf{X}^{pT}$, $\mathbf{D}^{p,q}$ 表示一个对角矩阵其对角元素被定义为 $\mathbf{D}_{ii}^{p,q} = \sum_{j=1}^n \mathbf{W}_{ij}^{p,q}$, $\mathbf{R}^{p,q} = \mathbf{X}^p \mathbf{W}^{p,q} \mathbf{X}^{qT} + \lambda' \mathbf{I}$ 。

将目标函数进一步简化为

$$f = \text{tr}(\mathbf{U}^T \mathbf{R} \mathbf{U}) \quad (10)$$

$$\mathbf{U} = [\mathbf{U}^1; \mathbf{U}^2; \dots; \mathbf{U}^N] \in \mathbb{R}^{Nd \times C} \quad (11)$$

$$\mathbf{R} = \begin{pmatrix} \mathbf{G}^1 & -\mathbf{R} & \dots & -\mathbf{R}^{1,N} \\ -\mathbf{R}^{2,1} & \mathbf{G}^2 & \dots & -\mathbf{R}^{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{R}^{N,1} & -\mathbf{R}^{N,2} & \dots & \mathbf{G}^N \end{pmatrix} \quad (12)$$

其中 $\mathbf{G}^k = \sum_{q \neq k} \mathbf{H}^{k,q} + \lambda \mathbf{I}$ 。这是松弛约束, $\mathbf{U}^{kT} \mathbf{M}^k \mathbf{U}^k = \mathbf{I}$, $k = 1, 2, 3, \dots, N$, $\sum_{k=1}^N \mathbf{U}^{kT} \mathbf{M}^k \mathbf{U}^{kT} = \mathbf{N} \mathbf{I}$, 优化问题可以修改为

$$\min_{\mathbf{U}} \text{tr}(\mathbf{U}^T \mathbf{R} \mathbf{U}) \quad (13)$$

$$\text{s. t. } \mathbf{U}^T \mathbf{M} \mathbf{U} = \mathbf{N} \mathbf{I}$$

上式中 $\mathbf{M} = \text{diag}(\mathbf{M}^1, \mathbf{M}^2, \dots, \mathbf{M}^N)$ 。式(9)表示的优化问题可以通过计算相应的最小特征值的特征向量解决:

$$\mathbf{R} \mathbf{u} = v \mathbf{M} \mathbf{u} \quad (14)$$

其中 v 为拉格朗日乘子, 在得到特征向量 $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_c$ 后, 对第 p 视域的第 c 个特征向量进行变换:

$$\mathbf{u}_c^p = \frac{\delta_p(\mathbf{u}_c)}{\|\delta_p(\mathbf{u}_c)\|_M}, \text{ 其中 } \delta_p(\cdot) \text{ 表示得到第 } p \text{ 个子}$$

向量和 $\|\mathbf{v}\|_M = \sqrt{\mathbf{v}^T \mathbf{M} \mathbf{v}}$ 。通过 $\mathbf{U}$ 代替 $\boldsymbol{\alpha}, \mathbf{M}^k$ 代替

$M^k, K^p D^{p,q} K^{qT} + \lambda' K$ 代替 $R^{p,q}, K^p (D^{p,q} + \eta D^{p,p} - \eta W^{p,p}) K^{pT}$ 代替 $H^{p,q}$ 。核扩展方案的解决方案可以通过解式(14)求得。

经过以上步骤,学习视域 p 和 q 的两个映射矩阵 U^p 和 U^q 从而得到映射模型,此模型可以寻找到一个潜在的公共空间,在这个空间中,不同摄像机视域下的同一行人的特征更加相似,而不同行人的特征差异更大。该模型应用在不相交的摄像机视域下的行人图像匹配中,能够保留更多判别性特征表示。

2 实验仿真研究

本实验在 VIPeR、PRID450S 和 CUHK01 三个数据库上进行,并在所有数据集上采用累积匹配特性(cumulative match characteristic, CMC)曲线及其匹配率排名对包括本算法在内的 9 种算法进行性能对比评估,其它 8 种算法分别为:局部费舍尔判别分析(local Fisher discriminant analysis, LFDA)^[7]、成对约束成分分析(PCCA)^[8]、支持向量机距离度量(support

vector machine metric learning, SVMML)^[12]、简单而且直接的度量学习(KISSME)^[13]、正则化成对约束成分分析(rPCCA)^[8]、核局部费舍尔判别分析(kLFDA)^[7]、边界费舍尔分析(marginal Fisher analysis, MFA)^[14]、内核交叉视图判别成分分析(kernel cross-view discriminant component analysis, KCVD-CA)^[15]。实验平台为处理器 AMD Athlon(tm) II X2 255 Processor 3.10GHz,内存 10.0G 以及 Windows7、64 位操作系统。

