中国科学院大学

University of Chinese Academy of Sciences

博士学位论文

跨摄像头行人重识别问题研究与应用

作者姓名：曹敏

指导教师：彭思龙 研究员 中国科学院自动化研究所

学位类别：工学博士

学科专业：模式识别与智能系统

培养单位：中国科学院自动化研究所

2019 年 12 月

Research and Application on Person Re-Identification across Camera Views

A dissertation submitted to the

University of Chinese Academy of Sciences

in partial fulfillment of the requirement

for the degree of

Doctor of Engineering

in Pattern Recognition and Intelligent System

By

Min Cao

Supervisor: Professor Silong Peng

Institute of Automation, Chinese Academy of Sciences

December, 2019

中国科学院大学
学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。本人完全意识到本声明的法律结果由本人承担。

作者签名：

日　　期：2019.12.19

中国科学院大学
学位论文授权使用声明

本人完全了解并同意遵守中国科学院大学有关保存和使用学位论文的规定，即中国科学院大学有权保留送交学位论文的副本，允许该论文被查阅，可以按照学术研究公开原则和保护知识产权的原则公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

涉密及延迟公开的学位论文在解密或延迟期后适用本声明。

作者签名：　　导师签名：

日　　期：2019.12.19　　日　　期：2019.12.20

摘 要

行人重识别旨在对同一行人由不同摄像机拍摄得到的行人图像进行匹配。对于实现视频监控系统智能化，行人重识别的自动化是其中必不可少的一个任务。伴随着视频监控系统的不断完善，行人重识别问题在计算机视觉和机器学习领域引起了研究人员的广泛关注。由于不同的摄像机拍摄到的行人存在姿态、光照、视角和背景等变化，造成了行人在不同摄像机下可能呈现不同的外貌特征，此外，同一摄像机拍摄到的不同行人可能具有相似的外貌特征，这些都导致行人重识别任务具有一定的挑战性。

为了解决这些问题，实现一个具有高性能表现的行人重识别系统，研究人员作出了不懈努力。通常，特征提取和度量学习是行人重识别系统的关键步骤，研究人员针对其中一个步骤或两个步骤进行研究，近几年，为了进一步提高行人重识别的性能，重排序成为了行人重识别系统的一个重要步骤，获得了研究人员的关注。本文针对行人重识别系统中的度量学习和重排序步骤进行了详细研究，其主要工作和创新点归纳如下：

(1) 对于实现行人重识别任务，一个有效的方法是通过同时最小化样本类内散度矩阵和最大化样本类间散度矩阵学习一个区分性良好的度量函数。然而，样本的特征向量的维数通常大于训练样本数量，从而导致样本类内散度矩阵是奇异矩阵，无法学习一个优良的度量函数。本文提出通过探索类内散度矩阵的伪逆，针对度量函数学习一个正交转化矩阵，来解决这个奇异问题。本文提出的方法具有两个优点：通过模型求解可以得到一个闭式解，且没有超参数需要调试。此外，针对行人重识别问题的非线性特征，本文发展了方法的核化版本；为了更高效的模型求解，本文发展了方法的快速版本。在实验中，本文作者验证了提出的方法对于解决奇异问题的有效性和优越性，分析了方法核化版本和快速版本的性能表现。在四个行人重识别数据集上的广泛对比实验，显示了提出的方法在性能方面的先进性。

(2) 对于行人重识别系统中的度量学习步骤，其输入是行人图像的特征向量，研究人员基于行人重识别的目标-正确匹配的行人样本对在排序列表中排在所有样本对的第一位，构造一个目标损失函数，通过优化求解得到区分性良好的度量

函数。行人图像的特征向量往往是基于图像的不同区域分块提取，伴随着这样的特征向量作为输入，大多数度量学习方法致力于学习一个区域通用的转换矩阵，也就是不同的区域特征共享一个同质转换，忽视了行人图像的空间结构和不同区域特征的分布差异性。因此本文提出了一个新颖的区域特定的度量学习方法，致力于学习区域特定的转换矩阵，分别对隶属于不同区域的子模型进行优化。而对于目标损失函数的构造，本文对行人重识别的目标进行数学语言的直接翻译，提出通过最小化正确匹配的行人样本对的距离与所有样本对中最小样本对的距离之间的差异，学习最优的特征映射函数。相比于其他目标损失函数，针对样本集的部分样本建模，本文提出的目标损失函数是通过最直接和直观的方式，对行人重识别中的排序目标进行建模，并且作用在了整个候选集样本上。通过在行人重识别数据集上的广泛实验，表明了本文提出的方法的有效性和先进性。

(3) 为了进一步提高行人重识别的性能，本文基于个人信息得到的排序结果，有效地利用样本的上下文信息进行重识别。从时空域来看，群体成员可以提供视觉线索帮助行人重识别。本文探讨了基于群体的行人重识别方法的本质，提出了一个广义的群体定义。同时，本文提出了一个对匹配机制用于测量广义群体间的距离，对于跨摄像头群体成员位置发生变化具有鲁棒性。基于个人信息得到的样本对的匹配值加权其广义群体间的匹配值，得到最终的样本对的匹配值，完成行人重识别。从欧式空间看，通过探索样本在欧式空间中的近邻样本信息，样本对的匹配值基于数据流形结构的测地线路径得到了更精确的估计。因此本文提出样本对的相似性受它们的近邻样本信息影响，提出的方法是基于假设-如果样本对中一个样本的近邻样本信息与另一个样本相似，则样本对中的样本相似于彼此。通过在不同的行人重识别数据集上的实验，验证了提出的方法可以实现性能增强，并且相比于同类型的基于重排序的方法在性能和效率上都具有优越性。

关键词：行人重识别，度量学习，重排序，奇异问题，区域特定的度量函数，排序损失函数，群体信息，近邻样本

Abstract

Person re-identification (re-id) aims to identify designated individuals from a large amount of pedestrian images across non-overlapping camera views and and is a critical task for the realization of an intelligent video monitoring system. With the development of video monitoring system, person re-id is favored by the academe in the field of computer vision and machine learning in recent years. However due to the large intra-class variations caused by the change in illumination, person pose and occlusion across views, person re-id is a challenging problem. In addition, the similarity in appearance among different people further increase its difficulty in real applications.

To address these challenges and achieve a high performance for the person re-id system, researchers have made continuous efforts. Most of existing methods focus on feature extraction and distance metric learning, which are the critical steps in person re-id system. In recent years, in order to further improve the performance of person re-id, re-ranking as an important step has recently gained attention. In this paper, distance metric learning and re-ranking are studied in details. The main works and innovations of the dissertation are as follows:

(1) For achieving person re-id, an effective way is to learn a discriminative metric by minimizing the within-class variance and maximizing the between-class variance simultaneously. However, the dimension of sample feature vector is usually greater than the number of training samples, as a result, the within-class scatter matrix is singular and the metric cannot be learned. In this paper, we propose to solve the singularity problem by employing the pseudo-inverse of the within-class scatter matrix and learning an orthogonal transformation for the metric. The proposed method can be effectively solved with a closed-form solution and no parameters required to tune. In addition, we develop a kernel version against non-linearity in person re-id, and a fast version for more efficient solution. In experiments, we prove the validity and advantage of the proposed method for solving the singularity problem in person re-id, and analyze the effectiveness of both kernel version and fast version. Extensively comparative experiments on

four person re-id benchmark datasets, show the state-of-the-art results of the proposed method.

(2) For distance metric learning, its input data is the feature vectors extracted on several regions of person image, then an objective function is constructed and solved by iterative optimization algorithms to satisfy the constraints on samples of the training set, and obtain a discriminative distance metric. With the features extracted on several regions of person image, most of distance metric learning methods have been developed in which the learnt cross-view transformations are region-generic, i.e all region-features share a homogeneous transformation. The spatial structure of person image is ignored and the distribution difference among different region-features is neglected. Therefore in this paper, we propose a novel region-specific metric learning method in which a series of region-specific sub-models are optimized for learning cross-view region-specific transformations. For the construction of the objective function, with the direct translation in math language for the goal of person re-id, we propose a novel metric learning method for person re-id to learn such an optimal feature mapping function, which minimizes the difference between the distance of matched pair and the minimum distance of all pairs, namely *Ranking Loss*. Compared with other loss functions, the proposed ranking loss optimizes the ultimate ranking goal in the most direct and intuitional way, and it directly acts on the whole gallery set efficiently instead of comparatively measuring in small subset. Extensive experiments on the datasets show the effectiveness of the proposed method compared to state-of-the-art methods.

(3) For enhance the performance of person re-id, in this paper we propose to utilize the contextual information of the sample for person re-id based on the initial ranking results obtained by the feature extraction and metric learning. For the view of the spatial-temporal domain, group members can provide visual clues for person re-id. For this, in this paper we discuss the essentials of group-based person re-id and relax the group definition towards a concept of "co-traveler set". Accordingly we propose a pair matching scheme to measure the distance between co-traveler sets, which tackles the problems caused by dynamic change of group across camera views. The final individual matching score is weighted by the obtained distance measurements between co-traveler

sets. For the view of the Euclidean space, the pairwise similarity can be computed more accurately by taking into account the contextual information of sample and the structure of the dataset manifold. For this, we propose a context-driven person re-id method in which the pairwise measure is determined by their contextual information provided through its neighbor samples. The main motivation of the proposed method relies on the conjecture that two sample are similar to each other if the contexts of them are similar to each other. Experiments were conducted on different person re-id datasets and shows the promising improvement of the proposed method compared with state-of-the-art re-ranking methods.

Keywords: Person Re-Identification, Metric learning, Re-Ranking, Singularity Problem, Region-Specific Metric Learning, Ranking Loss Function, Group Information, Neighborhood Sample

目 录

图形列表

表格列表

第 1 章 引言

1.1 研究背景

近年来，视频监控系统在城市与城镇中不断发展和完善。借助于这些视频监控系统，城市与城镇的安保工作得到了极大的帮助。然而，随着监控设备的不断增加，视频资源呈现了爆炸式的增长，利用这些海量视频辅助安保工作，对人力资本的需求非常高。比如，民警想通过监控视频网络寻找犯罪嫌疑人的逃跑轨迹，需要他（她）从海量视频当中查看可疑人员的行踪，这是一个非常费时费力的工作。因此，视频监控系统智能化迫在眉睫。通过对视频监控系统智能化，安保的工作效率可以得到大幅度的提高，继而可以最大化视频监控系统在安防中的作用。针对视频监控系统智能化，跨摄像头行人重识别是其中最重要的任务之一。

跨摄像头行人重识别（person re-identification）旨在对不存在重叠监控区域的不同摄像机中观察到的行人进行自动匹配，继而实现视频监控系统中行人追踪自动化。这是一个非常具有挑战性的任务。在公共区域里，监控摄像机在一天时间里可能会记录成千上万的行人。由于一部分行人穿衣风格相似，导致常常会出现不同的行人在监控视频中看起来相似的情况。另外，不同的摄像机存在着光照、拍摄姿态、分辨率和设置等不一样的问题，导致了同一个行人可能会在不同的摄像机下呈现不同的外貌特征。此外，在拥挤的公共场合，人与人之间往往存在着遮挡。这些都给行人重识别增加了困难。尽管存在诸多挑战，行人重识别仍然是一个具有研究价值的热门课题。

研究人员针对行人重识别问题的研究，最早可追溯到 1999 年，CAI 等人 [3] 首次提出解决跨摄像头行人追踪问题。尽管在当时并没有明确提出行人重识别这个概念，但是其所解决的问题与今天研究人员所定义的行人重识别本质上是一样的。由于在当时，城市与城镇中的视频监控系统并不发达，视频资源有限，人们对实现行人重识别自动化的需求不大，因此该问题没有引起研究人员的重视。从 2008 年开始，行人重识别问题受到了学术界越来越多的关注。伦敦玛丽女王大学（QMUL）Shaogang Gong 团队是最早研究该问题的团队之一 [4]。自 2014 年以来，借助于深度学习的发展，行人重识别技术的训练库趋于大规模化，

并广泛采用深度学习框架。随着高校、研究所以及一些厂商的研究持续深入，行人重识别技术得到了飞速发展。

具体地，目前研究人员对该问题的定义是：对于一个来自于摄像头 a 的行人图像（称之为询问图像，query image），和一个来自于与摄像头 a 没有重叠监控区域的摄像头 b 的行人图像库（称之为候选图像库，candidate images），研究人员需要针对这个询问图像，对候选图像库中的每个候选图像按照其与询问图像的相似度进行排序。如图1.1所示，对于从 Camera a 捕捉到的每个询问行人图像，研究人员需要对从 Camera b 捕捉到的所有候选行人图像按照其与这个询问行人图像的相似度进行排序。在此排序中，正样本候选图像（即，与询问行人图像为同一个人）位置越靠前，表明行人重识别方法效果越好。据此，行人重识别方法可以概括为以下三步：1）特征学习，2）度量学习，3）重排序。研究人员致力于其中某一个步骤或者几个步骤进行研究。本文作者针对度量学习步骤和重排序步骤，对行人重识别问题积极进行了相关理论与实际应用的研究。

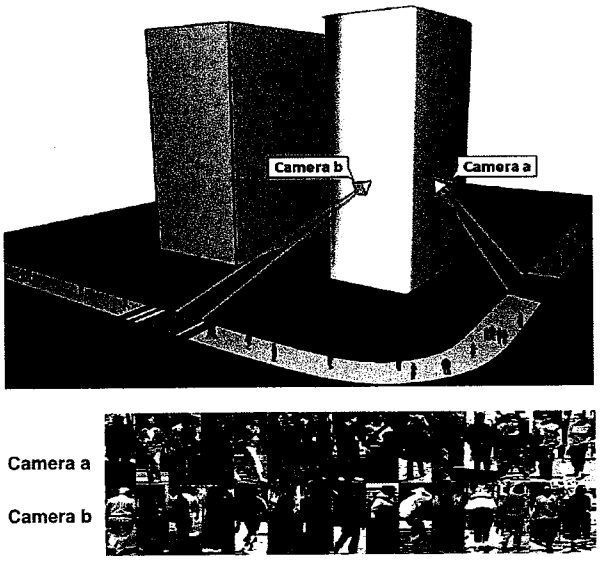


图 1.1 行人重识别问题的定义。

Figure 1.1 The definition of person re-identification.

1.2 研究内容及创新点

度量学习和重排序是行人重识别方法的两个关键步骤，本文的研究内容主要围绕这两个关键步骤展开。度量学习步骤致力于学习一个鲁棒的特征空间，使得在这个新的特征空间中，正样本对的相似值大于负样本对的相似值。为此本

文作者需要构造一个合理的目标函数，并利用训练样本求解这个新的特征空间。依照求解方法的不同，度量学习可以分为基于闭式解的度量学习和基于迭代学习的度量学习。本文作者针对这两种度量学习方法，分别提出了不同的行人重识别模型。重排序步骤致力于改进和优化通过特征提取和度量学习步骤得到的初始排序结果，它可以被看作是行人重识别方法的后处理，这一步骤可以实现重识别率的进一步提高。本文作者针对重排序步骤积极进行了研究工作，充分挖掘数据样本的上下文信息，并将该信息应用于重识别过程中，从而帮助重识别率的提高。具体地，本文的研究工作主要概括为以下三个方面：

（1）基于正交线性判别分析的行人重识别方法

在度量学习中，线性判别分析（Linear Discriminant Analysis，LDA）是最经典和受欢迎的方法。它具有无参数和闭式解的优势，要求训练样本的类内散度矩阵非奇异。然而在行人重识别问题中，样本数常常远远小于特征维数，导致训练样本的类内散度矩阵奇异，即，行人重识别是一个奇异问题。为此，本文作者提出基于正交线性判别分析的行人重识别方法。经典的 LDA 的目标函数是最小化类内散度矩阵，同时最大化类间散度矩阵：$L^* = \arg\max trace\{(L^T S_w L)^{-1} L^T S_b L\}$。而本文作者提出的基于正交线性判别分析的行人重识别方法的优化函数是：$L^* = \arg\max trace\{(L^T S_w L)^{+} L^T S_b L\}$，$S^+$ 表示 S 矩阵的伪逆运算。由于目标函数中的求逆改进成了求伪逆，本文作者解决了行人重识别的奇异问题。对于新提出的目标函数的求解，常规的方法是采用广义奇异值分解，该方法的计算量太大。为此，本文作者采用基于三个散度距离的同时对角化求解目标函数。

（2）基于区域特定和排序损失函数的行人重识别方法

为了更好的描述行人的空间信息，行人图像经常被划分为多个区域，用于提取行人特征，即：行人的特征向量是区域特征连接的。将这些考虑了空间结构的特征描述作为度量学习方法的输入，现存的大部分度量学习方法将这些特征作为整体对待，忽视了其不同区域之间的特征差异性。为此，本文作者提出了一种新的区域特定的度量学习方法，并成功地应用在了行人重识别任务中。此外，现有的基于度量学习的行人重识别方法，其度量函数主要可以分为三类：二分类度量，三元组度量和四元组度量。这些度量最终都可以实现基于度量学习的行人重识别的目标，即，正样本对的相似值比负样本对的相似值大。然而这三类用于行人重识别任务的度量函数都是通过一种相对间接的方法建模，是针对全样本集

中的部分集合建模，并且这些方法都需要合理的参数的设置。这些问题会导致模型的鲁棒性差，可操作性不强。为了解决上述问题，本文作者提出了一种鲁棒性强，高效的行人重识别方法。考虑到行人重识别的目标：正样本对的相似值比负样本对的相似值大，也就是正样本对排序在所有样本对中的第一名，本文作者对其目标直接翻译，构建一个新的排序损失函数用于行人重识别：最小化正样本对和最小样本对（即：所有样本对中对距离最小的样本对）之间的距离。相比于其它基于度量学习的行人重识别方法，该方法能更好地解决行人重识别问题，得到更高的识别率。

（3）基于上下文信息的行人重识别方法

目前大部分行人重识别工作的研究都是致力于个人信息的挖掘，凭借个人的外貌特征实现重识别。然而，由于在不同的摄像头拍摄到的行人图像，其光照、行人姿态、视角和背景等都会发生变化，从而导致行人的外貌可能会发生比较严重的变化。因此，如果仅仅依靠基于个人的外貌特征实现重识别，其性能是有限的。为了提高行人重识别的性能，本文作者提出利用样本的上下文信息辅助重识别。具体地，本文作者分别从时空域和欧氏空间提取了不同的上下文信息。相应地，其从时空域提取的上下文信息为行人的群体信息，从欧式空间提取的上下文信息为行人的 k 近邻样本信息。本文作者成功地将这些上下文信息引入行人重识别过程中，与行人的个人信息相结合，作为重识别的辅助信息，大大地提高了行人重识别率。

1.3 论文组织结构

行人重识别是计算机视觉领域中一个既具有研究价值又极具挑战性的热门课题，特征提取、度量学习、重排序是行人重识别方法最关键的三个步骤。本文针对行人重识别中的度量学习和重排序这两个步骤展开研究，全文共分为六章，具体安排如下：

第一章是绪论。首先介绍了行人重识别问题的研究背景及意义，分析了行人重识别的研究价值，并指出了其核心难点。然后阐述了本文的研究内容及创新点，最后总结了本文的组织结构。

第二章是研究综述。对行人重识别的研究现状进行了梳理，分析了代表性工作存在的问题以及本文作者提出的解决思路。

第三章是基于正交线性判别分析的行人重识别方法。首先分析了行人重识别中的奇异问题，进而引出对于该方法的研究动机。接着具体介绍了基于正交线性判别分析的行人重识别方法，包含模型的介绍、模型的求解、模型的核化版本的介绍及求解、模型的快速算法的介绍。然后呈现了详细的实验结果，包括与同类型方法的对比实验结果、模型的核化版本的实验结果、与最先进的行人重识别方法的对比实验结果、模型的快速算法的实验结果。最后对本章进行了总结和进一步展望。

第四章是基于区域特定和排序损失函数的行人重识别方法。首先分析了基于度量学习的行人重识别方法现存的问题，引出对于本章方法的研究动机。接着具体介绍了基于区域特定的行人重识别方法和基于排序损失函数的行人重识别方法，分别详述了其建模过程和求解过程。然后呈现了详细的实验结果，并进行了与相关方法的对比实验的分析。最后对本章进行了总结，更进一步探讨了方法的优缺点，对未来的研究工作进行了展望。

第五章是基于上下文信息的行人重识别方法。首先分析了仅仅致力于个人信息挖掘的行人重识别方法的问题，引出基于上下文信息的行人重识别方法的研究动机。接着具体阐述了基于广义群体信息的行人重识别方法和双边上下文驱动的渐进行人重识别方法。然后对其方法的实验结果进行了呈现和详细分析。最后对本章进行了总结和进一步的展望。

第六章是总结与展望。对全文工作进行了总结，并给出了下一步的工作设想。

第 2 章 行人重识别的研究现状

目前，行人重识别问题的研究都是基于假设：行人被不同摄像头捕捉的时间跨度较短，也就是行人的衣服和体态这些客观的表征在不同的摄像头下不存在变化。这些较稳定的表征可以作为目标的特征表示。而对于相隔数天的行人重识别问题，由于行人的服饰或者体态可能发生改变，使得问题变得更加复杂，无法通过外表特征进行重识别。尽管研究人员可以通过提取更鲁棒和更稳定的生物特征，比如人脸、步态、指纹等，实现相隔数天的行人重识别目的，然而由于目前的监控视频分辨率有限，导致这些生物特征难以提取。

在实际情况下，行人重识别任务的输入往往是监控视频片段。而针对一个给定的询问行人样本（query sample），人们所期望的输出是，与捕捉询问样本不同的摄像头所拍摄到的行人的排序列表，且与询问行人样本标签一致的正样本排列在第一位。因此一个行人重识别系统的一般流程是，对于输入的视频片段，首先进行行人检测，然后对于检测到的行人区域，进行特征提取和度量学习，得到行人图像之间的相似性匹配值，进而得到针对一个给定的询问行人样本的行人排序列表，最后对这个初始排序列表进行优化，得到最终的排序列表，作为行人重识别系统的输出。然而由于行人检测本身就是一个研究课题且行人检测的误差会严重影响行人重识别系统后面的步骤，因此对于行人重识别问题的研究，往往假定其输入是已经通过一个鲁棒的行人检测器 [5] 或者人工手动裁剪得到的行人图像。

综上所述，目前对于行人重识别问题的研究，其主要有三个步骤：特征提取，度量学习和重排序。研究人员致力于其中一个步骤或几个步骤进行研究。对于行人重识别方法的详细综述，感兴趣的读者可以参阅 [6, 7]。接下来本文作者将这三个行人重识别步骤的研究现状分别进行分析和总结，由于本文作者主要针对度量学习和重排序这两个步骤展开研究，因此对度量学习和重排序这两个步骤进行重点分析和总结。此外，近年来，随着深度学习技术的不断发展，越来越多的研究人员致力于利用深度学习技术解决行人重识别问题，因此本文作者也对基于深度学习的行人重识别方法进行一个总结和介绍。另外，近几年，研究人员也开始将目光转向解决针对具体实际情况的行人重识别问题，如可扩展的

行人重识别（scalable person re-identification）等，本文作者也对这类行人重识别方法进行一个介绍。

2.1 基于特征提取的行人重识别方法

基于特征提取的行人重识别方法主要致力于学习一个有效的鲁棒的特征表示，使其对于不同摄像头下呈现的行人外貌特征的变化具有鲁棒性。早期的研究致力于底层特征的挖掘，比如颜色和纹理特征，其相关的具有代表性的工作有[2, 8, 9]。这些方法是基于非监督学习的方法，因此其性能有限，常常需要搭配度量学习步骤，将提取到的特征通过监督学习转化到一个具有良好判别能力的特征空间。近几年，随着深度学习技术的不断发展，研究人员转向基于深度学习的特征提取。针对不同摄像头下拍摄到的行人图像存在的对不准、视角变化、尺度变化和背景干扰等问题，研究人员借助于深度学习技术提出了不同的解决方案 [10–13]，以获得有效的鲁棒的特征表示。为了解决行人图像中背景干扰的问题，研究人员首先采用二值分割掩模（binary segmentation masks）区分出行人图像的前景区域，然后分别学习前景区域和背景区域的特征 [14–17]。为了解决行人图像之间空间对不准的问题，研究人员往往将行人图像分块，然后针对每一块图像区域进行特征提取，最后将这些区域特征整合在一起，形成最终的特征表示[18, 19]。为了更进一步提高性能，（1）近几年研究人员广泛采用行人的语义信息定位行人的图像块，然后进行图像块区域的特征提取 [20, 21]；（2）一些方法基于来自于行人图像块区域或者行人姿态的带约束的注意力选择机制（constrained attention selection mechanisms）来隐式对齐对不准的图像对 [15, 22]；（3）还有一些方法同时探索全局特征和区域特征的优势，结合全局特征和区域特征作为最终的特征表示 [23–25]。全局特征主要捕捉了全图像中最有判别力的外表信息，然而可能忽视了最有判别力的局部信息的细节，而区域特征可以作为其信息的补充。在 [26] 中，一个多粒度的网络被提出，其中包含一个用于提取全局特征的网络分支和两个用于提取局部特征的网络分支；在 [27] 中，作者将图像的特征图严格地划分成多个局部条纹，然后提出一个最短路径损失函数用于对齐这些局部条纹，提取区域特征，最后区域特征与全局特征加权，得到最终的特征表达。

2.2 基于度量学习的行人重识别方法

对于基于度量学习的行人重识别方法，研究人员通常借助数据集的训练样本，学习有效的度量函数，使其正样本对的相似值大于负样本对的相似值。这些方法大致可以分为两类：基于闭式解的和基于迭代学习的。对于基于闭式解的方法，其通常与线性判别分析（Linear Discriminant Analysis，LDA）技术有关，通过同时最小化样本的类内方差和最大化样本的类间方差，构造其目标函数，然后通过方程求解，得到最优解。在基于迭代学习的方法中，研究人员基于目标：正样本对相似值大于负样本相似值，构造其目标函数，然后通过采用迭代优化算法，比如梯度下降法，获得最优的度量函数。

由于大量关于 LDA 的理论研究，以及 LDA 的无参调节和闭式解的优势，一些基于度量学习的行人重识别方法借助于 LDA 技术学习最优的度量函数。对于典型的 LDA，要求类内散度矩阵必须满足非奇异性，然而对于行人重识别问题，样本的特征向量维数通常大于样本数量，导致所有的散度矩阵都是奇异的，这就导致 LDA 技术不能直接应用在行人重识别问题中。为此，一些研究人员致力解决于行人重识别的奇异问题。Pedagadi 等人 [28] 采用两步法学习最优度量函数：主成分分析（Principal Component Analysis，PCA）和正则化局部 Fisher 判别分析（Regularized Local Fisher Discriminant Analysis，RLFDA）。Liao 等人 [9] 引入正则化线性判别分析（Regularized LDA，RLDA），解决行人重识别的奇异问题，正则化线性判别分析通过给类内散度矩阵加上一个偏置，达到消除类内散度矩阵的奇异性，因此如何选择一个合适数量级的偏置是需要解决的问题。Zhang 等人 [29] 提出基于零空间线性判别分析（Null Space LDA，NLDA）的行人重识别方法，零空间线性判别分析在类内散度矩阵的零空间寻找最优解，然而对于低维数据，类内散度矩阵的零空间可能为空，因此该方法可能无法应用在低维数据中。本文作者提出基于正交线性判别分析的行人重识别方法，有效的解决了行人重识别的奇异问题，并且该方法既可以应用在低维数据中，也可以应用在高维数据中。

