Neurocomputing 168 (2015) 550–559

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Person re-identification with multi-level adaptive correspondence models

Shi-Chang Shi [a], Chun-Chao Guo [a,c], Jian-Huang Lai [a,b,*], Shi-Zhe Chen [a], Xiao-Jun Hu [a]

[a] School of Information Science and Technology, Sun Yat-sen University, Guangzhou 510006, P.R. China
[b] Guangdong Key Laboratory of Information Security Technology, Guangzhou 510006, P.R. China
[c] SYSU-CMU Shunde International Joint Research Institute (JRI), Shunde 528300, P.R. China

ARTICLE INFO

Article history:
Received 10 December 2014
Received in revised form
8 April 2015
Accepted 20 May 2015
Communicated by Rongrong Ji
Available online 28 May 2015

Keywords:
Person re-identification
Misalignment problem
Adaptive correspondence
Multi-level visual cues
Data-driven accumulation

ABSTRACT

In this work, we present a multi-level adaptive correspondence model for person re-identification. Coarse segmentation and single level representation carry poorly discriminative information for generating a signature of a target, whilst fine segmentation with a fixed matching fashion is hindered severely by misalignment of corresponding body parts. We address such a dilemma through a multi-level adaptive correspondence scheme. Our approach encodes a pedestrian based on horizontal stripes in multi-level to capture rich visual cues as well as implicit spatial structure. Then dynamic correspondence of stripes within an image pair is conducted. Considering that manually selected weights in the final fusion stage is not advisable, we employ RankSVM to seek a data-driven fusion solution. We demonstrate the effectiveness of our method on two public datasets and another new dataset built for single shot re-identification. Comparisons with state-of-the-art re-identification methods show the superior performance of our approach.

1. Introduction

Person re-identification is a task of matching pedestrians across non-overlapping camera views based on appearance, which has drawn great interest recently [2–11]. It can be widely used in visual surveillance including human retrieval, human tracking, activity analysis and scene understanding, where the use of face clues is restricted due to low resolution or far field of view. As illustrated in Fig. 1, drastic appearance variations can be observed even that an image pair in different cameras belongs to the same identity. Therefore, a key issue for re-identification is to match images in a robust and discriminative way, removing the hindrance of appearance variations brought about by illumination, pose, viewpoint and low resolution. Our work is mainly motivated by the following several aspects.

Varieties of features for re-identification have been explored in previous studies [3,12–15], most of which are extracted from coarsely segmented person images in a single level. For instance, feature pool is built upon a few stripes [16,17], two-part models [3,12], or body-part-based models [18]. In spite of the fact that conventional coarse segmentation can alleviate adverse effects of pose variations to some extent, this fashion carries poorly discriminative information, throwing away rich visual and structural hints. On the contrary, fine segmentation of pedestrian images seems to provide a promising way to embed rich visual cues. Fig. 2 gives a realistic demonstration. As shown by coarse segmentation in red rectangles, different persons share similar feature representations. This tends to bring on wrong matches in the following steps easily. Green rectangles in fine segmentation demonstrate the great power to discriminate similar persons. Unfortunately, a new threaten named misalignment between corresponding stripes comes out (yellow rectangles). Many matched features from the same range of axis within an image pair do not represent corresponding parts of a person. As a result, performance degradation is inevitable in the fixed matching scheme with fine representation. It is a dilemma to incorporate merits of both coarse and fine representation, as well as resolve the misalignment problem. Meanwhile, neither fine segmentation nor conventional coarse segmentation can be sufficient to depict spatial structure of a target. Hence a multi-level model is conducive to generate a more discriminative signature, which has not been widely explored for re-identification. In this work, we jointly address the aforementioned dilemma through a multi-level adaptive correspondence scheme.

The main procedures of our method are summarized as below. First we segment the foreground images of all pedestrians into horizontal equal stripes in multi-level and extract diverse features from each stripe. Then we match the probe with each sample from gallery set employing an adaptive strategy to build stripe-based

* Corresponding author.
E-mail addresses: no1.shichang@gmail.com (S.-C. Shi), chunchaoguo@gmail.com (C.-C. Guo), stsljh@mail.sysu.edu.cn (J.-H. Lai), chenshizh@gmail.com (S.-Z. Chen), hux999m@gmail.com (X.-J. Hu).

http://dx.doi.org/10.1016/j.neucom.2015.05.072
0925-2312/

Fig. 1. Some examples from public datasets. (a) Samples from the VIPeR dataset [1]. (b) Samples from the SYSU-sReID dataset. Each column represents the same identity. As seen from examples, images of the same person captured from different camera views always undergo drastic variations of pose and viewpoint, which will result in heavy misalignments for further matching.