2.1 VIPeR 数据库

在行人再识别研究领域 VIPeR 数据库是最早被公开、应用广泛的行人再识别数据集。该数据集行人目标是从户外拍摄的原始视频序列中手动准确地裁剪出来的,共包含 632 个行人目标,每个行人目标包含两张图像,整个数据集共有 1264 张图像,其中每张图像的大小归一化为 128×48 ,其中存在大量光照变化的图像,有些图像光照变化还比较剧烈,如图 3 所示,其同一列的行人图像属于同一行人身份。



图 3 VIPeR 数据库中光照变化的行人图像部分示例

在进行实验时,将数据集分成两组:训练集和测试集,其中测试集包括查寻集和候选集,为与已有算法作公正对比,用先前工作普遍采用的评价框架,将数据集随机平均分成两部分,其中一部分作为实验训练集,另一部分作为测试集。

图 3 列出数据集中的部分样本,此后,在此数据库中利用 CMC 曲线对 9 种算法进行性能比较,其评估结果如图 4 和表 1 所示。因此证明本文算法能较好地应对光照变化和信息的完整提取,具有比较好的性能。

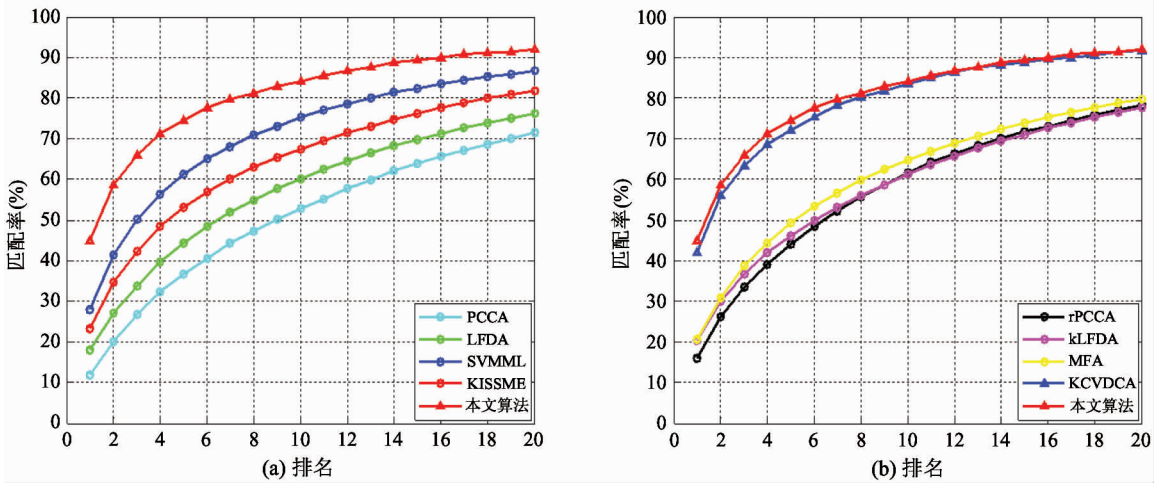


图 4 VIPeR 数据库上的 CMC 曲线比较

表 1 VIPeR 数据库上几种算法匹配率比较				
算法	$r = 1$	$r = 5$	$r = 10$	$r = 20$
PCCA ^[8]	11.98%	36.71%	52.10%	71.41%
LFDA ^[7]	17.96%	44.38%	60.09%	76.17%
SVMML ^[12]	28.03%	61.20%	75.16%	86.65%
KISSME ^[13]	23.35%	52.95%	67.51%	81.78%
rPCCA ^[8]	16.05%	44.09%	61.67%	78.24%
kLFDA ^[7]	20.24%	46.18%	61.16%	77.58%
MFA ^[14]	20.63%	49.36%	64.79%	79.79%
KCVDCA ^[15]	41.99%	72.15%	83.58%	91.78%
本文算法	44.91%	74.34%	84.21%	91.87%

2.2 PRID450S 数据库

PRID450S 数据库中同一个行人目标的两张图像存在强烈的光照变化,每个图像尺寸不相同,本文在进行实验时,将行人图像归一化为 160×60 。由于每个行人光照变化较大,相同行人的图像存在显著区别,实验数据分组与上节相同。

