Cost-effective features for re-identification in camera networks

Syed Fahad Tahir and Andrea Cavallaro

***Abstract*—Networks of smart cameras share large amounts of data to accomplish tasks such as re-identification. We propose a feature selection method that minimizes the data needed to represent the appearance of objects by learning the most appropriate feature set for the task at hand (person re-identification). The computational cost for feature extraction and the cost for storing the feature descriptor are considered jointly with feature performance in order to select cost-effective good features. This selection allows us to improve inter-camera re-identification while reducing the bandwidth that is necessary to share data across the camera network. We also rank the selected features in the order of effectiveness for the task to enable a further reduction of the feature set by dropping the least effective features when application constraints require this adaptation. We compare the proposed approach with state-of-the-art methods on the i-LIDS and VIPeR datasets and show that the proposed approach considerably reduces network traffic due to inter-camera feature sharing while keeping the re-identification performance at an equivalent or better level compared with the state of the art.**

***Index Terms*—Smart camera networks, person re-identification, data reduction, cost of features, feature selection.**

I. Introduction

SMART camera networks are composed of nodes that perform image processing locally and aim to transfer the minimum amount of data over the network to accomplish collaborative tasks such as object detection and tracking [1], [2]. The challenges involved in smart camera networks include reducing the amount of data to be processed and shared across the network, real-time operation and energy efficiency [3], [4]. Data reduction can be achieved by using image compression [5], [6] and metadata reduction, which are generally obtained via trade-offs between data-transmission rate and distortions [2].

An important example of the use of metadata (features) is for describing objects such as people for their re-identification across the network [7]–[9]. Multiple object features are combined and shared among cameras to improve performance [8]–[16]. Gabor and Schmid filters that define two kernels applied to the luminance channel to extract textures may be used, followed by learning and probabilistic approaches for feature optimization [14]–[17]. The histogram of the RGB [12], [13] and the HSV [10] color space are also used as features. Alternatively, the two chrominance channels of the YUV space are used [18] with a Gaussian Mixture Model to encode the most relevant color clusters. Features from multiple color spaces (RGB, YCbCr, and SV) and texture types (Gabor and Schmid) may be concatenated to increase their discriminative power [14]–[16]. However, after a certain number of feature concatenations, any additional features might decrease the re-identification performance. This problem can be addressed by feature selection. However, very little work has been done to consider the *cost* of a feature, such as, for example, the computational time for its extraction and the amount of data that is necessary for its storage. We argue that the cost of a feature should be considered jointly with its *performance*, namely its ability to represent and to discriminate an object for the task at hand. The cost constraint in feature selection becomes particularly important when the cost varies significantly across features to be shared among nodes of a smart camera network.

In this paper, we propose a Cost-and-Performance-Effective (CoPE) feature selection method that combines the cost of using features with their performance in order to identify the most inexpensive feature subset for person re-identification in a smart camera network. To the best of our knowledge, the proposed approach is the first to apply explicit feature selection for re-identification while considering the cost of features. Instead of optimizing the combined contribution of the best set of features, the most discriminative, well-performing and cost-effective features are selected by evaluating each feature individually and then by ranking the selected features based on their contribution to the task, thus making the approach scalable in transmitting data over the network. CoPE is evaluated in the person re-identification task using Direct Distance Minimization (DDM). The results are compared with two existing re-identification approaches, namely PRDC [15] and ASFI [17], and five feature selection approaches: Fisher score [19], Information gain [20], mRMR [21], ReliefF [22] and Bi-clusters [23]. We also demonstrate that CoPE features further improve the performance of learning methods for re-identification with rankSVM [14] and AdaBoost [16] by reducing the feature dimensions, the training time and by improving their effectiveness. The software of the proposed method is available at *http://www.eecs.qmul.ac.uk/~andrea/software.htm*.

The paper is organized as follows. Section II discusses the related work on feature selection and the features used in person re-identification. Section III analyzes the performance

The authors are with the Centre for Intelligent Sensing, Queen Mary University of London, E1 4NS, London, UK. (email: s.fahad.tahir@eecs.qmul.ac.uk; andrea.cavallaro@eecs.qmul.ac.uk). S.F. Tahir was supported by the Erasmus Mundus Joint Doctorate in Interactive and Cognitive Environments, which is funded by the Education, Audiovisual & Culture Executive Agency (FPA n° 2010-0012). This work was partially supported by the Artemis JU and UK Technology Strategy Board as part of the Cognitive & Perceptive Cameras (COPCAMS) project under GA number 332913.

and the cost of features, defines their combination strategy and discusses the implementation in multi-camera networks under constrained environments. In Sec. IV, the proposed approach is validated and compared with existing feature selection and re-identification approaches using the challenging i-LIDS and VIPeR datasets. Finally, Sec. V draws conclusions.

II. Related work

In this section, we discuss relevant feature selection approaches and the features used in the re-identification, while highlighting properties and inadequacies of current feature selection and re-identification approaches.

A. Feature selection

Feature selection aims at finding the most important features and their combinations for effectively describing and matching objects [24], [25]. Feature selection is an NP-hard problem. Approaches based on heuristics exist, which approximate the solution by exploiting problem-specific properties. Selection approaches produce a subset of features [26] and reduce redundancies among features [27]. Feature selection is also an important pre-processing step in machine learning to avoid over-fitting and to increase the effectiveness of learning. Features can be selected either based on *group performance* or on their *individual performance* [28]. The set of individually selected features may not collectively provide good classification performance due to the lack of information about feature correlation, while individual weak features may provide strong discriminatory power in a group [23]. However, individually selected features can perform well in constrained environments when some features need to be discarded adaptively because of user requirements or application constraints with resource-constrained devices; whereas in the case of feature grouping, the removal of a single feature may significantly reduce the effectiveness of the whole feature set.

A method for *ranking* of selected features according to their contribution to the tasks is presented in [29]. The similarity measures between each feature and all the rest are added to generate a feature score. The highest-scoring feature is selected and the process is repeated to choose the next relevant feature. The feature importance and similarity between features can be exploited with a greedy selection method [27] or boosted regression trees can be applied [30]. A hierarchical feature selection method is developed by using rankSVM along with a quality measure to predict the number of selected features [31]. The best-first search can be used to partition the features into subsets that are then combined to maximize the defined information retrieval measures [25]. The coherence between subgroups of data can also be used to rank features [23]. An approach based on cooperative game theory evaluates the power of each feature individually and within groups [32]. The structural similarity between data before and after feature selection is maintained and topological neighborhood information is used for computing the structural similarity [33]. An unsupervised feature ranking algorithm can discover Bi-clusters that are used to evaluate feature inter-dependencies, separability of instances and feature ranking [23]. This approach inherits some characteristics from ranking and wrappers, which use learning methods for feature selection and are classifier-dependent. A minimum-redundancy maximum-relevance (mRMR) based approach can be combined with a wrapper method to select a more compact subset from the candidate features [21]. A kernel-based feature selection criterion incorporates the kernel trick with the class separability measures [34], where the kernel parameters are automatically tuned by maximizing kernel class separability criteria. Feature selection based on a distance discriminant method converts the search problem of feature selection into feature ranking. The approach achieves feature selection performance comparable to exhaustive-search methods with a lower computational complexity [35]. Finally, hierarchical clustering is applied to select the optimal feature subset [28]. Table I summarizes and compares state-of-the-art feature selection methods.

TABLE I
State-of-the-art feature selection methods.

		[27]	[33]	[28]	[31]	[32]	[23]	[25]	[34]	[35]	[21]	[22]	[17]	CoPE
Selection approach	Best First Search	✓						✓						✓
	Structural Similarity		✓											
	Feature Cooperation			✓										
	Hierarchical Clustering			✓	✓									
	Game Theory					✓								
	Co-ordinate Ascent							✓						
	Kernel Class Separability								✓					
	Random forest												✓	
	Bi-clusters						✓							
	mRMR										✓			
	ReliefF											✓		
	Distance Discriminant									✓				✓
Dataset	Text Retrieval	✓												
	Medical Data	✓												
	UCI ML Benchmarks		✓	✓		✓	✓		✓	✓	✓			
	LETOR 4.0				✓									
	Handwriting Images								✓					
	Carnegie Mellon Datasets									✓				
	Bio-Informatics						✓				✓			
	UCI regression											✓		
	Surveillance Videos												✓	✓
Evaluation criteria	Performance	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Cost													✓

B. Person re-identification

Various feature types have been used for image-based person re-identification [7], [8]. Features can be extracted from the image using a Region Covariance Descriptor (RCD) [36] to preserve shape, location and color information also using a multi-scale quadtree descriptor [37]. An unsupervised clustering approach based on appearance attributes can be used to mine Attribute Sensitive Feature Importance (ASFI) [17], which are then combined with global features. Random forests can be used to group images of the same person into sub-clusters based on color and texture features. Features relevant to each sub-cluster are weighted to improve the re-identification rate. A spatio-temporal relationship is learned to find the probability of matching a person from one camera to another, coupled with an adaptive Brightness Transfer Function (BTF) to handle illumination changes [38].

Local Binary Patterns (LBP) are used to describe spatial patterns using normalized color intensities, followed by a pixel-based thresholded color distance to depict structural information [39]. The Mean Riemannian Covariance Grid [40] is used to generate a human signature from the detected persons using LBP on the head regions [41]. Histograms of Oriented Gra-

dients (HOGs), the Scale Invariant Feature Transform (SIFT) and weighted HSV histograms can be used for shape, texture and chromatic content to build a discriminative signature along with a weighted distance measure [42]. The mean color values from small image regions can be combined with the histogram of LBP to represent an image, and then pairwise sample differences are learned for re-identification [43]. Multi-shot re-identification approaches combine the information from multiple instances of the same object [11]. HSV histograms of the image-epitome group extract patches from multi-images with similar properties. Alternatively, the image can be divided into small components and the difference from an existing bag of components is considered. The difference vector is represented as a descriptor of the image [44]. For a detailed analysis of the state of the art in person re-identification and related taxonomy the reader is referred to [7]–[9].

Existing re-identication approaches exploit features for improving the re-identication rate without considering constraints on resource utilization, which significantly vary between features. This limits their usability and scalability in real-world applications. Unlike existing methods, the proposed approach, CoPE, selects good-performing less-expensive features.

III. Cost-and-performance-effective (CoPE) feature selection

Let $\mathbf{C} = \{C_n\}_{n=1}^{N}$ be a network of N cameras and $\mathbf{P}_n = \{P_{mn}\}_{m=1}^{M}$ be a set of M persons in the Field of View (FoV) of C_n. Each P_{mn} is represented by a feature set $\mathbf{F}_{mn} = \{\mathbf{f}_{mn}^r\}_{r=1}^{R}$ containing R feature types $\mathbb{f}^r$, where $r = 1 \cdots R$.

