Pattern Recognition 47 (2014) 1602–1615

Contents lists available at ScienceDirect

Pattern Recognition

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

On-the-fly feature importance mining for person re-identification

Chunxiao Liu [a,*], Shaogang Gong [b], Chen Change Loy [c]

[a] Tsinghua University, China
[b] Queen Mary University of London, United Kingdom
[c] The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 6 April 2013
Received in revised form
4 October 2013
Accepted 1 November 2013
Available online 13 November 2013

Keywords:
Person re-identification
Unsupervised salience learning
Feature importance
Random forest

ABSTRACT

State-of-the-art person re-identification methods seek robust person matching through combining various feature types. Often, these features are implicitly assigned with generic weights, which are assumed to be universally and equally good for all individuals, independent of people's different appearances. In this study, we show that certain features play more important role than others under different viewing conditions. To explore this characteristic, we propose a novel unsupervised approach to bottom-up feature importance mining on-the-fly specific to each re-identification probe target image, so features extracted from different individuals are weighted adaptively driven by their salient and inherent appearance attributes. Extensive experiments on three public datasets give insights on how feature importance can vary depending on both the viewing condition and specific person's appearance, and demonstrate that unsupervised bottom-up feature importance mining specific to each probe image can facilitate more accurate re-identification especially when it is combined with generic universal weights obtained using existing distance metric learning methods.

1. Introduction

A critical task in visual surveillance is to automatically associate individuals across different disjoint and large spaces at different times, known as re-identification, in order to facilitate cross-camera tracking of people and understanding their global behaviour in a wider context [21]. Typically, when a target (a probe) is observed in a view, the goal of person re-identification (re-id) is to discover the same person that appears at an arbitrary location and time from a crowd of people (gallery candidates) based on their appearance similarity to the probe image. Appearance-based person re-identification is a non-trivial problem owing to visual ambiguities and uncertainties caused by illumination changes, viewpoint and pose variations, and inter-object occlusions. To address this problem, most existing methods [9,5] combine different appearance features, such as colour and texture, to improve reliability and robustness in person matching.

Often, each type of visual features is represented by a bag-of-words scheme in the form of a histogram. Feature histograms are then concatenated with some weighting between different types of features in accordance to their perceived *importance*, i.e. based on some empirical assumed discriminative power of certain type of features in distinguishing the visual appearance of an individual from the others [25,31,23,10,12]. Moreover, an implied assumption for choosing a generic feature weighting scheme is that the underlying features used are also tolerant/invariant to camera view changes. To accommodate such feature importance selection criteria, existing techniques implicitly assume a feature weighting or a selection mechanism that is *generic*, by imposing weights (or a linear weight function) on certain feature types that are considered optimal in a universal sense, e.g. colour may be considered as the most stable and universally good (therefore more important) feature for discriminating people in crowded spaces subject to frequent occlusion and unknown viewpoint changes, rather typical re-identification scenarios. In this study, we refer such universal feature weights selection schemes as learning *top-down Generic Feature Importance* (GFI). They can be learned either through boosting [10], rank learning [25,28], or distance metric learning [31,12,23,14].

Human often relies on salient features for distinguishing one from the others, i.e. using the plaid pattern on the shirt to distinguish the man from the woman wearing red sweater in Fig. 1. Such bottom-up feature saliency is valuable for person re-identification but is often too subtle to be captured when computing feature importance using existing top-down GFI techniques. In this study, we propose a new and interesting perspective for person re-identification based on unsupervised feature importance mining. In particular, we investigate a different notion of

* Corresponding author.
E-mail addresses: lcx08@mails.tsinghua.edu.cn (C. Liu), sgg@eecs.qmul.ac.uk (S. Gong), ccloy@ie.cuhk.edu.hk (C.C. Loy).
URLS: http://www.eecs.qmul.ac.uk/~sgg/ (S. Gong), http://personal.ie.cuhk.edu.hk/~ccloy/ (C.C. Loy).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.11.001

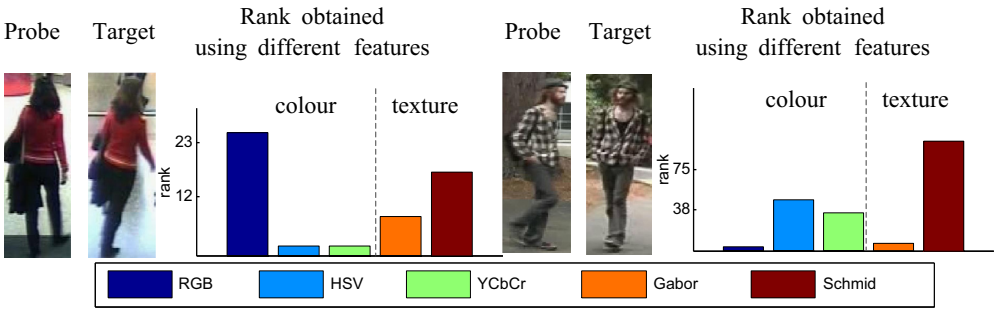

Fig. 1. A probe image and the target image, together with the rank of correct matching by using different feature types separately. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

