Pattern Recognition ■ (■■■■) ■■■–■■■

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

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

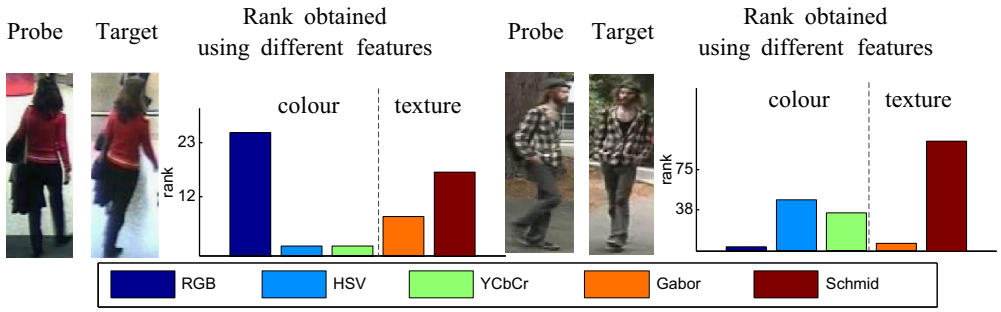

Fig. 1. A probe image and the target image, together with the rank of correct matching by using different feature types separately. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

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.

Please cite this article as: C. Liu, et al., On-the-fly feature importance mining for person re-identification, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.11.001

Person re-identification by image matching can benefit from integrating several types of visual features [9,5,25,31,20,10,27,1,26,2]. For instance, Farenzena et al. [9] combine weighted colour histogram, maximally stable local colour regions and structured patches for constructing a feature descriptor. Bazzani et al. [5] propose histogram plus epitome features as a human signature. Bak et al. [3] and Alahi et al. [1] combine local statistics of colour and gradient to construct a covariance descriptor. Wang et al. [27] introduce shape context along with colour histograms to capture more structural information.

Given the host of available appearance representations of colour, texture and shape, most existing distance metric learning based person re-identification methods take a GFI learning strategy [25,31,20,23,10,28,14]. Essentially, such techniques assume that certain features are universally more important in all circumstances, regardless of viewing condition changes between gallery and probe images and the specific visual appearance characteristics of a re-identification target person in the probe image. For example, the RankSVM method by Prosser et al. [25] aims to find a linear function to weight the absolute difference of samples via optimisation given pairwise relevance constraints. The Probabilistic Relative Distance Comparison (PRDC) model of Zheng et al. [31] maximises the probability of a pair of true match having a smaller distance than that of a wrong matched pair. The output is an orthogonal matrix that essentially encodes the universal importance of each feature. Then the learned feature importance is used universally for all the probe images.

There are other methods that extract important (salient) parts of a person for robust matching [19,9,8]. For instance, Farenzena et al. [9] select salient parts of a body figure by symmetry; Cheng et al. [8] exploit human salient body parts to enable more accurate visual correspondence. Their consideration of importance is different from ours in this study. In the aforementioned methods, the feature importance mining is spatial and confined within a single image, e.g. selecting certain body parts as important rather than the background region. In this study, we aim to discover unique visual properties of a person, not within an image, but relative to a dynamic crowd of people under unknown changing viewing conditions between different camera locations.

The method proposed by Schwartz and Davis [26] shares a similar spirit to our work, i.e. it aims to discover what is important given specific appearance. In contrast to the model of Schwartz and Davis [26] that requires labelled images to discover feature importance for a close-set of appearances, our method is fully unsupervised. Importantly, the proposed approach in this study is more adaptable in principle due to that the feature importance is mined by unsupervised learning of appearance attribute prototypes. This approach is designed not only to discover the feature importance from a training dataset off-line, but also to readily allow for computing feature importance on-the-fly given a specific probe image for re-identification.

3. Quantifying feature importance for Re-ID

A diagram that summarises our approach for bottom-up feature importance mining is depicted in Fig. 2. To address the challenge of both avoiding prohibitive feature importance mining from a training dataset and providing adaptive per probe specific feature importance selection on-the-fly, we formulate a novel method based on a cascaded clustering-classification random forest.

Specifically, in the training stage, a clustering forest is first employed to discover latent manifold clusters from a large set of unlabelled training images (Fig. 2(c)–(e)). The discovered clusters are considered as feature *prototypes*, which correspond to a visually meaningful set of appearance attributes. These prototypes are exploited to facilitate unsupervised bottom-up feature importance mining. The prototype discovery is critical that it avoids exhaustive search of feature importance against all the training images given in a probe image. Instead, it facilitates mining feature importance in a much smaller number of representative prototypes. To mine the feature importance of each prototype (Fig. 2(f), (g)), we formulate a classification forest to quantify the relevance of a feature variable to a prototype by examining its error gain in an information theoretic sense.

In the process of re-identifying a probe image, our method determines on-the-fly the bottom-up feature importance for the given probe image according to its mixture of prototype memberships inferred by a classification random forest.

3.1. Prototypes discovery

The first step of our method is to cluster a given set of unlabelled images into several representative *prototypes*, each of which composes images that are most likely to correspond to similar constitutions of multiple classes of appearance attributes, e.g. wearing colourful shirt, with backpack, dark jacket (Fig. 2(e)).

Formally, given an input of n unlabelled images $\{I_i\}$, where $i=1,\ldots,n$, feature extraction $f(\cdot)$ is first performed on every image to extract a D-dimensional feature vector, that is $f(I)=\mathbf{x}=(x_1,\ldots,x_D)^{\mathrm{T}}\in\mathbb{R}^D$ (Fig. 2(b)). We aim to discover a set of prototypes

$$c\in\mathcal{C}=\{1,\ldots,K\}, \tag{1}$$

i.e. low-dimensional manifold clusters that aim at grouping images $\{I\}$ with similar appearance attributes. Note that this unsupervised feature prototype discovery process is critical for enabling tractable feature importance mining (Section 3.2). In particular, performing an exhaustive feature importance mining against n images has a complexity of $O(n^2)$, while our approach takes $O(K^2)$ given K prototypes, where $K\ll n$.

We treat the prototype discovery problem as a graph partitioning problem, which requires us to first estimate the pairwise similarity between images and construct a similarity matrix for the training dataset. For addressing this problem, instead of using conventional Euclidean distance based similarity measure, we exploit a clustering random forest of a cascaded model for similarity matching [7,17]. This is because that a clustering forest can (1) avoid manual definition of distance function since the pairwise affinities are defined by the tree structure itself, and (2) select implicitly and automatically optimal features via optimisation of the well-defined clustering information gain function [7]. This property is desired to ensure noisy and possibly redundant feature variables to play a lesser role in constructing the pairwise similarity matrix, also referred to as an affinity matrix in the following.

A clustering forest is an ensemble of T_{cluster} clustering trees (Fig. 2(c)). Each clustering tree t defines a partition of the input samples $\mathbf{x}$ at its leaves, $l(\mathbf{x}):\mathbb{R}^D\rightarrow\mathcal{L}\subset\mathbb{N}$, where l represents a leaf index and $\mathcal{L}$ is the set of all leaves in a given tree. For each tree, we compute an $n\times n$ affinity matrix A^t, with each element A^t_{ij} defined as

$$A^t_{ij}=\exp^{-\mathrm{dist}^t(\mathbf{x}_i,\mathbf{x}_j)}, \tag{2}$$

where

$$\mathrm{dist}^t(\mathbf{x}_i,\mathbf{x}_j)=\begin{cases}0 & \text{if } l(\mathbf{x}_i)=l(\mathbf{x}_j)\\ \infty & \text{otherwise}\end{cases}. \tag{3}$$

Following Eq. (3), we assign the closest affinity=1 (distance=0) to samples $\mathbf{x}_i$ and $\mathbf{x}_j$ if they fall into the same leaf node, and affinity=0 (distance=∞) otherwise. To obtain a smooth forest

Please cite this article as: C. Liu, et al., On-the-fly feature importance mining for person re-identification, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.11.001

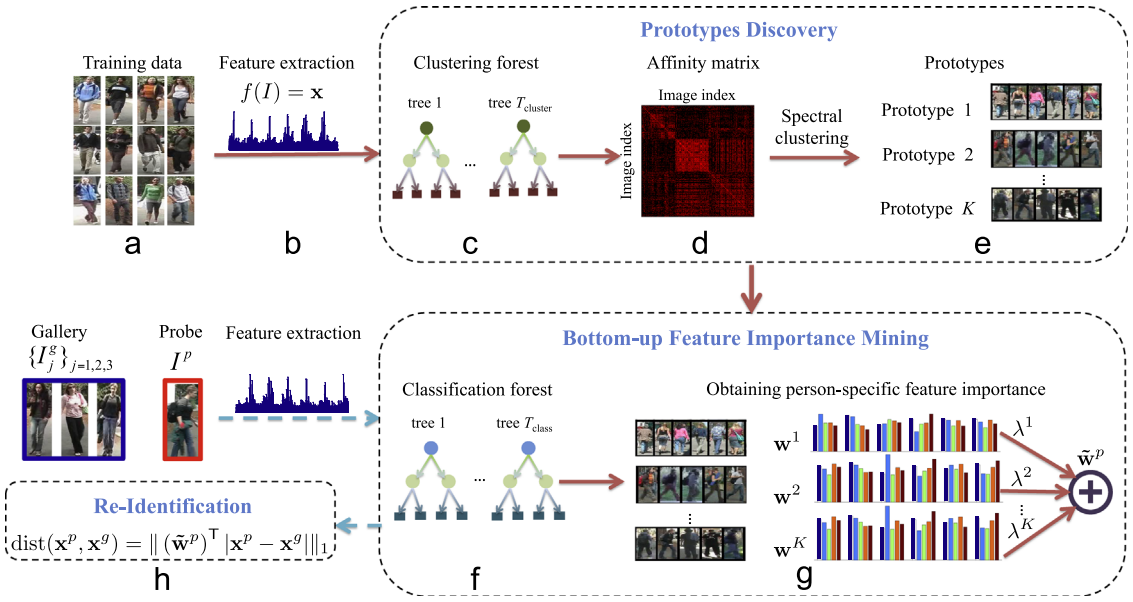

Fig. 2. An overview of the proposed bottom-up feature importance mining approach for person re-identification. Training steps are indicated by red solid arrows and testing steps are denoted by blue slash arrows. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

affinity matrix, we compute the final affinity matrix as

$$A = \frac{1}{T_{\text{cluster}}} \sum_{t=1}^{T_{\text{cluster}}} A^t, \tag{4}$$

and construct a normalised affinity matrix as

$$L = D^{-1/2} A D^{-1/2}, \quad \text{where } D_{ii} = \sum_{j=1}^{n} A_{ij}. \tag{5}$$

