Chapter 10
Evaluating Feature Importance for Re-identification

Chunxiao Liu, Shaogang Gong, Chen Change Loy and Xinggang Lin

Abstract Person re-identification methods seek robust person matching through combining feature types. Often, these features are assigned implicitly with a single vector of global weights, which are assumed to be universally and equally good for matching all individuals, independent of their different appearances. In this study, we present a comprehensive comparison and evaluation of up-to-date imagery features for person re-identification. We show that certain features play more important roles than others for different people. To that end, we introduce an unsupervised approach to learning a bottom-up measurement of feature importance. This is achieved through first automatically grouping individuals with similar appearance characteristics into different prototypes/clusters. Different features extracted from different individuals are then automatically weighted adaptively driven by their inherent appearance characteristics defined by the associated prototype. We show comparative evaluation on the re-identification effectiveness of the proposed *prototype-sensitive feature importance*-based method as compared to two generic weight-based *global feature importance* methods. We conclude by showing that their combination is able to yield more accurate person re-identification.

C. Liu (✉) · X. Lin
Tsinghua University, Beijing, China
e-mail: lcx08@mails.tsinghua.edu.cn

X. Lin
e-mail: xglin@mail.tsinghua.edu.cn

S. Gong
Queen Mary University of London, London, UK
e-mail: sgg@eecs.qmul.ac.uk

C. C. Loy
The Chinese University of Hong Kong, Hong Kong, China
e-mail: ccloy@ie.cuhk.edu.hk

S. Gong et al. (eds.), *Person Re-Identification*,
Advances in Computer Vision and Pattern Recognition,
DOI: 10.1007/978-1-4471-6296-4_10, © Springer-Verlag London 2014

10.1 Introduction

Visual appearance-based person re-identification aims to establish a visual match between two imagery instances of the same individual appearing at different locations and times under unknown viewing conditions which are often significantly different. Solving this problem is non-trivial owing to both very sparse samples of the person of interest, often a single example imagery to compare against, and the unknown viewing condition changes, including visual ambiguities and uncertainties caused by illumination changes, viewpoint and pose variations and inter-object occlusion [16, 27, 28]. In order to cope with sparsity of data and the challenging view conditions, most existing methods [8, 9, 17] combine different appearance features, such as colour and texture, to improve reliability and robustness in person matching. Typically, feature histograms are concatenated and weighted in accordance to their *importance*, i.e. their discriminative power in distinguishing a target of interest from other individuals.

Current re-identification techniques [19, 30, 33, 41] assume implicitly a feature weighting or selection mechanism that is *global*, i.e. a set of generic weights on feature types invariant to a population. That is, to assume a single weight vector (or a linear weight function) that is globally optimal for all people. For instance, one often assumes colour is the most important (intuitively so) and universally a good feature for matching all individuals. In this study, we refer such a generic weight vector as a Global Feature Importance (GFI) measure. They can be learned either through boosting [19], rank learning [33] or distance metric learning [41]. Scalability is the main bottleneck of such approaches as the learning process requires exhaustive supervision on pairwise individual correspondence from a known dataset.

Alternatively, we consider that certain appearance features are 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 a textureless bright red shirt, but texture information can be equally or more critical for a person wearing a plaid shirt (Fig. 10.1). It is therefore undesirable to bias all the weights to the features that are universally good for all individuals. Instead, feature weighting should be able to *selectively distribute different weights adaptively according to the informativeness of features given different visual appearance attributes under changing viewing conditions and for different people*. By visual appearance attributes, we refer to conceptually meaningful appearance characteristics of an individual, e.g. dark shirt, blue jeans.

In this study, we first provide a comprehensive review of various feature representations and weighting strategies for person re-identification. In particular, we investigate the roles of different feature types given different appearance attributes and give insights into what features are more important under what circumstances. We show that selecting features specifically for different individuals can yield more robust re-identification performance than feature histogram concatenation with GFI as adopted by [27, 37].

Fig. 10.1 Two examples of a pair of probe image against a target (gallery) image, together with the rank of correct matching by different feature types independently

It is non-trivial to quantify feature importance adaptively driven by specific appearance attributes detected on an individual. A plausible way is to apply supervised attribute learning method, i.e. training a number of attribute detectors to cover an exhaustive set of possible attributes, and then defining feature importance associated to each specific attribute. This method requires expensive annotation and yet the annotation obtained may have low quality due to inevitable visual ambiguity. Previous studies [10, 18, 29] have shown great potential in using unsupervised attributes in various computer vision problems such as object recognition. Despite that the unsupervised attributes are not semantically labelled or explicitly named, they are discriminative and correlated with human attribute perception.

Motivated by the unsupervised attribute studies, we investigate here a random forests-based method to discover *prototypes* in an unsupervised manner. Each prototype reveals a mixture of attributes to describe specific population of persons with similar appearance characteristics, such as wearing colourful shirt and black pants. With the discovered prototypes, we further introduce an approach to quantify the feature importance specific for an individual driven by his/her inherent appearance attributes. We call the discovered feature importance *Prototype-Sensitive Feature Importance* (PSFI). We conduct extensive evaluation using four different person re-identification benchmark datasets, and show that combining prototype-sensitive feature importance with global feature importance can yield more accurate re-identification without any extra supervision cost as compared to existing learning-based approaches.

10.2 Recent Advances

Most person re-identification methods benefit from integrating several types of features [1, 8, 9, 14, 17, 19, 26, 33, 36, 37, 41]. In [17], weighted colour histograms derived from maximally stable colour regions (MSCR) and structured patches are combined for visual description. In [8], histogram plus epitome features are proposed as a human signature. Essentially, they explore the combination of colour and texture properties on the human appearance but with more specific feature types. There are a number of reviews on features and feature evaluation for person re-identification

[1, 7]. In [1], several colour and covariance features are compared; whilst in [7], local region descriptors such as SIFT and SURF are evaluated.

A *global feature importance* scheme is often adopted in existing studies to combine different feature types by assuming that certain features are universally more important under any circumstances, regardless of possible changes (often significant) in viewing conditions between the probe and gallery views and specific visual appearance characteristics of different individuals. Recent advances based on metric learning or ranking [2, 19, 21, 30, 33, 41] can be considered as data-driven *global feature importance mining* techniques. For example, the ranking support vector machines (RankSVM) method [33] converts the person re-identification task from a matching problem into a pairwise binary classification problem (correct match vs. incorrect match), and 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) [41] 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 encodes the global importance of each feature. In essence, the learned global feature importance reflects the stability of each feature component across two cameras. For example, if two camera locations are under significantly different lighting conditions, the colour features will be less important as they are unstable/unreliable. A major weakness of this type of pairwise learning-based methods is their potential limitation on scalability since the supervised learning process requires exhaustive supervision on pairwise correspondence, i.e. the building of a training set is cumbersome as it requires to have for each subject a pair of visual instances. The size of such a pairwise labelled dataset required for model learning is difficult to be scaled up.

Schwartz and Davis [36] propose a feature selection process depending on the feature type and the location. This method, however, requires labelled gallery images to discover the gallery-specific feature importance. To relax such conditions, in this work we investigate a fully unsupervised learning method for adaptive feature importance mining which aims to be more flexible (attribute-driven) without any limitations to a specific gallery set. A more recent study in [34] explores prototype relevance for improving processing time in re-identification. In a similar spirit but from a different perspective, this study investigates salient feature importance mining based on prototype discovery for improving matching accuracy. In [23], a supervised attribute learning method is proposed to describe the appearance for each individual. However, it needs massive human annotation of attributes which is labour-intensive. In contrast, we explore in an unsupervised way to discover the inherent appearance attributes.

10.3 Feature Representation

Different types of visual appearance features have been proposed for person re-identification, including colour histogram [17, 22], texture filter banks [33], shape context [37], covariance [3, 4, 6] and histogram plus epitome [8]. In general, colour

information is dominant when the lighting changes are not severe, as colour is more robust to viewpoint changes as compared to other features. Although texture or structure information can be more stable under significant lighting changes, they are sensitive to changes in viewpoint and occlusion. As shown in [8, 17], re-identification matching accuracy can be improved by combining several features so as to gain benefit from different and complementary information captured by different features.