图 5 列出本数据集中的部分样本,此后,在此数据库中利用 CMC 曲线对 9 种算法进行性能比较,其评估结果如图 6 和表 2 所示。发现本算法能较好地应对光照变化和提取完整判别性信息,证明该算法的合理性和有效性,对于解决光照变化问题效果显著。



图 5 PRID450S 上光照变化的行人图像部分示例

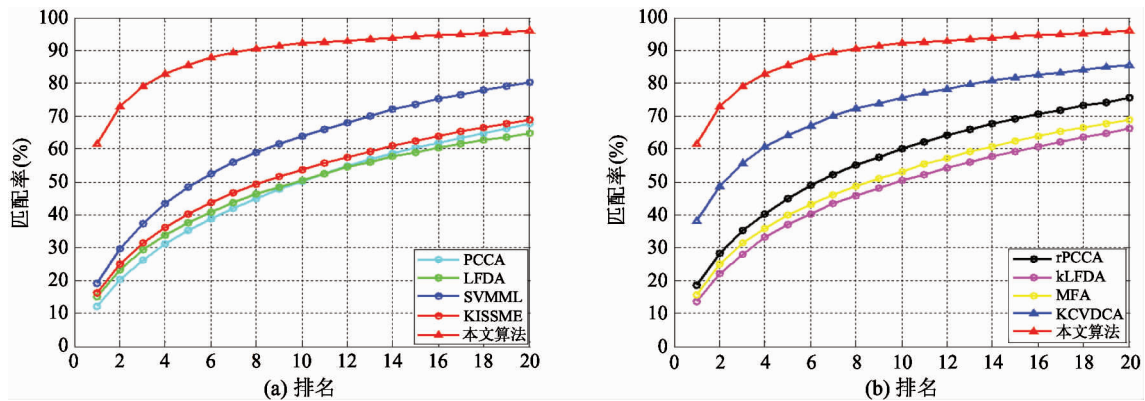


图 6 PRID450S 数据库上的 CMC 曲线比较

表 2 PRID450S 数据库上几种算法匹配率比较

算法	$r = 1$	$r = 5$	$r = 10$	$r = 20$
PCCA	12.18%	35.16%	50.20%	67.56%
LFDA	15.31%	37.01%	50.10%	64.40%
SVMML	19.22%	48.25%	63.92%	80.16%
KISSME	16.24%	40.30%	53.61%	68.77%
rPCCA	18.54%	45.02%	60.03%	75.49%
kLFDA	13.57%	37.11%	50.03%	66.14%
MFA	15.77%	39.85%	53.17%	68.90%
KCVDCa	38.04%	64.27%	75.60%	85.60%
本文算法	61.47%	85.60%	92.18%	95.91%

2.3 CUHK01 数据库

CUHK01 数据库中有 971 个行人,该数据集每个摄像机视域中包含同一行人的两张图像,共有 3884 张图像,图像归一化为 160×60 , 本文研究的是单发情况下的行人再识别,所以在每个摄像机视域下仅选取同一行人的一张图像来进行实验,即本

实验随机选取 1942 张行人图像。如图 7 所示 CUHK01 数据库中光照发生变化的行人图像示例。

图 7 列出数据集中的部分样本,此后,在此数据库中利用 CMC 曲线对 9 种算法进行性能比较,其评估结果分别如图 8 和表 3 所示。因此可以验证本算法也能较好地应对光照变化和信息的完整提取问题。

通过在 3 个公共数据集上的对比实验可知,局部最大概率特征经过映射后算法能够在 VIPeR、PRID450S、CUHK01 数据集上实现显著的性能提升。并且值得一提的是,该算法的 Rank 1 与其他算法相比均实现了较大的提升,第一匹配率表示的是模型的真正识别能力,这在实际应用中是很重要的优势之一,例如在刑事侦查这样典型的行人再识别技术应用场景中,Rank1 越高则在刑事侦查活动中就可以越快地锁定嫌疑人的行踪路线。