为了获得鲁棒的特征表示，研究人员往往针对行人图像的不同区域分别提取特征。因此在行人重识别系统中，度量学习这一步骤的输入往往是区域连接的特征。考虑到行人图像的结构以及在不同区域上提取到的特征的分布差异性，一些基于度量学习的行人重识别方法提出基于每个图像区域学习不同的度量函数。

Chen 等人 [30] 认为行人图像的每个区域都应该有其特定的相似度量，该相似度量应擅长处理在其所属区域中的类内差异，鉴于此，他们提出了一个包含有多个子相似度量的相似函数，并采用交替方向乘子法（Alternating Direction Method of Multipliers，ADMM）对模型进行求解。Chong 等人 [31] 分别针对行人图像的水平对不准、垂直对不准和腿部姿态建立模型，将一个全局度量模型和多个弹性度量模型综合起来以完成行人特征的表达。这些工作都是建立了一个比较复杂的模型，求解模型需要较多运行时间且占用较大内存。本文作者通过约束度量函数中的映射矩阵为块对角结构，提出基于区域特定的行人重识别方法，该方法简单且有效。

对于基于迭代学习的方法，其目标函数大致可以归为三类：基于二值分类损失、基于三元组损失和基于四元组损失。对于基于二值分类损失目标函数的方法：在 PCCA 方法中 [32]，优化目标函数，使得正样本对的距离值低于一个给定的阈值同时负样本对的距离值高于这个给定的阈值，该方法需要研究人员手动设置一个阈值；进一步，一个基于正则化相似度量的行人重识别方法 [33] 被提出，该方法采用自适应的阈值。对于基于三元组损失目标函数的方法：Ding 等人 [34] 构建了一个深度神经网络，其通过最大化正样本对距离和负样本对距离之间的距离，求解最优的度量函数；Zhou 等人 [35] 提出了一个 set to set（S2S）损失层，致力于最大化类内集合与类间集合的间距，并将其应用在了深度学习框架中。对于基于四元组损失目标函数的方法，Chen 等人 [36] 提出了一个四元组深度网络，解决行人重识别问题，相比于三元组损失，四元组损失的约束更强，以求实现更小的类内方差和更大的类间方差。此外，一些研究人员结合二值分类损失和三元组损失，并联合优化，从而获得更好的重识别率 [37, 38]。一些研究人员认为行人重识别任务是介于图像分类和目标检索之间的任务 [7]，因此提出同时采用分类损失和排序损失，如三元组损失，优化网络 [26, 39, 40]。

2.3 基于重排序的行人重识别方法

目前，大多数的行人重识别工作集中在特征提取和度量学习方面，这些工作通过充分挖掘行人自身的信息建立模型，从而获得排序结果。基于该初始排序结果，增加重排序步骤，可以有效地提高重识别率。这些方法大致可以分为两类：（1）通过在线的行人反馈，迭代更新重识别模型；（2）自动地挖掘样本之间

丰富的上下文信息。相比于基于反馈的方法，基于上下文的方法由于不需要任何手动的干涉具有更广泛的适用性。

Mang Ye 等人 [41] 认为对于不同的特征表达，同一个人应具有相同的排序结果，基于该假设对初始排序结果进行改进。Ye 等人 [42] 综合多个基于特征提取和度量学习方法得到的初始排序结果，进行排序优化，得到最优的排序结果。Bai 等人 [43] 融合多个度量函数，建立一个算法模型，实现无监督的重排序，并成功应用在了行人重识别问题上。这些重排序方法需要不同的基准算法得到的多个排序结果，因此将它们应用在现实场景中是不方便的。Wei Li 等人 [44] 通过对每对样本的近邻排序进行共性分析，重新优化了初始排序结果。Zhong 等人 [45] 通过比较样本对的 k 相互近邻的相似性，修正初始排序结果。相似地，在 [46] 中，作者提出 Expanded Cross Neighborhood (ECN) 距离度量，衡量样本对中这两个样本各自排序的相似性，从而重新计算样本对之间的相似值，进而对初始排序结果进行重排序。以上这三种方法都与排序结果之间的相似性匹配有关，在计算过程中，数据集中询问行人集合和候选行人（candidates）集合中的样本都要参与计算，然而在现实场景中，系统的输入往往是一个询问行人，以及一个候选行人集合，因此这三种方法也不方便应用在现实场景中。Bai 等人 [47] 提出通过基于流形的映射学习来实现性能的增强，由于该重排序需要训练学习，因此应用在现实场景中也是不方便的。还有一些研究人员依据图理论充分挖掘样本的上下文信息，构造了端到端的深度模型 [48–50]。比如，Yichao Yan 等人 [51] 提出一个上下文实例扩展模型（contextual instace expansion module），该模型可以自动搜索和过滤得到有用的上下文信息，然后作者建立了一个图学习框架，有效地探索了上下文对用于更新图像之间的相似值。Yantao Shen 等人 [52] 提出了一个 similarity-guided 图神经网络模型，一个询问样本（probe image）和所有候选样本（gallery image）构成一个图，用于表示样本对之间的关系，然后利用这个图通过一个端到端的方式更新样本对之间的关系。相比之下，本文作者充分探索了样本在其欧氏空间里的上下文信息，提出了双边上下文驱动的渐进行人重识别重排序方法，该方法不需要训练学习，并且只是利用了一个询问行人和候选行人集合的信息，更符合现实场景的设定。

此外，一部分研究人员致力于基于群体信息的行人重识别方法的研究。严格来说，这些方法中的一部分方法并不能归类为基于重排序的行人重识别方法中，

因为它们并不是继特征提取和度量学习步骤之后的操作，而是与特征提取和度量学习步骤有效的融合在一起。但是它们仍然与基于重排序的行人重识别方法有共性，即，这些方法都是通过挖掘行人自身信息之外的信息：群体信息，来完成重识别率的提高。Zheng 等人 [53] 针对跨摄像头群体中行人位置改变以及跨摄像头存在光照和视角变化这两种情况，相应地提出了两种群体特征，然后结合个人特征完成行人重识别。在 Cai 等人提出的基于群体的行人重识别方法中 [54]，群体特征通过协方差统计得到，然后结合行人的外表特征实现行人重识别。以上这些方法都是手动检测得到群体，然后对群体信息进行建模。Li 等人 [55] 采用仿射传播（Affinity Propagation, AP）聚类算法实现群体的自动检测，进而对检测到的群体进行建模，辅助行人重识别。Chen 等人 [56] 提出宽泛的群体概念，因此不需要进行群体检测，通过行人之间的时间差，定位时空域下的显著行人，辅助行人重识别。近几年，随着深度学习的发展，一些研究人员借助于深度学习技术，提出基于群体信息的行人重识别方法 [51, 57]。本文作者针对基于群体信息的行人重识别，也进行了深入研究，提出了基于广义群体信息的行人重识别方法。

2.4 基于深度学习的行人重识别方法

近些年，深度学习技术被广泛地应用在许多计算机视觉任务中。它作为一个强有力的工具，可以在不需要任何手动设计的特征的情况下，通过端到端的方式处理各种各样的计算机视觉任务。受到这项技术的启发，研究人员开始致力于采用深度学习技术解决行人重识别问题。参考文献 [58]，这些基于深度学习的行人重识别方法大致可以被分为六类：识别深度模型（identification deep model）、验证深度模型（verfication deep model）、基于距离度量的深度模型（distance metric-based deep model）、基于部件的深度模型（part-based deep model）、基于视频的深度模型（video-based deep model）和基于数据增广的深度模型（data augmentation-based deep model）。接下来，本文作者对这六类方法进行一个简单介绍。

识别深度模型 [59–61] 把行人重识别任务视为一个分类问题。在训练阶段，该模型充分利用数据集的标签信息，帮助指导提高训练精度；在测试阶段，向模型输入一张行人图像，然后通过计算，输出该图像所对应的标签信息。由于识别深度模型的训练目标与测试目标（也就是行人重识别的目标）不一致，因此对于

测试结果的准确性具有一定的影响。

验证深度模型 [62, 63] 把行人重识别任务视为一个二分类问题。模型的输入是一对行人图像，输出是这对行人图像之间的相似值，该相似值作为这对行人图像是正样本对还是负样本对的衡量。验证深度模型的输入是图像对，它仅仅利用了数据集的弱标签信息，因此对于训练结果的准确性有一定的影响。近几年，研究人员结合识别深度模型和验证深度模型实现行人重识别 [64, 65]，取得了不错的效果。

基于距离度量的深度模型 [66, 67] 的目标是学习一个最优的特征表达，使得正样本对的距离尽可能的小而负样本对的距离尽可能的大。最经典的基于距离度量的深度模型是三元组深度模型 [68]。该模型的输入是一个三元组单元 $I_i =< I_i^1, I_i^2, I_i^3 >$，其中 I_i^1 和 I_i^2 是正样本对，I_i^1 和 I_i^3 是负样本对。对于每一个三元组单元 I_i，模型致力于学习一个最优的特征表达，使得 I_i^1 和 I_i^2 的距离小于 I_i^1 和 I_i^3 的距离。基于距离度量的深度模型有效地挖掘了不同行人图像之间的关系，训练目标与测试目标一致。然而，模型需要构建三元组单元作为其输入，需要对所有三元组单元进行计算，使得训练效率较低。此外，模型只利用了数据集的弱标签信息。

大部分最新的基于深度学习的行人重识别方法都是采用基于部件的策略学习最优的特征映射 [10, 22, 69, 70]。像这样的方法可以被归类为基于部件的深度模型。局部视觉线索符合人类的视觉习惯，而局部视觉信息之间进行互补可以构成全局视觉信息。基于部件的深度模型结合局部特征和全局特征解决行人重识别问题。不过这类模型存在一些局限性：（1）把多个局部分枝加入到深度模型中，增加了模型的复杂度，降低了训练效率。（2）大多数基于部件的深度模型仅仅考虑了局部级别的信息，而忽略了像素级别的信息。（3）部件之间的相互关系也是一个有效的信息，然而目前大部分基于部件的深度模型都忽略了这个信息。

对于基于视频的行人重识别任务，近几年研究人员也采用了深度学习技术来解决这一任务，研究人员将用于解决基于视频的行人重识别任务的深度学习模型称之为基于视频的深度模型 [52, 71–73]。

在目前的行人重识别数据集中，行人的图像数量仍然是有限的。比如，对于大规模行人重识别数据集 CUHK03 [62]，Market1501 [47] 和 DukeMTMC [74]，一个行人的平均图像数量分别是 9.6，17.2 和 23.5。使用这样的数据集训练深度模

型可能会导致模型过拟合。为此，一些研究工作尝试通过深度模型扩展行人重识别数据集中的样本数量。比如，通过生成对抗网络（generative adversarial network (GAN)）生成新的样本图像。对于这样的深度模型，研究人员称之为基于数据增广的深度模型 [75, 76]。

2.5 针对具体问题的行人重识别方法

不同的行人重识别数据集是基于不同的拍摄条件采集得到的，现如今大多数最先进的行人重识别方法针对一个行人重识别数据集进行模型训练，得到的最优的模型在该数据集上的性能表现可以达到很好，然而在一个新的数据集上性能表现常常欠佳。为了使得学习到的最优的模型在实际问题中表现良好，近几年一些研究人员致力于基于非监督域适应（unsupervised domain adaptation）的行人重识别研究。比如，一些模型 [77–81] 采用基于图像合成的 GAN[82] 或者域对齐（domain alignment）。具体地，在一些模型中，研究人员首先利用源域（source domain）中的带标签样本学习一个深度网络，作为初始特征提取器，然后，在目标域（target domain）中进行非监督聚类修正这个初始的深度网络 [83]。这些方法存在一个缺陷：在修正初始网络时没有充分利用目标域中的监督样本信息。为此，一些域对齐方法在修正初始网络时，同时利用目标域中的非监督样本信息和监督样本信息 [77, 79, 84]。

最近，可扩展的行人重识别方法受到了越来越多的研究人员的关注。这类方法的提出的是为了减少模型扩展的成本。为了最小化对于标签样本的需求或者可以选择特定的数据用于训练，研究人员提出了非监督学习 [77–79, 83–87]、迁移学习 [75, 77, 84]、小样本学习 [29, 88] 和积极学习 [89–91] 用于解决行人重识别问题。为了减少模型的训练计算成本，研究人员提出了一些快速自适应的方法（fast adaptation）[47, 92]。为了减少模型的测试计算成本，研究人员提出了一些轻量级的快速检索模型 [93] 和二进制表现模型（binary representation）[94, 95].

2.6 本章小结

本章对行人重识别的研究现状进行了总结和介绍。目前，针对行人重识别的研究工作主要有三个步骤：特征提取、度量学习和重排序，本章分别从这三个方面介绍了行人重识别的相关方法。此外，随着近几年深度学习的不断发展，研究

人员致力于利用深度学习技术解决行人重识别问题，本章也对基于深度学习的行人重识别方法进行了总结和介绍。另外，针对一些具体问题的行人重识别方法，本章也进行了总结。

第 3 章 基于正交线性判别分析的行人重识别算法

3.1 引言

基于度量学习的行人重识别方法主要致力于学习一个合适的度量函数，或者一个具有良好判别性能的低维特征空间。在这些度量学习方法中，线性判别分析（Linear Discriminant Analysis, LDA）由于其无参数需要调试和闭式解的特点，成为了最典型和受欢迎的方法之一。LDA 旨在学习一个最优的线性变换，通过这个线性变换，可以得到一个新的低维的特征空间。在这个空间中，样本的类内方差最小，类间方差最大。对于典型的 LDA，样本的类内散度矩阵必须是非奇异矩阵。然而，对于行人重识别问题，样本的特征向量维度通常大于样本数量，从而导致样本的所有的散度矩阵都是奇异矩阵。这意味是行人重识别是一个奇异问题（或者称为欠采样问题），典型的 LDA 无法直接应用在行人重识别问题中。为了解决这个问题，研究人员提出了一些方法。比如，在 [9, 28] 中，正则化 LDA（Regularized LDA，RLDA）被有效地应用在了行人重识别问题上，RLDA 通过给类内散度矩阵加上扰动项，使类内散度矩阵成为一个非奇异矩阵。尽管它解决了奇异问题，但是研究人员需要通过参数调试选择一个合适的扰动值，并且 RLDA 存在退化特征值问题（degenerate eigenvalue problem），也就是一部分的特征向量可能具有同样的特征值，这就导致了 RLDA 得到的解可能并不是最优的解。Zhang 等人提出将零空间 LDA（Null Space LDA, NLDA）应用到行人重识别问题中。NLDA [29] 旨在在类内散度矩阵的零空间计算最优的具有判别力性能的特征空间，但是对于低维数据样本，类内散度矩阵的零空间为空，因此 NLDA 可能无法应用在低维数据样本中。

鉴于此，本文作者提出利用伪逆 LDA（Pseudo-inverse LDA, PLDA）解决行人重识别的奇异问题。PLDA 采用类内散度矩阵的伪逆来克服奇异问题，它的解等价于采用最小二乘法得到的近似解。相比于以上提到的 LDA 变体，PLDA 不仅解决了奇异问题，同时避免了以上 LDA 变体存在的缺陷。表3.1展示了这些 LDA 变体的目标函数。对于求解 PLDA，传统的方法是采用广义奇异值分解（Generalized Singular Value Decomposition, GSVD）[96]，其缺点是需要付出高昂的计算代价。取而代之，本文作者通过对类内散度矩阵、类间散度矩阵和总体散度矩阵同时对

表 3.1 LDA 变体的目标函数。S_w 和 S_b 分别表示样本的类内散度矩阵和类间散度矩阵。$(S)^+$ 表示矩阵 S 的伪逆。L^* 表示最优变换。

Table 3.1 The objective function of various variants of LDA. S_w and S_b denote the within-class scatter matrix and the between-class scatter matrix, respectively. $(S)^+$ denotes the pseudo-inverse of the matrix S. L^* denotes the optimal transformation.

Method	Objective function
LDA	$L^* = \arg\max trace\{(L^T S_w L)^{-1} L^T S_b L\}$
Regularized LDA	$L^* = \arg\max trace\{(L^T (S_w + \lambda I) L)^{-1} L^T S_b L\}$
Null space LDA	$L^* = \arg\max_{L^T S_w L=0} trace\{L^T S_b L\}$
Pseudo-inverse LDA	$L^* = \arg\max trace\{(L^T S_w L)^{+} L^T S_b L\}$

角化实现 PLDA 的求解 [97]。计算得到的最优的具有判别性能的向量彼此正交，因此也可以将该方法称之为正交 LDA（Orthogonal LDA, OLDA）。OLDA 是专门针对解决奇异问题的算法，它具有详尽的理论分析 [97]。此外，OLDA 具有以下三点重要特征：1）闭式解，2）没有需要调试的参数，3）对于样本中存在的噪声具有更好的鲁棒性，这些特征对于解决行人重识别问题都是非常友好的。

考虑到行人重识别问题中数据非线性分布的特点，本文作者进一步提出了方法的核化版本。它结合了 OLDA 和基于核学习技术的优点，可以获得更高的重识别率。此外，模型求解过程中涉及到特征值分解（eigen-decomposition）的计算。对于高维数据，相比于 QR 分解，特征值分解的运算量大。因此，本文作者进一步提出了方法的快速版本，方法的快速版本是采用 QR 分解代替特征值分解。本文作者在章节3.3将详细分析本文作者提出的方法的计算复杂度、运行时间和性能表现。

3.2 算法模型

3.2.1 问题描述

输入一个询问行人（query person）和一个候选人集合（candidate set），其中询问行人和候选人集合中的行人分别被两个不存在重叠监控区域的监控摄像头捕捉。本文作者的目标是通过建立模型以及求解模型，得到询问行人与每个候选人之间的相似度量值，对这些度量值进行降序排列，进而得到针对询问行人的候

表 3.2 模型中一些重要的数学符号的说明。

Table 3.2 The description of important notations used in the paper.

符号	描述	符号	描述
m	原始特征空间的特征维度	n	训练样本数量
d	新的特征空间的特征维度	X	训练样本的特征矩阵
k	类别数量	S_w	类内散度矩阵
S_b	类间散度矩阵	S_t	总体散度矩阵
r_b	S_b 矩阵的秩	r_t	S_t 矩阵的秩
L	变换矩阵	–	–

选人的排序列表。在训练阶段，通过本文作者提出的方法，学习得到一个最优的特征变换。在测试阶段，通过这个最优特征变换，本文作者将行人图像的原始特征向量转换到新的特征空间，然后计算行人图像特征向量之间的欧式距离。

考虑到在接下来的算法模型描述中存在许多数学符号，为了方便起见，本文作者在表3.2呈现了一些重要的数学符号说明。

3.2.2 典型的线性判别分析模型

在介绍本文作者提出的方法之前，本文作者首先简要介绍典型的 LDA 模型。

对于训练集中的 n 个行人，每个行人的特征向量为 m 维，这 n 个行人的特征向量构成一个特征矩阵 $X = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n] \in \mathbb{R}^{m\times n}$。典型的 LDA 模型是计算一个线性变换矩阵 $L \in \mathbb{R}^{m\times d}$ $(m > d)$，通过 L，可以把第 i 个行人的特征向量 $\boldsymbol{x}_i$ 映射到一个新的位于 d 维空间的特征向量 $\boldsymbol{y}_i$：

$$\boldsymbol{y}_i = L^T \boldsymbol{x}_i. \tag{3.1}$$

接下来，定义矩阵

$$\begin{aligned} H_w &= \frac{1}{\sqrt{n}} \left[X_1 - \boldsymbol{c}^{(1)}(\boldsymbol{e}^{(1)})^T, \ldots, X_k - \boldsymbol{c}^{(k)}(\boldsymbol{e}^{(k)})^T\right], \\ H_b &= \frac{1}{\sqrt{n}} \left[\sqrt{n_1}(\boldsymbol{c}^{(1)} - \boldsymbol{c}), \ldots, \sqrt{n_k}(\boldsymbol{c}^{(k)} - \boldsymbol{c})\right], \\ H_t &= \frac{1}{\sqrt{n}}(X - \boldsymbol{c}\boldsymbol{e}^T), \end{aligned} \tag{3.2}$$

$X_i \in \mathbb{R}^{m\times n_i}$ $(i = 1, \ldots, k)$ 是第 i 类样本的特征向量，$\sum_{i=1}^{k} n_i = n$，$\boldsymbol{e}^{(i)} = [1, \ldots, 1]^T \in$

$\mathbb{R}^{n_i}$，$\boldsymbol{e} = [1, \ldots, 1]^T \in \mathbb{R}^n$，$\boldsymbol{c}^{(i)} = \frac{1}{n_i} X_i \boldsymbol{e}^{(i)}$ 和 $\boldsymbol{c} = \frac{1}{n} X \boldsymbol{e}$ 分别是第 i 类样本的质心和全部样本的质心。

则类内散度矩阵 S_w、类间散度矩阵 S_b 和总体散度矩阵 S_t 可以被表示为

$$S_w = H_w H_w^T, \quad S_b = H_b H_b^T, \quad S_t = H_t H_t^T. \tag{3.3}$$

通过方程(3.2)和(3.3)，很容易验证得到 $S_t = S_b + S_w$。

LDA 模型的目标是找到一个最优的变换矩阵 L^*，使得在由 L^* 转化得到低维空间中，类内样本之间的分布更紧凑，同时类间样本之间的分布更松散：

$$L^* = \arg\max_L trace\{(L^T S_w L)^{-1} L^T S_b L\}. \tag{3.4}$$

这个优化问题等价于解决广义特征值问题 $S_b \boldsymbol{l} = \lambda S_w \boldsymbol{l}$。$S_w^{-1} S_b$ 的 $k-1$ 个最大特征值对应的特征向量即为最优的变换矩阵 L^*。但这个解的前提条件是 S_w 是一个非奇异矩阵。而在行人重识别问题中，通常 $m > n$，这就导致 S_w 奇异，典型的 LDA 模型无法应用在行人重识别问题。因此，本文作者提出通过考虑 S_w 的伪逆来解决奇异问题。通过对三个散度矩阵的同时对角化，求解最优的线性变换 L^*。

3.2.3 基于伪逆线性判别分析的正交变换学习

首先定义新的优化问题

$$L^* = \arg\max_L trace\{(L^T S_t L)^+ L^T S_b L\}, \tag{3.5}$$

$(L^T S_t L)^+$ 表示 $(L^T S_t L)$ 的伪逆，由于 $S_t = S_b + S_w$，因此 $(L^T S_t L)^+$ 等价于 $(L^T S_w L)^+$。

本文作者通过同时对角化三个散度矩阵解决上述优化问题。

理论 1. *H_t 的奇异值分解（Singular Value Decomposition, SVD）表示为 $H_t = U \Sigma V^T$，其中 $\Sigma = \begin{bmatrix} \Sigma_t & 0 \\ 0 & 0 \end{bmatrix}$，$\Sigma_t \in \mathbb{R}^{r_t \times r_t}$，$r_t = rank(S_t)$。令 $U = [U_1, U_2]$，其中 $U_1 \in \mathbb{R}^{m \times r_t}$，$U_2 \in \mathbb{R}^{m \times (m - r_t)}$。令 $T = \Sigma_t^{-1} U_1^T H_b$，计算 T 的 SVD，有 $T = P \widetilde{\Sigma} Q^T$。那么，$G = U_1 \Sigma_t^{-1} P$ 可以用来实现三个散度矩阵的同时对角化。*

证明. 根据 H_t 的奇异值分解，可以得到

$$
\begin{aligned}
&S_t = H_t H_t^T = U\Sigma V^T (U\Sigma V^T)^T = U\Sigma V^T V\Sigma^T U^T = U\begin{bmatrix} \Sigma_t^2 & 0 \\ 0 & 0 \end{bmatrix} U^T, \\
&\Rightarrow \quad U^T S_t U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} S_t \begin{bmatrix} U_1 & U_2 \end{bmatrix} = \begin{bmatrix} \Sigma_t^2 & 0 \\ 0 & 0 \end{bmatrix}, \\
&\Rightarrow \quad U_1^T S_t U_1 = \Sigma_t^2, \\
&\Rightarrow \quad \Sigma_t^{-1} U_1^T S_t U_1 \Sigma_t^{-1} = I_t.
\end{aligned} \tag{3.6}
$$

因为 $S_t = S_w + S_b$, 则

$$
\Sigma_t^{-1} U_1^T S_w U_1 \Sigma_t^{-1} + \Sigma_t^{-1} U_1^T S_b U_1 \Sigma_t^{-1} = I_t. \tag{3.7}
$$

令 $T = \Sigma_t^{-1} U_1^T H_b$，计算 T 的奇异值分解 $T = P\widetilde{\Sigma} Q^T$. 则

$$
\Sigma_t^{-1} U_1^T S_b U_1 \Sigma_t^{-1} = \Sigma_t^{-1} U_1^T H_b H_b^T U_1 \Sigma_t^{-1} = P\widetilde{\Sigma} Q^T Q \widetilde{\Sigma}^T P^T = P\Sigma_b P^T, \tag{3.8}
$$

其中 $\Sigma_b = \widetilde{\Sigma}^2$。因为 P 是一个正交矩阵，则有

$$
P^T \Sigma_t^{-1} U_1^T S_b U_1 \Sigma_t^{-1} P = \Sigma_b. \tag{3.9}
$$

因此，参考方程(3.7)和(3.8)，可以得到

$$
\begin{aligned}
&\Sigma_t^{-1} U_1^T S_w U_1 \Sigma_t^{-1} = I_t - P\Sigma_b P^T, \\
&\Rightarrow \; P^T \Sigma_t^{-1} U_1^T S_w U_1 \Sigma_t^{-1} P = I_t - \Sigma_b = \Sigma_w.
\end{aligned} \tag{3.10}
$$

由于 P 是一个正交矩阵，那么方程(3.6)的最后一行可以被重新表达为

$$
P^T \Sigma_t^{-1} U_1^T S_t U_1 \Sigma_t^{-1} P = I_t. \tag{3.11}
$$

现在，结合方程(3.9)，(3.10)和(3.11)，定义 $G = U_1 \Sigma_t^{-1} P$，则

$$
G^T S_t G = I_t, \quad G^T S_b G = \Sigma_b, \quad G^T S_w G = \Sigma_w. \tag{3.12}
$$

通过 $G = U_1 \Sigma_t^{-1} P$，可以实现三个散度矩阵的同时对角化。

□

根据理论 1，重新表达 $L^T S_t L$ 和 $L^T S_b L$ 为

$$\begin{aligned} L^T S_t L &= L^T (G^{-1})^T (G^T S_t G) G^{-1} L = \widetilde{L}^T I_t \widetilde{L} = \widetilde{L}^T \widetilde{L}, \\ L^T S_b L &= L^T (G^{-1})^T (G^T S_b G) G^{-1} L = \widetilde{L}^T \Sigma_b \widetilde{L}, \end{aligned} \tag{3.13}$$

其中

$$\widetilde{L} = G^{-1} L. \tag{3.14}$$

则可以得到

$$F = trace\{(L^T S_t L)^+ L^T S_b L\} = trace\{(\widetilde{L}^T \widetilde{L})^+ \widetilde{L}^T \Sigma_b \widetilde{L}\} = trace\{(\widetilde{L}\widetilde{L}^+)^T \Sigma_b (\widetilde{L}\widetilde{L}^+)\}. \tag{3.15}$$

然后计算 $\widetilde{L}$ 的奇异值分解：

$$\widetilde{L} = M \begin{bmatrix} \Sigma_l & 0 \\ 0 & 0 \end{bmatrix} N^T. \tag{3.16}$$

其中 $\Sigma_l \in \mathbb{R}^{r_l \times r_l}$ with $r_l = rank(\widetilde{L})$。因为 M 和 N 都是正交矩阵，Σ_l 是对角矩阵，则有

$$\widetilde{L}\widetilde{L}^+ = M \begin{bmatrix} \Sigma_l & 0 \\ 0 & 0 \end{bmatrix} N^T N \begin{bmatrix} \Sigma_l^{-1} & 0 \\ 0 & 0 \end{bmatrix} M^T = M \begin{bmatrix} I_l & 0 \\ 0 & 0 \end{bmatrix} M^T. \tag{3.17}$$

那么，方程(3.15) 可以被重新表达：

$$\begin{aligned} F &= trace\{M \begin{bmatrix} I_l & 0 \\ 0 & 0 \end{bmatrix} M^T \Sigma_b M \begin{bmatrix} I_l & 0 \\ 0 & 0 \end{bmatrix} M^T\} \\ &= trace\{\begin{bmatrix} I_l & 0 \\ 0 & 0 \end{bmatrix} M^T \Sigma_b M \begin{bmatrix} I_l & 0 \\ 0 & 0 \end{bmatrix}\} \\ &= trace\{M_l^T \Sigma_b M_l\}, \end{aligned} \tag{3.18}$$

其中 M_l 是矩阵 M 的前 r_l 列。

现在，优化问题(3.5)变成了 $trace\{M_l^T \Sigma_b M_l\}$ 的最大化。对于这个最大化问题的求解，本文作者首先引入一个引理 [98]。

引理 1. 对于任意一个满足 $A^T A = I$ 的矩阵 $A \in \mathbb{R}^{m \times q}$ $(q \leq m)$ 和一个半正定矩阵 $J \in \mathbb{R}^{m \times m}$，$trace(A^T J A) \leq \lambda_1 + \cdots + \lambda_q$。当 $A = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_q] E$，等式满足，即 $trace(A^T J A) = \lambda_1 + \cdots + \lambda_q$。其中 $\boldsymbol{h}_i$ 是矩阵 J 的第 i 大的特征值 λ_i 所对应的特征向量，$E \in \mathbb{R}^{q \times q}$ 是一个任意的正交矩阵。

由于篇幅有限，引理 1 的证明不具体展示，感兴趣的读者可以查阅 [98]。

根据引理 1，可以得到 $F = trace\{M_l^T \Sigma_b M_l\} \leq \lambda_1 + \cdots + \lambda_{r_b}$，其中 $r_b = rank(S_b)$。当 $M_l = \begin{bmatrix} E \\ 0 \end{bmatrix}$ 时，F 取得最大值。其中 $E \in \mathbb{R}^{r_b \times r_b}$ 是一个满足 $E^T E = I$ 的任意矩阵。把 M_l 代入方程(3.16)，由于有 $r_b = r_l$，则可以得到 $\widetilde{L} = \begin{bmatrix} E\Sigma_l N^T \\ 0 \end{bmatrix}$。因为 E 和 N 都是正交矩阵，Σ_l 是对角矩阵，本文作者表示 $A = E\Sigma_l N^T$ 作为一个任意非奇异矩阵。

最后，结合方程(3.14)，可以得到方程(3.5)的优化问题的解为 $L^* = GA$.

