Neurocomputing 000 (2017) 1–12

Contents lists available at ScienceDirect

Neurocomputing

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

Perceptual hash-based feature description for person re-identification☆

Wen Fang[a], Hai-Miao Hu[a,b,*], Zihao Hu[a], Shengcai Liao[c], Bo Li[a,b]

[a] Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China
[b] State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China
[c] Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

ARTICLE INFO

Article history:
Received 11 April 2017
Revised 7 July 2017
Accepted 9 July 2017
Available online xxx

Communicated by Jiwen Lu

Keywords:
Person re-identification
Image pyramid
Regional statistics
Hierarchical feature description
The perceptual hash algorithm (PHA)

ABSTRACT

Person re-identification is one of the most important and challenging problems in video surveillance systems. For person re-identification, feature description is a fundamental problem. While many approaches focus on exploiting low-level features to describe person images, most of them are not robust enough to illumination and viewpoint changes. In this paper, we propose a simple yet effective feature description method for person re-identification. Starting from low-level features, the proposed method uses perceptual hashing to binarize low-level feature maps and combines several feature channels for feature encoding. Then, an image pyramid is built, and three regional statistics are computed for hierarchical feature description. To some extent, the perceptual hash algorithm (PHA) can encode invariant macro structures of person images to make the representation robust to both illumination and viewpoint changes. On the other hand, while a rough hashing may be not discriminative enough, the combination of several different feature channels and regional statistics is able to exploit complementary information and enhance the discriminability. The proposed approach is evaluated on seven major person re-identification datasets. The results of comprehensive experiments show the effectiveness of the proposed method and notable improvements over the state-of-the-art approaches.

1. Introduction

Person re-identification describes the process of pedestrian images observed across camera views at different locations and times based on visual features [1]. Appearance-based person re-identification is a very challenging task because the appearance of an individual undergoes significant changes due to variations of illumination, poses and viewpoints across non-overlapping camera views. Image resolution, camera settings and background clutter will increase the difficulty of person re-identification. It can be observed from Fig. 1 that the appearance of an individual varies greatly in different camera views.

The re-identification method usually requires two stages. First, a reliable and distinctive descriptor is constructed to describe both the query and the gallery images. Second, the adapted distance measures are used to calculate the similarity between the query and each of the gallery images, and the similarity is used to find the correct match among a large number of gallery images.

Development of a representation construct is a critical and challenging problem in person re-identification. There has been a great deal of work focused on constructing the descriptor. Methods in the recent literature of person re-identification exploit low-level features such as color [41], shape and filter [6,22], spatial structure [49] or combinations thereof [6,32], because they can be relatively easily obtained from the image. Usually, the constructed descriptor should be robust to various changes, such as changes in illumination, viewpoint, background clutter, occlusion and image resolution. Nevertheless, despite extensive research, finding a method to construct the descriptor for person re-identification is still a largely unsolved problem. This is because most low-level feature representations are either insufficiently robust to illumination changes, especially with noisy background, or insufficiently discriminative for viewpoint variations. For example, color is sensitive to lighting variations, and texture features are subject to the variations in viewpoint and pose. Thus, such low-level features are not effective for re-identification.

In this article, we propose a simple yet effective feature description method to address the above issue. Our descriptor is named MSHF, which is short for multi-statistics on hash feature map.

The proposed MSHF descriptor includes two steps. In the first step, low-level color and gradient features of the image are extracted, and each low-level feature is quantified into 2 scales by the perceptual hash algorithm (PHA) [58]. Since a rough hashing

☆ This work was partially supported by the National Key Research and Development Program of China (Grant no. 2016YFC0801003), and the National Natural Science Foundation of China (No. 61370121).

