

指导教师： 杨 涛

提交时间： 2016/3/19

# CVPR2015 Paper Translation

No: 01

姓名： 戎思安

学号： 2013302557

班号： 10011304

# 通过局部最大事件与度量学习

## 实现人的重新识别

廖胜才, 胡杨, 朱翔宇, 和李子青

生物识别技术与安全研究中心, 模式识别国家实验室, 自动化研究所, 中国科学院, 中关村东路 95 号, 北京, 100190, 中国。

{scliao, yhu, xiangyu. zhu, szli}@nlpr.ia.ac.cn

### 摘要

在监控视频中, 为了对人的存在进行自动搜索, 人的再识别是一项重要的技术。对人的重新识别、特征表示和度量学习, 有两个基本的问题是关键的。一个有效的特征表示应该稳健地反映光照和视点的变化, 和度量的判别应该学会与各种人物形象相匹配。在这份研究报告中, 我们提出了一个有效的称作局部最大事件表示 (LOMO) 的特征表示方法, 和一个称作交叉视图二次判别分析 (XQDA) 的子空间与度量学习方法。

LOMO 特征分析了局部特征的水平发生, 和使发生事件最大化, 形成一个稳定的表示, 而不是视点的变化。除此以外, 为了处理光照的变化, 我们采用同态滤波变换和尺度不变结构算子。为了学习度量判别式, 我们提出了通过交叉视图二次判别分析来学习在低维子空间表达式, 并且同时 QDA (二次判别分析) 度量被用来在衍生子空间中学习。我们还为 XQDA 提出一种实用的计算方法, 以及它的规则化。对四组具有挑战性的人的重新识别数据库的试验, 这些数据库是 VIPeR, QMUL GRID, CUHK Campus, 和 CUHK03, 显示了所采用的方法对在这四个数据库中使用最先进技术的 1 级别识别率分别提高了 2.2%, 4.88%, 28.91% 和 31.55%。

### 1. 介绍

人的重新识别是从图库中寻找对探测器来说具有相同身份的人的问题。这是一个具有挑战性的问题, 因为由于在光照, 姿势或视点, 和闭塞中级别大的变化。许多方法已经被采用来进行人的重新识别, 已经大大推进了这一领域。

对人的重新识别、特征表示和度量学习来说, 有两个基本的问题是关键的。一个有效的特征表示应该稳健地反映光照和视点的变化, 和度量的判别应该学会与各种人物形象相匹配。已采取许多努力沿着双向来解决人的重新识别的挑战。对于特征表示, 几种有用的方法已被采用, 例如, 局部特征的集成 (ELF), SDALF, kBiCov, fisher 矢量 (LDFV), 特征匹配, 和中级滤波。这些手工制作或者学习基于描述符的已经比稳健的特征表示起到了令人印象深刻的改进, 并且推进了人的重新识别的研究。然而, 对于人的重新识别来说, 如何设计或者学习稳健的特征仍然是一个公开未解决的问题。

关于人的重新识别的另一方面是如何学习一个鲁棒的距离或者近似函数来解决复杂的匹配问题。许多度量学习算法已经在这方面被考虑采用。在操作中, 许多以前的度量学习方法在度量学习上展示了一个有两个阶段的进展, 就是, 要素分析原则 (PCA) 在 PCA 子空间已被执行。然而, 这个包含个阶段的进程也许不是在低维空间进行度量学习最优的方法, 因为不同类的样本也许在第一阶段后变杂乱

了。

在这份报告中，我们提出了一个有效的叫做局部最大发生度 (LOMO) 的特征识别，和一个叫做交叉视图二次判别分析的子空间与度量学习方法。LOMO 的特征是分析局部特征水平发生度，和为了阻止视点改变最大化的形成稳定的表示。除此之外，我们发现采用同态滤波来解决在人的识别的光照变化是有用的。为了学习度量判别，我们提出通过交叉视图二次判别分析来学习低维子空间的判别，与此同时，在派生子空间学习 QDA (二次判别分析) 度量。我们展示这个问题能被广义 Rayleigh 商确切地表示，并且一个封闭解决方案能由广义特征值分解得到。我们还为 XQDA 提出了一个实用的计算方法，以及它的正规化和维数选择。所提出的方法显示能够在四个公共数据库中进行人的重新识别是有效和高效的，并且我们也证明所提出的部分如何引起改进。

## 2. 相关工作

许多现有的用来进行人的重新识别的方法试图去创建一个鲁棒的特征表示，这个表示在各种状况下既与众不同又能稳健地来描述一个人的相貌。Gray 和 Tao 提出使用 AdaBoost (自适应提升算法) 来从颜色和纹理特征集中选择好的功能。Farenzena 等人提出局部特征的对称性驱动 (SDALF) 的方法，其中的对称性与非对称性被认为是用来解决视点变化。Ma 等人将 Fisher 矢量中的局部描述符转变成产生一种图像的全局表示。Cheng 等人利用图案结构，其中的部分基础颜色信息和颜色位移被认为是用来人的重新识别。现在，为了人的重新识别，显著的信息已经被研究，引起新的特征表示。一个叫做 regionlets 的方法已被提出，此方法是从三个随机的区域选择一个最大槽用来在变形下进行目标检

测。相比之下，我们提出在所有的水平子窗口之中使每个局部模式的发生最大化从而解决视点变化。

除了鲁棒性质的特征之外，度量学习已经被广泛应用于人的重新识别。Zheng 等人提出 PRDC 算法，用来优化相对距离比较。Hirzer 等人提出在 Mahalanobis (马氏; 马哈拉诺比斯) 的度量学习放宽 PSD (Phase Sensitive Demodulator 相位灵敏调解器) 的约束，并且得到的简化规则仍然展示出有前途的性能。Li 等人提出局部自适应判决功能 (LADF) 的用来人的认证的学习方法，它可以被看作是距离度量的联合模型和局部自适应的阈值规则。Prosser 等人制定了人的重新识别问题作为排名问题，并且运用 RankSVM 来学习子空间。在当地专家中，被认为学习共同特征空间来进行人的重新鉴定是跨越意见。