对于这个解当中的任意非奇异矩阵 A 的选择，一个最简单的选择是 $A = I$，对应地，$L^* = G$。结合方程(3.12)，可以发现 $L^{*T} S_t L^* = I_t$。这意味着求解得到的最优的变换矩阵中任意两个向量都是彼此 S_t 正交的。通过这样的变换矩阵 L^* 得到的新的特征向量彼此是不相关的。选择 $A = I$，其所对应的模型称之为不相关 LDA 模型（Uncorrelated LDA，ULDA）。然而由于通过 ULDA 模型得到的新的特征空间存在最小冗余信息，导致模型可能出现过拟合的现象，以及对数据中的噪声比较敏感。

鉴于此，本文作者考虑 A 的其它选择。对矩阵 G 进行 QR 分解 $G = Q_g R_g$，选择 $A = R_g^{-1}$。对应地，得到 $L^* = Q_g$，且满足 $L^{*T} L^* = I$。该模型称之为 OLDA 模型。对于上述的 ULDA 模型可能存在的过拟合问题，OLDA 模型有效地避免这个问题。

总而言之，本文作者首先采用 PLDA 模型通过训练样本学习了一个最优的变换矩阵 L^*，该变换矩阵中的向量彼此正交。然后通过方程(3.1)把变换矩阵 L^* 作用在测试样本的特征向量上，得到新的具有良好判别性能的特征向量。最后计算新的特征向量之间的欧式距离，得到行人图像之间的距离度量值。算法1描述了本文作者提出的方法。

3.2.4 正交变换学习的核化版本

通过算法1，本文作者学习得到了一个最优的线性变换矩阵，但它只能用于解决线性问题。然而由于行人外貌特征的非线性特点，在大多数情况下，行人重识别是一个非线性问题。为了使得提出的方法更适合行人重识别问题，本文作者进一步提出了方法的核化版本，并在这一节进行详细介绍。

算法 1 学习一个应用于行人重识别问题的正交变换

输入：

1: 训练样本的特征矩阵 X 和其对应的标签信息。

2: 测试样本的特征向量 $\boldsymbol{x}_i^{test}$ 和 $\boldsymbol{x}_j^{test}$。

输出：$\boldsymbol{x}_i^{test}$ 和 $\boldsymbol{x}_j^{test}$ 之间的距离度量值 $D(\boldsymbol{x}_i^{test}, \boldsymbol{x}_j^{test})$。

步骤：

1: 根据方程3.2构造矩阵 H_b 和 H_t。

2: 计算 H_t 的特征值分解 $H_t = U\begin{bmatrix} \Sigma_t & 0 \\ 0 & 0 \end{bmatrix}V^T$.

3: 计算 T 的特征值分解 $T = P\widetilde{\Sigma}Q^T$，其中 $T = \Sigma_t^{-1}U_1H_b$，U_1 是矩阵 U 的前 r_t 列，$r_t = rank(H_tH_t^T)$。

4: 计算 G 的 QR 分解，其中 $G = U_1\Sigma_t^{-1}P$。

5: 令 $L^* = Q_g$，它就是优化问题3.5的最优解。

返回：$D(\boldsymbol{x}_i^{test}, \boldsymbol{x}_j^{test}) = \left\|L^{*T}\boldsymbol{x}_i^{test} - L^{*T}\boldsymbol{x}_j^{test}\right\|^2$;

通过一个非线性映射函数 Φ，原始的特征空间可以被映射到一个新的高维核空间：$\boldsymbol{x} \in \mathbb{R}^m \rightarrow \boldsymbol{\Phi}(\boldsymbol{x}) \in \mathbb{R}^r$ $(r \gg m)$。新的特征矩阵可以表示为 $\Phi(X) = \left[\boldsymbol{\Phi}(\boldsymbol{x}_1), \ldots, \boldsymbol{\Phi}(\boldsymbol{x}_n)\right] \in \mathbb{R}^{r\times n}$。

S_w^Φ、S_b^Φ 和 S_t^Φ 表示在核空间中的三个散度矩阵，则有

$$\begin{aligned} S_w^\Phi &= \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_i}[\boldsymbol{\Phi}(\boldsymbol{x}_i^j) - \boldsymbol{\Phi}(\boldsymbol{c}^{(j)})][\boldsymbol{\Phi}(\boldsymbol{x}_i^j) - \boldsymbol{\Phi}(\boldsymbol{c}^{(j)})]^T \\ S_b^\Phi &= \frac{1}{n}\sum_{j=1}^{k} n_i[\boldsymbol{\Phi}(\boldsymbol{c}^{(j)}) - \boldsymbol{\Phi}(\boldsymbol{c})][\boldsymbol{\Phi}(\boldsymbol{c}^{(j)}) - \boldsymbol{\Phi}(\boldsymbol{c})]^T \\ S_t^\Phi &= \frac{1}{n}\sum_{i=1}^{n}[\boldsymbol{\Phi}(\boldsymbol{x}_i) - \boldsymbol{\Phi}(\boldsymbol{c})][\boldsymbol{\Phi}(\boldsymbol{x}_i) - \boldsymbol{\Phi}(\boldsymbol{c})]^T, \end{aligned} \tag{3.19}$$

其中 $\boldsymbol{\Phi}(\boldsymbol{c}^{(j)})$ 和 $\boldsymbol{\Phi}(\boldsymbol{c})$ 分别是核空间中第 j 类样本的质心和核空间中全部样本的质心。

直接计算映射 $\boldsymbol{\Phi}(\boldsymbol{x})$，然后运行算法1，是困难且计算昂贵的。取而代之，可以把内积运算替换成核函数 $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{\Phi}(\boldsymbol{x}_i)\cdot\boldsymbol{\Phi}(\boldsymbol{x}_j)$，将数据隐式地映射到核空间中。

对于向量 $\boldsymbol{l}^{\boldsymbol{\Phi}}$，满足 $\boldsymbol{l}^{\boldsymbol{\Phi}} = \sum_{i=1}^{n}\alpha_i\boldsymbol{\Phi}(\boldsymbol{x}_i)$，其中系数向量 $\boldsymbol{\alpha} = \left[\alpha_1, \ldots, \alpha_n\right]^T$。则

有

$$\begin{aligned}(l^{\Phi})^T S_b^{\Phi} l^{\Phi} &= \alpha^T [K(B-O)(B-O)^T K]\alpha \\ (l^{\Phi})^T S_t^{\Phi} l^{\Phi} &= \alpha^T [K(I-O)(I-O)^T K]\alpha,\end{aligned} \tag{3.20}$$

其中，$K = \Phi(X)^T \Phi(X)$，$B = diag(B_1, \cdots, B_k) \in \mathbb{R}^{n\times n}$，$B_i \in \mathbb{R}^{n_i \times n_i}$ $(i = 1, \ldots, k)$ 且 B_i 的所有项都为 $\frac{1}{n_i}$，$O \in \mathbb{R}^{n\times n}$ 且 O 的所有项都为 $\frac{1}{n}$，I 是一个 $n \times n$ 的单位矩阵。

接下来本文作者对方程(3.20)进行详细推导。

根据方程(3.19)，可以得到

$$\begin{aligned}(l^{\Phi})^T S_w^{\Phi} l^{\Phi} &= (l^{\Phi})^T \sum_{j=1}^{k} \sum_{i=1}^{n_i} [\Phi(x_i^j) - \Phi(c^{(j)})][\Phi(x_i^j) - \Phi(c^{(j)})]^T l^{\Phi}, \\ (l^{\Phi})^T S_t^{\Phi} l^{\Phi} &= (l^{\Phi})^T \sum_{i=1}^{n} [\Phi(x_i) - \Phi(c)][\Phi(x_i) - \Phi(c)]^T l^{\Phi}.\end{aligned} \tag{3.21}$$

把 $l^{\Phi} = \sum_{i=1}^{n} \alpha_i \Phi(x_i)$ 代入方程(3.21)，有

$$\begin{aligned}(l^{\Phi})^T S_w^{\Phi} l^{\Phi} &= \alpha^T \Phi(X)^T \sum_{j=1}^{k} \sum_{i=1}^{n_i} [\Phi(x_i^j) - \Phi(c^{(j)})][\Phi(x_i^j) - \Phi(c^{(j)})]^T \Phi(X)\alpha \\ &= \alpha^T \Gamma_w^{\Phi} \Gamma_w^{\Phi^T} \alpha, \\ (l^{\Phi})^T S_t^{\Phi} l^{\Phi} &= \alpha^T \Phi(X)^T \sum_{i=1}^{n} [\Phi(x_i) - \Phi(c)][\Phi(x_i) - \Phi(c)]^T \Phi(X)\alpha \\ &= \alpha^T \Gamma_t^{\Phi} \Gamma_t^{\Phi^T} \alpha,\end{aligned} \tag{3.22}$$

其中

$$\begin{aligned}\Gamma_w^{\Phi} &= \sum_{j=1}^{k} \sum_{i=1}^{n_i} [\Phi(X)^T \Phi(x_i^j) - \frac{1}{n_i} \sum_{q=1}^{n_i} \Phi(X)^T \Phi(x_q^j)] \\ &= \sum_{j=1}^{k} \sum_{i=1}^{n_i} [k(X, x_i^j) - \frac{1}{n_i} \sum_{q=1}^{n_i} k(X, x_q^j)] \\ &= KI - KB, \\ \Gamma_t^{\Phi} &= \sum_{i=1}^{n} [\Phi(X)^T \Phi(x_i) - \frac{1}{n} \sum_{q=1}^{n} \Phi(X)^T \Phi(x_q)] \\ &= \sum_{i=1}^{n} [k(X, x_i) - \frac{1}{n} \sum_{q=1}^{n} k(X, x_q)] \\ &= KI - KO.\end{aligned} \tag{3.23}$$

$I \in \mathbb{R}^{n\times n}$ 是一个单位矩阵，$B = diag(B_1, \cdots, B_k) \in \mathbb{R}^{n\times n}$，$B_i \in \mathbb{R}^{n_i \times n_i}$ 是一个所有项都等于 $\frac{1}{n_i}$ 的矩阵，$O \in \mathbb{R}^{n\times n}$ 是一个所有项都等于 $\frac{1}{n}$ 的矩阵。

继而有

$$\begin{aligned}(l^{\Phi})^T S_w^{\Phi} l^{\Phi} &= \alpha^T [K(I-B)(I-B)^T K]\alpha, \\ (l^{\Phi})^T S_t^{\Phi} l^{\Phi} &= \alpha^T [K(I-O)(I-O)^T K]\alpha.\end{aligned} \tag{3.24}$$

又因为 $S_t^{\Phi} = S_b^{\Phi} + S_w^{\Phi}$，则

$$(l^{\Phi})^T S_b^{\Phi} l^{\Phi} = \alpha^T [K(B-O)(B-O)^T K]\alpha. \tag{3.25}$$

至此，完成了对方程(3.20)的推导。

根据方程(3.20)，三个散度矩阵被核化为

$$K_w = K(I-B)(I-B)^T K, \quad K_b = K(B-O)(B-O)^T K, \quad K_t = K(I-O)(I-O)^T K. \tag{3.26}$$

现在，可以重新表达优化问题(3.5)为

$$\mathcal{A}^* = \arg\max_{\mathcal{A}} trace\{(\mathcal{A}^T K_t \mathcal{A})^+ \mathcal{A}^T K_b \mathcal{A}\} \tag{3.27}$$

其中 $\mathcal{A} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_d]$。

然后参考算法1，就可以计算得到在核空间中的 d 个正交的具有良好判别性能的向量了。特别地，在测试阶段，通过 $\boldsymbol{y}_i^{test} = (L^{\Phi})^T \cdot \boldsymbol{\Phi}(\boldsymbol{x}_i^{test}) = \mathcal{A}^{*T} k(X, \boldsymbol{x}_i^{test})$，得到新的特征向量。

3.2.5 正交变换学习的快速版本

在这一节，本文作者呈现提出的方法的快速计算版本。具体地，本文作者将计算过程中的特征值分解替换为 QR 分解。

重新表达特征向量 $X = [X_1, X_2, \ldots, X_k]$，$X_i \in \mathbb{R}^{m \times n_i}$ $(i = 1, \ldots, k)$ 是第 i 类样本的特征矩阵。

首先，计算 X 的经济型 QR 分解（economy-size QR decomposition），$X = Q_x R_x$，其中 $Q_x \in \mathbb{R}^{m \times n}$，$R_x \in \mathbb{R}^{n \times n}$。相比于全 QR 分解（full QR decomposition）$X = QR$，其中 $Q \in \mathbb{R}^{m \times m}$，$R \in \mathbb{R}^{m \times n}$，经济型 QR 分解只需要计算 Q 的前 n 列，R 的前 n 行即可。

令 Z 为置换矩阵，用来交换单位矩阵 $I \in \mathbb{R}^{n \times n}$ 的第 i 列和第 $(\sum_{j=1}^{i-1} n_j + 1)$ 列，H_i $(i = 1, \ldots, k)$ 和 H 分别为向量 $[1, \ldots, 1]^T \in \mathbb{R}^{n_i}$ 和 $\left[\sqrt{n_1}, \ldots, \sqrt{n_k}\right]^T$ 的 Householder 变换。则有

$$R_x \begin{bmatrix} H_1 & & \\ & \ddots & \\ & & H_k \end{bmatrix} Z \begin{bmatrix} H & \\ & I \end{bmatrix} = [R_1, R_2, R_3], \tag{3.28}$$

其中 $R_1 \in \mathbb{R}^{n \times 1}$，$R_2 \in \mathbb{R}^{n \times (k-1)}$，$R_3 \in \mathbb{R}^{n \times (n-k)}$。

计算 $[R_2, R_3]$ 的 rank-revealing QR 分解（也称为列主元经济 QR 分解，economic QR decomposition with column pivoting）

$$[R_2, R_3]\, P_{qr} = \widetilde{Q}\widetilde{R} = \left[\widetilde{Q}_1, \widetilde{Q}_2, \widetilde{Q}_3\right] \begin{bmatrix} \widetilde{R}_{11} & \widetilde{R}_{12} \\ 0 & \widetilde{R}_{22} \\ 0 & \widetilde{R}_{32} \end{bmatrix}, \tag{3.29}$$

算法 2 正交线性变换学习的快速版本

输入：

1: 训练样本的特征矩阵 X 和其对应的标签信息。

2: 测试样本的特征向量 $\boldsymbol{x}_i^{test}$ 和 $\boldsymbol{x}_j^{test}$。

输出：$\boldsymbol{x}_i^{test}$ 和 $\boldsymbol{x}_j^{test}$ 之间的距离度量值 $D(\boldsymbol{x}_i^{test}, \boldsymbol{x}_j^{test})$。

步骤：

1: 计算 X 的经济型 QR 分解 $X = Q_x R_x$。

2: 构造矩阵 Z，H 和 $H_i\ (i = 1, \ldots, k)$，根据方程(3.28)计算 $[R_1, R_2, R_3]$。

3: 计算 $[R_2, R_3]$ 的 rank-revealing QR 分解 $[R_2, R_3] = \left[\widetilde{Q}_1, \widetilde{Q}_2, \widetilde{Q}_3\right] \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \\ 0 & 0 \end{bmatrix}$。

4: 计算 QR 分解 $\begin{bmatrix} R_{12} \\ R_{22} \end{bmatrix} R_{22}^T = \bar{Q} \begin{bmatrix} \bar{R} \\ 0 \end{bmatrix}$。

5: 令 $L^* = Q_x \left[\widetilde{Q}_1, \widetilde{Q}_2\right] \bar{Q}$，它是优化问题(3.5)的最优解。

返回：$D(\boldsymbol{x}_i^{test}, \boldsymbol{x}_j^{test}) = \left\| L^{*T}\boldsymbol{x}_i^{test} - L^{*T}\boldsymbol{x}_j^{test} \right\|^2$；

其中 P_{qr} 是一个置换矩阵，$\widetilde{R}_{11}$ 是一个上三角矩阵，$\widetilde{R}_{32} \approx 0$，$\widetilde{Q}_1 \in \mathbb{R}^{n \times r_2}$，$\widetilde{Q}_2 \in \mathbb{R}^{n \times r_3}$，$\widetilde{Q}_3 \in \mathbb{R}^n$，$r_2 = rank(R_2)$，$r_3 = rank(R_3)$，$\widetilde{R}_{11} \in \mathbb{R}^{r_2 \times (k-1)}$，$\widetilde{R}_{12} \in \mathbb{R}^{r_2 \times (n-k)}$。令

$$\begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \widetilde{R}_{11} & \widetilde{R}_{12} \\ 0 & \widetilde{R}_{22} \\ 0 & \widetilde{R}_{32} \end{bmatrix} P_{qr}^T. \tag{3.30}$$

则可以获得 $[R_2, R_3]$ 的 QR 分解。

相应地，$\begin{bmatrix} R_{12} \\ R_{22} \end{bmatrix} R_{22}^T$ 的 QR 分解为

$$\begin{bmatrix} R_{12} \\ R_{22} \end{bmatrix} R_{22}^T = \bar{Q} \begin{bmatrix} \bar{R} \\ 0 \end{bmatrix}. \tag{3.31}$$

则优化问题(3.5)的解为 $L^* = Q_x \left[\widetilde{Q}_1, \widetilde{Q}_2\right] \bar{Q}$。对于详细的证明，感兴趣的读者可查阅 [99]。

正如前所述，本文作者采用了 QR 分解代替特征值分解，求解优化问题(3.5)。算法2总结了提出的方法的快速版本，相比于算法1，算法2的计算复杂度更低。本文作者将会在章节3.3对算法的计算复杂度进行详细分析。

3.3 实验与分析

3.3.1 数据集与实验设置

1）行人重识别数据集简介

本文作者分别在 4 个被广泛使用的行人重识别基准数据集：VIPeR [1]，PRID2011 [100]，CUHK01 [101]，CUHK03 [62] 上进行实验。

VIPeR 数据集中包含 632 个行人图像对，每一对图像中的行人都是从两个不存在重叠监控区域的户外监控摄像头中捕捉到的，因此图像之间存在视角、光照和姿态等方面比较大的变化。遵循标准设置，本文作者归一化所有的图像为 128×48 的尺寸，以比率 1：1 随机划分数据集为训练集和测试集两部分。

PRID2011 数据集由来自两个不同的监控摄像头拍摄到的行人图像组成。这两个摄像头分别捕捉到了 385 和 749 个行人，其中只有 200 个行人在两个摄像头中都出现过。所有的行人图像都被归一化到 128×64 的尺寸。本文作者使用数据集的单图像版本（single shot version）进行实验。本文作者从 200 个在两个摄像头中都出现过的行人图像中选择 100 个行人图像对，作为训练集；剩余的 100 个行人图像对中的一个摄像头拍摄的行人组成测试集的询问行人集合（probe set），而这 100 个行人图像对中的另一个摄像头拍摄的行人，以及数据集中其余 549 个行人图像组成测试集的候选行人集合（gallery set）。由于测试集的候选行人集合中有许多干扰样本，因此这个数据集对于行人重识别问题来说是非常具有挑战性的数据集。

CUHK01 数据集是行人重识别数据集中样本数量比较多的一个数据集。该数据集中的图像都是来自于校园监控摄像头拍摄到的图像，总共收集了 971 个行人的 3,884 张图像。每个行人在每个摄像头下捕获到了 2 张图像。所有图像都归一化到 160×60 的尺寸。一个摄像头捕捉行人的侧面，另一个摄像头捕捉行人的正面或者背面。本文作者采用与 [29] 同样的设置：多图像版本（multi-shot setting），且 485/486 的训练集/测试集划分进行实验。

CUHK03 数据集是样本数量最多的行人重识别数据集之一。它包含有 1,360 个行人的 13,164 张图像。这些图像来自于 6 个校园监控摄像头拍摄到的监控视频，每个行人在其中的两个摄像头下被捕捉到，且在每个摄像头下平均捕获到 4.8 张图像。这个数据集分为 CUHK03(manual) 数据集和 CUHK03(detected) 数据集，CUHK03(manual) 数据集中的行人图像是通过手动裁剪得到的，CUHK03(detected) 数据集中的行人图像是通过行人检测模型 Deformable-Part-Model (DPM) [5] 得到的，它更贴近现实生活中的行人重识别系统。基于数据集的单图像设置（single shot setting），1,260 张图像和剩余的 100 张图像分别作为训练集和测试集。所有的图像都被归一化为 128×48 的尺寸。

2）实验协议

本文作者采用累计匹配特性曲线（Cumulative Matching Characteristics，CMC）作为性能评测指标，Rank k 表示正样本排序在前 k 个位置的概率。除了 CUHK03 数据集，每个数据集随机划分训练集和测试集 10 次进行实验，CUHK03 数据集的训练集/测试集划分为 20 次。本文作者取多次实验结果的平均性能作为最终结果，并报告 CMC 曲线中的 Rank 1、Rank 5 和 Rank 10。

3）参数设置

在提出的方法中，本不需要调节任何参数。然而由于对方法进行了核化，因此必须选择合适的核函数参数。本文作者通过 2-折交叉验证选择这些参数。具体地，对于每一个随机划分得到的训练集/测试集，本文作者随机选择训练集中 90% 的样本作为新的训练集，其余的 10% 作为验证集。另外，在实验中，为了实验对比，本文作者采用了 PCA 技术，90% 的能量被保存作为降维之后的数据。

4）行人图像特征选择

为了对提出的方法进行一个全面的评估，在实验中，本文作者分别采用了三个特征描述符作为行人图像特征的表达：Histogram of Intensity Pattern & His-

togram of Ordinal Pattern (HIPHOP) [102]，Local Maximal Occurrence (LOMO) [9] 和 Gaussian Of Gaussian (GOG) [2]。HIPHOP 特征描述符是借助 AlexNet [103] 深度学习框架得到的深度行人外貌特征表示[1]。LOMO 特征描述符分析了图像中每个局部特征的水平方向的出现频率，该描述符对于跨摄像头视觉变化造成的行人图像变化具有一定的鲁棒性。GOG 描述符通过利用分层高斯分布统计量表示行人图像，分别从不同的颜色通道提取特征：GOG_{RGB}、GOG_{Lab}、GOG_{HSV} 和 GOG_{RnG}。图3.1总结了特征维数和 4 个行人重识别数据集中样本数量的统计量。对于以上提到的所有特征描述符，其特征维数都大于 4 个行人重识别数据集中样本的数量，这表明奇异问题确实存在于行人重识别问题中。为了进一步验证提出的方法对于不存在奇异问题的行人重识别的性能，以及 NLDA 对于低维数据的性能，本文作者对 GOG 特征描述符进行了 PCA 降维，表示为 GOG_{PCA}，并将该特征描述符也用于实验中。从图3.1，可以看到，在所有行人重识别数据集中，GOG_{PCA} 的维数都小于样本数量，这表明了问题的非奇异性。

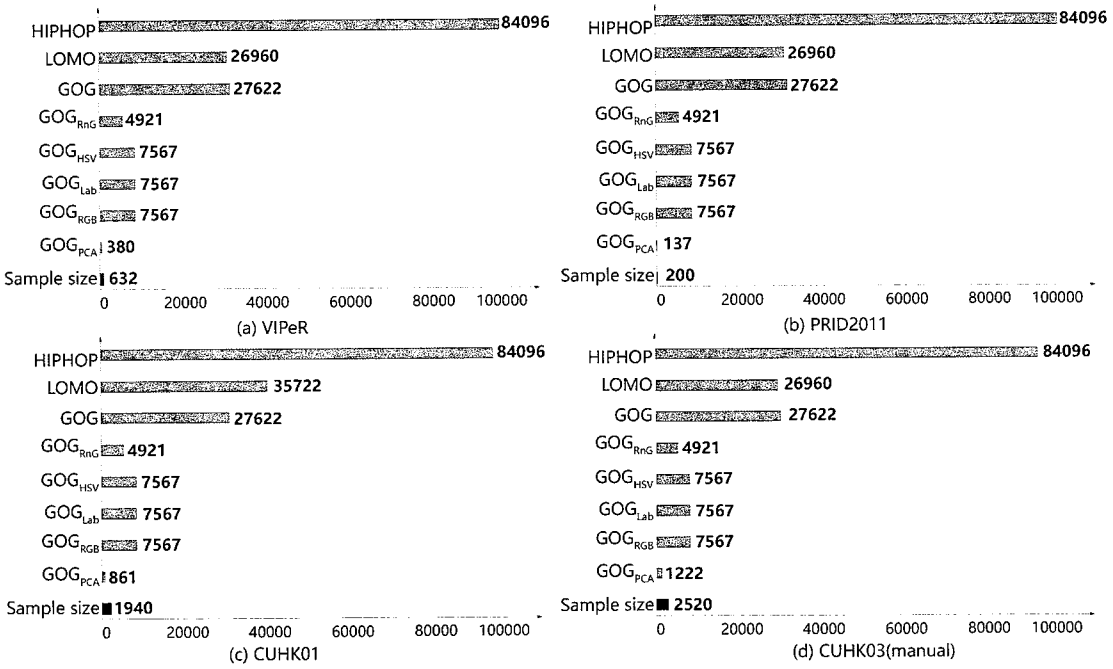


图 3.1 特征维数和 4 个行人重识别数据集中样本数量的统计量。

Figure 3.1 The statistics of dimensionality of feature and sample size on four datasets.