In this study, we investigate a mixture of commonly used colour, structure and texture features for re-identification, similar to those employed in [19, 33], plus a few more additional local structure features. In particular, the following range of imagery features are considered:

- **Colour Histogram**: HSV colour histogram is employed in [8, 17, 36]. Specifically, in [17] they generate a weighted colour histogram according to pixel's location to the vertical symmetry axes of the human body. The intuition is that central pixels should be more robust to pose variations. HSV is effective in describing the bright colours, such as red, but not robust to neutral colour as the hue channel is undefined. An alternative representation is to combine the colour histograms from several complementary colour spaces, such as HSV, RGB, and YCbCr [19, 21, 33, 41].
- **Texture and Structure**: Texture and structure patterns are commonly found on clothes, such as the plaid (see Fig. 10.1) or the stripes (see Fig. 10.5b) on a sweater. Possible texture descriptors include Gabor and Schmid filters [19, 33] or local binary patterns (LBP) [39]. As to structure descriptor, histogram of gradient (HOG) [15] that prevails in human detection is considered in [1, 36, 37]. As these texture and structure features are computed on the intensity image, they play an important role in establishing correspondence when colour information degrades under drastic illumination changes and/or change of camera settings.
- **Covariance**: Covariance feature has been reported to be effective in [4, 5, 20, 24]. It has three advantages: (1) it reflects second-order regional statistical property discarded by histogram; (2) different feature types such as colour and gradient can be readily integrated; (3) it is versatile with no limitation to the region's shape, suggesting its potential to be integrated with most salient region detectors.

In this study, we divide a person image into six horizontal stripes (see Fig. 10.2). This is a generic human body partitioning method that is widely used in existing methods [33, 41] to capture distinct areas of interest. Alternative partitioning schemes, such as symmetry segmentation [8] or pictorial model [14], are also applicable. A total of 33 feature channels including RGB, HSV, YCbCr, Gabor (8 filters), Schmid (13 filters), HOG, LBP and Covariance are computed for each stripe. For the first five types of features, each channel is represented by a 16-dimensional vector. A detailed explanation of computing the former 5 features can be found in [33]. For HOG feature, each strip is further divided into 4×4 pixels cell and each cell is represented by a 9-dimensional gradient histogram, yielding a 36-dimensional feature vector for each strip. For LBP feature, we compute a 59-dimensional local binary pattern histogram on the intensity image. As for covariance feature for a given strip

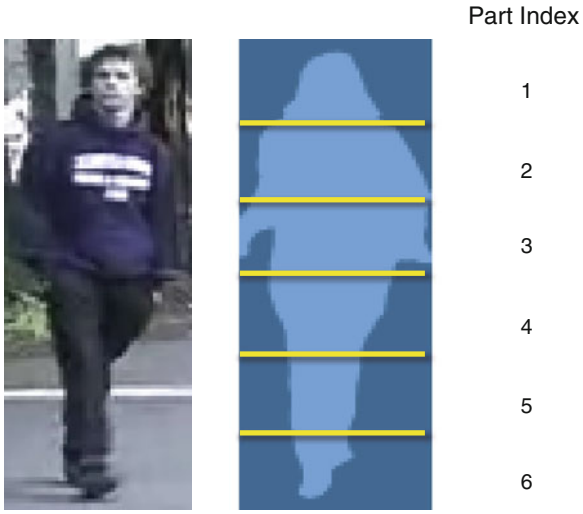

Fig. 10.2 A spatial representation of human body [33, 41] is used to capture visually distinct areas of interest. The representation employs six equal-sized horizontal strips in order to capture approximately the head, upper and lower torso and upper and lower legs

$R \subset I$, let $\{\mathbf{z}_m\}_{m=1\ldots M}$ be the feature vectors extracted from M pixels inside R. The covariance descriptor of region R is derived by

$$\mathbf{C}_R = \frac{1}{M-1} \sum_{m=1}^{M} (\mathbf{z}_m - \mu)(\mathbf{z}_m - \mu)^{\mathsf{T}}$$

where μ denotes the mean vector of $\{\mathbf{z}_m\}$. Here we use the following features to reflect information of each pixel $\mathbf{z} = [H, S, V, I_x, I_y, I_{xx}, I_{yy}]$ where H, S, V are the HSV colour values. The first-order (I_x and I_y) and second-order (I_{xx} and I_{yy}) image derivatives are calculated through the filters $[-1, 0, 1]^{\mathsf{T}}$ and $[-1, 2, -1]^{\mathsf{T}}$, respectively. The subscript x or y denotes the direction for filtering. Thus the covariance descriptor is a 7×7 matrix. While in this form covariance matrix cannot be directly combined with other features to form a single histogram representation. Hence, we follow the approach proposed by [20] to convert the 7×7 covariance matrix $\mathbf{C}$ into *sigma points*, expressed as follows:

$$\mathbf{s}_0 = \mu \tag{10.1}$$

$$\mathbf{s}_i = \mu + \alpha(\sqrt{\mathbf{C}})_i \tag{10.2}$$

$$\mathbf{s}_{i+d} = \mu - \alpha(\sqrt{\mathbf{C}})_i, \tag{10.3}$$

where μ is the mean value of sample data and $(\sqrt{\mathbf{C}})_i$ denotes the i-th column of the covariance matrix square root. Parameter α is a scalar weight for the elements in $\mathbf{C}$ and is set to $\alpha = \sqrt{2}$ for Gaussian data. Thus, the vector form of covariance

feature can be obtained by concatenation of all sigma points, in our case resulting in a 105-dimensional vector. Therefore, it allows for integration of other feature channels into one compact feature vector.

10.4 Unsupervised Mining of Feature Importance

Given the range of features included in our feature representation, we consider an unsupervised way to compute and evaluate a bottom-up measurement of feature importance driven by intrinsic appearance of individuals. To that end, we propose a three-step procedure as follows: (1) automatic discovery of feature prototypes by exploiting clustering forests; (2) prototype-sensitive feature importance mining by classification forests; (3) determining the feature importance of a probe image on-the-fly adapting to changes in viewing condition and inherent appearance characteristics of individuals. An overview of the proposed approach is depicted in Fig. 10.3.

Our unsupervised feature importance mining method is formulated based on random forests models, particularly the clustering forests [25] and classification forests [12]. Before introducing and discussing the proposed method, we briefly review the two forests models.

10.4.1 Random Forests

Random forests [12] are a type of decision trees constructed by an ensemble learning process, and can be designed for performing either classification, clustering or regression tasks. Random forests have a number of specific properties that make it suitable for the re-identification problem. In particular

1. It defines the pairwise affinity between image samples by the tree structure itself, therefore, avoiding manual definition of distance function.
2. It selects implicitly optimal features via optimisation of the well-defined information gain function [12]. This feature selection mechanism is beneficial to mitigating noisy or redundant visual features in our representation.
3. It performs empirically well on high-dimensional input data [13], a problem that is typical in person re-identification problem.

In addition to the three aforementioned characteristics, there are other attractive general properties in random forests such as it approximates the Bayes optimal classifier [35], it handles inherently multiple-class problem and it provides probabilistic outputs.

