

基于外观统计特征融合的人体目标再识别

曾明勇^{*①} 吴泽民^{①②} 田畅^① 付毅^① 揭斐然^②

^①(解放军理工大学通信工程学院 南京 210007)

^②(中国航空工业集团光电控制技术重点实验室 洛阳 417009)

摘要: 人体目标再识别是视频监控等应用的关键问题之一。该文从外观统计特征融合的角度,利用人体的颜色和结构信息,基于空间直方图和区域协方差两种优秀的统计描述方法,研究了再识别问题的特征构建和测度选择等内容。构建特征时从图像多个层次的统计区域中提取了多类互补性较好的统计向量,设计测度时使用了简单的 l_1 距离进行加权组合。两类统计方式融合而成的再识别方法不需要进行预处理和监督性训练过程。该文进行了广泛的实验比较和分析,验证了该文方法优异的识别性能和较强的实用性能。

关键词: 人体目标再识别; 特征融合; 空间直方图; 区域协方差

中图分类号: TP391.41

文献标识码: A

文章编号: 1009-5896(2014)08-1844-08

DOI: 10.3724/SP.J.1146.2013.01389

Fusing Appearance Statistical Features for Person Re-identification

Zeng Ming-yong^① Wu Ze-min^{①②} Tian Chang^① Fu Yi^① Jie Fei-ran^②

^①(College of Communications Engineering, PLA University of Science and Technology, Nanjing 210007, China)

^②(Science and Technology on Electro-Optic Control Laboratory, Aviation Industry Corporation of China, Luoyang 417009, China)

Abstract: Person re-identification is among the key issues in video surveillance. From the viewpoint of fusing appearance statistical features, human color and structure information are exploited; two statistical descriptors named spatiogram and region covariance are both explored on feature designing and metric choosing. Several complimentary feature vectors are extracted from a proper number of hierarchical image layers and regions. The simplest l_1 norm distance is chosen to form the proposed weighted combining distance. The fused method with such two descriptors requires neither preprocessing nor supervised training. Extensive experiments by comparisons and analysis show that the proposed method not only achieves the state-of-the-art re-identification performance, but also enjoys a great applicability.

Key words: Person re-identification; Feature fusing; Spatiogram; Region covariance

1 引言

近年来,人体目标再识别技术得到了广泛的研究和应用,在视频网络监控、多目标跟踪等方面发挥了重要作用^[1]。2013年Google公司推出的可穿戴计算设备Google Glass^[2]已具备人体再识别功能,它通过眼镜的高清摄像头,可从超市等场所迅速识别出想要寻找的目标。人体目标再识别是指利用已经存在的人体目标图库,在场景或时间变化时,通过搜索图库,再次识别和确认当前所查询人体图片的目标身份^[3]。由于存在光照、视角、姿态、背景和遮

挡等极具挑战性的影响因素,人体再识别并不是一个简单的问题。

目前,人体再识别技术主要从3个方面来提高再识别率。第1类只采用基于图像外观特征的方法;第2类结合图像序列的空时关系进行推理;第3类结合生物特征,如人脸识别或步态识别。基于外观的人体再识别对先验知识要求最少,是本文的研究重点。外观方法目前主要存在两种不同的解决思路:(1)特征类方法,从外观特征出发,选取或设计类内具有一定不变性且类间具有较强区分性的描述特征;(2)测度类方法,从特征相似度或距离出发,在人为选择特征的基础上,有监督地学习和训练距离测度或分类器,使同类样本距离小而非同类样本距离大。此外,根据每个人在目标图库的图片数量为单帧或多帧,人体再识别还可分为单帧模式和多帧

2013-09-11 收到, 2013-12-09 改回

光电控制技术重点实验室和航空科学基金(20125186005)联合资助课题

*通信作者: 曾明勇 zengmingyong1987@gmail.com

模式，多帧模式通过融合同一个人的多幅图像能够获得更好的性能。

在特征方面，人体再识别主要使用的特征有颜色、结构、形状、梯度、纹理、关键点以及区域描述子等。特征提取前可能进行前景分割、人体部件检测等预处理过程，提取到的不同特征通常可以进行级联或组合，形成区分性更强的特征。SDALF 方法^[3]提取前景后将人体目标分割为头部、躯干和腿部，并使用加权 HSV 直方图、最大稳定颜色区域和重复纹理块 3 种特征进行组合识别。CPS 方法^[4]利用人体结构检测出人体的详细部位后再进行识别。AHPE 方法^[5]将人体分为上下身并计算多帧模式下的产生式缩影特征。eLDFV 方法^[6]使用 Fisher 向量编码图像的亮度和梯度信息，并和 SDALF 方法进行组合。文献[7]融合了方向梯度直方图和空间颜色直方图等多种特征。SCEFA 方法^[8]采用分层加权的 HSV 直方图和双向匹配的颜色区域特征，并用 Gabor 纹理模式描述检测到的关键点。文献[9,10]采用区域协方差的方式融合多种特征，SCR 方法^[9]在人体各部位上计算协方差，MRCG 方法^[10]在多帧模式下计算均值协方差。此外，部分文献通过寻找人体特性或显著性进行再识别，eSDC 方法^[11]通过非监督的学习方式寻找个体显著特征，并和 SDALF 方法进行组合增强。