feature importance in comparison to existing re-id studies, i.e. the discriminative power of intrinsic appearance attributes unique to each individual. We consider that certain appearance features can be more important than others in describing an individual and distinguishing him/her from other people. For instance, colour is more informative to describe and distinguish an individual wearing textureless bright red sweater, but texture information can be equally or more critical for a person wearing plaid shirt (Fig. 1). Hence, it is desired not to bias all the weights to some universally good features that are assumed to be stable for re-identifying all individuals. Instead, we wish to investigate an approach to *selectively distribute weights to person probe image specific feature subset given different appearance attributes of different people*.[1]

There are two clear distinctions between the conventional top-down and the proposed bottom-up feature importance mining. First, the conventional top-down GFI methods are supervised, i.e. the learning process requires exhaustive supervision on pairwise individual correspondence between camera pair. In contrast, the proposed bottom-up feature importance mining is fully unsupervised, i.e. without requiring manually labelled person identities in the training process. Second, the conventional top-down approach imposes weights on certain feature types that are considered optimal in a universal sense; while the bottom-up approach aims to discover a set of discriminative features and quantify their importance specific to each individual. From another perspective, the notion of bottom-up learning can also be interpreted as a process of unsupervised discovering latent attribute (see Section 3.1), which is largely different from existing top-down supervised attribute learning [16,15] that requires exhaustive human-specified attributes.

Formulating an unsupervised and on-the-fly importance sampling method for person re-identification is non-trivial. Firstly, what is unique or salient about a person against a large and dynamic crowd of people is somewhat difficult and subjective to quantify under different circumstances. Secondly, simultaneously identifying any and all salient features specific to each individual can be computationally prohibitive. Lastly, a model is required to not only discover a set of probe-specific important (salient) features, but also quantify automatically the importance of each feature type.

In this study, we investigate what features are more important for person re-identification under significantly changing viewing conditions. In particular, we show that selecting features adaptively for different individuals yield more robust re-identification performance than feature histogram concatenation with uniform weighting [27,21]. Motivated by this observation, we formulate a fully unsupervised approach to on-the-fly bottom-up feature importance mining driven by learning to classify the probe person's appearance attributes. Two methods for computing the bottom-up feature importance are proposed and evaluated: *Prototype-Specific Feature Importance* (PSFI) and *Individual-Specific Feature Importance* (ISFI).

To avoid a potentially prohibitive feature importance mining process, our model is designed to first discover, by unsupervised clustering, inherent visual appearance attribute *prototypes*, in order to yield more meaningful and compact groupings of image samples of different people in a training pool. From this unsupervised learning of appearance attribute based prototypes, we formulate a principled method to quantify bottom-up feature importance specific to each probe image re-identification based on introducing an error gain criterion from classifying the probe image by learned attribute prototypes using a random forest.

The contributions of this study are two-fold:

1. While most existing person re-identification methods focus on supervised top-down feature importance learning, we provide empirical evidence to support the view that some benefits can be gained from unsupervised bottom-up feature importance mining guided by a person's appearance attribute classification. To the best of our knowledge, this is the first study that systematically investigates the role of different feature types in relation to appearance attributes for person re-identification.
2. We formulate a novel unsupervised approach for on-the-fly mining of person appearance attribute-specific feature importance. Specifically, we introduce the concept of learning grouping of appearance attributes for guiding bottom-up feature importance mining. Moreover, we define an error gain based criterion to systematically quantify feature importance for the process of re-identification of each specific probe image.

Extensive experiments conducted on three benchmarking re-identification datasets demonstrate that person re-identification can benefit from complementing existing supervised learning based top-down generic feature importance weighting approaches with the unsupervised learning based bottom-up feature importance mining approach investigated in this study.

2. Related work

Person re-identification is typically defined as the task of matching and ranking pedestrian across non-overlapping camera views. This task is related to the tracking-by-identification problem [22,6], which aims to re-identify people across trajectory fragments in multiple cameras with overlapping fields of view. Often, person-specific appearance and motion cues are exploited for tracks association to prevent identity switches. In this study, we focus on person re-identification across non-overlapping views.

[1] Similar to that of Layne et al. [16], we refer attributes as appearance characteristics of individuals, e.g. dark shirt, blue jeans, carrying-object, backpack.