除了一个新的特征表示，被采用的 XQDA 算法是主要与贝叶斯人脸，KISSME，线性判别分析 (LDA)，局部 Fisher 判别分析 (LF)，和 CFML 相关。XQDA 可以被看作是贝叶斯人脸和 KISSME 的扩展，这其中一个是判别的子空间和度量可以被进一步了解到。LF 方法适用于 FDA 一起与 PCA 和 LPP 得到一个低维判别子空间。这 CFML 算法目的在不同的问题，但是因为 XQDA 学会了相似的子空间。然而，无论 LF 还是 CFML，都在派生子空间使用欧几里得距离，尽管提出的方法考虑一个判别子空间和一个综合指标。对于传统的 LDA 方法，尽管 XQDA 共享一个类似的广义 Rayleigh 商的公式，这等于是说明，它们基本上不等同。

## 3. 局部最大发生度的特征

### 3.1. 光照变化的处理

颜色是用来描述人的图像一个重要的特征。然而，通过相机的光照条件

是很不同的，并且各个相机的设置可能也不同。因此，从不同相机视角关于同一个人的感知颜色也许会有很大的变化。例如，图 1(a)展示一些来自 VIPeR 数据库的样本图像。它可以被看出同一个人的图像经过两个摄像机拍摄在光照和色表上有很大的不同。



图 1. (a) 图像例子对来自 VIPeR 数据库。(b) 在(a)中经过 Retinex (同态滤波)处理过的图像。在同一图像列中代表同一个人。

在这份报告中，我们提出了运用同态滤波算法来预处理人的图像。同态滤波考虑人类的亮度和色彩感知。它的目的是产生和人类观测场景相一致的彩色图像，尤其是增强在阴影区域的细节。

我们根据结合了同时在动态范围压缩下的小规模同态滤波和色调再现下的大规模的同态滤波，来实行尺度变换算法。结果，此算法自动同时处理颜色的恒常性和动态范围压缩，得到了一个好的近似于人的视觉感知。特别地，我们使用双尺度的中心/环绕的同态滤波，其中 $\sigma = 5$ 和 $\sigma = 20$ 。除此之外，我们自动计算增益/偏移参数，这样，所产生的线性强度在[0,255]之间。

图 1(b)显示了一些经过我们同态滤波处理过的图像的例子。与图像 1(a)相比，它可以观测到通过不同相机的同一个人的经过同态滤波后的图像在亮度和颜色上有更好的一致性。这使得人的重新识别比使用原始图像要容

易得多。以经过同态滤波后的图像，我们应用 HSV(HSV 色彩模型(Hue 色度, Saturation 饱和度, Value 纯度))颜色直方图来提取颜色特征。

除了颜色描述之外，我们也在光照不变纹理描述中应用规模不变局部三元模式 (SILTP) [26]描述符。SILTP 是一个在著名的局部二元模式(Local Binary Pattern, LBP) [37]的改进算子。事实上，LBP 在单调灰度变换下有很出色的不变性，但是对图像噪声而言，它不具有鲁棒的健壮稳定性。SILTP 通过引入规模不变局部比较公差来改进 LBP，得到强度级别变化的不变性和图像噪声的鲁棒性。

### 3.2. 视点变化的处理

行人在不同的摄像机中经常出现在不同的视点。例如，一个人在一个摄像机中是正面视图，可能在另一个摄像机下是后视图。因此，在不同的视点下进行人的匹配也是困难的。为了解决这个问题，[39,49]提出将在一个人的图像平均分成六个横条纹，和一个单一的直方图计算在每一个条纹中。这个特征在视点不变人的表示中已经取得了成功[39,49,27]。然而，它可能会在条纹中丢失空间细节特征，从而影响它的判别力。

我们提出使用滑动窗口来描述一个人的图像的局部细节。特别地，我们使用大小为  $10 \times 10$  的子窗口，在  $128 \times 48$  的图像中使用 5 个像素的重叠步骤来定位局部补丁。在每个子窗口中，我们提取 SILTP 直方图的两个级别

( $SILTP_{4,3}^{0.3}$ 和 $SILTP_{4,5}^{0.3}$ )，和一个  $8 \times 8 \times$

$8$ 的槽的联合 HSV 直方图。每个直方图槽在子空间中代表一种模式的出现概率。为了解决视点的变化，我们在同一水平位置下检查每个子窗口，并且在这些子窗口下使每个模式 (即相同的直方图槽) 的局部发生度最大化。由此产生的直方图在视点变化中得到一些

不变性，并且同时捕获人的局部区域特征。图 2 显示所提出的 LOMO 的特征提取步骤。

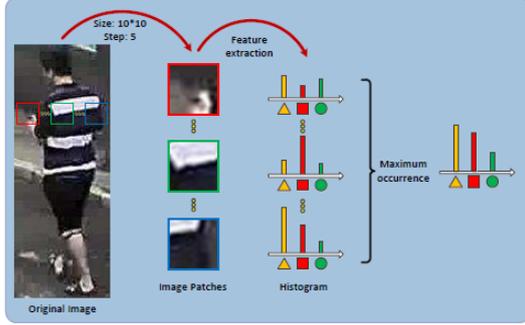


图 2。LOMO 特征提取方法的示意图。

为了进一步考虑多尺度的信息，我们创建了一个三棱锥模型表示，通过两个  $2 \times 2$  局部平均集合操作对原始的  $128 \times 48$  的图像进行下采样，并且重复上述特征提取步骤。通过连接所有的经过计算的局部最大事件，我们最终的描述符是  $(8 * 8 * 8 \text{ 颜色槽} + 34 * 2 \text{ SILTP 槽}) * (24+11+5 \text{ 水平组}) = 26,960$  的尺寸。最后，我们应用对谁转换来抑制大的槽值，并且使 HSV 和 SILTP 特征规范化为单位长度。因为我们只使用了简单的 HSV 和 SILTP 特征，对计算来说所提出的特征提取方法是有效的（参见 5.5.4 节）。

## 4. 交叉视图二次判别分析

### 4.1. 贝叶斯人脸和 KISSME 重温

考虑一个样本的差异  $\Delta = X_i - X_j$ 。

如果  $y_i = y_j$ ， $\Delta$  被称作内省差异，但如果

如果  $y_i \neq y_j$  [36]，它被称作外省差异。于是，变化的两类可以被定义：内省变化

$\Omega_I$  和外省变化  $\Omega_E$ 。因此，在这种方式下可以通过区分上述两类，来解决多级分类的问题。Moghaddam 等人 [36] 提出通过多元高斯分布来是两类中的每一个模型化。通过被定义的  $\Omega_I$  和  $\Omega_E$  作