在距离测度方面，测度学习的本质是在计算特征向量距离时估计各分量距离的尺度或权重，选取或突出那些区分性强的特征分量。ELF 方法^[12]通过 AdaBoost 分类器选取颜色和纹理特征。COSMATI 方法^[13]在协方差测度空间利用相关性选择显著区域和外观特征。ERSVM 方法^[14]采用基于相对距离排序的 SVM 分类器集成多个特征向量权重。文献[15]采用多核 SVM 学习的方法融合颜色和视觉单词直方图。其它方法主要学习基于对称半正定矩阵的马氏测度，如采用大边近邻分类器的 LMNN-R 方法^[16]和采用相对距离比较的 PRDC 方法^[17]。LF 方法^[18]则先对高维特征进行 PCA 降维，再基于局部 Fisher 判别分析进行降维映射，获得了非常高的识别性能。

与特征类方法相比，测度学习方法对特征的选择要求较低，通常能取得更好的再识别效果。但其训练的时间和空间复杂度较高，在训练样本不足时存在过度拟合等现象，且场景和图库变化时需要重新训练。特征类方法则需要人为地寻找和设计较好的特征，方法的性能依赖于特征的优劣程度。本文在特征类方法的思路下设计和构造优秀的特征及描述，同时也讨论特征的非学习距离测度，突出方法的实用性。本文主要基于颜色和结构信息，融合两

种外观统计特征描述方法进行人体再识别。此外，本文从应用的角度设计了更合理的再识别性能评估指标。最后本文在两个公共图库上进行了广泛的实验，验证本文方法的优异性能。

2 融合外观特征

本文主要基于颜色和结构信息提出了两种外观统计特征提取方法。第 1 种是基于多种颜色通道的空间直方图特征 (Multi-Channel based Spatio-Histogram, MCSH)，第 2 种是融合多种颜色和梯度信息的金字塔型协方差特征 (Multi-Statistics on Pyramid of Covariance, MSPC)。两种方法均不需要进行前景分割和部件检测，均在图像多个层次或区域上计算统计特征并级联成特征向量，距离测度采用简单的非监督 l_1 测度。本文融合这两种不同的统计方式进行人体再识别，融合方法记为 HSCD (Hybrid Spatiogram and Covariance Descriptor)。

通常，人体再识别假设人在短时间内着装相同，因此颜色是人体识别最常用且最有效的特征。另外人体作为一种特殊的目标，人体结构对于人体识别也非常重要，将人体图片进行合理的多层次多区域划分能更好地反映出人体的结构信息。由于空间直方图和区域协方差都能有效地统计和融合区域内的多种信息，本文基于上述出发点探讨和设计相应的统计特征及距离测度，并将这两种基于类似特征但统计方式截然不同的方法进行融合，下面分别详细介绍。

2.1 多通道空间直方图(MCSH)

颜色直方图能够表现目标区域的整体特征，对视角和形状变化具有一定的不变性，但由于它完全忽略了颜色的空间结构，降低了模型的分辨能力。空间直方图可将颜色直方图和高阶的空间矩进行结合，因此具有更强的描述能力^[19]。

2.1.1 空间直方图 图像区域 I 的二阶空间直方图可表示为

$$S_I^{(2)}(b) = \langle n_b, \mu_b, \Sigma_b \rangle, \quad b = 1, 2, \dots, B \quad (1)$$

其中， B 为量化级数， n_b 为颜色直方图， μ_b 和 Σ_b 分别为均值矢量和协方差矩阵，分别定义为

$$\left. \begin{aligned} n_b &= \sum_{k=1}^N \delta_{kb} \\ \mu_b &= \frac{1}{n_b} \sum_{k=1}^N c_k \delta_{kb} \\ \Sigma_b &= \frac{1}{n_b} \sum_{k=1}^N (c_k - \mu_b)(c_k - \mu_b)^T \delta_{kb} \end{aligned} \right\} \quad (2)$$

其中， N 为图像区域的像素总数； $c_k = (x, y)^T$ 为像素 k 的空间坐标；如果像素 k 在量化级数 b 内，则

$\delta_{bb} = 1$, 否则为 0。直方图的距离测度通常采用巴氏距离或直方图交集方法, 空间直方图的测度需要将空间相似度融入到直方图测度计算中。对于两个空间直方图(\mathbf{S}, \mathbf{S}'), 其相似度为^[19]

$$\left. \begin{aligned} \rho(\mathbf{S}, \mathbf{S}') &= \sum_{b=1}^B \varphi_b \sqrt{n_b n'_b} \\ \varphi_b &= \eta \exp \left\{ -\frac{1}{2} (\boldsymbol{\mu}_b - \boldsymbol{\mu}'_b)^T \widehat{\boldsymbol{\Sigma}}_b^{-1} (\boldsymbol{\mu}_b - \boldsymbol{\mu}'_b) \right\} \end{aligned} \right\} \quad (3)$$