[1]在提取 HIPHOP 特征时，遵循标准设置 [102]，所有图像被归一化为 227×227 的尺寸。

3.3.2 性能比较

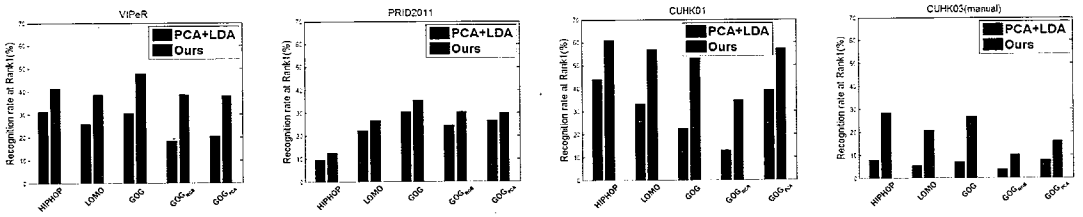
3.3.2.1 与各种 LDA 变体的性能比较

对于提出的方法，本文作者在 VIPeR，PRID2011，CUHK01 和 CUHK03(manual) 数据集上与各种 LDA 变体进行性能比较。相比较的方法有 PCA+LDA [104]、RLDA [105]、ULDA [97] 和 NLDA [106]。为了全面的比较，HIPHOP、LOMO、GOG 和 GOG_{RGB} 特征描述符分别被用来表示行人图像特征。在这些情况下，行人重识别是一个奇异问题。此外，为了验证提出的方法对于不存在奇异问题的行人重识别的性能，以及 NLDA 对于低维数据的性能，本文作者也采用了 GOG_{PCA} 特征描述符表示行人图像特征。

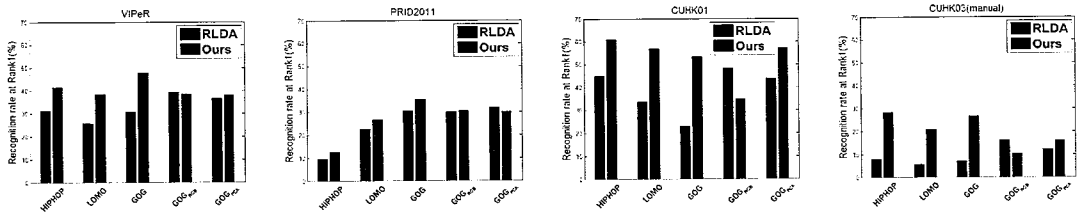
图3.2和表3.3展示了对比实验的结果。可以看到：1）在所有的数据集上，对于所有的特征描述符，本文作者提出的方法都优于 PCA+LDA 和 ULDA。这表明和这两种方法相比，本文作者提出的方法更适合解决行人重识别问题。2）除了低维数据，比如 GOG_{RGB}，本文作者提出的方法在性能方面显著优于 RLDA。对于 RLDA 算法，需要仔细调节参数以求获得最好的结果。相反，对于本文作者提出的方法，不需要调节任何参数。3）在大多数情况下，NLDA 具有与本文作者提出的方法相似的性能表现。然而，对于低维数据，比如 GOG_{RGB}，本文作者提出的方法具有比 NLDA 更好的性能表现。特别地，NLDA 在大数据集 CUHK01 和 CUHK03(manual) 上性能表现较差，其 Rank 1 分别仅仅只有 0.2% 和 1.0%。这表明了 NLDA 算法无法应用在低维数据样本中。4）在 VIPeR，CUHK01 和 CUHK03(manual) 数据集上，对于 GOG_{PCA} 特征描述符，本文作者提出的方法性能最佳。这表明了本文作者提出的方法对于解决不存在奇异问题的行人重识别的优越性。

3.3.2.2 与方法核化版本的性能比较

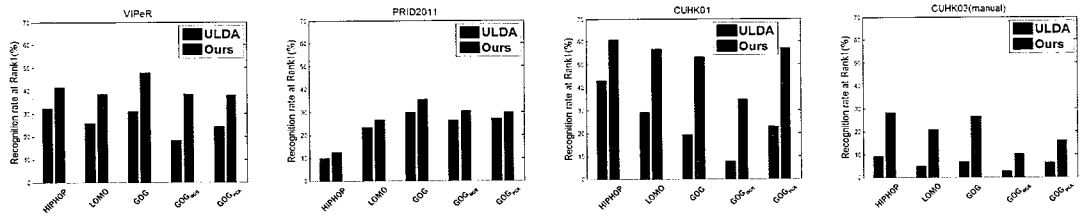
本小节通过将不同的核函数应用在提出的方法中，考察其在行人重识别数据集 VIPeR，PRID2011，CUHK01 和 CUHK03(manual) 上的性能。在这个实验中，HIPHOP，LOMO 和 GOG 特征描述符分别用来表示行人图像特征。此外，对 HIPHOP，LOMO 和 GOG 特征描述符分别得到的距离度量值进行加和作为最终的距离度量值，本文作者也呈现了其对应的结果，以验证针对多个特征描述符融合的情况，本文作者提出的方法的性能表现。



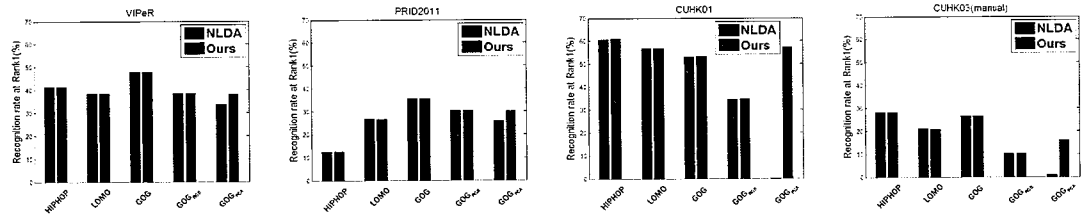
(a) comparison with PCA+LDA



(b) comparison with RLDA



(c) comparison with ULDA



(d) comparison with NLDA

图 3.2 与各种 LDA 变体的 Rank 1 比较结果。

Figure 3.2 Comparison of Rank 1 with various variants of LDA.

表 3.3 与各种 LDA 变体的性能比较结果。最好的结果用粗体显示。

Table 3.3 Comparison with various variants of LDA. The best results are shown in boldface.

	Methods	VIPeR		PRID2011		CUHK01		CUHK03(m)	
		r=1	r=5	r=1	r=5	r=1	r=5	r=1	r=5
HIPHOP	PCA+LDA	31.3	54.9	9.6	23.2	44.1	61.9	7.9	23.0
	RLDA	31.4	55.0	9.5	23.3	44.9	62.9	7.8	22.9
	ULDA	32.5	56.6	9.9	24.9	43.1	61.0	9.1	22.5
	NLDA	**41.5**	69.7	**12.5**	27.9	60.7	81.2	**28.3**	52.6
	Ours	**41.5**	69.7	**12.5**	27.9	**60.9**	81.1	**28.3**	52.6
LOMO	PCA+LDA	25.7	50.0	22.3	41.8	33.2	51.2	5.4	17.4
	RLDA	25.7	50.4	22.6	41.3	33.7	52.0	5.6	17.2
	ULDA	25.8	49.5	23.4	41.4	29.2	48.2	4.9	16.0
	NLDA	**38.5**	68.3	**26.9**	50.0	**56.9**	78.0	**20.9**	47.7
	Ours	**38.5**	68.3	26.7	50.2	56.8	78.1	20.8	47.7
GOG	PCA+LDA	30.4	55.0	30.3	50.3	22.6	39.6	7.1	17.2
	RLDA	30.9	55.8	30.5	50.3	23.1	39.9	7.1	17.3
	ULDA	31.1	55.3	30.0	50.7	19.1	34.3	6.6	16.6
	NLDA	**47.8**	78.6	**35.6**	57.8	53.2	74.3	**26.7**	54.7
	Ours	**47.8**	78.6	**35.6**	57.8	**53.3**	74.5	**26.7**	54.7
GOG_{RGB}	PCA+LDA	18.3	37.8	24.7	45.6	12.9	25.8	3.8	11.4
	RLDA	**39.4**	72.1	30.0	51.7	**48.4**	70.6	**16.0**	35.8
	ULDA	18.4	36.8	26.5	43.8	7.7	17.2	2.5	9.8
	NLDA	38.6	70.7	**30.3**	52.3	34.5	56.5	10.2	27.1
	Ours	38.6	70.7	**30.3**	52.3	34.6	56.9	10.2	27.1
GOG_{PCA}	PCA+LDA	20.3	40.3	26.5	46.1	39.2	60.9	7.9	19.8
	RLDA	36.5	67.7	**32.0**	55.4	44.0	65.5	12.2	29.0
	ULDA	24.2	44.7	27.1	45.6	22.8	38.9	6.4	15.6
	NLDA	33.7	64.6	25.9	48.8	0.2	1.0	1.0	5.0
	Ours	**37.9**	68.8	30.0	49.9	**57.2**	78.1	**16.0**	39.9

相关的核函数有：

• Gaussian：一个典型的径向基核，对于数据中的噪声具有良好的抗干扰性，其数学表达式为 $k(\boldsymbol{x},\boldsymbol{y})=e^{-\frac{\|x-y\|^2}{2\sigma^2}}$。

• Sigmoid：一个“S”形核函数，目前广泛应用于深度学习中，其数学表达式为 $k(\boldsymbol{x},\boldsymbol{y})=\tanh(\alpha\langle\boldsymbol{x},\boldsymbol{y}\rangle+\beta)$。

• Cauchy：一个来源于 Cauchy 分布的核函数，由于其广泛的定义域，这个核函数可以应用于原始维度很高的数据上，其数学表达式为 $k(\boldsymbol{x},\boldsymbol{y})=\frac{1}{\|x-y\|^2/\sigma+1}$。

• ANOVA：属于径向基核函数一族，其数学表达式为 $k(\boldsymbol{x},\boldsymbol{y})=e^{(-\sigma(x-y)^2)^\gamma}$。

• Multiquadric (MQ)：二次有理核，其数学表达式为 $k(\boldsymbol{x},\boldsymbol{y})=\sqrt{\|\boldsymbol{x}-\boldsymbol{y}\|^2+c^2}$。

• Inverse Multiquadric (IM)：逆多元二次核，其核矩阵满秩，数学表达式为 $k(\boldsymbol{x},\boldsymbol{y})=\frac{1}{\sqrt{\|x-y\|^2+c^2}}$。

对比实验结果如表3.4所示。据此，本文作者有如下观察：1）在大多数情况下，相比于本文作者提出的方法的非核化版本，方法的核化版本都可以实现性能的提高。在 CUHK03(manual) 数据集中，采用 Gaussian 核函数，且是基于多个特征描述符融合的情况，本文作者提出的方法对于 Rank 1 指标实现了最大性能提升，提高了 33.7%。然而，在 VIPeR 和 PRID2011 数据集中，通过核化，方法的性能出现了下降。比如，在 PRID2011 数据集中，采用 Sigmoid 核函数，且基于 GOG 特征描述符，本文作者提出的方法的于 Rank 1 指标下降了 0.7%。这可能是由于不同规模的数据集具有不同的非线性情况。对于大数据集，由于数据集中收集有大量的行人图像样本，因此这类数据集可以较为准确地描述行人重识别问题的非线性，那么对方法进行核化可以实现性能的增强；而对于小数据集，由于数据集中的样本数量较少，因此行人重识别问题的非线性没有得到准确描述，那么对方法进行核化对于性能的提高没有帮助。2）在所有数据集中，对于任何一种特征描述符，本文作者提出的方法的核化版本不论是采用哪种核函数，其性能大致相同。这表明了本文作者提出的方法对于核函数变化的鲁棒性。3）在本文作者提出的方法中，相比于基于单个特征描述符，多个特征描述符融合的情况可以显著地提高性能。具体地，在 VIPeR，PRID2011，CUHK01 和 CUHK03(manual) 数据集上，分别实现了 4.4%，5.0%，8.7% 和 8.0% 的提高。

表 3.4 本文作者提出的方法的非核化版本与核化版本的性能比较。

Table 3.4 Comparison of our methods for non-kernel version and kernel version with different kernel functions.

	Methods		VIPeR		PRID2011		CUHK01		CUHK03(m)	
			r=1	r=5	r=1	r=5	r=1	r=5	r=1	r=5
GOG	Ours(w/o)		47.8	78.6	35.6	57.8	53.3	74.5	26.7	54.7
	Ours(kernel)	Gaussian	49.1	80.6	33.3	57.9	70.8	88.3	59.8	87.7
		Sigmoid	47.2	80.2	34.9	58.4	70.8	88.4	58.0	86.6
		Cauchy	49.2	80.8	32.7	56.7	71.3	88.7	58.8	86.9
		ANOVA	48.6	79.9	34.7	58.0	71.1	88.6	57.8	86.6
		MQ	48.9	80.1	33.4	57.8	71.2	88.6	59.2	87.1
		IM	48.8	80.0	33.3	58.1	71.2	88.8	59.4	86.9
LOMO	Ours(w/o)		38.5	68.3	26.7	50.2	56.8	78.1	20.8	47.7
	Ours(kernel)	Gaussian	41.7	71.2	27.3	50.2	68.0	88.1	48.4	80.4
		Sigmoid	42.5	73.5	27.2	50.4	68.1	88.7	49.3	81.7
		Cauchy	41.4	71.1	26.6	49.7	67.6	87.7	47.9	80.2
		ANOVA	41.5	71.4	27.2	50.3	67.9	87.8	49.3	81.8
		MQ	41.1	70.8	26.9	50.6	66.8	86.3	45.2	77.5
		IM	41.5	71.1	27.2	50.4	67.7	87.5	49.3	80.8
HIPHOP	Ours(w/o)		41.5	69.7	12.5	27.9	60.9	81.1	28.3	52.6
	Ours(kernel)	Gaussian	44.2	73.2	12.3	27.4	65.4	84.9	44.8	75.8
		Sigmoid	44.4	72.9	13.0	29.3	64.9	83.9	43.5	72.8
		Cauchy	44.2	73.3	13.0	28.3	65.3	84.7	43.0	75.7
		ANOVA	44.0	73.4	13.6	28.6	65.0	84.1	46.2	76.8
		MQ	44.1	72.8	12.7	28.5	64.9	84.0	44.2	74.6
		IM	44.2	73.0	12.7	28.5	63.8	85.0	46.2	78.2
Fusion	Ours(w/o)		52.0	82.0	40.6	65.6	55.0	81.6	34.8	70.4
	Ours(kernel)	Gaussian	52.8	82.8	37.8	63.2	79.7	94.7	68.5	92.7
		Sigmoid	44.2	74.6	36.0	59.0	76.1	93.0	57.8	87.1
		Cauchy	52.9	82.0	36.9	62.3	79.8	94.7	64.4	91.0
		ANOVA	50.9	80.4	35.5	60.6	80.0	94.5	66.4	91.7
		MQ	53.6	82.7	38.3	63.1	78.1	93.6	65.1	89.6
		IM	49.3	77.8	32.8	57.6	73.4	91.3	62.1	88.2

表 3.5 与其它行人重识别方法在 VIPeR 数据集上的性能比较结果。最好的结果用粗体显示。

Table 3.5 Comparison with state-of-the-art methods on VIPeR dataset. The best results are shown in boldface.

	Methods	r=1	r=5	r=10
LOMO	ITML [107]	24.7	49.8	63.0
	LMNN [108]	29.4	59.8	73.5
	KISSME [109]	34.8	60.4	77.2
	kCCA [110]	30.2	62.7	76.0
	MFA [111]	38.7	69.2	80.5
	kLFDA [111]	38.6	69.2	80.4
	XQDA [9]	40.0	68.1	80.5
	MLAPG [112]	40.7	–	82.3
	CRAFT [102]	**42.3**	**74.7**	**86.5**
	Ours	41.7	71.2	83.3
	DMLLV [31]	50.4	80.5	88.7
	PML&LSL [113]	46.5	69.3	80.7
	PDC [114]	51.3	74.1	84.2
	MuDeep [64]	43.0	74.4	85.8
	DM^3 [115]	42.7	74.3	85.1
	CSPL [116]	51.3	81.7	90.2
	Ours(Fusion)	**52.8**	**82.8**	**91.3**

3.3.2.3 与其它行人重识别方法的性能比较

在这一节，比较本文作者提出的方法和其它行人重识别方法的性能。由于大多数的基于度量学习的方法采用 LOMO 特征描述符作为行人图像的特征表达，因此为了公平起见，对于与基于度量学习的方法的比较，本文作者采用 LOMO 特征描述符。除此之外的性能比较，本文作者采用多个特征描述符融合的模式（简称为 our(fusion)）进行比较。另外，在所有对比实验中，均采用 Gaussian 核函数核化本文作者提出的方法[2]。

1) VIPeR

对于 VIPeR 数据集，本文作者首先进行与基于度量学习的方法的性能比较。正如表3.5所示，本文作者提出的方法其性能表现排为第二名，略次于性能表现最好的方法 CRAFT[102]，它致力于学习一个视角特定的特征变换。然而，与其

[2]通过 2-折交叉验证，Gaussian 核函数中的参数 σ 设置为，对于 VIPeR 数据集，$\sigma = 1$，对于 PRID2011 数据集，$\sigma = 2$，对于 CUHK01 和 CUHK03 数据集，$\sigma = 0.5$。

表 3.6 与其它行人重识别方法在 PRID2011 数据集上的性能比较结果。最好的结果用粗体显示。

Table 3.6 Comparison with state-of-the-art methods on PRID2011 dataset. The best results are shown in boldface.

	Methods	r=1	r=5	r=10
LOMO	kCCA [110]	14.3	37.4	47.6
	MFA [111]	22.3	45.6	57.2
	kLFDA [111]	22.4	46.5	58.1
	XQDA [9]	26.7	49.9	61.9
	Ours	**27.3**	**50.2**	**62.0**
	Metric Ensembles [117]	17.9	39.0	50.0
	M3TCP [69]	22.0	–	47.0
	CrowdPSE [118]	21.1	46.7	60.0
	DMLLV [31]	27.8	48.4	59.5
	Ours(Fusion)	**37.8**	**63.2**	**73.4**

它不局限于基于度量学习的先进的行人重识别方法相比，本文作者提出的方法在 Rank 1、Rank 5 和 Rank 10 均取得了最高的重识别率。

2) PRID2011

同样地，在 PRID2011 数据集上，本文作者进行与其它基于度量学习的行人重识别方法以及与其它先进的行人重识别方法的性能比较。对比实验结果如表3.6所示。在所有的对比方法中，本文作者提出的方法取得了最高的重识别率。特别地，本文作者提出的方法比 M3TCP [69] 在 Rank 1 上高 15.8%。M3TCP[3]借助深度学习框架学习行人图像的全局特征和局部特征。这表明了，目前基于深度学习的方法在小数据集上性能表现欠佳。

3) CUHK01

相比于以上两个数据集，CUHK01 数据集收集了更多的样本。尽管如此，如图3.1所示，在 CUHK01 数据集中，奇异问题仍然存在。本文作者提出的方法与其它基于度量学习的方法以及与其它先进的重识别方法进行性能比较，结果见表3.7。可以看出，无论是与基于度量学习的方法的比较，还是与先进的方法的比较，本文作者提出的方法在 Rank 1、Rank 5 和 Rank 10 上均取得了最好的结

[3]在 M3TCP 中，需要学习百万以上的参数，调试三个超参数；而在本文作者提出的方法中，对于采用多个特征描述符融合的模式，仅仅需要学习大约六万的参数，调试一个核参数。

表 3.7 与其它行人重识别方法在 CUHK01 数据集上的性能比较结果。最好的结果用粗体显示。

Table 3.7 Comparison with state-of-the-art methods on CUHK01 dataset. The best results are shown in boldface.

	Methods	r=1	r=5	r=10
LOMO	kCCA [110]	56.3	80.7	87.9
	MFA [111]	54.8	80.1	87.3
	kLFDA [111]	54.6	80.5	86.9
	XQDA [9]	63.2	83.9	90.0
	MLAPG [112]	64.2	–	90.8
	CRAFT [102]	65.4	85.3	90.5
	Ours	**68.0**	**88.1**	**93.1**
	Metric Ensembles [117]	53.4	76.4	84.4
	M3TCP [69]	53.7	84.3	91.0
	DMLLV [31]	65.0	85.6	91.1
	PML&LSL [113]	53.5	82.5	91.2
	DM^3 [115]	49.7	77.3	86.1
	CSPL [116]	72.0	88.6	92.8
	SGLE [119]	70.9	89.8	93.5
	Ours(Fusion)	**79.7**	**94.7**	**97.3**

果。特别地，本文作者提出的方法比基于深度学习的方法 M3TCP [69] 在 Rank 1 上高 26%。

4）CUHK03

这是行人重识别最大的数据集之一。然而，从图3.1可以看出，在这个数据集上，奇异问题仍然存在。在 CUHK03(manual) 和 CUHK03(detected) 上，分别比较本文作者提出的方法和其它传统的行人重识别方法的性能，以及和其它基于深度学习的方法的性能。对比实验的结果如表3.8所示。相比传统的行人重识别方法，本文作者提出的方法在 CUHK03(manual) 数据集上取得了最好的结果，在 CUHK03(detected) 数据集上取得了最好的 Rank 10 结果。从整体上看，基于深度学习的方法性能指标优于传统的方法。在基于深度学习的方法中，需要学习大概千万级的参数，从数据当中可以获得更多的信息，因此训练得到的模型具有更好的泛化能力。毫无疑问，这是基于一个前提：模型在大数据集中进行训练。此外，还有多个超参数需要调试。相比于基于深度学习的方法，本文作者提出的方法更简单，需要学习的参数很少，且仅仅只有一个核参数需要调试，同时

表 3.8 与其它行人重识别方法在 CUHK03 数据集上的性能比较结果。最好的结果用粗体显示。

Table 3.8 Comparison with state-of-the-art methods on CUHK03 dataset. The best results are shown in boldface.

	Methods	CUHK03(manual)			CUHK03(detected)		
		r=1	r=5	r=10	r=1	r=5	r=10
Deep learning	IDLA [63]	54.7	87.6	94.0	54.7	86.5	93.9
	EDM [120]	61.3	88.9	96.4	52.1	82.9	91.8
	LSTM S-CNN [121]	–	–	–	57.3	80.1	88.3
	MSCAN [122]	74.2	94.3	97.5	68.0	91.0	95.4
	DSPL [123]	73.16	92.26	96.54	–	–	–
	MuDeep [64]	**76.9**	**96.1**	**98.4**	**75.6**	**94.4**	**97.5**
Traditional	kLFDA [111]	48.2	59.3	66.4	–	–	–
	XQDA [9]	52.2	82.2	92.1	46.3	78.9	88.6
	MLAPG [112]	58.0	87.1	94.7	51.2	83.6	92.1
	SSSVM [124]	57.0	84.4	90.9	–	–	–
	GOG [2]	67.3	91.0	96.0	**65.5**	**88.4**	93.7
	Ours	**68.5**	**92.7**	**97.0**	60.8	86.6	**94.2**

在 CUHK03(manual) 数据集上，Rank 10 达到了 97%，仅比最好的结果低 1.4%。

3.3.2.4 与方法快速版本的性能比较

在章节3.2.5中，本文作者介绍了提出的方法的快速计算版本。在这一小节，本文作者详细分析其算法的计算复杂度、性能表现和运行时间。表3.9展示了本文作者提出的算法的非快速版本和快速版本之间的计算复杂度的对比结果。略去低阶项，本文作者提出的算法的非快速版和快速版本的计算复杂度分别为 $18mn^2 + n^3$ 和 $5mn^2 + 6n^3$。快速版本的计算复杂度比非快速版本降低了一半。表3.10展示了性能表现和运行时间的对比结果。该实验的结果是基于 GOG 作为特征描述符，其运行实验的计算机的配置是：载有一个 4GHzx8 的内核和 16GB 的随机存取存储器（RAM）。从表中可以看出，快速版本的性能表现与非快速版本大体一致，同时其运行时间是非快速版本的一半。

表 3.9 本文作者提出的方法的非快速版本与快速版本的计算复杂度比较结果。

Table 3.9 Comparison with fast version for our method in computational complexity.

	Algorithm1	Algorithm2
Step1	$O(mn)$	$4mn^2 - \frac{4}{3}n^3$
Step2	$14mn^2 - 2n^3$	$O(n^2)$
Step3	$2mr_tk + 14r_tk^2 - 2k^3$	$2nr_t^2 + \frac{2}{3}r_t^3$
Step4	$2mr_t^2$	$2r_tk(n-k) + 4kr_t^2 + \frac{2}{3}k^3 - 2r_tk^2$
Step5	$4mr_b^2 - \frac{4}{3}r_b^3$	$2(m+r_t)nr_b$
Overall	$18mn^2 + n^3$	$5mn^2 + 6n^3$

表 3.10 本文作者提出的方法的非快速版本与快速版本的性能表现和运行时间比较结果。

Table 3.10 Comparison with fast version for our method in accuracy and run time.