We adopt a self-tuning spectral clustering method [24] to partition the weighted graph into K prototypes, with the model order K being estimated automatically through analysing the eigenvectors of the normalised affinity matrix L [24,29]. Subsequently, each unlabelled training probe image $\{I_i\}$ is assigned as a member of a prototype c_i as shown in Fig. 2(e). More examples of prototypes are given in Figs. 6–8.

It is worth pointing out that there is no guarantee that in clustering only a single cluster or prototype contains a particular appearance attribute. The reason is that we are characterising persons whose appearance is likely to be partially similar with others. Therefore we do not expect the automatically discovered clusters or prototypes to be totally different at each member. In practice, we would discover a few distinct clusters, each of which contains members that are consistent in appearance. Inevitably, we would also obtain some other clusters that house images which are less representative in a given dataset. As such, the purity of the obtained classes cannot be guaranteed. Nevertheless, empirically we found that the presented application is not sensitive to the uniqueness and purity constraints.

3.2. Quantifying feature importance of prototypes

As discussed in Section 1, unlike the generic feature importance that is assumed to be universally good for all people under all viewing conditions, bottom-up probe-specific feature importance is designed to be specific to a person characterised by his/her unique appearance attributes undergone viewing condition changes. To achieve that, we first compute the feature importance revealed for each prototype driven by the shared attributes among the images clustered into this prototype. Then we determine for a given probe image its feature importance according to its mixture of memberships among the prototypes (Fig. 2(g)).

We consider that each prototype c has its own attribute-sensitive weighting $\mathbf{w}^c = (w_1^c, \ldots, w_D^c)^\top$, of which high value is assigned to unique features of that prototype. For example, in the first prototype shown in Fig. 2(e), colour features gain higher weights, reflecting higher feature importance, than others since the members in the prototype exhibit richer appearance with bright colour but relatively lesser expression in texture as compared to other prototypes. It is not difficult to see that allocating higher weights to the colour features allows us to better distinguish this prototype from the others.

Based on this principle, we wish to compute the importance of a feature according to its ability in discriminating different prototypes. Specifically, we train a classification random forest [7] using $\{\mathbf{x}\}$ as inputs and treating the associated prototype labels $\{c\}$ as classification outputs (Fig. 2(f)). For each tree t, we reserve $\frac{1}{3}$ of the original training data as out-of-bag (oob) validation samples. First, we compute the classification error $\varepsilon_d^{c,\,t}$ for every dth feature in prototype c. Then we randomly permute the value of the dth feature in the oob samples and compute the $\tilde{\varepsilon}_d^{c,\,t}$ on the perturbed oob samples of prototype c. The importance of the dth feature of prototype c is then computed as an error gain

$$w_d^c = \frac{1}{T_{\text{class}}} \sum_{t=1}^{T_{\text{class}}} (\tilde{\varepsilon}_d^{c,\,t} - \varepsilon_d^{c,\,t}), \tag{6}$$

where T_{class} is the total number of trees in the classification forest. Higher value in $w_d{}^c$ indicates higher importance of the dth feature in prototype c. Intuitively, the dth feature is important if perturbing its value in the samples causes a drastic increase in classification error gain, which suggests its critical role in discriminating different prototypes.

3.3. On-the-fly feature importance inference

In the previous step we compute the feature importance $\mathbf{w}^c$ for each prototype c but not for a specific individual's probe image. In this section, we explain our approach for computing the

Please cite this article as: C. Liu, et al., On-the-fly feature importance mining for person re-identification, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.11.001

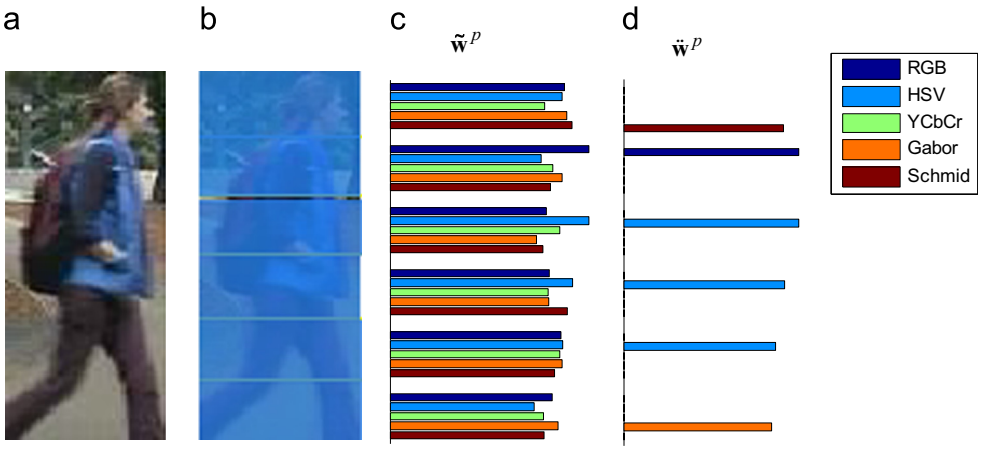

Fig. 3. An example to show maximal-weight selection. (a) Input image; (b) strip partition; (c) bottom-up feature importance; (d) after maximal-weight selection.

bottom-up feature importance for an unseen probe image on-the-fly driven by its appearance (Fig. 2(h)).

Firstly, we extract a feature vector, $\mathbf{x}^p$ to represent the unseen image. We then determine its mixture of memberships to prototypes, $\boldsymbol{\lambda}=(\lambda^1,\ldots,\lambda^K)^{\mathsf{T}}$ by classifying $\mathbf{x}^p$ using the classification forest learned in the previous step (see Section 3.2).

$$\lambda^c=\frac{1}{T_{class}}\sum_{t=1}^{T_{\text{class}}}p_t(c|\mathbf{x}^p), \tag{7}$$

where $p_t(c|\mathbf{x}^p)$ represents the posterior of tree t. The λ^c represents the average frequency of the given $\mathbf{x}^p$ being assigned to prototype c across all the trees. The mixture of prototype memberships realistically reflects the fact that each individual is a multi-mode composition of various visual characteristics.

Given the mixture of prototype memberships, $\boldsymbol{\lambda}$, we propose two methods to obtain the bottom-up feature importance for the unseen probe image:

Prototype-Specific Feature Importance (PSFI)[2]: The feature importance of $\mathbf{x}^p$ is defined as follows:

$$\tilde{\mathbf{w}}^p=(\mathbf{w}^{c^*}|c^*=\underset{c\in\{1,\ldots,K\}}{\arg\max}\lambda^c). \tag{8}$$

Individual-Specific Feature Importance (ISFI): The feature importance of $\mathbf{x}^p$ is defined as follows:

$$\tilde{\mathbf{w}}^p=\sum_{c=1}^{K}\lambda^c\mathbf{w}^c. \tag{9}$$

In PSFI an individual is assumed to be directly associated to one single prototype. This assumption may be too restrictive since different individuals would have different appearance attributes though the differences can be subtle. The ISFI relaxes this assumption. Specifically, each person is allowed to hold different degrees of membership to all the prototypes. In comparison to PSFI, the ISFI offers further intuitions about what features are unique for each specific individual.

3.4. Feature importance in re-identification ranking

To obtain the matching ranks of $\mathbf{x}^p$ against a gallery of images, we compute a feature importance weighted $\ell 1-$norm distance between $\mathbf{x}^p$ and a feature vector of the jth gallery image $\mathbf{x}_j^g$ as follows:

$$\text{dist}(\mathbf{x}^p,\mathbf{x}_j^g)=\|(\tilde{\mathbf{w}}^p)^{\mathsf{T}}|\mathbf{x}^p-\mathbf{x}_j^g|\|_1, \tag{10}$$

where $\tilde{\mathbf{w}}^p$ is computed by either Eq. (8) or (9). The ranks are obtained by sorting $\text{dist}(\mathbf{x}^p,\mathbf{x}_j^g)$ in an ascending order, that is a smaller distance which results in a higher rank (higher visual similarity).

3.5. Fusion with generic feature importance

Contemporary methods [25,31] learn a generic weight function a priori (i.e. off-line) assuming the stability of feature elements across cameras. We now investigate possible benefit in improving re-identification accuracy from the fusion of the proposed bottom-up feature importance vector $\tilde{\mathbf{w}}^p$, and a top-down generic feature weight matrix $\mathbf{V}$ obtained from [25,31].

The main objective of fusion is to combine the benefits of both approaches. In particular, the top-down approach is capable of capturing the global environmental viewing condition changes which cannot be derived from the unsupervised bottom-up method discussed so far; whereas the proposed bottom-up approach discovers valuable salient information specific to individual.