其中, 直方图采用巴氏测度, φ_b 为空间相似性, η 为高斯归一化常数, $\widehat{\boldsymbol{\Sigma}}_b^{-1} = \boldsymbol{\Sigma}_b^{-1} + (\boldsymbol{\Sigma}'_b)^{-1}$ 。

2.1.2 多通道多区域级联的空间直方图 传统的空间直方图通常在图片全局区域采用单类颜色空间进行计算, 本文针对人体再识别的特定应用对其做出适当改造和增强。在区域方面采用多层次多尺度的统计区域, 在像素特征方面引入描述各个区域的多通道颜色通道, 力求更准确地反映人体的颜色和结构信息。

通常人体检测技术获取的人体图像并不完全准确, 即使对于准确的检测结果, 由于检测时使用矩形框来截取人体目标, 仍然会存在一定的背景干扰。另外运动人体的手脚等边缘图像信息也不一定能为识别带来好处, 因此本文只采用图片的中间部分作为第 1 层的人体区域。其它层次的区域采用重叠或非重叠的方式进行密集型划分, 并采用不同的空间直方图量化级数。以 VIPeR 图库^[1]的图片为例, 详细的分层区域划分如图 1 所示, 图中从左至右依次为原图、网格划分图以及自顶向下的 4 层区域划分。其中第 3 层采用重叠型密集划分, 4 层划分中每个区域进行直方图统计的量化级数分别设为 32, 32, 16 和 8。

在像素统计特征方面, 本文除了使用常用的 HSV 颜色空间, 还使用包含色差通道的 YCbCr 颜色空间以及可消除光照影响的归一化 RGB 颜色空间。其中, HSV 和 YCbCr 空间直接由第 1 层区域的原始 RGB 空间转换得到。为了进一步消除光照影响, 提取归一化 RGB 空间前先对原始 RGB 空间的 3 个通道进行直方图均衡, 再按照式(4)的定义进行计算。



图1 MCSH方法分层区域划分

$$\left. \begin{aligned} nR &= R/(R+G+B) \\ nG &= G/(R+G+B) \\ nB &= B/(R+G+B) \end{aligned} \right\} \quad (4)$$

颜色通道中 Y 和 V 均表示亮度, 本文只选用 Y 通道, 在计算 Y 通道时采用直方图均衡后的 RGB 空间进行亮度转换。这样 3 种颜色空间共选出 8 个颜色通道 $\{Y, C_b, C_r, H, S, nR, nG, nB\}$ 。每个通道都在图 1 所示的各层区域上提取空间直方图并进行级联, 8 个通道再依次级联形成总量化级数为 2432 的空间直方图 \mathbf{S} 。

2.1.3 MCSH 特征优化和测度选取 获取空间直方图 \mathbf{S} 后, 测度的选取十分重要。采用式(3)测度的最大问题是计算复杂, 存在大量矩阵求逆运算, 且部分矩阵可能不可逆。从实验效果来看, 式(3)空间相似度的引入也并没能显著地提升直方图的性能(参见本文 4.1.1 节)。

在实际场景中, 尽管摄像机拍摄人体图片时可能存在较大的视角差异, 但摄像机一般都位于与人体高度相当或略高的水平面上。因此, 利用空间信息时可以去掉图像的水平空间信息, 只保留图像的垂直空间信息。故本文将式(1)改为

$$\left. \begin{aligned} \mathbf{S}(b) &= \langle n_b, \mu_{by}, \sigma_{by} \rangle, b = 1, 2, \dots, B \\ \mathbf{S} &= \{ \mathbf{h} = (n_b); \boldsymbol{\mu}_y = (\mu_{by}); \boldsymbol{\sigma}_y = (\sigma_{by}), b = 1, 2, \dots, B \} \end{aligned} \right\} \quad (5)$$

其中, μ_{by} 和 σ_{by} 分别代表图像区域第 b 级直方图的像素点在 y 方向上的位置均值和标准差, $\mu_{by} = \boldsymbol{\mu}_b(2)$, $\sigma_{by} = \boldsymbol{\Sigma}_b(2,2)$, 这两种垂直空间信息对于相同个人的图片具有较好的不变性。

