摘　　要

行人再识别是监控视频分析和理解过程中的一个十分重要的研究点，在智能监控、体感游戏和多媒体检索等领域都有非常大的应用价值。经过不同机构和研究者们的努力，行人再识别的性能得到了重大的改善，并使得它能更加适用于真实监控场景。但是，由于人体姿态多样性、摄像机视角变化、光照变化和复杂背景等因素的存在，人体的表观信息往往发生巨大的变化，从而给开发高效、鲁棒和可用的识别算法带来了巨大的挑战。

以往的行人再识别工作存在如下问题：1）在有监督地学习跨视角行人度量时，都只利用了行人身份的信息，而没有充分挖掘同一行人不同视角图像的内在联系；同时，这些只利用两种不同视角学习得到的距离度量模型，其泛化能力没有在多视角且大规模的数据库上进行验证。2）在无监督情况下，很多算法没有考虑行人局部表观的对齐问题，使得这些算法不能对遮挡、姿态变化和视角变化产生鲁棒性，且不适用于真实监控环境。

本文围绕这些行人再识别中的问题和难点，研究了如何自动学习“视角变化鲁棒的距离度量”以及在无监督情况下如何进行“行人局部表观的对齐”。具体地，本论文的主要贡献总结如下：

- 提出了一种视角自适应的度量学习方法

  为了处理多视角下的行人再识别问题，本文提出了一种视角自适应的度量学习方法。与以往只学习单个距离度量矩阵的方法不同，本文的方法在一个公共空间里同时学习了多个视角相关的度量矩阵。在对不同视角的成对样本进行距离度量时，能够自动选取与这两个视角对应的度量矩阵，用于匹配样本。具体地，给定一对样本（或特征），首先分别估计它们的视角向量（每一维表示属于一种视角的概率值）；然后自动的生成特定于这两个样本的一个全新度量。为了更好的达到这个目标，估计得到的视角向量与样本特征结合，被自动编码成一种增广特征的表示；最终，这种视角变化鲁棒的度量方法可采用经典马氏距离度量的计算方式来获得行人距离。进一步，本文还构建了一个大规模的多视角行人数据库用于性能验证。实验表明，本文提出的视角自适应度量学习方法有效地利用了行人多视角信息来获得对视角变化的鲁棒性，并且在多视角的大规模行人数据库上取得了性能的提升。

- 提出了一种基于推土机距离的行人局部表观对齐方法

  由于行人受视角、光照以及姿态变化的影响很大，不同时间和地点出现的行人往往伴随着剧烈的表观变化。很多已有的方法通过学习判别性特征或者距离度量来

获得对这些变化的鲁棒性。然后，要从真实的监控场景中获得标注的行人身份信息是十分困难的。本文提出了一种新的无监督行人匹配方法，即局部化约束的推土机距离，来自动学习成对行人样本间的最优距离。在此方法中，高斯混合模型被用于表示行人的局部表观特征。通过采用局部约束开发推土机距离的稀疏属性，不同高斯混合模型的成分高斯之间能够很自然地进行部分匹配，来达到行人局部表观的对齐。这种部分匹配策略的采用，使得此方法能够对较大的视角变化和姿态变化更加鲁棒。通过在两个公开数据库上进行实验，表明基于推土机距离的行人局部表观方法优于已有的无监督行人识别算法。

关键词：行人再识别，视角自适应的度量学习，推土机距离，高斯混合模型，局部表观对齐

Research on Person Re-identification

Yan Canxiang (Computer Applied Technology)
Advised by Professor Shan shiguang

Person re-identification is a very impotent research point in the field of video analysis and understanding. It is very valuable for intelligent surveillance, somatosensory game, multimedia retrieval and so on. With the hard working of different institutes and researchers, the performance of person re-identification achieves a lot of improvement, which makes proposed algorithms more applicable to real surveillance scenarios. However, due to many existing factors, such as pose variations, viewpoint change, illumination change, background clutter and so on, the change of local appearance is very huge, which makes development of robust and practical algorithms very challenging.

The existing methods of person re-identification have the following problems: 1) When learning cross-view distance metrics in a supervised way, they only make use of identity information of pedestrians. However the relationship between images from different camera views of an individual is not exploited fully. 2) The generalization ability of traditional metric learning-based methods, which learns metric from data only with two views, lacks multi-view and large-scale dataset for validation. 3) In unsupervised situations, many approaches always ignore alignment problem of person local appearance. The misalignment will make those methods neither robust to occlusion, pose variations and viewpoint change nor practical for real surveillance environment. To address this problem, two kinds of useful ways are researched in this thesis, i.e., metric learning method robust to changes of view and unsupervised alignment of person local appearance. Specifically, the main contributions of this thesis are summarized as follows:

- **Propose a view-adaptive metric learning method**

  In addressing the multi-view person re-identification problem, we propose a view-adaptive metric learning (VAML) method. Different from traditional metric learning methods, VAML attempts to adopt different metrics adaptively for different sample pairs under varying views. Specifically, given a pair of samples (or features extracted), VAML firstly estimates their view vectors (consisting of probabilities being each view) respectively, and then adaptively generates a new metric for the two samples. To better achieve this goal, we elaborately encode the automatically estimated view vector into an augmented representation of the input feature,

with which the distance can be analytically computed as done in traditional metric learning methods. Furthermore, we also contribute a new large-scale and multi-view pedestrian dataset for performance validation. Extensive experiments show that the proposed method achieves much higher accuracies than that of state of the art methods on the new dataset.

- **Propose an Earth Mover's Distance based method for alignment of local appearance**

  Person re-identification is a challenging problem due to drastic variations in viewpoint, illumination and pose. Many existing approaches learn discriminative features or distance metrics in a supervised way to handle those variations. However, it is hard to obtain data with labeled identities from real surveillance network. In this paper, we propose a novel unsupervised method, called locality-constrained Earth Mover's Distance (LC-EMD), to learn the optimal distance measure between a pair of person images. In our method, Gaussian mixture models (GMMs) are learned as signatures to represent local appearance. By exploiting the sparse property of EMD with locality constrains, partial matching between component Gaussians of the GMMs are developed naturally, which makes our method more robust to misalignment caused by large viewpoint and pose variations. The proposed method is validated on two publicly available datasets and achieves better performance than existing state-of-art methods.

Keywords: Person Re-identification，View-adaptive Metric Learning，Earth Mover's Distance，Gaussian Mixture Models，Alignment of Local Appearance