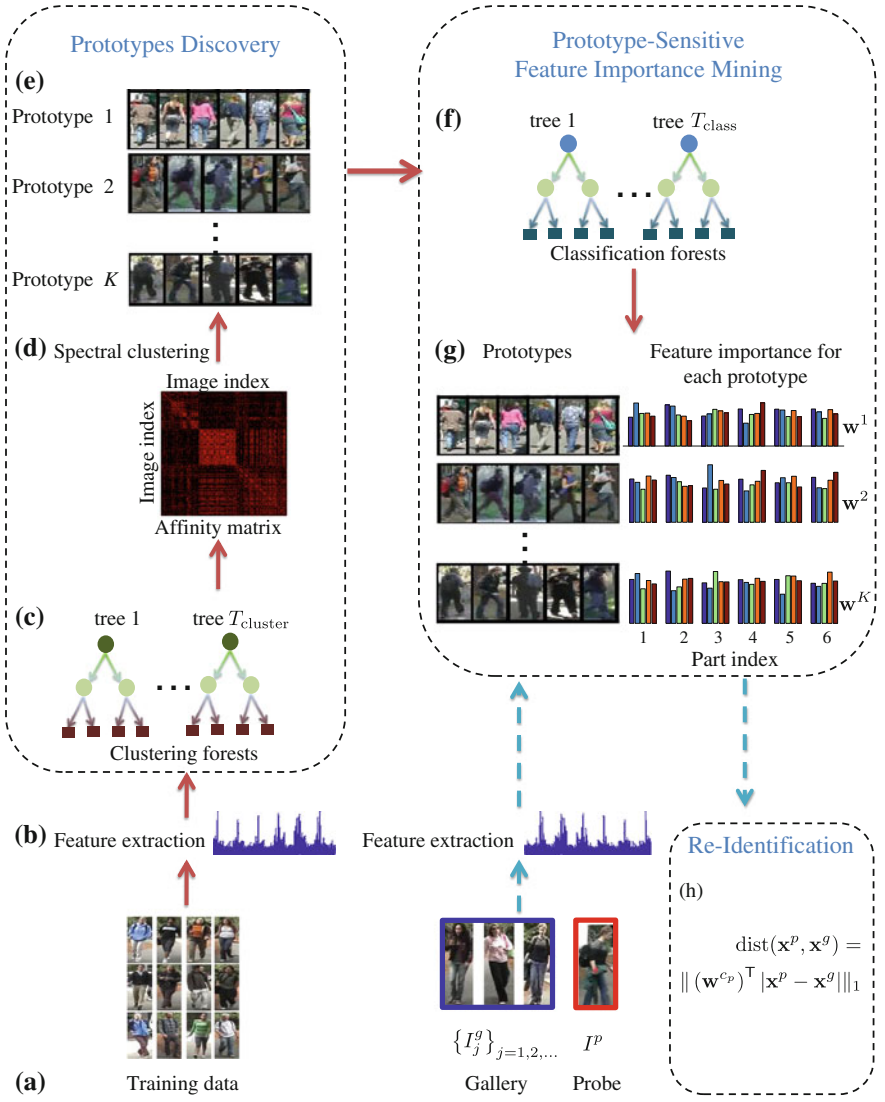

Fig. 10.3 Overview of prototype-sensitive feature importance mining. Training steps are indicated by *red solid arrows* and testing steps are denoted by *blue slash arrows*

Classification Forests

A common type of random forests is the classification forests. Classification forests [12, 35] consists of a set of T_{class} binary decision trees $\mathcal{T}(\mathbf{x}) : \mathcal{X} \rightarrow \mathcal{R}^K$, where $\mathcal{X} = \mathbb{R}^D$ is the D-dimensional feature space and $\mathcal{R}^K = [0, 1]^K$ represents the space of class probability distribution over the label space $\mathcal{C} = \{1, \ldots, K\}$. During testing, given an unseen sample $\mathbf{x}^* \in \mathbb{R}^D$, each decision tree produces a posterior $p_t(c|\mathbf{x})$,

and a probabilistic output from the forests can be obtained via averaging

$$p(c|\mathbf{x}^*) = \frac{1}{T_{\text{class}}} \sum_{t}^{T_{\text{class}}} p_t(c|\mathbf{x}^*). \tag{10.4}$$

The final class label c^* can be obtained as $c^* = \text{argmax}_c\, p(c|\mathbf{x}^*)$.

In the learning stage, each decision tree is trained independently from each other using a random subset of training samples, i.e. bagging [12]. Typically, one draws $\frac{2}{3}$ of the original training samples randomly for growing a tree, and reserves the remaining as *out-of-bag* (oob) validation samples. We will exploit these oob samples for computing importance of each feature (Section 10.4.3).

Growing a decision tree involves an iterative node splitting procedure that optimises a binary split function of each internal node. We define the split function as:

$$h(\mathbf{x}, \boldsymbol{\theta}) = \begin{cases} 0 & \text{if } \mathbf{x}_{\boldsymbol{\theta}_1} < \boldsymbol{\theta}_2 \\ 1 & \text{otherwise} \end{cases}. \tag{10.5}$$

The above split function is parameterised by two parameters: (i) a feature dimension $\boldsymbol{\theta}_1 \in \{1, \ldots, D\}$, and (ii) a threshold $\boldsymbol{\theta}_2 \in \mathbb{R}$. Based on the outcome of Eq. (10.5), a sample $\mathbf{x}$ arriving at the split node will be channelled to either the left or right child nodes.

The best parameter $\boldsymbol{\theta}^*$ is chosen by optimising

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta} \in \Theta}{\text{argmax}}\, \Delta\mathcal{I}, \tag{10.6}$$

where Θ is a randomly sampled set of $\{\boldsymbol{\theta}^i\}$. The information gain $\Delta\mathcal{I}$ is defined as follows:

$$\Delta\mathcal{I} = \mathcal{I}_p - \frac{n_l}{n_p}\mathcal{I}_l - \frac{n_r}{n_p}\mathcal{I}_r, \tag{10.7}$$

where p, l and r refer to a splitting node, the left and right child, respectively; n denotes the number of samples at a node, with $n_p = n_l + n_r$. The $\mathcal{I}$ can be computed as either the entropy or Gini impurity [11]. Throughout this paper we use the Gini impurity.

Clustering Forests

In contrast to classification forests, clustering forests does not require any ground truth class labels for learning. Therefore, it is suitable for our problem of unsupervised prototype discovery. Clustering forests consists of T_{cluster} decision trees whose leaves define a spatial partitioning or grouping of the data. Although the clustering forests is an unsupervised model, it can be trained using the classification

forests optimisation routine by following the pseudo two-class algorithm proposed in [12, 25]. In particular, in each splitting node we add n_p uniformly distributed pseudo points $\bar{\mathbf{x}} = \{\bar{x}_1, \ldots, \bar{x}_D\}$, with $\bar{x}_i \sim \mathbf{U}(x_i | \min(x_i), \max(x_i))$ into the original data space. With this strategy, the clustering problem becomes a canonical classification problem that can be solved by the classification forests training method discussed above.

10.4.2 Prototype Discovery

Now we discuss how to achieve feature importance mining through a clustering-classification forests model. First, we describe how to achieve prototype discovery by clustering forests (Fig. 10.3a–e). In contrast to a top-down approach to specifying appearance attributes and mining features to support each attribute class independently [23], in this study we investigate bottom-up approach to discovering automatically representative clusters (prototypes) corresponding to similar constitutions of multiple classes of appearance attributes.

To that end, we first perform unsupervised clustering to group a given set of unlabelled images into several *prototypes* or clusters. Each prototype is composed of images that possess similar appearance attributes, e.g. wearing colourful shirt, with backpack, dark jacket (Fig. 10.3e). More precisely, 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 $f(I) = \mathbf{x} = (x_1, \ldots, x_D)^\mathsf{T} \in \mathbb{R}^D$ (Fig. 10.3b). We wish to discover a set of prototypes $c \in \mathcal{C} = \{1, \ldots, K\}$, i.e. low-dimensional manifold clusters that group images $\{I\}$ with similar appearance attributes.

We treat the prototype discovery problem as a graph partitioning problem, which requires us to first estimate the pairwise similarity between images. We adopt the clustering forests [12, 25] for pairwise similarity estimation. Formally, we construct clustering forests as an ensemble of T_{cluster} clustering trees (Fig. 10.3c). 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. Now for each tree, we are able to compute an $n \times n$ affinity matrix A^t, with each element A^t_{ij} defined as

$$A^t_{ij} = \exp^{-\text{dist}^t(\mathbf{x}_i, \mathbf{x}_j)}, \tag{10.8}$$

where

$$\text{dist}^t(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} 0 & \text{if } l(\mathbf{x}_i) = l(\mathbf{x}_j) \\ \infty & \text{otherwise} \end{cases}. \tag{10.9}$$

Following Eq. (10.9), we assign 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 $= \infty$) otherwise. To obtain a smooth forests affinity matrix, we compute the final affinity matrix as

$$A = \frac{1}{T_{\text{cluster}}} \sum_{t=1}^{T_{\text{cluster}}} A^t. \tag{10.10}$$

Given the affinity matrix, we perform spectral clustering algorithm [31] to partition the weighted graph into K prototypes. Thus, each unlabelled probe image $\{I_i\}$ is assigned to a prototype c_i (Fig. 10.3e). In this study, K is the cluster number and pre-defined, but its value can be readily estimated automatically using alternative methods such as [32, 38].

10.4.3 Prototype-Sensitive Feature Importance

In this section, we discuss how to derive the feature importance for each prototype generated by the previous prototype discovery. As discussed in Sect. 10.1, unlike the global feature importance that is assumed to be universally good for all images, prototype-sensitive feature importance is designed to be specific to prototype characterised by different appearance characteristics. That is, each prototype c has its own prototype-sensitive weighting or feature importance (PSFI)

$$\mathbf{w}^c = \left(w_1^c, \ldots, w_D^c\right)^{\mathsf{T}}, \tag{10.11}$$

of which high value should be assigned to unique features of that prototype. For example, texture features gain higher weights than others if the images in the prototype have rich textures but less bright colours.

Based on the above consideration, we compute the importance of a feature according to its ability in discriminating different prototypes. The forests model naturally reserves a validation set or out-of-bag (oob) samples for each tree during bagging (Sect. 10.4.1). This property permits a convenient and robust way of evaluating the importance of individual features.

