摘要

随着城市视频监控系统规模的不断增长，如何从海量监控图像视频数据中“快速找到想要的目标或对象”就成为视频分析与理解技术领域所面临的重大挑战。对象再标识技术是解决这一监控视频数据中“找不到”难题的关键：它通过将多个监控摄像头所监测到的、位于不同位置的同一个目标（如行人）关联起来，从而形成一个准确和稳定的人体（行人）身份。基于视觉表观特征的相似性匹配是目前对象再标识技术所采用的主要技术路线；但当视觉表观特征不可靠时（例如嫌疑人进行了化妆），基于行为特征（如步态）的分析就成为人体再标识问题的可选解决方案。然而，由于不同摄像机视角差异和监控场景不同，以及行人背包、携物、穿大衣等因素，导致了在不同监控摄像机视域下行人前景剪影形态特征会产生剧烈的变化，因而基于步态识别的行人再标识问题同样具有极大的技术挑战性。针对多角度、多协变量等复杂情况，本文[①]对基于步态识别的行人再标识技术进行了研究，取得的主要创新成果如下：

第一，针对训练数据缺失情况的多角度步态识别问题，提出了基于视角特征恢复模型（VFRM）的步态识别算法。一般地，解决多角度步态识别问题通常需利用大量对象所有角度的步态特征来训练角度转换模型（VTM）。然而在现实场景中，很难同时采集大量对象的多个角度数据。因此，本文将这种“多角度训练数据不完全”情况建模为“训练数据缺失”问题，并针对此问题提出了视角特征恢复模型 VFRM，可以在多角度训练数据不完全的情况下来学习得到健壮的角度转换模型 VTM。VFRM 利用基于测地距离的 K 近邻方法度量行人临近关系，并通过度量得到行人 K 近邻填补对象缺失角度下的步态特征。当应用 VTM 模型到现实场景中的步态识别问题时，本文还提出基于步态前景剪影自相关函数的步态周期提取算法，并提出步行轨迹拟合算法（WTF）来进行自动角度识别。在 PKU 及 CASIA-B 数据集上的实验结果表明，WTF 算法能够实现高准确率的自动角度识别，且基于 VFRM 的步态识别算法能在多角度训练数据不完全的情况下实现对现实场景中的不同视角步态进行有效识别。

第二，针对存在多协变量影响的复杂场景下步态识别问题，提出了基于瑞士轮赛制的多层排序模型。除角度变化以外，行人对象是否携包、对象衣鞋款式等多种条件变化也会影响步态识别的准确率，本文将这些会引起步态变化进而影响识别准确率的条件变化统称为“协变量”。针对多种未知协变量情况下的步态识别问题，本文采用多种特征来对行人的步态进行描述，并将步态识别建模为排序问题，即给定一个查询图

① 本研究得到国家重点基础研究计划项目（编号：2015CB351806）和国家自然科学基金（编号：61390515，61035001，61471042）资助。

像，以排序而非绝对分数的形式从候选图像库中返回匹配的结果。为了更好地组合不同的特征得到更好的再标识结果，本文提出了基于瑞士轮多回合竞争模型的级联排序算法：将一系列对应于不同特征或特征组合的排序器以级联方式进行组合，其中前面的排序器用于快速删除大量不正确的样例，而后续的排序器则以计算开销为代价使用更多的特征来删除不正确的样例；通过类似的多回合竞争，最终能高效地进行人体再标识。在室内人为控制数据集 CASIA-B，室外人为控制数据集 Soton Large 以及室外完全无人为控制实际监控场景数据集 PKU 三个数据集上的实验结果表明，基于瑞士轮赛制的多层排序模型能有效提升各种不同场景下的步态识别性能。

第三，针对监控视频场景中行人再标识需求，实现了一个基于步态的行人再标识系统。该系统包括目标对象步态特征提取和对象再标识两个子模块，目标对象步态特征提取模块通过前景检测对象并提取目标对象剪影、步态周期和步行角度，最终融合步态特征送入对象再标识模块。对象再标识模块在候选对象收据集中寻找与目标对象最为匹配的前三名再标识结果，将结果显示在系统界面中以完成用户指定的再标识任务。本文通过多个数据场景测试证明了系统良好的应用性能。

综上所述，本文针对基于步态识别的行人再标识技术所面临的两个重要问题进行了研究，同时对基于步态识别的行人再标识技术在实际监控视频场景中的应用作了初步的探索。本文首创性地提出训练数据缺失情况下的多角度步态识别问题，以及首次将瑞士轮排序模型应用在多协变量步态识别问题中，并通过大量实验展示了所提模型的可行性和有效性。此外，本论文的研究工作为在基于步态识别的行人再标识领域开展更为广泛的深入研究奠定了基础。