* Corresponding author at: Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China.
E-mail address: Frank0139@163.com (H.-M. Hu).

http://dx.doi.org/10.1016/j.neucom.2017.07.019
0925-2312/

Fig. 1. Images of the same person from two different camera views.

may be not discriminative enough, the combination of various different feature channels is able to exploit complementary information and enhance the discriminability. Then, a hash feature map, in which each pixel is represented as a binary number, is generated by the combined quantified low-level features of each pixel. In the second step, three complementary regional statistical features are extracted from the center area of this hash feature map. These features include the histogram, the mean vector and the co-occurrence matrix. The simplest way to describe a hash feature map is by its raw pixel values, which were used for many years in computer vision. However, this feature is not robust to various changes and nonrigid motion, and the dimension value is high [36]. To describe the image more effectively, we extract three regional statistical features from the hash feature map to describe people. Furthermore, the center area of the hash feature map is selected, and the noisy background information can be discarded. Finally, since the MSHF descriptor and low-level features provide very different types of information, we combine the MSHF descriptor with other low-level features to enhance the performance. Additionally, since the metric learning algorithms improve the performance of person re-identification in the recent literature, we also use metric learning to further enhance the performance of the person re-identification.

Compared with the state-of-the-art descriptors, the proposed descriptor is different in the following three ways. First, the computational complexity of our descriptor is relatively low, since the perceptual hash algorithm (PHA) quantifies each low-level feature on 2 scales. Second, the proposed MSHF descriptor is tolerant to both illumination and viewpoint changes; to some extent, we use PHA to encode some invariant macro structures of person images. Furthermore, MSHF is also robust to background variations, because we extract the center area of the hash feature map and discard the noisy background information.

However, it is important to note that the MSHF descriptor makes very different use of the perceptual hash algorithm. First, in the image retrieval, the perceptual hash algorithm only uses the gray information of the image, while the MSHF descriptor employs different types of low-level features in the person re-identification. Because a rough hashing may be not discriminative enough, the combination of several different feature channels and regional statistics is used to exploit complementary information and enhance the discriminability. Second, the perceptual-hash-based similarity is defined as the difference between the string descriptors of two different images. The MSHF descriptor extracts three complementary features from the hash feature map. These three complementary descriptors are then concatenated to form the image signature, and the similarity of two different images is obtained by simply computing the l_1 vector distance between their descriptors.

The remainder of this paper is organized as follows. In the next section, we review the related works on person re-identification. The proposed descriptor is presented in detail in Section 3. Section 4 compares the performance of our strategy with those of state-of-the-art algorithms on benchmark datasets, including VIPeR, CAVIAR4REID, i-LIDS, ETHZ, GRID, CUHK01, and Market-1501. Finally, Section 5 concludes the paper.

2. Related works

Early published work on re-identification dates back to 2003 [1]. With the development of pattern recognition, computer vision [37,26,59] and machine learning, person re-identification became a hot topic in academia and has received extensive attention from researchers since 2008 [14,33,34].

Person re-identification algorithms generally fall into two categories [1], namely, the unsupervised algorithms and the supervised algorithms. The unsupervised algorithms [2–5] rely on a robust feature description, which is a set of distinguishing characteristics that describe the appearance of pedestrians from various camera views. Typical descriptors that have been proposed include color, texture, shape, edges, and semantic attributes. The supervised algorithms [6–11] employ learning techniques for descriptor extraction and matching. They require labeled samples for training.

2.1. Unsupervised methods

Bazzani et al. [2] proposed a descriptor called Symmetry-Driven Accumulation of Local Features (SDALF). This method combines three features as a human signature, including maximally stable color regions (MSCR), weighted color histograms (WCH) and recurrent high-structured patches (RHSP). However, Cheng et al. [22] observed that the performance will not be too heavily degraded if the RHSP are removed. Bazzani et al. [18], through epitomic analysis, extracted the Histogram Plus Epitome (HPE) to describe the appearance of a person. However, this multi-shot

Please cite this article as: W. Fang et al., Perceptual hash-based feature description for person re-identification, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.07.019