通过去除性能较差的水平空间信息, 特征的提取速度和存储空间都得到了优化。对特征优化后的空间直方图 \mathbf{S} , 本文将测度设计为所含 3 类特征向量 l_i 距离的加权组合

$$d(\mathbf{S}, \mathbf{S}') = w_1 \|\mathbf{h} - \mathbf{h}'\|_1 + w_2 \|\boldsymbol{\mu}_y - \boldsymbol{\mu}'_y\|_1 + w_3 \|\boldsymbol{\sigma}_y - \boldsymbol{\sigma}'_y\|_1 \quad (6)$$

其中, $\mathbf{S} = \{ \mathbf{h}; \boldsymbol{\mu}_y; \boldsymbol{\sigma}_y \}$ 和 $\mathbf{S}' = \{ \mathbf{h}'; \boldsymbol{\mu}'_y; \boldsymbol{\sigma}'_y \}$ 分别代表两张图片的优化空间直方图。特征的权重在 VIPeR 图库上使用前 100 对人体图片进行交叉验证实验, 分别确定为 0.5, 0.2 和 0.3。经测试验证, 实际上这 3 类特征的相对重要性在各类场景中基本保持不变, 因此后文实验均采用上述权重。

2.2 多统计性金字塔协方差(MSPC)

协方差描述子是一种与直方图截然不同的统计描述子, 它通过计算图像区域内多种特征的协方差矩阵来融合多种特征^[20]。本节基于上节对人体再识

别的若干认识,使用协方差替换空间直方图对区域进行统计描述,区域采用多层金字塔型区域,像素特征添加梯度信息。

2.2.1 协方差特征提取 本文 MSPC 特征提取时同样采用多层多区域统计模式,采用金字塔型统计区域。以 iLIDS 图库^[21]的图片为例,详细分层区域如图 2 所示,图中从左到右依次为原图、网格划分图以及 3 层金字塔型区域划分。



图2 MSPC方法分层区域划分

在图 2 中需要计算每个区域的协方差矩阵。对于某个图像区域 R , 其协方差矩阵 C_R 可计算为^[20]

$$C_R = \frac{1}{n-1} \sum_{i=0}^n (z_i - \mu)(z_i - \mu)^T \quad (7)$$

其中, n 为区域的总像素, z_i 为使用映射函数 F 从区域第 i 个像素提取的 d 维特征向量, $\mu = \sum_{i=1}^n z_i / n$ 为区域均值特征向量。本文采用的像素特征映射函数 F 为

$$F(x, y) = [y, Y, C_b, C_r, H, S, nR, nG, Y_x, Y_y, Y_{xx}, Y_{yy}]^T \quad (8)$$

其中, y 为像素点的垂直空间坐标, 采用的颜色通道与 2.1.2 节相同(此处去除冗余的 nB 颜色通道), 后面添加的 4 个特征分别为亮度通道在 x 和 y 方向上的一阶二阶梯度。

2.2.2 协方差多统计性和测度选取 与直方图描述子不同, 协方差矩阵在计算测度时不位于欧式空间, 不能直接使用常用的欧式空间距离进行计算。文献[20]中提出的测度方法是目前最常用的协方差测度

$$\rho(C_1, C_2) = \sqrt{\sum_{i=1}^d \ln^2 \lambda_i(C_1, C_2)} \quad (9)$$

其中, d 为协方差的统计特征总数, $\lambda_i(C_1, C_2)$ 为协方差矩阵 C_1 和 C_2 的广义特征值。

对于包含多个区域协方差矩阵的图像特征集, SCR 方法首先基于上述测度累积两帧图像所有对应区域的协方差距离, 然后从中剔除前 20% 的最大距离, 形成两帧图像的最终距离。这样的测度方式对于较多的区域协方差而言计算量极大, 从实验效果来看识别性能也不够理想(参见本文 4.1.1 节)。

文献[22]提出对协方差矩阵进行 Cholesky 分

解, 以将协方差转换为欧式空间下的 Sigma 集。Sigma 集能够近似地保持协方差的统计特性, 本文受其启发提出从协方差矩阵抽取多种统计向量来进行区域描述。选取的统计向量包括特征均值向量、特征标准差向量、特征相关系数向量和 Sigma 向量。

这 4 类统计向量的构造如下: 特征均值向量即式(7)中的 d 维向量 μ ; 提取协方差矩阵主对角线上的方差值进行开方运算, 得到 d 维标准差向量; 将协方差转换成相关系数矩阵, 然后取其下三角部分(不含主对角线)进行向量化, 形成 $d(d-1)/2$ 维相关系数向量; 将 d 阶协方差矩阵 C 进行 Cholesky 分解 $C = LL^T$, 获得对称阵 L 后向量化其下三角部分, 形成 $d(d+1)/2$ 维的 Sigma 向量^[22]。根据式(8)的映射函数可知 $d = 12$, 将 4 类统计向量分别在图 2 所有的 15 个区域上进行级联, 分别形成 180, 180, 990 和 1170 维度的级联统计特征向量, 分别记为 f_1, f_2, f_3, f_4 。

