

Discriminant Context Information Analysis for Post-Ranking Person Re-Identification

Jorge García, *Student Member, IEEE*, Niki Martinel, *Student Member, IEEE*, Alfredo Gardel, *Member, IEEE*, Ignacio Bravo, *Member, IEEE*, Gian Luca Foresti, *Senior Member, IEEE*, and Christian Micheloni, *Member, IEEE*

Abstract—Existing approaches for person re-identification are mainly based on creating distinctive representations or on learning optimal metrics. The achieved results are then provided in the form of a list of ranked matching persons. It often happens that the true match is not ranked first but it is in the first positions. This is mostly due to the visual ambiguities shared between the true match and other “similar” persons. At the current state, there is a lack of a study of such visual ambiguities which limit the re-identification performance within the first ranks. We believe that an analysis of the similar appearances of the first ranks can be helpful in detecting, hence removing, such visual ambiguities. We propose to achieve such a goal by introducing an unsupervised post-ranking framework. Once the initial ranking is available, content and context sets are extracted. Then, these are exploited to remove the visual ambiguities and to obtain the discriminant feature space which is finally exploited to compute the new ranking. An in-depth analysis of the performance achieved on three public benchmark data sets support our beliefs. For every data set, the proposed method remarkably improves the first ranks results and outperforms the state-of-the-art approaches.

Index Terms—Person re-identification, re-ranking, post-ranking, context information.

I. INTRODUCTION

THE recent advancement and the price plummet of imaging sensor technologies have remarkably increased the adoption of video analytics systems for various application ranging from home to border surveillance. However, it is a matter of fact that, due to the amount of human supervision, privacy concerns, and maintenance costs involved, it is still not possible to deploy a camera network covering all the

Manuscript received March 22, 2016; revised September 2, 2016 and November 30, 2016; accepted December 30, 2016. Date of publication January 15, 2017; date of current version February 17, 2017. This work was supported in part by the University of Alcalá through the IDEA Project under Grant CCG2015/EXP-041 and in part by the University of Udine through the PRESNET Project under Grant 1889/2016. The associate editor coordinating the review of this manuscript and approving it for publication was Mr. Pierre-Marc Jodoin. (*Jorge García and Niki Martinel are co-first authors.*)

J. García, A. Gardel, and I. Bravo are with the Department of Electronics, University of Alcalá, 28871 Madrid, Spain.

N. Martinel, G. L. Foresti, and C. Micheloni are with the Department of Mathematics, Computer Science and Physics, University of Udine, 33100 Udine, Italy.

This paper has supplementary downloadable multimedia material available at <http://ieeexplore.ieee.org> provided by the authors. This includes an evaluation of the proposed solution with different features. It also provides experiments on two additional datasets that, due to space constraints, have been excluded from the main paper. This material is 3 MB in size.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2017.2652725

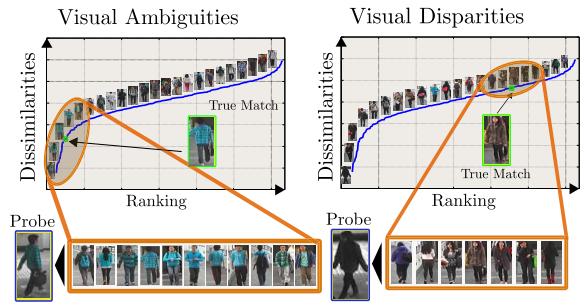


Fig. 1. Example of visual ambiguities and visual disparities. Persons images in the first ranks share visual ambiguities, while persons images in the last ones have visual disparities. In this work, the objective is to conduct an analysis on the visual ambiguities and remove them such that the initial ranking can be improved. (Best viewed in color).

areas of a wide environment. Thus, only a small portion of it is monitored by cameras which often have non-overlapping fields-of-view (FoVs). As a result, there exist blind areas from which no information can be directly obtained. This raises the need for methods able to link the information acquired between the covered areas such that high-level semantics can be obtained. In particular, one of the most currently attractive issues that such blind areas have introduced is the problem of re-associating a same person that is moving in a wide environment and who might be detected at a different location and time. This is known as the person re-identification problem. Countless applications like multi-camera tracking [1], [2], situational awareness [3], [4], and multi-camera event detection [5]–[7] would benefit.

The inherent conditions of a multi-camera network make the person re-identification an open and challenging problem. Indeed, to cover larger portions of the monitored environment the deployed cameras are often installed on poles, roofs, etc. This yields to low quality footages acquired by different viewpoints and with different photometric settings. In addition, when a person moves across the disjoint FoVs, his/her appearance undergoes significant illumination and color variations. The non-rigid shape of the human body, as well as background clutter, introduce further challenges.

In the recent past, the research community endeavored to overcome the aforementioned issues by proposing different methods based on: (i) discriminative signatures (e.g. [8]–[10]), (ii) feature transformations (e.g. [11]–[13]) and (iii) metric learning (e.g. [14]–[16]). These approaches use similar

pipelines, but they differ from how the person visual appearance is modeled and/or how the match is computed. Under the considered wide environment scenarios, due to the location of the cameras, the image resolution and the size of the region of interest that include the persons, biometric analysis-based methods are commonly not adopted.

In spite of all such efforts, the currently achieved performance are not satisfactory and sufficient to provide systems able to autonomously solve the re-identification problem. As a result, there is still the need of human supervision. Indeed, the re-identification problem is usually cast as a ranking problem whose results, in form of a ranking list of matched persons, are presented to the end user for the final judgment. The majority of the works proposed so far assume that the provided ranking list is optimal and it is suitable for end user inspection. It is our belief that such an initial ranking is not optimal for the task and it is just a first step to remove the majority of the possible mismatches. Additional inspections on the ranked matches can be applied to refine the output in such a way that the correct match will have higher probability to be found in the first ranks. Hence, the current work is founded on the idea that a ranking, achieved by any algorithm, contains valuable information which can be further exploited to improve the rank of the true match (e.g., true match within the first 1-10 ranks).

As shown in Fig. 1, in a common ranking list of matches two main regions can be identified. In the first (i.e., low) ranks region, the dissimilarities between gallery images and the probe are very low due to the shared visual ambiguities. In the last (i.e., high) ranks region, the dissimilarities between gallery images and the probe are very high due to the visual disparities [17] that are introduced by variations in viewpoints, pose, illumination changes, etc. It is often the case that existing methods have not collected enough ability to identify the relevant information hidden within the visual ambiguities to best rank the true match. This motivates a study of the visual ambiguities occurring at first ranks such that discriminative information can be used to better rank the true match. For such a purpose, we propose an unsupervised post-ranking framework. To find the visual ambiguities and remove them, the concepts of content and context information carried by the initial ranking are coined. The content information is the set of gallery persons that have low dissimilarity with respect to the probe. The context information is the set of gallery persons that have low dissimilarity with both the probe and a person of the content information. In this way, the context information leads to extract the global appearance shared by the probe and the content information, thus the visual ambiguities. Then, this is removed before re-ranking. We named such a framework discriminant context information analysis (DCIA).

The rest of the paper is organized as follow. In section II, state-of-the-art works in the field of person re-identification are discussed. An overview of our approach is given in section III. In section IV, we delve deeper into the details of the proposed DCIA framework. The superior performance of our approach over recent state-of-the-art methods is shown in section V. Finally, conclusions are drawn in section VI.

II. RELATED WORK

In the following, we briefly overview existing approaches as classified in recent surveys [18], [19].