To take advantages of both approaches, we adopt a weighted sum method as follows:

$$\text{dist}_{\text{fusion}}(\mathbf{x}^p,\mathbf{x}^g)=\alpha\|(\ddot{\mathbf{w}}^p)^{\mathsf{T}}|\mathbf{x}^p-\mathbf{x}^g|\|_1+(1-\alpha)\|\mathbf{V}^{\mathsf{T}}|\mathbf{x}^p-\mathbf{x}^g|\|_1, \tag{11}$$

where α is a parameter that controls the weight between the top-down and bottom-up feature importances, and $\ddot{\mathbf{w}}^p$ is a post-processing of $\tilde{\mathbf{w}}^p$. The details are given in the following paragraph.

We observe that instead of using the original weight values of $\tilde{\mathbf{w}}^p$ as it is for fusion, selectively keeping its high weights while suppressing the less prominent weights leads to a more robust fusion. The reason of doing this is intuitive: (1) preserving the most salient features that are stable across camera views, and (2) suppressing the weights of the remaining features in $\tilde{\mathbf{w}}^p$ allows us to discard less discriminative features during fusion, so that their weighting can be fully handled by top-down generic feature weight matrix $\mathbf{V}$, which is more robust in coping with global viewing condition changes. To that end, we employ a maximal-weight selection function $\mathbb{M}$ to automatically adapt the weight values of $\tilde{\mathbf{w}}^p$, that is $\mathbb{M}:\tilde{\mathbf{w}}^p\rightarrow\ddot{\mathbf{w}}^p\in\mathbb{R}^D$. In particular, for each spatially local segment of a person image (see Fig. 3), we retain the feature channel with the largest weight, while suppress the weight values of other feature channels to 0 in $\tilde{\mathbf{w}}^p$. Note that the maximal-weight selection is not performed when using PSFI/ISFI alone without the merits from supervision.

We shall show in the following experiments that such a combined feature importance distance measure can improve both unsupervised bottom-up feature importance mining from on-the-fly individual visual appearance changes and supervised top-down generic feature importance weighting learned off-line from a labelled dataset between camera views.

[2] This method was presented in our earlier version of this work [18].

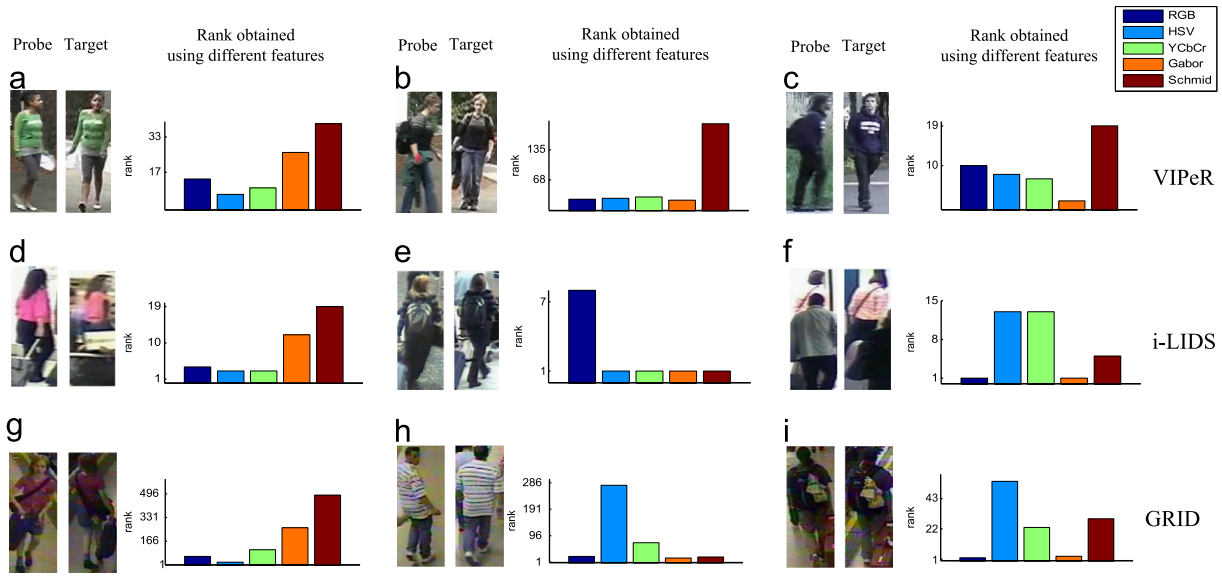

Fig. 4. In each subfigure, we show a probe image and the groundtruth target image, together with the rank of correct re-identification matches by using different isolated feature types respectively.

4. Experiments

A primary aim of this study is to investigate what features are important in different circumstances – a comprehensive evaluation on this is presented in Section 4.2. Next, we present the results from automatic unsupervised prototype discovery in Section 4.3. We then compare in Section 4.4 different feature importance measures computed by the unsupervised bottom-up solution and two top-down generic feature importance weighting methods, which are RankSVM [25] and PRDC [31]. And we evaluate the fusion of the bottom-up and generic feature importance mining methods in Section 4.5. Finally we present the sensitivity tests of several critical parameters in Section 4.6.

4.1. Experimental settings

Datasets: Three publicly available person re-identification benchmark datasets were used for our experiments, including VIPeR [10], i-LIDS Multiple-Camera Tracking Scenario (i-LIDS) [30] and QMUL underGround Re-IDentification (GRID) [21].

- *VIPeR*: The dataset contains 632 persons, each of which has two images captured in outdoor views. The dataset is challenging due to drastic appearance difference between most of the matched image pairs caused by viewpoint variations and large illumination changes at outdoor environment.
- *i-LIDS*: The dataset was captured in a busy airport arrival hall using multiple cameras. It contains 119 people with a total of 476 images, with an average of four images per person. Apart from the illumination changes and pose variations, many images in this dataset are also subject to severe inter-object occlusion.
- *GRID*: The challenging GRID dataset was captured from 8 disjoint camera views installed in a busy underground station. It is divided into a probe and a gallery sets. The probe set contains 250 person, while the gallery set contains 1025 person in which an additional 775 persons were collected who do not match any images in the probe set. The dataset is challenging due to severe inter-object occlusion and large viewpoint variations.

Feature representation: We employ a mixture of colour and texture histograms similar to those employed in [25,31,23,18,12]. Specifically, we divide an image of a person equally into six horizontal stripes to roughly capture the head, upper and lower torsos, and leg regions (see Fig. 3(b)). Alternatively, one can segment an image into patches with smaller size to conduct a finer-scale part-based search. We consider 8 colour channels (RGB, HSV and YCbCr)[3] and 21 texture filters (8 Gabor filters and 13 Schmid filters) applied to the luminance channel. Then in each stripe feature extracted from each channel is represented by a 16-dimensional histogram. Concatenating all the feature channels results in 2784-dimensional feature vector for each image. Note that our method is not restricted to the aforementioned feature representation. Other more elaborative features can be readily used, such as the covariance feature [3,4], Haar [11], maximally stable colour regions [9], and epitome features [5].

Evaluation: For each dataset, we select images of p person to build the test set, and the remaining as validation and training partitions. In the test set of each trial, we choose one image from each person randomly to set up the test gallery set and the remaining images are used as probe images. The testing process is as follows: given a probe set and a gallery set, each image of the probe set is matched with the images of the gallery. Thus, a ranking for every image in the gallery with respect to the probe is obtained.

We quantify re-identification performance using three standard measures, i.e. matching rate at rank-r, cumulative matching characteristic (CMC) curve [10], and area under the CMC curve (AUC). Matching rate at rank r measures the expectation of finding the correct match in the top r matches. The CMC curve plots this value for all r and AUC summarises the curve: higher AUC is better. In our experiments all reported performance is averaged over 10 trials.

4.2. Comparing feature effectiveness for Re-ID

We consider that certain features are more important than others in describing an individual and distinguishing him/her from other people. To validate this hypothesis, we analyse the matching performance of using different features individually.

[3] Since HSV and YCbCr share similar luminance/brightness channel, dropping one of them results in a total of 8 channels.

We first provide some visual examples in Fig. 4 (also presented in Fig. 1) to compare the ranks returned by using different feature types. It is observed that no single feature type is able to constantly outperform the others. For example, for individuals wearing textureless but colourful and bright clothing (Fig. 4(a), (d) and (g)), the colour features yielded a higher rank. For person wearing clothing with rich texture (Fig. 4(b), (e), (f) and (h)), logo (e.g. Fig. 4(c)) or backpack (e.g. Fig. 4(i)), texture features especially the Gabor features tend to dominate. These examples indicate that certain features can be more informative than others given different appearance attributes.

A more complete evaluation of the effects of different features on re-identification performance is presented in Fig. 5. In general, HSV and YCbCr colour features exhibit very close performances, and are much superior over all other features. The Schmid texture feature is least effective when used alone. This observation of colours being the most informative features supports similar conclusions drawn from early studies [10].

One may consider concatenating all the features together, assuming that different features could complement each other leading to better performance. Nevertheless, we found that naively concatenating all the feature histograms with uniform (identical) weighting does not necessarily yield a better performance, and sometimes even worse than using a single feature type, as shown by the 'Concatenated Features' performance in Fig. 5. These results suggest that a more selective feature weighting is necessary based on the level of informative of each feature variable.