关键词：基于步态的行人再标识，多角度步态识别，多协变量步态识别

Person Re-identification with Gait Recognition

Lan Wei (Computer Applied Technology)
Directed by Prof. Tiejun Huang

ABSTRACT

In recent years, surveillance cameras have been widely deployed almost everywhere in the city. To automatically recognize or search a suspicious target/object from large and massive video monitoring image data, pose a major challenge for video data processing and understanding. Person re-identification technology is the key to solve the problem of "matching" the target in video surveillance data: it matches the same accurate and stable identifier to all the instances of the same object from various visual surveillance scenarios. Similarity matching based on visual appearance features is the mainstream of the current methods. However, behavior-based analysis (e.g. human gait) has been shown to be an efficient optional biometric measure for person re-identification problem especially when the visual appearance features are unreliable (e.g. when the suspect has been made up). Nevertheless, various factors can affect human gait, including variations of view, walking speed, carrying an object and even wearing different types of shoes, because the human gait silhouettes can be different under the various situations. To address the problem of multi-view gait recognition and multi-covariate conditions gait recognition, this thesis[①] investigates the technology of person re-identification with gait recognition. The main contributions of this thesis can be summarized as follows:

Firstly, this thesis presents a View Feature Recovering Model (VFRM) to generate the View Transformation Model (VTM) for multi-view gait recognition with incomplete training data. VTM is used to tackle the multi-view problem by transforming gait features from across views to a common viewing angle. However, VTM must use the data of subjects crossing all views to train the pre-constructed model, which might be unsuitable for the real applications. We address the problem of "incomplete training data" for the situation under which not all subjects' features from all viewing angles are available, and propose VFRM using the Geodesic distance based K-Nearest Neighbor (GKNN) algorithm to measure the neighborhood between two pedestrians. Moreover, this thesis formulates gait cycle

① This work is partially supported by grants from the National Basic Research Program of China under grant 2015CB351806, and the National Natural Science Foundation of China under contract No. 61390515, No. 61035001 and No. 61471042.

extraction method and a walking trajectory fitting (WTF) algorithm to recognize walking view of a gait sequence automatically. Experimental results on CASIA-B and PKU database has demonstrated the effectiveness of the proposed WTF and VFRM.

Secondly, this thesis proposes a Swiss-system based cascade ranking model for multi-covariate conditions gait recognition. Besides changes in the viewing angles, various covariate conditions, including viewing angles, walking speed, carrying an object and wearing different types of shoes, make the gait features less distinguishable. This thesis formulates the changes conditions making the gait features less distinguishable as gait recognition with "multi-covariate conditions". To address this problem, five such state-of-art gait features encoding both static, time-variant and motion information are extracted in this thesis to describe the gait feature, and the following gait feature matching problem is re-formulated as a bipartite ranking problem. A Swiss multi-round competition system of cascade ranking model is developed for multi-feature ensemble learning. In this system, a series of rankers are applied to every pair of matches between probe and gallery gait sequences. The initial ranker eliminates a large number of incorrect examples with very little processing, while the subsequent layers eliminate other incorrect ones by using more features at the cost of additional computation. Thus through such as multi-round competition, the effectiveness and efficiency of cascade ranking model can be improved remarkably. Extensive experiments are performed on three benchmarking datasets, including the indoor CASIA database, the outdoor Soton Large dataset, and the actual monitoring database. The experimental results demonstrate that our method outperforms several state-of-the-art methods remarkably.

Thirdly, to meet the demand of person re-identification problem under surveillance scenarios, a person re-identification system, Person ReID System, is developed in the thesis. This system consists of mainly two modules: the subject gait feature extraction model and person re-identification model. In these two modules, the gait silhouettes, gait cycles and walking angles are extracted to get the subjects' gait features. Then person re-identification algorithm is applied to find the best matches of the subject in gallery data. These matching objects, along with the ranking, are used to provide. With the pre-recorded video, indoor and outdoor real-time surveillance monitoring data, we prove the availability of the system.

To sum up, this thesis investigates two important issues: multi-view and multi-covariate conditions situation, in person re-identification with gait recognition. Moreover, tentative studies have been carried out on the application of gait-based person re-identification in

practical visual surveillance scenarios. To the best of our knowledge, this thesis presents a systematic study on how to solve the problem of multi-view gait recognition with incomplete training data, and presents the Swiss-system based cascade ranking model for multi-covariate conditions gait recognition for the first time. Moreover, this thesis demonstrates the feasibility and effectiveness of proposed methods with extensive experiments. This will spark a great interest of research in the related communities in years to come.

KEY WORDS: Gait-based person re-identification, Multi-view gait recognition, Multi-covariate gait recognition