Applied Computing and Informatics (2016) 12, 142–153

Saudi Computer Society, King Saud University

Applied Computing and Informatics

(http://computer.org.sa)
www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Ensemble of different approaches for a reliable person re-identification system

Loris Nanni [a], Matteo Munaro [a], Stefano Ghidoni [a], Emanuele Menegatti [a], Sheryl Brahnam [b,*]

[a] Department of Information Engineering at the University of Padua, Via Gradenigo, 6-35131 Padova, Italy
[b] Computer Information Systems Department at Missouri State University, Springfield, MO 65804, USA

Received 16 December 2014; revised 2 February 2015; accepted 17 February 2015
Available online 7 March 2015

KEYWORDS

Person re-identification;
Texture descriptors;
Ensemble;
Color space;
Depth map

Abstract An ensemble of approaches for reliable person re-identification is proposed in this paper. The proposed ensemble is built combining widely used person re-identification systems using different color spaces and some variants of state-of-the-art approaches that are proposed in this paper. Different descriptors are tested, and both texture and color features are extracted from the images; then the different descriptors are compared using different distance measures (e.g., the Euclidean distance, angle, and the Jeffrey distance). To improve performance, a method based on skeleton detection, extracted from the depth map, is also applied when the depth map is available. The proposed ensemble is validated on three widely used datasets (CAVIAR4REID, IAS, and VIPeR), keeping the same parameter set of each approach constant across all tests to avoid overfitting and to demonstrate that the proposed system can be considered a general-purpose person re-identification system. Our experimental results show that the proposed system offers significant improvements over baseline approaches. The source code used for the approaches tested in this paper will be available at https://www.dei.unipd.it/node/2357 and http://robotics.dei.unipd.it/reid/.

1. Introduction

Person re-identification is the task of recognizing a given individual when he or she is viewed across any number of non-overlapping views in a distributed network of cameras or at different time instants when captured by a single camera. Research in person re-identification is motivated by the need of automating many surveillance activities in airports, metro stations, etc. This task requires the creation of a model recording macroscopic characteristics, as many of the classic biometric cues (facial appearance and gait characteristics) are often not

* Corresponding author. Tel.: +1 417 8739979.
E-mail addresses: nanniemg@dei.unipd.it (L. Nanni), munaroemg@dei.unipd.it (M. Munaro), ghidoniemg@dei.unipd.it (S. Ghidoni), sbrahnam@missouristate.edu (S. Brahnam).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.aci.2015.02.002
2210-8327

available due to the low frame-rates and resolutions of many surveillance cameras. Appearance-based, non-collaborative scenarios are challenging because the system must measure the similarity between two person-centered bounding boxes and correctly identify the same person despite changes in illumination, pose, background, occlusions, and the variability in camera resolutions and viewpoints. New advances, however, such as using 3D sensors [16], are making it possible to extract some soft-biometric features such as a person's 3D shape, height, and the lengths of limbs.

This paper targets short-term re-identification, which aims at recognizing people within relatively short time frames, thus relying on the assumption that the person is wearing the same clothing during the training and testing phases. Unlike tracking, we assume that no motion information is available for this task.

In the literature on this topic, the features that are most commonly exploited are color, texture and shape. For instance, in [7,6,8], the body of each target is divided into smaller parts and evaluated with multiple color histograms, one for each part. Even though this method is simple and effective, it fails in the case of strong illumination changes. Texture-based and shape-based approaches, such as [1,11,24], usually make use of local features, which provide a detailed description of targets. These approaches exploit descriptors evaluated on a set of keypoints to generate the signature of a target. The performance of this method is thus dependent on the capability of the keypoint detector to select stable features. In [17] a texture-based signature is proposed that consists of local descriptors computed around the principal joints of the human body. To detect the body joints, 3D data from consumer depth sensors and state-of-the-art skeletal tracking algorithms are exploited. The resulting Skeleton-based Person Signature (SPS) has proved to be very robust in the presence of strong illumination changes. The main drawback of this approach, however, is its dependency on the skeletal tracker; when this fails to recognize the body pose, the provided signature is meaningless. For a recent survey on person re-identification, see [23].

In this paper we improve the performance of state-of-the-art person re-identification systems using an ensemble of methods combined by weighted sum rule. The different systems utilize different color spaces and several texture and color features for describing the images. To the best of our knowledge, this is the first work in which several different state-of-the-art person re-identification systems, and their variants, are combined to obtain a more robust approach.

To demonstrate the generality of our system, we validate our approach on the following well-known datasets: CAVIAR4REID, IAS, and VIPeR. Moreover, we test our system on a dataset derived from VIPeR, which we call VIPeR45 because it contains 45 image pairs from VIPeR that focus on some of the most difficult samples to re-identify images of persons containing strong pose changes, for instance, or wearing very similar clothing. VIPeR45 was created because person re-identification performance was tested in [7] using a dataset that was built in a similar fashion (i.e., using 45 difficult image pairs extracted from VIPeR); the human subjects obtained a Rank(1) of 75% and a Rank(10) of ~100% [7]. Thus, it is possible for other researchers in person re-identification to use VIPeR45 for approximately comparing the performance of their computer vision systems with the performance of human beings at this same task. The VIPeR45 dataset will be available at http://robotics.dei.unipd.it /reid/.

The remainder of this paper is organized as follows. In Section 2 we describe the base approaches used in our system and provide details of our weighted ensemble. In Section 3, we describe the datasets used in our experiments, and in Section 4 we provide the experimental results. Finally, in Section 5 we summarize the significance of our work and highlight some future directions of exploration.

2. Methods

In this work we compare and combine different recent state-of-the-art person re-identification systems, viz. a representation that combines biologically inspired features and covariance descriptors, called gBiCov [15], Symmetry-Driven Accumulation of Local Features (SDALF) [8], Custom Pictorial Structures (CPS) [7] based on chromatic content and color displacement (CCD), Color Invariants (CI) [12], and the Skeleton-based Person Signature (SPS) technique [17]. Moreover, we propose variants of such approaches, obtained by varying the features used for describing the images and by using different distance measures. Each of these state-of-the-art systems, our variants, and the different color spaces, distance measures (specifically, the Jeffery Divergence measure, which obtains the best performance), and the color and texture descriptors used in our approaches are described in this section.

The following descriptors (detailed in Section 2.8) are tested:

- Color: Color descriptor proposed in [3].
- WLD: Weber's Law Descriptor proposed in [5].
- LPQT: Local Phase Quantization from Three Orthogonal Plane proposed in [18].
- VLPQ: Volume Local Phase Quantization proposed in [19].

The best approach (see Fig. 1) is obtained by combining several methods (detailed in this section) that utilize different characteristics and can be described as follows:

- Convert the RGB image to XYZ.
- Extract the pictorial structures (PS) from both the RGB and XYZ image.
- Find skeleton joints from the RGB image in the 3D domain using the tracker.
- Extract gBiCov and SDALF from the RGB image: several descriptors are used to describe the region found by PS and the area around the skeleton joints.
- Match the two images using an appropriate distance measure: different Matching Functions (MF) are used in the different methods.
- Combine the set of matching scores by sum rule.

It is important to note the methods composing the ensemble schematized in Fig. 1 all work in parallel, i.e., each method is performed independently of the others. The scores of each method are simply summed (after normalization to mean 0 and std = 1). Moreover, the proposed ensemble uses no optimization algorithm: we simply combine the best methods for optimizing the average performance on the tested datasets.