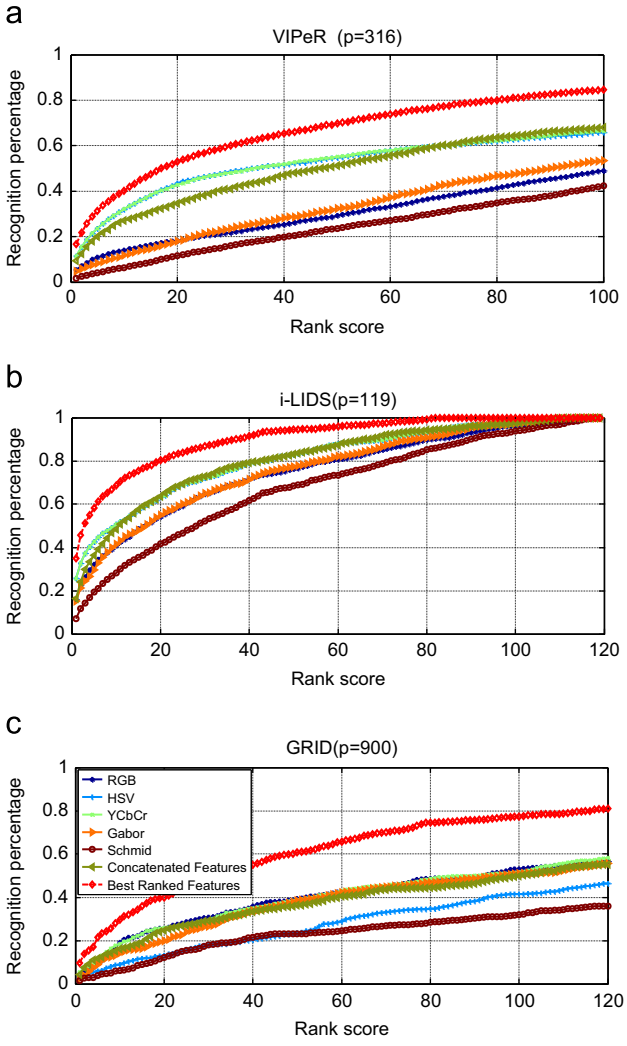

Fig. 5. The CMC performance comparison of using different features on the VIPeR, i-LIDS, and GRID datasets. 'Concatenated Features' refer to the concatenation of all feature histograms with uniform (i.e. identical) weighting. In the 'Best Ranked Features' strategy, ranking for each individual is selected based on the best feature that returned the highest rank during matching.

In the 'Best Ranked Features' strategy, the final rank is obtained by automatically selecting the best feature that returned the highest rank for each individual, e.g. selecting HSV feature for Fig. 4(a) while choosing Gabor feature for Fig. 4(c). As expected, the 'Best Ranked Features' strategy yields the best performance, i.e. 13.97%, 11.31%, and 14.31% improvement of AUC on the VIPeR, i-LIDS, and GRID datasets, respectively, in comparison to the 'Concatenated Features'.

This verification demonstrates that for each individual in most cases there exists certain type of features (or the 'Best Ranked Feature') which can achieve a high rank, and selecting such 'Best Ranked Feature' is critical to a better matching rate. Based on the analysis from Fig. 4, these 'Best Ranked Features' generally show consistency with the appearance attributes for each individual. Therefore, the results suggest that the overall matching performance can potentially be boosted by weighting features selectively according to the inherent appearance attributes.

4.3. Evaluation of prototype discovery

To weigh features selectively in accordance to the individual appearance attributes and to achieve efficient bottom-up feature importance mining, our method first discovers prototypes, i.e. low-dimensional manifold clusters that model similar appearance attributes.

To enrich the diversity of appearance characteristics available for more robust prototype discovery, we borrow additional unlabelled samples from different data sources so that the training set size of each dataset achieves 700. The additional images of VIPeR and i-LIDS are borrowed from each other to make up the balance, since the illumination and viewpoint of both datasets are similar. For the GRID dataset, the additional images are obtained from other camera views in the same underground station. Different datasets inherently contain different number of prototypes. Our method described in Section 3.1 automatically discovers the numbers as 11, 12, and 11 for VIPeR, i-LIDS, and GRID respectively. We set the number of trees in our model as $T_{\text{cluster}} = T_{\text{class}} = 200$. The minimum forest node size, which implicitly influences the depth of each tree in the forest, is set to 1. From our sensitivity tests presented in Section 4.6, we observe that the final re-identification performance is not sensitive to the setting of these forests' parameters.

Some examples of prototype discovered on the VIPeR, i-LIDS, and GRID datasets are depicted in Figs. 6, 7, and 8 respectively. Each colour-coded row represents a prototype. A short list of possible attributes discovered in each prototype is given in the figure caption. Note that these inherent attributes are neither pre-defined nor pre-labelled, but automatically discovered by the unsupervised clustering forest in our cascaded model. As shown by the example members in each prototype, images with similar attributes are categorised into the same cluster. For instance, a majority of members in the 5th prototype of VIPeR can be characterised with bright and high contrast colour appearance. In the first prototype of VIPeR, the key attributes are 'carrying backpack' and 'side pose'. A similar visual consistency in prototype can be observed in the i-LIDS and GRID datasets. Note that some prototypes, however, have lower purity as they also accommodate images whose appearance is less representative and frequent in the dataset.

In general, the results demonstrate that our method is capable of generating reasonably good clusters of inherent attributes,

Please cite this article as: C. Liu, et al., On-the-fly feature importance mining for person re-identification, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.11.001

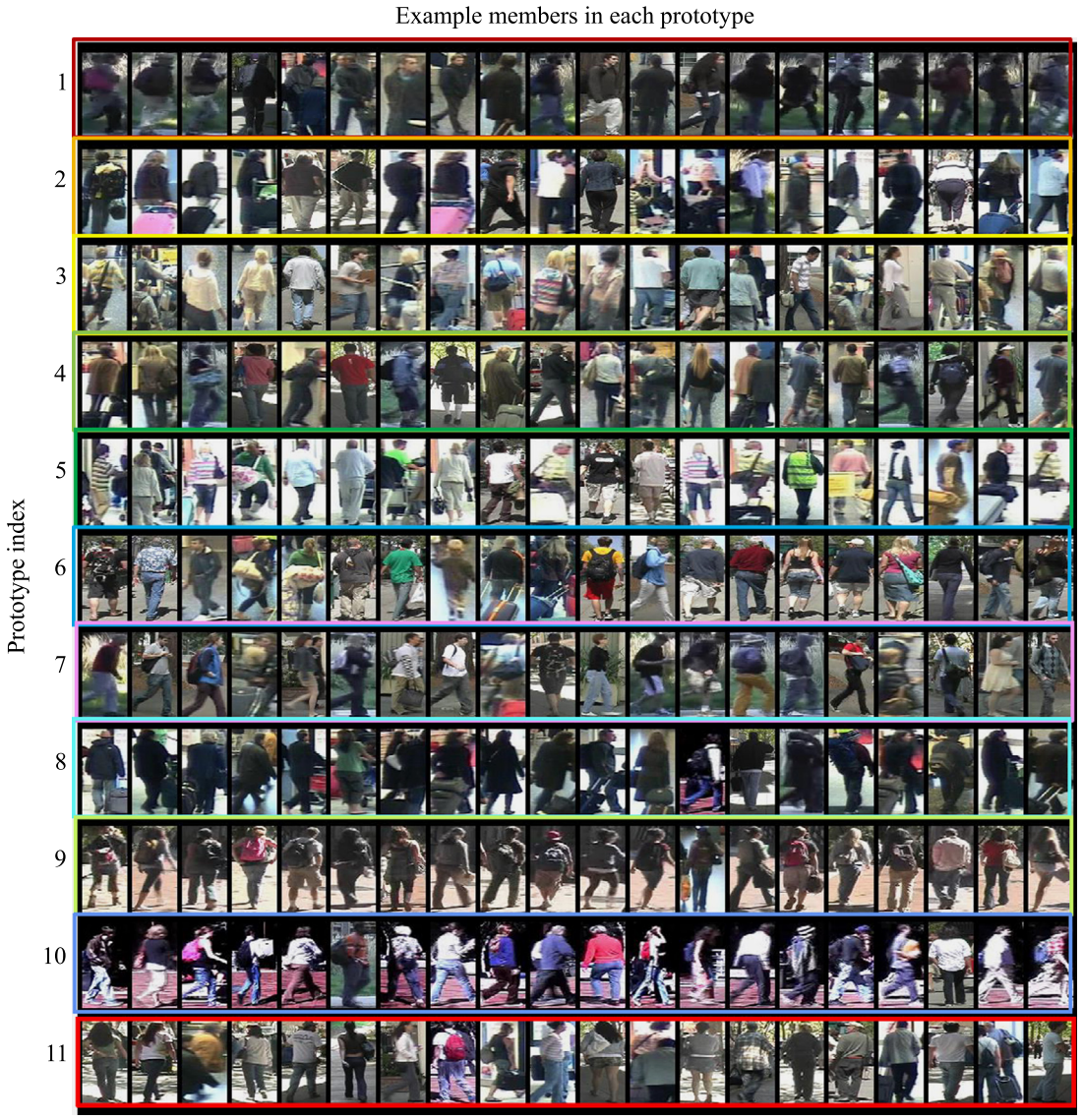

Fig. 6. Examples of prototype discovered in VIPeR dataset with some unlabelled images borrowed from i-LIDS. Each prototype represents a low-dimensional manifold cluster that models similar appearance attributes. Each image row in the figure shows a few examples of images in a particular prototype, with their interpreted unsupervised attributes listed as follows: (1) dark coat, dark trousers, side pose; (2) dark coat, with luggage; (3) bright shirt with texture; (4) jeans; (5) colourful jacket with texture, bright trousers; (6) colourful shirt, with bag; (7) dark trousers, side pose; (8) dark coat, dark trousers; (9) dark shirt, bright trousers, back pose; (10) colourful shirt, jeans, side pose; (11) bright shirt, dark trousers. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

which can be employed in subsequent step for bottom-up feature importance mining.