为两类，这相当于 QDA 模型。此外，它还被发现无论是  $\Omega_I$  还是  $\Omega_E$  都具有零均值。由此产生的算法叫做贝叶斯人脸，可被应用于面部识别。有趣的是，在 [18] 中，Köstinger 等人通过两个高斯分布的对数似然比的检验，也得出类似的叫做 KISSME 的方法，并将它应用于人的重新识别。

正式的，贝叶斯人脸和 KISSME 算法都被制定如下。在零均值的高斯分布下，观测的  $\Delta$  在  $\Omega_I$  和  $\Omega_E$  中的似然比定义为：

$$P(\Delta|\Omega_I) = \frac{1}{(2\pi)^{d/2} |\Sigma_I|^{1/2}} e^{-\frac{1}{2} \Delta^T \Sigma_I^{-1} \Delta}, \quad (1)$$

$$P(\Delta|\Omega_E) = \frac{1}{(2\pi)^{d/2} |\Sigma_E|^{1/2}} e^{-\frac{1}{2} \Delta^T \Sigma_E^{-1} \Delta}, \quad (2)$$

其中  $\Sigma_I$  和  $\Sigma_E$  分别是  $\Omega_I$  和  $\Omega_E$  协方差矩阵，并且  $n_I$  和  $n_E$  在两类中表示样本的数目。通过应用贝叶斯法则和对数似然比的测试，决策函数可以被简化为：

$$f(\Delta) = \Delta^T (\Sigma_I^{-1} - \Sigma_E^{-1}) \Delta, \quad (3)$$

所以在  $x_i$  和  $x_j$  之间派生距离函数为：

$$d(x_i, x_j) = (x_i - x_j)^T (\Sigma_I^{-1} - \Sigma_E^{-1}) (x_i - x_j). \quad (4)$$

因此，学习距离函数对应估计协方差矩阵  $\Sigma_I$  和  $\Sigma_E$ 。

### 4.2. XQDA

通常情况下，原始的特征维度  $d$  是大的，并且低维空间  $\mathbb{R}^r$  ( $r < d$ ) 是进行分类的优先选择。[36] 建议分别分解  $\Sigma_I$  和  $\Sigma_E$  来降低维度。在 [18] 中，应用了 PCA，然后  $\Sigma_I$  和  $\Sigma_E$  在 PCA 子空间中被估计。然而，这两种方法并不是最佳的，因为降维并未考虑距离度量学习。

在这份报告中，我们将贝叶斯人脸和 KISSME 方法扩展到度量学习，其中我们考虑到通过交叉数据学习子空间  $W = (w_1, w_2, \dots, w_r) \in \mathbb{R}^{d \times r}$ ，并且同时在  $r$  维子空间学习距离函数用来交叉视图相似性的度量。假设我们有

一个关于  $c$  类交叉视图训练集  $\{X, Z\}$ , 其中  $X = (x_1, x_2, \dots, x_n) \in \mathbb{R}^{d \times n}$  包含来自一个视图的一个  $d$  维空间的  $n$  个样本,  $Z = (z_1, z_2, \dots, z_m) \in \mathbb{R}^{d \times m}$  包含在相同的  $d$  维空间却是来自其它视图的  $m$  个样本。这个交叉视图匹配问题源于许多的应用, 像合成脸的识别[25]和视角不同的人的重新识别[10]。注意在单个视图匹配方案中  $Z$  和  $X$  是相同的。

考虑一个  $W$  子空间, 距离函数相等式(4)在  $r$  维子空间中被计算为

$$d_W(x, z) = (x - z)^T W (\Sigma_I'^{-1} - \Sigma_E'^{-1}) W^T (x - z), \quad (5)$$

其中  $\Sigma_I' = W^T \Sigma_I W$ ,  $\Sigma_E' = W^T \Sigma_E W$ 。因此, 我们需要来学习核心矩阵  $M(W) = W (\Sigma_I'^{-1} - \Sigma_E'^{-1}) W^T$ 。然而, 直接优化  $d_W$  是困难的, 因为  $W$  包含在两个逆矩阵中。

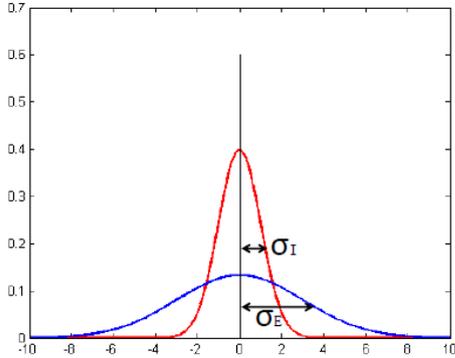


图 3。在一个投影维度中  $\Omega_I$  和  $\Omega_E$  的分布。

回想  $\Omega_I$  和  $\Omega_E$  有零均值, 然后依据  $w$  给出, 这两类投影样本仍然集中在零, 然而也许会有不同的方差, 就像图 3 展示的那样。在这种情况下, 被用来推导 LDA 的传统的 Fisher 准则就不再适合了, 因为这两类有相同的平均值。然而, 方差  $\sigma_I$  和  $\sigma_E$  仍然可以用来区分这两类。因此, 我们可以优化投影方向  $w$ , 这样  $\sigma_E(w)/\sigma_I(w)$  被最大化。注意  $\sigma_I(w) =$

$w^T \Sigma_I w$  和  $\sigma_E(w) = w^T \Sigma_E w$ , 因此, 目标  $\sigma_E(w)/\sigma_I(w)$  对应于广义 Rayleigh 商

$$J(w) = \frac{w^T \Sigma_E w}{w^T \Sigma_I w}. \quad (6)$$

最大化的  $J(w)$  相当于

$$\max_w w^T \Sigma_E w, \text{ s.t. } w^T \Sigma_I w = 1, \quad (7)$$

可以通过在 LDA 中类似的广义特征值分解问题来得到解决。就是说,  $\Sigma_I^{-1} \Sigma_E$  最大的特征值就是  $J(w)$  的最大值。此外, 该方案的对  $w_1$  正交和对应  $J(w)$  的第二大值是  $\Sigma_I^{-1} \Sigma_E$  的第二大特征值的特征向量, 等等。因此, 通过  $W = (w_1, w_2, \dots, w_r)$ , 我们学习判别子空间, 以及在学习子空间的距离函数, 就像在等式(5)定义的。我们称这衍生算法交叉视图二次判别分析(XQDA)来反映它的与 QDA 的联系和交叉视图度量的输出。

