Video-Based Person Re-Identification With Unregulated Sequences

Wenjun Huang, National Engineering Research Center for Multimedia Software, Wuhan University, Wuhan, China

Chao Liang, National Engineering Research Center for Multimedia Software, Wuhan University, Wuhan, China

https://orcid.org/0000-0002-8287-8655

Chunxia Xiao, School of Computer Science, Wuhan University, Wuhan, China

Zhen Han, School of Computer Science, Wuhan University, Wuhan, China

ABSTRACT

Video-based person re-identification (re-id) has recently attracted widespread attentions because extra space-time information and more appearance cues in videos can be used to improve the performance of image-based person re-id. Most existing approaches equally treat person video images, ignoring their individual discrepancy. However, in real scenarios, captured images are usually contaminated by various noises, especially occlusions, resulting in a series of unregulated sequences. Through investigating the impact of unregulated sequences to feature representation of video-based person re-id, the authors find a remarkable promotion by eliminating noisy sub sequences. Based on this interesting finding, an adaptive unregulated sub sequence detection and refinement method is proposed to purify original video sequence and obtain a more effective and discriminative feature representation for video-based person re-id. Experimental results on two public datasets demonstrate that the proposed method outperforms the state-of-the-art work.

KEYWORDS

Adaptive Weighting, Noise Detection, Recognition, Sequence Stability Measure, Sparse Construction

INTRODUCTION

Person re-identification, which aims at identifying a person of interest among different cameras, has become increasingly popular in the community due to its critical role in many surveillance, security and multimedia applications. Currently, major efforts towards this problem focus on the still-image-based scenario, in which each person has only one image available per camera view. Many methods have been developed to either extract discriminative features (Liao et al., 2015; Matsukawa et al., 2013; Satta et al., 2013; Shen et al., 2013) or learn effective distance metric (Hirzer et al., 2012; Köstinger et al., 2012; Liang et al., 2015; Liao et al., 2015; Yang et al., 2016) for this problem.

In spite of great research progress achieved for the still-image-based task, the real-world re-id performance is hindered by limited information extracted from a single image. Such still-image-based person re-id ignores the temporal information among person images, which leads poor feature representation of person. In practical surveillance systems, persons are always recorded by videos, which means that there are multiple consecutive frames available for an individual in each camera's view field. Thus, it is intuitive to use such sequential images to improve re-id performance, which directly motivates the investigation of video-based person re-id.

Recently, impressive research progress has been reported in video-based person re-id. However, most approaches mentioned above generally assume that all images in each sequence are of equal importance, losing sight of their difference caused by the interference of various noises. Take iLIDS-VID dataset (Wang et al. 2014) in Figure 1 as an example, person images are always flooded with various noises, such as object occlusions or background clutters, resulting in highly noisy unregulated sequences. In our preliminary comparative experiment conducted on 199-pair unregulated person sequences of iLID-VID dataset, average matching accuracy on original unregulated video sequences is only 7%, ten percent lower than that obtained on filtered clean video sequences.

Occlusion is a very common case in video-based applications, taking dataset ETHZ (Schwartz et al. 2010), iLIDS-VID (Wang et al. 2014) and OTB (Wu et al., 2013) as examples, we count the number of video sequences that have heavy occlusions in these public video datasets, as shown in Table 1. From Table 1, we can see that: (1) Occlusion is universal in real world scenarios. (2) In the case of the occlusions, `long-term occlusion' (e.g., the target person is occluded by another person in the whole sequence, which makes there are no clean sub sequences of target person) accounts for only a few parts. In this paper, we focus on the `temporary occlusion' (e.g., another person passes by the target person temporarily or the target person is occluded by surrounding objects temporarily), which means that in most cases, there are still some clean sub sequences even if part of the sequence is occluded.

In this field, Wang et al. (Wang et al., 2014; Wang et al., 2016) first noticed the quality discrepancy problem of different person sequences, and an optical flow-based algorithm was raised to detect walking cycles to divide a video sequence into different sub fragments. Then, a ranking model was proposed to select and match video fragment pairs. However, (1) it is hard to obtain a reliable optical flow estimation without considering the occlusion (Ayvaci et al., 2010). Thus, the algorithm will eventually generate noisy sub sequences when occlusion occurs. (2) It uniformly exploits all video

Figure 1. An example illustrating that person images are always highly noisy in practical situations

Table 1. Occlusion condition investigation

Dataset	Occlusion	Long-Term Occlusion	Temporary Occlusion
ETHZ	45.21%	15.15%	84.85%
OTB	58.05%	10.03%	89.97%
iLIDS-VID	66.33%	10.05%	89.95%

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/video-based-person-re-identification-with-unregulated-sequences/246838

Related Content

Fingerprint Liveness Detection Based on Fake Finger Characteristics
Gian Luca Marcialis, Pietro Coliand Fabio Roli (2013). *Emerging Digital Forensics Applications for Crime Detection, Prevention, and Security (pp. 1-17).*
www.irma-international.org/chapter/fingerprint-liveness-detection-based-fake/75660

Extended Time Machine Design using Reconfigurable Computing for Efficient Recording and Retrieval of Gigabit Network Traffic
S. Sajan Kumar, M. Hari Krishna Prasadand Suresh Raju Pilli (2011). *Cyber Security, Cyber Crime and Cyber Forensics: Applications and Perspectives (pp. 168-177).*
www.irma-international.org/chapter/extended-time-machine-design-using/50721

Locally Square Distortion and Batch Steganographic Capacity
Andrew D. Ker (2011). *New Technologies for Digital Crime and Forensics: Devices, Applications, and Software (pp. 144-160).*
www.irma-international.org/chapter/locally-square-distortion-batch-steganographic/52850

Holistic Analytics of Digital Artifacts: Unique Metadata Association Model
Ashok Kumar Mohan, Sethumadhavan Madathiland Lakshmy K. V. (2021). *International Journal of Digital Crime and Forensics (pp. 78-100).*
www.irma-international.org/article/holistic-analytics-of-digital-artifacts/283128

Identifying the Use of Anonymising Proxies to Conceal Source IP Addresses
Shane Miller, Kevin Curranand Tom Lunney (2021). *International Journal of Digital Crime and Forensics (pp. 1-20).*
www.irma-international.org/article/identifying-the-use-of-anonymising-proxies-to-conceal-source-ip-addresses/279371