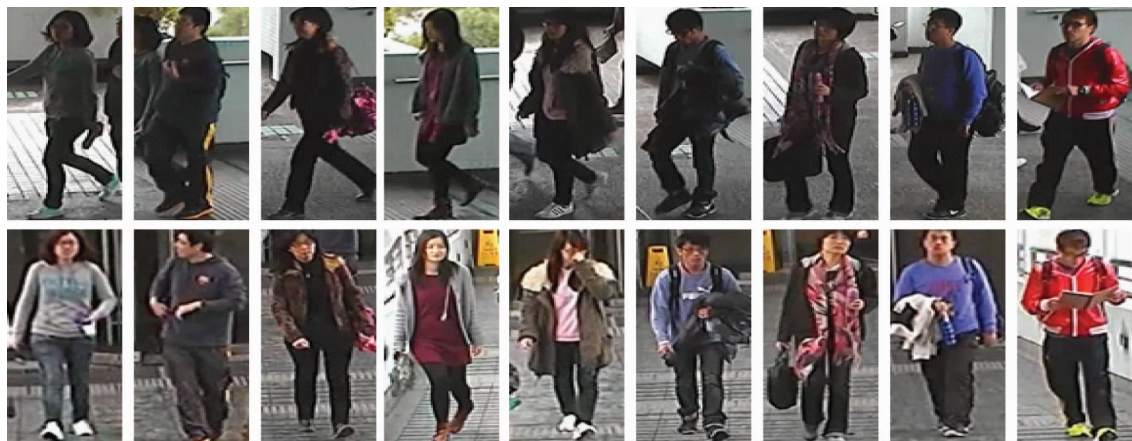


图 7 CUHK01 中光照变化的行人图像部分示例

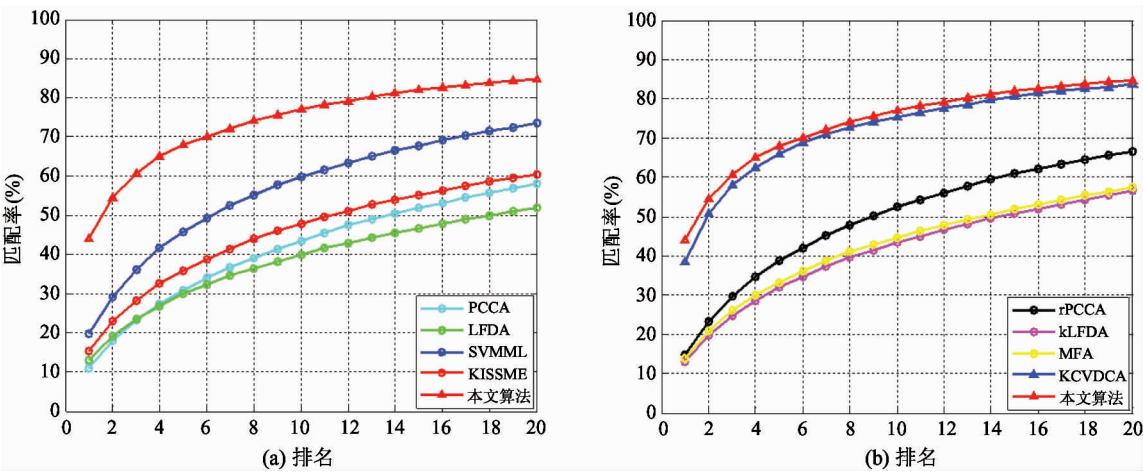


图 8 CUHK01 数据库上的 CMC 曲线比较

表 3 CUHK01 数据库上几种算法匹配率比较

算法	$r = 1$	$r = 5$	$r = 10$	$r = 20$
PCCA	10.99%	30.87%	43.49%	57.93%
LFDA	13.11%	29.83%	39.96%	51.86%
SVMML	19.66%	45.77%	59.69%	73.42%
KISSME	15.49%	35.84%	47.90%	60.48%
rPCCA	14.71%	38.69%	52.34%	66.65%
kLFDA	13.07%	31.87%	43.29%	56.44%
MFA	14.00%	33.07%	44.61%	57.48%
KCVDCA	38.43%	66.05%	75.37%	83.68%
本文算法	43.97%	67.98%	77.10%	84.77%

实际中获得好性能的前提是有足够多的训练样本,但在距离度量学习的方法中,训练数据需要人工标定,获取成本比较高。由于实验需要,本文选定的数据库规模比较小,包含的行人图像较少。而本文其他的距离度量方法例如 KISSME 方法需要估计高斯分布的模型参数,这在训练数据不足的情况下很难估计出准确的参数组合;kLFDA 算法的距离函数优化过程需要耗费较长的时间,且在只有少量训练样本的情况下此算法容易产生较为严重的过拟合问题。相比之下,本算法在数据库较小的情况下将特征进行映射后再进行距离度量学习,能够使最终得到的相似度量函数具有更高的区分性,得到优秀的识别效果。另外本算法的特征提取计算量并不大,距离度量学习针对训练样本来说虽然计算量不小,但是一旦学习完毕,对于测试样本来说距离度量是固定的,因此速度是比较快的。

