



研究与开发

引入垂直全局表观约束的中层特征行人再识别

王亚东, 张荣, 蒋桢慧, 郭立君

(宁波大学信息科学与工程学院, 浙江 宁波 315211)

摘要: 行人再识别问题中, 包含语义信息的中层特征能够提供更强的判别力。由于中层特征也采用局部匹配方式, 与底层特征一样存在由于不同行人部分表观区域比较相似而产生误匹配问题。考虑到行人几乎都处于站立姿态, 同一行人在垂直方向上的表观序列比不同行人的更相似, 提出了在中层特征的基础上引入行人垂直全局表观约束, 并融合底层稠密块匹配的识别方法。实验结果表明, 算法在最具挑战的公用 VIPeR 数据库和 CUHK01 数据库上, 均取得了比现有方法更高的命中率。

关键词: 行人再识别; 中层特征; 垂直表观约束

中图分类号: TP391

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017023

Person re-identification using mid-level features with vertical global appearance constraint

WANG Yadong, ZHANG Rong, JIANG Huihui, GUO Lijun

College of Information Science and Engineering, Ningbo University, Ningbo 315211, China

Abstract: Mid-level features with the semantic information can provide stronger discrimination in the person re-identification than low-level features. But like the low-level features, the mid-level features also use local feature matching methods and easily leads mismatch problem when different pedestrians have similar appearance features in some local areas. Considering the same pedestrians are more similar than different pedestrians in the vertical direction since pedestrians are almost always in a standing position, global vertical appearance constraints was introduced. Furtherly, a method for person re-identification was proposed which fuses the low-level densely patch-matching and the mid-level features with the global vertical appearance constraints. Experimental results show that the proposed algorithm can achieve a higher hit rate than the existing methods on the most challenging public VIPeR database and CUHK01 database.

Key words: person re-identification, mid-level feature, apparent constraint

收稿日期: 2016-11-25; 修回日期: 2017-01-06

通信作者: 张荣, zhangrong@nbu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61175026); 浙江省自然科学基金资助项目 (No.LY17F030002); “信息与通信工程”浙江省重中之重学科开放基金资助项目 (No.xkx11516, No.xkx11521); 宁波大学胡岚博士基金资助项目 (No.ZX2013000319)

Foundation Items: The National Natural Science Foundation of China(No.61175026), Zhejiang Provincial Natural Science Foundation of China(No.LY17F030002), Zhejiang Open Foundation of the Most Important Subjects(No.xkx11516, No.xkx11521), Dr.Hu Lan Foundation of Ningbo University(No.ZX2013000319)

1 引言

行人再识别^[1,2]是指在非重叠多摄像头监控系统中, 判别一个视频中出现的行人是否与另一个视频中的行人为同一个人。该技术在智能视频监控领域有着重要的应用, 但如今仍存在许多问题难以克服。如图 1 所示, 同一列图像中的行人为同一行人, 他们在不同摄像头中会受到如光照不同、视角变换、姿势变化、部分遮挡等因素的影响^[3], 导致同一行人的不同图像在视觉表现上差异较大, 在进行再识别时容易产生错误匹配。建立一个对上述影响因素具有顽健性且命中率高的行人再识别方法, 已成为该领域研究的一个热点。



图1 CUHK01 数据库中的行人图像

经过对近几年研究的分析, 本文所提方法主要考虑以下两个方面。

(1) 在现有的研究中, 底层特征在行人再识别领域已经取得了较好的效果。如参考文献[4-6]等利用底层稠密块匹配, 计算稠密块的突出性作为匹配的约束完成行人再识别, 在识别结果上相

比其他底层特征识别方法有了较大的提升。考虑到底层特征缺乏语义信息, 在行人再识别任务中无法兼顾准确性和不变性, 研究人员提出了基于中层特征的识别方法。中层特征在底层特征的基础上构建, 包含一定的语义信息, 具有兼顾准确性和局部改变不变性等特点。考虑到底层特征与中层特征各自的特点及互补性, 本文提出融合底层稠密块匹配及中层特征的方法, 以取得更好的识别效果。

(2) 由于底层特征与中层特征都采用局部特征匹配方式, 均存在由于不同行人部分表现区域比较相似而产生误匹配问题。通过对行人图像的观察, 可以发现行人几乎都处于站立姿势, 而且相同行人在垂直方向上的全局表现信息比不同行人的更加相似。如图 2 所示, 图 2(a)、图 2(b)为同一行人, 图 2(b)、图 2(c)为不同行人。这 3 张图像的部分区域比较相似, 如每幅图像中方框标记块。从图 2 中可以发现, 同一行人图 2(a)、图 2 (b)标记块各自对应的垂直表现信息(左侧一列)比不同行人图 2(b)、图 2(c)标记块各自对应的垂直表现信息更加相似。因此, 提出了引入垂直全局表现约束的中层特征行人再识别方法, 该方法在考虑两幅行人图像对中层特征匹配的同时, 引入了中层特征在各自行人图像中垂直方向上的全局表现约束, 并融合了取得不错效果的底层稠密块匹配方法, 从而有效地提高了再识别的命中率。



图2 稠密块表现信息序列



2 研究现状

最近目标行人再识别的工作主要集中在两个方面：特征的设计与选择、度量学习。

在特征的设计与选择方面，较早的方法是采用如颜色、纹理等全局特征直接进行再识别，但这类方法在视角及姿态变化的行人图像上识别效率并不是很高。因此，一些研究提出了将行人图像划分为块，利用块之间的相似性来进行匹配的方法。如 Farenzena 等人^[7]提出了累积对称局部特征（symmetry-driven accumulation of local feature, SDALF）的方法，主要利用了行人图像处理视图中的对称性变化。Cheng 等人^[8]利用图像结构来估计行人结构，并根据不同的身体部位来计算视觉特征，以应对姿势变化。参考文献[4]中将行人的图像平均划分为多个块并提取各块的颜色特征和 SIFT 特征，然后利用无监督的方法学习各块的显著性特征并进行识别。另外，Prosser 等人^[9]将行人再识别问题转化为排名问题，利用改进的 SVM 算法 RankSVM 来学习相似的行人对，也取得了不错的效果。但是，随着研究的深入，底层特征缺乏语义信息、判别能力比较弱等缺点逐渐显露出来，而近几年在一些视觉模型上中层特征得到了许多应用。如 Singh 等人^[10]和 Jain 等人^[11]通过块聚类、纯度测量和分数检测来学习中层特征用于场景检测和动作识别。考虑到中层特征出色的效果，最近也被用于行人再识别领域，Layne 等人^[12]提出了一种对中级语义属性的选择和加权的方法来描述行人。Song 等人^[13]利用行人的属性来修剪主题模型，并通过贝叶斯决策来进行行人匹配，然而学习行人的属性需要人工标注图像的属性标签，因为每个行人的图像可能有超过 50 个属性，所以标记匹配行人标签的代价是非常昂贵的。Zhao 等人^[14]提出一种自动学习具有判别性的中层过滤器的方法，这种方法不需要对行人属性进行标注，通过累计中层过滤器相应的匹配得分