4.4. Bottom-up versus top-down generic feature importance

It is interesting to first analyse which features are regarded as important by different importance measures. In the following experiment we compare the feature importance measures produced by two generic feature importance (GFI) methods, i.e. the RankSVM [25] and the PRDC [31] (see Section 2 for details), and the bottom-up feature importance mining method. The GFI-based approaches are trained using labelled images, and the results are averaged over 10-fold cross validation. We fix the penalty parameter in RankSVM to 100 and used the default parameter values for PRDC as in [31] for all the datasets.

Fig. 9 shows examples to highlight the feature importance values discovered by different methods at different body regions. On the left-most pane we show the feature importance discovered by both the RankSVM and the PRDC,[4] followed by that inferred by ISFI in the middle-pane, and PSFI in the right-most pane. Each region in the silhouette/actual images is masked with the labelling colour of the most dominant feature type. In the feature importance plot, we show in each region the importance of each type of the features, of which the values are derived by summing the weight of all the histogram bins that belong to this type.

We first compare the top-down generic feature importance with the bottom-up feature importance, i.e. PSFI and ISFI. In general, the GFI methods emphasise more on the colour features for all the regions, whereas the texture features are assigned higher weights in the leg region than the torso region. The same feature importance assignment is applied equally to all images regardless of the appearance of individuals. In contrast, the

[4] For PRDC, we only show the first orthogonal projection learned by the algorithm, i.e. the most dominant feature importance.

Please cite this article as: C. Liu, et al., On-the-fly feature importance mining for person re-identification, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.11.001

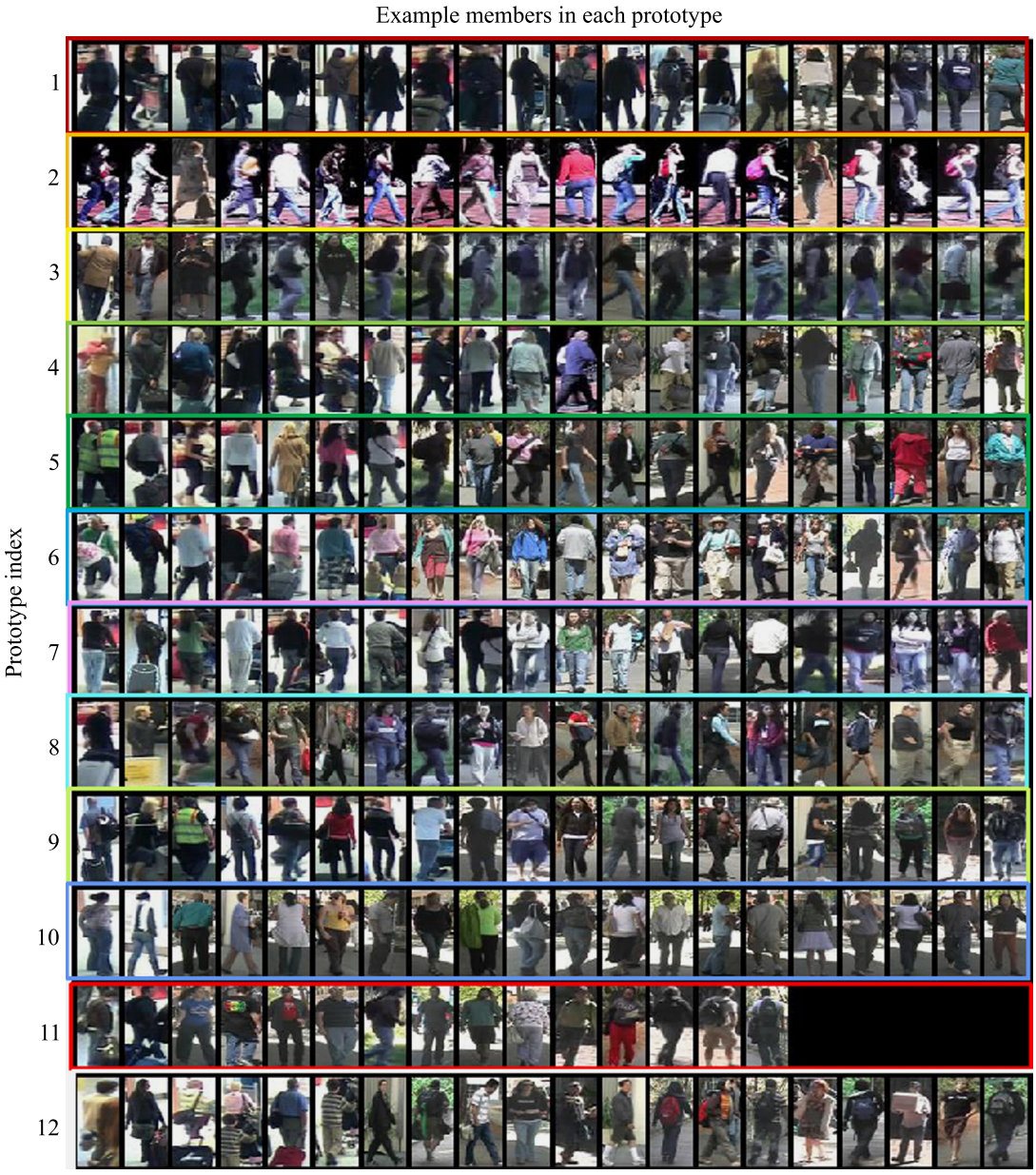

Fig. 7. Examples of prototype discovered in i-LIDS dataset with some unlabelled images borrowed from VIPeR. Their interpreted unsupervised attributes listed as follows: (1) dark coat with luggage; (2) colourful shirt, jeans; (3) dark coat, dark trousers, side pose, with backpack; (4) dark coat; (5) colourful jacket, dark trousers; (6) bright jacket with texture; (7) bright shirt, bright jeans; (8) dark jacket with texture; (9) shirt with texture, dark trousers; (10) colourful shirt; (11) colourful shirt with texture; (12) shirt with stripe pattern. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

bottom-up feature importance discovered by both the PSFI and the ISFI is more attribute-sensitive. For example, for image regions with bright and distinct colour appearance, e.g. Fig. 9(a)-3&4 and Fig. 9(b)-1&2, the colour feature types in the torso region are allocated higher weights than the texture feature types. For image regions that exhibit rich texture pattern, such as Fig. 9(a)-1 with stripes on the jumper, Fig. 9(b)-4 with floral-skirt, the relative importance of texture features increases. For instance, in Fig. 9(b)-4, the weight of the Gabor feature type in the forth region is 12.37% higher than that observed in the second region.

We analyse subsequently the differences between PSFI and ISFI. Note that in each row of Fig. 9 we show two example members from a particular prototype in the middle pane and the associated feature importance of that prototype, that is the PSFI in the right-most pane. As revealed by the selected example pairs from the same prototype, although the PSFI is capable of assigning higher weight to certain common attributes, such as bright shirt (e.g. Fig. 9(a)-3&4) or bright coat (e.g. Fig. 9(b)-1&2), it is not able to distinguish further those examples in the same prototype. For instance, PSFI fails to discover the fact that the third region of Fig. 9(b)-1 has more structured texture pattern than that of Fig. 9 (b)-2. In contrast, the ISFI is marginally better in mining the subtle uniqueness of specific individual, as it is able to assign higher weight to the texture feature type in the third region of Fig. 9(b)-1.

4.5. Further evaluations

Evaluating bottom-up feature importance: As shown in Table 1, in comparison to the baseline uniform weighting method, the bottom-up feature importance mining gives improved matching

Please cite this article as: C. Liu, et al., On-the-fly feature importance mining for person re-identification, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.11.001

