Engineering Science and Technology, an International Journal xxx (2017) xxx–xxx

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Engineering Science and Technology, an International Journal

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

Full Length Article

An unsupervised meta-graph clustering based prototype-specific feature quantification for human re-identification in video surveillance

Aparajita Nanda, Pankaj Kumar Sa, Suman Kumar Choudhury

Department of Computer Science and Engineering, National Institute of Technology Rourkela, Rourkela 769008, India

ARTICLE INFO

Article history:
Received 1 December 2016
Revised 4 March 2017
Accepted 9 March 2017
Available online xxxx

Keywords:
Human re-identification
Video surveillance
Prototype discovery
Clustering

ABSTRACT

Human re-identification is an emerging research area in the field of visual surveillance. It refers to the task of associating the images of the persons captured by one camera (probe set) with the images captured by another camera (gallery set) at different locations in different time instances. The performance of these systems are often challenged by some factors—variation in articulated human pose and clothing, frequent occlusion with various objects, change in light illumination, and the cluttered background are to name a few. Besides, the ambiguity in recognition increases between individuals with similar appearance. In this paper, we present a novel framework for human re-identification that finds the correspondence image pair across non-overlapping camera views in the presence of the above challenging scenarios. The proposed framework handles the visual ambiguity having similar appearance by first segmenting the gallery instances into disjoint prototypes (groups), where each prototype represents the images with high commonality. Then, a weighing scheme is formulated that quantifies the selective and distinct information about the features concerning the level of contribution against each prototype. Finally, the prototype specific weights are utilized in the similarity measure and fused with the existing generic weighing to facilitates improvement in the re-identification. Exhaustive simulation on three benchmark datasets alongside the CMC (Cumulative Matching Characteristics) plot enumerate the efficacy of our proposed framework over the counterparts.

1. Introduction

Human re-identification aims at establishing the visual correspondence between two image instances of an individual appearing at different locations at different time instances. It has a wide range of applications in video surveillance such as pedestrian tacking, theft avoidance, activity recognition, crowd monitoring and so forth. A re-identification problem usually faces a number of challenges owing to illumination change, unusual pose variation, partial occlusion, poor camera resolution, cluttered background etc; few such challenges are well described in Fig. 1. Most of the existing methods apply three variant strategies to counter the above challenges. (i) the first group of methods focus on designing discriminative feature signature that remains invariant to the above challenges [1–5]. In general, color and texture features, in terms of histogram representation, are preferred for such invariant representation [6–8]. The feature signature, thus obtained, are then combined with the patch matching schemes [9,10], reference set [11], and salience learning [12]. (ii) the second approach is based on brightness transformation function (BTF) on the visual features captured from non-overlapping camera FoV [13–15]. The current state-of-the-art schemes have shown that the BTF is not unique and varies on several other factors [16,14,17]. (iii) the third one focuses on distance metric learning which aims to learn a distance metric to transform the feature samples from original space to another feature space by exploiting the inter-class and intra-class variations [18,16]. Few popular metrics include the Mahalanobis metric learning [19], considering multiple metrics [20], equivalence constraints [18] etc.

Motivations– Scalability is the primary concern in the aforementioned approaches since they need an extensive learning of pairwise images and require the determination of the number of attributes in advance. In other words, all the aforementioned approaches focus either on finding an invariant complementary feature set or learning a distance metric in supervised manner. However, none of these schemes address the problem with visual ambiguity that may arise owing to look-alike individuals; most of these methods fail to discriminate between individuals with similar attire pattern.

E-mail addresses: aparajita1.nanda@gmail.com (A. Nanda), PankajKSa@nitrkl.ac.in (P.K. Sa), sumanchoudhury.nitr@gmail.com (S.K. Choudhury)

http://dx.doi.org/10.1016/j.jestch.2017.03.001
2215-0986/

Fig. 1. Challenging issues in the appearances of an individual captured from disjoint camera views. Images in a column denote the same person. (a) Image-pairs with viewpoint variations, (b) image-pairs with variation in appearance and illumination, (c) image-pairs with arbitrary pose change and clutter background, and (d) image-pairs with similar color attire.