进行再识别，并取得了显著的效果。但该方法中的行人匹配建立在局部图像块对每个中层过滤器响应上，由于在计算中层过滤器响应度的匹配得分时只考虑了人体的基本结构约束（如头部、躯干、腿部等），这会使得只考虑对应的局部区域内中层过滤器响应度匹配得分时，两个行人相似度很高。但实际上，相同区域内对同一过滤器得分相似的图像块不一定意味着是同一行人，还应该考虑这样的图像块在各自人体的全局表现约束上是否一致，即此图像在各自图像中垂直表现序列是否相似。同样地，参考文献[15,16]主要运用块或像素的约束关系来确定人体部位关系，依然会由于不同行人的部分表现比较相似而导致匹配得分较高，最终陷入局部匹配而产生误判。本文考虑到行人一般处于站立姿态，同一行人的不同图像在垂直方向上的全局表现信息较为相似，将在已取得良好效果的中层特征的基础上引入垂直全局表现信息，对原中层特征匹配得分的有效性予以约束，从而在一定程度上避免相似行人误匹配的情况发生。

在度量学习方面，Zheng 等人^[17]提出了概率相对距离比较（person reidentification as a relative distance comparison, PRDC）模型，该算法通过对同一个人的两幅图像计算的距离较小的概率大于不同人计算距离较小的概率来进行目标行人再识别。Li 等人^[18]开发了一个局部自适应决策函数（learning locally-adaptive decision functions, LADF）模型，通过联合距离度量和局部自适应阈值规则，取得了良好的效果。

3 引入表现约束的中层特征再识别

该部分将分为 3 个阶段介绍引入垂直全局表现约束的中层特征行人再识别方法。首先，介绍学习中层特征的过程，中层特征是利用底层稠密块构建的，主要包括选取判别能力和泛化能力较强的稠密块（以下简称有效块），对选取的有效块

聚类以及基于聚类结果训练中层过滤器3个步骤。其次,计算垂直全局表观约束,以对中层特征匹配得分的有效性予以约束。最后,利用加入垂直全局表观约束的中层特征匹配得分,融合底层稠密块匹配得分进行行人再识别。

3.1 学习中层过滤器

中层过滤器是在底层特征上构建的,通过聚类和训练所获得的能够描述一定视觉信息的过滤器。而行人图像在各中层过滤器的响应度将组成该行人的中层特征。以下将介绍如何学习中层过滤器并计算行人的中层特征。

3.1.1 有效块的选取

训练一个有效而又可靠的中层过滤器非常重要的一点是选取合理的样本,即选取具有辨别能力和泛化能力的稠密块。通过实验验证,这种有效块的一个特点是在图像库中拥有相似视觉信息的块不是太多也不是太少,因此能够代表一类行人所特有的视觉信息。为了选取这种有效块,将计算每个稠密块在图像库中的最近邻稠密块所组成的集合,并计算该稠密块与其最近邻稠密块所组成的集合中前 N_p 个稠密块的距离和。较小的距离和说明此稠密块在图像库中存在较多与其相似的稠密块,属于最普通、最常见的块,这种稠密块在行人再识别中的辨别能力较弱。相反较大的距离和说明此稠密块在图像库中存在较少与其相似的稠密块,尽管这样的稠密块由于稀少可能使其辨别能力较强,但这种稠密块的泛化能力较弱;合适的距离和说明此稠密块在图像库中存在部分与其相似的稠密块,拥有一定的辨别能力和泛化能力,能够有效地描述图像库中一组行人所共有的表观属性。因此,可以利用此距离和作为得分来表示稠密块的辨别能力和泛化能力,具体符号约定如下。

记 $X^{A,u} = \{x_{m,n}^{A,u}\}$ 为 A 摄像头下第 u 幅图像的稠密块特征集合,其中 $x_{m,n}^{A,u}$ 表示第 u 幅图像的 m 行

n 列的稠密块特征($m=1, \dots, M, n=1, \dots, N, u=1, \dots, U$)。同理有 $X^{B,v} = \{x_{i,j}^{B,v}\}$ 。那么,对于每一个稠密块 $x_{m,n}^{A,u}$ 在摄像头 B 下所有图像的最近邻块所组成的集合 $X_{NN}(x_{m,n}^{A,u})$ 可以表示为:

$$X_{NN}(x_{m,n}^{A,u}) = \{x | x = \arg \min_{x_{i,j}^{B,v}} d(x_{m,n}^{A,u}, x_{i,j}^{B,v}), x_{i,j}^{B,v} \in \partial(x_{m,n}^{A,u}, x_{i,j}^{B,v}), v=1, \dots, V\}, \quad (1)$$

$$\partial(x_{m,n}^{A,u}, x_{i,j}^{B,v}) = \{x_{i,j}^{B,v} | j=1, \dots, N, i = \max(0, m-h), \dots, \min(M, m+h)\} \quad (2)$$

其中, $d(x_{m,n}^{A,u}, x_{i,j}^{B,v})$ 表示 $x_{m,n}^{A,u}$ 和 $x_{i,j}^{B,v}$ 之间的欧氏距离。 $\partial(x_{m,n}^{A,u}, x_{i,j}^{B,v})$ 表示稠密块 $x_{m,n}^{A,u}$ 在 B 摄像头下第 v 幅图搜索匹配块的集合范围, V 表示 B 摄像头下图像的总个数。由于检测框会在人体区域有一定的空间浮动,为了减少空间变化对识别的影响,因此这里定义一个变量 h 来表示其在垂直方向上的浮动搜索范围。经实验验证^[4],当 h 设置过小时,由于垂直方向未对齐,稠密块便无法搜索到最佳匹配块;当 h 设置过大时,又可能会造成不同部位稠密块的误匹配。对实验结果对比分析,取 $h=2$ 。