Existing *discriminative signature based methods* exploited multiple local and global features [20]–[23] to compute the persons representations. These representations were combined with reference sets [24], patch matching strategies [25], saliency learning [26], [27] and joint attributes [28]–[30]. An interesting bag-of-words approach, together with a benchmark dataset, was proposed in [31]. Both labeled and unlabeled data [32] as well as collaborative representations [33] were used to boost performance. Learning architectures [8], [34], [35] and multiple frame analysis methods [9], [36] were also explored to extract the most relevant features.

An earliest work on *features transformation* for re-identification was proposed in [37] where a unique brightness transfer function (BTF) computed between features was used to match persons across camera pairs. Recent works [11]–[13], [38] considered that the transformation is not unique and it depends on several factors. The case where no labeled data for training is available was also studied in [39] by means of a transfer learning framework.

The authors of [40] proposed a *metric learning* framework which minimizes the distance between features of pairs of true matches, while maximizing the same between pairs of wrong matches. Performance were improved by learning a relaxed Mahalanobis metric [41], by considering multiple metrics [15], [42]–[44] in a transfer learning set up [45], or by relying on equivalence constraints [16], [46], [47]. Other methods have focused on local distance comparison problems [14], [48], sparse pairwise constraints [49] and human-in-the-loop solutions [50]. Network-wise solutions were also investigated in [51].

Post-ranking methods for person re-identification is a relatively unexplored area. Earliest works following the post-ranking approach exploited boosting techniques for feature selection [52] or additional cues coming from soft biometrics [53]. However, in such a case, there is the need to acquire reliable biometric features which is generally a challenge in surveillance scenarios. In [54], ranked lists computed for multiple probe persons were exploited to refine a single probe ranking. Therefore, the approach works only if additional rankings (minimum 3 or 4), besides the one obtained for the current probe, are available. Bidirectional ranking [55] and a saliency-based matching scheme [56] were also introduced. In the former case, first direction is usual ranking of the probe with the gallery. Second direction is the ranking obtained by matching each gallery with the probe and the rest of the gallery. Hence, differently from our approach, the whole gallery for post-ranking is considered, and no focus is placed on the visual ambiguities shared between first ranks. In the latter, the saliency similarity is computed between the probe and the gallery only. Then, such similarities are adopted to revise the initial ranking within a local gallery window.

The post-ranking optimization was also studied by including human feedback in the loop. The end user had to identify

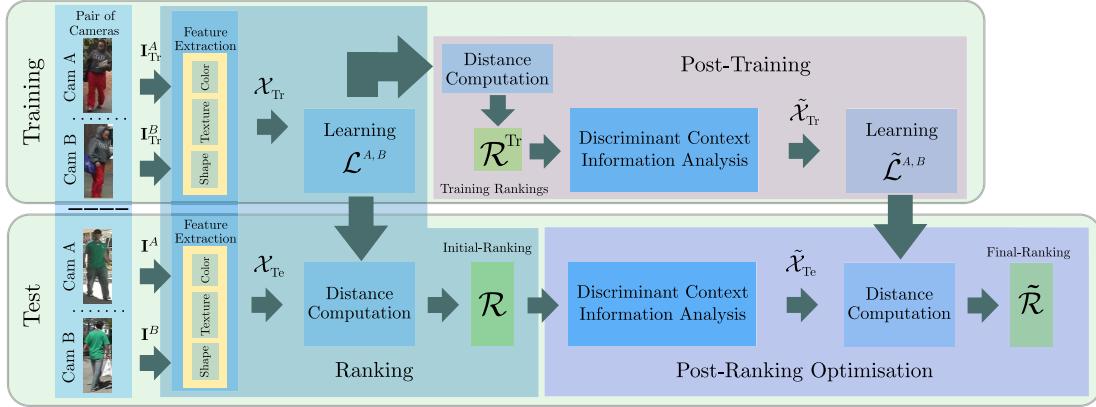


Fig. 2. Overview of the proposed re-identification framework consisting of three modules: ranking, post-training and post-ranking optimisation. The ranking module defines the basis for the approach. This can be substituted with existing re-identification methods. The Discriminant Context Information Analysis (DCIA) considered in the post-training and the post-ranking optimisation modules represents the main contribution of this work. (Best viewed in color).

both similar and dissimilar samples to adapt the metric model [57], to provide relative feedback for the classifier training [58], or to select a single strong negative feedback to refine the ranking [17] in the deployment stage. Similar and dissimilar samples can be picked by user within local regions [59]. In contrast to all such methods, we propose a single-shot approach that does not require human intervention.

Finally, a slightly different approach was recently introduced in [10]. An iterative extension to sparse discriminative classifiers was adopted to ensure that the best candidates are ranked at each iteration. However, such a method did not directly consider the content and the context similarities of ranked individuals. It attacks the problem by analyzing the reconstruction error and by partially ranking the gallery in terms of similarity to the probe.

An earlier version of this work appeared in [60]. This paper introduces: (i) additional technical details; (ii) a clearer explanation of the contribution; and (iii) deeper experimental evaluations.

To conclude, three main differences between the proposed approach and all such existing works can be highlighted: (i) we are not considering the whole gallery for re-ranking; (ii) the approach does not require human intervention both in the training and in the re-identification stages; (iii) most importantly, the proposed approach is the only one studying the visual ambiguities shared between first ranks to improve re-identification performance.

III. OVERVIEW OF THE APPROACH

The architecture of the proposed re-identification framework is shown in Fig. 2. It consists of three main modules: ranking, post-training and post-ranking optimization.

The ranking computation module resembles commonly used re-identification pipelines and it defines the basis for our approach. Let \mathcal{I}_{Tr} be the set of training image pairs ($\mathbf{I}_{\text{Tr}}^A, \mathbf{I}_{\text{Tr}}^B$) acquired by disjoint cameras A and B . To model the appearance of each image, a vector $\mathbf{x} \in \mathbb{R}^d$ of concatenated features including color, texture and shape is extracted. To each image pair corresponds a label y which assumes value



Fig. 3. Ranking results and corresponding content sets. Gallery images in the first ranks show visual ambiguities with the probe. The probe dissimilarities clustering process determines which gallery images should be included in the content sets (orange rectangles). This enables us to focus on the visual ambiguities which are shared with the true match (green rectangles).

+1 if the person in the two images is the same (positive image pair), -1 otherwise (negative image pair). The set \mathcal{X}_{Tr} containing $|\mathcal{I}_{\text{Tr}}|$ pairs of training feature vectors, together with the corresponding pair labels, is used by a learning algorithm to build the model $\mathcal{L}^{A,B}$ which optimally separates the sets of positive and negative pairs.

The same set of vectors and the trained model are input to the post-training module. This is essential to tune the method on the new features that are created after removing visual ambiguities. It starts by computing the set of training rankings \mathcal{R}^{Tr} , i.e., a ranking for each training probe sample in camera A is computed over training gallery samples in camera B . Then, the information carried by such rankings is exploited by the discriminant context information analysis module which performs the following steps for each probe. First, the content set is obtained from a subset of the persons in the initial ranking. Then, for each person in such a set the context set is extracted. Both the content and context sets are exploited to identify the visual ambiguities which are finally removed from the corresponding feature vectors. The set of transformed feature vectors pairs, denoted as $\tilde{\mathcal{X}}_{\text{Tr}}$, carries the discriminative information of the relevant persons.