Datasets	**Methods**	r=1	r=5	r=10	Time(s)
VIPeR	Ours	47.8	78.6	88.4	2.1952
	Ours_{fast}	47.9	78.6	88.3	1.2910
PRID2011	Ours	35.6	57.8	69.0	0.4919
	Ours_{fast}	35.4	57.7	69.2	0.2484
CUHK01	Ours	53.3	74.5	81.8	13.4156
	Ours_{fast}	53.3	74.5	81.8	7.8118
CUHK03(manual)	Ours	26.7	54.7	69.8	25.4500
	Ours_{fast}	26.6	54.8	69.6	12.7273

3.4 本章小结

本章针对行人重识别中的奇异问题，提出了一个基于伪逆 LDA 的行人重识别方法，该方法基于三个散度矩阵的同时对角化计算一个最优的正交变换。此外，考虑到行人重识别的非线性，本文作者提出在核空间学习这个正交变换，给出了方法的非线性模型，进一步提高了方法的性能；同时，本文作者也发展了方法的快速版本，进一步提高了方法的效率。对于模型的求解，本文作者给出了闭式解，并且在提出的方法中没有超参数需要调试，这些都体现了本文作者提出的方法的简单性和高效性。在 4 个行人重识别数据集上的对比实验展示了本文作者提出的方法优于其它先进的行人重识别方法。同时，本文作者也进行了与方法的核化版本和快速版本的实验对比，证明了这些版本的优越性。

奇异问题普遍存在于图像分类和检索任务中。为了更精确的表达图像的内容，从图像中提取的特征向量往往维度很高，同时由于对图像样本进行标注是一件非常耗时的工作，导致训练集的样本数量有限，通常远远小于特征向量维数。这种情况与行人重识别的情况非常相似，因此本文作者提出的方法对于解决图像分类和检索任务中奇异问题可以带来一些启示。在提出的方法中，借助于标签样本，可以学习得到最优的正交变换。然而，在某些实际情况中，标签样本数量非常有限，未标签样本数量庞大。鉴于此，在未来的工作中，本文作者将致力于通过半监督的学习方式解决行人重识别的奇异问题。比如，可以借助于标签传递技术，将标签样本的标签信息传递给未标签样本，发展一个半监督学习的方法。

第 4 章　基于区域特定和排序损失函数的行人重识别方法

现阶段，行人重识别算法的输入通常是经过手动裁剪或自动行人检测器得到的行人图像。根据这些输入图像，研究人员设计算法进行图像的特征提取，其目的是提取一个对行人外貌特征跨摄像头变化鲁棒的特征表达。对于每一张图像，通过特征提取步骤得到一个特征向量，之后将图像的特征向量作为度量学习步骤的输入。在度量学习步骤中，研究人员致力于计算特征向量之间的距离。采用欧式距离是最直接的一种方式，然而对于通过非监督方法提取到的特征向量，它的可区分性往往是非常弱的，采用欧式距离度量特征向量之间的距离往往无法得到令人满意的结果。因此研究人员希望利用带标签信息的样本，进行监督学习，得到一个合适的度量，即学习一个度量，使得正样本对的相似值大于负样本对的相似值。

度量学习大致可以分为两类，基于闭式解和基于迭代学习。在基于闭式解的度量学习中，研究人员往往利用线性判别分析（Linear Discriminant Analysis，LDA）等相关技术，最小化类内方差同时最大化类间方差，得到一个闭式解，从而求得最优的度量。本文作者提出了相关的算法：基于正交线性判别分析的行人重识别算法，并在第3章对其进行了详细介绍。在基于迭代学习的度量学习中，研究人员往往构造一个目标函数，然后通过迭代优化算法，如梯度下降法，优化目标函数，使其学习一个最优的度量，满足正样本对的相似值大于负样本对的相似值。本文作者提出了相关的算法：基于区域特定和排序损失函数的行人重识别方法，并在本章进行详细介绍。

4.1　基于区域特定的行人重识别方法

4.1.1　引言

对于行人重识别任务，其第一个步骤往往是特征提取，然后将提取的特征作为第二个步骤-度量学习的输入，通过度量学习，求解得到一个最优的度量，根据此度量，就可以计算行人图像之间的相似值了。

为了提取行人特征，使得对不同摄像机下行人的外貌变化有一定的鲁棒性，研究人员设计了各种各样的特征提取方法 [102, 125]。这些特征的维数一般都

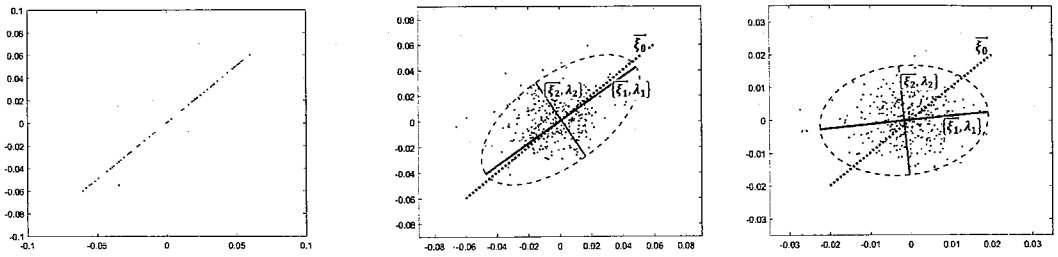
非常大，无法直接用于度量学习，因此大部分的行人重识别工作都会在度量学习之前对提取到的特征进行降维。但是由于低质量的行人图像和不完美的特征提取步骤，提取得到的特征往往良莠不齐。为此本文作者提出基于主成分分析（Principal Component Analysis（PCA））的特征选择方法。对每一维特征，进行PCA 操作，根据其得到的特征值和特征向量，判断该维度的特征是否具有良好的可区分性，将具有较好区分性的特征保留，其余删除，从而实现降维和改善特征质量的目的。

为了更好地描述行人的空间信息，行人图像经常被划分为多个区域，用于提取行人特征，即：行人的特征向量是区域特征连接的。这些考虑到空间结构的特征描述常作为度量学习的输入，但是现存的大部分度量学习方法都是将这些特征作为整体对待，忽视了其不同区域之间的特征差异性。本文作者提出了一种新的区域特定的度量学习方法，对于不同区域的特征描述，考虑到其分布差异性，学习区域特定的度量函数。基于这样的度量函数，得到的新的特征空间中的特征，保持了其区域特征连接的性质。此外，相比于其它度量学习方法，基于区域特定的行人重识别方法需要学习的参数减少了，一定程度地遏制了过拟合现象的发生。

4.1.2 算法模型

对于两个分别来自于不同摄像机的行人样本 p_i 和 g_j（$i = 1,\ldots,N$，$j = 1,\ldots,N$），$i = j$ 时表明 p_i 和 g_j 是同一个行人（称之为正样本对），否则 p_i 和 g_j 是不同的行人（称之为负样本对）。通过特征提取，可以得到样本的特征向量 $\mathbf{x}_i^p$ 和 $\mathbf{x}_j^g$，特征向量的维数 d_x 通常非常大，因此本文作者首先对样本的特征向量进行 PCA 降维。经过 PCA 降维后的 d_y（$d_y < d_x$）维特征向量分别表示为 $\mathbf{y}_i^p$ 和 $\mathbf{y}_j^g$。$\mathbf{y}_i^p$ 和 $\mathbf{y}_j^g$ 的每一维特征具有不同的可区分性，为了提高特征的质量，本文作者选择可区分性最好的前 d_z（$d_z < d_y$）维特征作为新的特征表示。

如果特征向量 $\mathbf{y}$ 满足 $\mathbf{y}_i^p = \mathbf{y}_i^g$ $(i = 1,\ldots,N)$ 和 $\mathbf{y}_i^p \neq \mathbf{y}_j^g$ $(i \neq j)$，本文作者称其为一个完美的特征向量。完美的特征向量的任意第 k 维，$\{(\mathbf{y}_i^p(k), \mathbf{y}_i^g(k))\}_{i=1}^N$ 的散度图分布应该是一条斜率为 45° 的直线，如图4.1(a) 所示。然而在实际中，由于低质量的行人图像和不完美的特征提取步骤等干扰存在，提取到的特征不可能是这样完美的特征，$\{(\mathbf{y}_i^p(k), \mathbf{y}_i^g(k))\}_{i=1}^N$ 的散度图分布往往是一个椭圆点集，如图4.1(b) 所示。椭圆点集越扁平，并且其主轴离 45° 斜率的直线越近，这样的特



(a) Ideally discriminative feature　　(b) The extracted features

图 4.1 （a）完美的特征向量的分布图；（b）VIPeR 数据集 [1] 中样本的 GOG 特征向量 [2] 的分布图。（b）的左图和右图分别是具有良好可区分性的特征和较差可区分性的特征的分布图。紫色圆点的 x 轴和 y 轴分别代表一个样本在两个不同摄像机下的特征。

Figure 4.1 The distributions of (a) the ideally discriminative feature and (b) GOG descriptor [2] on VIPeR dataset [1] in one dimension. In (b), the distribution at left is about the feature with good discriminative power, the one at right is about the feature with poorly discriminative power. The x-axis and y-axis of purple dots represent the features of the same people from two different cameras, respectively.

征越接近完美特征。根据此观察，选择特征。

具体地，对于第 k 维特征，本文作者应用 PCA 到 $P = \{(\mathbf{y}_i^p(k), \mathbf{y}_i^g(k))\}_{i=1}^N$，计算得到特征向量 ξ_1、ξ_2，和其相对应的特征值 λ_1、λ_2 $(\lambda_1 > \lambda_2)$。特征向量和特征值分别反应了椭圆点集的轴和长度。则椭圆点集的扁平程度通过以下方程度量 [126]：

$$f_k(\lambda_1, \lambda_2) = \frac{\lambda_1 - \lambda_2}{\lambda_1}, \tag{4.1}$$

$f_k \in [0, 1]$，当 $f = 0$ 时，表示椭圆点集是一个圆形。在图4.1(b) 中，左图和右图的 f 分别为 0.37 和 0.15。主轴与斜率为 45° 直线的角度通过以下方程度量：

$$g_k(\xi_0, \xi_1) = \arccos\frac{\langle \xi_0, \xi_1 \rangle}{\|\xi_0\| \, \|\xi_1\|}, \tag{4.2}$$

g_k 的单位为弧度。在图4.1(b) 中，左图和右图的 g_k 分别为 0.07 和 0.66。

现在，定义第 k 维特征的可区分性：

$$\rho_k = \frac{f_k(\lambda_1, \lambda_2)}{g_k(\xi_0, \xi_1) + \varepsilon}, \tag{4.3}$$

ε 是一个非常小的值，其作用是避免方程的分母等于 0。

如果第 k 维特征的可区分性度量值 $\rho_k > \tau$，保留这维特征，否则删除。经过

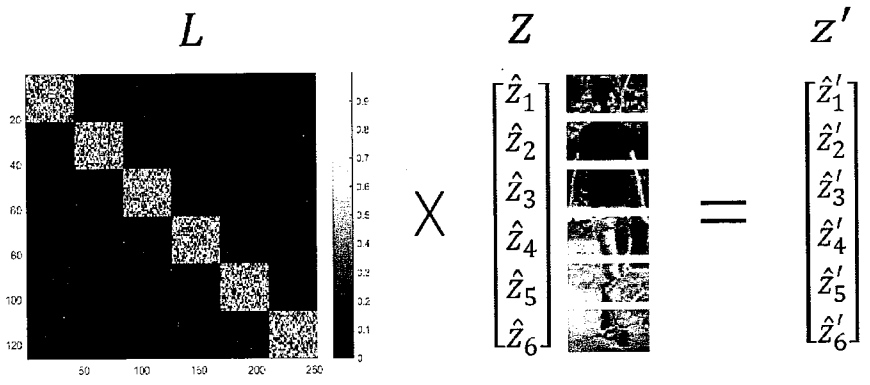


图 4.2 区域特定的度量学习模型的图示。为了方便起见，假设提取特征时，行人图像被划分为了 6 个水平条纹区域。

Figure 4.2 The illustration of the proposed region-specific metric learning model. We assume that the person image is divided into 6 horizontal strips for convenience.

筛选之后的特征，用 $\mathbf{z}$ 表示。对于参数 τ，本文作者在实验部分4.1.3章节将会进行详细讨论。

对特征向量进行预处理之后，本文作者学习一个区域特定的度量函数。让 $D(\mathbf{z}_i^p, \mathbf{z}_j^g) = \left\| L(\mathbf{z}_i^p - \mathbf{z}_j^g) \right\|_2^2$ 表示特征向量 $\mathbf{z}_i^p$ 和 $\mathbf{z}_j^g$ 之间的距离，本文作者致力于学习一个最优的映射矩阵 L，满足

$$D(\mathbf{z}_i^p, \mathbf{z}_j^g) \begin{cases} < C & i = j \\ > C & i \neq j \end{cases} \quad i, j = 1, \ldots, N. \tag{4.4}$$

C 是一个超平面参数，在实验中设置为 $C = 1$。

具体地，通过最小化下面的目标函数学习最优的 L：

$$\min_{L} \sum_{i=1,j=1}^{N} l_\beta(k_{ij}(D(\mathbf{z}_i^p, \mathbf{z}_j^g) - 1)) + \alpha \|L\|_F^2, \tag{4.5}$$

当 $(\mathbf{z}_i^p, \mathbf{z}_i^g)$ 是正样本对时，$k_{ij} = 1$，否则 $k_{ij} = -1$。α 是一个正则化参数。$l_\beta(x) = \frac{1}{\beta}\log(1 + e^{\beta x})$ 是合页损失函数（hinge loss function）$h(x) = \max\{x, 0\}$ 的光滑近似函数[1]。在实验中，本文作者设置 $\alpha = 0.01$，$\beta = 3$。

为了处理行人图像中的空间对不准问题以及更好地描述行人图像的空间信息，特征通常是基于几个不同区域提取的，即 $\mathbf{z} = [\hat{\mathbf{z}}_1 \ldots \hat{\mathbf{z}}_S]$，$\hat{\mathbf{z}}_s$ $(i = s, \ldots, S)$ 是第 s 个区域提取得到的特征，本文作者希望学习一个最优的映射矩阵 L，对

[1] $\lim_{\beta \to \infty} l_\beta(x) = h(x)$

于每一个区域有其特定的子映射，即满足图4.2所示。重新表达映射矩阵 $L=\begin{bmatrix} L_{11} & \cdots & L_{1S} \\ \vdots & \ddots & \vdots \\ L_{S1} & \cdots & L_{SS} \end{bmatrix}$，希望其非对角线的子块为0，即满足 $L=diag(L_{11},L_{22},\ldots,L_{SS})$。
因此构建如下优化模型：

$$\begin{aligned} &\min_{L} \sum_{i=1,j=1}^{N} l_{\beta}(k_{ij}(D(\mathbf{z}_i^p,\mathbf{z}_j^g)-1))+\alpha\|L\|_F^2 \\ &s.t \quad L_{ij}=\mathbf{O},\ i\neq j,\ i,j=1,\ldots,S \end{aligned} \tag{4.6}$$

算法 3 RSL+Pre 行人重识别方法

输入：

1: 训练集样本特征向量和标签信息 $\{(\mathbf{x}_i^p,\mathbf{x}_j^g),y_{ij}\}_{i,j=1}^{N}$。

2: 测试集样本特征向量 $\{(\mathbf{u}_i^p,\mathbf{u}_j^g)\}_{i,j=1}^{N_t}$。

输出：$D(\mathbf{u}_i^p,\mathbf{u}_j^g)$。

步骤：

1: 初始化 $s=1$。

2: 对第 s 个区域提取得到的特征向量进行 PCA 降维。

3: 初始化 $k=1$。

4: 应用 PCA 到 $P=\{(\hat{\mathbf{y}}_s)_i^p(k),(\hat{\mathbf{y}}_s)_j^g(k)\}_{i,j=1}^{N}$。

5: 通过方程(4.3)计算 ρ_k。

6: 如果 $\rho_k<\tau$，移除第 k 维特征。

7: 判断 $k\leq d_y^s$，若是，$k=k+1$ 且返回第 4 步继续循环，否则跳出循环，执行下一步。

8: 输出训练集样本新特征 $\{((\hat{\mathbf{z}}_s)_i^p,(\hat{\mathbf{z}}_s)_j^g),y_{ij}\}_{i,j=1}^{N}$ 和测试集样本新特征 $\{(\hat{\mathbf{v}}_s)_i^p,(\hat{\mathbf{v}}_s)_j^g\}_{i,j=1}^{N}$。

9: $\{(\hat{\mathbf{z}}_s)_i^p,(\hat{\mathbf{z}}_s)_j^g\}_{i,j=1}^{N}$ 作为输入，通过方程(4.5)计算 L_{ss}。

10: 判断 $s\leq S$，若是，$s=s+1$ 且返回第 2 步继续循环，否则跳出循环。

11: 计算 $L=diag(L_{11},L_{22},\ldots,L_{SS})$。

12: 计算 $D(\mathbf{u}_i^p,\mathbf{u}_j^g)=\left\|L(\mathbf{v}_i^p-\mathbf{v}_j^g)\right\|_2^2$。

返回：$D(\mathbf{u}_i^p,\mathbf{u}_j^g)$;

通过以上建模，不同区域的特征学习不同的度量映射，它们是彼此独立的。

通过度量映射之后，新的特征 $\mathbf{z}'$ 也是区域连接的，保留了原始特征的结构。此外，映射矩阵 L 需要学习的参数也大大的减少了，一定程度地抑制了模型的过拟合问题。

在实验中，由于方程(4.6)的约束是非凸的，因此本文作者分别对划分后的每个图像区域进行最小化模型(4.5)，同时优化，求得块对角的映射矩阵 L。此外，考虑到行人重识别中样本图像的非线性问题，本文作者对模型进行了核化。具体地，本文作者采用高斯核函数 $K(x,y)=e^{-\frac{\|x-y\|^2}{2\sigma^2}}$。

概括起来，本文作者提出的方法是一个区域特定的度量学习方法。本人先对输入特征进行了预处理操作，具体地，通过移除可区分性差的特征实现特征降维及改善特征质量的目的，然后将新的特征向量作为后续的度量学习步骤的输入，完成区域特定的度量学习。在接下来的内容里，本文作者将提出的方法简称为RSL+Pre。算法3总结了 RSL+Pre 行人重识别方法。

4.1.3 实验与分析

4.1.3.1 数据集与实验设置

本文作者在 3 个公开的行人重识别数据集 VIPeR [1]，PRID450S [127] 和 QMUL GRID [128] 上评估提出的 RSL+Pre 行人重识别方法的性能。VIPeR 和 PRID450S 数据集中分别有 632 个行人样本图像和 450 个行人样本图像，这些样本图像是由两个不存在重叠拍摄区域的户外摄像机捕捉拍摄到的，对于每个行人，不同的摄像机分别捕捉拍摄到一张图像。GRID 数据集中的样本图像是由 8 个不存在重叠拍摄区域的摄像机捕捉拍摄到的。该数据集中有 250 个行人样本，由其中 2 个不同的摄像机分别捕捉拍摄到一张图像，剩余的 775 个样本仅仅由其中一个摄像机捕捉拍摄到一张图像，这 775 个样本用来扩展候选人集合（gallery set）样本数量。

对于行人特征的提取，本文参考 [32]。把行人图像划分为 14 个不重叠的区域。对于每个区域，本文作者分别计算 RGB、HSV 和 YCrCb 颜色空间的颜色直方图，以及基于局部二值模式（Local Binary Patterns，LBP）的纹理直方图，并将其归一化为单位 L1 范数。然后连接所有这些直方图，形成最终的 6,020 维特征向量。该特征向量作为 RSL+Pre 方法的输入。此外，本文作者也采用 GOG 特征描述符 [2] 作为行人图像的特征表达。行人图像被划分为 7 个水平条纹区域，对于每个区域，GOG 特征描述符建模像素特征的均值和协方差信息，其中像素

表 4.1 与最先进的行人重识别方法在 VIPeR、PRID450S 和 GRID 数据集上的性能比较。最好的结果和次好的结果（%）分别用红色和蓝色进行了标注。

Table 4.1 Performance comparison with the state-of-the-art methods on VIPeR, PRID450S and GRID datasets. The best and second best results (%) are respectively shown in red and blue.

	VIPeR		PRID450S		GRID	
Methods	r=1	r=10	r=1	r=10	r=1	r=10
CSL [129]	34.8	82.3	44.4	82.2	-	-
LOMO+XQDA [9]	40.0	80.5	61.4	91.0	16.6	41.8
Kernel HPCA [130]	39.4	85.1	52.2	92.8	-	-
LSSCDL [124]	42.7	84.3	60.5	88.6	22.4	51.3
SCSP [30]	53.5	91.5	58.8	91.3	24.2	54.1
OL-MANS [131]	45.0	85.0	-	-	30.2	49.2
PAR [125]	48.7	85.1	-	-	-	-
LML [31]	50.4	88.7	64.5	92.1	19.8	46.5
Spindle [25]	53.8	83.2	67.0	89.0	-	-
RSL+Pre	31.8	75.3	45.6	83.2	13.8	43.2
RSL+Pre†	43.3	85.7	55.6	89.7	18.1	49.6
RSL+Pre‡	49.8	89.3	68.1	94.0	25.1	57.2

特征包括颜色、纹理和像素位置信息。GOG 特征描述符有 27,622 维。在本文作者提出的方法中，对于采用 GOG 特征描述符作为行人图像的特征表达，简称为 RSL+Pre†。

在实验中，本文作者把所有图像归一化为 128×48 的尺寸。与行人重识别常用的性能评测指标一致，本文作者采用累计匹配特性曲线（Cumulative Matching Characteristics，CMC）评测提出的 RSL+Pre 方法的性能。本文作者重复进行 10 次训练集和测试集的随机划分，然后进行实验，相应地，得到 10 个 CMC 结果。本文作者取其平均值，并展示平均 CMC 曲线的 Rank 1，Rank 10 和 Rank 20 的结果。对于颜色 + 纹理直方图描述符，阈值参数 τ 设置为 0.1，对于 GOG 特征描述符，通过交叉验证获得 τ 的取值。在4.1.3.3小节中，本文作者将会详细讨论参数对于性能的影响。

4.1.3.2 性能比较

在所有这 3 个数据集上，本文作者提出的方法与最先进的行人重识别方法的性能进行比较。在这些比较的方法中，CSL [129]、LOMO+XQDA [9]、Kernel HPCA [130]、LSSCDL [124]、SCSP [30]、OL-MANS [131] 和 LML [31] 是基于度量学习的行人重识别方法，PAR [125] 和 Spindle [25] 采用深度学习技术提取特征。表4.1展示了比较结果。表格中，对于提出的方法的变体 RSL+Pre‡，本文作者采用 GOG 作为特征描述符，用零空间线性判别分析（Null Space Linear Discriminant Analysis，NLDA）代替方程(4.5)的模型学习映射矩阵 L。

从表4.1中，可以看出，RSL+Pre‡ 在性能方面优于大多数方法。在 VIPeR 数据集上，SCSP 在 Rank 10 取得了最高的重识别率，但是在其它 2 个数据集上 RSL+Pre‡ 要优于 SCSP。在 GRID 数据集上，OL-MANS 比 RSL+Pre‡ 在 Rank 1 上高 5.1%，除此之外，RSL+Pre‡ 比 OL-MANS 具有绝对的性能优势。相似地，Spindle 在 VIPeR 数据集的 Rank 1 上实现了最高的重识别率，但是除此之外，RSL+Pre‡ 具有更高的重识别率。另外，RSL+Pre 和 RSL+Pre† 相比其它方法，在性能方面的竞争力较弱，这主要是因为在 RSL+Pre 和 RSL+Pre† 中，本文作者采用的特征描述符和度量学习基础模型的性能表现较弱。通过采用可区分性更好的特征描述符 GOG 和性能表现更好的度量学习基础模型 NLDA，RSL+Pre‡ 获得了更好的性能。这也从另一个角度证明了本文作者提出的区域特定的度量学习的有效性。

4.1.3.3 讨论

1）特征预处理的性能表现

通过本文作者提出的特征预处理步骤，特征的可区分性得到了提高，同时降低了特征的维度。本文作者对提出的 RSL+Pre† 方法和去掉特征预处理步骤的方法（简称为 RSL†）进行性能比较。表4.2展示了结果。实验结果显示，对特征进行预处理，既降低了特征维数，同时也提高了行人重识别率。

2）参数分析

对于特征预处理步骤，需要选择一个合适的 τ。τ 取值过大，就会导致更多有用的特征被去除；τ 取值过小，导致特征预处理步骤可能对于提高重识别率没有太多帮助。本文作者选取 $\tau = 0.05, 0.1, 0.5, 0.9, 1.3$ 进行实验，如表4.3所示，当

表 4.2 基于不同的特征预处理的方法在 VIPeR 数据集上的性能比较。

Table 4.2 Performance comparison with the methods based on different variants of feature pre-processing on VIPeR dataset.

	r=1	r=10	r=20	dimension
RSL†	44.5	86.1	92.7	4417
RSL+Pre†	44.5	87.0	93.0	3953

表 4.3 基于不同的 τ 值，提出的 RSL+Pre 方法在 VIPeR 数据集上的性能表现。

Table 4.3 Performance comparison with the proposed RSL+Pre methods based on different threshold τ on VIPeR dataset.

τ	r=1	r=10	r=20
0.05	44.5	86.9	92.8
0.1	44.5	87.0	93.0
0.5	42.2	85.7	92.9
0.9	39.6	83.5	91.9
1.3	38.7	81.5	90.5

$\tau = 0.1$ 时，RSL+Pre† 取得最好的结果。

3）区域特定的度量学习与区域通用的度量学习的比较

如前所述，对于不同的区域，特征的分布可能不同，因此本文作者提出一个区域特定的度量学习方法，提高行人重识别性能。现在，本文作者验证区域特定的度量学习的有效性。本文作者提出的方法与区域通用的度量学习方法（简称为 RGL+Pre†）进行比较，实验结果如表4.4所示。相比区域通用的度量学习方法 RGL+Pre†，通过区域特定的度量学习方法 RSL+Pre†，在 Rank 1 上的性能提高了 17.5%。

4）对于区域块的数量的选择

在模型(4.6)中，映射矩阵 L 中非 0 子矩阵块的数量由行人图像被划分的区域数量 S 决定。如果在提取特征时行人图像被划分为了 S 个区域，则在 L 中，需要学习 S 个子映射矩阵 $\{L_{ss}\}_{s=1}^{S}$。子映射矩阵也可以成为与映射矩阵 L 相似的块对角的结构，也就是 $L_{ss} = diag(\{L_{ss}\}_{11}, \{L_{ss}\}_{22}, \ldots, \{L_{ss}\}_{pp})$。本文作者选取 $p = 2, 4$ 进行实验，采用 GOG 特征描述符，该描述符是基于行人图像被划分为 7 个区域提取特征得到的。因此 $p = 2$ 和 $p = 4$ 分别意味着 $S = 14$（简称为

表 4.4 对于学习映射矩阵 L，基于不同数量的区域块，度量学习方法在 VIPeR 数据集上的性能表现。

Table 4.4 Performance comparison with the metric learning methods based on different number of blocks learned in projection matrix L on VIPeR dataset.

	r=1	r=10	r=20
RGL+Pre†	27.0	74.5	86.0
RSL+Pre†	44.5	87.0	93.0
RSL+Pre†_14	41.5	85.5	92.8
RSL+Pre†_28	34.9	79.2	89.2

RSL+Pre†_14）和 $S = 28$（RSL+Pre†_28）。表4.4展示了实验结果。可以看出，随着映射矩阵 L 中块的数量的增加，性能并没有得到提高。这是因为 $S = 14$ 和 $S = 28$ 没有遵循真实的特征描述符结构，同一区域里，特征的分布应该是一致的。

4.2 基于排序损失函数的行人重识别方法

4.2.1 引言

现有的基于度量学习的行人重识别方法，其目标损失函数主要分为三类：二分类损失（binary classification loss）[32, 33]，三元组损失（triplet loss）[34, 113]和四元组损失（quadruplet loss）[36]。在二分类损失函数中，其目标是学习最优的特征映射函数，使得正样本对的距离小于一个预设的阈值，负样本对的距离大于该阈值；对于三元组损失函数，学习最优特征映射函数，希望对于每一个询问行人（query person），其正样本对的距离小于其对应的负样本对的距离；而四元组损失函数，是在三元组损失函数的基础上，提出了更强的约束，希望对于每一个询问行人，其正样本对的距离小于数据集中所有负样本对的距离。尽管这些损失函数最终都可以实现基于度量学习的行人重识别方法的训练目标，即，正样本对的距离比负样本对的距离小，但是他们都是通过一种相对间接的方法建模，针对全样本集中的部分集合建模，并且这些方法都需要参数的设置。这会导致模型的鲁棒性差，可操作性不强。

为了解决上述问题，本文作者提出了一种鲁棒性强且高效的行人重识别方法。考虑到基于度量学习的行人重识别方法的训练目标-正样本对的距离比负样

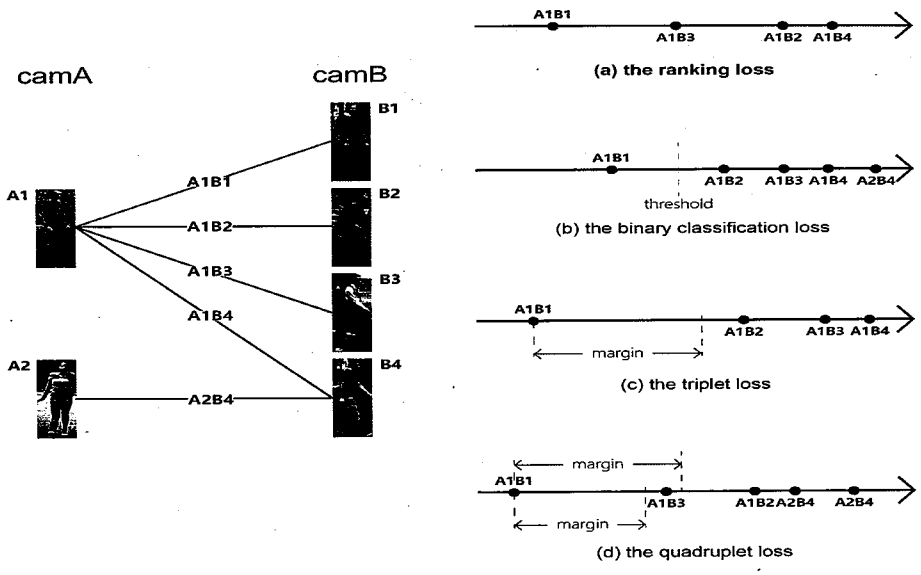


图 4.3 损失函数中正样本对距离和负样本对距离之间的关系。蓝色的线和点代表正样本对，红色的代表负样本对。

Figure 4.3 The relationships between positive pairs'distances and negative pairs'distances for loss funtions. The blue line and point represent the positive pair and the red ones represent the negative pair.