To characterize a feature we consider a performance vector and a cost vector. The performance vector, $\mathbf{\Pi}_r$, measures the discriminating ability of $\mathbb{f}^r$ in person re-identification; whereas the cost vector, $\mathbf{\Psi}_r$, measures the extraction time and the storage size associated to $\mathbb{f}^r$. $\mathbf{\Pi}_r$ and $\mathbf{\Psi}_r$ are iteratively combined to generate cost-effective well-performing features.

A. Feature performance

Let a training set be composed of $\mathbf{P}_1 = \{P_{k1}\}_{k=1}^{M}$ and $\mathbf{P}_2 = \{P_{m2}\}_{m=1}^{M}$, where M is the number of persons visible in both C_1 to C_2. The same value of k and m represents the same person. The feature sets $\mathbf{F}_{k1} = \{\mathbf{f}_{k1}^r\}_{r=1}^{R}$ and $\mathbf{F}_{m2} = \{\mathbf{f}_{m2}^r\}_{r=1}^{R}$ are extracted from person P_{k1} and P_{m2}, respectively. We measure the performance of a feature by analyzing the similarity of the representation of the same person in two views using feature $\mathbb{f}^r$ as well as the similarity with the other $M-1$ people.

We define the performance vector $\mathbf{\Pi}_r$ representing the performance of $\mathbb{f}^r$ on M persons. We measure the similarity between two instances $\mathbf{f}_{k1}^r$ and $\mathbf{f}_{m2}^r$ of $\mathbb{f}^r$ by a relative matching distance $D(\cdot)$, which receives as input a feature pair and returns the feature similarity d_{mk}^r between P_{k1} and P_{m2}:

$$d_{mk}^r = D(\mathbf{f}_{k1}^r, \mathbf{f}_{m2}^r). \tag{1}$$

For each P_{m2}, we measure M distances from $\mathbf{P}_1$. Each d_{mk}^r is then normalized ($0 \leq \hat{d}_{mk}^r \leq 1$) as

$$\hat{d}_{mk}^r = \frac{d_{mk}^r - \min_k d_{mk}^r}{\max_k d_{mk}^r - \min_k d_{mk}^r}, \tag{2}$$

where $\min_k d_{mk}^r$ and $\max_k d_{mk}^r$ are, respectively, the minimum and the maximum distances of P_{m2} from $\mathbf{P}_1$ using $\mathbb{f}^r$. The set of M normalized distances $\hat{\mathbf{d}}_m^r = \{\hat{d}_{mk}^r\}_{k=1}^{M}$ contains one distance corresponding to the same person in C_1 and C_2 ($\hat{d}_{mm}^r$: *correct* matching distance) and $M-1$ distances of P_{m2} from the instances of other persons in C_1 ($\mathbf{\Omega}_m^r$: the set of *incorrect* matching distances).

In the ideal case, a feature $\mathbb{f}^r$ is considered *well-performing* for P_{m2} if the distance in $\mathbb{f}^r$ between the correct matching pair is smaller than the minimum value of distances in $\mathbf{\Omega}_m^r$. However, in real-world re-identification scenarios the ideal condition is fulfilled only for a limited number of people. Therefore, we relax the criterion and use the median value of the incorrect distances, $\widetilde{\mathbf{\Omega}}_m^r$, and measure the performance score Π_{rm} as

$$\Pi_{rm} = \frac{\hat{d}_{mm}^r}{\widetilde{\mathbf{\Omega}}_m^r}. \tag{3}$$

The condition $\hat{d}_{mm}^r < \widetilde{\mathbf{\Omega}}_m^r$ leads to $0 \leq \Pi_{rm} < 1$ in Eq. 3. The smaller Π_{rm}, the better the performance. $\Pi_{rm} \geq 1$ indicates that $\mathbb{f}^r$ performs poorly. For each $\mathbb{f}^r$, we define the performance vector $\mathbf{\Pi}_r$ using M persons as

$$\mathbf{\Pi}_r = \{\Pi_{rm}\}_{m=1}^{M}, \tag{4}$$

where each element Π_{rm} corresponds to the performance score of $\mathbb{f}^r$ for a single person in the training data. $\mathbb{f}^r$ with $\min_m \Pi_{rm} \geq 1$ for all M persons are discarded before performing the feature selection thus resulting in $\hat{R} \leq R$ remaining features. We then define the $\hat{R} \times M$ performance matrix $\mathbf{\Delta}$ as

$$\mathbf{\Delta} = \left[\Pi_{rm}\right]_{\hat{R} \times M}, \tag{5}$$

where $r = 1, \ldots, \hat{R}$ and $m = 1, \ldots, M$. $\mathbf{\Pi}_r$ is the r^{th} row representing the performance of $\mathbb{f}^r$ for $\mathbf{P}_2$, while χ_m is the m^{th} column representing the performance comparison of P_{m2} for $\hat{R}$ features. $\mathbf{\Delta}$ is further analyzed jointly with the cost of features as discussed next.

B. Feature cost

We define the cost vector of feature $\mathbb{f}^r$ by considering two components, namely its storage size, β_{mn}^r, and the computational time for feature extraction, Γ_{mn}^r. The range of both components can vary largely across different $\mathbb{f}^r$. To avoid one of the two components dominating the other for feature selection in case of large values (e.g. large storage size or high computational time), we define the cost vector, $\mathbf{\Psi_r}$, as the inverse of the average of the two cost components:

$$\mathbf{\Psi}_r = (\Psi_{\Gamma r}, \Psi_{\beta_r}) = \left(\frac{\alpha MN}{\sum_{n=1}^{N}\sum_{m=1}^{M} \Gamma_{mn}^r}, \frac{(1-\alpha)MN}{\sum_{n=1}^{N}\sum_{m=1}^{M} \beta_{mn}^r} \right), \tag{6}$$

where $\alpha \in [0, 1]$ is a weight that helps accounting for cases when one constraint is more important than the other (e.g. when a limited storage space is available with no constraints on the computational time, or vice versa). In this paper we consider the byte as the unit for β_{mn}^r and the millisecond

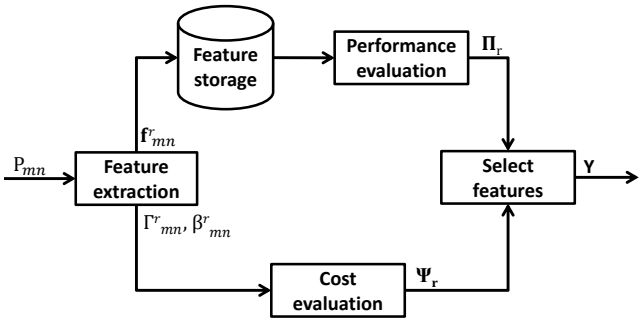


Fig. 1. Block diagram of the proposed Cost-and-performance-effective (CoPE) feature selection approach.

(ms) as the unit for Γ^r_{mn}. We measure the magnitude of the cost vector by calculating the Euclidean norm $||\mathbf{\Psi}_r||$, where $||\mathbf{\Psi}_r|| \in [0, 1]$. The larger $||\mathbf{\Psi}_r||$, the cheaper the feature. Note that as this cost score is obtained by combining two independent components in Eq. 6, new cost constraints can be included as additional independent components of the vector.

C. Feature selection

We perform a competitive feature selection such that the least costly features exhibiting the best performance are selected by exploiting $\mathbf{\Psi}_r$ and $\mathbf{\Pi}_r$ of each $\mathbb{f}^r$ (Fig. 1). We define a vector $\mathbf{V}$ that contains the elements $\Pi_{rm} \leq 1$ from $\mathbf{\Delta}$ sorted in ascending order. We divide $\mathbf{V}$ into bins I_i, where $1 \leq i \leq \hat{R}$ and each bin contains M performance scores such that in the best case a single feature with the best performance for all the M persons can be selected in a single iteration. A set $\mathbf{\Phi}_{ir}$ is defined which contains the performance scores Π_{rm} within I_i for each $\mathbb{f}^r$.

Figure 2 shows an example of performance matrix $\mathbf{\Delta}$, and highlights the vector $\mathbf{V}$, the bin I_i and the set $\mathbf{\Phi}_{ir}$. We iteratively traverse each bin I_i until all performance scores in $\mathbf{V}$ are exploited for feature selection. Cost is considered jointly with performance to select a cheaper feature when comparable results can be obtained by the features in the set. We calculate the combined importance score S_{ir} of each $\mathbb{f}^r$ within I_i as

$$S_{ir} = \frac{|\mathbf{\Phi}_{ir}|\ ||\mathbf{\Psi}_r||}{\widetilde{\mathbf{\Pi}}_r}, \tag{7}$$

where $||\mathbf{\Psi}_r||$ is the Euclidean norm and $\widetilde{\mathbf{\Pi}}_r$ is the median of the values in $\mathbf{\Pi}_r$. $|\mathbf{\Phi}_{ir}|$ is the cardinality of $\mathbf{\Phi}_{ir}$, which represents the number of persons for which $\mathbb{f}^r$ has the performance scores within I_i. S_{ir} combines the cost with the performance within the bin I_i such that $\mathbb{f}^r$ with the least cost, the highest overall performance and the maximum number of Π_{rm} within the interval I_i gets the maximum importance score S_{ir}. As the bins are sorted in decreasing order of performance, the best performing feature can be selected as

$$r^* = \arg\max_r S_{ir}, \tag{8}$$

where r^* is the ID of the feature with the highest combined importance score S_{ir}.