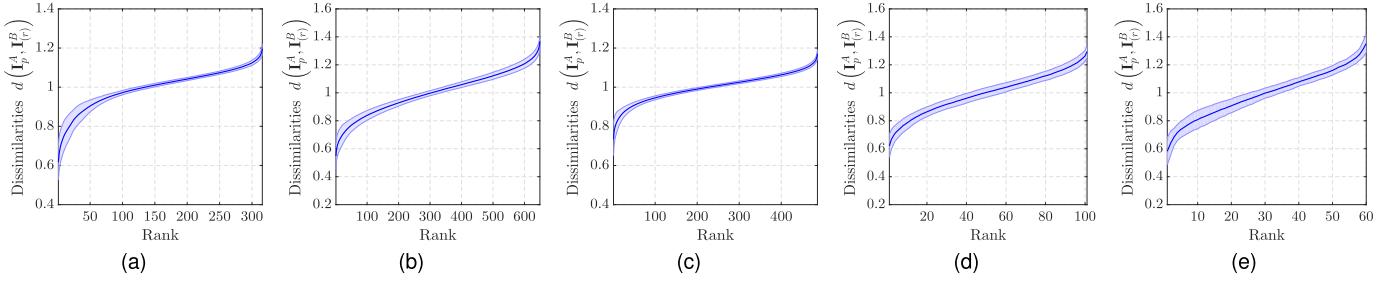


Fig. 4. Ranks and corresponding dissimilarities averaged over all the probe images in the (a) VIPeR, (b) PRID, (c) CUHK02, (d) 3DPeS, and (e) i-LIDS datasets. The blue shaded area corresponds to \pm the standard deviation. Results show a trend that is shared among all the datasets which can be exploited to identify the content set.

Finally, a learning algorithm is trained with such transformed feature vectors to separate between positive and negative pairs. The obtained model is denoted as $\tilde{\mathcal{L}}^{A,B}$.

In the re-identification phase, features extracted from a probe and every gallery image are collected in the set \mathcal{X}_{Te} of feature vectors pairs. This, together with the model $\mathcal{L}^{A,B}$, is used to compute the initial ranking \mathcal{R} . Then, the post-ranking optimization module exploits the discriminant context information analysis, hence the content and context sets extracted using \mathcal{R} , to identify the visual ambiguities. These are removed from the corresponding feature vectors and the final ranking $\tilde{\mathcal{R}}$ is computed by evaluating the set of transformed feature vectors pairs $\tilde{\mathcal{X}}_{\text{Te}}$ with the learned model $\tilde{\mathcal{L}}^{A,B}$.

IV. DISCRIMINANT CONTEXT INFORMATION ANALYSIS

A. Preliminaries and Definitions

Let $\mathcal{A} = \{\mathbf{I}_p^A\}_{p=1}^N$ be the set of N probe images and $\mathcal{B} = \{\mathbf{I}_g^B\}_{g=1}^M$ be the set of M gallery images. Given a probe image \mathbf{I}_p^A its initial ranking is defined as $\mathcal{R}_p = \{\mathbf{I}_{(r)}^B\}_{r=1}^M$ where the gallery images \mathbf{I}_g^B are sorted depending on the dissimilarity to the probe image \mathbf{I}_p^A . Notice that, here, as well as in the following, the subscript is used without parenthesis to denote a generic gallery person identity g , and it is enclosed within parenthesis to indicate its position in a ranking list. In other words, $d(\mathbf{I}_p^A, \mathbf{I}_{(r)}^B) < d(\mathbf{I}_p^A, \mathbf{I}_{(r+1)}^B)$, where $r = 1, \dots, M - 1$. Such an order is computed on the basis of a dissimilarity measure $d(\mathbf{I}_p^A, \mathbf{I}_g^B)$ which is the result of the application of the model $\mathcal{L}^{A,B}$ on the feature vector pair $(\mathbf{x}_p^A, \mathbf{x}_g^B)$. $\mathcal{R} = \{\mathcal{R}_p\}_{p=1}^N$ denotes the set of such initial rankings computed for all the N probes.

Our aim is to improve the rank of the true match in each \mathcal{R}_p . Towards this objective we first select the *content set*, i.e., a subset of gallery images $\mathcal{B}^{\text{cn}} \subseteq \mathcal{B}$ whose elements belong to the first ranks. Then, we compute the *context set* $\mathcal{B}^{\text{cx}} \subseteq \mathcal{B}$, which contains gallery images that have small dissimilarity with respect to either the probe or an image in the content set.

B. Content Analysis

Existing methods try to locate the true match in the first ranking positions out from a large set of possible gallery matches. As shown in Fig. 3, the visual ambiguities bring false matches in the first ranks, often before the true match.

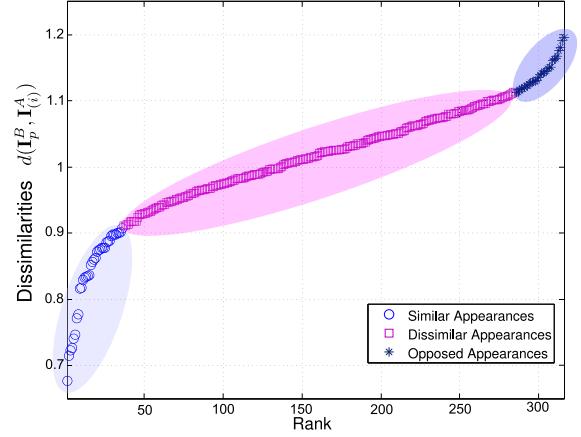


Fig. 5. An example of the results obtained by applying k-means clustering algorithm to the dissimilarities computed between a probe and all the gallery images. Similar, dissimilar and opposed appearance clusters are represented by blue circles, magenta squares and dark blue stars, respectively.

To detect such ambiguities, we propose to analyze the information carried by the content set of a given probe \mathbf{I}_p^A . Elements in such a set are selected from the top m positions in the initial ranking \mathcal{R}_p .

To select such m images, we propose a dynamic method that does not require m to be fixed *a priori* but can vary for every probe. The selection process is inspired by the shape of the dissimilarities vs ranks plots. Fig. 4 shows that there exists a significant trend among all the datasets highlighting that: (i) at first ranks, dissimilarities with the probe image increases abruptly, then flatten (first elbow); (ii) from the first elbow, dissimilarities grow linearly till reaching high ranks, where they finally start increasing significantly (second elbow). According to such trend we have identified three classes of gallery images (see Fig. 5): (i) similar appearance class (C_{sa}), which corresponds to gallery images with dissimilarities located before the first elbow; (ii) dissimilar appearance class (C_{da}) corresponding to gallery images having dissimilarities located in-between the two elbows and (iii) opposed appearance class (C_{oa}), which corresponds to all the other gallery images.

To identify gallery images sharing visual ambiguities with a probe, C_{sa} has to be computed. Towards this objective, we propose to use the k-means clustering algorithm to partition \mathcal{B} into the three aforementioned sets, which minimize the