获取 4 类统计特征向量后, MSPC 方法的测度同样设计为 4 类特征向量 l_i 距离的加权组合

$$d(F, F') = \sum_{i=1}^4 w_i \|f_i - f'_i\|_1 \quad (10)$$

其中, $F = \{f_i, i = 1, 2, \dots, 4\}$ 和 $F' = \{f'_i, i = 1, 2, \dots, 4\}$ 分别代表两张图片的 MSPC 特征。特征的权重分别确定为 0.2, 0.1, 0.35 和 0.35。这 4 类特征在识别中的相对重要性在各类场景也基本保持不变, 后文实验均采用上述权重。

2.3 融合特征描述(HSCD)

上面提出的两种统计方法都使用了颜色和结构信息, 但它们截然不同的特征统计方式使它们具有融合增强的潜质。MCSH 特征是在多种颜色通道上级联多区域的空间和直方图信息, 就颜色通道而言属于纵向级联; MSPC 特征则是在多个区域上级联多种颜色和梯度通道的综合信息, 就颜色通道而言属于横向融合。

因此, 本文 HSCD 方法将上述两种统计方法进行融合, 采用加权组合的方式

$$d(I, I') = w_1 d(S, S') + w_2 d(F, F') \quad (11)$$

其中, $I = \{S, F\}$ 和 $I' = \{S', F'\}$ 分别代表两张图片的两类统计特征, 权重均设为 0.5。

3 再识别性能评估

目前文献通常采用累积匹配特性曲线(Cumulative Matching Characteristic, CMC)^[1]来评估人体再识别性能。CMC 曲线反映的是在目标图库前 n 个最相似候选目标内找到探测图片正确匹配的概率。不同方法的 CMC 曲线间可能存在交叉, 目

前没有统一的方式来综合评估 CMC 曲线的性能。通常 CMC 曲线下的归一化面积(normalized Area Under CMC, nAUC)能反映出 CMC 曲线的整体性能。文献[18]最近提出一种基于信息熵的指标 PUR, 该指标实际上反映的仍然是 CMC 曲线的整体性能。但目前多数文献只画出了 CMC 曲线的前面部分, 在没有代码和 CMC 数据的情况下也难和其它方法进行基于整体指标的性能比较。

事实上, 仅仅比较曲线的整体性能可能并不利于实际应用。例如, 目标重捕通常选取目标库少量的最相似候选目标, 尤其是相似度最高的候选目标。因此, 从应用的角度出发, 本文认为 CMC 曲线真正有意义的部分是曲线前面部分。故本文提出使用 CMC 曲线的 Rank-1, Rank-10%和 nAUC-10% 3 类指标合成实用性综合指标, 记为 ACI(Application based CMC Index)。其中, Rank-1 指标反映最相似候选目标刚好为探测图片目标的概率; Rank-10% 指标反映在图库目标总数 10%的候选目标内找到探测图片正确匹配的概率; nAUC-10%指标代表 CMC 曲线在 Rank-1 和 Rank-10%范围间的归一化面积。对所有参与比较的 CMC 曲线, 3 类指标分别 t_i 归一后再加权相加, 合成反映相对实用性能的 ACI 综合指标, 3 类指标的权重分别设为 0.25, 0.25, 0.5。采用 ACI 指标能够更有效地比较目前多数文献的 CMC 曲线, 而且能够更合理有效地反映出实际应用的识别性能。

4 实验结果及分析

本文在两个常用图库上进行了广泛的实验比较。本文方法在 Intel Xeon 2 GHz 服务器平台上使用 Matlab 实现, 实验首先在 VIPeR 图库上进行单帧模式下的性能分析和比较, 然后在 i-LIDS 图库上进一步验证多帧模式下的性能。

4.1 VIPeR 图库实验

VIPeR 图库是再识别文献中使用最多的单帧模式标准图库, 它包含 632 对不同视角、姿态和光照条件下的行人。本文采用文献[1]提出的 5 轮 2 折交叉验证方法计算平均的 CMC 曲线。实验采用与文献[6]相同的测试集划分, 每轮实验从图库选出 316 对行人, 其中一半用作目标图库, 另一半用作探测集。

4.1.1 HSCD 方法分析

本文 HSCD 方法由 MCSH 和 MSPC 两种方法融合而成, 实验首先考察这两种方法的性能和方法内部特征的性能。为了验证方法在计算时间上的优势, MCSH 和 MSPC 方法分别考察了式(3)测度^[10]

和 SCR 方法测度^[9]的识别性能和计算时间。实验还将 HSCD 方法和两种公开源码的方法(SDALF 特征方法^[9]和 PRDC 测度方法^[17])进行了性能和时间上的综合比较。实验结果如图 3 所示, 3 个子图分别画出了 MCSH, MSPC 和 HSCD 方法的相关曲线, 子图右下角标注处列出了方法的内部特征和部分方法的时间指标(即图库所有图片的特征提取时间 T_1 , 特征距离计算时间 T_2 以及距离测度训练时间 T_3)。