Let $\mathbf{Y}_{12}$ be the list of selected features for C_1 and C_2. If $\mathbb{f}^{r^*} \notin \mathbf{Y}_{12}$ then $\mathbb{f}^{r^*}$ is appended in $\mathbf{Y}_{12}$. The list $\mathbf{Y}_{12}$ is

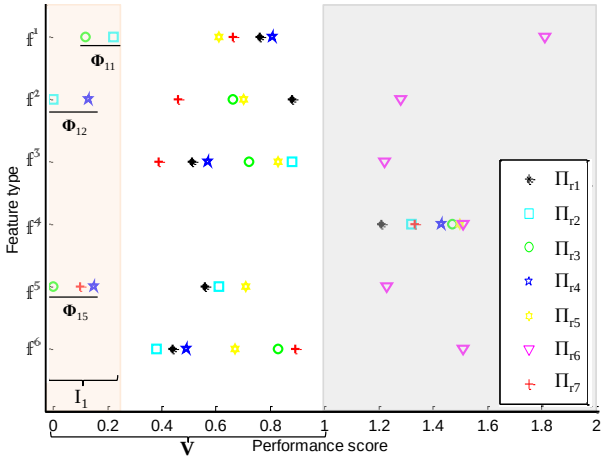


Fig. 2. Example of performance matrix $\mathbf{\Delta}$ containing performance score values Π_{rm} (horizontal axis) obtained for R features (vertical axis) and M persons (color coded), where $r = 1 \cdots 6$, $m = 1 \cdots 7$, $R = 6$ and $M = 7$. $\mathbf{V}$ contains Π_{rm} in the range $[0, 1]$. The bin I_i within $\mathbf{V}$ contains M values of Π_{rm}. The values within I_i are spread among R features such that each feature has a set $\mathbf{\Phi}_{ir}$ containing Π_{rm} and $\sum_{r=1}^{R} |\mathbf{\Phi}_{ir}| = M$. The bin I_1 for $i = 1$ is illustrated in the figure. For each feature $\mathbb{f}^r$ we define a set $\mathbf{\Phi}_{1r}$, e.g. when $r = 5$, $\mathbb{f}^5$ contains $\mathbf{\Phi}_{15} = \{\Pi_{53}, \Pi_{57}, \Pi_{54}\}$, $|\mathbf{\Phi}_{15}| = 3$ and $\sum_{r=1}^{6} |\mathbf{\Phi}_{1r}| = 7$; $\mathbb{f}^4$ is discarded because $\min_m \Pi_{4m} \geq 1$; person $m = 6$ is discarded because $\min_r \Pi_{r6} \geq 1$.

progressively filled with $\mathbb{f}^{r^*}$ in order of importance. However, repeating the selection process of the next best feature may result in performance overlapping. *Performance overlapping* occurs when the performance scores of data points (persons) already considered in the previously selected features are used to measure the importance score S_{ir} of a new feature. To avoid performance overlapping, we remove from $\mathbf{V}$ all the performance scores χ_m of a person P_{mn} for which at least one performance score Π_{rm} exists in the set $\mathbf{\Phi}_{ir^*}$ of the selected feature $\mathbb{f}^{r^*}$. Person P_{mn} is included in the set $\mathbf{Z}^*$ containing the persons that have already taken part in the selection of $\mathbb{f}^{r^*}$, given as $\mathbf{Z}^* \cup P_{mn}\ \forall\ \Pi_{rm} \in \mathbf{\Phi}_{ir^*}$. We then repeat the process for selecting the next best feature. Each selected feature is now representative of a unique subset of data and a selected new feature now increases the diversity in the feature set by covering a wider range of data.

Feature selection continues within the same bin I_i until all performance scores have been utilized for the selection. Then we move to the next bin in $\mathbf{V}$. The algorithm stops when all persons in the training data are exhausted ($|\mathbf{Z}^*| = M$) or when all features are selected ($\langle \mathbf{Y}_{12} \rangle = \hat{R}$, where $\langle \cdot \rangle$ counts the elements in the list). In the former case we obtain a subset of features. In the latter case the method returns the complete feature set with features ranked in order of importance.

Note that because the selected features in $\mathbf{Y}_{12}$ are ranked by decreasing importance, the feature set can be further reduced by dropping the IDs of the least important features should the constraints of the application become more restrictive. CoPE is summarized in Algorithm 1.

D. Discussion

Feature selection using CoPE is performed once using training data when a camera network is set-up. Then each

Algorithm 1 CoPE feature selection

M : total number of persons;
$\hat{R}$: number of features;
C_n : n^{th} camera in the network;
$\mathbb{f}^r$: r^{th} feature in the feature set;
P_{mn} : m^{th} person in C_n;
$\mathbf{\Psi}_r$: cost vector of $\mathbb{f}^r$;
Π_{rm} : performance score value for P_{mn} using $\mathbb{f}^r$;
$\mathbf{\Pi}_r$: performance vector for $\mathbb{f}^r$;
$\mathbf{\Delta}$: performance matrix;
χ_m : $\hat{R}$ performance scores for P_{mn};
$\mathbf{V}$: vector containing sorted values of $\Pi_{rm} \leq 1$ from $\mathbf{\Delta}$;
I_i : i^{th} bin with values from $\mathbf{V}$;
S_{ir} : combined importance score of $\mathbb{f}^r$ in I_i;
$\mathbf{Y}_{nq}$: list of selected features for C_n and C_q;
$\mathbf{Z}^*$: set of people taking part in the selection;
$\mathbf{\Phi}_{ir}$: set of Π_{rm} in $\mathbf{V}$ within I_i for $\mathbb{f}^r$;
$\langle \cdot \rangle$: number of elements in the list;
$| \cdot |$: cardinality of a set;

$\mathbf{Z}^* = \phi$, $\mathbf{Y}_{nq} = \phi$
while $|\mathbf{Z}^*| \leq M$ or $\langle \mathbf{Y}_{nq} \rangle \leq \hat{R}$ **do**
  while $1 \leq i \leq \hat{R}$ **do**
    for $r = 1$ to $\hat{R}$ **do**
      $\mathbf{\Phi}_{ir} = \Pi_{rm}$ in $\mathbf{V}$ within I_i for $\mathbb{f}^r$
    end for
    for $r = 1$ to $\hat{R}$ **do**
      calculate S_{ir} using Eq. 7
    end for
    get r^* using Eq. 8 ▷ ID of selected feature
    if $\mathbb{f}^{r^*} \notin \mathbf{Y}_{nq}$ **then**
      append $\mathbb{f}^{r^*}$ to $\mathbf{Y}_{nq}$
    end if
    remove χ_m from $\mathbf{V}$; $\forall\ \Pi_{rm} \in \mathbf{\Phi}_{ir^*}$
    $\mathbf{Z}^* = \mathbf{Z}^* \cup P_{mn}$; $\forall\ \Pi_{rm} \in \mathbf{\Phi}_{ir^*}$
    remove $\mathbf{\Pi}_{r^*}$ from $\mathbf{V}$
  end while
end while

camera locally stores the list of selected features $\mathbf{Y}_{nq}$ for each neighboring camera C_q. C_n and C_q are neighbors if an object exiting the FoV of a camera is expected to enter the FoV of the second camera without passing through the FoV of a third camera. If a new camera is added to the network, the training is performed pair-wise between the new camera and its neighbors only [45]. Note that features selected for a camera pair may not always be appropriate for another camera pair because of differences in illumination conditions and camera pose with respect to the targets. This approach, developed for camera pairs, is appropriate for distributed multi-camera settings where each camera communicates with its neighbors without a central control unit.

A performance matrix $\mathbf{\Delta}$ is generated for each camera pair, while the cost vector $\mathbf{\Psi}_r$ already takes into account N cameras and therefore remains the same. N cameras in the extreme case form a complete graph, where each camera has $N-1$ neighbors. The time complexity of feature selection for such a network is $N(N+1)/2$ times that of the feature selection for a camera pair. In the case of multiple neighboring cameras, it is also possible that the locally stored lists of selected features may together result in the extraction of the complete feature set. Such a scenario may occur in the case of neighbors located far away and reduce the benefits of feature selection.

When machine learning is used for re-identification, two training phases are involved, namely the training for feature selection and the training for learning the re-identification model. Each camera stores the trained models (weights) in addition to the selected feature IDs between (neighboring) camera pairs. The inclusion of learning models with the CoPE feature selection is independent of the feature selection itself, and there is an increase in the storage cost (fixed) due to the local storing of the trained models (and not because of the feature selection).

IV. Evaluation and comparisons

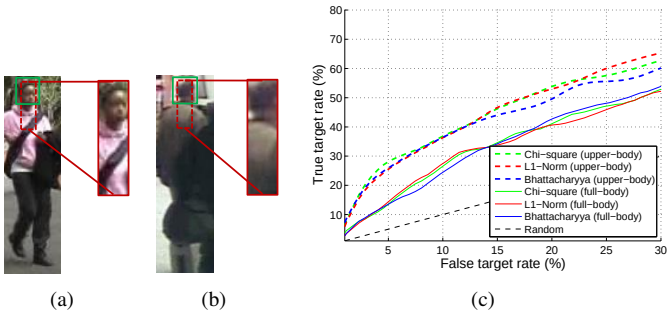
We evaluate CoPE on the initial feature set as in [14]–[17]. We test the re-identification capabilities with existing approaches based on DDM, namely Bhattacharyya distance [10], [16], L1-Norm [9] and Chi-square distance. CoPE with DDM is further compared with recent re-identification approaches such as Probabilistic Relative Distance Comparison (PRDC) [15] and Attribute-Sensitive Feature Importance (ASFI) [17]. We further compare CoPE with the following feature selection methods: Fisher score [19], Information gain [20], mRMR [21], ReliefF [22] and Bi-clusters [23] for re-identification using DDM (Bhattacharyya) and learning approaches (rankSVM [14] and AdaBoost [16]). Finally, we perform a cost-performance analysis for CoPE.

A. Evaluation criteria

We consider three validation criteria, namely cost of features, re-identification rate and feature budgeting. The *cost of features* is calculated for the initial feature set and then for the selected features to analyze improvements in data reduction and computational time. In addition, we evaluate the training time for feature selection. The *re-identification rate* for the association methods is compared with the initial feature set and then with the selected features using the Cumulative Matching Characteristics (CMC) curves [16]. The curves show the true target rates for given false target rates. The overall performance is also evaluated using the Area Under the CMC Curves (AUC). Finally, we consider *feature budgeting* in constrained environments to analyze the scalability of CoPE and the effects in terms of cost and performance of further feature reductions.

B. Experimental setup

The datasets for the evaluation are VIPeR [16] and i-LIDS [46]. These datasets present a mix of characteristics such as outdoor and indoor settings, variations in viewing angle, occlusions and illumination changes. The VIPeR dataset (outdoors) contains 632 pairs of persons from arbitrary viewpoints. The images of people in VIPeR are not occluded, but present significant appearance changes. In the i-LIDS dataset, C_1 is Camera 1 and C_2 is Camera 3. These two cameras represent non-overlapping views with considerable illumination changes. We represent each person with a single image per camera. In total, we have 348 people that go from C_1 to C_2. Single images of 124 people are manually extracted while exiting C_1 when the person is completely in the FoV of C_1 and the first image of a person on reappearance when (s)he is completely in C_2. People can be partially visible due



(a) (b) (c)

Fig. 3. Examples of upper-body stripes (cropped region) extracted based on the location of the head (green bounding box) Eq. 9 in (a) VIPeR and (b) i-LIDS. (c) Comparison of re-identification rates using the initial feature set from the full-body patch (solid lines) and the upper-body stripe (dashed lines), without feature selection, using i-LIDS ($M = 174$). For association three DDM approaches are compared.

to occlusions. For the remaining 224 people, we utilize the existing ([9], [15], [40])images of persons extracted from i-LIDS videos and select one image per person per camera, such that no person is repeated.