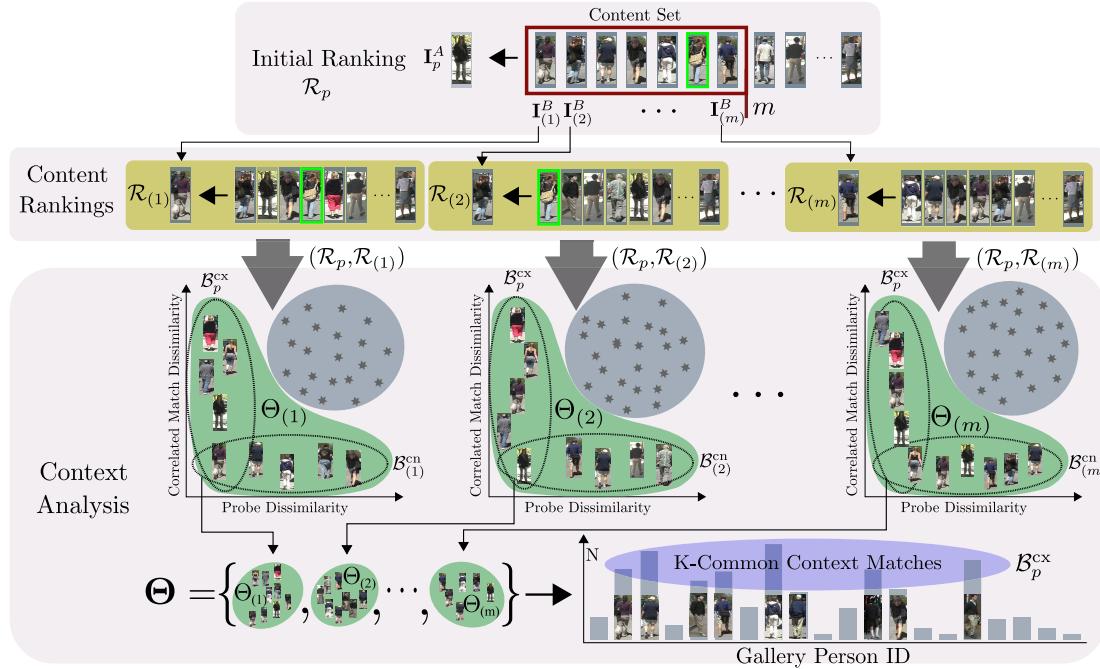


Fig. 6. Representation of content and context sets computed for a probe image. Gallery images inside the red rectangle represent the content set. Each of such images is used to compute a ranking with all the gallery images in \mathcal{B}^* . The obtained content rankings are then clustered on the basis of the similarity of each gallery with the image in the content set and the probe. The K -common gallery images in Θ are used to define the context set. (Best viewed in color).

following cost function

$$\sum_{j \in \{\text{sa}, \text{da}, \text{oa}\}} \sum_{\mathbf{I}_{(r)}^B \in \mathcal{C}_j} \|d(\mathbf{I}_p^A, \mathbf{I}_{(r)}^B) - \mu_j\|^2, \quad r = 1, \dots, M \quad (1)$$

where μ_j is the mean of the dissimilarities within the j -th partition. Before minimizing the cost function, we initialize the cluster means as $\mu_{\text{sa}} = d(\mathbf{I}_p^A, \mathbf{I}_{(1)}^B)$, $\mu_{\text{da}} = d(\mathbf{I}_p^A, \mathbf{I}_{(M/2)}^B)$ and $\mu_{\text{oa}} = d(\mathbf{I}_p^A, \mathbf{I}_{(M)}^B)$. Thus, once the minimization is concluded, the cluster C_{sa} contains the top- m best matches, hence defines the content (cn) set $\mathcal{B}_p^{\text{cn}} = \{\mathbf{I}_{(1)}^B, \dots, \mathbf{I}_{(m)}^B\}$.

Notice that, only persons whose images are in the content set are re-ranked. This is compliant with the assumption that the true match is probably located in the first rank positions.

C. Context Analysis

Context information can be defined as the object frequency appearance in a particular domain. In image retrieval, the context information is extracted from the set of images containing the target object [61]. We provide a similar definition for the person re-identification problem: the context information is extracted from the K -common nearest neighbors of the probe and a gallery in the content set.

Given a probe, we extract the context set by exploiting the content one (see Fig. 6). For each image $\mathbf{I}_{(r)}^B \in \mathcal{B}_p^{\text{cn}}$, $r = 1, \dots, m$, we compute $\mathcal{B}_{(r)}^{\text{cn}}$ exploiting the content ranking $\mathcal{R}_{(r)}$ which is obtained using the gallery set $\mathcal{B}^* = \mathcal{B} \setminus \mathbf{I}_{(r)}^B$. Given the content sets $\mathcal{B}_p^{\text{cn}}$ and $\mathcal{B}_{(r)}^{\text{cn}}$, we define $\Theta_{(r)} = \mathcal{B}_p^{\text{cn}} \cup \mathcal{B}_{(r)}^{\text{cn}}$. The elements in $\Theta_{(r)}$ are the images in \mathcal{B} having high similarity with either the probe \mathbf{I}_p^A or the gallery $\mathbf{I}_{(r)}^B$.

Then, the K images having highest frequency of appearance in $\Theta = \{\Theta_{(r)}\}_{r=1}^m$, i.e., the K -common nearest neighbors,

form the context (cx) set $\mathcal{B}_p^{\text{cx}}$. The threshold K has been introduced to reject gallery images that have low frequency of appearance, i.e. images not having strong similarity neither with the content set images nor with the probe. As it will be shown in the experiments, the performance degrades if such images are considered in the detection of the visual ambiguities.

D. Discriminant Information Analysis

The content and context sets include images with “similar” appearance. The goal of the following analysis is to detect, hence to remove, the visual ambiguities shared by these images. This allows us to focus on discriminant details that help to correctly locate the true match within the content set, thus improving its rank.

We believe that visual ambiguities correspond to the global appearance which mainly contributes to the dissimilarity computation. Motivated by this, we propose to represent such information with the principal components of the considered images. In particular, we apply principal component analysis (PCA) to the set of feature vectors extracted from images belonging to the content and context sets to find their common information. Then, this is removed from the such feature vectors. This yields to novel discriminant feature vectors that carry detailed information suitable for post-ranking.

More formally, given a probe image \mathbf{I}_p^A , let $\mathcal{B}_p^{\text{cn,cx}} = \mathcal{B}_p^{\text{cn}} \cup \mathcal{B}_p^{\text{cx}}$ be the set containing the gallery images having “similar” appearance with the probe and the images in the content set (without duplicates). To model the global appearance we first represent all such images in terms of feature vectors and define $\mathbf{X}_p = [\mathbf{x}_p^A, \mathbf{x}_1^{\text{cn,cx}}, \dots, \mathbf{x}_n^{\text{cn,cx}}]$ where \mathbf{x}_p^A and $\mathbf{x}_g^{\text{cn,cx}}$

TABLE I
DETAILS AND COMPARISON OF THE CONSIDERED PERSON RE-IDENTIFICATION BENCHMARK DATASETS. FOR CUHK02,
IN BRACKETS THE NUMBER OF PERSONS AND CAMERAS CONSIDERED TO COMPUTE THE REPORTED RESULTS

Dataset	Persons	Cameras	Additional Info
VIPeR [62]	632	2	Location: outdoor Challenges: viewpoint variation, light and color changes http://vision.soe.ucsc.edu/node/178
PRID [52]	934	2	Location: outdoor Challenges: illumination and color changes, similar appearances (visual ambiguities) http://lrs.icg.tugraz.at/download.php
CUHK02 [12]	1816 (971)	10 (2)	Location: outdoor Challenges: viewpoint variation, and attributes (bag-packs, handbags, folders, etc.) http://www.ee.cuhk.edu.hk/~xgwang/CUHK_identification.html

are feature vector representations of the probe image and of a generic gallery image in $\mathcal{B}_p^{\text{cn},\text{cx}}$. We assume that the feature matrix $\mathbf{X}_p \in \mathbb{R}^{d \times (n+1)}$, with $n = |\mathcal{B}_p^{\text{cn},\text{cx}}|$, is centered at the origin. Then, PCA is exploited to find the common appearance subspace $\mathbf{U} \in \mathbb{R}^{d \times k}$ represented by the first k principal components of \mathbf{X}_p . To identify the discriminant appearance details, we first project \mathbf{X}_p onto such a subspace. Then, the obtained global information is reconstructed back and removed from the original feature vectors as

$$\tilde{\mathbf{X}}_p = \mathbf{X}_p - \mathbf{U}\mathbf{U}^T\mathbf{X}_p \quad (2)$$

where each column of the resulting matrix $\tilde{\mathbf{X}}_p$ represents a discriminant feature vector $\tilde{\mathbf{x}}$.

Starting from the discriminant context information analysis method, we can now describe the post-training and post-ranking optimization.