Specifically, we train a classification random forests [12] using $\{\mathbf{x}\}$ as inputs and treating the associated prototype labels $\{c\}$ as classification outputs (Fig. 10.3f). To compute the feature importance, we first 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 $\widetilde{\varepsilon}_d^{c,\,t}$ on the perturbed oob samples of prototype c. The importance of the dth feature of prototype c is then computed as the error gain

$$w_d^c = \frac{1}{T_{\text{class}}} \sum_{t=1}^{T_{\text{class}}} (\widetilde{\varepsilon}_d^{c,\,t} - \varepsilon_d^{c,\,t}). \tag{10.12}$$

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, therefore suggests its critical role in discriminating between different prototypes.

10.4.4 Ranking

With the method described in Sect. 10.4.3, we obtain PSFI for each prototypes. This subsequently permits us to evaluate bottom-up feature importance of an unseen probe image, $\mathbf{x}^p$ on-the-fly driven by its intrinsic appearance prototype. Specifically, following Eq. (10.4), we classify $\mathbf{x}^p$ using the learned classification forests to obtain its prototype label c_p

$$c_p = \text{argmax}_c\, p(c|\mathbf{x}^p), \tag{10.13}$$

and obtain accordingly its feature importance $\mathbf{w}^{c_p}$ (Fig. 10.3h). Then we compute the distance between $\mathbf{x}^p$ against a feature vector of a gallery/target image $\mathbf{x}^g$ using the following function:

$$\text{dist}(\mathbf{x}^p, \mathbf{x}^g) = \|\left(\mathbf{w}^{c_p}\right)^{\mathsf{T}}|\mathbf{x}^p - \mathbf{x}^g|\|_1 \tag{10.14}$$

The matching ranks of $\mathbf{x}^p$ against a gallery of images can be obtained by sorting the distances computed from Eq. (10.14). A smaller distance results in a higher rank.

10.4.5 Fusion of Different Feature Importance Strategies

Contemporary methods [33, 41] learn a weight function that captures the global environmental viewing condition changes which cannot be derived from the unsupervised method described so far. Thus we investigate the fusion between the global feature weight matrix obtained from [33, 41] and our prototype-sensitive feature importance vector $\mathbf{w}$ to gain more accurate person re-identification performance.

In general, methods [33, 41] aim to optimise a distance metric so that a true match pair lies closer than a false match pair, given a set of relevance rank annotations. The distance metric can be written as

$$d(\mathbf{x}_i^p, \mathbf{x}_j^g) = (\mathbf{x}_i^p - \mathbf{x}_j^g)^{\mathsf{T}}\mathbf{V}(\mathbf{x}_i^p - \mathbf{x}_j^g). \tag{10.15}$$

The optimisation process involves finding a semi-positive definite global feature weight matrix $\mathbf{V}$. There exist several global feature weighting methods, most of them differing by different constraints and optimisation schemes they use (see Sect. 10.2 for discussion).

To combine our proposed prototype-sensitive feature importance with the global feature importance, we adopt a weighted sum scheme as follows:

$$\text{dist}_{\text{fusion}}(\mathbf{x}^p, \mathbf{x}^g) = \alpha \| \left(\mathbf{w}^{c_p}\right)^{\mathsf{T}} |\mathbf{x}^p - \mathbf{x}^g| \|_1 + (1-\alpha) \| \mathbf{V}^{\mathsf{T}} |\mathbf{x}^p - \mathbf{x}^g| \|_1, \quad (10.16)$$

where $\mathbf{V}$ is the global weight matrix obtained from Eq. (10.15) and α is a parameter that balances global and prototype-sensitive feature importance scores. We found that setting α in the range of $[0.1, 0.3]$ gives stable empirical performance across all the datasets we tested. We fix it to 0.1 in our experiments. 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. We shall show in the following evaluation that this fused metric is able to benefit from both feature importance mining from individual visual appearance changes, whilst taking into account the generic global environmental viewing condition changes between camera views.

10.5 Evaluation

In Sect. 10.5.2, we first investigate the re-identification performance of using different features given individuals with different inherent appearance attributes. In Sect. 10.5.3, the qualitative results of prototype discovery are presented. Sect. 10.5.4 then compares feature importance produced by the proposed unsupervised bottom-up prototype discovery method and two top-down GFI methods, namely RankSVM [33] and PRDC [41]. Finally, we report the results on combining the bottom-up and the top-down feature importance mining strategies.

10.5.1 Settings

We first describe the experimental settings and implementation details.

Datasets Four publicly available person re-identification datasets are used for evaluation. They are VIPeR [19], i-LIDS Multiple-Camera Tracking Scenario (i-LIDS) [40], QMUL underGround Re-IDentification (GRID) [27] and Person Re-IDentification 2011 (PRID2011) [20]. Example images of these datasets are shown in Fig. 10.4. More specifically,

1. The VIPeR dataset (see Fig. 10.4a) contains 632 persons, each of which has two images captured in two different 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 (see also Fig. 10.5a, b).
2. The i-LIDS dataset (see Fig. 10.4b) 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 illumination changes and pose

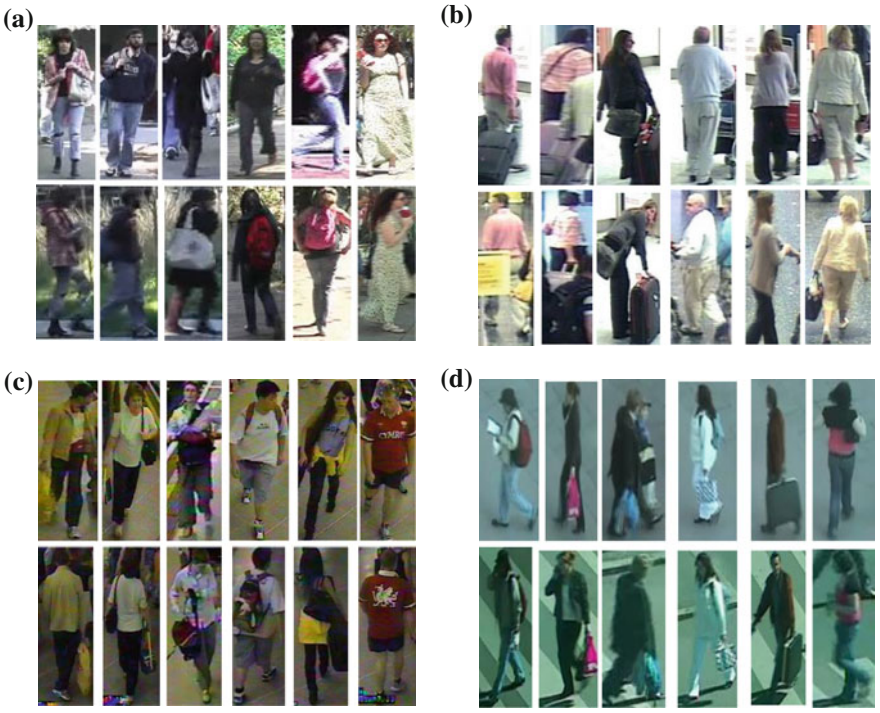

Fig. 10.4 Example images of different datasets used in our evaluation. Each column denotes an image pair of the same person. Note the large appearance variations within an image pair. In addition, note the unique appearance characteristics of different individuals, which can potentially be used to discriminate him/her from other candidates. **a** VIPeR, **b** i-LIDS, **c** GRID, **d** PRID2011

variations, many images in this dataset are also subject to severe inter-object occlusion (see also Fig. 10.5c, d).

3. The GRID dataset (see Fig. 10.4c) was captured from eight disjoint camera views installed in a busy underground station. It was divided into probe and gallery sets. The probe set contains 250 persons, whilst the gallery set contains 1,025 persons 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, large viewpoint variations and poor image quality (see also Fig. 10.5e, f).
4. The PRID2011 dataset (see Fig. 10.4d) was captured from two outdoor cameras. We use the single-shot version in which each person is only associated with one picture in a camera. The two cameras contains 385 and 749 individuals separately, within which the first 200 persons have two views. The challenge lies in severe lighting changes caused by the sunlight (see also Fig. 10.5g, h).

A summary of these datasets is given in Table 10.1.