本对的距离小，本文作者对其目标直接翻译，构建一个新的排序损失函数用于行人重识别：最小化正样本对和最小样本对（即：所有样本对中对距离最小的样本对）之间的距离。本文作者提出的模型在具体求解过程中，包括两个贡献：1）在损失函数中，使用 p 范数作为 min 函数的光滑近似，参数 p 控制了正样本对和负样本对之间的距离，有助于模型的泛化能力；2）在优化过程中，仅仅采用了一小部分的样本对参与计算，在准确率没有损失的情况下，有效地提高了算法的效率。相比于其它基于度量学习的行人重识别方法，该方法能更好地解决行人重识别问题，得到更高的识别率。图4.3图示了二分类损失函数、三元组损失函数、四元祖损失函数以及本文作者提出的排序损失函数的正样本对距离和负样本对距离之间的关系。

4.2.2 算法模型

给定一个拍摄于某一摄像机下的行人图像集 $V=\{v_1,...,v_N\}$，对于这个集合中的行人，本文作者称之为询问行人；同时给定一个拍摄于不同摄像机下的行人图像集 $C=\{c_1,...,c_M\}$，对于这个集合的行人，本文作者称之为候选集行人。标注 V 和 C 集合中行人图像之间的关系（正样本对或负样本对），获得标签信

息，得到训练集。那么，对于每个询问行人，在集合 C 中一定存在其正样本集 $C^{i+}=\{c_j^{i+}\in C\,|j=1,2,\ldots,M^{i+}\}$ 和负样本集 $C^{i-}=\{c_j^{i-}\in C\,|j=1,2,\ldots,M^{i-}\}$。

提取行人图像的特征后，本文作者建立排序损失函数模型。基于度量学习的行人重识别的目标是，希望学习一个最优的特征映射函数，使得训练集中的正样本对的距离小于所有负样本对的距离。基于此目标，本文作者将其直接翻译成数学语言，构建模型：

$$\min_L f(L)=\sum_{i=1}^{N}\sum_{c_j^{i+}\in C^{i+}}(d_{v_i,c_j^{i+}}-\min_{c_m\in\{c_j^{i+}\}\cup C^{i-}}d_{v_i,c_m}), \tag{4.7}$$

其中，$d_{v,c}$ 是行人 v 和 c 的距离度量函数，

$$d_{v,c}=\left\|L(x_v-x_c)\right\|^2, \tag{4.8}$$

x_v 和 x_c 是行人 v 和 c 的特征向量。

从方程(4.7)中，可以看到，构建的模型由于 min 函数的存在，导致目标函数不连续且不可导，从而影响模型的求解。为此，本文作者对模型进行优化，具体地，本文作者采用 p 范数作为 min 函数的光滑近似：

$$(\sum_{c_m\in\{c_j^{i+}\}\cup C^{i-}}(d_{v_i,c_m})^p)^{\frac{1}{p}}\approx\min_{c_m\in\{c_j^{i+}\}\cup C^{i-}}d_{v_i,c_m}, \tag{4.9}$$

那么，优化后的排序损失函数模型计算公式为：

$$\min_L f(L)=\sum_{i=1}^{N}\sum_{c_j^{i+}\in C^{i+}}(d_{v_i,c_j^{i+}}-(\sum_{c_m\in\{c_j^{i+}\}\cup C^{i-}}(d_{v_i,c_m})^p)^{\frac{1}{p}}). \tag{4.10}$$

优化后的模型，不仅继承了原模型的优点，同时获得了良好的求解性质。其中的参数 p 控制了正样本对和负样本对之间的距离，有效地控制了模型的鲁棒性。接下来本文作者详细解释为什么模型(4.10)具有以上这些优势。为此，本文作者展开方程(4.9)：

$$\begin{aligned}(\sum_{c_m\in\{c_j^{i+}\}\cup C^{i-}}(d_{v_i,c_m})^p)^{\frac{1}{p}}&=[(d_{v_i,c_j^{i+}})^p+(d_{v_i,c_1^{i-}})^p+\ldots+(d_{v_i,c_{M^{i-}}^{i-}})^p]^{\frac{1}{p}}\\&=[(d_{v_i,c_j^{i+}})^p(1+(\frac{d_{v_i,c_1^{i-}}}{d_{v_i,c_j^{i+}}})^p+\ldots+(\frac{d_{v_i,c_{M^{i-}}^{i-}}}{d_{v_i,c_j^{i+}}})^p]^{\frac{1}{p}}.\end{aligned} \tag{4.11}$$

把方程(4.11)代入方程(4.10)，本文作者可以看到当且仅当对于每个询问行人 v_i 及相应的正样本 c_j^{i+}，$(\frac{d_{v_i,c_t^{i-}}}{d_{v_i,c_j^{i+}}})^p=0\ (\forall t=1,\ldots,M^{i-})$ 时，目标函数 $f(L)$ 达到其最小值。因为本文作者想要实现的是对于每个询问行人 v_i 及相应的正样本 c_j^{i+}，

$d_{v_i,c_j^{i+}} < d_{v_i,c_t^{i-}}$ $(\forall t = 1,\ldots,M^{i-})$，这意味着对于所有项需要满足 $\frac{d_{v_i,c_t^{i-}}}{d_{q_i,c_j^{i+}}} > 1$，而只有当 $p < 0$ 时，才有可能实现。由此可见，只要选择一个合适的 p，具有光滑和连续差分良好性质的模型(4.10)可以实现与模型(4.7)一样的作用。

基于这个基本的约束 $p < 0$，有两种情况均可以实现 $(\frac{d_{v_i,c_t^{i-}}}{d_{v_i,c_j^{i+}}})^p = 0$:

1) 当 $p \to -\infty$，$\frac{d_{v_i,c_t^{i-}}}{d_{v_i,c_j^{i+}}}$ 只需要略大于 1，此时意味着正样本对与负样本对距离较小；

2) 当 $p \to 0$，$\frac{d_{v_i,c_t^{i-}}}{d_{v_i,c_j^{i+}}}$ 需要远远大于 1，此时意味着正样本对与负样本对距离较大。

由此可见，当优化模型(4.10)的时候，p 的取值控制着正样本对和负样本对之间的距离，本文作者可以根据对此距离的需求，灵活地设置 p 的值。

本文作者认为，当正样本对和负样本对之间的距离大的时候，模型具有良好的泛化能力。因此，在实验中 p 应该取一个相对较大的值（$p \to 0$）。然而，当 p 越来越接近 0 的时候，在优化模型(4.10)时，为了满足 $(\frac{d_{v_i,c_t^{i-}}}{d_{v_i,c_j^{i+}}})^p = 0$，会使得 $\frac{d_{v_i,c_t^{i-}}}{d_{v_i,c_j^{i+}}} \to -\infty$，这意味着在优化过程中分母项 $d_{v_i,c_j^{i+}} \to 0$，对分子项 $d_{v_i,c_t^{i-}}$ 的约束越来越小，从而失去了对正样本对和负样本对之间距离的控制。综上所述，为了避免过拟合，和失去对正样本对和负样本对之间距离的控制，一个安全的选择是设置 $p \to 0$ 但不要太接近 0。据此，$p = -5$ 是一个最优选择。在实验部分第4.2.3章节，本文作者将会提供基于不同 p 值，提出的方法的性能表现，来验证以上关于参数 p 的分析。

本文作者采用基于线性搜索的梯度下降法求解排序度量函数模型。在每次迭代中，对于每个询问行人 v_i $(i = 1,...,N)$，仅采用其样本对集合中前 k 个距离最小的样本对参与运算（$k < M^{i-} + 1$），即在第 t 次迭代时，排序函数模型计算公式如下：

$$f_t(L) = \sum_{i=1}^{N} \sum_{c_j^{i+} \in C^{i+}} [d_{v_i,c_j^{i+}} - (\sum_{c_n \in \Omega_t^{i,j,k}} (d_{v_i,c_n})^p)^{\frac{1}{p}}], \tag{4.12}$$

其中，$\Omega_t^{i,j,k} \subseteq (\{c_j^{i+}\} \cup C^{i-})$ 是集合 $\{c_j^{i+}\} \cup C^{i-}$ 中前 k 个与 v_i 组成的样本对距离最小的样本，其样本对距离是通过第 $t-1$ 次迭代得到的 L_{t-1} 计算得到。相

应地，本文作者计算第 t 次迭代的梯度值：

$$\frac{\partial f_t(L)}{\partial L} = \sum_{i=1}^{N} \sum_{c_j^{i+} \in C^{i+}} [\pi_L(v_i, c_j^{i+}) - \rho(v_i, c_j^{i+}) \sum_{c_n \in \Omega_t^{i,j,k}} (d_{v_i,c_n})^{p-1} \pi_L(v_i, c_n)], \tag{4.13}$$

其中 $\pi_L(v,c)$ 是距离度量函数 $d_{v,c}$ 关于 L 的导数，$\pi_L(v,c) = 2L(x_v - x_c)(x_v - x_c)^T$，$\rho(v_i, c_j^{i+}) = (\sum_{c_n \in \Omega_t^{i,j,k}} (d_{v_i,c_n})^p)^{\frac{1}{p}-1}$ 是一个常数。对于方程中的参数 k，在实验中设置 $k = 2$。本文作者将会在实验部分第4.2.3章节详细分析参数 k 对性能的影响，验证 $k = 2$ 是一个最佳的取值，可以实现与 $k = M$ 几乎一样的性能，同时具有更短的算法运行时间。

在测试阶段，对于输入的拍摄于某一摄像头下的询问行人图像，本文作者采用方程(4.8)的距离度量函数和经过以上优化算法求解得到的最优解 L，依次与拍摄于其它摄像头下的行人图像库中的图像进行距离计算。根据计算得到的距离对行人图像库中的图像进行排序，找到其正样本，完成跨摄像头行人重识别。为了方便起见，在下文，对于提出的基于排序损失函数的行人重识别方法，本文作者简称为 R-Loss 方法。

4.2.3 实验与分析

4.2.3.1 数据集与实验设置

本文作者在 2 个公开的行人重识别数据集 VIPeR [1] 和 CUHK01 [101] 上评估提出的 R-Loss 方法的性能。对于 CUHK01 数据集，其单图像版本（single-shot）CUHK01(M1) 和多图像版本（multi-shot）CUHK01(M2) 在实验中均被采用。此外，本文作者分别采用 3 种特征描述符 Local Maximal Occurrence (LOMO) [9]，Gaussian Of Gaussian (GOG) [2] 和 Histogram of Intensity Pattern & Histogram of Ordinal Pattern (HIPHOP) [102] 进行实验。累计匹配特性曲线（Cumulative Matching Characteristics，CMC）作为性能评测指标，每个数据集随机划分训练集和测试集 10 次进行实验，并取 10 次实验结果的平均值作为最终结果。

4.2.3.2 方法分析

1）p-norm vs. min

对于模型(4.7)，本文作者采用 p 范数作为 min 函数的光滑近似，得到最终的优化模型(4.10)。由此所带来的优势，本文作者已经在第4.2.2节进行了理论分析。在这一小节，本文作者验证分析结果。本文作者采用 GOG 作为特征描述符，

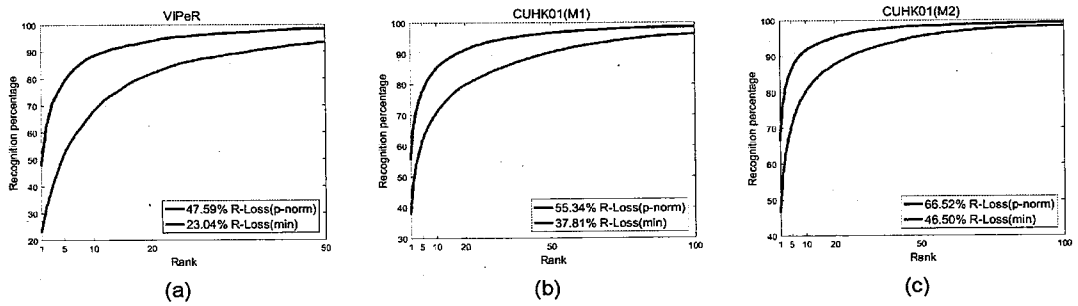


图 4.4 基于 p 范数的 R-Loss 方法和基于 min 函数的 R-Loss 方法的 CMC 曲线对比。

Figure 4.4 Comparison of CMC curves for the proposed method by using p-norm function and minimum function.

图4.4展示了基于 p 范数的 R-Loss 方法(R-Loss(p-norm))和基于 min 函数的 R-Loss 方法 (R-Loss(min)) 的对比结果。在所有数据集上，相比于 R-Loss(min)，R-Loss(p-norm) 的性能得到了大幅度提高，具体地，其 Rank 1 在 VIPeR、CUHK01(M1) 和 CUHK01(M2) 数据集上分别提高了 24.55%、17.53% 和 20.02%。这表明了，相比于 R-Loss(min)，对 R-Loss(p-norm) 模型进行求解可以收敛到一个更好的解。

2) 参数 p

对于参数 p 的选择，在第4.2.2节，本文作者呈现了一个详细的分析，并得出结论：$p \to 0$ 但不要太接近 0，最终认为 $p = -5$ 是一个最优选择。在这一小节，本文作者对基于不同 p 取值的 R-Loss 方法的性能进行比较。本文作者分别采用 LOMO 和 GOG 作为特征描述符，图4.5展示其在 VIPeR 和 CUHK01 数据集上的实验结果。从图中可以看到，当 $p \to -\infty$ 时，所有数据集上的重识别率都逐渐下降，具体地，当 $p < -10$ 时，所有数据集上的重识别率快速下降。这表明了 p 取值太低会影响方法的性能。此外，也可以看到当 $p = -1$ 时，CUHK01(M1) 和 CUHK01(M2) 数据集上的 Rank 1 快速下降，尽管 VIPeR 数据集上的 Rank 1 表现稳定，设置 $p = -1$ 仍然有一定的风险。因此，设置 $p = -5$ 是一个最佳选择。

3) 参数 k

对于模型(4.10)的求解，本文作者介绍了一个简化算法寻找其最优解，即，在优化算法的每一次迭代中，对于每个正样本对，仅采用前 k 个距离最小的样本对参与计算。在这一小节，本文作者调查 k 的取值对于方法性能和运行时间的影响。本文作者采用 LOMO 和 GOG 作为特征描述符。表4.5展示了实验结果。从表中可以看出，k 的取值变化对于 R-Loss 方法的性能基本没有影响，而随着 k 值

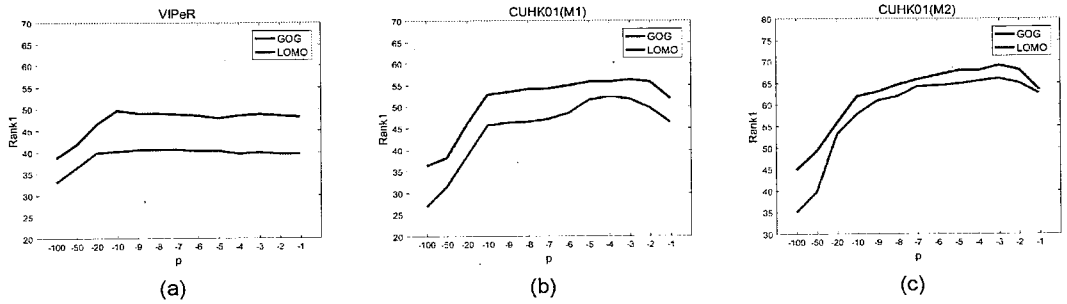


图 4.5 基于不同 p 取值的 R-Loss 方法的性能对比。

Figure 4.5 The performances of Rank1 for the proposed method with different value of p.

表 4.5 k 的取值对于提出的 R-Loss 方法的性能和运行时间的影响。运行时间指的是优化算法迭代一次的时间。$k = M$ 表示在优化过程中采用全部样本对参与计算。

Table 4.5 Effects of k on the performance and running time of the proposed R-Loss method. Running time refers to one iteration time of optimization algorithm. $k = M$ represents the case for using all samples in optimization.

k	VIPeR				CUHK01(M1)				CUHK01(M2)			
	GOG		LOMO		GOG		LOMO		GOG		LOMO	
	r=1	T(s)	r=1	T(s)	r=1	T(s)	r=1	T(s)	r=1	T(s)	r=1	T(s)
2	47.6	0.4	39.5	0.4	55.3	9.1	51.6	9.4	67.9	9.3	64.8	9.3
5	47.8	0.8	39.5	0.4	56.0	9.3	51.9	9.4	69.1	9.5	65.8	9.5
10	48.8	0.8	39.8	0.4	56.2	9.6	51.7	9.6	68.5	9.6	66.3	9.7
M	46.6	1.5	38.3	1.4	54.3	32.8	50.4	32.7	65.8	32.4	64.2	32.6

的变大，算法的运行时间逐渐变长。当 $k = M$ 时，全部样本对参与计算，使得优化过程变得更加复杂，导致性能下降。综上所述，$k = 2$ 是一个最优的选择。

4.2.3.3 与其它基于度量学习的方法的性能比较

本文作者已经分析了提出的基于排序损失函数的行人重识别方法相比于基于二分类损失函数、三元组损失函数和四元组损失函数的方法的优势。现在，本文作者对提出的基于排序损失函数的方法和这些方法进行性能比较。为了一个公平的比较，在对比实验中，本文作者采用 GOG 作为特征描述符。对比结果见表4.6，本文作者提出的 R-Loss 方法在性能上优于其它所有的对比方法。

表 4.6 与其它基于度量学习的方法在 VIPeR 和 CUHK01 数据集上的性能比较。最好的结果（%）用红色标注。

Table 4.6 Comparison with methods based on other loss function on VIPeR and CUHK01 datasets. The best results (%) are respectively shown in red.

Methods	VIPeR			CUHK01(M1)			CUHK01(M2)		
	r=1	r=5	r=10	r=1	r=5	r=10	r=1	r=5	r=10
binary	46.2	77.3	87.8	32.3	51.3	60.9	37.0	59.6	69.8
binary(smooth)	46.5	78.4	88.4	51.4	76.1	84.1	62.6	84.7	90.2
triplet	30.1	62.4	76.9	45.5	69.3	78.1	54.7	78.1	86.1
quadruplet	33.0	66.1	77.9	36.0	59.1	68.8	44.1	69.0	78.5
R-Loss	47.6	78.9	88.4	55.3	77.5	85.1	66.5	86.1	91.6

4.2.3.4 与最先进的方法的性能比较

本文作者分别在 2 个行人重识别数据集 VIPeR 和 CUHK01 数据集上进行与最先进的方法的性能比较。首先，表4.7列举了在 VIPeR 数据集上的性能比较。本文作者提出的方法分别与 13 个先进的行人重识别算法进行了比较，可以看出，本文作者的方法达到了最高的性能指标。

然后，本文作者在 CUHK01(M2) 数据集上进行了性能对比，如表4.8所示。可以看出，最好的性能由 MC-PPMN 方法 [134] 达到，这是一个基于深度学习的方法，但本文作者提出的方法在所有传统方法中，具有绝对的优势。

4.2.3.5 讨论

本文作者将提出的基于排序损失函数的行人重识别方法应用于深度学习中。在具体计算中，本文作者提出的模型无需再采用 p 范数作为 min 函数的光滑近似。本文作者采用 ResNet-50 [138] 作为基础模型，三元组损失与本文作者提出的排序损失分别作为损失函数进行在行人重识别数据集 Market1501 [139] 和 DukeMTMC [74] 上的对比实验[2]。实验结果表4.9所示。在所有的数据集上，本文作者提出的基于排序损失的方法比基于三元组损失的方法在 Rank 1 识别率上高大约 7%。

[2]对于 Market1501 和 DukeMTMC 数据集的介绍，请参见章节5.2.3。

表 4.7 与最先进的方法在 VIPeR 数据集上的性能比较。基于深度学习的方法中最好的结果加粗标注，传统的方法中最好的结果用红色标注。

Table 4.7 Comparison with the state-of-the-art methods on VIPeR dataset. The best results for deep learning based and traditional methods (%) are respectively shown in boldface and red.

	Method	Reference	r=1	r=5	r=10	r=20
Deep learning	IDLA	2015CVPR[63]	34.81	63.60	75.63	84.49
	Deep Ranking	2016TIP [132]	38.40	69.20	81.30	90.40
	TCP	2016CVPR [69]	47.80	74.70	84.80	91.10
	PDC	2017ICCV [114]	51.27	74.05	84.18	91.46
	DeepAlign	2017ICCV [125]	48.70	74.70	85.10	**93.00**
	EBG	2018CVPR [16]	**51.90**	74.40	84.80	90.20
	MLS	2018CVPR [133]	50.10	73.10	84.35	–
	MC-PPMN	2018AAAI [134]	50.13	**81.17**	**91.46**	–
Traditional	WARCA	2016ECCV[135]	40.22	68.16	80.70	91.14
	TMA	2016ECCV [89]	48.19	–	87.65	93.54
	PatchM&LocalM	2017PR [113]	46.50	69.30	80.70	–
	MVLDML+	2018TIP [136]	50.00	79.20	88.50	94.70
	GCT	2018AAAI [137]	49.40	77.60	87.20	94.00
	R-Loss	Our	53.01	83.07	90.82	96.27

表 4.8 与最先进的方法在 CUHK01 数据集上的性能比较。基于深度学习的方法中最好的结果加粗标注，传统的方法中最好的结果用红色标注。

Table 4.8 Comparison with the state-of-the-art methods on CUHK01 dataset. The best results for deep learning based and traditional methods (%) are respectively shown in boldface and red.

	Method	Reference	r=1	r=5	r=10	r=20
Deep learning	IDLA	2015CVPR [63]	47.50	71.60	80.30	87.50
	TCP	2016CVPR [69]	53.70	84.30	91.00	96.30
	Deep Ranking	2016TIP [132]	50.40	70.00	84.80	92.00
	DeepAlign	2017ICCV [125]	75.00	93.50	95.70	**97.70**
	MC-PPMN	2018AAAI [134]	**78.95**	**94.67**	**97.64**	–
Traditional	WARCA	2016ECCV [135]	65.64	85.34	90.48	95.04
	PatchM&LocalM	2017PR [113]	53.50	82.50	91.20	96.10
	GCT	2018AAAI [137]	61.90	81.90	87.60	92.80
	MVLDML+	2018TIP [136]	61.40	82.70	88.90	93.90
	R-Loss	Our	73.66	90.64	94.14	96.87

表 4.9 基于深度学习框架，与基于三元组损失的方法的性能比较。

Table 4.9 Comparison with triplet loss based method under deep learning scheme.