那么,一个稠密块辨别能力和泛化能力的得分可以表示如下:

$$s(x_{m,n}^{A,u}) = \sum_{k=1}^{N_p} d_k(X_{NN}(x_{m,n}^{A,u})) \quad (3)$$

其中, $d_k(X_{NN}(x_{m,n}^{A,u}))$ 表示最近邻块所组成的集合中排名第 k 的稠密块与块 $x_{m,n}^{A,u}$ 的欧氏距离, N_p 表示稠密块的最近邻集合中排在前面的稠密块的个数,实验发现^[7],稠密块与最近邻块的距离随着排名的增长而增大,而达到一定排名后,距离基本处于同一水平,考虑到最近邻块的有效性,本文实验中 $N_p = 0.3V$ 。

为了方便后续实验,将对得分进行量化处理。考虑到行人具有一定的结构性,为了防止行人不同部位的稠密块由于较为相似混合在一起,本文首先将行人图像重叠地分为 N_y 层,每层与相邻两层各重叠一行,如图3所示。然后,根据式(2)得到的稠密块得分把各层的稠密块平均量化为



N_l 个等级。一个稠密块所处的 l 等级代表着此稠密块的辨别能力和泛化能力, l 等级越大, 辨别能力越强, 泛化能力越弱; 相反 l 等级越小, 辨别能力越弱, 泛化能力越强。

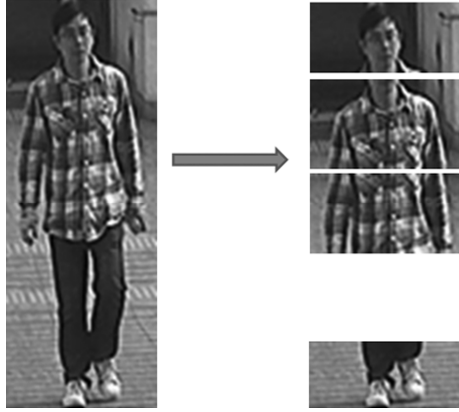


图3 行人图像垂直方向重叠分层

具体计算式如下:

$$S_y^A = \begin{cases} \{x_{m,n}^{A,u} | m = (y-1) \times \left\lfloor \frac{M}{N_y} \right\rfloor + 1, \dots, \\ y \times \left\lfloor \frac{M}{N_y} \right\rfloor + 1, n = 1, \dots, N, u = 1, \dots, U\}, y = 1 \\ \{x_{m,n}^{A,u} | m = (y-1) \times \left\lfloor \frac{M}{N_y} \right\rfloor, \dots, \\ y \times \left\lfloor \frac{M}{N_y} \right\rfloor + 1, n = 1, \dots, N, u = 1, \dots, U\}, 1 < y < N_y \\ \{x_{m,n}^{A,u} | m = (y-1) \times \left\lfloor \frac{M}{N_y} \right\rfloor, \dots, \\ y \times \left\lfloor \frac{M}{N_y} \right\rfloor, n = 1, \dots, N, u = 1, \dots, U\}, y = N_y \end{cases} \quad (4)$$

$$S_{y,l}^A = \{x | x \in S_y^A, s^{pAUC}(x) \in \left[s_{y,\min}^A + (l-1) \frac{\Delta S_y^A}{N_L}, s_{y,\min}^A + l \frac{\Delta S_y^A}{N_L} \right] \} \quad (5)$$

其中, $s_{y,\min}^A = \min\{s^{pAUC}(x) | x \in S_y^A\}$, $s_{y,\max}^A = \max\{s^{pAUC}(x) | x \in S_y^A\}$, $\Delta S_y^A = s_{y,\max}^A - s_{y,\min}^A$, $l = 1, \dots, N_L$ 。实验中 $N_y = 10$, $N_l = 10$ 。为了获取更好的辨别能力, 选取处于中间及之上等级的稠密块^[7], 即 $l = [6, 10]$ 。

3.1.2 有效块的聚类

一个性能良好的中层过滤器是根据一组描述特有视觉信息的稠密块训练获得的。虽然根据对每一层有效块进行量化分级, 但是在每一层的每一级内包含不同视觉信息的有效块依然混合在一起。因此将对每一层内每一级的有效块进行聚类, 尽量使含有相同视觉信息的稠密块聚在一起。这里采用聚类粒度从粗到细来构建层次树^[19], 对层次树进行修剪以找到相关有效块集合。那么, 对于一个分级内的有效块进行聚类的具体方法是, 由式(5)得到的一个分级的有效块集合 $S_{y,l}^A$ 作为根节点, 建立一个子节点为 O_l 、最大深度为 D_l 的层次聚类树^[7]。为了学习到一个能描述特定的视觉信息, 并产生紧凑响应的过滤器, 将只保留叶子节点, 并设置阈值范围对叶子节点进行裁剪, 当叶子节点中的有效块数目处于 $T_{\max}$ 与 $T_{\min}$ 之间, 将作为聚类结果, 记作 $\{Node_k\}_{k=1}^{N_{node}}$, N_{node} 表示聚类的节点数目。在本文实验中 $O_l = 4$, $D_l = 10$ 。

3.1.3 训练 SVM 中层过滤器

获取到所需的聚类结果后, 目标是训练出能从行人库中辨别出同一行人图像, 并对因摄像头变换而引起的姿势、视角和亮度变化有顽健性的过滤器。因此, 训练准确描述一个聚类结果 $Node_k$ 所表示的视觉信息的过滤器, 以此聚类结果中的所有稠密块作为正样本 X_k^+ , 从与此聚类结果同一等级的其他聚类结果中随机抽样的稠密块作为负样本 X_k^- 。为了确保不同视角的同一行人的图像对过滤器的响应程度保持一致, 仅仅依靠上述正负样本还不能学习到具有较强顽健性和较强判别性的过滤器, 因此将手工标注此聚类结果中所有稠密块对应同一行人不同图像的匹配块作为辅助正样本 X_k^{aux+} , 这样可以使得同一行人的不同图像对过滤器的响应程度保持一致, 另外

将此聚类结果中所有稠密块所匹配到相似行人（利用参考文献[4]的算法）的稠密块作为辅助负样本 X_k^{aux-} ，可以在一定程度上避免相似行人对此过滤器拥有较高的响应度。那么，利用构建好的样本 $\{X_k^+, X_k^-, X_k^{aux+}, X_k^{aux-}\}$ 训练一个线性 SVM 过滤器 $\{w_k, b_k\}$ ，由于 $Node_k$ 属于第 y_k 层，所以相应的 SVM 过滤器将被约束在这一层，因此，聚类结果 $Node_k$ 对应的 SVM 过滤器可以表示为 $\{w_k, b_k, y_k\}_{k=1}^{N_{node}}$ 。行人对中层过滤器的响应度是通过 max-pooling 技术计算该过滤器所属层对应的所有稠密块获得的。A 摄像头下第 u 幅图像（即行人 u ）对每一个过滤器的响应度记为 $f^{A,u} = \{f_k^{A,u}\}_{k=1}^{N_{node}}$ ，并对其进行 L2 范数归一化和稀疏化处理后所得到的 $\hat{f}^{A,u}$ 作为行人 u 的中层过滤器响应度，即行人 u 的中层特征。