The head and the upper-body are the most frequently visible and recognizable parts of a person (Fig. 3 (a, b)) in case of surveillance settings and multiple people in the scene. We assume that the person detection phase is solved using a head detector [47], [48] resulting in a bounding box $B_{mn} = (x, y, w, h)$, for the head of person P_{mn} in C_n, where x and y are the x-y coordinates of the top left corner, w is the width and h is the height of the bounding box. From a given bounding box B_{mn} (in this paper manually defined), a vertical stripe of the upper-body is generated using the approach described in [9]:

$$P_{mn} = f(B_{mn}) = [x + w/4, y + h/4, w/2, h * 2]. \quad (9)$$

The area of the defined upper-body shape is 8 times smaller than that of normalized full-body patch.

We use the commonly employed color and texture features [14]–[17] for the comparison. Existing approaches use a 2784-dimensional feature vector by dividing the full-body person image into a set of stripes (6) and then concatenating the corresponding features from each stripe. We reduce the size of the object representation by extracting features from the defined upper-body shape only (better suited for crowded scenes). Each feature is a 12-bin histogram of a color channel or a filtered image of the defined shape. Nine color channels (R, G, B, Y, Cb, Cr, H, S, V) from RGB, YCbCr and HSV color spaces are used. We divide the upper-body patch into two halves and extract the features. Instead of concatenating the extracted features, we add the corresponding bins. This operation results in a single feature set with twice the weight to the upper region of the defined shape. For texture, 8 Gabor and 13 Schmid filters are applied on the Y channel of the defined shape. The parameters of the Gabor and Schmid features are indicated in Fig. 4. Figure 3 (c) shows the re-identification results in i-LIDS for the complete feature set extracted from the defined upper-body and the full-body patches without feature selection. It can be observed that a better re-identification rate can be achieved with the upper-body shape in the case of occlusions and crowd.

The data generated for each feature is encoded using the lossless data compression algorithm 'deflate', which combines LZ77 and Huffman coding. The average storage size and computational cost of a single feature (per person) is calculated for each camera. For each dataset, we apply the two-fold cross validation. The experiments are carried out using Matlab 7.11 on a 3.3 GHz dual core desktop system with 3 GB of RAM.

C. Analysis and discussion

Figure 4 shows the storage size β^r and the extraction time Γ^r of the 30 features used. A fixed number of bins is used for histograms. However β^r varies between 29 and 56 bytes because the data encoding is applied before the feature storage. Γ^r of the feature extraction varies between 16 and 60 ms. The overall computational time and the storage size required by a single camera C_n for $\hat{R}$ features and M persons is given by $\mathbf{T}_n$ as

$$\mathbf{T}_n = \left(\sum_{m=1}^{M} \sum_{r=1}^{\hat{R}} \Gamma^r_{mn}, \sum_{m=1}^{M} \sum_{r=1}^{\hat{R}} \beta^r_{mn} \right). \quad (10)$$

For the comparisons, we measure the normalized cost $\mathbb{E}$ of the selected feature sets from the cost score $\mathbf{\Psi}_r$ (Eq. 6) as

$$\mathbb{E} = \frac{\sum_{\mathbb{F}^{r^*} \in \mathbf{Y}_{nq}} 1/||\mathbf{\Psi}_{r^*}||}{\sum_{r=1}^{\hat{R}} 1/||\mathbf{\Psi}_r||}, \quad (11)$$

where $\mathbf{Y}_{nq}$ is the list of selected features. The set of 30 features has the maximum cost $\mathbb{E}_{max} = 1$. We consider 316 and 174 persons in VIPeR and i-LIDS, respectively.

Figure 5 shows the analysis of CoPE on i-LIDS with three selection criteria: (i) feature selection as a function of performance only keeping the cost fixed (CoPE-FC); (ii) feature selection considering both cost and performance of a feature (CoPE); and (iii) feature selection as a function of cost only while keeping the performance fixed (CoPE-FP). The cost of the selected features is the highest for CoPE-FC, since the selection is carried out based on performance only. However, in the absence of cost constraints CoPE-FC is also able to achieve the highest re-identification rate using DDM. Both the cost and the re-identification rate of the selected features are reduced with CoPE. When varying the cost parameter α in CoPE, while the composition of the selected features remains similar their order changes (top three rows of the table in Fig. 5 (a)). Since cost and performance are independent in the feature selection, varying α does not affect the performance of a feature. The selected features may vary based on the requirement of a system controlled by α, i.e. for good-performing features with a limited extraction time $\alpha = 1$, and for features with limited storage size and good performance $\alpha = 0$. Note that a limited extraction time may not imply a higher storage size (and vice versa). The smallest cost for CoPE is obtained when there is an equal contribution of computational time and storage size ($\alpha = 0.5$).

In CoPE-FP, although performance is not used for feature selection, in order not to obtain a sorted list of all features

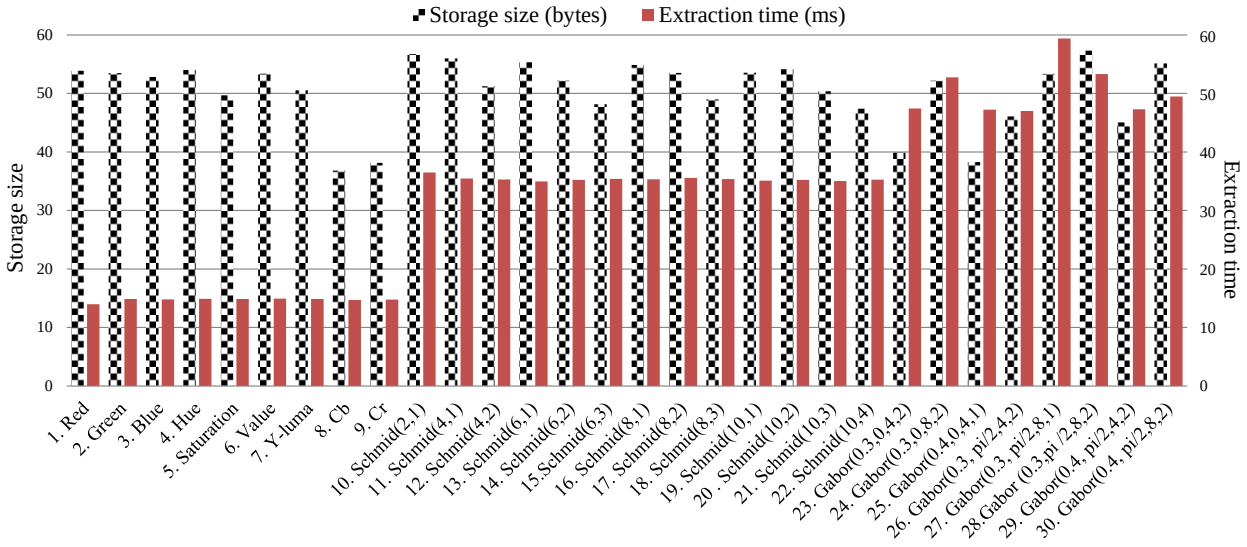
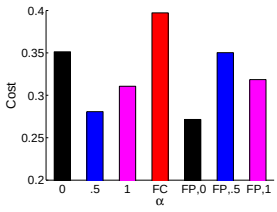


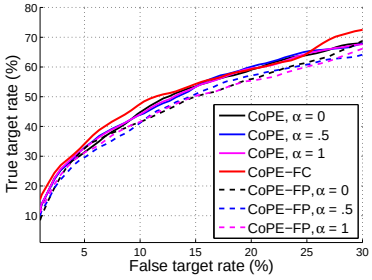
Fig. 4. Storage size and the extraction time for each feature in the initial feature set (listed on the horizontal axis with their IDs and parameters). For Gabor $(\gamma, \theta, \lambda, \sigma^2)$ and Schmid (σ, τ) filters, γ is the aspect ratio, θ is the angle in radians, λ the wavelength of the sinusoidal factor, σ the standard deviation and τ the number of cycles. The vertical axes represent the storage size (bytes) and the feature extraction time (ms) for a single person within a camera FoV.

	α	Feature IDs as in Fig. 4
CoPE	0	7 25 9 10 8 19 3 30 4 29 21
	0.5	7 3 9 8 25 11 4 19 21 29
	1	7 3 9 8 11 25 4 19 21 29 16
CoPE-FC	-	7 10 25 9 19 3 8 30 4 27 21 29 5
CoPE-FP	0	8 9 25 29 22 15 7 3 19
	0.5	1 8 9 5 3 7 4 22 15 21 19 25 29
	1	1 8 9 3 2 7 5 4 13 21 19 25 29

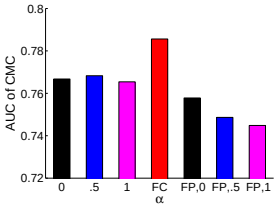
(a) Feature IDs i-LIDS



(b) Cost $\mathbb{E}$ i-LIDS



(c) CMC i-LIDS



(d) AUC i-LIDS

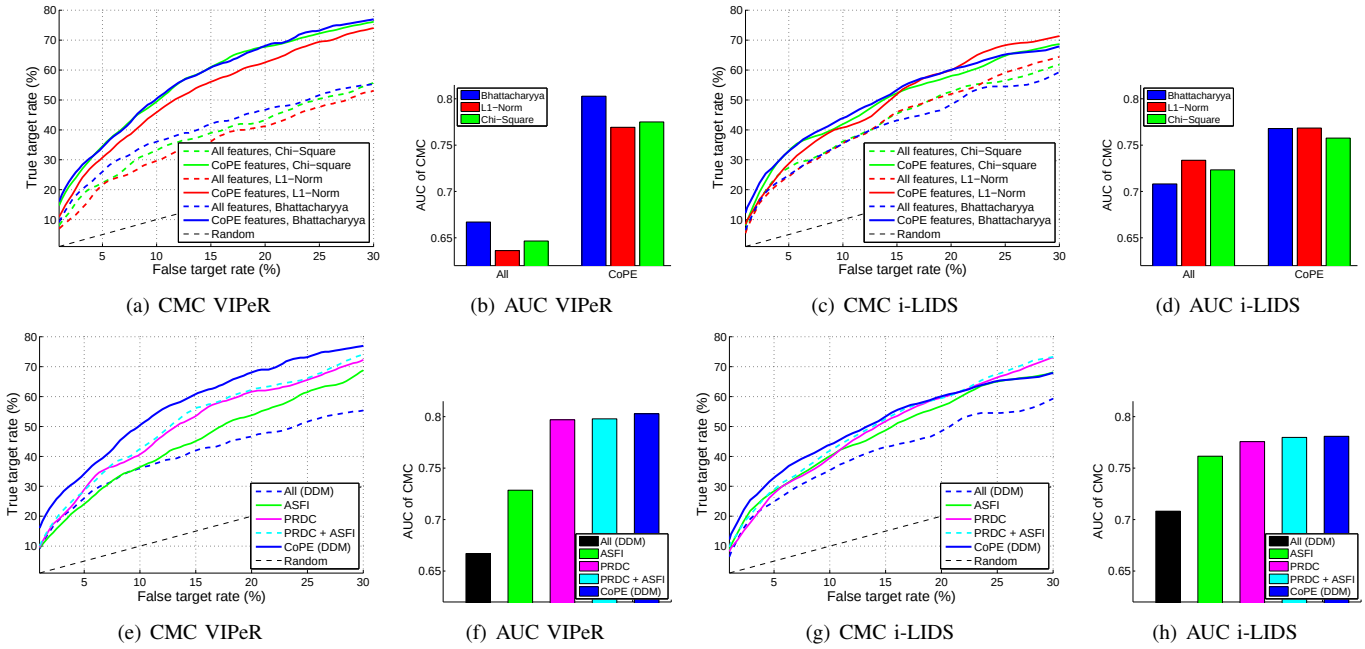
Fig. 5. Analysis of feature selection with varying selection criteria: CoPE (both cost and performance), CoPE-FC (fixed cost, performance only) and CoPE-FP (fixed performance, cost only), and with varying α values: $\alpha = 0$ (black), $\alpha = 0.5$ (blue), $\alpha = 1$ (magenta) in i-LIDS ($M = 174$) in terms of (a) selected features, (b) normalized cost $\mathbb{E}$, (c) CMC curves, and (d) AUC of CMC curves.