图 3(a)表明在 MCSH 方法的 3 类特征中, 直方图信息优于空间信息, y 方向的二阶空间信息优于一阶信息。式(3)空间直方图的传统测度方法计算时间极长, 性能却略差于直方图方法。图 3(b)表明 MSPC 方法中的均值、相关系数和 Sigma 向量性能相当, 均优于二阶方差信息。MSPC 方法的 4 类特征存在极强的互补性, 组合后其性能得到了很大提升, 但其协方差传统测度方法同样存在计算时间长和性能较差的问题。

图 3(c)表明 MCSH 和 MSPC 两种统计方法具有较强的互补性, 融合后的 HSCD 方法优于 SDALF 特征方法和 PRDC 测度方法。在时间方面, 本文方法不需要训练, 且特征提取和测度计算的时间耗费非常小。SDALF 方法依赖分割检测等预处理, 特征提取时间最长; PRDC 方法则训练时间长、训练时耗费内存量大(10 G 左右内存)。

4.1.2 当前方法比较 实验进一步和目前已有的特征和测度类方法进行比较。实验结果如图 4 所示, 其中比较方法的 CMC 曲线均从相应文献或代码中获得, 图中竖直虚线代表目标库 Rank-10%(即 32)的位置。

表 1 列出了图 4 中 CMC 曲线的各项性能指标, 包括排序为 1, 10%和 50 时的识别率以及 nAUC-10%和 ACI 综合指标。

根据图 4 曲线和表 1 性能指标可得到如下结论。(1)在特征类方法中, HSCD 方法各项性能指标均优于其它方法, 包括目前最新的 SCEAF 多特征组合方法和增强型 eSDC 方法; (2)即使与测度类方法相比, HSCD 仍然优于多数方法, CMC 曲线综合性能只略差于最新的 LF 测度方法。但本文 HSCD 方法的 Rank-1 指标达到 29.3%, 高出 LF 约 21%, 导致其 ACI 指标略高于 LF; (3)与特征类方法相比, 测度类方法主要增强了 CMC 曲线的后面部分, 测度方法在训练时存在的过度拟合等现象可能是导致曲线 Rank-1 取值较低的原因。

4.2 i-LIDS 图库实验

i-LIDS 图库是常用的多帧模式人体再识别图库, 该图库包含某机场 119 个行人的 476 张图片(每

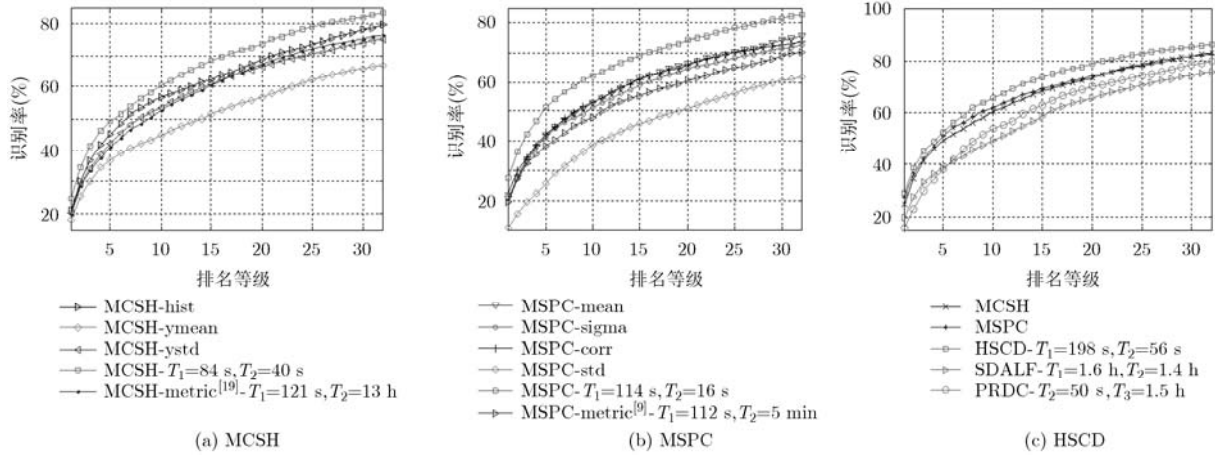


图 3 VIPeR 图库上 HSCD 方法的分析比较

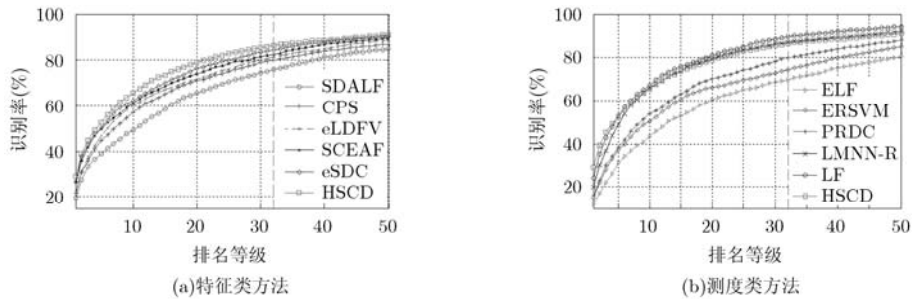


图 4 VIPeR 图库上各方法的 CMC 曲线

表 1 VIPeR 图库上各方法的识别率比较(%)