行人 u 与行人 v 的中层特征匹配得分可以定义为：

$$s_f = \sum_{k=1}^{N_{node}} s_{f_k} \quad (6)$$

$$s_{f_k} = \exp\left(\frac{-(\hat{f}_k^{A,u} - \hat{f}_k^{B,v})^2}{\sigma_f^2}\right) \quad (7)$$

其中， s_{f_k} 表示行人 u, v 对过滤器 k 的响应度匹配得分， $\hat{f}_k^{A,u}$ 为行人 u 在第 k 个中层过滤器上的响应度， $\hat{f}_k^{B,v}$ 为行人 v 在第 k 个中层过滤器上的响应度， σ_f 表示高斯函数的频宽。

3.2 垂直全局表观约束

虽然中层过滤器考虑到了一定的空间约束信息，即在过滤器上加入了所属层的空间限制，但这种约束是对过滤器进行了局部空间约束，当不同行人图像的部分区域比较相似时，会由于部分过滤器响应度的匹配得分较高而产生错误匹配。考虑到行人在摄像头中一般处于站立姿态，同一行人的不同图像在垂直方向上的表观信息更加相似。因此，将在原中层特征的基础上引入垂直全局表观约束，利用同一行人的不同图像在垂直方向上的表观信息更加相似这一特性，对中层特征

匹配得分的有效性施以约束，在一定程度上解决相似行人误匹配的问题。

一个行人的中层过滤器响应度是利用 max-pooling 技术进行计算的。如图 4 所示，以计算行人对第一个过滤器的响应度为例。图 4 中行人 u 为被稠密分块的简易图，稠密块为白色的表示背景，其他颜色表示前景。被黑色框围上的前 3 行稠密块为第一层，那么计算第一层内所有的稠密块与第一个中层过滤器（此过滤器 $y_k=1$ ）的响应度，以稠密块中的最大响应度作为行人的中层过滤器响应度，对应的稠密块为最大响应度稠密块，即图 4 中行人 u 第一行中颜色最深的稠密块。由图 2 所示，不同行人的表观中部分区域较为相似，导致对同一个描述这一视觉信息的中层过滤器都拥有较高的响应度，但是不同行人在垂直方向上的表观信息存在差异，可以利用这种差异进行约束。为了抵抗部分遮挡、姿势变化、光线、衣服皱纹等因素对表观信息的影响，对图 4 中最大响应度稠密块在自身前景图像的每一行都搜索一个与其最相似的稠密块，即图 4 中每行画斜线的稠密块，这些画斜线的稠密块所组成的序列即图 2 中的垂直表观序列。对应的，此最大响应度稠密块与这些画斜线的稠密块的距离所组成的序列，称为此最大响应度稠密块的表观信息序列。

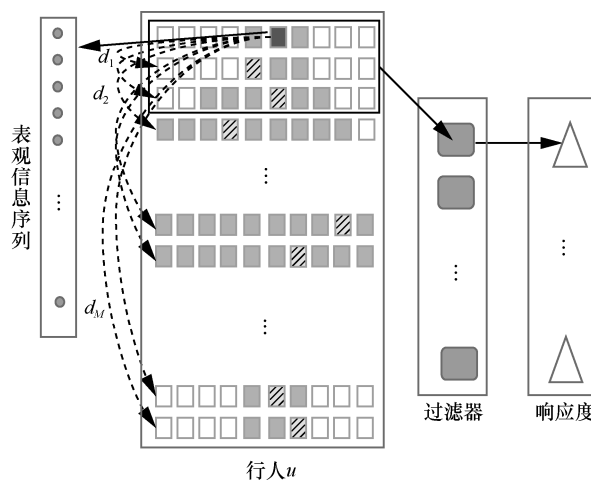


图4 中层过滤器响应度及表观信息序列计算流程

设稠密块 $x_{m,n}^{A,u}$ 是行人 u 对一个中层过滤器的



最大响应度稠密块, 那么稠密块 $x_{m,n}^{A,u}$ 在自身图像的垂直表观信息序列 $d_{m,n}^{A,u}$ 定义如下:

$$d_{m,n}^{A,u} = \{d_{m,n,1}^{A,u}, d_{m,n,2}^{A,u}, \dots, d_{m,n,k}^{A,u}, \dots, d_{m,n,M}^{A,u}\},$$

$$1 \leq m \leq M, 1 \leq n \leq N, 1 \leq k \leq M \quad (8)$$

其中, $d_{m,n,k}^{A,u} = \min\{d(x_{m,n}^{A,u}, x_{k,1}^{A,u}), \dots, d(x_{m,n}^{A,u}, x_{k,j}^{A,u}), \dots, d(x_{m,n}^{A,u}, x_{k,N}^{A,u})\}$ 表示稠密块 $x_{m,n}^{A,u}$ 在行人自身前景图像中的第 k 行搜索到的一个最相似稠密块所求得的距离, $d(x_{m,n}^{A,u}, x_{k,j}^{A,u})$ 表示稠密块 $x_{m,n}^{A,u}$ 与第 k 行的第 j 个稠密块之间的欧氏距离。

同理, 另一行人 v 对上述同一中层过滤器的最大响应度稠密块为 $x_{i,j}^{B,v}$, 那么其垂直表观信息序列 $d_{i,j}^{B,v}$ 定义如下:

$$d_{i,j}^{B,v} = \{d_{i,j,1}^{B,v}, d_{i,j,2}^{B,v}, \dots, d_{i,j,k}^{B,v}, \dots, d_{i,j,M}^{B,v}\},$$

$$1 \leq i \leq M, 1 \leq j \leq N, 1 \leq k \leq M \quad (9)$$

当这两个最大响应度稠密块来源于同一个行人的不同图像时, 那么在各自图像中的垂直表观信息更相似, 即垂直表观信息序列 $d_{m,n}^{A,u}$ 、 $d_{i,j}^{B,v}$ 之间的欧氏距离更小。为了将这种垂直表观信息的相似性作为对应中层过滤器响应度匹配有效性的约束, 这里将表观信息差异转化为权值问题来量化处理, 计算方式如下:

$$w(x_{m,n}^{A,u}, x_{i,j}^{B,v}) = \sum_{k=1}^M (|d_{m,n,k}^{A,u} - d_{i,j,k}^{B,v}|) \quad (10)$$