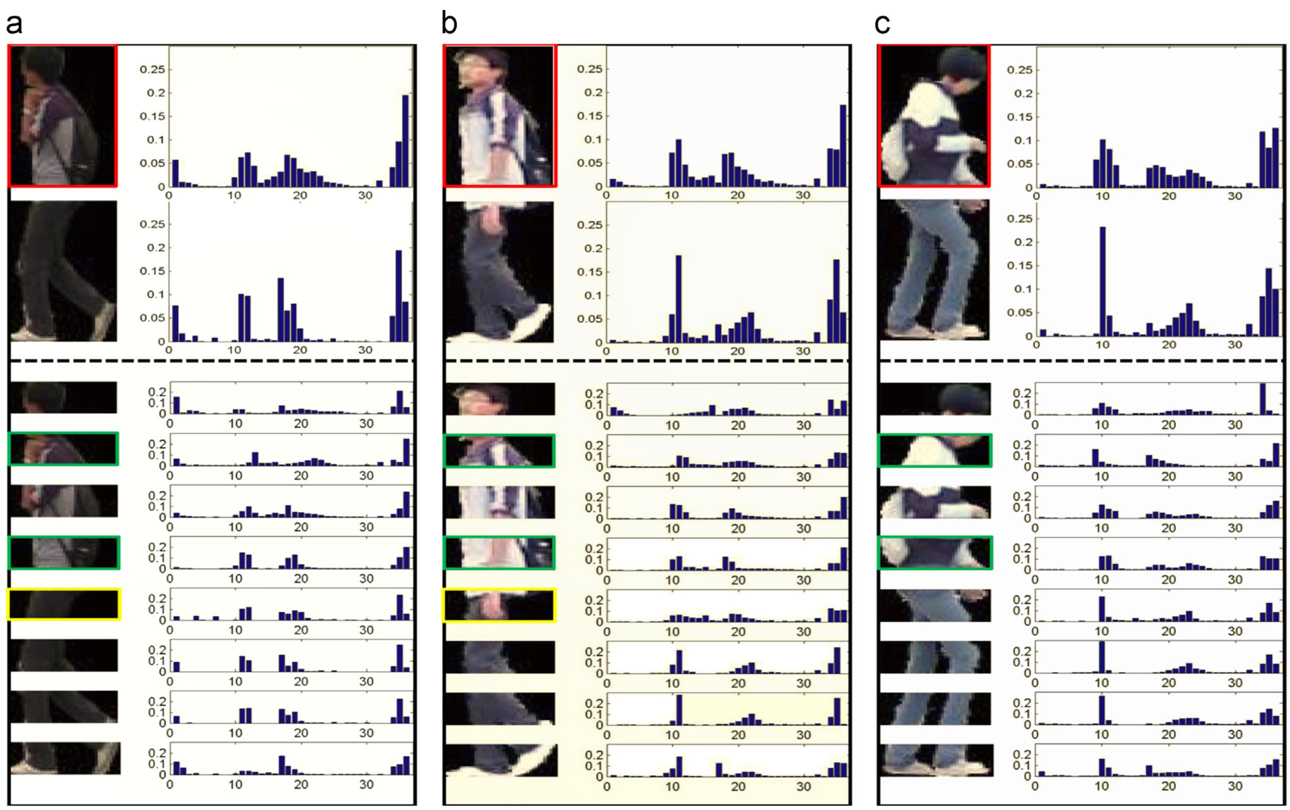

Fig. 2. The dilemma of whether following coarse segmentation or fine segmentation for human representation. This is an example from real-world scenarios. (a) is a target pedestrian in the probe set with the same identity of (b), while (c) belongs to another person in the gallery. Representation with the two strategies, i.e. coarse segmentation and fine segmentation are listed in the top and bottom, respectively, which are separated by dotted lines. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

feature correspondences. Finally, all correspondence scores in multi-level are accumulated in the RankSVM framework [19] to obtain final similarity measure of a pair of candidates.

The contributions of this work are three-fold:

- An adaptive correspondence model is developed to jointly satisfy the need of both coarse segmentation and fine segmentation as well as to tackle the misalignment problem. Misalignment is among major causes of performance degradation. However, to our best knowledge, it has not yet been adequately explored within the field of re-identification.
- A multi-level presentation is proposed in our model which employs a coarse-to-fine strategy and embeds spatial structure of a target implicitly. Compared with conventional single-level presentation, our approach carries richer and more discriminative information to re-identify a person.
- A data-driven strategy is explored to accumulate and fuse correspondence scores, which is shown to be more qualified than selecting weights of component scores manually.

The rest of this paper is organized as follows. Related work is discussed in Section 2. Section 3 introduces the proposed approach, including representation, matching score calculation and accumulation. The experimental evaluations and analysis are demonstrated in Section 4. Finally conclusions and future work are given in Section 5.