### 4.3. 实用计算

$\Sigma_I$  和  $\Sigma_E$  的两个协方差矩阵的计算分别要求  $O(Nkd^2)$  和  $O(nmd^2)$  乘法运算,  $N = \max(m, n)$ , 并且  $k$  代表在每个类中的图像的数目。为了减少计算, 我们表示为

$$n_l \Sigma_l = \tilde{X} \tilde{X}^T + \tilde{Z} \tilde{Z}^T - S R^T - R S^T, \quad (8)$$

$$\text{其中 } \tilde{X} = (\sqrt{m_1} x_{11}, \sqrt{m_1} x_{12}, \dots, \sqrt{m_1} x_{n1}, \dots, \sqrt{m_c} x_{cn}), \quad \tilde{Z} = (\sqrt{n_1} z_{11}, \sqrt{n_1} z_{12}, \dots, \sqrt{n_1} z_{m1}, \dots, \sqrt{n_c} z_{cm}), \quad S = (\Sigma_{y_i=1} x_i,$$

$$\Sigma_{y_i=2} x_i, \dots, \Sigma_{y_i=c} x_i), \quad R = (\Sigma_{l_j=1} z_j,$$

$$\Sigma_{l_j=2} z_j, \dots, \Sigma_{l_j=c} z_j), \quad y_i \text{ 和 } l_j \text{ 是类标签,}$$

$n_k$  是在  $X$  中类  $k$  样本的数目。除此之外,

$$n_E \Sigma_E = m \tilde{X} \tilde{X}^T + n \tilde{Z} \tilde{Z}^T - s r^T - r s^T -$$

$$n_l \Sigma_l, \quad (9)$$

其中  $s = \Sigma_{i=1}^n x_i$  和  $r = \Sigma_{j=1}^m z_j$ 。上述简化

显示了  $\Sigma_l$  和  $\Sigma_E$  的计算都降低到  $O(Nd^2)$ 。可以观测到的是,  $\Sigma_l$  和  $\Sigma_E$  可以直接从样本平均值和每个类与所有类的协方差计算得到, 所以必要实际计算在许多其它度量学习算法中的样本区别  $mn$  对。

另外实际的问题是,  $\Sigma_l$  可能是奇数, 导致  $\Sigma_l^{-1}$  不能被计算。因此, 为  $\Sigma_l$  的对角元素增加一个小的正则化是必要的, 就像 LDA 中在相似问题里经常做的那样。这可以使  $\Sigma_l$  的估计更光滑和具有鲁棒性。经验性的, 我们发现, 当所有的样本都被标准化为单位长度, 作为正则化的 0.001 的值可以被经常使用, 以此来改善结果。

最后, 这仍然有个关于选择衍生 XQDA 子空间的维度的问题。在实际应用中, 这应该有一个考虑项是拥有一个低维的子空间以此来确保处理速度。除了这个考虑, 我们发现具有被选择的比 1 大的特征值的  $\Sigma_l^{-1} \Sigma_E$  是一个用来决定维数的有用的签名。这是因为  $\Sigma_l^{-1} \Sigma_E$  的特征值对应于图 3 中的  $\sigma_E / \sigma_l$ , 并且  $\sigma_E < \sigma_l$  可能提供有用的判决信息。

## 5. 实验

### 5.1. 在 VIPeR 中的实验

VIPeR [9] 是一个具有挑战性的人的重新识别数据库, 已经被基准广泛应用于基准评测。它包含 632 对人的图像, 这些图像是通过室外环境中的一对相机捕获的。VIPeR 的图像包括在背景下, 光照, 和视点下的巨大变化。图 1(a) 展示了一些来自 VIPeR 数据库的图像示例对。所有的图像都缩放至  $128 \times 48$  像素大小。在这个数据库中被广泛采用的实验方案是将这 632 对图像随机的划分成用于训练的一半和另外用于测试的一半, 并且将这个步骤重

复 10 次以此来得到平均的表现。我们在我们的实验中遵循这个步骤。

#### 5.1.1. 度量学习算法的比较

我们评估所提出的 XQDA 算法和一些度量学习算法, 包括欧几里得距离, 经过纯正对训练的马氏距离 [18], LMNN v2.5 [43], ITML [4], KISSME [18], 和 RLDA [45], 具有相同的 LOMO 特征。对于比较算法, PCA 是第一个被用来使特征维数减少到 100。被提出的 XQDA 算法和 RLDA 也是学习一个 100 维的子空间。由此产生的累积匹配特性 (CMC) 曲线在图 4(a) 中显示。可以看到所提出的方法比这个比较度量学习算法要好。这表明 XQDA 成功地学习了一个判别子空间以及一个有效的度量。除此之外, 我们也考察不同的子空间维度, 性能是如何变化的, 在图 4(b) 中显示。可以看到 XQDA 在所有维度下一贯表现的很好。

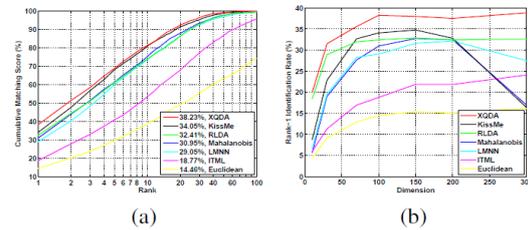


图 4. 在 VIPeR 数据库中通过有相同 LOMO 特征的度量学习算法的比较 [9] (P=316)。 (a) CMC 特征曲线, 其特征降至 100 维数。 (b) 具有不同子空间维度的 1 级别识别速率。

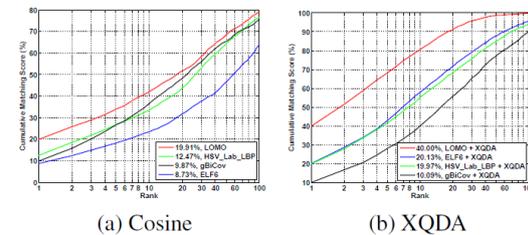


图 5. 在 VIPeR 数据库中的 CMC 曲线和 1 级别识别速率 [9] (P=316), 通过将所提出的 LOMO 特征与三种可用的特征比较, ELFF [50], HSV + Lab + LBP [18],