表 1 (a) 特征类方法性能指标

方法	文献	$R=1$	$R=10\%$	$R=50$	nAUC-10%	ACI
SDALF	CVPR 2010 ^[3]	19.87	75.79	84.84	57.34	79.42
CPS	BMVC 2010 ^[4]	21.84	80.00	87.00	62.68	86.10
eLDFV	ECCV 2012 ^[6]	22.34	81.24	88.82	64.35	88.07
SCEAF	CVPR 2013 ^[8]	26.49	82.60	89.50	66.43	93.47
eSDC	CVPR 2013 ^[11]	26.74	84.50	90.50	67.97	95.32
HSCD	本文	29.30	86.46	90.90	70.69	100.0

表 1 (b) 测度类方法性能指标

方法	文献	$R=1$	$R=10\%$	$R=50$	nAUC-10%	ACI
ELF	ECCV 2008 ^[12]	12.00	70.00	80.30	50.94	65.07
ERSVM	BMVC 2010 ^[14]	15.00	74.40	85.00	56.78	72.91
PRDC	PAMI 2013 ^[17]	15.66	80.00	88.00	60.09	77.33
LMNN-R	ACCV 2010 ^[16]	20.00	87.00	92.00	69.85	89.75
LF	CVPR 2013 ^[18]	24.18	89.50	94.12	72.18	95.63
HSCD	本文	29.30	86.46	90.90	70.67	98.12

个行人平均 4 张图片)。该图库不仅存在行人视角、姿态和光照方面的变化，还存在较严重的部分遮挡

现象。

本文在该图库上进一步验证 HSCD 方法在单帧

和多帧模式下的性能。单帧模式下每个行人随机选取一张图片构成目标库,所有剩余图片构成探测集。多帧模式下每个行人随机选取两张图片构成目标库,再从剩余图片随机选取两张构成探测集(对于只有 2 张或 3 张图片的行人,分别随机选取 1 张和 2 张放入目标库,其余放入探测集)。单帧和多帧模式均进行 100 次实验,实验结果如图 5 所示。

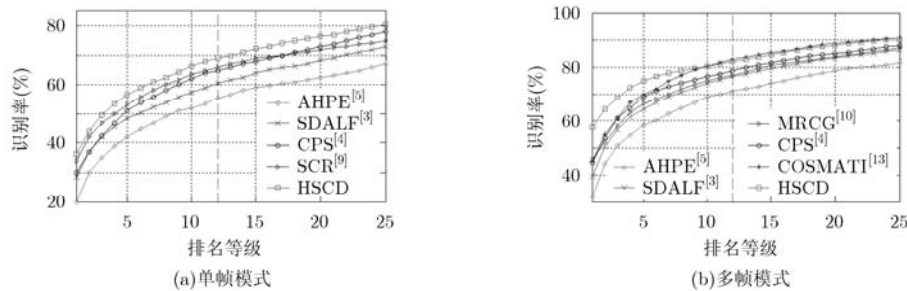


图 5 i-LIDS 图库上各方法的 CMC 曲线

以上实验结果表明,本文提出的 HSCD 方法在单帧和多帧模式下均可取得很好的识别性能。与其它方法相比,本文方法的 ACI 应用型指标更高,计算时间更少,更利于在实际应用中进行实时的人体再识别。

5 结束语

本文成功构造了基于空间直方图和协方差的两种外观统计描述方法,然后融合得到了性能优异的 HSCD 方法。该方法除了识别性能上的优势,其它优势还包括:(1)属于特征类方法,不需要进行监督性训练;(2)特征性能优异,MCSH 和 MSPC 方法均不需要进行分割检测等预处理步骤,特征提取迅速;(3)测度简单,本文将传统测度替换为简单的 l_1 距离加权组合测度,更利于高效的距离计算;(4)实用,本文方法 ACI 应用性能指标优异,有利于实时的视觉应用。

鉴于人体再识别存在的挑战性和对其它应用的重要性,下一步可能的研究工作还包括:(1)考察与颜色和结构信息具有较强互补性的其它特征,并寻找其它优秀的特征描述方法;(2)探讨针对各影响因素下的处理策略,如颜色和光照处理等;(3)基于本文特征考察和设计合适的测度学习方法;(4)将再识别方法引入到多机跟踪等应用领域,将人体再识别推广到其它目标的再识别问题,考察方法的实用性能和推广能力。