3 结论

本文针对剧烈光照变化和不能提取完整判别性信息问题,研究局部最大概率特征和映射模型学习相结合的算法,表明本算法对光照变化具有鲁棒性并能提取完整判别性信息,可显著地提高匹配率,对提高行人再识别算法性能具有优越性。

但值得注意的是,学习映射模型时,需要针对特定摄像机配对来进行学习,需要对大量摄像机进行配对处理,处理数据大大增加,如何学习一个具有良好推广能力的映射模型方法同时又能克服不同摄像机下行特征分布不一致问题值得进一步深入研究。并且由于本算法主要目的是解决行人再识别过程中存在的背景变化和相机视域差异等问题,所以选用的是视域差异较大且背景较为繁乱的数据库,但是这 3 个数据库规模均比较小,训练复杂深度神经网络结构具有一定困难,因此本文算法没有采用深度学习方法进行研究,但在后续的基于视频的行人再识别研究中,将会展开深度学习的视频行人再识别研究。

参考文献

[1] Zhao R, Ouyang W L, Wang X G. Learning mid-level filters for person re-identification[C]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014. 144-151

[2] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification; a benchmark [C]. In: Proceedings of

- IEEE International Conference on Computer Vision, Santiago, Chile, 2015. 1116-1124
- [3] Su C, Yang F, Zhang S L, et al. Multi-task learning with low rank attribute embedding for person re-identification [C]. In: Proceedings of IEEE International Conference on Computer Vision, Santiago, Chile, 2015. 3739-3747
 - [4] Weinberger K Q, Saul L K. Distance metric learning for large margin nearest neighbor classification[J]. *Journal of Machine Learning Research*, 2009, 10(1): 207-244
 - [5] Zhang Y, Li B H, Lu H C, et al. Sample-specific SVM learning for person re-identification[C]. In: Proceedings of IEEE Computer Vision and Pattern Recognition, Nevada, USA, 2016. 1278-1287
 - [6] Liao S C, Hu Y, Zhu X Y, et al. Person re-identification by local maximal occurrence representation and metric learning[C]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015. 2197-2206
 - [7] Pedagadi S, Orwell J, Velastin S, et al. Local fisher discriminant analysis for pedestrian re-identification [C]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Portland, USA, 2013. 3318-3325
 - [8] Mignon A, Jurie F. Pcca: a new approach for distance learning from sparse pairwise constraints[C]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Rhode Island, USA, 2012. 2666-2672
 - [9] Xiong F, Gou M, Camps O, et al. Person re-identification using kernel-based metric learning methods[C]. In: Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 2014. 1-16
 - [10] Ahmed E, Jones M, Marks T K. An improved deep learning architecture for person re-identification[C]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015. 3908-3916
 - [11] Park S, Moon B, Yu S, et al. Low-light image enhancement using variational optimization-based retinex model [C]. In: Proceedings of IEEE International Conference on Consumer Electronics, Las Vegas, USA, 2017. 70-71
 - [12] Li Z, Chang S Y, Liang F, et al. Learning locally-adaptive decision functions for person verification [C]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Portland, USA, 2013. 3610-3617
 - [13] Koestinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints [C]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Rhode Island, USA, 2012. 2288-2295
 - [14] Yan S C, Xu D, Zhang B Y, et al. Graph embedding and extensions: a general framework for dimensionality reduction[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(1): 40-51
 - [15] Chen Y C, Zheng W S, Lai J H, et al. An asymmetric distance model for cross-view feature mapping in person re-identification[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017, 27(8): 1661-1675

Person re-identification based on local maximal occurrence feature and mapping model learning

Hu Zhengping, Zhang Minjiao, Li Shufang, Ren Dawei

(School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004)

Abstract

The technique of person re-identification is studied. Aiming to solve the problem of low re-identification rate caused by pedestrians' own difference and the difference of camera view, a person re-identification algorithm based on local maximal occurrence features and the mapping model is proposed. Firstly, the algorithm extracts the local maximal occurrence features from pedestrians' images, so as to overcome illumination changes and get complete image information. Secondly, it learns the cross-view mapping model to convert pedestrian features to eliminate characteristic differences of different camera views. Finally, the features will be sorted by distance metric learning. The experimental results show that the proposed person re-identification algorithm is effective enough to obtain a more complete discriminant feature representation so that it can improve the matching accuracy for person re-identification.

Key words: person re-identification, mapping model learning, local maximal occurrence feature