based on cost, we remove the people from the training data for which the selected minimum cost feature has good performance so that the algorithm stops when all the people in the training data are exhausted. For $\alpha = 0$ the order of selection is controlled by the storage size, while for $\alpha = 1$ the features with the shortest extraction times (color features) are selected first (see Fig. 4 for time and size). An interesting case is when the features with IDs 8 and 9 are selected for all three values of α, since these features have both the shortest computational time and the smallest storage size. In contrast, the feature with $ID = 1$ has the shortest extraction time and a large storage size. This makes it the first feature with $\alpha = 0.5$ and 1, while it is not selected with $\alpha = 0$. The order of performance of the selected features for the three criteria is as follows $CoPE\text{-}FP < CoPE < CoPE\text{-}FC$. The rest of the evaluation is performed for $\alpha = 0.5$, in order to have an equal contribution from the storage size and the extraction time.

1) CoPE vs all-features: Table II shows the storage size and the computational time for the features extracted from each person observed in one camera using Eq. 10. We compare the results of the initial feature set with that of the three non-unique sets of selected features obtained by CoPE using three similarity measures: Bhattacharyya distance, L1-Norm and Chi-square distance in Eq. 1. For VIPeR, the number of selected features are 6, 6 and 8, respectively, for the three similarity measures that reduce the storage size per camera to 11%, 18% and 24% of the total size (466.43 KB) of the initial 30 features. In i-LIDS, 10, 8 and 9 features are selected for the three similarity measures that respectively reduce the storage size to 31%, 23% and 27% of the storage requirement for the initial feature set (256.83 KB). Similarly, the computational

TABLE II
STORAGE SIZE, COMPUTATIONAL TIME AND NORMALIZED COST OF THE INITIAL FEATURE SET PER CAMERA USED IN EXISTING RE-IDENTIFICATION APPROACHES COMPARED WITH COPE FEATURES OBTAINED FOR THREE SIMILARITY MEASURES (IN EQ. 1).

Dataset	Distance as in Eq. 1	Total features	Feature IDs as in Fig. 4	Size (KB)	Time (sec)	Cost $\mathbb{E}$ (Eq. 11)
VIPeR (M=316)	-	30	1-30	466.43	314.83	1.00
	Bhattacharyya	6	4 8 9 5 7 25	82.46	38.34	0.16
	L1-Norm	6	8 9 4 7 5 12	86.46	34.56	0.16
	Chi-Square	8	8 9 4 5 7 18 17 25	114.07	60.75	0.23
i-LIDS (M=174)	-	30	1-30	256.83	173.35	1.00
	Bhattacharyya	10	7 3 9 8 25 11 4 19 21 29	80.74	47.70	0.29
	L1-Norm	8	3 9 7 25 8 4 18 29	61.90	35.48	0.24
	Chi-Square	9	7 9 3 8 25 6 15 29 16	70.97	41.64	0.26



(a) CMC VIPeR (b) AUC VIPeR (c) CMC i-LIDS (d) AUC i-LIDS

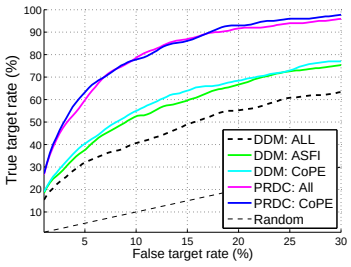
(e) CMC VIPeR (f) AUC VIPeR (g) CMC i-LIDS (h) AUC i-LIDS

Fig. 6. Person re-identification comparison using CMC curves representing true target rate for the top 30% of false target rate, and the AUC of the CMC curves in VIPeR ($M = 316$) and i-LIDS ($M = 174$) datasets. (a-d) CoPE features selected using three similarity measures, namely Bhattacharyya distance (blue), L1-Norm (red) and Chi-square (green) in Eq. 1, compared with the initial complete feature set using DDM approaches for re-identification. (e-h) Existing re-identification approaches: PRDC [15] (magenta) and ASFI [17] (green) using the complete feature set are compared with DDM (Bhattacharyya) [16] using CoPE features (blue) and the complete feature set (black).

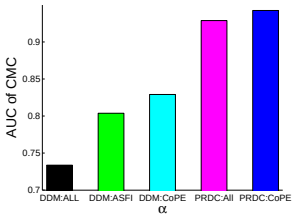
time of feature extraction per camera is reduced significantly. In the VIPeR dataset, the computational time is reduced to 12%, 10% and 19% for the three similarity measures, respectively. In the case of i-LIDS, the computational time is reduced to 27%, 20% and 23%. It can also be observed that the normalized cost $\mathbb{E}$ of the selected CoPE features is reduced more in VIPeR than in i-LIDS because mostly the color features are selected in VIPeR. The color features are fast to extract with less or comparable storage size (Fig. 4) and perform better than texture features. In VIPeR, we reduce the cost $\mathbb{E}$ of the feature set to 20%, while in i-LIDS we reduce it to 33% of the initial feature set.

Figure 6 (a-d) compares the re-identification rate for the three DDM approaches with the state of the art. In a DDM approach, two persons are considered correctly matched for re-identification, if their obtained feature sets have the minimum matching distance between them. The performance of the selected features is measured in terms of improvement of the re-identification rate of DDM approaches compared to that of using the initial feature set. CMC curves highlight the true target rate for the first 30% of false target rates (the most important part of CMC for evaluation). In VIPeR, a higher re-identification rate is obtained using the selected features. For example, at 20% false target rate in the CMC curves, the true target rate is above 65% for selected features compared to the initial feature set with true target rates between 40% to 50% for all the three measures. Due to the limited illumination changes between cameras, mostly color features are selected (Table II). In i-LIDS, both color and texture features are selected. The re-identification results for association using the selected features are improved and in some points are comparable to that of using all features. The AUC shows that the features selected using all the three similarity measures have overall better performance than that of the initial feature set. The highest re-identification rate is obtained when the features are selected using the Bhattacharyya distance. Therefore, in the following experiments we use the Bhattacharyya distance as a similarity measure while comparing with existing re-identification and feature selection approaches.

Figure 6 (e-h) shows the performance comparison of DDM (Bhattacharyya) using CoPE with two recent state-of-the-art re-identification approaches: PRDC [15] and ASFI [17]. The extracted features from the upper-body patch are given as input to PRDC and ASFI. In both i-LIDS and VIPeR, a better or comparable re-identification performance is achieved by CoPE with less storage and computational requirements. CMC curves show a higher re-identification rate for CoPE especially at lower false target rates. CoPE outperforms PRDC and ASFI (AUC in the case of VIPeR), with a cost of 20% of the initial feature set used in these methods. In i-LIDS comparable results can be observed at 33% the cost. The results of PRDC and ASFI in the original papers used the full-body patches and a large 2784-dimensional feature set. Therefore we also include a comparison with the reported results while performing the CoPE feature selection on the larger feature set and full-body patches. Figure 7 shows the cost-performance comparison on ViPER, which has less occlusions and thus justifies the use of the full patch for person description. It can be observed from the CMC curves and the AUC that CoPE selected features with DDM show a better re-identification rate than ASFI with a 73% reduction in the storage size and a 77% reduction in the extraction time. Finally, the use of



(a) CMC VIPeR



(b) AUC VIPeR

Features	Size (KB)	Time (sec)	Cost $\mathbb{E}$
All	8.54	5.89	1
CoPE	2.34	1.35	0.24

(c) Cost $\mathbb{E}$

Fig. 7. Person re-identification comparison using full-body patches and 2784-dimensional feature vector on VIPeR ($M = 316$). (a) CMC curves and (b) AUC of CMC curves obtained by existing re-identification approaches: PRDC [15] (magenta) and ASFI [17] (green). The results are compared with CoPE using DDM (cyan) and PRDC (blue) for association. (c) The storage size and the computational time of the extracted features.

the CoPE features as input to PRDC further improves the re-identification rate of PRDC at 24% the cost of the complete feature set.

2) CoPE vs feature selection methods: We compare CoPE with 5 existing feature selection and ranking methods, namely Fisher score [19], Information gain [20], mRMR [21], ReliefF [22] and Bi-clusters [23]. Since these are single-objective feature selection approaches, for comparision we perform feature selection using the performance only while keeping the cost fixed (CoPE-FC). The similarities between the feature pairs obtained using Eq. 1 along with the assigned labels as correct/incorrect matches are given as input to the feature selection methods. Feature selection methods return a ranked list of features and a weight vector in the case of Fisher score, Information gain and ReliefF methods, while mRMR and Bi-clusters return only a ranked feature list.

Table III shows the training time for feature selection and the obtained features ranked in order of importance for re-identification. Training time is useful to understand the feasibility of the single time set-up off-line process and becomes crucial as the size of the network increases. The training time is measured using 316 and 174 people in the VIPeR and i-LIDS, respectively. With VIPeR, CoPE takes 0.76 seconds, 5 times less than the next shortest training time by the Fisher score. ReliefF requires the maximum time (124.70 seconds) for training, while Bi-cluster could not be trained for VIPeR even after 25 days. With i-LIDS, the training time of Bi-clusters is nearly 20 hours. Therefore, in a larger camera network Bi-clusters may not be applicable for feature selection. CoPE and CoPE-FC take 0.30 and 0.20 seconds, respectively. The Fisher score takes 0.15 seconds. As the dataset size almost doubles from i-LIDS to VIPeR, the time requirement for Fisher Score is increased by nearly 24 times, whereas others are only 3 times longer. With the smallest ratio and minimum training time CoPE is desirable for feature selection in a camera network.