1) *Post-Training*: The discriminant context information analysis described in the previous sections is performed for each probe p , thus for each ranking $\mathcal{R}_p^{\text{Tr}} \in \mathcal{R}^{\text{Tr}}$. Then, new parameters of the model $\tilde{\mathcal{L}}^{A,B}$ are learned by using a transformed training set $\tilde{\mathcal{X}}_{\text{Tr}} = \{\tilde{\mathcal{X}}_p\}_{p=1}^N$. $\tilde{\mathcal{X}}_p = \{(\tilde{\mathbf{x}}_p^A, \tilde{\mathbf{x}}_g^{\text{cn},\text{cx}}; y_{p,g}) | g = 1, \dots, n\}$ where $\tilde{\mathbf{x}}_{p,1} = \tilde{\mathbf{x}}_p^A$ and $\tilde{\mathbf{x}}_{p,1+g} = \tilde{\mathbf{x}}_g^{\text{cn},\text{cx}}$ are respectively the transformed feature vectors of the probe and of a generic gallery image $\mathbf{I}_g^{\text{cn},\text{cx}} \in \mathcal{B}_p^{\text{cn},\text{cx}}$. $\tilde{\mathbf{x}}_{p,k}$ denotes the k -th column vector of $\tilde{\mathbf{X}}_p$ and $y_{p,g}$ is the pairwise label associated to probe p and gallery g .

2) *Post-Ranking Optimization*: The post-ranking optimization is performed during the re-identification phase. Let us consider the ranking \mathcal{R} computed using the original feature vectors pairs in \mathcal{X}_{Te} extracted from a test probe and every gallery image. Then, the discriminant context information analysis is exploited to find the content and context sets, hence to obtain the discriminative feature vector set $\tilde{\mathcal{X}}_{\text{Te}}$. The final ranking $\tilde{\mathcal{R}}$ is obtained by sorting the dissimilarities $d(\cdot, \cdot)$ computed by means of $\tilde{\mathcal{L}}^{A,B}$ for each pair in $\tilde{\mathcal{X}}_{\text{Te}}$.

V. EXPERIMENTAL RESULTS

In this section, we report on the performance of our framework on three publicly available benchmark datasets.¹

¹Due to the page limit constraints, experimental results on two additional datasets, namely 3DPeS and i-LIDS, have been included in the supplementary material.



Fig. 7. 15 Image samples from VIPeR dataset. Each column corresponds to a pair of images of the same person captured by the two different cameras.

First, we introduce the characteristics of the selected datasets as well as the experimental settings. Then, we provide an extensive set of experiments to demonstrate that our method is able to improve the initial performance achieved by four existing approaches. Finally, comparisons with state-of-the-art methods are provided to show the superior performance of our solution.

For each experiment, we run 10 trials using independent random splits to select train and test person IDs. Results averaged over these 10 trials are shown in terms of Cumulative Matching Characteristics (CMC) curve [62]. The CMC curve is a plot of the recognition performance versus the rank score and represents the expectation of finding the correct match within top k matches. A single randomly selected image is used for datasets having more than one image of a same person in each camera.

A. Datasets

The most relevant characteristics of each dataset are provided in Table I and discussed in the following. We have selected such datasets because they provide many challenges faced in real scenarios like strong illumination changes, viewpoint variations, occlusions and background clutter.

1) *VIPeR Dataset*: The VIPeR dataset [62] is considered the most challenging one for person re-identification. It contains images of 632 persons viewed by two non-overlapping cameras. The 1264 images have severe lighting variation and background clutter. In addition, image pairs have viewpoint variations of 90, 180 and 270 degrees (see Fig. 7 for a few samples). Following the commonly adopted procedure [46], [62], [63], the dataset has been split into two disjoint subsets of 316 persons, one for training and one for testing, respectively.



Fig. 8. 15 Image samples from PRID dataset. Each column corresponds to a pair of images of the same person captured by the two different cameras.



Fig. 9. 15 Image samples from CUHK02 dataset. Each column corresponds to a pair of images of the same person captured by the two different cameras.

2) PRID Dataset: The PRID dataset [52] contains 1134 images acquired by two disjoint cameras, named camera *A* and camera *B*. 385 persons appears in camera *A* and 749 in camera *B*, but only 200 persons are contained in both cameras. This dataset comes with numerous persons with similar appearance, hence the visual ambiguities are higher than in the VIPeR. For the evaluation, we have adopted the same protocol proposed in [41]: persons from camera *A* have been used as probes and persons from camera *B* as gallery. Among the 200 persons appearing in both cameras (see Fig. 8 for some examples), we have randomly selected 100 persons for training and 100 for testing. The remaining 549 persons appearing only in camera *B* are referred to as the “distractors”. We provide results for the case where distractors are included in the gallery.

3) CUHK02 Dataset: The CUHK02 dataset [12] contains images sensed by five disjoint camera pairs each of which have footages of 971, 306, 107, 193 and 239 persons, respectively. Each person has two images in each camera view. The dataset has been acquired in a campus environment and, at the current state, is one of the largest re-identification datasets. In addition, due to pose variations and lighting changes that occurs between camera pairs it is considered a challenging dataset. To evaluate our approach, we have used the same protocol introduced in [12] and selected the camera pair having 971 persons to compare the performance with state-of-the-arts methods (image pair samples belonging to such camera pair are depicted in Fig. 9). The dataset has been split into two sets, one for training and one for testing, which contain 485 and 486 persons, respectively.

B. Implementation Details and Settings²

To model the person appearance we have used the same representation proposed in [63]. Feature vectors extracted from the resized 64×128 images are represented by isotropic Gaussian weighted color histograms extracted from 8 horizontal stripes. From each stripe, 24-bins histograms have been

extracted from the Hue-Saturation (HS), RGB and Lab channels. Then, histogram of oriented gradients (HOG) quantized in 4 bins, and local binary patterns (LBP) sampled from a grid with cell size equal to 16 pixels, are concatenated to form the final 4842-dimensional feature vector.

To evaluate the performance of the proposed framework we have selected four baseline models, namely KCCA [63], KISSME [46], LADF [11] and Euclidean distance to compute the matches for the initial and final rankings. In our notation, $\mathcal{L}^{A,B}$ and $\tilde{\mathcal{L}}^{A,B}$ define the set of parameters learned by such models trained with input feature vectors pairs in \mathcal{X}_{Tr} and with discriminative feature vectors pairs in $\tilde{\mathcal{X}}_{\text{Tr}}$, respectively.

C. Parameters for Global Information Computation

DCIA relies on several components to discover the global information which is later removed to improve the true match position within the first ranks. In the following, an analysis of each of such components is conducted. The experiments have been carried out to see how much influential to the final results is the selection of the m , k , and K parameters. In addition, experiments separately considering the content and context information contributions have been conducted.

All the experiments have been run using the KCCA baseline model and image pairs from the VIPeR dataset.

Fig. 10 shows two examples of the global information removal effects. The figure shows more insights regarding the benefits of the proposed discriminant information analysis. First row shows a case where the separability between the true match and false ones is improved. Notice that the distance between the true match (green square) and the probe image (red circle) decreases while distances with content set images increase. Images without global information (shown in Fig. 10b) reveal that discriminant information (in this case the red bag) helps to improve the rank of the true match in the post-ranking process (see Fig. 10c). In contrast, second row shows a case where the rank of the true match degrades. Due to a strong perspective variation of the person (front and back view), discriminant information of the probe is missing on the corresponding correct match. Thus, false matches having similar discriminant information with the probe improve their rank after global information removal (see Fig. 10f).