In contrast to these approaches, we present a prototype specific feature quantification model to counter the above problem alongside the generic challenges. The primary idea is to first partition the gallery crowd into a number of consensus prototypes (groups) with high intra-prototype-similarity and high inter-prototype deviation. The notion of prototype formation is to reduce the miss match rate for each probe where it is being compared within its look-alike set of images by means of an appropriate weighting similarity measure; more commonality attributes are given less weight and less commonality attributes are given higher weight to find the exact match within the inlier-like images. Besides, for each probe, the prototype specific weights are combined with the existing generic weights [21,7] to facilitates the improvement in re-identification. The proposed model does not require any manual individual-labeling during prototype formation process. The contributions of this framework are threefold:

- A cluster ensemble paradigm is employed to segment the input gallery into a set of consensus prototypes; the primary concern in prototype discovery is to tackle the visual ambiguity owing to similar attire pattern in appearance, and thereby reducing the mismatch rate in recognition.
- A weighing scheme is incorporated in our framework in which a weight vector is defined for each prototype quantifying the contributions of the feature attributes. Compared with the conventional generic feature weights, our approach holds more selective and distinctive quantification strategies.
- A learning model is employed to find the closest prototype for any given probe. The weight vector of the corresponding prototype is employed in the similarity measure for finding the best possible match within the prototype-images; the existing generic feature weights are also fused in the similarity function to improve the recognition accuracy.

The paper organization is enumerated as follows. Related work on human re-identification is depicted in Section 2. Our proposed framework is detailed in Section 3. The experimental results are analyzed in Section 4 followed by conclusion in Section 5.

2. Related work

The task of human re-identification aims at matching the individuals observed across disjoint camera views. The existing works mainly focus on appearance based feature extraction and distance learning.

2.1. Feature extraction and descriptor formation

The major concern in human re-identification by image matching is to combine various appearance based features [1,6,7,22,23,21,24]. For example, Farenzena et al. accumulate the local features such as the weighted color histograms, maximally stable color regions, and structured patches for the formation of feature descriptor [1]. The ensemble of spatial and color features are integrated and the feature subset is optimized by boosting technique to improve the discriminative power of the descriptor [6]. Wang et al. introduce the shape and appearance context with the color histograms to capture the structural information [22]. Bak et al. propose the stable region based statistical informations of color and gradients to construct a covariance descriptor [25,26]. Similarly, the set of feature descriptors such as color, texture, and shape are combined to form the feature descriptor [27]. Ma et al. consider seven dimensional local features including color, texture and spatial structure to represent an image and the distribution of the features are modeled by Gaussian model [3]. In another work, Ma et al. combine the biologically inspired features and the covariance descriptor into a single representation [4]. In recent work, Wu et al. formulate a re-identification model by integrating the camera viewpoint and human pose information [5]. Subsequently, Lisanti et al. propose an approach based on iterative re-weighted sparse model by considering the color histogram where the pixel's distributions is weighted by a kernel centered on the image [2].

There exist some more methods that consider the salient parts of individual's image for robust re-identification [28,1,29]. For instance, Cheng et al. consider the human salient body parts to provide visual correspondence [29]. Satta et al. introduce the multiple components matching by considering the different body parts and it is further extended to faster matching re-identification [30]. In a recent work, the salient body parts are represented with patches of cluster [31]. Similarly, Liu et al. investigate on the importance of features with respect to visual circumstances where selective features are weighted to represent the discriminative nature in appearance [8]. Xie et al. formulate the visual spatial learning and incorporate the region based matching [32]. Martinel et al. propose a kernelized graph-based approach to represent the salience points on the person's image and used as weights in feature extraction process [33]. The above approaches are concerned with the spatial feature importance within a single image. Hence, selecting the significant body parts are important than the background region.

Please cite this article in press as: A. Nanda et al., An unsupervised meta-graph clustering based prototype-specific feature quantification for human re-identification in video surveillance, Eng. Sci. Tech., Int. J. (2017), http://dx.doi.org/10.1016/j.jestch.2017.03.001