Features In Sect. 10.5.2, we employ all the feature types discussed in Sect. 10.3 for a comprehensive evaluation of their individual performance in person

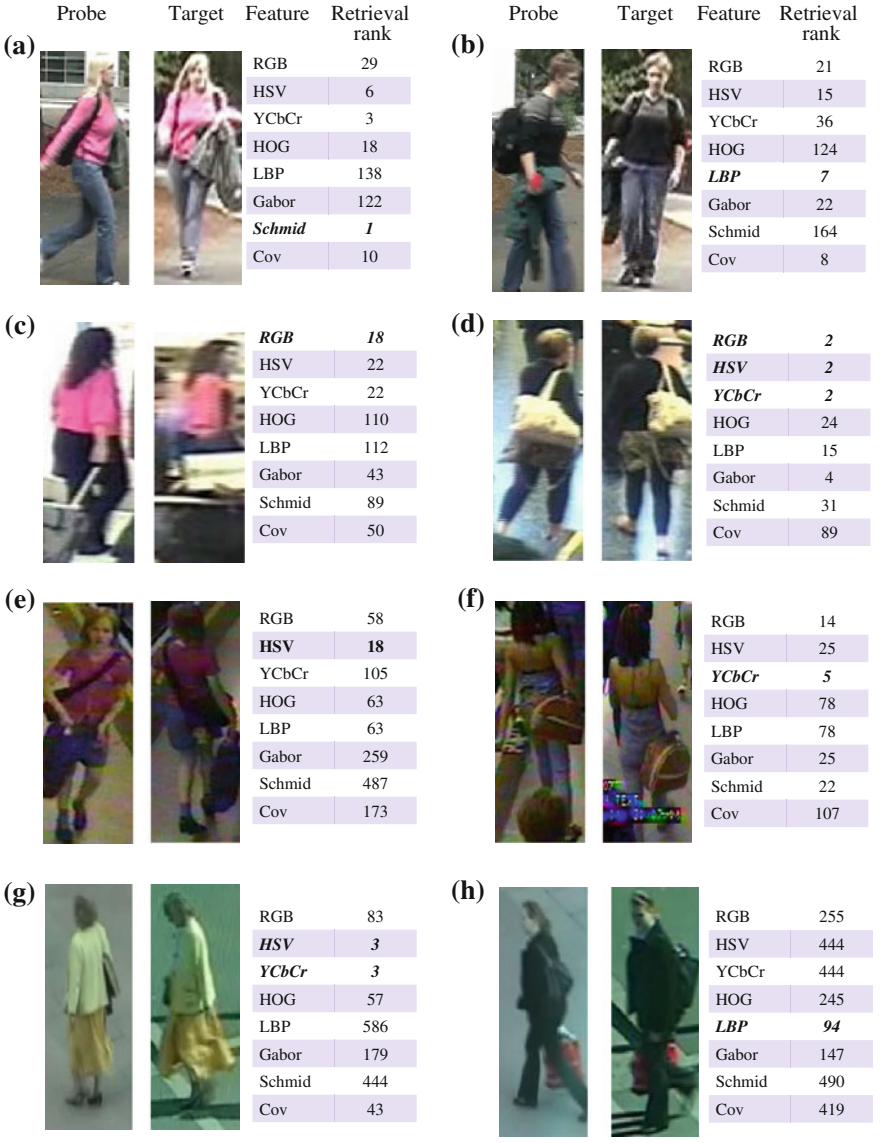

Fig. 10.5 Feature effectiveness in re-identification—in each subfigure, we show the probe image and the target image, together with the rank of correct matching by using different feature types separately

re-identification. In Sect. 10.5.3, we select from the aforementioned feature channels to form a feature subset, which is identical to those used in existing GFI mining methods [30, 33, 41]. Having the similar set of features allows a fair and comparable evaluation against the methods. Specifically, we consider 8 colour

Table 10.1 Details of the VIPeR, i-LIDS, GRID and PRID2011 datasets

Name	Environment	Resolution	#probe	#gallery	Challenges
VIPeR	Outdoor	48×128	632	632	Viewpoint and illumination changes
i-LIDS	Indoor airport arrival hall	An average of 60×150	119	119	Viewpoint and illumination changes and inter-object occlusion
GRID[a]	Underground station	An average of 70×180	250	1050	Inter-object occlusion and viewpoint variations
PRID2011[b]	Outdoor	64×128	385	749	Severe lighting change

[a] 250 matched pairs in both views
[b] 200 matched pairs in both views

channels (RGB, HSV and YCbCr)[1] and the 21 texture filters (8 Gabor filters and 13 Schmid filters) applied to luminance channel [33]. Each channel is represented by a 16-dimensional vector. Since we divide the human body into six strips and extract features for each strips, concatenating all the feature channels from all the strips thus results in a 2,784-dimensional feature vector for each image.

Evaluation Criteria We use $\ell 1$-norm as the matching distance metric. The matching performance is measured using an averaged cumulative match characteristic (CMC) curve [19] over 10 trials. The CMC curve represents the correct matching rate at the top r ranks. We select all the images of p person to build the test set. The remaining data are used for training. 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.

Implementation Details For prototype discovery, the number of cluster K is set to 5 for the i-LIDS dataset and 10 for the other three datasets, roughly based on the amount of training samples in each of the datasets. As for the forests' parameters, we set the number of trees of clustering and classification forests as $T_{\text{cluster}} = T_{\text{class}} = 200$. In general, we found that better performance is obtained when we increase the number of trees. For instance, the average rank 1 recognition rates on VIPeR dataset are 8.32 %, 9.56 % and 10.00 % when we set T_{cluster} to 50, 200 and 500, respectively. The depth of a tree is governed by two criteria—a tree will stop growing if the node size reaches 1, or the information gain is less than a pre-defined value.

10.5.2 Comparing Feature Effectiveness

We assume that certain features can be more important than others in describing an individual and distinguishing him/her from other people. To validate this hypothesis,

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

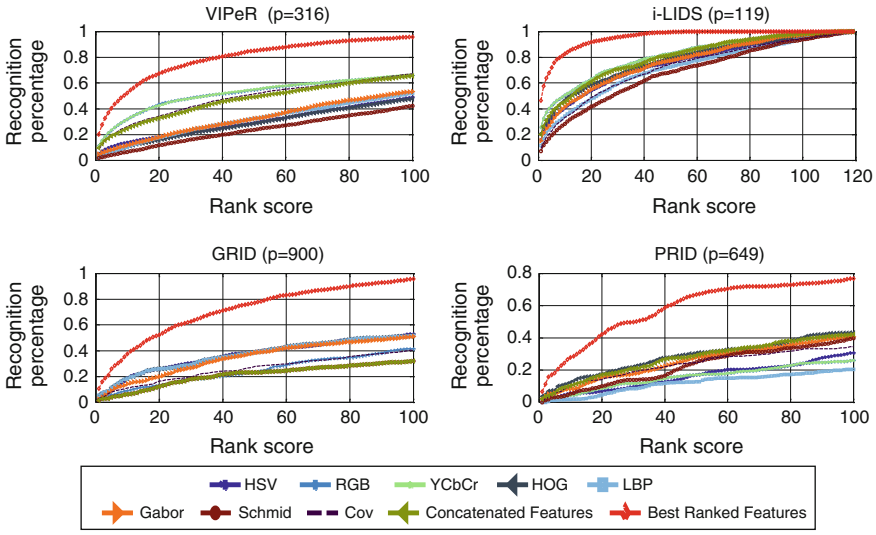

Fig. 10.6 The CMC performance comparison of using different features on various datasets. 'Concatenated Features' refers to the concatenation of all feature histograms with uniform weighting. In the 'Best Ranked Features' strategy, ranking for each individual was selected based on the best feature that returned the highest rank during matching. Its better performance suggests the importance and potential of selecting the right features specific to different individuals/groups

we analyse the matching performance of using different features individually as a proof of concept.

We first provide a few examples in Fig. 10.5 (also presented in Fig. 10.1) 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, e.g. Fig. 10.5a, c and g, the colour features generally yield a higher rank. For persons wearing clothing with rich texture (on the shirt or skirt), e.g. Fig. 10.5b and d, texture features especially the Gabor features and the LBP features tend to dominate. The results suggest that certain features can be more informative than others given different appearance attributes.

The overall matching performance of using individual feature types is presented in Fig. 10.6. In general, HSV and YCbCr features exhibit very close performances, which are much superior over all other features. This observation of colours being the most informative features agreed with the past studies [19]. Among the texture and structure features, the Gabor filter banks produce the best performance across all the datasets. Note that the performance of covariance feature can be further improved when combined with a more elaborative region partitioning scheme, as shown in [5].

One may consider concatenating all the features together, with the hope that these features could complement each other leading to better performance. From our experiments, we found that a naive concatenation of all feature histograms with uniform weighting does not necessary yield better performance (sometimes even

worse than using a single feature type), as shown by the 'Concatenated Features' performance in Fig. 10.6. The results suggest a more careful feature weighting is necessary based on the level of informativeness of each feature.