和 gBiCov[33]。

### 5.1.2 特征比较

接下来,我们将提出的 LOMO 特征与其它三个可用的人的重新识别特征相比较。第一个特征叫做局部特征的合成(ELF),在[10]中提出,后来经过[39,40]修改。我们在[50]使用结果,通过 ELF6 表示,在六个平等分配的水平条纹直方图中计算八色通道(RGB,HSV,和 YCbCr)和 21 纹理过滤(8 种 Gabor 过滤和 13 种 Schmid 过滤)被用来进行直方图的表示。其它的特征在[18]中被提出,其中应用了 HSV,和 Lab 颜色特征,通过 LBP 提取纹理特征。第三个特征叫做gBiCov<sup>1</sup>是生物启发特征(BIF)和协方差描述符的组合。我们采用直接余弦相似度和 XQDA 算法来比较特征的四种不同种类,所得的在图 5 的 CMC 曲线中。为了保证一致性,在接下来的实验中,我们自动地决定 XQDA 子空间维数通过接受  $\Sigma_I^{-1}\Sigma_E$  所有的特征值是大于 1 的,如前面所讨论的那样。从图 5(a)中可以看到原始的 LOMO 特征优于其它的现有的特征。更重要的是,图 5(b)显示通过 XQDA 的帮助,性能的改进更为显著。既然这些种类的特征在颜色合成和纹理信息上是相似的,通过提出的 LOMO 特征进行的改进主要是基于光照和视点变化处理的具体考虑。

<sup>1</sup> 我们使用作者的实现(在网址 <http://vipl.ict.ac.cn/members/bpma>)和默认参数,其可能没有反映出最好的状态。

图 6。在 VIPeR 数据库中 CMC 曲线和 1 级别识别速率[9](P=316),通过将提出的 LOMO+XQDA 方法与其它的最新算法比较。

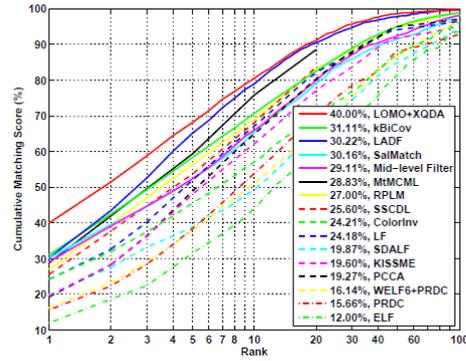


表 1。与 VIPeR 数据库中最先进的结果的比较(P=316)。所列出的是在 1,10 和 20 级别的累积匹配分数(%)。

Method	rank=1	rank=10	rank=20	Reference
LOMO+XQDA	40.00	80.51	91.08	Proposed
SCNCD	37.80	81.20	90.40	2014 ECCV [44]
kBiCov	31.11	70.71	82.45	2014 IVC [33]
LADF	30.22	78.92	90.44	2013 CVPR [24]
SalMatch	30.16	65.54	79.15	2013 ICCV [46]
Mid-level Filter <sup>a</sup>	29.11	65.95	79.87	2014 CVPR [48]
MtMCML	28.83	75.82	88.51	2014 TIP [34]
RPLM	27.00	69.00	83.00	2012 ECCV [14]
LDFV	26.53	70.88	84.63	2012 ECCVW [32]
SSCDL	25.60	68.10	83.60	2014 CVPR [28]
ColorInv	24.21	57.09	69.65	2013 TPAMI [19]
LF	24.18	67.12	82.00	2013 CVPR [38]
SDALF	19.87	49.37	65.73	2013 CVIU [2]
KISSME	19.60	62.20	77.00	2012 CVPR [18]
PCCA	19.27	64.91	80.28	2012 CVPR [35]
WELF6+PRDC	16.14	50.98	65.95	2012 ECCVW [27]
PRDC	15.66	53.86	70.09	2013 TPAMI [50]
ELF	12.00	44.00	61.00	2008 ECCV [10]

\*请注意[48]报道通过将他们的方法与 LADF 融合,有 43.39%的 1 级别的精度[24]。融合不同的方法通常可以改进性能。事实上,我们也试着将我们的方法与 LADF 融合,得到了 50.32%的 1 级别的识别速率。

### 5.1.3 与最先进结果的比较

最后,我们将所提出的方法的性能与在 VIPeR 数据库中被报道过的最先进的结果进行比较,这在图 6 和表 1 中总结。四种方法是, SCNCD[44], kBiCov[33], LADF[24], 和 SalMatch[46]至今在 VIPeR 数据库报道了最好的性能,在 1 级别中超过 30%。从表 1 中可以看到所提出的算法在 1 级别中得到的 40%的新的最先进状态,比第二好的 SCNCD 表现超过 2.2%。

## 5.2. 在 QMUL GRID 中的实验

QMUL 的隐蔽的重新识别(GRID)数据集[31]是另一个具有挑战性的人的重新识别试验台，然而却还没有很大程度的关注。GRID 数据集是从地铁站中 8 台不相干的摄像机视图中抓取的。它包含 250 组行人图像对，每一对包含从不同摄像机视图中获得的同一个人的两张图像。除此之外，还有 775 张额外的不属于这 250 个人的图像，这可以用来扩充图库集。GRID 中的样本图像可以在图 7 中找到。可以看出这些图像有差的图像质量和低分辨率，并且包含光照和视点的巨大变化。



图 7. 来自 GRID 数据库中的样本图像对[31]。在同一列中的图像代表同一个人。

10 个随机试验的实验设置是为 GRID 提供数据集。对每个试验，125 对图像被用来训练，剩下来的 125 对图像，和 775 张背景图像一样用来测试。在[27]中被描述的 ELF6 特征集是提供用来发展机器学习算法。我们首先将所提出的方法应用在 GRID 所提供的特征集中。这导致的“ELF6+XQDA”的结果在表 2 中列出。我们将来自[30]的可用结果进行比较，其中使用了相同的特征集。表 2 中所展示的结果表明了所提出的联合的降维和度量学习方法优于其它的像 RankSVM[39]，PRDC[49]和 MRank[30]的距离学习算法，除了 XQDA 的 1 级别精确性比 MRank-RankSVM 略差。