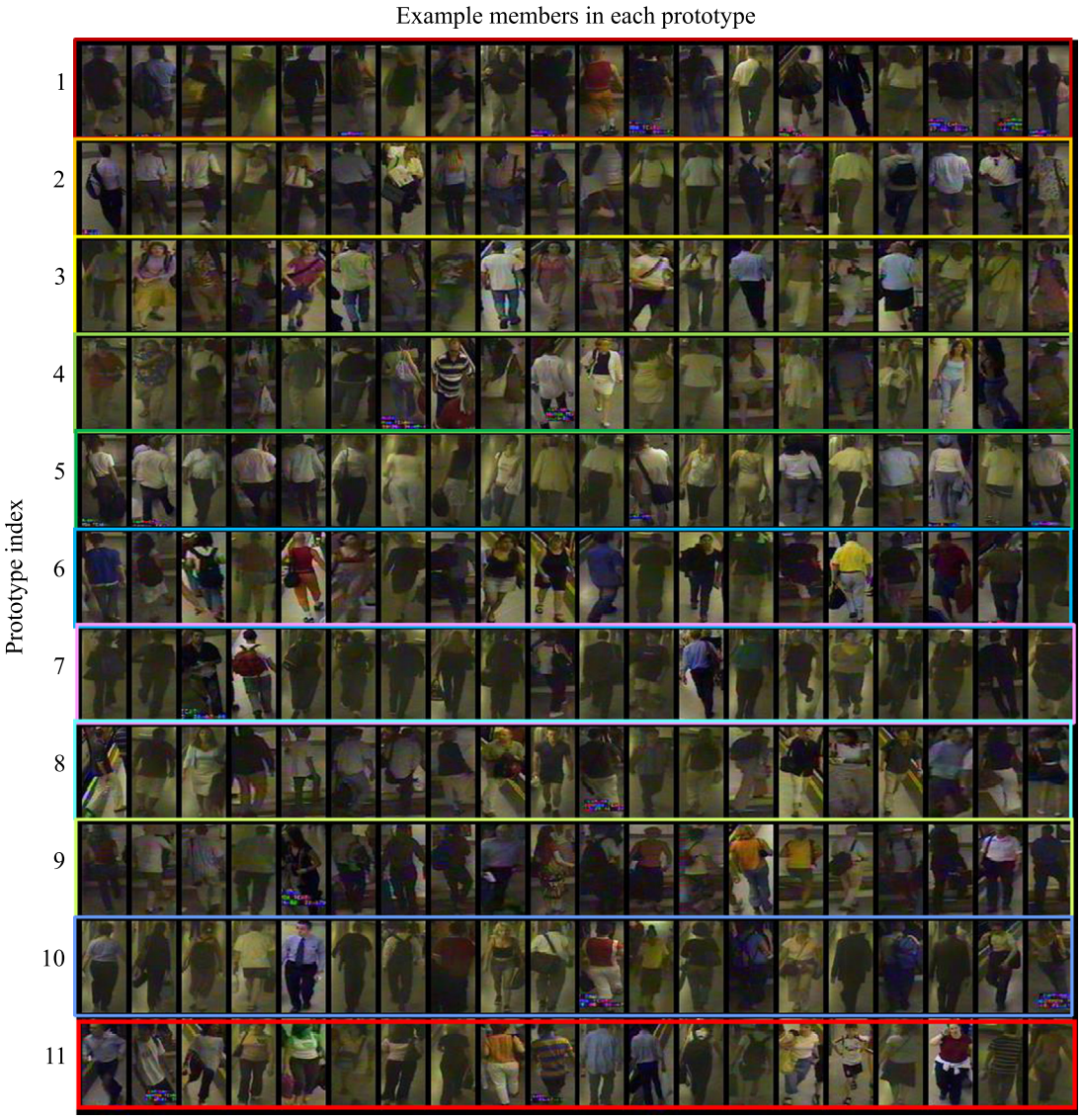

Fig. 8. Examples of prototype discovered in GRID dataset, with their interpreted unsupervised attributes listed as follows: (1) dark coat, dark trousers, with backpack; (2) bright shirt, dark trousers, with backpack; (3) colourful shirt with texture; (4) bright shirt, bright trousers; (5) white shirt, dark trousers, back pose; (6) colourful shirt; (7) dark coat, dark trousers; (8) bright trousers; (9) colourful shirt, dark trousers; (10) bright shirt, dark trousers; (11) bright shirt with texture. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

rate on all the datasets. The improvements on both the VIPeR and GRID datasets are statistically significant at 5% significance level.[5] Owing to the better capability of representing uniqueness of specific individual, the performance of ISFI is better to that obtained by using PSFI, as indicated by the relative increase of $\sim 2.5\%$ in rank 1 matching rate averaged across all three datasets.

Note that we borrow unlabelled images from different external data sources to facilitate the prototype discovery process (see Section 4.3). Without the additional unlabelled data, the unsupervised prototype clustering suffers from insufficient data for capturing the statistics of the population. In particular, we observe a performance drop of 7.26% and 3.55% in AUC of PSFI and ISFI, respectively, when no additional unlabelled data are used. The results suggest that PSFI and ISFI can greatly benefit from the freely available unannotated samples, even from different data sources.

Evaluating the fusion of top-down and bottom-up feature importance: In this experiment we evaluate the fusion of top-down and bottom-up feature importance (Section 3.5). We use a separate validation partition to obtain the value of α in Eq. (11). In particular, the values are set to 0.2, 0.4, and 0.2 for VIPeR, i-LIDS, and GRID respectively. We shall provide sensitivity test on this parameter in Section 4.6.

Table 2 summarises the results. It is evident from the table that the proposed fusion approach improves the RankSVM and PRDC baselines. In particular, the improvements yielded by all combination variants are statistically significant at 5% significance level, except the combinations with RankSVM on the challenging GRID dataset (despite higher averaged matching rates are observed).The relatively limited improvement may be caused by the nature of the GRID dataset, in which many people tend to wear clothing with a similar style and colour. This largely increases the difficulty in discovering meaningful and distinctive prototypes, leading to poorer weight estimation in both PSFI and ISFI. Adopting more elaborative features may overcome this issue.

It is evident that on their own the top-down supervised generic feature weighting outperforms bottom-up unsupervised feature

[5] We employ Wilcoxon signed-rank test in all the significant tests in this paper. In particular, we quantify the improvement significance in terms of the AUC of top 30 ranks. In general, performance gains on these top ranks are regarded important in person re-identification application.

Please cite this article as: C. Liu, et al., On-the-fly feature importance mining for person re-identification, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.11.001

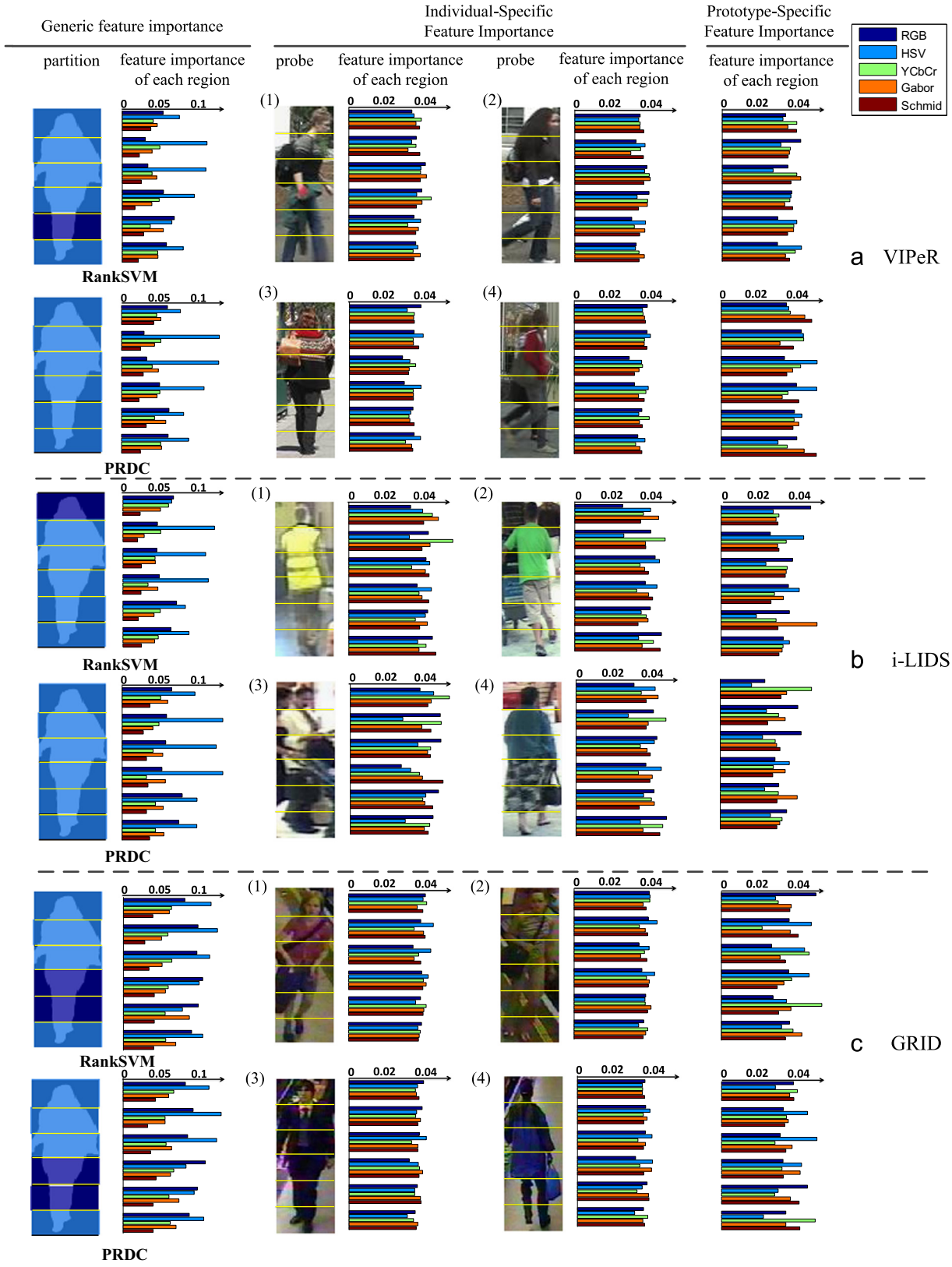

Fig. 9. Examples of comparison of the generic feature importance/weights by RankSVM [25] and PRDC [31] against the bottom-up feature importance mining.

importance mining. However, a combined weighting (Eq. (11)) improves both models. This suggests that the benefits from top-down and bottom-up feature importance to re-identification are not exclusive and can play a complementary role.

4.6. Pre-processing and parameter sensitivity test

Parameter for combining top-down and bottom-up feature importance α: Fig. 10 shows the sensitivity test results of α (see Eq. (11)

Please cite this article as: C. Liu, et al., On-the-fly feature importance mining for person re-identification, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.11.001

Table 1
Comparison of top rank matching rate (%) on VIPeR, i-LIDS and GRID datasets, between PSFI/ISFI and uniform weighting method. r is the rank and p is the size of gallery set. We use a superscript * beside the dataset on which the improvements are statistically significant.