In the 'Best Ranked Features' strategy, the final rank is obtained by selecting the best feature that returned the highest rank for each individual, e.g. selecting HSV feature for Fig. 10.5e whilst choosing LBP feature for both Fig. 10.5b and h. As expected, the 'Best Ranked Features' strategy yields the best performance, i.e.37.80 %, 21.92 %, 15.28 % and 48.97 % improvement of AUC (area under curve) on the VIPeR, i-LIDS, GRID and PRID2011 datasets, respectively, in comparison to 'Concatenated Features'. The recognition rates at top ranks has been significantly increased across all the datasets. For example, on the i-LIDS dataset, the 'Best Ranked Features' obtains 92.02 % versus 56.30 % of concatenated features at rank 20.

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. 10.5, in general these 'Best Ranked Features' show consistency with the appearance attributes for each individual. Therefore, the results suggest that the overall matching performance can be boosted potentially by weighting features selectively according to the inherent appearance attributes.

10.5.3 Discovered Prototypes

It is non-trivial to weigh features in accordance to their associated inherent appearance attributes. We formulate a method to first discover prototypes, i.e. low-dimensional manifold clusters that aim to correlate features contributing towards similar appearance attributes.

Some examples of prototypes discovered from the VIPeR dataset are depicted in Fig. 10.7. Each colour-coded row represents a prototype. A short list of possible attributes discovered/interpreted in each prototype is given in the caption. Note that these inherent attributes are neither pre-defined nor pre-labelled, but discovered automatically by the unsupervised clustering forests (Sect. 10.4.2).

As shown by the example members in each prototype, images with similar attributes are likely to be categorised into the same cluster. For instance, a majority of images in the second prototype can be characterised with bright and high contrast attributes. In the fourth prototype, the key attributes are 'carrying backpack' and 'side pose'. These results demonstrate that the formulated prototype discovering mechanism is capable of generating reasonably good clusters of inherent attributes, which can be employed in subsequent step for prototype-sensitive feature importance mining.

Fig. 10.7 Examples of prototypes discovered automatically from the VIPeR dataset. 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) white shirt, dark trousers; (2) bright and colourful shirt; (3) dark jacket and jeans; (4) with backpack and side pose; (5) dark jacket and light colour trousers; (6) dark shirt with texture, back pose; (7) dark shirt and side pose; (8) dark shirt and trousers; (9) colourful shirt jeans; (10) colourful shirt and dark trousers

10.5.4 Prototype-Sensitive Versus Global Feature Importance

Comparing Prototype-Sensitive and Global Feature Importance The aim of this experiment is to compare different feature importance measures computed by existing GFI approaches [33, 41] and the proposed PSFI mining approach. The RankSVM [33] and PRDC [41] (see Sect. 10.1) were evaluated using the authors' original code. The global feature importance scores/weights were learned using the labelled images, and averaged over tenfold cross-validation. We set the penalty parameter C in RankSVM to 100 for all the datasets and used the default parameter values for PRDC.

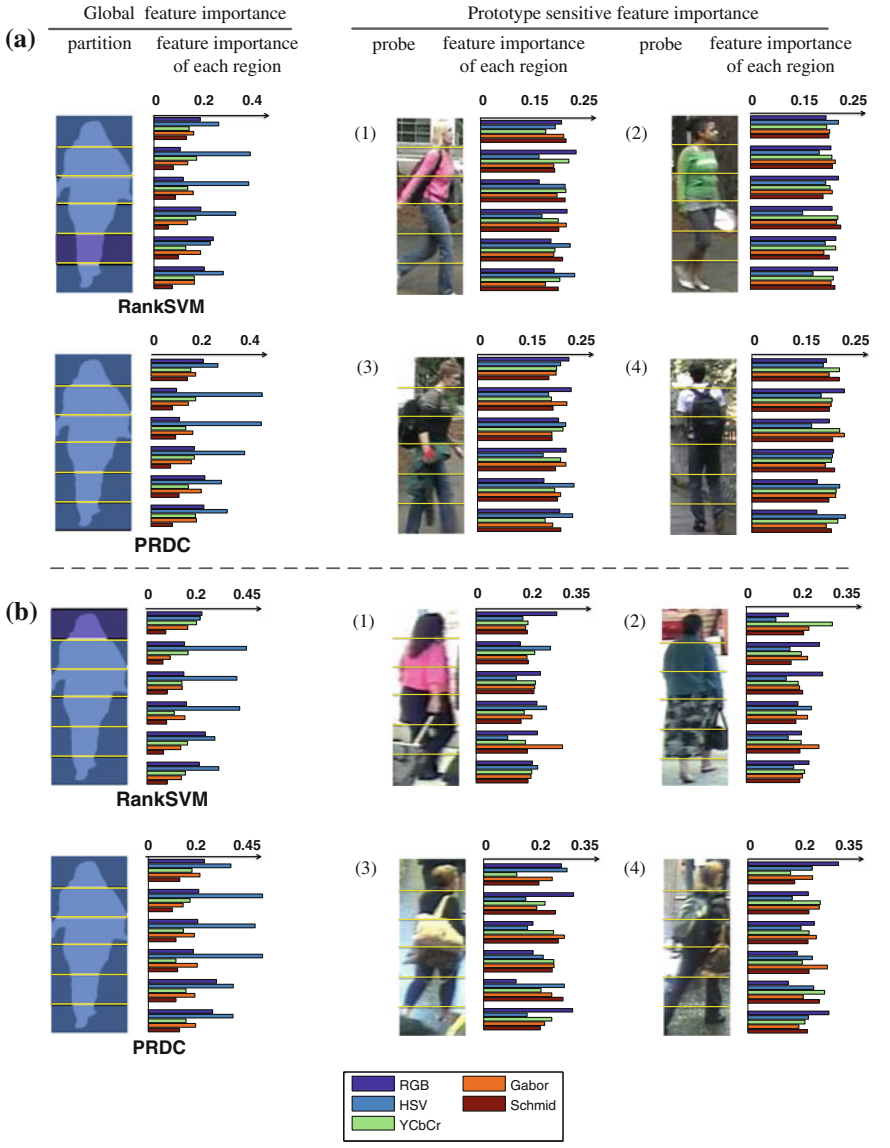

Fig. 10.8 Comparison of global feature importance weights produced by RankSVM [33] and PRDC [41] against those by prototype-sensitive feature importance. These results are obtained from the VIPeR and i-LIDS datasets

The left pane of Figs. 10.8 and 10.9 shows the feature importance discovered by both the RankSVM and PRDC. For PRDC, we only show the first learned orthogonal projection, i.e.feature importance. Each region in the partitioned silhouette images are masked with the labelling colour of the dominant feature. In the feature importance