我们还试验所提出的特征提取方法，和将相同的 XQDA 算法用于度量学

习。这对应的结果在表 2 的最后一行。这比较显示了新功能改善了 1-10 级别的性能。特别地，一个 4.32%性能提升可以用来 1 级别的精确性。这表明了新功能可用来帮助减少内部类的变化，所以相同的人可以在更高的级别中被识别。

请注意上述的方法训练的都是摄像机中的独立的一般模型。在[34]中的研究显示了可以通过摄像机的网络信息来进行性能提升。换句话说，他们的方法 MtMCML 训练不同的度量，每一个都是给定的摄像机视图对。我们还依靠已知的双摄像头来采用这种方法并训练几个度量。列举在表 3 中的结果显示，同时与 ELF6 功能的所提出的方法仅仅比 MtMCML 提升了 1 级别的精确度，和新的 LOMO 功能所提出的方法明显的优于 MtMCML。然而，在实践中我们不建议这种训练方式，因为基于评测的摄像机视图通常是看不见的，并且它不容易用来对新的摄像机视图进行标记数据，以此来再训练算法。

表 2. 与 GRID 中最先进的结果的比较 ( $p=900$ )，没有摄像机的网络信息。红色和蓝色数值分别是最好和第二好的结果。

Method	rank=1	rank=10	rank=20
ELF6 + L1-norm [30]	4.40	16.24	24.80
ELF6 + RankSVM [39]	10.24	33.28	43.68
ELF6 + PRDC [50]	9.68	32.96	44.32
ELF6 + MRank-RankSVM [30]	12.24	36.32	46.56
ELF6 + MRank-PRDC [30]	11.12	35.76	46.56
ELF6 + XQDA	10.48	38.64	52.56
LOMO + XQDA	16.56	41.84	52.40

Method	rank=1	rank=10	rank=20
ELF6 + MtMCML [34]	14.08	45.84	59.84
ELF6 + XQDA	16.32	40.72	51.76
LOMO + XQDA	18.96	52.56	62.24

表 3. 与 GRID 中最先进的结果的比较 ( $p=900$ )，有摄像机的网络信息。红色和蓝色数值分别是最好和第二好的结果。

### 5. 3. 在香港中文大学校园的实验

香港中文大学数据集是在校园环境中的两台摄像头捕捉的。与上述数据集的不同是，在此数据集中的图像是有更高的分辨率。香港中文大学数据集包含 971 人，在每个摄像头下每个人有两张图像。摄像机 A 捕获行人的正面视图或后视图，而摄像机 B 捕获侧视图。所有的图像缩放至  $160 \times 60$  像素。人被分成 485 份用来训练和 486 份进行测试(多镜头)。结果在图 8 中展示。我们的方法在很大程度上由于现有的最先进的方法。迄今报道的最好的 1 级别的识别速率是 34.30%[48]，而我们已经获得的速率是 63.21%，提升了 28.91%。

#### 5.4. 在香港中文大学 3 号校区的实验

香港中文大学 3 号校区的数据集 [23] 中包含 1,360 名行人的 13,164 张图

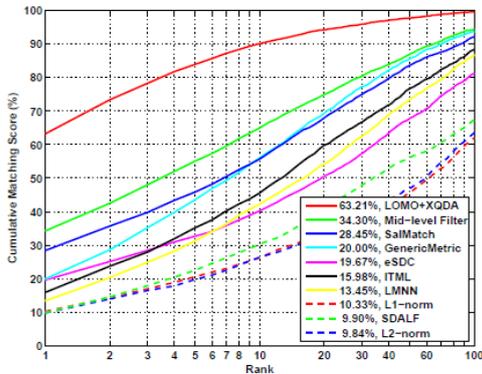


图 8. 在香港中文大学校园数据库中的多镜头 CMC 曲线和 1 级别识别速率 [22]( $P=486, M=2$ )。比较的结果来自[48]。

像。它是目前最大的公开的可用的人的重新识别的数据集。香港中文大学 3 号校区数据集是经过几个月的有六个监控摄像头捕捉而成的，每个人是由两个不相关的摄像头观察的，且在每个视图中平均有 4.8 张图像。此外行人的图像的手动剪裁，进行样本检测的最先进的行人探测器也被提供。这是

一个更现实的考虑到错位，闭塞和身体部分丢失的设置。

我们运行我们的和[23]相同设置的算法。这就是说，数据集被分成 1,160 人的训练集和 100 人的测试集。实验是进行 20 份随机拆分来进行的，并且所有的 CMC 曲线是由单次设定来进行的。在标记和检测的设置下的 1 级别的各种算法的识别速率在表 4 中展示。所提出的方法通过使用标记边界盒和自动地检测边界盒取得了 52.20%和 46.25%的 1 级别的识别速率，分别地，很明显地优于最先进的方法 DeepReID[23]，在标记设置上提升了 31.55%，在检测设置上提升了 26.36%。

表 4. 在香港中文大学 3 号校区数据库 [23] 中通过标记和检测设置进行的与最先进的 1 级别识别速率(%)的比较 ( $P=100$ )。比较结果来自[23]。

	Labeled	Detected
LOMO+XQDA	<b>52.20</b>	<b>46.25</b>
DeepReID [23]	20.65	19.89
KISSME [18]	14.17	11.70
LDML [11]	13.51	10.92
eSDC [47]	8.76	7.68
LMNN [43]	7.29	6.25
ITML [4]	5.53	5.14
SDALF [2]	5.60	4.87

#### 5.5. 所提出的方法的分析

为了更好地理解所提出的方法，我们在许多方面对它进行分析：同态滤波的作用，局部最大度量发生操作的任务，子空间维数的影响，和运行时间。在 VIPeR 数据库中进行分析，通过使用 316 人的随机抽样的训练集，和剩下的人组合而成的测试集。

##### 5.5.1 同态滤波的作用

我们通过将提出的 LOMO 功能经过同态滤波的处理和未处理，所得结果在图 9(a)和(b)中显示。这一比较分别是通过使用直接余弦相似性度量和

XQDA 算法来完成的。从图 9(a)中可以看到,进行直接匹配,性能可以通过应用同态滤波转换明显地得到提升,得到没有使用同态滤波的 1 级别 12.97% 的准确度,和使用同态滤波的 20.25% 的准确度。这个结果显示了通过不同摄像头的同态滤波有助于得出一个一致的衍颜色表示,也可以从图 1(b)中看到。然而,从图 9(b)中可以看到两个作用通过 XQDA 提升到相似的性能。这显示了 XQDA 能够应对光照的变化而学习到鲁棒性的度量。

