Exploiting feature representations through similarity learning, post-ranking and ranking aggregation for person re-identification

Julio C.S. Jacques Junior, Xavier Baró, Sergio Escalera

PII: S0262-8856(18)30137-9
DOI: doi:10.1016/j.imavis.2018.08.001
Reference: IMAVIS 3709

To appear in: *Image and Vision Computing*

Received date: 8 October 2017
Revised date: 11 April 2018
Accepted date: 13 August 2018

Please cite this article as: Julio C.S. Jacques Junior, Xavier Baró, Sergio Escalera , Exploiting feature representations through similarity learning, post-ranking and ranking aggregation for person re-identification. Imavis (2018), doi:10.1016/j.imavis.2018.08.001

Exploiting feature representations through similarity learning, post-ranking and ranking aggregation for person re-identification

Julio C. S. Jacques Junior[a,b,*], Xavier Baró[a,b], Sergio Escalera[c,b]

[a] *Faculty of Computer Science, Multimedia and Telecommunication - Universitat Oberta de Catalunya, Spain*
[b] *Computer Vision Center - Universitat Autònoma de Barcelona, Spain*
[c] *Department of Mathematics and Informatics - University of Barcelona, Spain*

Abstract

Person re-identification has received special attention by the human analysis community in the last few years. To address the challenges in this field, many researchers have proposed different strategies, which basically exploit either cross-view invariant features or cross-view robust metrics. In this work, we propose to exploit a post-ranking approach and combine different feature representations through ranking aggregation. Spatial information, which potentially benefits the person matching, is represented using a 2D body model, from which color and texture information are extracted and combined. We also consider background/foreground information, automatically extracted via Deep Decompositional Network, and the usage of Convolutional Neural Network (CNN) features. To describe the matching between images we use the polynomial feature map, also taking into account local and global information. The Discriminant Context Information Analysis based post-ranking approach is used to improve initial ranking lists. Finally, the Stuart ranking aggregation method is employed to combine complementary ranking lists obtained from different feature representations. Experimental results demonstrated that we improve the state-of-the-art on VIPeR and PRID450s datasets, achieving 67.21% and 75.64% on top-1 rank recognition rate, respectively, as well as obtaining competitive results on CUHK01 dataset.

Keywords: person re-identification, similarity learning, feature fusion, post-ranking, ranking aggregation.

1. Introduction

Person re-identification is the task of assigning the same identifier to all instances of a particular individual captured in a series of images or videos, even after the occurrence of significant gaps over time or space. It has a wide range of applications, most of them focused on surveillance and forensic systems. Even though the proposed models and reported results in this field have considerably advanced in recent years [1, 2, 3], this task still presents open challenges, mainly due to the influence of numerous real-world factors such as illumination problems, occlusions, camera settings, as well as factors associated with the dynamics of the human being, like the great variety of appearance features, pose variations and strong visual similarity between different people. These difficulties are often compounded by low resolution images or poor quality video feeds with large amounts of unrelated information, making re-identification even harder.

As related in [4], given a query person image, in order to find the correct matches among a large set of candidate images captured by different cameras, two crucial problems have to be addressed. First, good image features are required to represent both the query and the gallery images. Second, suitable distance metrics are indispensable to determine whether a gallery image contains the same individual as the query image. An ideal measurement is a matching rule that yields higher matching score for the image pairs belonging to the same person than the pairs belonging to different persons, which can be a big challenge if images are captured by different views/cameras with different setups and illumination conditions (*i.e.*, a typical scenario found in person re-identification, usually not handled by direct distance metric comparison). As highlighted in [5], similarity measurements which are learned (*e.g.*, [6, 7]) from training samples generally enjoy better accuracy performance than learning free methods [8]. Note that the goal of metric learning algorithms is to take advantage of prior information in form of labels over simpler though more general similarity measures [9]. The achieved results are then provided in the form of a list of ranked matching persons. It often happens that the true match is not ranked first but it is in the first positions. This is mostly due to the visual ambiguities shared between the true match and other "similar" persons [10].

In order to address the re-identification problem, existing methods exploit either feature representation [11, 12, 13] or metric learning [9, 7]. In feature representation, robust and discriminative features are constructed

*Corresponding author
Email address: juliojj@gmail.com (Julio C. S. Jacques Junior)