In Table III, each selection approach returns a different ranking order of features, since there exists no unique feature subset to solve the same task. If two features show an identical performance then any of the two can be selected. In performing a cost-aware feature selection, CoPE returns a subset of well-performing cost-effective features until any further addition in the cost of features does not improve performance. In VIPeR, most feature selection methods including CoPE and CoPE-FC return similar sets with color features in the top ranks. CoPE and CoPE-FC returns the same set of 6 features because of the similarity in the selection procedure. In i-LIDS, 10 features are selected by CoPE, while 13 features are selected by CoPE-FC. We fix the number of selected features for the existing methods to be equal to the number of features selected by CoPE-FC (a comparison with varying number of selected features can be seen in Fig. 11). We pick the top 13 features in i-LIDS and the top 6 features in VIPeR from the ranked features of the existing approaches. Figure 8 (a, d) shows the normalized cost $\mathbb{E}$ (Eq. 11) of the obtained selected features. Even after fixing the number of selected features, $\mathbb{E}$ for CoPE features remains the smallest. mRMR features show the highest cost in both datasets, while those of Fisher score, Information gain and Bi-clusters have costs comparable with that of CoPE-FC. In VIPeR, the CoPE feature set contains all color features because of the limited illumination changes, while in i-LIDS both color and texture features are selected. CoPE selects the color features first and then the texture, resulting in the least $\mathbb{E}$ of 0.15 and 0.30 in VIPeR and i-LIDS, respectively.

Figure 8 (b, c, e, f) shows the re-identification performance of the selected features using DDM (Bhattacharyya) as the association method. In both VIPeR and i-LIDS, the selected features using CoPE and CoPE-FC reach the highest re-identification rate. In i-LIDS, CoPE-FC reaches the highest performance in the absence of the cost constraints. Unlike the existing approaches based on overall performance only, CoPE selects features by iteratively relaxing the performance score S_{ir}, thus achieving cost as well as performance advantages.

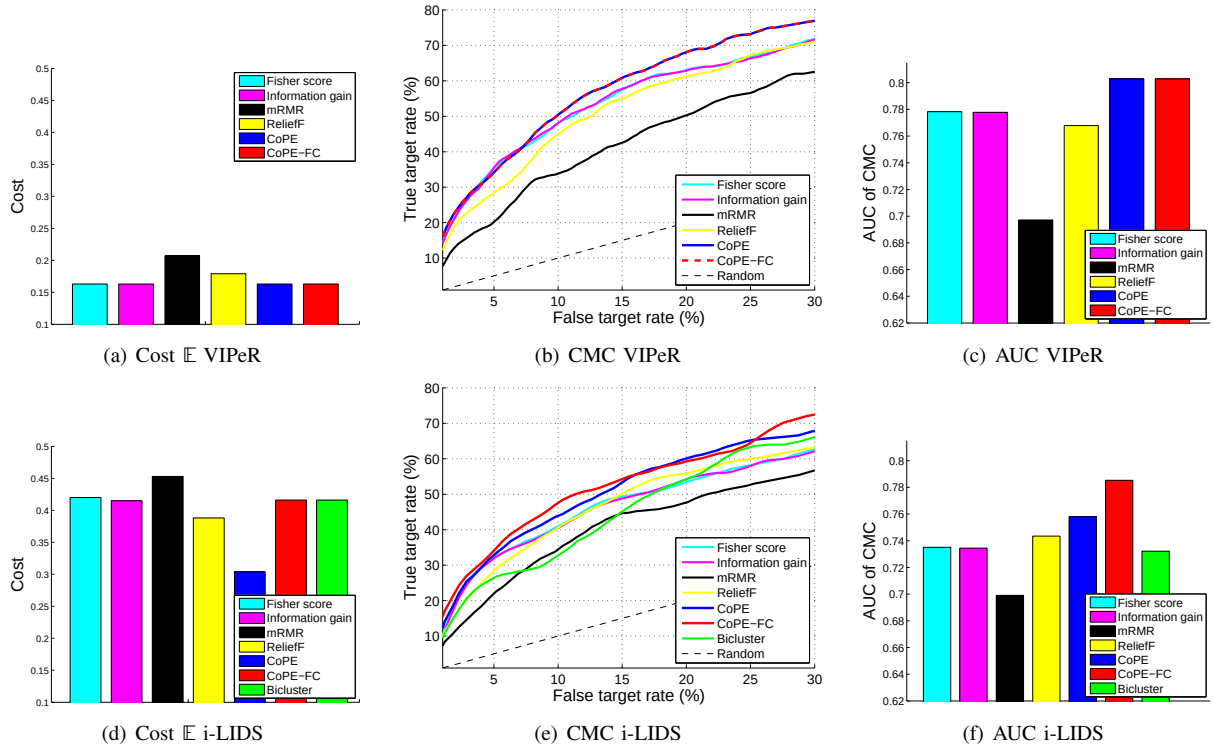
Finally, features are selected on one dataset and tested on the other to analyze the amount of degradation in the results. We compare CoPE with two feature selection approaches, namely Fisher score and Information gain, which have the highest performance in the cross validation within the same dataset. Fig. 9 shows the cross-data robustness of selected features. In VIPeR, the performance of CoPE is degraded less compared to the other two methods. The performance of features selected using VIPeR deteriorates at a greater rate in i-LIDS, which is a more challenging dataset. The results are degraded at a comparable rate for all the feature selection approaches, since almost the same 6 color features are selected by the three feature selection approaches.

3) CoPE with learning models: The top-ranked selected features are used as input to the two learning methods, namely rankSVM [14] and AdaBoost [16] for re-identification, which apply implicit feature selection by weighting the feature set. In these cases, feature selection may be used to remove poorly performing features as a pre-processing step to improve the effectiveness of learning methods. Since the features are rearranged and weighted within the specific learning method, the order of selection is not important and only the difference in the selected features affects the performance. We compare

TABLE III

THE TRAINING TIMES AND THE RANKING ORDER OF FEATURES FOR RE-IDENTIFICATION USING FISHER SCORE [19], INFORMATION GAIN [20], MRMR [21], RELIEFF [22] AND BI-CLUSTERS [23] AS FEATURE SELECTION METHODS COMPARED WITH COPE AND COPE-FC USING VIPER ($M = 316$) AND I-LIDS ($M = 316$).

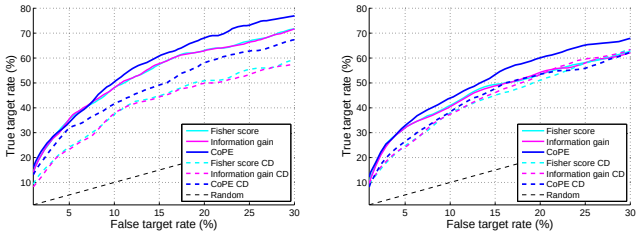
	Training time (sec)		Ranking order (Feature IDs as in Fig. 4)		
Feature selection	**VIPeR**	**iLids**	**Ratio**	**VIPeR**	**iLids**
Fisher score	3.67	0.15	24.47	4 8 9 5 1 2 6 7 3 17 14 13 11 10 23 30 16 12 18 21 26 27 24 28 22 29 19 25 15 20	3 4 7 6 2 10 11 9 1 13 19 30 16 20 17 26 14 27 12 8 28 23 21 24 18 25 15 29 22 5
Information gain	14.10	4.99	2.82	4 8 9 5 6 2 1 3 7 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30	1 7 3 6 11 10 20 13 9 16 4 17 2 14 19 12 30 26 27 15 18 28 21 23 8 25 24 29 22 5
mRMR	29.06	9.54	3.04	4 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 9 8 5 7 6 3 2 1	1 30 26 23 7 6 2 24 28 3 27 29 21 17 14 22 15 18 20 12 10 19 16 11 13 25 8 4 9 5
ReliefF	124.70	44.98	2.77	4 8 9 5 27 25 24 2 13 29 6 1 28 30 10 7 3 26 23 22 11 15 16 19 12 21 20 18 17 14	8 4 9 25 13 22 23 16 7 26 1 5 18 2 29 15 21 24 19 28 30 6 27 14 12 20 3 11 17 10
Bi-clusters	-	72000	-	-	9 8 24 25 12 14 5 11 4 15 10 21 17 16 6 1 28 13 26 23 7 3 20 18 2 19 30 27 29 22
CoPE	0.76	0.30	2.53	4 8 9 5 7 25	7 3 9 8 25 11 4 19 21 29
CoPE-FC	0.52	0.20	2.60	4 8 9 5 7 25	7 10 25 9 19 3 8 30 4 27 21 29 5



(a) Cost $\mathbb{E}$ VIPeR (b) CMC VIPeR (c) AUC VIPeR

(d) Cost $\mathbb{E}$ i-LIDS (e) CMC i-LIDS (f) AUC i-LIDS

Fig. 8. (a, d) Normalized cost, (b, e) CMC curves and (c, f) AUC of CMC curves obtained for re-identification by applying DDM to the features selected using the Fisher score [19] (cyan), Information gain [20] (magenta), mRMR [21] (black), ReliefF [22] (yellow) and Bi-clusters [23] (green), CoPE (blue) and CoPE-FC (red) on (Top row) VIPeR ($M = 316$) and (Bottom row) i-LIDS ($M = 172$).