Method	VIPeR ($p=316$)*				i-LIDS ($p=50$)				GRID ($p=900$)*			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
Uniform weight	9.43	20.03	27.06	34.68	30.40	55.20	67.20	80.80	4.40	11.68	16.24	24.80
PSFI	**10.32**	**23.10**	**32.18**	**45.57**	30.40	**56.20**	**68.00**	**81.60**	**4.96**	**14.32**	**20.24**	**26.56**
ISFI	**10.63**	**24.02**	**32.18**	**44.40**	30.20	**57.00**	**67.60**	**82.00**	**5.20**	**14.80**	**20.32**	**26.56**

Table 2
Comparison of top rank matching rate (%) on VIPeR, i-LIDS and GRID datasets, between the generic feature weighting and the combination methods. r is the rank and p is the size of gallery set. Note that on GRID dataset only the improvements from the combinations of PRDC are statistically significant. We use a superscript * beside the dataset on which the improvements are statistically significant.

Method	VIPeR ($p=316$)*				i-LIDS ($p=50$)*				GRID ($p=900$)*			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
RankSVM	14.87	37.12	50.19	65.66	29.80	57.60	73.40	84.80	10.24	24.56	33.28	43.68
PSFI+RankSVM	**15.76**	**38.70**	**51.36**	**66.84**	**32.60**	**60.00**	**73.40**	**86.40**	**10.32**	**25.36**	**33.52**	**43.84**
ISFI+RankSVM	**16.46**	**38.76**	**51.36**	**67.18**	**31.60**	**58.40**	**73.80**	**86.40**	**10.72**	24.56	**33.52**	**44.16**
PRDC	16.01	37.09	51.27	65.95	31.40	57.00	70.20	83.00	9.68	22.00	32.96	44.32
PSFI+PRDC	**16.99**	**38.10**	**52.37**	**66.84**	**33.60**	**60.80**	**73.00**	**85.60**	**10.24**	**23.44**	**34.80**	**45.44**
ISFI+PRDC	**17.12**	**38.96**	**52.94**	**67.34**	**35.00**	**59.80**	**72.80**	**85.00**	9.60	**23.04**	**33.92**	**46.08**

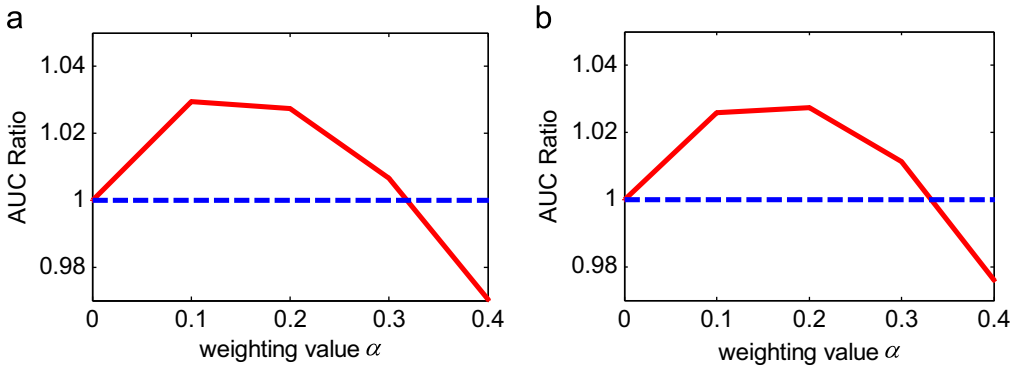

Fig. 10. α sensitivity test. (a) Combination of ISFI and RankSVM on the VIPeR dataset; (b) combination of ISFI and PRDC on the VIPeR dataset.

on the VIPeR dataset. The AUC ratio with respect to a specific value of α is computed by dividing the area under the CMC curve top 30 ranks obtained from the combined measure by that obtained when we set $\alpha=0$ (i.e. only GFI is activated). The higher ratio indicates a better performance of the combined feature importance weighting (Eq. (11)). These results show that setting α in the range of [0.1, 0.3] generally improves both top-down and bottom-up feature importance weighting on the VIPeR dataset. Similar α sensitivity test results are observed on i-LIDS and GRID datasets, where the best ranges are [0.3, 0.6] and [0.1, 0.3], respectively.

Note that setting a small α implies a high emphasis on the global weight derived from supervised learning. This is reasonable since performance gain in re-identification still has to rely on the capability of capturing the global viewing condition changes, which requires supervised weight learning.

Evaluating the effect of maximal-weight selection: This scheme automatically adapts the original weight values of $\tilde{\mathbf{w}}^p$ for more robust fusion (see Section 3.5). In Fig. 11, we compare ISFI+PRDC with and without applying the scheme, in terms of their respective AUC improvement over the baseline PRDC method. Clearly while ISFI+PRDC with the maximal-weight selection rarely performs worse than that without selection, the potential improvement is in general promising. We observe similar results on other ISFI/PSFI and RankSVM/PRDC combinations.

Forest parameters for prototype generation: We evaluate the sensitivity of the number of trees T_{cluster}, node size in the clustering forest, and the number of prototypes K during the prototype generation, using the CMC curve of ISFI as our performance measure and VIPeR as the test dataset. As shown in Fig. 12 (a), only a slight performance increase is obtained from introducing more trees to the forest, at a price of higher computational burden. The re-identification performance is equally insensitive to the node size and the number of prototypes as shown in Fig. 12 (b) and (c).

Pre-processing with foreground mask: In our experiment, the features are extracted from the whole image to ensure consistency with the experimental settings applied in both [25] and [31]. This is also stemmed from a practical consideration that finding accurate foreground regions is non-trivial in real-world scenario. However, intuitively the features, and subsequently the prototypes, would be less influenced by the background region if the feature extraction is performed on the human body with a foreground mask imposed. To evaluate this assumption, we applied an ellipsoid mask and treated the internal region of the ellipse as foreground area, as shown in Fig. 13(a). Alternatively, other segmentation techniques such as STEL [13] can be used to discard the background. The same testing protocol is adopted as described in Section. 4.1. As shown in Fig. 13(b), both the ISFI and PSFI

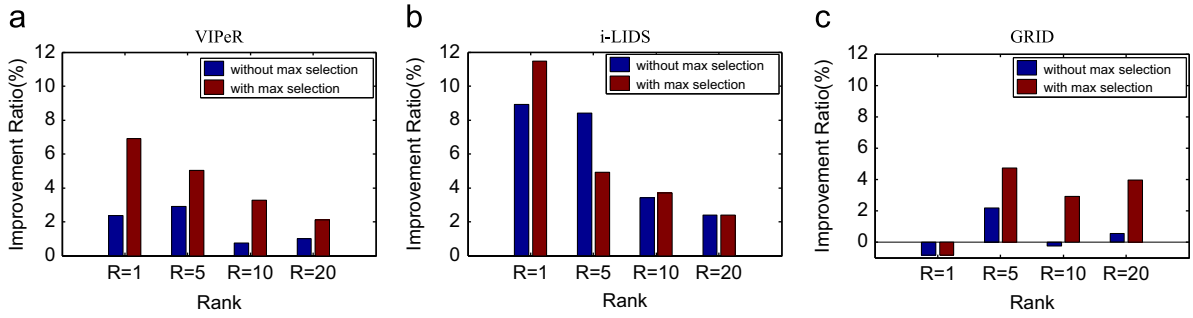

Fig. 11. The effect of applying maximal-weight selection scheme when combining bottom-up and top-down feature importance. We compare the performance between ISFI+PRDC with and without applying the scheme, in terms of AUC improvement over the baseline PRDC.

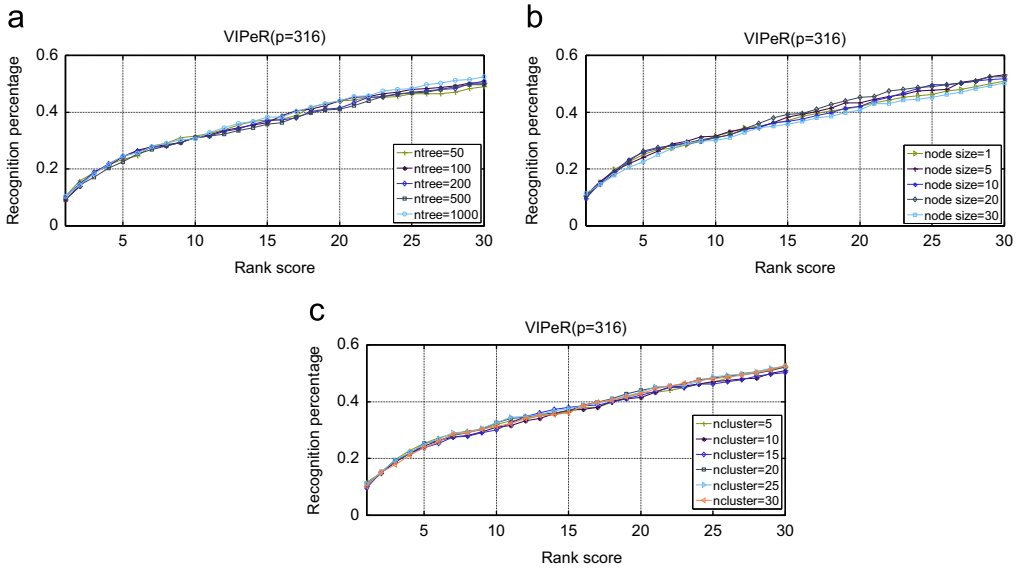

Fig. 12. Sensitivity of parameters in prototype generation, including (a) the number of trees and (b) the node size in the clustering forest, and (c) the number of clusters or prototypes.