其中, $w(x_{m,n}^{A,u}, x_{i,j}^{B,v})$ 表示两个行人对同一个中层过滤器响应度匹配度的权值。当两行人对应的最大响应度稠密块在各自图像中的垂直表观信息越相似时, $w(x_{m,n}^{A,u}, x_{i,j}^{B,v})$ 权值越小, 即表示两个行人可能是来自同一个行人的不同图像。

由于中层过滤器响应度匹配得分越高, 两行人越可能是同一行人, 为了与之一致, 将权值 $w(x_{m,n}^{A,u}, x_{i,j}^{B,v})$ 做如下转化:

$$w'(x_{m,n}^{A,u}, x_{i,j}^{B,v}) = \exp\left(-\frac{w(x_{m,n}^{A,u}, x_{i,j}^{B,v})^2}{2\delta_{pw}^2}\right) \quad (11)$$

经过转化后, 两个最大响应度稠密块在各自图像中的垂直表观信息越相似, 对应中层过滤器响应度匹配得分给予越大的权值 w' , 与两行人对此中层

过滤器响应度匹配计算出的得分越大保持一致。

引入垂直全局表观约束的中层特征匹配得分计算流程如图 5 所示, 行人图像 u 、 v 是图像库中

2017023-7

原始图像, 为了避免背景对识别的影响, 采用了 Yang 等人^[20]的姿势评估算法, 较快地从图像中提取粗略的前景(人体部分)。然后, 利用训练好的过滤器计算行人的中层过滤器响应度, 并计算对应最大响应度稠密块的表观信息距离序列。最后, 将表观信息距离序列的相似度作为响应度匹配的约束, 从而得到中层特征匹配得分。具体计算式为:

$$s_f' = \sum_{k=1}^{N_{node}} s_{f_k}' \quad (12)$$

$$s_{f_k}' = s_{f_k} \cdot w'(x_k^{A,u}, x_k^{B,v}) \quad (13)$$

其中, s_{f_k}' 表示引入垂直全局表观约束后行人图像 u 、 v 对过滤器 k 的响应度匹配得分, $x_k^{A,u}$ 表示行人 u 在第 k 个中层过滤器对应的最大响应度稠密块, $x_k^{B,v}$ 表示行人 v 在第 k 个中层过滤器对应的最大响应度稠密块。

3.3 融合匹配得分

底层稠密块匹配方法^[4]取得了良好的效果, 而且能够弥补中层特征所丢失的一些信息, 融合底层特征与中层特征能起到互补的作用, 因此在最终匹配过程中融合底层稠密块匹配的得分, 从而取得更好的识别效果。两幅图像的最终匹配得分由底层稠密块匹配得分和引入垂直表观全局约束的中层特征得分融合共同决定, 得分越高表示两幅图像越有可能是同一行人。

底层稠密块匹配得分的具体计算式如下:

$$s_p = \sum_{i=1}^{MN} s_{p_i} \quad (14)$$

$$s_{p_i} = \exp\left(\frac{-d(x_{p_i}^{A,u} - x_{p_i}^{B,v})}{\sigma_p^2}\right) \cdot w_{p_i} \quad (15)$$

其中, s_{p_i} 表示行人 u 、 v 对应的匹配块之间得分, $x_{p_i}^{A,u}$ 与 $x_{p_i}^{B,v}$ 表示行人 u 、 v 对应的匹配稠密

块, w_{p_i} 是对应的突出性权重。

那么, 最终得分式描述如下:

$$\text{Score}(I_u^A, I_v^B) = w^T \Phi(x^{A,u}, x^{B,v}, \hat{f}_*^{A,u}, \hat{f}_*^{B,v}) \quad (16)$$

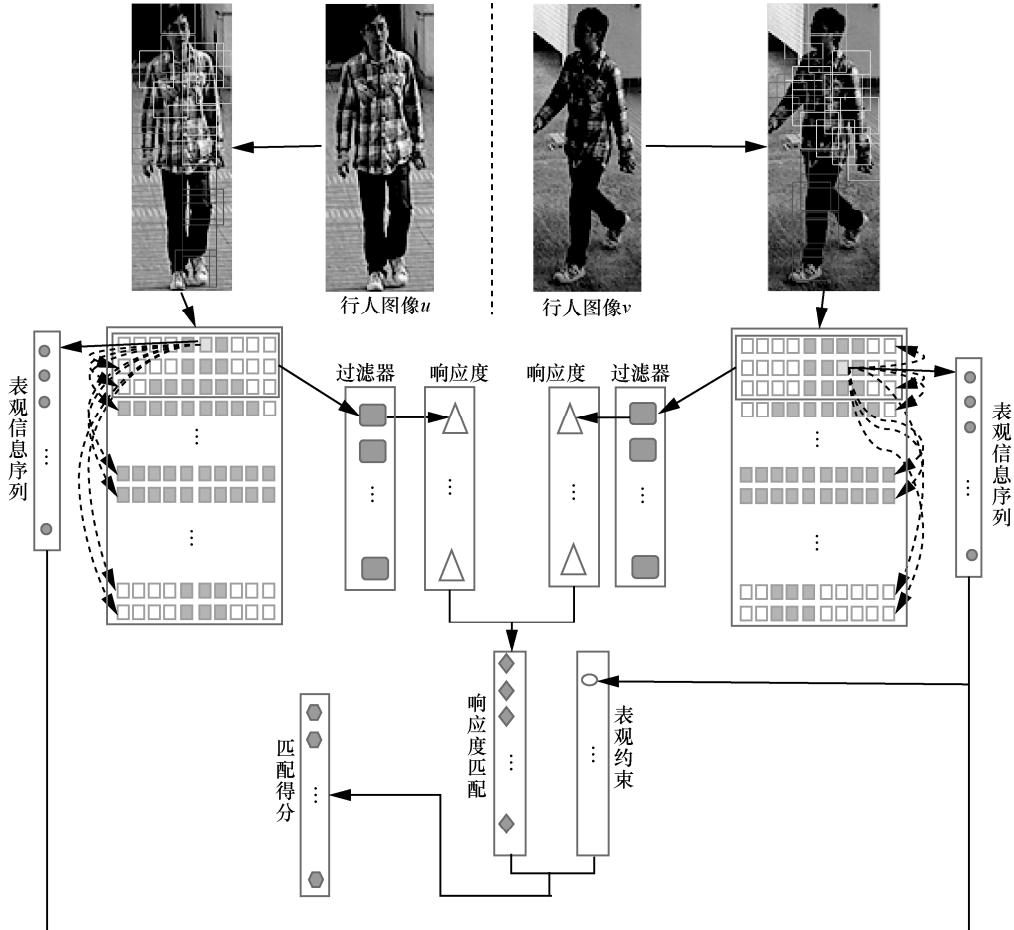


图5 中层特征匹配计算流程

$$\begin{aligned} & \Phi(x^{A,u}, x^{B,v}, \hat{f}_*^{A,u}, \hat{f}_*^{B,v}) \\ &= [s_p(x^{A,u}, x^{B,v}), s_f(\hat{f}_*^{A,u}, \hat{f}_*^{B,v})] \end{aligned} \quad (17)$$

$$s_p(x^{A,u}, x^{B,v}) = [s_{p_1}, \dots, s_{p_{MN}}] \quad (18)$$

$$s_f(\hat{f}_*^{A,u}, \hat{f}_*^{B,v}) = [s_{f_1}, \dots, s_{f_{N_{\text{node}}}}] \quad (19)$$

