Neurocomputing 212 (2016) 65–74

Contents lists available at ScienceDirect

Neurocomputing

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

Resolution adaptive feature extracting and fusing framework for person re-identification

Hua Yang [a,b,*], Xinyu Wang [a,b], Ji Zhu [a,b], Wenqi Ma [a,b], Hang Su [c]

[a] Institute of Image Communication and Network Engineering of Shanghai Jiao Tong University, Shanghai 200240, China
[b] Shanghai Key Lab of Digital Media Processing and Transmission, Shanghai 200240, China
[c] Department of Computer Science and Technology, Tsinghua University, Beijing, China

ARTICLE INFO

Article history:
Received 18 November 2015
Received in revised form
9 March 2016
Accepted 23 March 2016
Available online 5 July 2016

Keywords:
Person re-identification
Resolution adaptive framework
Feature fusing
Space constrained weight
Selected candidate

ABSTRACT

Person re-identification has gained an increasing attention in both of academic research communities and industrial labs in recent years. It remains as a challenging problem due to the extracting and matching of reliable and distinctive features under different camera views across a wide spatial and temporal scope. To address these issues, we propose a resolution adaptive method by extracting and fusing the global and local features within a unified framework. Specifically, global features and local features are extracted separately in two image scales which are constructed in a cascaded way. After extracting HS and HOG features and ranking in low scale, we choose top k percent as a candidate subset, meanwhile we obtain another person subset by LPQ (Local Phase Quantization) Face detection. The union of these two candidate subsets is used for high scale processing in which wHSV (weighted Hue Saturation Value), LSCF (Local Spatial Constrain Feature) and MSCR (Maximally Stable Colour Regions) local features are adopted. Afterwards, both the global and local features extracted are fused with an unsupervised query adaptive method, based on which person re-identification is conducted with a high accuracy. Experiments are conducted on two real world datasets: ETHZ1, 2, 3, and our own dataset from high resolution cameras in real roads and campus scenes. Experimental results demonstrate that the proposed method outperforms the conventional methods in terms of both accuracy and efficiency.

1. Introduction

Person re-identification across different views of cameras is a fundamental task in automated video surveillance. Despite many efforts have been made during the past decades, it remains as a considerable challenge for visually detecting and recognizing a person at different time and space locations. First, the resolution of current monitored cameras is not high enough to implement reliable person verification based on biometrics. Second, as the transition time between disjoint cameras varies greatly for difficult individuals, it is hard to impose accurate temporal and spatial constraints. Third, the visual appearance features, which are extracted mainly from the clothing and shapes of people, are usually weak indistinctive for matching people. To further compound this problem, a person's appearance can vary significantly from different view angles, lighting, background clutter and occlusion. Finally, low computing complexity of re-identification framework is imperative for a large video surveillance networks application.

Recent years have witnessed a sustained research in automatic re-identification challenges with the forensic analysis of video from multi-camera networks, which aims to alleviate the burden on limited human resources and augment human capabilities. Most of the algorithms and system designs typically develop computer vision and machine learning techniques to solve the problems of feature representation, model matching and context based analysis [1].

In practical surveillance applications, there exist essential issues in the person re-identification framework, e.g., camera setting (viewing angle, image resolution) and image quality (illumination change, body pose, occlusion and other large appearance variations), as is shown in Fig. 1. We observe that different image resolution scales of one person supplies information contents signified by different feature types. Generally, some global features such as colour and shape are discriminative when the resolution scale level is lower; in the high resolution images, the local features such as texture and soft biometric (like face) are more discriminative. However, computation cost is heavier for images with higher resolution given a certain set of algorithms. It is desirable to

* Corresponding author at: Institute of Image Communication and Network Engineering of Shanghai Jiao Tong University, Shanghai 200240, China.
E-mail address: hyang@sjtu.edu.cn (H. Yang).

http://dx.doi.org/10.1016/j.neucom.2016.03.096
0925-2312/

Fig. 1. Examples of challenges encountered in person re-identification. Each group represents a typical aspect of the challenges from two scales: original and 1/2 down sampling scale. Images (c) and (d) are from ETHZ dataset, images (a), (b) and (e) come from our real High Definition camera network.

develop a more efficient resolution adaptive system for person re-identification, which consists of feature selection and model matching. A competitive re-identification system should be with the discriminative power, generalization capability and scalability with low computational complexity. On the one hand, the discriminative power demands that the selective signature should be unique or as distinctive as possible to deal with the intra and inter-class variations; the generalization capability means a model can be trained once and then applied to a variety different camera configurations from different locations; and scalability with low computational complexity requires that the system can adapt a topologically complex and large camera network. On the other side, camera resolution, computational load and other implementation issues prevent or limit the applications of many specific algorithms. In this paper, we propose a novel resolution adaptive framework to address how to trade off between these two issues.

This paper focuses on effectively using the benefits of high resolution images and reducing the complexity in the foundation of improving the accuracy of re-identification. To perform re-identification under the HD monitor cameras, we propose a re-identification framework, in which global features and local features are extracted respectively in different image resolution scales based on their scale behaviours. Specifically, the global features are represented by the histogram whose matching performance is not related to the image scale directly while the local features perform better on the higher image scale since they need more feature details to evaluate similarity of image pairs. For high scale level, not all adopted local features such as face, wHSV (weighted Hue Saturation Value) [2], SURF [3] and MSCR (Maximally Stable Colour Regions) [4] are equally important or useful for re-identification. If the resolution of represented person is high enough and view point is proper enough to obtain the face part, the face feature could be the most discriminative signature as a kind of biometric feature. Otherwise, the performance will decrease if we cannot detect effective face and still take the head part for granted. So we take advantage of the unsupervised query-adaptive mechanism [5] to estimate and select the weighted local features with complementary nature, which contributes significantly to the performance of whole re-identification framework.

The contributions of this framework can be summarized as follows:

- We propose a resolution adaptive re-identification framework in which global and local features can be extracted in proper image scales based on the properties. The cascaded work flow first extracts global features on low resolution image scalearticle5, and then choose the top K percent candidate persons. Afterwards, local features processing on high resolution scale (wHSV, SURF, MSCR); and the face feature is also extracted when faces for candidates are clearly visible. The system not only benefits from High Definition cameras by exploiting more complementary information, but also improves computing efficiency of the system greatly.
- To benefit from the high resolution images and overcome large alterations in person visual appearance for long period, we adopt biometric face feature on high scale level as one of the local features to complement biometric information and give a richer signature.
- We adopt a Local Spatial Constrain Feature (LSCF) based on [6] by combining the LAB colour feature into SURF descriptor under the spatial constraint of different body parts.
- Based on unsupervised query-adaptive learning mechanism [5], we fuse local features including face LPQ [7], MSCR, wHSV, LSCF and global features (HOG and HS) according to feature effectiveness on both the high scale and low scale levels, in which no prior knowledge on the topic of the query image is provided and no large amount of labelled samples are needed.

The rest of the paper is organized as follows. In Section 2, recent related work for re-identification is analyzed and problem context is indicated. In Section 3, the details of the proposed framework which we refer to as Resolution Adaptive Framework (RAF) is described. Section 4 demonstrates the experimental