参考文献

[1] Gray D, Brennan S, and Tao H. Evaluating appearance

从图 5 的 CMC 曲线可看出, HSCD 方法在单帧和多帧模式下均优于其它方法。单帧模式下 HSCD 的 Rank-1 为 36%,多帧模式下上升到 58%,表明了 HSCD 方法在多帧模式下显著的增强作用。在多帧模式下,只有基于协方差测度学习的 COSMATI 方法能在 CMC 曲线的后半段略优于 HSCD,其它方法均明显差于 HSCD。

models for recognition, reacquisition, and tracking[C]. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance. Rio de Janeiro, Brazil, 2007: 41-47.

[2] Wang H, Bao X, Choudhury R R, *et al.* Insight: recognizing humans without face recognition[C]. Proceedings of the 14th Workshop on Mobile Computing Systems and Applications, Georgia, USA, 2013, 7.

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

[4] Cheng D S, Cristani M, Stoppa M, *et al.* Custom pictorial structures for re-identification[C]. British Machine Vision Conference, Dundee, UK, 2011, 6.

[5] 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.

[6] Ma B, Su Y, and Jurie F. Local descriptors encoded by fisher vectors for person re-identification[C]. European Conference on Computer Vision, Florence, Italy, 2012: 413-422.

[7] 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.

Fan Cai-xia, Zhu Hong, Lin Guang-feng, *et al.* Person re-identification based on multi-features[J]. *Journal of Image and Graphics*, 2013, 18(6): 711-717.

[8] Hu Y, Liao S, Lei Z, *et al.* Exploring structural information and fusing multiple features for person re-identification[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Portland, USA, 2013: 794-799.

- [9] Bak S, Corvee E, Brémond F, *et al.*. Person re-identification using spatial covariance regions of human body parts[C]. IEEE International Conference on Advanced Video and Signal Based Surveillance, Boston, USA, 2010: 435-440.
- [10] Bak S, Corvee E, Bremond F, *et al.*. Multiple-shot human re-identification by mean riemannian covariance grid[C]. IEEE International Conference on Advanced Video and Signal-Based Surveillance, Klagenfurt, Austria, 2011: 179-184.
- [11] Zhao R, Ouyang W, and Wang X. Unsupervised salience learning for person re-identification[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Portland, USA, 2013: 4321-4328.
- [12] Gray D and Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features[C]. European Conference on Computer Vision, Marseille, France, 2008: 262-275.
- [13] Bak S, Charpiat G, Corvée E, *et al.*. Learning to match appearances by correlations in a covariance metric space[C]. European Conference on Computer Vision, Florence, USA, 2012: 806-820.
- [14] Prosser B, Zheng W S, Gong S, *et al.*. Person re-identification by support vector ranking[C]. British Machine Vision Conference, Wales, UK, 2010, DOI: 10.5244/c.24.21.
- [15] 陈方, 许允喜. 视频监控中基于在线多核学习的目标再识别[J]. 光电工程, 2012, 39(9): 65-71.
Chen Fang and Xu Yun-xi. People re-identification based on online multiple kernel learning in video surveillance[J]. *Opto-Electronic Engineering*, 2012, 39(9): 65-71.
- [16] Dikmen M, Akbas E, Huang T S, *et al.*. Pedestrian recognition with a learned metric[C]. Asian Conference on Computer Vision, Queenstown, New Zealand, 2010: 501-512.
- [17] Zheng W S, Gong S, and Xiang T. Reidentification by relative distance comparison[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(3): 653-668.
- [18] Pedagadi S, Orwell J, Velastin S, *et al.*. Local fisher discriminant analysis for pedestrian re-identification[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Portland, USA, 2013: 3318-3325.
- [19] Birchfield S T and Rangarajan S. Spatiograms versus histograms for region-based tracking[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005: 1158-1163.
- [20] Tuzel O, Porikli F, and Meer P. Region covariance: a fast descriptor for detection and classification[C]. European Conference on Computer Vision, Graz, Austria, 2006: 589-600.
- [21] Zheng W, Gong S, and Xiang T. Associating groups of people[C]. British Machine Vision Conference, London, UK, 2009: 1-11.
- [22] Hong X, Chang H, and Shan S. Sigma set: a small second order statistical region descriptor[C]. IEEE Conference on Computer Vision and Pattern Recognition, Miami, USA, 2009: 1802-1809.
- 曾明勇：男，1987年生，硕士生，研究方向为图像处理、计算机视觉。
- 吴泽民：男，1973年生，副教授，博士，研究方向为数据融合、图像识别。
- 田畅：男，1963年生，教授，博士生导师，研究方向为数据链系统、计算机网络、图像分析。