其中, I_u^A 、 I_v^B 分别表示待识别图像与目标库中的图像, w^T 表示融合各部分得分的权重, 本文通过 rankSVM^[5]得到。

4 实验结果及分析

本文算法将在两个公用数据库上进行实验

验证, 分别为 VIPeR 和 CUHK01 数据库^[2]。VIPeR 数据库是现在用来评估行人再识别最为广泛的数据库, 而 CUHK01 数据库相比 VIPeR

数据库拥有更多的图像。两个数据库所包含的图像都是低分辨率, 光照、姿态、背景都有变化, 都是极具挑战性的数据库。实验结果将采用标准的累积匹配特征 (cumulative matching characteristic, CMC) 曲线进行评估描述, CMC 曲线是指在候选行人库中检索目标行人, 前 r 个检索结果中命中正确匹配结果的比率。 r 如果比较小, 在实际应用中就可以通过人眼进行辅助识别出目标, 大大减少人力物力, 当 $r=1$ 时, 即该算法的真实识别能力。

采用基于稠密块的特征提取策略, 稠密块定



义为块大小为 $10 \text{ dpi} \times 10 \text{ dpi}$ ，以步长 4 个像素点向右或向下移动。在每个稠密块中的 LAB 通道中提取 32 维的颜色直方图特征和 128 维的 SIFT 特征，为了更好地获得轮廓与细节的判别信息，每个通道都进行了降采样（降采样因子分别为 0.5、0.75）。颜色直方图和 SIFT 特征通过 L2 范数进行归一化处理，最终结合为 672 维的特征作为一个稠密块的特征。

4.1 VIPeR 数据库的测试结果

VIPeR 数据库是从两个不重叠摄像头在户外环境捕获的包含 632 个行人的图像库，其中每个行人都有两张不同视角的图像，且大部分行人的两幅图像的视角变化超过了 90° 。所有的图像都归一化到 $128 \text{ dpi} \times 48 \text{ dpi}$ 。

实验中为与其他算法进行公平的比较，将随机选取 316 个行人图像（632 幅）作为训练集，余下的 316 个行人作为测试集。测试时，摄像头 A 中的图像先作为查询库，摄像头 B 中的图像作为候选行人库。每个查询图像与候选行人库中的每幅图像都要匹配，从而得到相应的匹配排名。然后，再交换查询库和候选行人库。为获得稳定的识别率，以上产生训练集和测试集的过程将重复 10 次，并取 10 次实验结果的平均值作为最终的实验结果。

为了充分地验证本算法的性能，将其与近几年比较有代表性的算法进行了比较，对比算法的实验结果采用相关文献提供的实验结果或文献中提供的公开代码进行的仿真结果。图 6 是根据各算法结果绘制的 CMC 曲线，表 1 给出了图 6 中各算法在 $r=1$ 、5、10、20 时的匹配率。

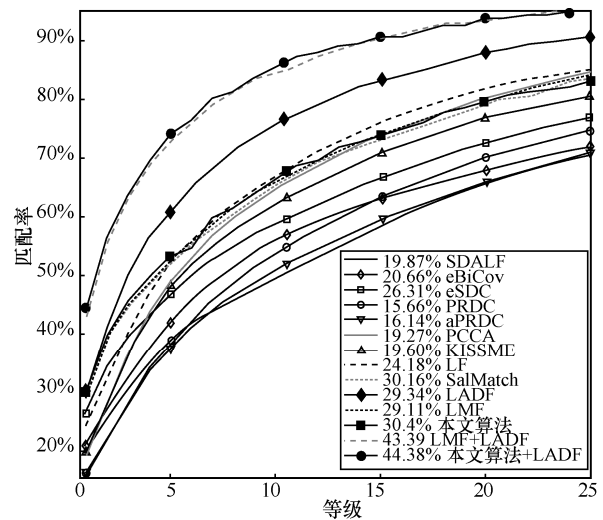


图 6 CMC 曲线比较 (VIPeR)

2017023-9

表 1 VIPeR 库各算法部分匹配率比较

方法	$r=1$	$r=5$	$r=10$	$r=20$
SDALF	19.87%	38.89%	49.37%	65.73%
eSDC	26.31%	46.60%	58.86%	72.76%
PRDC	15.66%	38.42%	53.86%	70.09%
PCCA	19.27%	48.89%	64.91%	80.28%
KISSME	19.60%	48%	62.20%	77%
LF	24.18%	52%	67.12%	82%
SalMatch	30.16%	52.31%	65.54%	79.14%
LADF	29.34%	61.04%	75.98%	88.10%
LMF	29.11%	52.34%	65.95%	79.87%
本文算法	30.04%	53.27%	66.21%	79.50%
LMF+LADF	43.29%	73.04%	84.87%	93.70%
本文算法+LADF	44.38%	74.14%	85.56%	93.97%

从图 6 的 CMC 曲线及表 1 中可以看出，本文的方法取得最好的识别结果。其中 SDALF^[8]、SalMatch^[5]、eSDC^[4]及 LMF^[7]方法为特征设计方法，本文方法与 SDALF 及 SalMatch^[5]相比都有一定的提升，与本文方法所融合的底层显著性稠密块匹配方法 eSDC^[4]相比提升了 4% 左右，与未加入表观约束的中层特征匹配方法 LMF^[7]相比提高了 1% 左右。值得注意的是，LMF^[7]方法通过与当时识别效果最好的方法 LADF^[18]结合后 rank-1 命中率达到 43.39%，因此将本文方法与 LADF^[18]进行结合，可以看到 rank-1 达到 44.38%，依然高于结合了 LADF^[18]的 LMF^[7]方法，说明垂