Since DCIA learns a new subspace merging the probe and gallery features, we want to analyze the impact of using features that can be highly inconsistent (e.g., probe and gallery features may have large variations as they come from different cameras). To help with the analysis we used the same process proposed in [17] that, through visual expansion, synthesizes the probe into the gallery feature space, thus reducing feature inconsistency. The experiments presented in Fig. 11 aim to show the importance of including or not the probe in the subspace learning with respect to the visual expansion approach. Results shows that, while visual expansion improves the baseline performance, including the probe is more relevant for DCIA. As a matter of fact, DCIA with probe inclusion has better results than DCIA with visual expansion and without probe inclusion. This behavior is justified by the fact that DCIA wants to find the global appearance shared by both content and context set images and the probe. Such an

²Code is available at <https://github.com/iN1k1>

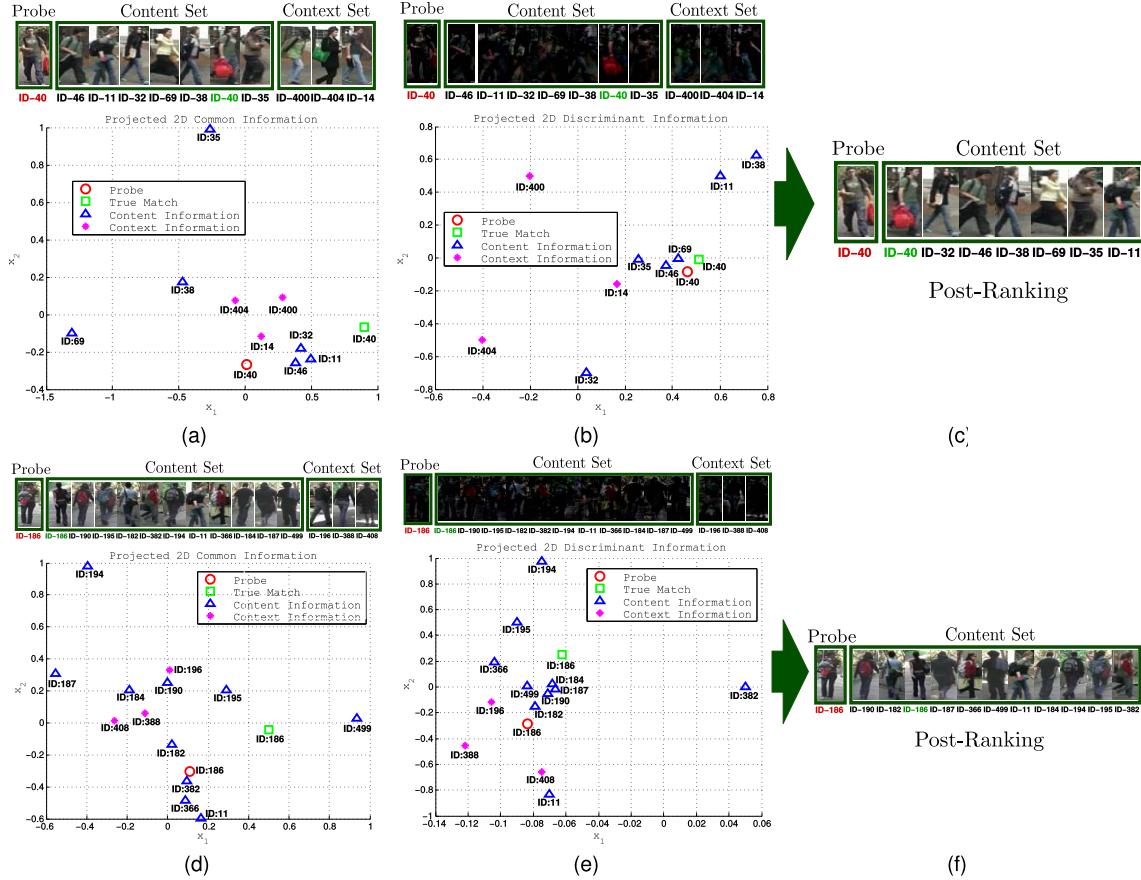


Fig. 10. Global information removal effects. First row represents a case where the separability between the true match and false ones is improved, while second row represents an opposite case. (a) and (d) show the content and context sets and their images projected onto the first two principal components of the original *rgb*-based PCA subspace together with the probe. (b) and (e) show the content and context sets with global information removed and their images projected onto the PCA subspace. (c) and (f) show the qualitative results after post-ranking considering all the *rgb* components. The distances in the charts are just indicative, not describing the effective distance used to compute the final ranking. (*Best viewed in color*).

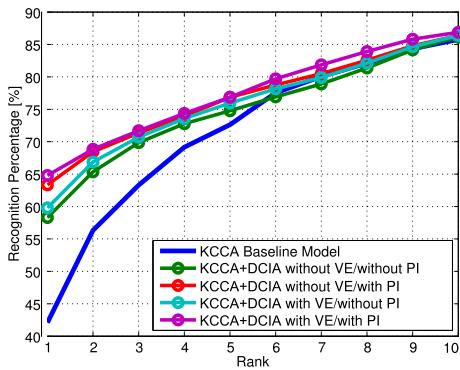


Fig. 11. Analysis of the DCIA performance with respect to probe inclusion in the PCA learning subspace. VE means visual expansion [17] and PI means probe image.

appearance is seen as the information that fools the initial matching, thus it is the information that DCIA aims to remove. Anyhow, the DCIA process slightly benefits from visual expansion, hence it has been used in the next experiments.

In DCIA, only the content set is re-ranked. To select the size of the content set, the m parameter is used. This is dynamically selected on the basis of the shape of dissimilarities distributions shown in Fig. 4. To show that such dynamical

TABLE II
ANALYSIS OF THE KCCA+DCIA RANK-1 PERFORMANCE WITH FIXED AND DYNAMIC SELECTION OF THE CONTENT SET SIZE m . RESULTS ON THE VIPeR DATASET HAVE BEEN COMPUTED FOR DIFFERENT TRAIN/TEST SIZES. BEST RESULTS ARE IN BOLD

Train/Test Size	158/63	158/316	158/474	316/158	316/316
Baseline	48.41	27.69	26.27	53.80	42.08
Dynamic m (average \pm std)	71.02 (3.8 \pm 2.1)	51.64 (6.6 \pm 5.0)	45.05 (5.2 \pm 3.8)	78.06 (7.1 \pm 5.3)	64.78 (6.4 \pm 5.1)
Fixed $m = 3$	67.46	44.30	37.13	67.09	59.49
Fixed $m = 4$	68.25	48.10	39.56	68.67	59.97
Fixed $m = 5$	64.67	33.23	23.20	59.37	50.45
Fixed $m = 7$	59.80	31.44	20.81	53.23	44.27
Fixed $m = 10$	53.63	26.64	19.36	40.00	29.62
Fixed $m = 15$	46.25	21.42	11.68	29.84	22.15
Fixed $m = 20$	33.49	16.23	10.00	21.20	12.47

selection depends on the gallery size and its content we have conducted the following analysis. We have run the experiments with different fixed values and the dynamic selection of m on different train/test sets sizes. The achieved rank-1 performances are shown in Table II. Results demonstrate that, when m is small DCIA generally improves the baseline performance. This is because in the majority of the cases the true match and the content set share the visual ambiguities that are

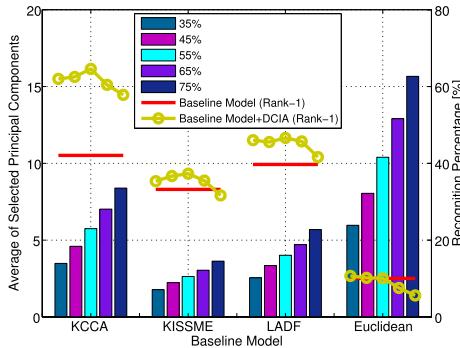


Fig. 12. Performance of the DCIA framework varying the k parameter. The value of k is computed with respect to the specified PCA retained variance. Values of k (bars height) have been computed by averaging the values obtained for the 316 probes in the VIPeR datasets. Baseline and Baseline+DCIA rank-1 performances are shown by the red and yellow lines with respect to the underlying k parameter value.

correctly exploited by DCIA. When larger values of m are used, the global information considers both visual disparities and ambiguities. Thus, DCIA removes the most discriminative information, hence the baseline performance worsens. The dynamic selection approach allows more flexibility in the choice of the content set size as shown by the average size of m and the corresponding standard deviation. This allows a better identification of the images that share relevant visual ambiguities with the probe. The achievement of the best re-identification performances substantiates such a solution.