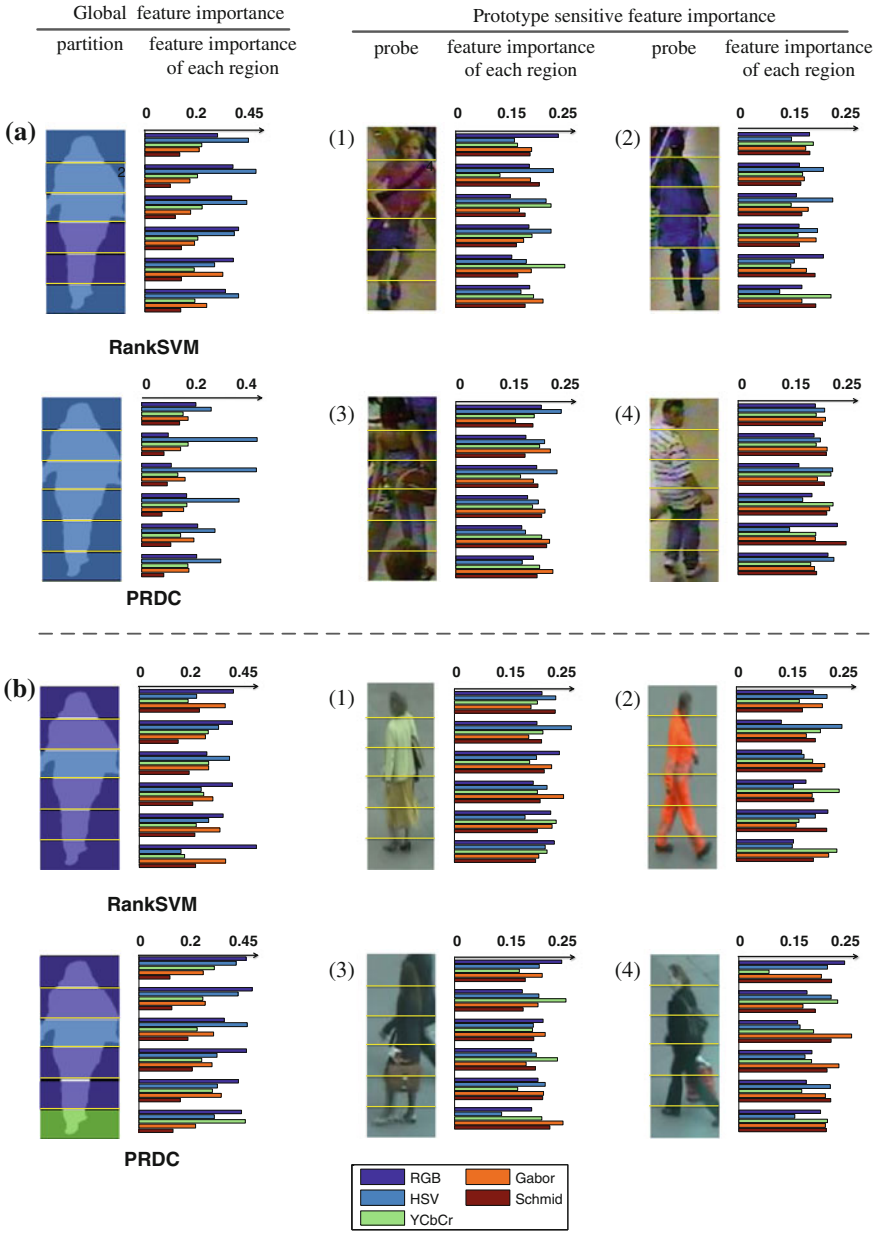

Fig. 10.9 Comparison of global feature importance weights produced by RankSVM [33] and PRDC [41] against those by prototype-sensitive feature importance. These results are obtained from the GRID and PRID2011 datasets

Table 10.2 Comparison of top rank matching rate (%) on the four benchmark datasets. r is the rank and p is the size of gallery set

Methods	VIPeR ($p = 316$)					i-LIDS ($p = 50$)				
	$r = 1$	$r = 5$	$r = 10$	$r = 20$		$r = 1$	$r = 5$	$r = 10$	$r = 20$	
GFI [27, 37]	9.43	20.03	27.06	34.68		30.40	55.20	67.20	80.80	
PSFI	9.56	22.44	30.85	42.82		27.60	53.60	66.60	81.00	
RankSVM [33]	14.87	37.12	50.19	65.66		29.80	57.60	73.40	84.80	
PSFI + RankSVM	**15.73**	**37.66**	**51.17**	**66.27**		**33.00**	**58.40**	**73.80**	**86.00**	
PRDC [41]	16.01	37.09	**51.27**	**65.95**		32.00	58.00	71.00	83.00	
PSFI + PRDC	**16.14**	**37.72**	50.98	**65.95**		**34.40**	**59.20**	**71.40**	**84.60**	
Methods	GRID ($p = 900$)					PRID2011 ($p = 649$)				
	$r = 1$	$r = 5$	$r = 10$	$r = 20$	$r = 50$	$r = 1$	$r = 5$	$r = 10$	$r = 20$	$r = 50$
GFI [27, 37]	4.40	11.68	16.24	24.80	36.40	3.60	6.60	9.60	16.70	31.60
PSFI	5.20	12.40	19.92	28.48	40.80	0.60	2.00	4.00	7.30	14.20
RankSVM [33]	10.24	24.56	33.28	43.68	**60.96**	4.10	8.50	**12.50**	18.90	31.70
PSFI + RankSVM	**10.32**	**24.80**	**33.76**	**44.16**	60.88	**4.20**	**8.90**	**12.50**	**19.70**	**32.20**
PRDC [41]	**9.68**	22.00	**32.96**	44.32	64.32	**2.90**	**9.50**	15.40	23.00	38.20
PSFI + PRDC	9.28	**23.60**	32.56	**45.04**	**64.48**	**2.90**	9.40	**15.50**	**23.60**	**38.80**

plot, we show in each region the importance of each type of features. The importance of a certain feature type is derived by summing the weight of all the histogram bins that belong to this type. The same steps are repeated to depict the prototype-sensitive feature importance on the right pane.

In general, the global feature importance emphasises 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. This weight assignment for feature importance mining is applied universally to all images. In contrast, the prototype-sensitive feature importance is more adaptive to changing viewing conditions and appearance characteristics. For example, for image regions with colourful appearance, e.g.Figs. 10.8a-1 and 10.9b-2, the colour features in torso region are assigned with higher weights than the texture features. For image regions with rich texture, such as the stripes on the jumper (Fig. 10.8a-3), floral skirt (Fig. 10.8b-2) and bag (Figs. 10.8a-4, 10.8b-4, 10.9b-3 and 10.9b-4), the importance of texture features increase. For instance, in Fig. 10.8b-2, the weight of Gabor feature in the fifth region is 36.7 % higher than that observed in the third region.

Integrating Global and Prototype-Sensitive Feature Importance As shown in Table 10.2, in comparison to the baseline GFI [27, 37], PSFI yields improved matching rate on the VIPeR and GRID datasets. No improvement is observed on the i-LIDS and PRID2011 datasets. A possible reason is the small training size in the i-LIDS and PRID2011 dataset, which leads to suboptimal prototype discovery. This can be resolved by collecting more unannotated images for unsupervised prototype discovery. We integrate both global and prototype-sensitive feature importance following the method described in Sect. 10.4 by setting $\alpha = 0.1$. An improvement as much

as 3.2 % on rank 1 matching rate can be obtained when we combine our method with RankSVM [33] and PRDC [41] on these datasets. It is not surprising to observe that the supervised learning-based approaches [33, 41] outperform our unsupervised approach. Nevertheless, the global approaches benefit from slight bias of feature weights driven by specific appearance attributes of individuals. The results suggest that these two kinds of feature importance are not exclusive, but can complement each other to improve re-identification accuracy.

10.6 Findings and Analysis

In this study, we investigated the effect of feature importance for person re-identification. We summarise our main findings as follows:

Mining Feature Importance for Person Re-Identification Our evaluation shows that certain appearance features are more important than others in describing an individual and distinguishing him/her from other people. In general, colour features are dominant, not surprisingly, for person re-identification and outperform the texture or structure features, though illumination changes may cause instability in the colour features. However, texture and structure features take greater effect when the appearances contain noticeable local statistics, caused by bag, logo and repetitive patterns.

Combining various features for robust person re-identification is non-trivial. Naively concatenating all the features and applying uniform global weighting to them does not necessarily yield better performance in re-identification. Our results show a tangible indication that instead of biasing all the weights to features that are presumably good for all individuals, distributing selectively some weights to informative feature specific to certain appearance attributes can lead to better re-identification performance.

We also find that the effectiveness of prototype-sensitive feature importance mining is dependent on the quantity and quality of training data, in terms of the available size of the training data and the diversity of underlying attributes in appearance, i.e. sufficient and non-biased sampling in the training data. First, as shown in the experiment on the i-LIDS dataset, a sufficient number of unlabelled data are desired to generate robust prototypes. Second, 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. For example, in the PRID2011 dataset, images with rich structural and texture features are rare. Therefore, the derived feature importance scores for those features are prone to be erroneous.

Hierarchical Feature Importance for Person Re-Identification The global feature importance and prototype-sensitive feature importance can be seen organising themselves in a hierarchical structure, as shown in Fig. 10.10. Specifically, the global feature importance exploited by existing rank learning [33] or distance learning method [21, 41] learns a feature weighting function to accommodate the

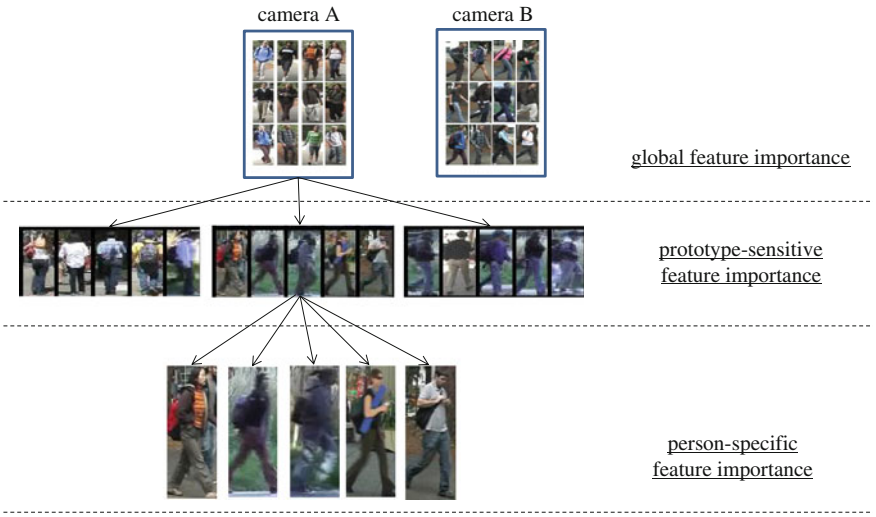

Fig. 10.10 Hierarchical structure of feature importance. Global feature importance aims at weighing more on those features that remain consistent between cameras from a statistical point of view. Prototype-sensitive feature importance emphasises more on the intrinsic features which can discriminate a given prototype from the others. Person-specific feature importance should be capable of distinguishing a given person from those who are categorised into the same prototype

feature inconsistency between different cameras, caused by illumination changes or viewpoint variations. The discovered feature weights can be treated as feature importance in the highest level of the hierarchy, without taking specific individual appearance characteristics into account. Whilst the prototype-sensitive feature importance aims to emphasise more on the intrinsic feature properties that can discriminate a given prototype from the others. Our study shows that these two kinds of feature importance in different levels of the hierarchy can be complementary to each other in improving re-identification accuracy.