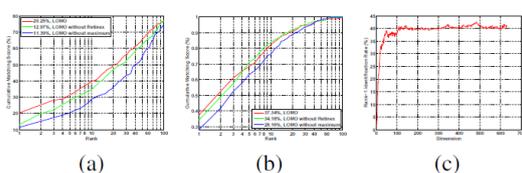


图 9. CMC 曲线在有同态滤波和没有同态滤波和局部最大度量操作的与提出的作用的比较((a)余弦和(b)XQDA)。(c)由于 LOMO 作用的 XQDA 算法, 1 级别在不同的子空间维数下的准确度。

### 5.5.2 局部最大事件的作用

人的重新识别的性能很大程度上受到视点变化的影响,这应该在功能设计或者分类器学习中处理。所提出的局部最大事件作用功能提取是一个战略,其朝向姿势或者视点鲁棒性功能的表示。通过比较具有或者不具有局部最大事件操作的所提出的功能,我们发现这个操作有很大的用来提升交叉视觉人的重新识别的作用,就像在图 9(a)和(b)中显示的。不具有局部最大事件操作的,通过应用余弦相似度量度的 1 级别准确度(图 9(a))是 11.39%,然而采用这个战略的,1 级别的准确度提升到 20.25%。当进一步使用 XQDA(图 9(b)),性能间隙减少了,然而所提出的具有局部最大事件操作的功能仍然比未使用此方法的表现得相当好。

### 5.5.3 子空间维数

对于所提出的 XQDA 算法,学习子空间的维数在性能上有影响。这种影响在图 9(c)中显示,是在 VIPeR 数据集中通过不同的子空间维数应用 XQDA 而得到的。大致地,性能是通过维数的提升而提升的,然而在 100 维度后变得稳定。因此,决定一个自控间的维度的数目是不困难的。我们使用一个自动的方式来接受所有  $\Sigma_I^{-1} \Sigma_E$  的比 1 大的特征值的规定,在所有的实验中工作得很好。然而,还可以考虑到计算的复杂性而选择一个小的值。可以在图 9(c)中观察到,当子空间的维数比 16 的时候,1 级别的准确度始终大于 30%。

表 5. 度量学习算法的训练时间(秒)。

LMNN	ITML	RLDA	ADL	LMN	GMN
1.28	1.34	1.23	30.18	50.58	
XQDA	KISSME	VLDA	LMN	GMN	

### 5.5.4 运行时间

度量学习算法的训练时间的比较情况在表 5 中显示(包括子空间的学习时间)。训练时间是在 VIPeR 数据集中的 10 个随机训练中取平均值。所有的算法在 MATLAB 中实施过。LMNN 算法拥有 MEX 功函数在 C 或者 C++中实施以此来加速计算。此训练在拥有 Intel i5-2400 @3.10GHz 中央处理器的台式机上运行。表 5 展示了 KISSME, RLDA, 和 XQDA 算法,具有封闭式解决方案,是非常有效的,然而 ITML 和 LMNN,需要迭代优化,是耗费时间的。

除此之外,我们还评估所提出的功能提取的运行时间。在处理  $128 \times 48$  人的图像,LOMO 特征提取平均每张图像需要 0.012 秒,是非常有效的。代码也在 MATLAB 上实施,通过使用经过同态滤波后的 MEX 函数。考虑到关于所提出的 LOMO 功能和 XQDA 算法的有效性和高效性,我们推出了为未来研究的代码<sup>2</sup>和人的重新识别的标准。

## 6. 总结和未来的工作

在这份报告中，我们已经提出了用来人的重新识别的一种有效的和高效的方法。我们已经提出了一个叫做 LOMO 的高效描述，已经展示了能够应对视点改变的光照变化。我们还提出一个叫做 XQDA 的子空间和度量学习的方法，可以被配置成广义 Rayleigh 商，和一个可以通过广义特征值分解得到的封闭形式的方法。实际的用来对 XQDA 计算问题已经讨论过，包括简化计算，正规化，和维度选择。在四个具有挑战性的人的重新识别数据库，VIPeR, QMUL GRID, CUHK Campus 和 CUHK03 中进行的实验，显示了所提出的方法将最先进的 1 级别的识别速率分别提升了 2.2%，4.88%，28.91% 和 31.55%。由于 LOMO 功能具有前景的表现，这用来研究其他局部特征将会是有趣的（例如 Gabor，其他颜色描述符等等）或者用来人的重新识别的拥有相同 LOMO 想法的功能编码方法。有趣的是看到 XQDA 应用在其他交叉视图匹配问题上，如合成的人脸的识别。

<sup>2</sup>[http://www.cbsr.ia.ac.cn/users/scliao/projects/lomo\\_xqda/](http://www.cbsr.ia.ac.cn/users/scliao/projects/lomo_xqda/)

## 致谢

这项工作是由中国国家自然科学基金项目 #61203267, #61375037, #61473291, 国家科技支持项目 #2013BAK02B01, 中国社科院科学项目编号 KGZD-EW-102-2, 和正宗研发资金支持。

## 参考

[1] B. Alipanahi, M. Biggs, A. Ghodsi, et al. Distance metric learning vs. fisher

discriminant analysis. In *International conference on Artificial intelligence*, 2008. 2

- [2] L. Bazzani, M. Cristani, and V. Murino. Symmetry-driven accumulation of local features for human characterization and re-identification. *Computer Vision and Image Understanding*, 117(2):130–144, 2013. 1, 6, 7
- [3] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino. Custom pictorial structures for re-identification. In *BMVC*, volume 2, page 6, 2011. 2
- [4] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In *Proceedings of the 24th international conference on Machine learning*, pages 209–216. ACM, 2007. 1, 2, 5, 7
- [5] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja. Pedestrian recognition with a learned metric. In *Computer Vision—ACCV 2010*, pages 501–512. Springer, 2011. 1, 2
- [6] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *CVPR*, pages 2360–2367, 2010. 2
- [7] N. Gheissari, T. B. Sebastian, and R. Hartley. Person reidentification using spatiotemporal appearance. In *CVPR (2)*, pages 1528–1535, 2006. 2
- [8] S. Gong, M. Cristani, S. Yan, and C. C. Loy. *Person Re-Identification*. Springer, 2014. 1
- [9] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *IEEE International workshop on performance evaluation of tracking and surveillance*, 2007. 2, 5,