直全局表观约束有效地提升了再识别的命中率。对于其他距离度量学习方法 LF^[21]、KISSME^[22]、PRDC^[23]及 PCCA^[24]等方法相比本文方法也有较大的提升。

4.2 CUHK01 数据库的测试结果

CUHK01 数据库是从两个不重叠摄像头在校园环境捕获的包含 971 个行人的图像库，每个人在每个摄像头下都捕获两张图像，共 4 张图像。行人的前两张图像捕获行人的正面和背面，行人的后两张图像捕获不同的侧面视角。行人的这 4 张图像在视角、姿态和光照上均有明显的变化。所有图像都归一化到 160 dpi × 60 dpi。

为了实验的公平性，实验中随机选取 486 个行人图像（1 994 幅）作为训练集，剩余 485 个行人图像作为测试集，为获得稳定的识别率，测试时，摄像头 A 中的图像先作为查询库，摄像头 B 中的图像作为候选行人库。每个查询图像与候选行人库中的每幅图像都要匹配，从而得到相应的匹配排名。然后，再交换查询库和候选行人库。以上产生训练集和测试集的过程将重复 10 次，并取 10 次实验结果的平均值作为最终的实验结果。

同样，为了充分地验证本算法的性能，将与最近几年比较有代表性的算法进行比较，实现结果采用文献

给出的结

果，或根据文献中提供的公开代码进行的仿真结果。图 7 是根据各算法的实验结果绘制的 CMC 曲线，表 2 给出了图 7 中各算法在 $r=1$ 、5、10、20 时的命中率。

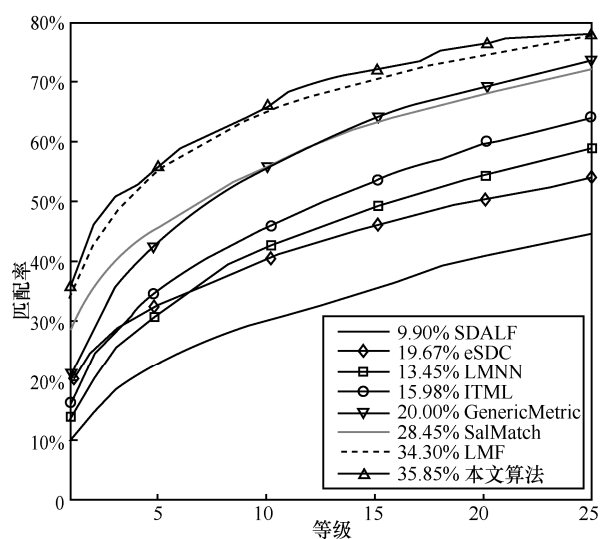


图 7 CMC 曲线比较(CUHK01)

表 2 CUHK01 库各算法部分匹配率比较

方法	$r=1$	$r=5$	$r=10$	$r=20$
SDALF	9.90%	22.57%	30.33%	41.03%
eSDC	19.67%	32.71%	40.29%	50.58%
LMNN	13.45%	31.33%	42.25%	54.11%
ITML	15.98%	35.22%	45.60%	59.81%
GenericMetric	20%	43.58%	56.04%	69.27%
SalMatch	28.45%	45.85%	55.67%	67.95%
LMF	34.30%	55.06%	64.95%	74.94%
本文算法	35.85%	55.80%	65.89%	76.39%

从图 7 的 CMC 曲线及表 2 可以看出，本文算法相比其他算法有较明显的提升。其中 SDALF^[8]、SalMatch^[5]、eSDC^[4] 及 LMF^[7]方法为特征设计方法，本文方法与 SDALF^[8]、SalMatch^[5] 及 eSDC^[4] 等方法相比提升较为明显，与未加入表观约束的中层特征匹配方法 LMF^[7]相比也提高了 1% 左右，rank-1 的命中率达到 35.85%。对于其他距离度量方法 LMNN^[25]、ITML^[25] 及 GenericMetric^[25] 相比都有较大的提升。

从实际应用价值角度来看，行人再识别中匹配结果的排名越靠前，应用价值越大，这可用匹配结果的等级排名来说明。图 8 给出了本文算法及 LMF 算法在 CUHK01 库上的可视化等级排名



对比结果,图8中左边一列为待识别行人,右边为识别匹配结果等级排名,越靠左排名越靠前,其中方框为正确匹配候选目标。从图8中可以看出,本文算法所识别的正确匹配候选目标更加靠前,在实际应用中具有更高的价值。

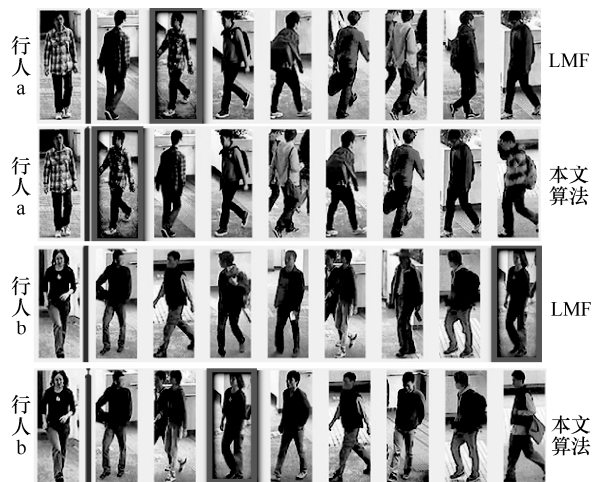


图8 CUHK01库上的可视化rank排名对比结果

另外,本文算法在CUHK01上的结果明显优于VIPeR,根据实验进行分析,一个原因是CUHK01所包含的样本数量更多,且分辨率稍高于VIPeR数据库,可以训练出更好的中层过滤器。而在实际应用中,样本复杂性更高且样本数量充足,因此本文算法的优势将会更加明显。

5 结束语

行人再识别是一个实际应用价值较高同时具有挑战性的问题。中层特征拥有底层特征所没有的语义信息和更好的辨别能力,结合中层特征和底层特征能起到很好的互补作用。而且现有的底层特征方法和中层特征方法都没有考虑人体在垂直方向上的全局表观约束。因此,提出了在中层特征的基础上引入了垂直全局表观约束,并融合底层稠密块特征匹配方法,一定程度上解决相似行人由于局部误匹配而影响识别效果的问题,提

高了行人再识别的命中率。VIPeR和CUHK01这两个公共数据集里的图像都存在光照变化、行人姿态变化、视角变化以及遮挡等情况,但该算法在这两个数据集上的识别效果都比已有算法更优,验证了本文算法的有效性,并且有一定的普适性。

参考文献:

- [1] VEZZANI R, BALTIERI D, CUCCHIARA R. People reidentification in surveillance and forensics: a survey[J]. ACM Computing Surveys, 2013, 46(2):1-37.
- [2] WANG X, ZHAO R. Person re-identification: system design and evaluation overview[M]. Berlin: Springer-Verlag, 2014: 351-370.
- [3] 黄凯奇, 陈晓棠, 康运锋, 等. 智能视频监控技术综述[J]. 计算机学报, 2015, 38(6): 1093-1118.
HUANG K Q, CHEN X T, KANG Y F, et al. Intelligent visual surveillance: a review[J]. Chinese Journal of Computers, 2015, 38(6): 1093-1118.
- [4] ZHAO R, OUYANG W, WANG X. Unsupervised salience learning for person re-identification[C]//CVPR, June 25-27, 2013, Portland, Oregon, New Jersey: IEEE Press, 2013: 3586-3593.
- [5] ZHAO R, OUYANG W, WANG X. Person re-identification by salience matching[C]// IEEE International Conference on Computer Vision, December 1-8, 2013, Sydney, Australia. New Jersey: IEEE Press, 2013: 2528-2535.
- [6] 陈普强, 郭立君, 张荣, 等. 基于全局空间约束块匹配的目标人体识别[J]. 计算机研究与发展, 2015(3): 596-605.
CHEN P Q, GUO L J, ZHANG R, et al. Patch matching with global spatial constraints for person re-identification[J]. Journal of Computer Research and Development, 2015(3): 596-605.
- [7] FARENZENA M, BAZZANI L, PERINA A, et al. Person re-identification by symmetry-driven accumulation of local features[C]//CVPR, June 13-18, 2013, San Francisco, California, USA. New Jersey: IEEE Press, 2010: 2360-2367.
- [8] CHENG D S, CRISTANI M, STOPPA M, et al. Custom pictorial structures for re-identification[C]//BMVC, August 22-September 2, 2011, Dundee, USA. New Jersey: IEEE Press, 2011, 1(2): 6.
- [9] PROSSER B, ZHENG W S, GONG S, et al. Person re-identification by support vector ranking[C]//BMVC, April 4-9, 2010, New York, USA. New Jersey: IEEE Press, 2010: 6.
- [10] SINGH S, GUPTA A, EFROS A A. Unsupervised discovery of mid-level discriminative patches[M]. Berlin: Springer Heidelberg, 2012: 73-86.
- [11] JAIN A, GUPTA A, RODRIGUEZ M, et al. Representing videos using mid-level discriminative patches[C]//CVPR, June 25-27, 2013, Portland, Oregon, New Jersey: IEEE Press, 2013: 2571-2578.
- [12] LAYNE R, HOSPEDALES T M, GONG S, et al. Person

- re-identification by attributes[C]//BMVC, April 11–17, 2012, Surrey, USA. New Jersey: IEEE Press, 2012: 8.
- [13] LIU X, SONG M, ZHAO Q, et al. Attribute-restricted latent topic model for person re-identification[J]. Pattern Recognition, 2012, 45(12): 4204-4213.
- [14] ZHAO R, OUYANG W, WANG X. Learning mid-level filters for person re-identification[C]//IEEE Conference on Computer Vision and Pattern Recognition, June 24–27, 2014, Columbus, Ohio, USA. New Jersey: IEEE Press, 2014: 144-151.
- [15] MA K, BEN-ARIE J. Vector array based multi-view face detection with compound exemplars[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), October 11–13, 2012, Providence, USA. New Jersey: IEEE Press, 2012: 3186-3193.
- [16] HU Y, LIAO S, LEI Z, et al. Exploring structural information and fusing multiple features for person re-identification[C]// CVPR, June 25–27, 2013, Oregon, Portland. New Jersey: IEEE Press, 2013: 794-799.
- [17] ZHENG W S, GONG S, XIANG T. Reidentification by relative distance comparison[J]. IEEE transactions on Pattern Analysis and Machine Intelligence, 2013, 35(3): 653-668.
- [18] LI Z, CHANG S, LIANG F, et al. Learning locally-adaptive decision functions for person verification[C]// Conference on Computer Vision and Pattern Recognition, June 25-27, 2013, Portland, Oregon. New Jersey: IEEE Press, 2013: 3610-3617.
- [19] ZHANG W, WANG X, ZHAO D, et al. Graph degree linkage: agglomerative clustering on a directed graph[C]//European Conference on Computer Vision, April 11–14, 2012, Berlin, German. Berlin: Springer, 2012: 428-441.
- [20] YANG Y, RAMANNAN D. Articulated pose estimation with flexible mixtures-of-parts[C]// IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2011, Colorado Springs, USA. New Jersey: IEEE Press, 2011: 1385-1392.
- [21] PEDAGADI S, ORWELL J, VELASTIN S, et al. Local fisher discriminant analysis for pedestrian re-identification[C]// CVPR, June 25–27, 2013, Oregon, Portland. New Jersey: IEEE Press, 2013: 3318-3325.
- [22] KÖSTINGER M, HIRZER M, WOHLHART P, et al. Large scale metric learning from equivalence constraints[C]// Conference on Computer Vision and Pattern Recognition, Providence, June 16–21, 2012, Providence, RI, USA. New Jersey: IEEE Press, 2012: 2288-2295.
- [23] ZHENG W S, GONG S, XIANG T. Person re-identification by probabilistic relative distance comparison[C]// Conference on Computer Vision and Pattern recognition, June 20–25, 2011, Colorado Springs, USA. New Jersey: IEEE Press, 2011: 649-656.
- [24] MIGNON A, JURIE F. PCCA: a new approach for distance learning from sparse pairwise constraints[C]// Conference on Computer Vision and Pattern Recognition, June 16–21, 2012, Providence, RI, USA. New Jersey: IEEE Press, 2012: 2666-2672.
- [25] LI W, ZHAO R, WANG X. Human reidentification with transferred metric learning[C]//Asian Conference on Computer Vision, March 1–5, 2012, Berlin, German. Berlin: Springer, 2012: 31-44.

[作者简介]



王亚东 (1990–), 男, 宁波大学信息科学与工程学院硕士生, 主要研究方向为计算机视觉与模式识别。



张荣 (1974–), 女, 博士, 宁波大学副教授, 主要研究方向为数字取证与信息安全。



蒋桢慧 (1993–), 女, 宁波大学信息科学与工程学院硕士生, 主要研究方向为计算机视觉与模式识别。



郭立君 (1970–), 男, 博士, 宁波大学教授, 主要研究方向为计算机视觉与模式识别、移动互联网及其应用。