Methods	Market1501				DukeMTMC			
	r=1	r=5	r=10	mAP	r=1	r=5	r=10	mAP
triplet	81.8	93.0	95.8	65.6	70.8	83.7	87.5	53.2
R-Loss	87.4	94.7	97.1	73.4	78.4	89.8	92.6	64.2

4.3 本章小结

度量学习是解决行人重识别问题的一个重要步骤。现阶段针对基于度量学习的行人重识别方法，主要有两类：基于闭式解和基于迭代学习。针对基于迭代度量学习的行人重识别方法，本文作者在本章提出了两种方法。

首先，本文作者提出了一个新颖的基于区域特定的行人重识别方法。在进行度量学习之前，首先进行了特征预处理。对于特征预处理，本文作者利用 PCA 技术选择判别能力最好的特征，从而达到优化特征表达能力同时降维的作用；对于区域特定度量学习，这是基于行人图像不同区域特征分布存在差异的考虑，从而学习一个最优的特征转化矩阵，该转化矩阵是一个块对角矩阵，即，针对每一个区域特征，学习一个对应的子特征转化。本文作者在行人重识别数据集 VIPeR、PRID450S 和 GRID 上进行了实验，验证了提出的方法在性能方面具有先进性。未来，本文作者将考虑如何将特征转化矩阵的块对角约束融入在模型求解中，使得模型可以同时学习每个区域的特征转化，从而获得更优的性能。

其次，本文作者提出了一个基于排序损失函数的行人重识别方法。通过优化损失函数，使得正样本对的距离是所有样本对距离的最小值。本文作者提出采用 p 范数近似损失函数中的最小值，从而获得一个光滑且连续可导的损失函数，使得模型更易求解。此外，在优化过程中，本文作者提出了简化求解，对于每个询问样本，仅采用两个样本对进行优化，可以达到与采用全部样本对进行优化同等的性能，更重要的是，大大地提高了算法的效率。在 VIPeR 和 CUHK01 行人重识别数据集上，本文作者进行了全面的性能评估，展示了提出的方法可以获得具有竞争力的结果。在未来，本文作者将会考虑将排序损失函数应用于深度学习框架中，使得其在大数据集上获得更优的性能。

第 5 章　基于上下文信息的行人重识别方法

特征提取和度量学习是行人重识别的两个关键步骤。目前，行人重识别的研究工作主要针对这两方面展开。这些工作旨在充分挖掘行人自身的信息，并基于此进行建模，完成跨摄像头的行人重识别。这类方法可以称为基于内容的行人重识别方法。然而由于跨摄像头导致摄像机视角、光照、以及行人姿态发生变化，行人自身的信息也会发生变化，因此仅仅基于行人自身的信息进行重识别，算法性能是有限的。

为了进一步提高行人重识别的性能，本文作者提出了基于上下文信息的行人重识别方法。在利用行人自身的信息进行跨摄像头的行人重识别任务的基础上，还增加了行人样本的上下文信息辅助重识别任务。具体地，本文作者分别从时空域和欧式空间提取不同的上下文信息，相应，提出了基于广义群体信息的行人重识别方法和双边上下文驱动的渐进行人重识别方法。接下来，本文作者对这两种方法进行详细介绍。

5.1　基于广义群体信息的行人重识别方法

5.1.1　引言

近几年，随着智能视频监控的广泛应用，行人重识别引起了科研界越来越多的关注。对于解决该问题，研究人员主要致力于：以个人为研究对象，强劲的行人特征的表达和合适的距离度量的学习。然而，由于跨摄像头，光照、行人姿态、视角和背景等都会发生变化，导致行人的外貌发生比较严重的变化。此外，监控摄像网络每天会捕捉成千上万的行人，很多行人之间的外貌是相似的。因此，如果仅仅基于个人信息的特征提取和度量学习，行人重识别的性能一定是受限的。

为了提高行人重识别的性能，一个直观的解决方案是利用上下文信息，比如，场景中的其它行人。实验研究表明大约 50%-70% 的行人（取决于具体的环境）是与他人结伴而行的 [140]。如果两个摄像机相隔不是很远，相同的群体结构将会再次出现在相邻的摄像机，那么这样的群体结构就可以为行人重识别提供线索和帮助。图5.1展示了一个直观的例子。(a) 是询问行人（query person），(b)-(d) 是候选人（candidates）。从图中可以看出，候选人穿着相似的衣服，仅根

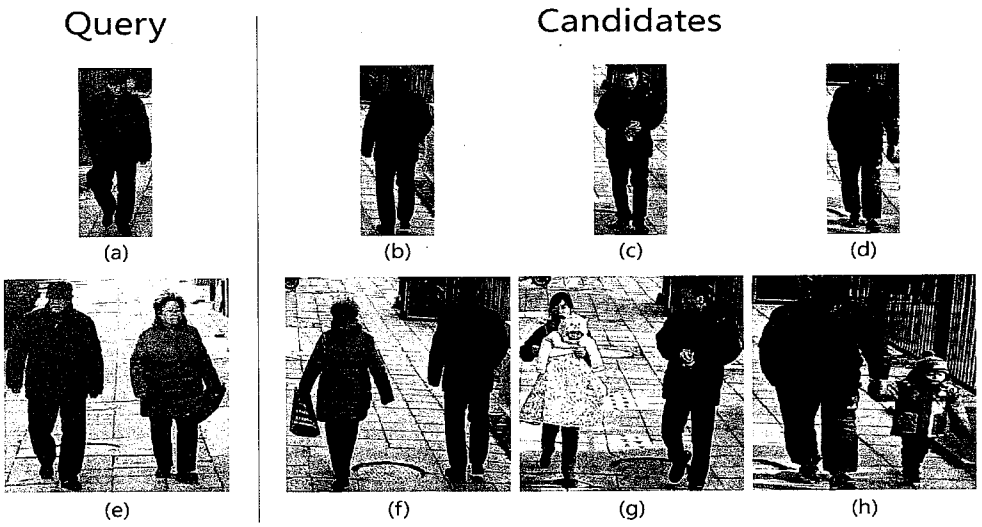


图 5.1 借助群体信息进行重识别的一个直观例子。(a)-(d) 展示了询问行人和与询问行人外貌相似的候选行人；(e)-(h) 展示了其相对应地的具有不同外貌特征的群体图像。

Figure 5.1 An intuitive example of re-identifying query person with the help of the group context. (a)-(d) show the query person and the candidates with similar appearance; (e)-(h) show the corresponding group images with different appearance.

据第一行图像的信息，从候选人中识别出与询问行人 (a) 相匹配的候选人 (b) 是比较困难的。然而，当引入群体信息 (e)-(h) 时，由于询问行人和候选人旁边的群体成员外貌的差异，就会比较容易地判断出 (f) 与 (e) 相匹配，从而确定 (a) 的匹配人。

近几年，研究人员提出了一些基于群体的行人重识别方法 [53–55, 141]，在一定程度地提高了行人重识别率。在这些方法中，群体被定义为一群具有相似速度和相近距离的人。群体的引入对行人重识别的准确率的提高主要在于：1）群体建立了询问行人与其周围行人的邻域关系；2）这些关系提供了额外的视觉信息；3）跨摄像头，这些关系是保持稳定的。群体引入行人重识别的成功，启发了本文作者：如果更多的询问行人与他人建立稳定的关系，更好的行人重识别率就可以达到。

基于此，本文提出了基于广义群体信息的行人重识别方法。具体的，该方法有两点主要贡献。

首先，本文作者扩展群体到一种稳定但更宽泛的关系，提出了广义群体的概念：一群有相似行走速度的行人。相比于传统的群体定义，由于本文作者对于群

体内行人之间的距离放松了约束，从而使得更多稳定关联被引入，对于重识别起到了积极的作用。

此外，无论是传统的群体还是本文提出的广义群体，这些群体关系的特征表示也应当确保跨摄像头的群体稳定性。许多基于群体的行人重识别方法，采用群体成员之间的空间关系和群体成员数量来表示群体特征。然而，群体是高度非刚性的，跨摄像头时群体成员之间的空间关系很有可能发生变化，群体成员数量也有可能发生变化。为了减少稳定关系在特征方面的不稳定的表达，本文提出了基于广义群体的对匹配机制。待匹配的人与该人所在群体内的其它成员分别组合，构成对（pair）。群体的匹配，从之前的提取群体特征进而特征之间进行匹配，转换为提取对的特征进而进行对的特征匹配。对匹配机制对于群体成员之间的空间关系和群体成员数量的变化有很好的鲁棒性。

5.1.2 算法模型

首先本文作者对监控视频中的行人进行广义群体检测。

在一段视频中，行人 p 的行走轨迹用一个三元组集合表示：$(\boldsymbol{s}_t, \boldsymbol{v}_t, t)$，$\boldsymbol{s}_t$ 和 $\boldsymbol{v}_t$ 分别表示行人在第 t 帧的位置向量和速度向量，帧 $t \in [T_1, T_2]$，T_1 和 T_2 分别表示行人 p 的起始帧和终止帧。相应地，可以算出行人 p 在时间 $[T_1, T_2]$ 的平均速度。

对于两个来自于同一个摄像头的行人 p_i 和 p_j，其中 $i \neq j$，$T_1^i \leq T_1^j$，$T_2^i \leq T_2^j$，他们在时间 $[T_1, T_2]$ 的平均速度分别为 $\boldsymbol{v}_i$ 和 $\boldsymbol{v}_j$。定义他们的对测量特征（pairwise measuring features）：

- 速度差：

$$v_{ij} = \|\boldsymbol{v}_i - \boldsymbol{v}_j\|,$$

- 行走方向差：

$$\theta_{ij} = arccos\left(\frac{\boldsymbol{v}_i \cdot \boldsymbol{v}_j}{\|\boldsymbol{v}_i\| \, \|\boldsymbol{v}_j\|}\right),$$

- 距离差：

$$\rho_{ij} = \frac{\|\boldsymbol{v}_i\| \cdot |T_1^i - T_1^j| + \|\boldsymbol{v}_j\| \cdot |T_2^i - T_2^j|}{2}.$$

则对于行人 p_i，其广义群体集合定义为：

$$C_i = \{p_j | v_{ij} < \tau_v, \theta_{ij} < \tau_\theta, \rho_{ij} < \tau_\rho, i \neq j\}. \tag{5.1}$$

τ_v、τ_θ 和 τ_ρ 是阈值参数。τ_v 和 τ_θ 应该取一个较小的值，相对地，τ_ρ 应该取一个较大的值。

群体检测之后，可以得到摄像机 A 捕捉到的行人 p_i^A 的群体集合 C_i^A。定义样本对 $P_{im}^A = (p_i^A, p_m^A)$，$p_m^A \in C_i^A$。相似地，对于摄像机 B 捕捉到的行人 p_j^B，群体集合表示为 C_j^B，则有 $P_{jn}^B = (p_j^B, p_n^B)$，$p_n^B \in C_j^B$。本文作者采用加权和计算 P_{im}^A 和 P_{jn}^B 的距离：

$$d(P_{im}^A, P_{jn}^B) = \alpha s(p_i^A, p_j^B) + (1-\alpha)s(p_m^A, p_n^B), \tag{5.2}$$

其中，α 表示权重（$0 \le \alpha \le 1$），$s(\cdot,\cdot)$ 是个人特征的距离度量。

对于行人 p，其对应的群体集合为 C，则 p 的样本对集合 $\mathcal{P} = \{(p,q)|q \in C\}$。如果摄像机 A 中的行人 p_i^A 和摄像机 B 中的行人 p_j^B 是正确的匹配，那么他们的群体也应该是匹配的，即 C_i^A 和 C_j^B 之间的距离度量值应该比较小。衡量 C_i^A 和 C_j^B 之间的距离等同于计算 $\mathcal{P}_i^A$ 和 $\mathcal{P}_j^B$ 的距离，C_i^A 和 C_j^B 之间的距离定义为：

$$s(C_i^A, C_j^B) = d(\mathcal{P}_i^A, \mathcal{P}_j^B). \tag{5.3}$$

根据集合匹配理论，本文作者计算 $\mathcal{P}_i^A$ 和 $\mathcal{P}_j^B$ 之间的距离：

$$d(\mathcal{P}_i^A, \mathcal{P}_j^B) = \min_{n,m} \left\{ d(P_{im}^A, P_{jn}^B) \mid P_{im}^A \in \mathcal{P}_i^A, P_{jn}^B \in \mathcal{P}_j^B \right\}. \tag{5.4}$$

相比于其它基于群体的行人重识别方法，本文作者使用样本对匹配机制和样本对集合距离衡量两个群体之间的相似性，从而有效地避免了由于群体内成员位置互换和群体成员数量变化所造成的群体无法匹配的问题。

如果行人 p_i^A 和 p_j^B 都是单独行动，不属于任何群体，则通过方程(5.1)的群体定义，他们的广义群体集合 C_i^A 和 C_j^B 都为空集。在这种情况下，p_i^A 和 p_j^B 分别与自身形成虚拟的样本对，$P_{ii}^A = (p_i^A, p_i^A)$，$P_{jj}^B = (p_j^B, p_j^B)$。然后本文作者仍然采用方程(5.4)计算 C_i^A 和 C_j^B 之间的距离，可以看出，此时 C_i^A 和 C_j^B 之间的距离退化为了行人 p_i^A 和 p_j^B 之间的距离。

最后，对于询问行人 p_i^A 和候选人 p_j^B，计算他们之间的距离：

$$d(p_i^A, p_j^B) = \tilde{s}(C_i^A, C_j^B) \cdot s(p_i^A, p_j^B), \tag{5.5}$$

其中

$$\tilde{s}(C_i^A, C_j^B) = \mathcal{N}(s(C_i^A, C_j^B)) \tag{5.6}$$

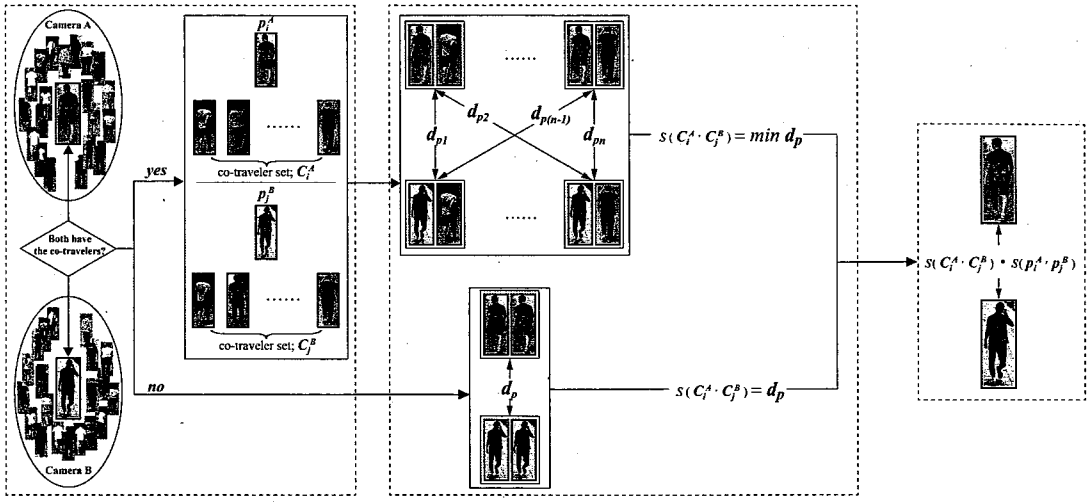


图 5.2 提出的基于广义群体信息的行人重识别方法的流程图。

Figure 5.2 Overview of the proposed co-traveler set based framework for person re-id.

是询问行人 p_i^A 的群体集合 C_i^A 和候选人 p_j^B 的群体集合 C_j^B 之间归一化后的距离值。$\mathcal{N}(\cdot)$ 表示线性函数归一化操作（min-max normalization operator），把数值归一化到 [0.1, 1] 的范围。$s(p_i^A, p_j^B)$ 是通过基于内容的行人重识别方法获得的行人之间的距离度量值。

在广义群体检测之后，每个行人都会获得一个标签信息：属于群体或不属于群体。对于询问行人 p_i^A 和候选人 p_j^B，有两种情况：1）这两个人具有相同的标签信息，也就是，p_i^A 和 p_j^B 都属于群体，或都不属于任何群体；2）这两个人具有不同的标签信息，也就是，其中一个人属于群体，而另一个不属于群体。显而易见，相比于第二种情况，第一种情况下的 p_i^A 和 p_j^B 更有可能是同一个行人（称之为正样本对）。因此，本文作者在方程(5.5)增加一个惩罚因子 λ 来区分以上这两种情况：

$$d(p_i^A, p_j^B) = \lambda \cdot \tilde{s}(C_i^A, C_j^B) \cdot s(p_i^A, p_j^B). \tag{5.7}$$

如果 p_i^A 和 p_j^B 具有相同的标签信息，则 $\lambda = 1$，否则，$\lambda = C$。C 是一个常数（$C > 1$）。通过引入 λ，属于群体的询问行人倾向于匹配同样属于群体的候选人，而不属于任何群体的询问行人倾向于匹配同样单独行动的候选人。

图5.2展示了基于广义群体信息的行人重识别方法的具体流程。本文作者将提出的方法简称为 CTS 方法（Co-Traveler Set based framework）。

5.1.3 实验与分析

本文作者分别在 3 个公开的行人重识别视频数据集：i-LIDS MCTS [53]，NLPR_MCT [55]，PRID2011 [100] 和一个新的数据集 CYBJ-G 评估提出的 CTS 方法。

i-LIDS MCTS 数据集的监控视频信息来源于一个繁忙时段的机场到达厅。Zheng 等人 [53] 从该数据集中提取了 65 个不同的群体，共有 274 张群体图像，这些图像的尺寸大小不一。65 个不同群体中的大多数群体有 4 张群体图像，这些图像来源于不同的摄像机，或者来源于同一个摄像机但拍摄于不同的位置。本文作者把这些群体图像的尺寸归一化到 182×60。这些群体存在严重的遮挡以及群体内成员跨摄像头位置变化的情况，因此这个数据集具有一定的挑战性。

NLPR_MCT 数据集包含三个子数据集。本文作者采用数据集 1 和数据集 2 中的视频信息（分辨率：320×240，帧率：20）进行实验。在数据集 1 和数据集 2 中，分别有 73 个行人和 106 个行人被两个不同的监控摄像头捕捉。在实验中，图像坐标和世界坐标之间的映射关系已经通过离线计算获得。

PRID2011 数据集中，两个不同的摄像机分别拍摄到了 385 个行人和 749 个行人。在实验中，本文作者仅采用在两个摄像机镜头中都出现过的 200 个行人的图像。广义群体检测所需要的相关信息，本文作者已经通过离线的行人追踪和摄像机标定算法获得。

CYBJ-G 是本文作者建立的数据集。该数据集包含有 194 个行人信息，每个行人的图像来源于两个位于住宅小区的监控摄像机，一个摄像机拍摄行人的前面，另一个摄像机拍摄行人的背面。在每个摄像机下，每个行人的数据是一个裁剪过的行人图像和对应的视频片段的序列图像，这些序列图像有 $9-199$ 帧不等。裁剪过的行人图像大小尺寸不等，本文作者将其归一化为 384×144。通过离线行人追踪和摄像机标定，广义群体检测所需要的相关信息已经获得。CYBJ-G 数据集的一些例子参见图5.3。

本文作者采用累计匹配特性曲线（Cumulative Matching Characteristics，CMC）评估和比较不同方法的性能。大多数基于内容的行人重识别方法都可以作为 CTS 方法的基准方法。对于需要训练模型的基准方法，本文作者随机划分数据集 10 次。具体地，一个连续的时间窗口随机滑动收集连续时间出现的行人构成测试集，这样做保证了测试集样本的时间连续性；其余的样本作为训练集。在实验

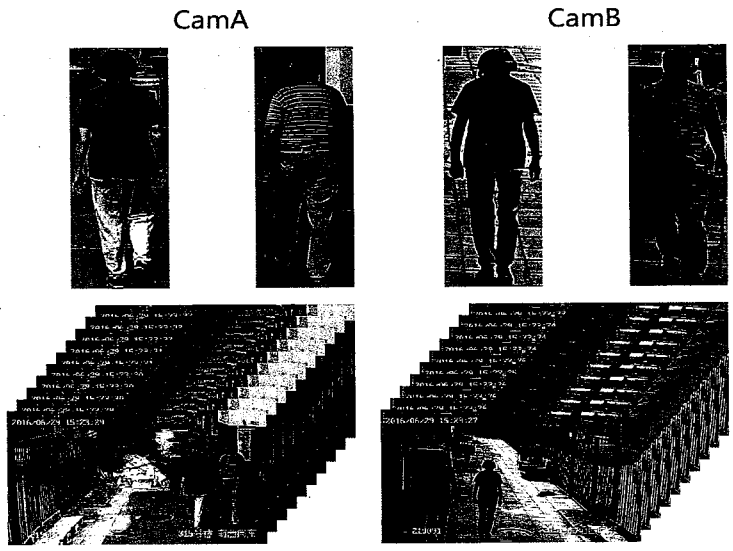


图 5.3 CYBJ-G 数据集的例子。第一行是裁剪过的行人图像，第二行是其对应的视频片段的序列图像。

Figure 5.3 Examples of CYBJ-G dataset: the cropped images of pedestrians are shown in the top row and their corresponding video clips in the bottom row.

中，方程(5.2)中的 α 设为 0.5，方程(5.7)中的惩罚因子 C 设为 2。

5.1.3.1 与其它基于群体信息的方法的性能比较

在这一节，本文作者在数据集 i-LIDS MCTS 和 NLPR_MCT 上比较本文作者提出的 CTS 方法与其它基于群体信息的方法：CRRRO [53] 和 SCGF [55] 的性能。对于 NLPR_MCT 数据集，关于群体检测的参数，本文作者分别设置为 $\tau_v = 0.38m/s$，$\tau_\theta = 10^\circ$ 和 $\tau_\rho = 10m$。为了公平比较，本文作者采用 SDALF [8] 作为基准方法。

对比结果如表5.1所示。大多数基于群体的方法都提高了基准方法的性能，展示了群体信息对于提高行人重识别性能的有效性。在 i-LIDS MCTS 和 NLPR MCT d1 数据集上，本文作者提出的 CTS 方法在 Rank = 1; 5; 10; 20 上优于所有比较的方法。在 NLPR MCT d2 数据集上，本文作者提出的方法在 Rank 1 上获得了最好的性能表现。

5.1.3.2 与先进的方法的性能比较

在表5.2，本文作者展示了提出的 CTS 方法与其它经典的和先进的基于内容的行人重识别方法的性能比较结果。实验在 PRID2011 和 CYBJ-G 数据集上进行。

表 5.1 与其它基于群体信息的行人重识别方法在 i-LIDS MCTS、NLPR_MCT 数据集 1 (d1)、NLPR_MCT 数据集 2 (d2) 上的性能比较。结果展示了 Rank = 1; 5; 10; 20 的匹配率 (%)。最好的结果和次好的结果分别用红色和蓝色标注。

Table 5.1 Comparison with group-based methods on i-LIDS MCTS, NLPR_MCT dataset 1 (d1) and dataset 2 (d2). Results are shown as matching rates (%) at Rank = 1; 5; 10; 20. The best and second results are shown in red and blue.

Methods	i-LIDS MCTS			NLPR_MCT d1			NLPR_MCT d2		
Rank	r=1	r=5	r=10	r=1	r=5	r=10	r=1	r=5	r=10
SDALF [8]	16.0	31.0	38.4	22.0	53.0	74.0	37.0	62.0	73.0
CRRRO [53]	12.5	28.4	36.4	25.0	57.0	76.0	41.0	67.0	78.0
SCGF [55]	-	-	-	25.0	64.0	80.0	44.0	72.0	82.0
CTS	22.1	40.6	48.5	26.0	67.1	82.2	44.3	66.0	79.2

表 5.2 与其它经典的和先进的行人重识别方法在 PRID2011 和 CYBJ-G 数据集上的性能比较。最好的结果和次好的结果 (%) 分别用红色和蓝色标注。

Table 5.2 Comparison with the classic and existing state-of-the-art methods on PRID2011 and CYBJ-G. The best and second best results (%) are respectively shown in red and blue.

	PRID2011				CYBJ-G			
Methods	r=1	r=5	r=10	r=20	r=1	r=5	r=10	r=20
SDALF [8]	4.1	20.6	31.6	41.9	32.2	57.2	69.8	79.9
Salience [142]	25.8	43.6	52.6	62.0	40.8	64.2	75.5	83.4
LOMO+XQDA [9]	39.0	68.0	83.0	91.0	67.2	87.4	91.7	95.2
SCSP [30]	12.7	32.7	51.0	66.0	21.7	39.1	50.0	67.4
DNS [29]	38.4	66.6	79.0	92.1	58.2	84.2	91.5	94.1
GOG [2]	59.2	83.5	92.2	96.8	76.5	93.3	97.2	98.3
DTDL [143]	41.0	70.0	78.0	86.0	-	-	-	-
PaMM [144]	45.0	72.0	85.0	92.5	-	-	-	-
STA [145]	64.1	87.3	89.9	92.0	-	-	-	-
SCSP+CTS	16.3	39.8	53.2	68.8	33.8	54.8	63.3	74.7
DNS+CTS	53.9	83.5	92.6	98.0	81.7	91.5	92.4	95.1
GOG+CTS	75.8	92.5	96.2	98.6	91.9	98.0	98.7	99.3

Methods		Query	Candidates					Rank
Baseline (GOG)								5
		scores	0.2483	0.3043	0.3191	0.3467	0.3722	
CTS (Ours)	Pair matching							1
		scores	0.1000	0.1573	0.2469	0.3071	0.3345	
	Individual matching							
		scores	0.0372	0.0479	0.0831	0.1038	0.1156	

图 5.4 本文作者提出的 CTS 方法应用在 CYBJ-G 数据集上的结果。基于广义群体的对匹配（在图中简称为 Pair matching）和加权的个人匹配（在图中简称为 Individual Matching）的匹配结果和排名名次展示在了图中。候选人下面的分数是询问行人与该候选人的匹配距离值。本文作者在图中展示了排名在前 5 名的候选人图像和相应的样本对图像。与询问行人相匹配的正样本候选人用红色框标注。

Figure 5.4 One example of re-id results from baseline method GOG and our proposed method. The matching results and ranks of co-traveler set based pair matching (Pair matching for short in figure) and weighted individual matching (Individual matching for short in figure) are displayed. The scores below the candidate are matching distances between query person (pair) and that candidate. We display the image of the top five candidate people and pairs in each grid. The ground truth matching is labeled by red boxes.

对于 CYBJ-G 数据集，本文作者设置群体检测相关的参数 $\tau_v = 0.38m/s$，$\tau_\theta = 10°$，$\tau_\rho = 10m$。本文作者采用 SCSP、DNS 和 GOG 方法作为 CTS 的基准方法。

从表5.2，可以看到，1）应用 CTS 到基准方法，可以实现性能增强，其中，GOG 作为基准的 CTS 实现了最好的性能表现，优于所有其它的方法；2）基准方法的性能越好，通过 CTS，可以实现越好的性能增强；3）在 PRID2011 数据集上，通过 CTS，Rank 1 增加了 3.6% ~ 16.2%，在 CYBJ-G 数据集上，Rank 1 增加了 12.1% ~ 23.5%，其增加幅度大于在 PRID2011 数据集上的增加幅度。这是因为在 CYBJ-G 数据集中，存在更多稳定的群体关系。

图5.4展示了 CTS 方法应用在 CYBJ-G 数据集上的一个例子。采用基准方法，与询问行人匹配的正样本候选人排在了第 5 名，然而，借助于群体信息，通过应用 CTS，正样本排在了第 1 名。

5.2 双边上下文驱动的渐进行人重识别方法

5.2.1 引言

行人重识别本质上可以看作是一个检索任务。给定一个询问行人和一个候选行人集，研究人员的目标是对候选行人集中所有候选行人进行排序，得到关于询问行人的排序结果。研究人员希望与询问行人相匹配的候选行人（也称为正样本）可以排在列表的最前面。目前主流的方法是提取行人图像的特征，并进行度量学习，然后根据个人之间的相似值进行升序排名得到排序结果 [2, 46, 113]。这是一个典型的基于内容的图像检索系统（Content-Based Image Retrieval，CBIR），仅仅根据询问行人和候选人之间的相似值决定他们的关系，无法充分揭示数据的流形结构 [146]。但是，通过考虑样本的上下文信息和数据的流形结构，可以得到更准确的样本对的相似值。图5.5验证了这个观点，相比于基于内容的方法，基于上下文的方法考虑了数据的流形结构，得到了正确的检索结果。

鉴于此，本文作者引入上下文样本集合辅助样本对的相似值计算。正如图5.6所示，即便两个目标样本彼此之间互不相似，但是存在另一个上下文样本，同时相似于这两个目标样本，那么本文作者认为这两个目标样本彼此相似。通过利用这些上下文样本计算样本对的相似值，本文作者有效地捕捉了数据的流形结构。因此，本文作者提出了一个双边上下文驱动的渐进行人重识别方法。

5.2.2 算法模型

给定一个询问行人 p 和一个候选行人集合 $G=\{g_i \mid i=1,\cdots,N\}$，通过基于内容的行人重识别算法，本文作者可以得到一个初始排序 $\mathcal{R}_p^o=\{g_1^o,g_2^o,\cdots,g_N^o\}$，其中，$S_{p,g_1^o}^o > S_{p,g_2^o}^o \cdots > S_{p,g_N^o}^o$，$g_i^o \in G$（$i=1,2,\cdots,N$），$S_{p,g_i^o}^o$ 表示 p 和 g_i^o 的相似值。本文作者通过充分考虑样本的上下文信息，优化初始排序 $\mathcal{R}_p^o$，使得更多的正样本排在列表的最前面。

不失一般性，借助 p 和 g 的上下文信息，本文作者重新计算 p 和 g 的相似值。本文作者分别基于一阶 k 近邻（first order k-nearest neighbor，一阶 k-nn）和二阶 k 近邻（second order k-nearest neighbor，二阶 k-nn）得到两种类型的上下文

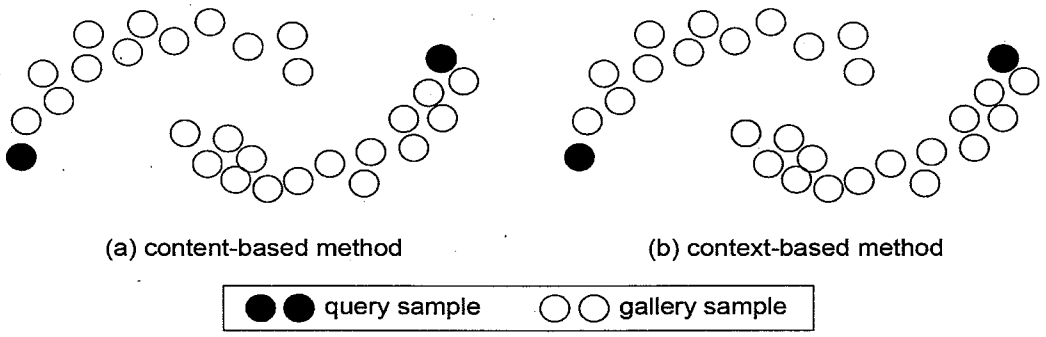


图 5.5 基于内容的方法和基于上下文的方法得到的检索结果的对比。每个候选人样本根据与询问行人样本的相似度标注颜色。

Figure 5.5 Comparison of retrieval results produced by (a) content-based method and (b) context-based method. The color of each gallery sample is marked according to the similarity to the probe sample.