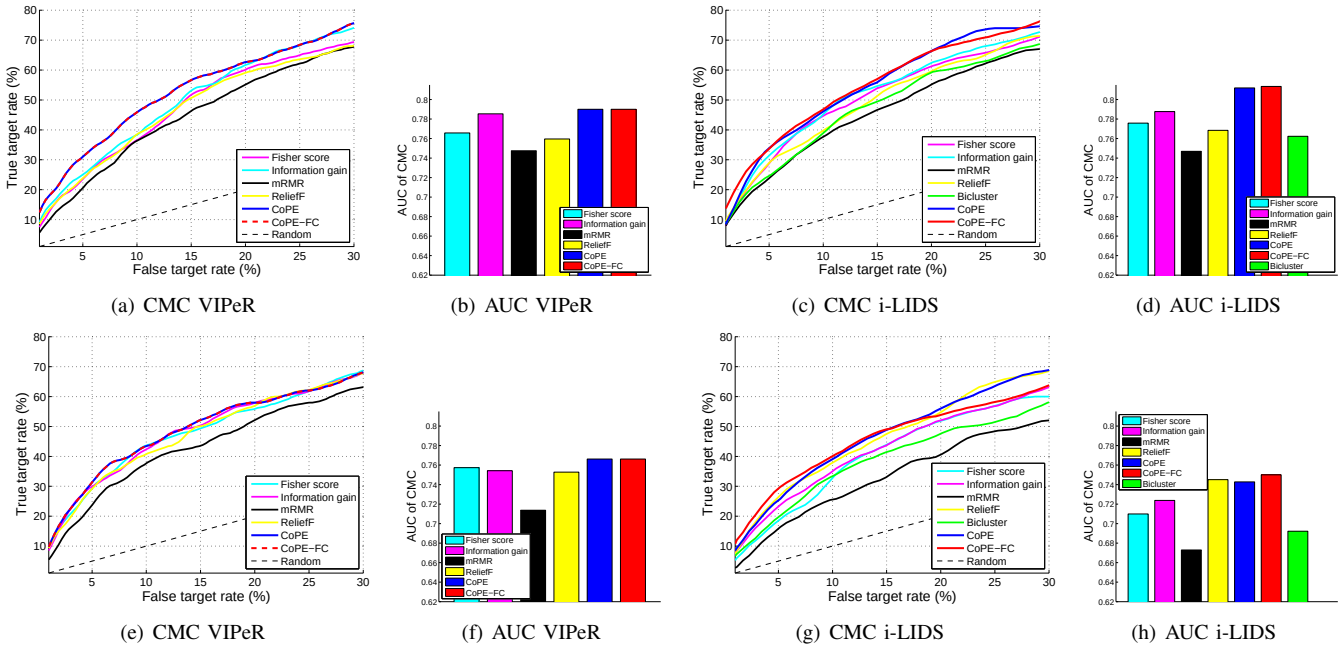
(a) CMC VIPeR (b) CMC i-LIDS

Fig. 9. Cross Data (CD) performance comparison in re-identification for the top-two performing existing feature selection approaches, namely Fisher score [19] (cyan) and Information gain [20] (magenta); and CoPE (blue). The CMC curves are obtained by (a) feature selection on i-LIDS ($M = 174$) and testing on VIPeR ($M = 316$) and by (b) feature selection on VIPeR ($M = 316$) and testing on i-LIDS ($M = 174$).

the performance of RankSVM and AdaBoost with their default settings. The comparisons are performed with and without feature selection keeping the same settings, which may not be optimal. However, the improvement in the results can be observed after the feature selection by the proposed approach.

RankSVM assigns relative weights to the input features based on the combined contributions in the feature set. Figure 10 (a-d) shows that rankSVM has a better re-identification rate for both VIPeR and i-LIDS using the features selected by CoPE compared to those from existing feature selection methods. The variation in re-identification rates using the selected features from different approaches is smaller in VIPeR than in i-LIDS because mostly the same color features are selected (Table III). With i-LIDS, the features selected by different methods (and the re-identification rate) vary in their composition. The best performance of CoPE-FC in the true target rate (CMC curves) is almost 15% higher than that of mRMR at the same false target rate, followed by CoPE with a slightly smaller re-identification rate because of the additional cost constraints. However CoPE remains higher than existing feature selection approaches. Also the obtained AUCs are highest for CoPE and CoPE-FC.

AdaBoost combines multiple weak classifiers/features to improve the matching performance. Figure 10 (e-h) shows the performance for AdaBoost. In both VIPeR and i-LIDS, the



(a) CMC VIPeR (b) AUC VIPeR (c) CMC i-LIDS (d) AUC i-LIDS

(e) CMC VIPeR (f) AUC VIPeR (g) CMC i-LIDS (h) AUC i-LIDS

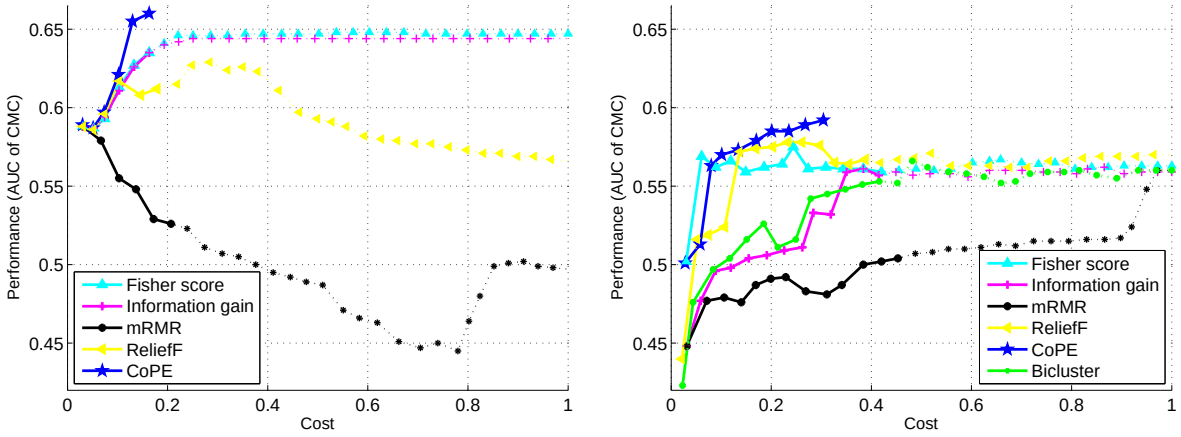
Fig. 10. CMC curves and AUC of CMC curves for re-identification using learning methods: (a-d) rankSVM [14] and (e-h) AdaBoost [16] applied to the features selected by CoPE (blue) and CoPE-FC (red) and existing methods: Fisher score [19] (cyan), Information gain [20] (magenta), mRMR [21] (black), ReliefF [22] (yellow) and Bi-clusters [23] (green), using (a, b, e, f) VIPeR ($M = 316$); and (c, d, g, h) i-LIDS ($M = 174$).

features selected by CoPE have an overall better or comparable re-identification rate than existing feature selection methods. In VIPeR, similarly to the rankSVM, CMC curves show a smaller re-identification rate variation among existing methods because of the limited number of selected features (i.e. only 6). In i-LIDS, the variation in performance between CoPE and existing feature selection methods is high as the number of selected features is increased (up to 13). AdaBoost has a better learning ability in i-LIDS than in VIPeR. The performance on the CMC curves, especially in the starting part, shows that CoPE and CoPE-FC are able to remove noisy features more effectively than existing feature selection methods thus resulting in a better re-identification rate. In Fig. 10 (g), the CMC curve for ReliefF shows a marginal improvement of up to 2% in true target rate between 20% and 25% of false target rates at the expense of more costly features than that of CoPE (Table: III). In CoPE because of the cost constraints, we may observe a drop in the performance in a few instances in favor of cost reduction and an overall performance improvement. Overall CoPE-FC remains the highest (AUC) followed by CoPE and ReliefF features.

Since learning algorithms are dependent on the training data in addition to the selected features, in challenging scenarios the performance of learning methods can be reduced. In re-identification, a single person may exhibit several pose and illumination changes, while we can only extract a few patches thus resulting in an under sampled data representation [15]. For example, in VIPeR (CMC curves in Fig. 6 (a) in comparison with Fig. 10 (a, e)), the performance of learning methods is slightly reduced. In Fig. 6 (a) we can see that after CoPE feature selection the performance is improved (almost double compared to using the initial feature set). A further improvement through a learning method will require a more robust training set.

4) CoPE and feature budgeting: In a constrained environment, a further reduction of the feature set might be necessary. In such cases the performance needs to be reduced in a predictable manner (feature budgeting). Figure 11 shows the cost vs performance comparison of feature selection methods for re-identification using DDM. The performance is measured using the area under the first half of the CMC curves and the cost is measured using Eq. 11. In CoPE, a consistent increase in performance and cost can be observed with the addition of each new feature. Since the most important features for re-identification are selected first, the rate of improvement in the performance is high at the beginning and the performance monotonically increases as the cost increases, the most desirable behavior in feature budgeting.

In VIPeR (Fig. 11 (a)), the performance of Fisher score and Information gain becomes constant after selection of up to 9 features due to minimal weighting to the lower ranked features. However, the low ranked features keep increasing the cost of the feature set. Such feature selection represents the majority of data with similar properties while neglecting the features with discriminating capability for small amounts of data. The mRMR feature selection produces a monotonically decreasing performance after reaching a high performance point because of the ranking only strategy. Since VIPeR requires up to 6 discriminant features as selected by CoPE, the additional features result in redundant information and the performance decreases (mRMR) or remains constant (Fisher and Information gain), while the cost increases. In i-LIDS dataset (Fig. 11 (b)), Fisher score shows a non-monotonically increasing performance at the start and while selecting the second feature shows a higher performance than CoPE because of the selection of a comparatively costly feature (with Feature ID=4). However,



(a) VIPeR (b) i-LIDS

Fig. 11. Cost vs performance analysis of CoPE (blue star) and existing feature selection methods: Fisher score [19] (cyan plus), Information gain [20] (magenta triangle), mRMR [21] (black circle), ReliefF [22] (yellow triangle) and Bi-clusters [23] (green circle) using DDM (Bhattacharyya) for re-identification on (a) VIPeR and (b) i-LIDS. Features are added in order of decreasing performance. Solid lines show the number of selected features equal to those generated with CoPE. Dotted lines extend the cost vs performance comparison for the feature excluded by CoPE. The cost (horizontal axis) is measured using (Eq. 11) and the performance (vertical axis) is measured as the area under the first-half of the CMC curves. At each marker point a new feature is added.

as new features are added, the performance starts decreasing, while CoPE preserves a balance between cost and performance, which results not only in a monotonically increasing performance but also in the highest performance with the smallest cost when the same number of features are used. The specific feature (with ID=4) is selected by CoPE at a later stage when its cost justifies the performance. A non-monotonically increasing performance is observed in the Information gain and Bi-clusters, however their performance is lower than that of CoPE as the cost increases.

This evaluation shows how CoPE can select, in the correct order, less expensive and well-performing features. Improved or comparable performance compared to existing selection approaches is achieved for DDM and learning methods for re-identification with cost-effective features.

V. Conclusions

We proposed a feature selection approach that identifies the most appropriate features for person re-identification. The amount of data stored for each feature and the computational time for its extraction are used jointly with their performance to generate an overall feature score. The best features are selected in a defined range of scores to reduce the performance overlap, a measure of similarity among features. A further reduction of the selected features is made possible to account for additional operational constraints (e.g. limited resources). The proposed method decreases both the amount of data generated per feature set and the amount of time needed for the extraction of the selected feature set, up to 80% in VIPeR and up to 70%, in i-LIDS dataset without compromising the re-identification rate compared to existing re-identification approaches (PRDC [15] and ASFI [17]). We also demonstrated that CoPE improves the performance of other learning-based re-identification approaches such as those based on rankSVM [14] and AdaBoost [16] compared to the existing feature selection methods. Future work includes implementing the proposed approach on actual smart cameras and to extract the head locations using the trained detectors.