- 6
- [10] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *European Conference on Computer Vision*, 2008. 1, 2, 4, 5, 6
- [11] M. Guillaumin, J. Verbeek, and C. Schmid. Is that you? Metric learning approaches for face identification. In *International Conference on Computer Vision*, 2009. 2, 7
- [12] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In *ICDSC*, pages 1–6, 2008. 2
- [13] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer, 2009. 2
- [14] M. Hirzer, P. M. Roth, M. Kostinger, and H. Bischof. Relaxed pairwise learned metric for person re-identification. In *European Conference on Computer Vision*. 2012. 1, 2, 6
- [15] Y. Hu, S. Liao, Z. Lei, D. Yi, and S. Z. Li. Exploring structural information and fusing multiple features for person reidentification. 2
- [16] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *Image Processing, IEEE Transactions on*, 6(7):965–976, 1997. 2
- [17] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell. Properties and performance of a center/surround retinex. *Image Processing, IEEE Transactions on*, 6(3):451–462, 1997. 2
- [18] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012. 1, 2, 3, 4, 5, 6, 7
- [19] I. Kviatkovsky, A. Adam, and E. Rivlin. Color invariants for person reidentification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(7):1622–1634, 2013. 6
- [20] E. H. Land and J. McCann. Lightness and retinex theory. *JOSA*, 61(1):1–11, 1971. 2
- [21] W. Li and X. Wang. Locally aligned feature transforms across views. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013. 2
- [22] W. Li, R. Zhao, and X. Wang. Human reidentification with transferred metric learning. In *Asian Conference on Computer Vision*, 2012. 7
- [23] W. Li, R. Zhao, T. Xiao, and X. Wang. DeepReID: Deep filter pairing neural network for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2014. 7
- [24] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith. Learning locally-adaptive decision functions for person verification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013. 1, 2, 5, 6
- [25] S. Liao, D. Yi, Z. Lei, R. Qin, and S. Z. Li. Heterogeneous face recognition from local structures of normalized appearance. In *International Conference on Biometrics*, 2009. 4
- [26] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikainen, and S. Z. Li. Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes. In *Proceedings of*

- IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, June 2010. 3
- [27] C. Liu, S. Gong, C. C. Loy, and X. Lin. Person reidentification: what features are important? In *Computer Vision—ECCV 2012. Workshops and Demonstrations*, pages 391–401. Springer, 2012. 3, 6
- [28] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, and J. Bu. Semi-supervised coupled dictionary learning for person reidentification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014. 6
- [29] Y. Liu, Y. Shao, and F. Sun. Person re-identification based on visual saliency. In *Intelligent Systems Design and Applications (ISDA), 2012 12th International Conference on*, pages 884–889. IEEE, 2012. 2
- [30] C. C. Loy, C. Liu, and S. Gong. Person re-identification by manifold ranking. In *IEEE International Conference on Image Processing*, volume 20, 2013. 6
- [31] C. C. Loy, T. Xiang, and S. Gong. Multi-camera activity correlation analysis. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1988–1995. IEEE, 2009. 6
- [32] B. Ma, Y. Su, and F. Jurie. Local descriptors encoded by fisher vectors for person re-identification. In *European Conference on Computer Vision Workshops*, 2012. 1, 2, 6
- [33] B. Ma, Y. Su, and F. Jurie. Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image and Vision Computing*, 32(6):379–390, 2014. 1, 5, 6
- [34] L. Ma, X. Yang, and D. Tao. Person re-identification over camera networks using multi-task distance metric learning. *IEEE Transactions on Image Processing*, 2014. 6, 7
- [35] A. Mignon and F. Jurie. Pcca: A new approach for distance learning from sparse pairwise constraints. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012. 6
- [36] B. Moghaddam, T. Jebara, and A. Pentland. Bayesian face recognition. *Pattern Recognition*, 33(11):1771–1782, 2000. 2, 3, 4
- [37] T. Ojala, M. Pietikäinen, and D. Harwood. “A comparative study of texture measures with classification based on feature distributions”. *Pattern Recognition*, 29(1):51–59, January 1996. 3
- [38] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian. Local fisher discriminant analysis for pedestrian re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013. 1, 2, 6
- [39] B. Prosser, W.-S. Zheng, S. Gong, T. Xiang, and Q. Mary. Person re-identification by support vector ranking. In *BMVC*, 2010. 2, 3, 5, 6
- [40] R. Vezzani, D. Baltieri, and R. Cucchiara. People reidentification in surveillance and forensics: A survey. *ACM Computing Surveys (CSUR)*, 46(2):29, 2013. 1
- [41] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. H. Tu. Shape and appearance context modeling. In *ICCV*, pages 1–8, 2007. 2
- [42] X. Wang, M. Yang, S. Zhu, and Y. Lin. Regionlets for generic object detection. In *IEEE International Conference on Computer Vision*, 2013. 2
- [43] K. Weinberger, J. Blitzer, and L. Saul.

- Distance metric learning for large margin nearest neighbor classification. *Advances in neural information processing systems*, 2006. 1, 2, 5, 7
- [44] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li. Salient color names for person re-identification. In *Proceedings of the European Conference on Computer Vision*, 2014. 5, 6
- [45] J. Ye, T. Xiong, Q. Li, R. Janardan, J. Bi, V. Cherkassky, and C. Kambhampettu. “Efficient model selection for regularized linear discriminant analysis”. In *Proceedings of the ACM Conference on Information and Knowledge Management*, pages 532–539, 2006. 5
- [46] R. Zhao, W. Ouyang, and X. Wang. Person re-identification by salience matching. In *International Conference on Computer Vision*, 2013. 1, 2, 5, 6
- [47] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person re-identification. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3586–3593. IEEE, 2013. 2, 7
- [48] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014. 1, 6, 7
- [49] W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by probabilistic relative distance comparison. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 649–656. IEEE, 2011. 1, 2, 3, 5, 6
- [50] W.-S. Zheng, S. Gong, and T. Xiang. Reidentification by relative distance comparison. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(3):653–668, 2013. 5, 6