Finally, the reported rank-1 performances show that the proposed method is not sensitive to the gallery set size. Indeed, even though the rank-1 values vary with the different tests, we can notice that the improvement with respect to the baseline method ranges always around 20%. This means that the baseline method is sensitive to the gallery size, not DCIA.

Fig. 12 shows the performance of the DCIA framework for different values of k and different baseline models. We have selected such values on the basis of the amount of variance retained in representing the common appearance PCA subspace. For each baseline model and for each value of k (height of the bars), the chart includes a comparison on the rank-1 recognition percentage between the baseline model (red line) and its re-ranking with DCIA (yellow line). Results show that, for all the baseline models, DCIA achieves the best results when 55% of the variance is retained. It is interesting to notice that DCIA performances degrade when higher levels of variance are retained. This is due to the fact that, in such cases, not only the visual ambiguities, but also the discriminative details are removed from the original feature vectors. Following such results, in the successive experiments, we have selected the k principal components that account for 55% of the variance when representing the common appearance subspace.

In Fig. 13, we report on the performance achieved by using the context information within the DCIA framework. Recognition percentages for the first 10 ranks are shown for different values of K . Experiment with $K = 0$ corresponds to the performance obtained without the usage of context information. Thus, in the former case only the content information is considered while in the latter all the context information

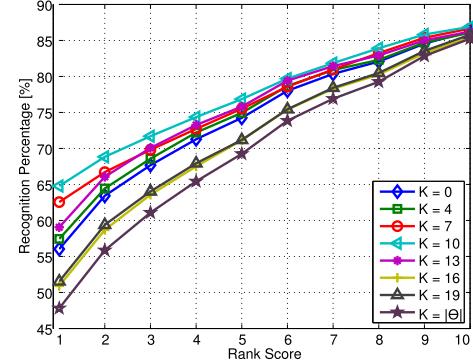


Fig. 13. Performance of the DCIA framework varying the K parameter. $K = 0$ corresponds to the KCCA+DCIA performance obtained without the usage of context information. $K = |\Theta|$ indicates that all gallery images included in the context relevant cluster are considered.

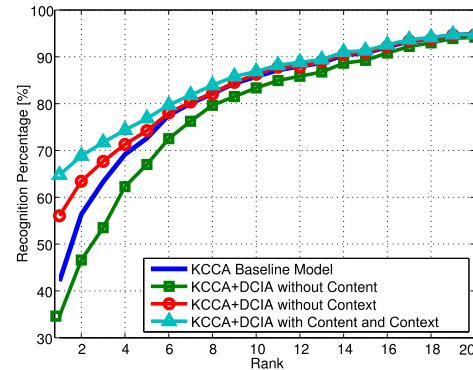


Fig. 14. Performance obtained by separately considering the content and context information. Results have been computed on the VIPeR dataset considering the KCCA baseline.

is exploited to compute the discriminant feature vectors. As shown in Fig. 13, values of $K > 0$ yield to superior performance, thus showing the benefits of the context information. In particular, the best rank 1 recognition percentage is obtained using $K = 10$. However, higher values of K , such as 16 or 19, lead to worse performance. This is more evident when all the context information is considered ($K = |\Theta|$). This degradation in the performance is motivated by the fact that images with dissimilar global appearance are considered in the discriminant information analysis. Thus, not only the visual ambiguities are removed, but also useful information which allows to discriminate between persons having different appearance.

Following such results, we have used $K = 10$ to compute the performance presented in this section. Anyhow, we believe that the K parameter can be dynamically computed by performing histogram analysis. This is demanded to future works.

To conclude, since DCIA to find the global common appearance relies on the content and context information, we have conducted the following analysis to see how much these contribute to the achieved performance. Results in Fig. 14 show the DCIA performances considering or not the content and context information. When the content set is not considered (green line) the visual ambiguities are not exploited thus global information removal yields to deteriorated post-ranking performance. For this experiment, recognition percentages are

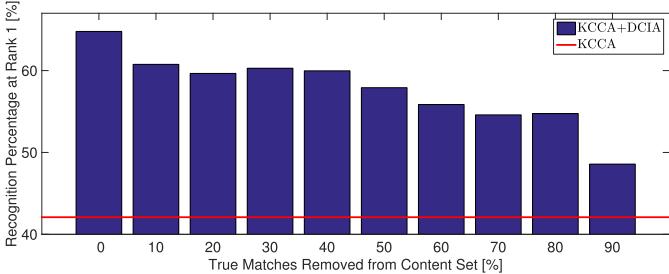


Fig. 15. Rank 1 recognition performance on VIPeR dataset computed with respect to the percentage of true matches randomly removed from the content sets generated during training. The red line shows the KCCA baseline performance.

even lower than the baseline model. Instead, if DCIA considers just the content information (red line), the baseline performance is improved. This is why removing visual ambiguities shared by the probe and the content information improves the discrimination. Such an effect is even more evident when the visual ambiguities between the content and the context sets are also removed (cyan line).

To summarize, results show that global common information is optimally captured when both the content and context sets are considered.

D. Baseline Model Error Tolerability

The performance of the proposed DCIA framework relies on the assumption that the true match is located in the content set. In the re-identification phase, if this is not included in the content set, it is not re-ranked, hence performance remains unaltered. However, if the true match is excluded from the content set in the training phase, then the post-ranking model is trained with many negative samples and few positive ones. Its exclusion from the content set during training can be seen as an “error” made by the baseline model. To see what is the amount of “errors” that DCIA can tolerate we have removed 10, 20, . . . , 80, 90% of the true matches that are included in the content sets generated during training.

The results of such a process are shown in Fig. 15. Depicted performances demonstrate that, up to an error of 40%, the rank 1 recognition percentage is still higher than 60%. A significant degradation is obtained when the error is 90% which corresponds to train the post-training model with a very large portion of negative samples. Such results substantiate the robustness and tolerability of our framework to the presence of a significant portion of “errors” that the initial model may introduce.

E. Analysis of the Approach With Different Baseline Models

To demonstrate that the DCIA framework can be used to improve the first ranks performance of existing methods we have applied our solution on four different baseline models.