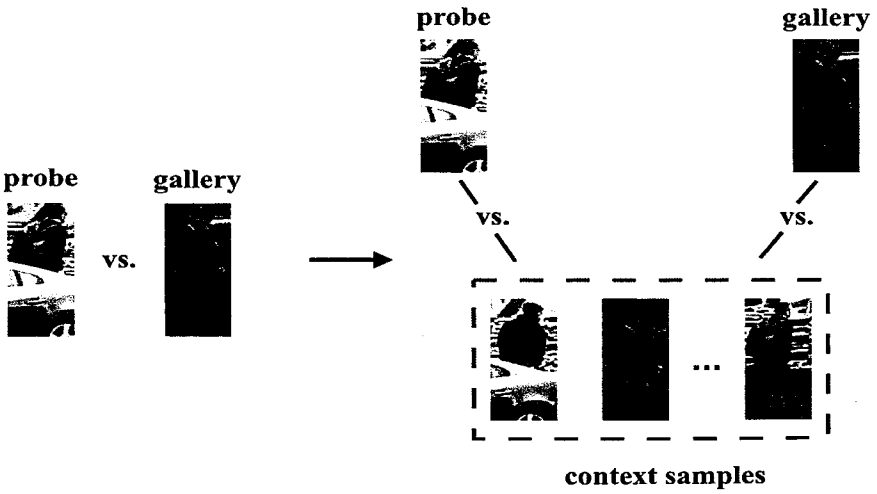


图 5.6 本文作者提出的方法的主旨。

Figure 5.6 Illustration of the gist of the proposed method.

信息。二阶上下文信息首先被用来优化初始排序 $\mathcal{R}_p^o$，然后在此基础上，利用一阶上下文信息进一步微调结果。整个过程是一个渐进的优化过程。

接下来，本文作者首先介绍 p 的上下文样本的定义，然后介绍 p 的上下文样本的权重计算，以及 p 的上下文样本与 g 的相似值计算。关于 g 的上下文样本定义、g 的上下文样本权重计算以及 g 的上下文样本与 p 的相似值计算，与 p 的计算一致。最后，本文作者介绍双边上下文驱动的针对初始排序 $\mathcal{R}_p^o$ 的优化计算。

本文作者借助 k-nn 算法定义样本的上下文。

对于询问行人样本 p，其上下文通过初始排序结果可以很容易得到，即排序结果的前 k 个样本作为其上下文。本文作者定义 p 的一阶上下文：

$$C(p,k)=\left\{g_1^o,g_2^o,\cdots,g_k^o\right\}, \tag{5.8}$$

其中 $|C(p,k)|=k$，$|\cdot|$ 表示集合中样本的数量。

更进一步，本文作者提出了基于二阶 k-nn 计算 p 的上下文。具体地，本文作者定义 p 的二阶上下文：

$$C(p,k_0,k)=\left\{C(g_1^o,k),C(g_2^o,k),\cdots,C(g_{k_0}^o,k)\right\}, \tag{5.9}$$

其中 $\left|C(p,k_0,k)\right|=k_0\times k$。二阶上下文是由 p 的排序结果的前 k_0 个样本的一阶上下文组成。

相比于一阶上下文，二阶上下文考虑了基于多阶的样本，可以提供更可靠和更有效的上下文信息。因此，在利用上下文信息计算样本对相似值的时候，本文作者首先采用二阶上下文改善初始排序，然后采用一阶上下文对结果进行微调，得到最终的优化结果。

值得注意的是，二阶上下文中的有效信息量主要取决于前 k_0 个样本，因为如果前 k_0 个样本中有一个样本与目标样本不相似，那么就相当于引入了 k 个干扰样本到二阶上下文。为了确保上下文的质量，k_0 应该取小值，相反，k 可以设置为较大的值。为了阅读方便，接下来本文作者用 $C_p=\left\{\rho_1,\rho_2,\cdots,\rho_{|C_p|}\right\}$ 和 $C_g=\left\{\varrho_1,\varrho_2,\cdots,\varrho_{|C_g|}\right\}$ 分别表示 p 的上下文和 g 的上下文。

对于上下文样本，与目标样本更相似的上下文样本，提供了更可靠的信息，所以在计算样本对相似值的时候，这类样本应该给予更大的权重。为此，本文作者根据上下文样本与目标样本的相似度，对上下文样本分配权重。

对于询问样本 p 的一阶上下文 $C(p,k)$ 中的样本 ρ_i（$i=1,2,\cdots,k$），本文作者根据 p 和 ρ_i 的相似性计算其权重。具体地，本文作者通过 p 和 ρ_i 在对方的排序列表所处的位置衡量 p 和 ρ_i 的相似性，进而计算 ρ_i 的权重：

$$w_{\rho_i}=\frac{1}{l_p(\rho_i)+l_{\rho_i}(p)+max(l_p(\rho_i),l_{\rho_i}(p))}, \tag{5.10}$$

其中 $l_p(\rho_i)$ 表示排序列表 C_p 中样本 ρ_i 的位置，$l_{\rho_i}(p)$ 表示排序列表 C_{ρ_i} 中样本 p 的位置。

对于询问样本 p 的二阶上下文 $C(p,k_0,k)$ 中的样本 ρ_i（$i=1,2,\cdots,k_0\times k$），根据方程(5.9)的定义，已经隐含分配了不同的权重。如果一个样本 ρ_i 与 p 相似，ρ_i 很有可能不仅属于集合 $C(g_m^o,k)$，也属于集合 $C(g_n^o,k)$（$m\neq n<k_0$）。那么根据方程(5.9)的定义，在 $C(p,k_0,k)$ 中就会有多个 ρ_i。相比于在集合 $C(p,k_0,k)$ 中出现次数少于 ρ_i 的样本，ρ_i 的信息被多次利用辅助重识别任务。总之，在二阶上下文中，样本在集合中出现的次数反映了这个样本与目标样本的相似程度，也是权重的直接表达。

通过以上计算，本文作者可以得到 p 和 g 的上下文集合 C_p 和 C_g，以及这些集合当中的上下文样本的权重。接下来，本文作者致力于询问行人的上下文样本和候选人之间的相似值计算。

本文作者根据排序信息计算相似值，定义候选人样本 g 和询问行人样本 p 的上下文样本 ρ_i 的相似值度量为：

$$r_{g,\rho_i}=\frac{1}{l_g(\rho_i)+l_{\rho_i}(g)+max(l_g(\rho_i),l_{\rho_i}(g))}. \tag{5.11}$$

然后结合样本 ρ_i（$i=1,2,\cdots,|C_p|$）的权重，本文作者计算候选人 g 和询问行人 p 的上下文集合 C_p 的相似值：

$$S_{g,C_p}=\sum_{i=1}^{|C_p|}w_{\rho_i}\cdot r_{g,\rho_i}. \tag{5.12}$$

类似地，本文作者也可以得到询问行人 p 和候选人 g 的上下文集合 C_g 的相似值：

$$S_{p,C_g}=\sum_{i=1}^{|C_g|}w_{\rho_i}\cdot r_{p,\rho_i}. \tag{5.13}$$

最后，本文作者重新计算询问行人 p 和候选人 g 的相似值，然后实现对初始排序 $\mathcal{R}_p^o$ 的优化。首先，本文作者借助二阶上下文集合 $C(p,k_0,k)$ 和 $C(g,k_0,k)$ 计

算 p 和 g 的相似值：

$$\hat{S}_{p,g} = S_{p,\hat{C}_g} + S_{g,\hat{C}_p},\ \hat{C}_p = C(p, k_0, k),\ \hat{C}_g = C(g, k_0, k). \tag{5.14}$$

对于来自于候选集中的每个候选人，本文作者根据方程(5.14)计算与询问行人 p 的相似值，然后就可以得到一个优化后的关于 p 的排序列表。在此基础上，本文作者利用一阶上下文集合 $C(p,k)$ 和 $C(g,k)$ 计算最终的 p 和 g 的相似值：

$$S_{p,g} = S_{p,C_g} + S_{g,C_p},\ C_p = C(p, k),\ C_g = C(g, k). \tag{5.15}$$

5.2.3 实验与分析

在本文作者提出的方法中，样本的上下文信息用于提高行人重识别的性能。因此，本文作者在三个大规模行人重识别数据集上进行实验，包括 Market1501 数据集 [139]、DukeMTMC 数据集 [74] 和 CUHK03 数据集 [62]。在这些数据集中，每个候选人都有多张图像，保证了样本的上下文可以为提高重识别性能提供有效信息。

Market1501。这个数据集中的行人来自于 2－6 个摄像机的捕捉，共有 1,501 个行人的 32,668 张图像。这些行人图像都是通过 DPM（Deformable Part Model）检测器 [5] 自动检测得到。751 个行人的 12,936 张图像作为训练集，其余的 750 个行人的 19,732 张图像作为测试集。本文作者基于单图像评估设置（single-query evaluation setting）进行实验。

DukeMTMC。这个数据集中有 1,812 个行人的 36,411 张手动裁剪好的图像，这些图像由 8 个摄像机捕捉。其中有 1,404 个行人被超过 1 个摄像机捕捉，而其余 408 个行人仅仅被 1 个摄像机捕捉，这 408 个行人的图像作为测试集中候选集的干扰样本。遵循其它方法对于该数据集的训练集和测试集的划分，702 个行人的 16,522 张图像用来作为训练集，其余 702 个行人的 2,228 张图像作为测试集的询问行人集，另外 17,611 张图像作为测试集的候选行人集。

CUHK03。在这个数据集中，共有 1,467 个行人的 14,096 张图像，它们来自 6 个摄像机。每个行人被其中 2 个摄像机捕捉，每个摄像机平均捕捉到 4.8 张图像。这个数据集分为两个版本：CUHK03(labeled) 和 CUHK03(detected)，分别代表行人图像是通过手动裁剪得到和自动检测得到。其中 767 个行人的图像作为训练集，其余的 700 个行人的图像作为测试集。

本文作者采用两个评估指标评价方法的性能：累计匹配特性曲线（Cumulative Matching Characteristics，CMC）和平均精度均值（mean average precision，mAP）。

接下来，本文作者在 3 个数据集上进行性能比较。本文作者提出的方法致力于利用样本上下文信息对基于内容的重识别方法得到的排序结果进行优化，从而达到提高行人重识别性能的目标，因此本文作者的方法也可以称为基于重排序的行人重识别方法。本文作者分别与基于内容的重识别方法以及基于重排序的重识别方法进行性能比较。为了进行公平的比较，本文作者采用 PCB [19] 作为所有基于重排序方法的基准方法。特别地，对于基于重排序的行人重识别方法，本文作者认为运行时间和 CMC、mAP 一样，也是一个重要的评价指标，因此与基于重排序的方法的比较，本文作者不仅比较 CMC 的 Rank 1 和 mAP 值，同时也比较算法的运行时间[1]。

在表5.3，本文作者展示了与 12 个基于内容的方法以及 3 个基于重排序的方法在 Market1501 数据集上的性能比较结果。与基于内容的方法相比，本文作者提出的方法在 Rank 1 和 mAP 上取得了最好的性能表现。与基于重排序的方法相比，本文作者提出的方法在 Rank 1 和 mAP 上获得了第二名，Rank 1 和 mAP 分别比性能表现最好的 ECN(rank-dist) [46] 方法低 0.4% 和 3.2%。然而，本文作者提出的方法在运行时间上具有绝对的优势。ECN(rank-dist) 方法运行了 13926.2 秒[2]，获得了最好的性能：Rank 1 为 94.6%；本文作者提出的方法仅运行了 9.3 秒，Rank 1 为 94.2%，几乎与表现最好的方法 ECN(rank-dist) 的 Rank 1 持平。

在 DukeMTMC 数据集上的比较结果，如表5.4所示。本文作者提出的方法在 Rank 1 上明显优于其它方法，在 mAP 上仅略低于 ECN(rank-dist) [46] 和 k-reciprocal [45] 方法 4.8% 和 0.1%。对于运行时间，本文作者提出的方法比 ECN(rank-dist) 和 k-reciprocal 快了 16 – 449 倍[2]。

在 CUHK03(labeled) 和 CUHK03(detected) 数据集上的比较结果如表5.5所示。对于 CUHK03(labeled)，本文作者提出的方法在 Rank 1 上取得了最好的结果，在 mAP 上取得了第二名的结果。对于 CUHK03(detected)，本文作者提出的方法在 Rank 1 和 mAP 上仅次于性能表现最佳的 ECN(rank-dist) 方法 0.5% 和 1.8%。然

[1]本文作者运行实验在一个载有 3.4GHz 的内核和 32GB 的随机存取存储器（RAM）的计算机上。

[2]ECN 方法在 Market1501 和 DukeMTMC 数据集上运行时，几乎占据了电脑 100% 的内存，从而导致了其非常慢的运行时间。

表 5.3 与最先进的方法在 Market1501 数据集上的性能比较。Time(s) 表示方法的运行时间，单位是秒。最好的结果和次好的结果分别用红色和蓝色加粗标注。

Table 5.3 Comparison among various state-of-the-art methods with the proposed method on the Market1501 dataset. Time(s) indicates the run time in seconds of the method. The best and second results are shown in red/bold and blue/bold.

Methods		Rank-1	mAP	Time(s)
Gated SCNN [147]	ECCV2016	65.9	39.6	-
PDC [114]	ICCV2017	84.1	63.4	-
LSRO [148]	ICCV2017	84.0	66.1	-
LML(S2S) [35]	TMM2018	65.3	39.8	-
MVLDML+ [136]	TIP2018	58.2	33.7	-
HA-CNN [10]	CVPR2018	91.2	75.7	-
MLFN [149]	CVPR2018	90.0	74.3	-
PABR [22]	ECCV2018	90.2	76.0	-
PN-GAN [150]	ECCV2018	89.4	72.6	-
GLIA [61]	ECCV2018	93.3	81.8	-
SGGNN [52]	ECCV2018	92.3	82.8	-
PCB (baseline) [19]	ECCV2018	93.1	80.4	-
k-reciprocal [45]	CVPR2017	94.2	88.6	**143.1**
ECN(orig-dist) [46]	CVPR2018	94.2	88.6	18970.7
ECN(rank-dist) [46]	CVPR2018	**94.6**	**92.1**	13926.2
Ours		**94.2**	**88.9**	**9.3**

表 5.4 与最先进的方法在 DukeMTMC 数据集上的性能比较。最好的结果和次好的结果分别用红色和蓝色加粗标注。

Table 5.4 Comparison of the proposed approach with various state-of-the-art methods on the DukeMTMC dataset. The best and second results are shown in red/bold and blue/bold.

Methods		Rank-1	mAP	Time(s)
LSRO [148]	ICCV2017	67.7	47.1	-
SVDNet [151]	ICCV2017	76.7	56.8	-
HAP2S [152]	ECCV2018	76.1	59.6	-
Pose-transfer [11]	CVPR2018	78.5	56.9	-
HA-CNN [10]	CVPR2018	80.5	63.8	-
MLFN [149]	CVPR2018	81.0	62.8	-
SPReID [153]	CVPR2018	82.0	73.3	-
PABR [22]	ECCV2018	82.1	64.2	-
Mancs [70]	ECCV2018	84.9	71.8	-
BDB+Cut [154]	ICCV2019	89.0	76.0	-
PCB (baseline) [19]	ECCV2018	85.0	72.4	-
k-reciprocal [45]	CVPR2017	88.4	**83.1**	**85.8**
ECN(orig-dist) [46]	CVPR2018	88.8	81.0	1471.9
ECN(rank-dist) [46]	CVPR2018	**90.2**	**87.8**	2426.6
Ours		**90.4**	83.0	**5.4**

表 5.5 与最先进的方法在 CUHK03 数据集上的性能比较。* 表示未正式发表的论文。最好的结果和次好的结果分别用红色和蓝色加粗标注。

Table 5.5 Comparison among various state-of-the-art methods with the proposed method on the CUHK03 dataset. * denotes an unpublished paper. The best and second results are shown in red/bold and blue/bold.

Methods	labeled			detected		
	Rank-1	mAP	Time(s)	Rank-1	mAP	Time(s)
TriNet+REDA* [155]	58.1	53.8	-	55.5	50.7	-
SVDNet [151]	40.9	37.8	-	41.5	37.3	-
HA-CNN [10]	44.4	41.0	-	41.7	38.6	-
Pose-Transfer [11]	45.1	42.0	-	41.6	38.7	-
Mancs [70]	69.0	63.9	-	65.5	60.5	-
PCB (baseline) [19]	57.8	54.0	-	54.9	50.7	-
k-reciprocal [45]	67.1	68.8	10.6	65.6	**66.3**	10.7
ECN(orig-dist) [46]	69.2	68.1	**5.3**	65.6	64.0	**5.2**
ECN(rank-dist) [46]	**69.4**	**71.2**	7.2	**66.4**	**67.9**	7.2
Ours	**70.1**	**70.3**	**2.1**	**65.9**	66.1	**2.1**

而，本文作者提出的方法仅需要非常短的运行时间。

5.3 本章小结

在本章，本文作者充分利用行人样本的上下文信息，实现行人重识别的性能增强。具体地，本文作者提出了基于广义群体信息的行人重识别方法和双边上下文驱动的渐进行人重识别方法。

对于基于广义群体信息的行人重识别方法，它与其它同类型的方法具有显著的差异，主要体现在：1）该方法基于一个更宽松的群体定义，本文作者称之为广义群体，它引入了更多的目标样本在时空域中的上下文信息，有益于性能的提高；2）该方法使用了有效的对匹配机制，进行广义群体的相似度量，该机制对于群体成员之间的空间关系和群体成员数量的变化具有很好的鲁棒性。基于广义群体信息的行人重识别方法可以采用大多数的基于内容的行人重识别方法作为基准方法，实验证明了在不同的行人重识别数据集中，提出的方法都可以提高重识别率。

对于双边上下文驱动的渐进行人重识别方法，这是一个具有一定推广性的

重排序方法，大多数基于内容的行人重识别方法都可以作为该方法的基准方法，并且通过提出的方法，重识别率可以得到进一步提高。具体地，双边上下文驱动的渐进行人重识别方法是基于假设：目标样本对中，任意一个目标样本在欧氏空间中的上下文集合与目标样本相似，那么目标样本对彼此相似。如此，在计算样本对的相似值时，同时考虑了数据集的流形结构。该方法简单且高效，可以处理现实情况中的大规模行人重识别任务。本文作者在三个大尺度的行人重识别数据集上的实验证明了该方法相比于其它方法的优越性。另外，相比于其它基于重排序的方法，双边上下文驱动的渐进行人重识别方法在计算效率上具有绝对的优势。

第 6 章 总结与展望

6.1 工作总结

跨摄像头行人匹配是行人重识别的任务。近几年，行人重识别问题受到了越来越多的关注，这个问题的兴起可以归功于两点：1）公共安全问题的需求增加；2）城市中摄像头网络的安装数量大幅增加。对于行人的跨摄像头追踪、行为分析和公共安全，行人重识别是一个非常有用的工具。在大规模监控中，行人长期追踪中的轨迹统计和提取，是行人重识别的一个应用。本论文针对行人重识别系统中两个重要步骤：度量学习和重排序，展开了详细研究。研究内容包括：1）针对行人重识别中的奇异问题，提出了基于正交线性判别分析的行人重识别方法；2）结合行人重识别具体问题，进行了具体分析，对度量学习中的目标函数进行了改进，提出了区域特定和排序损失函数的行人重识别方法；3）为了进一步提高重识别率，对样本的上下文信息进行了有效利用，提出了基于上下文信息的行人重识别方法。具体地，本文的主要工作如下：

（1）基于正交线性判别分析的行人重识别方法

在第 3 章中，针对行人重识别的奇异问题，本文作者提出了基于正交线性判别分析的行人重识别方法。具体地，在行人重识别问题中，样本的特征向量的维数通常大于样本的数量，导致样本的所有散度矩阵是奇异矩阵，典型的线性判别分析（Linear Discriminant Analysis，LDA）无法应用在行人重识别问题中。为此，本文作者提出利用伪逆 LDA（Pseudo-inverse LDA, PLDA）解决此问题，并通过同时对角化样本类内散度矩阵、类间散度矩阵和总体散度矩阵实现 PLDA 的求解。求解得到的最优转化向量彼此正交，因此该方法也可以称之为正交 LDA（Orthogonal LDA，OLDA）。此外，考虑到行人重识别是一个非线性问题，本文作者进一步发展了方法的核化版本，实现了性能的进一步提高，同时本文作者也发展了方法的快速版本，实现了算法效率的提高。通过在 4 个行人重识别数据集上的实验，证明了基于正交线性判别分析的行人重识别方法的性能优势，同时本文作者也进行了与方法核化版本、方法快速版本的对比实验，证明了这些版本的优越性。

（2）基于区域特定和排序损失函数的行人重识别方法

在第 4 章中，本文作者结合行人重识别的具体问题，针对性地提出了基于区域特定的行人重识别方法和基于排序损失函数的行人重识别方法。具体地，在特征提取阶段，为了使得提取到的特征可以更好的描述行人的空间信息，行人图像经常被划分为多个区域，然后针对每个区域提取特征，因此最终得到的特征向量是区域连接的。考虑到不同区域的特征存在差异性，本文作者针对每个不同区域，学习区域特定的度量函数，提出了基于区域特定的行人重识别方法。此外，在度量学习之前，本文作者对行人特征向量进行了预处理操作，通过利用主成分分析（Principal Component Analysis（PCA））技术筛选出区分能力最好的特征向量，实现了优化特征表达同时降维的作用。除此之外，本文作者针对行人重识别中构造目标损失函数时的目的进行了具体分析，发现现存的行人重识别方法中的目标损失函数都是通过一种相对间接的方法建模，针对样本集中的部分样本建模，导致模型鲁棒性差，可操作性不强。为此，本文作者对构造目标损失函数时的目的进行直接翻译，构建了一个新的排序损失函数，提出了基于排序损失函数的行人重识别方法。本文作者通过在行人重识别数据集上的实验，验证了基于区域特定和排序损失函数的行人重识别方法在性能方面的先进性。

（3）基于上下文信息的行人重识别方法

在第 5 章中，本文作者致力于借助样本的上下文信息，实现行人重识别性能的进一步提升。目前，大多数的行人重识别工作都是对行人的个人信息进行挖掘，基于此信息进行建模，实现重识别。为了更进一步的提高重识别率，本文作者对行人的上下文信息进行了挖掘，并与个人信息相结合，建立模型，实现重识别。具体地，本文作者分别从两个不同的角度挖掘样本的上下文信息。在时空域上，本文作者探索样本的上下文信息，即群体信息，提出了基于广义群体信息的行人重识别方法；在欧氏空间上，本文作者探索样本的上下文信息，即 k 近邻样本（k-nearest neighbor，k-nn），提出了双边上下文驱动的渐进的行人重识别方法。对于基于广义群体信息的行人重识别方法，相比于其它基于群体的行人重识别方法，该方法利用了更为宽泛的群体辅助重识别，引入了更多上下文信息，同时该方法采用对匹配机制进行群体匹配，该机制对于群体成员之间的空间关系和群体大小的变化有很好的鲁棒性。双边上下文驱动的渐进的行人重识别方法，本文作者在计算样本对之间的匹配值时，不仅考虑了样本对之间的相似关系，同时也考虑了样本对中一个样本与另一个样本的 k 近邻样本之间的相似关系，如

此一来，该方法有效地捕捉了数据的流形结构，计算的匹配值更可靠。本文作者在不同的行人重识别数据集对基于广义群体信息的行人重识别方法和双边上下文驱动的渐进的行人重识别方法进行了实验验证，证明了提出的方法相对于其它方法的优越性。

6.2 未来展望

尽管本文作者在行人重识别问题上进行了深入的研究，但是由于研究水平和时间的限制，本文工作仍然存在局限性，还有许多方法和理论需要完善和发展。接下来，可以从以下几个方面展开研究：

（1）对于基于正交线性判别分析的行人重识别方法，本文作者是为了解决行人重识别的奇异问题提出该方法。实际上，奇异问题普遍存在于图像分类和检索任务中。在现实生活中，对于行人重识别任务以及图像分类和检索任务，有时候会出现带有标签的样本数量有限，而未标签的样本数量庞大的情况。对于这种情况，基于正交线性判别分析的行人重识别方法无法通过少量的带有标签的样本学习到一个鲁棒的模型。因此，在未来的工作中，本文作者可以考虑进行半监督学习，解决奇异问题。比如，可以在模型学习之前，先通过标签传递技术，将标签样本的标签信息传递给未标签样本，增加标签样本的数量。

（2）对于基于区域特定和排序损失函数的行人重识别方法，尽管在基于区域特定的行人重识别方法中，本文作者将行人图像不同区域特征分布的差异性考虑其中，建立了一个新的目标函数，但是函数优化过程中，本文作者拆分了优化过程，对每个区域进行独立优化，得到的优化结果并不是全局最优解。在未来，本文作者考虑对函数进行整体优化，实现每个区域度量函数的同时优化，使得优化结果更加靠近全局最优解，从而实现更好的重识别率。而对于基于排序损失函数的行人重识别方法，目前在大数据集上与基于深度学习的方法相比，性能欠佳，在未来，本文作者考虑把排序损失函数应用于深度学习框架中，使其可以在大数据集上实现更优的性能表现。

（3）对于基于上下文信息的行人重识别方法，首先，对于基于广义群体信息的方法，在未来，本文作者将致力于通过借助深度学习技术，对群体信息进行更进一步的挖掘，将群体信息与个人信息有效结合，实现行人重识别性能的进一步提高。其次，对于双边上下文驱动的渐进的行人重识别方法，目前该方法对于初

始排序结果中排名非常靠后的正样本无法实现名次提升。这是因为跨摄像头视角等变化造成这类正样本与询问行人（query person）之间外貌特征差异太大，本文作者无法通过样本在欧式空间的上下文信息帮助其相互之间的正确匹配。在未来，本文作者将致力于这类困难正样本的重识别。