References

[1] H. Aghajan and A. Cavallaro, *Multi-Camera Networks: Principles and Applications*. Elsevier, 2009.
[2] V. Sulic, J. Pers, M. Kristan, and S. Kovacic, "Efficient feature distribution for object matching in visual-sensor networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 7, pp. 903 –916, 2011.
[3] Y. Charfi, N. Wakamiya, and M. Murata, "Challenging issues in visual sensor networks," *IEEE Wireless Commun. Mag.*, vol. 16, no. 2, pp. 44 –49, 2009.
[4] S. Soro and W. Heinzelman, "A survey of visual sensor networks," *Advances in Multimedia*, vol. 09, pp. 1–21, 2009.
[5] D.-U. Lee, H. Kim, M. H. Rahimi, D. Estrin, and J. D. Villasenor, "Energy-efficient image compression for resource-constrained platforms," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2100–2113, 2009.
[6] V. Lecuire, C. Duran Faundez, and N. Krommenacker, "Energy efficient image transmission in sensor networks," *International Journal of Sensor Networks*, vol. 4, no. 1, pp. 37–47, 2008.
[7] R. Vezzani, D. Baltieri, and R. Cucchiara, "People re-identification in surveillance and forensics: a survey," *ACM Computing Surveys*, vol. 46, no. 2, pp. 1–37, 2013.
[8] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher, "Appearance-based person reidentification in camera networks: problem overview and current approaches," *Journal of Ambient Intelligence and Humanized Computing*, vol. 2, no. 2, pp. 127–151, 2011.
[9] R. Mazzon, S. F. Tahir, and A. Cavallaro, "Person re-identification in crowd," *Pattern Recognition Letters*, vol. 33, no. 14, pp. 1828–1837, 2012.
[10] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. CVPR*, San Francisco, USA, June 2010.
[11] L. Bazzani, M. Cristani, A. Perina, and V. Murino, "Multiple-shot person re-identification by chromatic and epitomic analyses," *Pattern Recognition Letters*, vol. 33, no. 7, pp. 898–903, May 2012.
[12] Y. Cheng, W. Zhou, Y. Wang, C. Zhao, and S. Zhang, "Multi-camera-based object handoff using decision-level fusion," in *Int. congress on Image and Signal Processing*, Tianjin, China, October 2009.
[13] O. Javed, K. Shafique, Z. Rasheed, and M. Shah, "Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views," *Comput. Vis. Image Understand.*, vol. 109, no. Issue 2, pp. 146 – 162, February 2008.
[14] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by support vector ranking," in *Proc. BMVC*, Aberystwyth, UK, August 2010.

[15] W.-S. Zheng, S. Gong, and T. Xiang, "Re-identification by relative distance comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, 2013.
[16] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. ECCV*, Marseille, France, October 2008.
[17] C. Liu, S. Gong, C. C. Loy, and X. Lin, "Person re-identification: What features are important?" in *Proc. ECCV*, Firenze, Italy, October 2013.
[18] K. Jeong and C. Jaynes, "Object matching in disjoint cameras using a colour transfer approach," *Springer Journal of Machine Vision and Applications*, vol. 19, no. 5, pp. 88–96, September 2008.
[19] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. John Wiley & Sons, New York, 2001.
[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.
[21] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, 2005.
[22] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of relieff and rrelieff," *Machine Learning*, vol. 53, no. 1-2, pp. 23–69, Oct. 2003.
[23] Q. Huang, D. Tao, X. Li, L. Jin, and G. Wei, "Exploiting local coherent patterns for unsupervised feature ranking," *IEEE Trans. Syst., Man, Cybern. B*, vol. 41, no. 6, pp. 1471–1482, 2011.
[24] E. Guldogan and M. Gabbouj, "Feature selection for content-based image retrieval," *Signal, Image and Video Processing*, vol. 2, no. 3, pp. 241–250, 2008.
[25] V. Dang and W. B. Croft, "Feature selection for document ranking using best first search and coordinate ascent," in *Proc. SIGIR Workshop*, Geneva, Switzerland, July 2010.
[26] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1, pp. 273–324, 1997.
[27] X. Geng, T. yan Liu, T. Qin, and H. Li, "Feature selection for ranking," in *Proc. SIGIR*, Amsterdam, Netherlands, July 2007.
[28] T. Yan and Y. Hou, "An unsupervised feature selection method based on degree of feature cooperation," in *Proc. IEEE FSKD*, Shanghai, China, July 2011.
[29] H.-L. Wei and S. A. Billings, "Feature subset selection and ranking for data dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 162–166, 2007.
[30] F. Pan, T. Converse, D. Ahn, F. Salvetti, and G. Donato, "Feature selection for ranking using boosted trees," in *Proc. ACM CIKM*, New York, USA, November 2009.
[31] G. Hua, M. Zhang, Y. Liu, S. Ma, and L. Ru, "Hierarchical feature selection for ranking," in *Proc. World Wide Web*, Raleigh NC, USA, April 2010.
[32] X. Sun, Y. Liu, J. Li, J. Zhu, H. Chen, and X. Liu, "Feature evaluation and selection with cooperative game theory," *Pattern Recognition*, vol. 45, no. 8, pp. 2992–3002, 2012.
[33] S. Mitra, P. P. Kundu, and W. Pedrycz, "Feature selection using structural similarity," *Information Science*, vol. 198, no. 0, pp. 48–61, 2012.
[34] L. Wang, "Feature selection with kernel class separability," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1534–1546, 2008.
[35] J. Liang, S. Yang, and A. Winstanley, "Invariant optimal feature selection: A distance discriminant and feature ranking based solution," *Pattern Recognition*, vol. 41, no. 5, pp. 1429–1439, 2008.
[36] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," in *Proc. ECCV*, 2006, pp. 589–600.
[37] W. Ayedi, H. Snoussi, and M. Abid, "A fast multi-scale covariance descriptor for object re-identification," *Pattern Recognition Letters*, vol. 33, no. 14, pp. 1902 – 1907, 2012.
[38] K.-W. Chen, C.-C. Lai, P.-J. Lee, C.-S. Chen, and Y.-P. Hung, "Adaptive learning for target tracking and true linking discovering across multiple non-overlapping cameras," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 625 –638, 2011.
[39] G. Lian, J.-H. Lai, C. Suen, and P. Chen, "Matching of tracked pedestrians across disjoint camera views using ci-dlbp," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 7, pp. 1087 –1099, 2012.
[40] S. Bak, E. Corvee, F. Bremond, and M. Thonnat, "Multiple-shot human re-identification by mean riemannian covariance grid," in *Proc. IEEE AVSS*, Klagenfurt, Austria, September 2011, pp. 179 –184.
[41] E. Corvee, S. Bak, and F. Bremond, "People detection and re-identification for multi surveillance cameras," in *Proc. VISAPP*, February 2012, pp. 82–88.
[42] N. Martinel and C. Micheloni, "Re-identify people in wide area camera network," in *Proc. CVPR Workshop*, June 2012, pp. 31 –36.
[43] M. Hirzer, P. M. Roth, and H. Bischof, "Person re-identification by efficient impostor-based metric learning," in *Proc. IEEE AVSS*, September 2012, pp. 203 –208.
[44] R. Satta, G. Fumera, and F. Roli, "Fast person re-identification based on dissimilarity representations," *Pattern Recognition Letters*, vol. 33, no. 14, pp. 1838–1848, 2012.
[45] S. Khan and M. Shah, "Consistent labeling of tracked objects in multiple cameras with overlapping fields of view," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1355–1360, 2003.
[46] i LIDS, "Home office multiple camera tracking scenario definition (uk)," 2008. [Online]. Available: http://www.homeoffice.gov.uk/science-research/hosdb/i-lids/(lastaccessed:June,2013)
[47] S. Gong, T. Xiang, and S. Hongeng, "Learning human pose in crowd," in *Proc. ACM Multimedia*, Firenze, Italy, October 2010.
[48] H. Xu, P. Lv, and L. Meng, "A people counting system based on head-shoulder detection and tracking in surveillance video," in *Proc. IEEE ICCDA*, Qinhuangdao, China, June 2010, pp. 394–398.

Syed Fahad Tahir received his MS degree in Computer System Engineering from Ghulam Ishaq Khan Institute of Engineering Sciences and Technology (GIKI) in 2006 and BS degree in Computer Sciences from National University of Computer and Emerging Sciences (FAST-NU) in 2004. He served as Manager (Technical) in the Centre of Excellence in Science and Applied Technologies, a Government Research Organization of Pakistan. Since 2011, he has been with Queen Mary University of London, UK, and Alpen-Adria Universitat Klagenfurt, Austria, as a PhD researcher under the supervision of Prof. A. Cavallaro and Prof. B. Rinner. He was a NESCOM Fellow during his MS studies and was awarded Erasmus Mundus fellowship for his Double Doctorate. He has more than 6 journal and conference publications in the areas of re-identification in multi-camera networks, pattern analysis and object matching.

Andrea Cavallaro received his Ph.D. degree in electrical engineering from the Swiss Federal Institute of Technology, Lausanne, in 2002, and the Laurea (summa cum laude) degree in electrical engineering from the University of Trieste in 1996. He is Professor of multimedia signal processing and the Director of the Centre for Intelligent Sensing, Queen Mary University of London, U.K. He was a Research Fellow with British Telecommunications in 2004 and 2005 and was awarded the Royal Academy of Engineering Teaching Prize in 2007; three Student Paper Awards on target tracking and perceptually sensitive coding at the IEEE and the Best Paper Award at the IEEE AVSS 2009. He is an Area Editor for the IEEE SIGNAL PROCESSING MAGAZINE and an Associate Editor for the IEEE TRANSACTIONSON IMAGE PROCESSING. He is an elected member of the IEEE Signal Processing Society, Image, Video, and Multidimensional Signal Processing Technical Committee, and a Chair of its Awards Committee. He served as an elected member of the IEEE Signal Processing Society, Multimedia Signal Processing Technical Committee, as an Associate Editor for the IEEE TRANSACTIONS ON MULTIMEDIA and the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and as a Guest Editor for seven international journals. He was a General Chair for the IEEE/ACM ICDSC 2009, BMVC 2009, M2SFA2 2008, SSPE 2007, and the IEEE AVSS 2007. He was a Technical Program Chair of the IEEE AVSS 2011, the European Signal Processing Conference (EUSIPCO 2008), and WIAMIS 2010. He has published more than 130 journal and conference papers, one monograph on Video Tracking (Wiley, 2011) and three edited books: Multi-Camera Networks (Elsevier, 2009), Analysis, Retrieval and Delivery of Multimedia Content (Springer, 2012), and Intelligent Multimedia Surveillance (Springer, 2013).