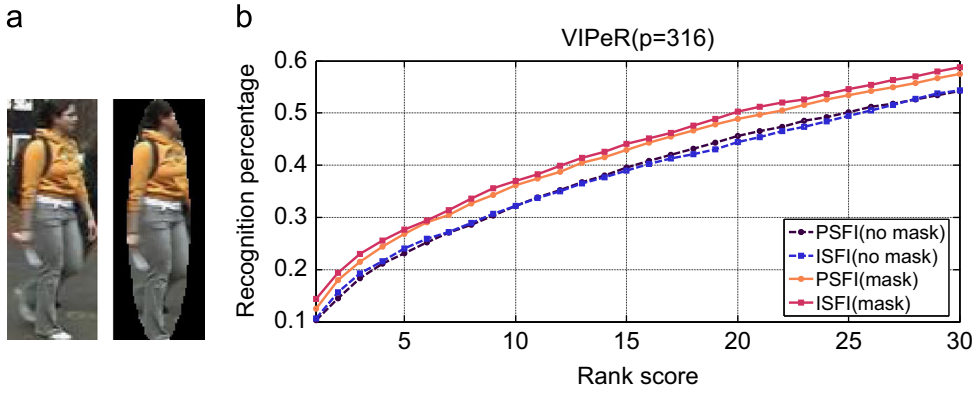

Fig. 13. The effect of foreground mask on recognition performance on VIPeR. (a) Left: probe image without mask; right: probe image with mask. (b) CMC curves of PSFI and ISFI.

methods enjoy an increase of performance when such a generic foreground mask is applied.

5. Discussion and conclusion

In this study, we have shown that certain appearance features can be more important than others in describing an individual and distinguishing him/her from other people. To that end, we proposed a novel method based on a cascaded clustering-classification random forest to perform unsupervised bottom-up feature importance mining driven by unsupervised appearance attribute-based prototype clustering. This approach complements existing person re-identification studies that focus on top-down supervised learning of generic feature weighting.

Experimental results on three benchmark datasets show a tangible indication that instead of biasing all the weights to features that are assumed universally good for all individuals

(to compensate and reflect the stability of each feature component across two cameras), computing selective feature weighting on-the-fly for each probe image can improve re-identification.

Importantly we found that the effectiveness of unsupervised bottom-up feature importance mining is dependent on both the quantity and quality of the unlabelled training data, in terms of the available size of the training data and the diversity of appearance attributes, i.e. sufficient and non-biased sampling of large diversity in population appearance in the training data can benefit significantly bottom-up feature importance mining for person re-identification. Firstly, as shown in the experiment, the sufficient number of unlabelled data is desired to generate robust prototypes. Secondly, it would be better to prepare a training set of unlabelled images that cover a variety of different prototypes, in order to have non-biased contributions from different feature types.

The results from this work raise an interesting question for further study, what is the best mechanism for unsupervised bottom-up feature importance mining? In this study, our approach explored explicitly unsupervised prototype discovery and classification error gain as the basis for bottom-up feature importance mining. Other alternatives can also be explored, e.g. exploiting different error gain measures such as outlier score from a large reference image set. Future work can also include the investigation of better feature selection fusion strategies for combining top-down generic weighting and bottom-up feature importance mining.

Conflict of interest

None declared.

References

[1] A. Alahi, P. Vandergheynst, M. Bierlaire, M. Kunt, Cascade of descriptors to detect and track objects across any network of cameras, Comput. Vision Image Understanding 114 (6) (2010) 624–640.
[2] S. Bak, E. Corvee, F. Brémond, M. Thonnat, Person re-identification using haar-based and DCD-based signature, in: IEEE International Conference on Advanced Video and Signal Based Surveillance, 2010, pp. 1–8.
[3] S. Bak, E. Corvee, F. Brémond, M. Thonnat, Person re-identification using spatial covariance regions of human body parts, in: IEEE International Conference on Advanced Video and Signal Based Surveillance, 2010, pp. 435–440.
[4] S. Bak, E. Corvee, F. Bremond, M. Thonnat, Multiple-shot human re-identification by mean riemannian covariance grid, in: IEEE International Conference on Advanced Video and Signal Based Surveillance, 2011, pp. 179–184.
[5] L. Bazzani, M. Cristani, A. Perina, V. Murino, Multiple-shot person re-identification by chromatic and epitomic analyses, Pattern Recognit. Lett. 33 (7) (2012) 898–903.
[6] H. Ben Shitrit, J. Berclaz, F. Fleuret, P. Fua, Tracking multiple people under global appearance constraints, in: IEEE International Conference on Computer Vision. IEEE, 2011, pp. 137–144.
[7] L. Breiman, Random forests, Machine Learning 45 (1) (2001) 5–32.
[8] D. Cheng, M. Cristani, M. Stoppa, L. Bazzani, V. Murino, Custom pictorial structures for re-identification, in: British Machine Vision Conference, 2011, pp. 68.1–68.11.
[9] M. Farenzena, L. Bazzani, A. Perina, V. Murino, M. Cristani, Person re-identification by symmetry-driven accumulation of local features, in: IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 2360–2367.
[10] D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, in: European Conference on Computer Vision, 2008, pp. 262–275.
[11] M. Hirzer, C. Beleznai, P. Roth, H. Bischof, Person re-identification by descriptive and discriminative classification. Image Anal. (2011), pp. 91–102.
[12] M. Hirzer, P. Roth, M. Köstinger, H. Bischof, Relaxed pairwise learned metric for person re-identification, in: European Conference on Computer Vision, 2012, pp. 780–793.
[13] N. Jojic, A. Perina, M. Cristani, V. Murino, B. Frey, Stel component analysis: modeling spatial correlations in image class structure, in: IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 2044–2051.
[14] M. Kostinger, M. Hirzer, P. Wohlhart, P.M. Roth, H. Bischof, Large scale metric learning from equivalence constraints, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2288–2295.
[15] C.H. Lampert, H. Nickisch, S. Harmeling, Learning to detect unseen object classes by between-class attribute transfer, in: IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 951–958.
[16] R. Layne, T. Hospedales, S. Gong, Person re-identification by attributes, in: British Machine Vision Conference, 2012.
[17] B. Liu, Y. Xia, P.S. Yu, Clustering through decision tree construction, in: International Conference on Information and Knowledge Management, 2000, pp. 20–29.
[18] C. Liu, S. Gong, C.C. Loy, X. Lin, Person re-identification: what features are important ? in: European Conference on Computer Vision Workshop on Person Re-identification, 2012, pp. 391–401.
[19] Y. Liu, Y. Shao, F. Sun, Person re-identification based on visual saliency, in: International Conference on Intelligent Systems Design and Applications, 2012, pp. 884–889.
[20] C.C. Loy, C. Liu, S. Gong, Person re-identification by manifold ranking, in: IEEE International Conference on Image Processing, 2013.
[21] C.C. Loy, T. Xiang, S. Gong, Time-delayed correlation analysis for multi-camera activity understanding, Int. J. Comput. Vision 90 (1) (2010) 106–129.
[22] R. Mandeljc, S. Kovačič, M. Kristan, J. Perš, Non-sequential multi-view detection, localization and identification of people using multi-modal feature maps, in: Asian Conference on Computer Vision, 2012, pp. 691–704.
[23] A. Mignon, F. Jurie, PCCA: a new approach for distance learning from sparse pairwise constraints, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2666–2672.
[24] P. Perona, L. Zelnik-Manor, Self-tuning spectral clustering, in: Neural Information Processing Systems, 2004, pp. 1601–1608.
[25] B. Prosser, W. Zheng, W. Gong, T. Xiang, Q. Mary, Person re-identification by support vector ranking, in: British Machine Vision Conference, 2010, pp. 21.1–21.11.
[26] W. Schwartz, L. Davis, Learning discriminative appearance-based models using partial least squares, in: The 22nd Brazilian Symposium on Computer Graphics and Image Processing, 2009, pp. 322–329.
[27] X.G. Wang, G. Doretto, T. Sebastian, J. Rittscher, P. Tu, Shape and appearance context modeling, in: IEEE International Conference on Computer Vision, 2007, pp. 1–8.
[28] Y. Wu, M. Minoh, M. Mukunoki, S. Lao, Set based discriminative ranking for recognition, in: European Conference on Computer Vision, Springer, 2012, pp. 497–510.
[29] T. Xiang, S. Gong, Spectral clustering with eigenvector selection, Pattern Recognit. 41 (March (3)) (2008) 1012–1029.
[30] W. Zheng, S. Gong, T. Xiang, Associating groups of people, in: British Machine Vision Conference, 2009, pp. 23.1–23.11.
[31] W. Zheng, S. Gong, T. Xiang, Re-identification by relative distance comparison, IEEE Trans. Pattern Anal. Mach. Intell. 35 (March (3)) (2013) 653–668.

Chunxiao Liu is a Ph.D. candidate in the Department of Electronics Engineering, Tsinghua University, China. He received his B.S. degree in the Department of Electronics and Information Engineering, Huazhong University of Science & Technology, Wuhan, in 2008. His research interests include human re-identification, tracking, camera network activity analysis, machine learning.

Shaogang Gong is a Professor of Visual Computation at Queen Mary University of London, a Fellow of the Institution of Electrical Engineers and a Fellow of the British Computer Society. He received his D.Phil in computer vision from Keble College, Oxford University in 1989. His work focusses on motion and video analysis; object detection, tracking and recognition; face and expression recognition; gesture and action recognition; visual behaviour recognition.

Chen Change Loy is a Research Assistant Professor in the Department of Information Engineering, the Chinese University of Hong Kong. Previously, he was a Post-doctoral Researcher at Vision Semantics Limited. He received the Ph.D. degree in Computer Science from Queen Mary University of London in 2010. His research interests include computer vision and machine learning, with focus on activity analysis and understanding in surveillance video.

Please cite this article as: C. Liu, et al., On-the-fly feature importance mining for person re-identification, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.11.001