Though the proposed prototype-sensitive feature importance is capable of reflecting the intrinsic/salient appearance characteristics of a given person, it still lacks the ability to differentiate the disparity between two different individuals who fall into the same prototype. Thus, it would be interesting to investigate person-specific feature importance that is unique to a specific person, which allows the manifestation of subtle differences among individuals belong to the same prototype.

References

1. Alahi, A., Vandergheynst, P., Bierlaire, M., Kunt, M.: Cascade of descriptors to detect and track objects across any network of cameras. Comput. Vis. Image Underst. **114**(6), 624–640 (2010)
2. Avraham, T., Gurvich, I., Lindenbaum, M., Markovitch, S.: Learning implicit transfer for person re-identification. In: European Conference on Computer Vision, First International Workshop on Re-Identification, pp. 381–390 (2012)

3. Bak, S., Corvee, E., Bremond, F., Thonnat, M.: Multiple-shot human re-identification by mean Riemannian covariance grid. In: IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 179–184 (2011)
4. Bak, S., Corvee, E., Brémond, F., Thonnat, M.: Person re-identification using haar-based and DCD-based signature. In: IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 1–8 (2010)
5. Bak, S., Corvee, E., Brémond, F., Thonnat, M.: Person re-identification using spatial covariance regions of human body parts. In: IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 435–440 (2010)
6. Bak. S., Charpiat, G., Corvée, E., Brémond, F., Thonnat, M.: Learning to match appearances by correlations in a covariance metric space. In: European Conference on Computer Vision, pp. 806–820 (2012)
7. Bauml, M., Stiefelhagen, R.: Evaluation of local features for person re-identification in image sequences. In: IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 291–296 (2011)
8. Bazzani, L., Cristani, M., Perina, A., Murino, V.: Multiple-shot person re-identification by chromatic and epitomic analyses. Pattern Recogn. Lett. **33**(7), 898–903 (2012)
9. Bazzani, L., Cristani, M., Murino, V.: Symmetry-driven accumulation of local features for human characterization and re-identification. Comput. Vis. Image Underst. **117**(2), 130–144 (2013)
10. Berg, T.L., Berg, A.C., Shih, J.: Automatic attribute discovery and characterization from noisy web data. In: European Conference on Computer Vision, pp. 663–676 (2010)
11. Breiman, L., Friedman, J., Stone, C., Olshen, R.: Classification and regression trees. Chapman and Hall/CRC, Boca Raton (1984)
12. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
13. Caruana, R., Karampatziakis, N., Yessenalina, A.: An empirical evaluation of supervised learning in high dimensions. In: International Conference on, Machine learning, pp. 96–103 (2008)
14. Cheng, D., Cristani, M., Stoppa, M., Bazzani, L., Murino, V.: Custom pictorial structures for re-identification. In: British Machine Vision Conference, pp. 68.1–68.11 (2011)
15. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. IEEE Comput. Vis. Pattern Recogn. **1**, 886–893 (2005)
16. Doretto, G., Sebastian, T., Tu, P., Rittscher, J.: Appearance-based person reidentification in camera networks: problem overview and current approaches. J. Ambient Intell. Humanized Comput. **2**(2), 127–151 (2011)
17. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: IEEE Conference Computer Vision and, Pattern Recognition, pp. 2360–2367 (2010)
18. Farhadi, A., Endres, I., Hoiem, D., Forsyth, D.: Describing objects by their attributes. In: IEEE Conference on Computer Vision and, Pattern Recognition, pp. 1778–1785 (2009)
19. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: European Conference on Computer Vision, pp. 262–275 (2008)
20. Hirzer, M., Beleznai, C., Roth, P., Bischof, H.: Proceedings of the 17th Scandinavian Conference on Image Analysis, Springer-Verlag, 91–102 (2011)
21. Hirzer, M., Roth, P., Köstinger, M., Bischof, H.: Relaxed pairwise learned metric for person re-identification. In: European Conference on Computer Vision, pp. 780–793 (2012)
22. Javed, O., Rasheed, Z., Shafique, K., Shah, M.: Tracking across multiple cameras with disjoint views. In: International Conference on Computer Vision, pp. 952–957 (2003)
23. Layne, R., Hospedales, T., Gong, S.: Person re-identification by attributes. In: British Machine Vision Conference (2012)
24. Liu, C., Wang, G., Lin, X.: Person re-identification by spatial pyramid color representation and local region matching. IEICE Trans. Inf. Syst. **E95-D**(8), 2154–2157 (2012)
25. Liu, B., Xia, Y., Yu, P.S.: Clustering through decision tree construction. In: International Conference on Information and, Knowledge Management, pp. 20–29 (2000)

26. Loy, C.C., Liu, C., Gong, S.: Person re-identification by manifold ranking. In: IEEE International Conference on Image Processing (2013)
27. Loy, C.C., Xiang, T., Gong, S.: Time-delayed correlation analysis for multi-camera activity understanding. Int. J. Comput. Vis. **90**(1), 106–129 (2010)
28. Loy, C.C., Xiang, T., Gong, S.: Incremental activity modelling in multiple disjoint cameras. IEEE Trans. Pattern Anal. Mach. Intell. **34**(9), 1799–1813 (2012)
29. Ma, S., Sclaroff, S., Ikizler-Cinbis, N.: Unsupervised learning of discriminative relative visual attributes. In: European Conference on Computer Vision, Workshops and Demonstrations, pp. 61–70 (2012)
30. Mignon, A., Jurie, F.: PCCA: A new approach for distance learning from sparse pairwise constraints. In: IEEE Conference Computer Vision and, Pattern Recognition, pp. 2666–2672 (2012)
31. Ng, A.Y., Jordan, M.I., Weiss, Y., et al.: On spectral clustering: analysis and an algorithm. Adv. Neural Inf. Process. Syst. **2**, 849–856 (2002)
32. Perona, P., Zelnik-Manor, L.: Self-tuning spectral clustering. Adv. Neural Inf. Process. Syst. **17**, 1601–1608 (2004)
33. Prosser, B., Zheng, W., Gong, S., Xiang, T.: Person re-identification by support vector ranking. In: British Machine Vision Conference, pp. 21.1–21.11 (2010)
34. Satta, R., Fumera, G., Roli, F.: Fast person re-identification based on dissimilarity representations. Pattern Recogn. Lett. **33**(14), 1838–1848 (2012)
35. Schulter, S., Wohlhart, P., Leistner, C., Saffari, A., Roth, P.M., Bischof, H.: Alternating decision forests. In: IEEE Conference Computer Vision and Pattern Recognition (2013)
36. Schwartz, W., Davis, L.: Learning discriminative appearance-based models using partial least squares. In: Brazilian Symposium on, Computer Graphics and Image Processing, pp. 322–329 (2009)
37. Wang, X.G., Doretto, G., Sebastian, T., Rittscher, J., Tu, P.: Shape and appearance context modeling. In: International Conference on Computer Vision, pp. 1–8 (2007)
38. Xiang, T., Gong, S.: Spectral clustering with eigenvector selection. Pattern Recogn. **41**(3), 1012–1029 (2008)
39. Zhang, Y., Li, S.: Gabor-LBP based region covariance descriptor for person re-identification. In: International Conference on Image and Graphics, pp. 368–371 (2011)
40. Zheng, W., Gong, S., Xiang, T.: Associating groups of people. In: British Machine Vision Conference, pp. 23.1–23.11 (2009)
41. Zheng, W., Gong, S., Xiang, T.: Re-identification by relative distance comparison. IEEE Trans. Pattern Anal. Mach. Intell. **35**(3), 653–668 (2013)