1) *Re-Identification Performance*: In Fig. 16, the performance achieved considering the VIPeR, PRID and CUHK02 datasets are shown. Each sub-figure includes eight CMC curves which have been computed by using baseline models alone (dashed lines) and by using them within our

DCIA framework (solid lines). Since only content set images are re-ranked –and those corresponds to first ranks, we provide results for the first 20 ranks only. At higher ranks, comparable results to the considered approaches are achieved. To highlight the performance gain at first three ranks, corresponding CMC performances are separately shown on the right side of each sub-figure.

a) *VIPeR*: In Fig. 16(a), results are computed for the VIPeR dataset. The KCCA baseline model obtains a recognition percentage of 42.08% for rank 1. When used within the DCIA framework, KCCA reaches a recognition percentage of 64.78%, thus increasing the baseline model performance by more than 20%. Similarly, using KISSME, LADF and Euclidean as baseline models, the recognition percentages increase from 33.22% to 37.18%, from 39.71% to 46.67% and from 10.02% to 11.07% for rank 1, respectively. Though the results using DCIA are better for each considered case, the most significant improvement is achieved by considering the KCCA baseline model. In such a case DCIA improves the performance up to the rank 14. Only the first 5 and 4 ranks are improved when the DCIA is applied over KISMME and LADF models. Similar results are obtained when the Euclidean distance is considered.

b) *PRID*: Results on the PRID dataset are shown in Fig. 16(b). The achieved performance demonstrates that DCIA yields to remarkable improvements for all the four baseline models. DCIA leads to a performance gain up to the first 14 ranks for LADF and the first 11 ranks for KCCA, KISSME and Euclidean. In particular, the recognition percentages increase from 15.0% to 32.5% for rank 1 using KCCA as baseline model, from 7.0% to 14.0% using KISSME, from 28.0% to 32.5% using LADF and from 5.5% to 9.5% using Euclidean. This is due to the fact that the PRID dataset has several persons that looks very similar to each other (i.e., it contains lots of visual ambiguities). Thus, more images are included in the content set and correctly re-ranked. This demonstrates that and our method correctly detects the visual ambiguities and is able to identify only the discriminative features which can be used to improve the initial ranking.

c) *CUHK02*: In Fig. 16(c), the performance of our framework have been computed considering the CUHK02 dataset. As for the other two datasets, DCIA improves the initial results achieved by the four baseline models. In particular, the most significant improvement is obtained using KCCA achieving a recognition percentage of 60.0% for rank 1. In such a case, the performance of the first 24 ranks are boosted. For KISSME, LADF and Euclidean, the recognition performances within the first 4, 6 and 7 ranks are improved, respectively.

2) *Discussion*: To provide an overview of the performance improvements achieved using the DCIA framework we have included Fig. 17. It shows the rank 1 performance comparisons for all the baseline methods and the 3 datasets used so far. Results demonstrate that, for every considered case, better recognition performance than the ones obtained by baseline models alone are achieved when the DCIA framework is used. In particular, the best performances are obtained using the KCCA model. Remarkable improvements can be observed for the VIPeR and CUHK02 datasets where the

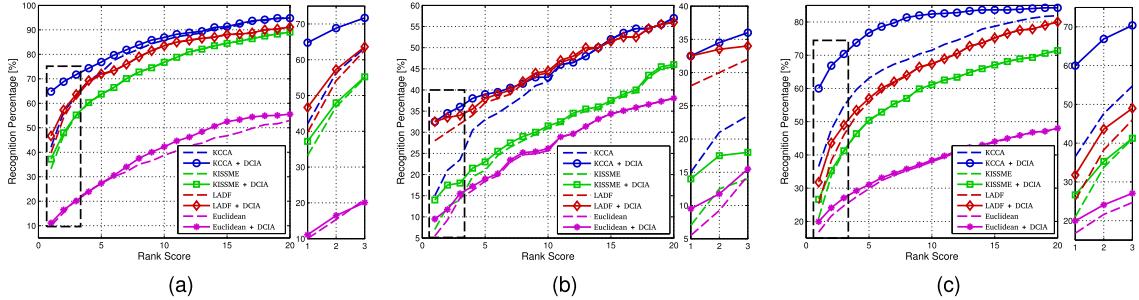


Fig. 16. CMC performance achieved using four different baseline models on: (a) VIPeR, (b) PRID and (c) CUHK02 datasets. Dashed lines show the performance of the baseline models. Solid lines have been computed by using such baseline models within the proposed DCIA framework.

TABLE III

RE-RANKING ANALYSIS COMPUTED FOR THE VIPeR DATASET USING THE FOUR BASELINE MODELS. RESULTS HAVE BEEN REPORTED AS AVERAGE VALUES AND STANDARD DEVIATIONS COMPUTED OVER THE 10 CONSIDERED TRIALS. AMONG ALL THE RE-RANKED PROBES (SECOND COLUMN), THE NUMBER OF IMPROVED OR DETERIORATED (CORRECT MATCH MOVES TO THE FIRST RANKS OR THE OPPOSITE) ONES ARE SHOWN IN THE LAST TWO COLUMNS. THIRD AND FOURTH COLUMNS SHOW THE AVERAGE AND THE MAXIMUM CONTENT SET SIZE (THIS IS EXPERIMENTALLY UPPER BOUNDED AT 25)

Baseline Model	Re-Ranked Probes	Content Set		Rankings	
		Average	Maximum	Improved	Deteriorated
KCCA	272.8 \pm 3.90[316]=86.33% \pm 1.23%	6.8 \pm 5.1	25.0 \pm 0.0	111.4 \pm 8.23	13.4 \pm 4.93
KISSME	104.4 \pm 11.72[316]=33.04% \pm 3.71%	2.5 \pm 0.9	6.8 \pm 1.1	21.2 \pm 6.38	0.8 \pm 1.30
LADF	103.0 \pm 30.02[316]=32.59% \pm 9.50%	2.8 \pm 1.2	8.4 \pm 1.1	26.4 \pm 7.46	3.6 \pm 3.29
Euclidean	303.0 \pm 1.58[316]=95.89% \pm 0.50%	15.2 \pm 8.0	25.0 \pm 0.0	64.0 \pm 4.69	62.2 \pm 5.72

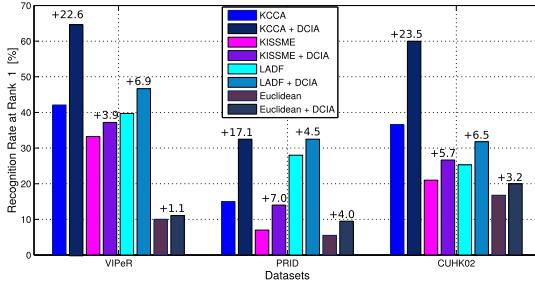


Fig. 17. Rank 1 recognition percentages obtained by the four baseline models used alone and within the DCIA framework. Results are provided for all the 3 considered datasets. The improvement achieved by using the DCIA framework is shown on the top of each bar.

KCCA baseline model recognition performances increase by 22.6% and 23.5%, respectively. A similar gain is obtained for all the baseline models on the PRID dataset.

Table III shows a more detailed analysis of the re-identification performance achieved for the VIPeR dataset. Results demonstrate that for all baseline models, the true match is correctly included in the content set, hence considerer for re-ranking, for at least 30% of the times. The dimension of the content set with respect to the size of the test set is less than 10%, which means that only first ranks are considered, i.e., those that share visual ambiguities. KCCA+DCIA yields to the largest number of improved rankings. Specifically, more than 111 probes have been correctly moved to first ranks, while only 13 have been wrongly moved to higher ones. This phenomena is mainly driven by the fact that KCCA brings the correct match into rank 1 for the 99.9% of the

training data. The other considered methods do not have such a behavior and the 100% recognition rate is usually achieved at higher ranks, namely at about rank 170 for KISSME, rank 95 for LADF and rank 280 using the Euclidean metric. As a consequence, the KCCA post-training model, i.e., DCIA, is always applied on training data which has images that include the true match as well as false matches that all together share visual ambiguities. For the other methods many probes do not have the correct match in the corresponding post-training gallery set. Thus, the post-training model exploits a large set of negative samples which limits its generalization performance.

During test, for KCCA, many correct matches (i.e. about 70% rank-5) are included in the content set. All of these, for the post-ranking phase, are in a similar situation to the one occurred during training. Thus, we have that 65% out of 70% reaches rank 1. For the other baseline methods, these similar situations do not take place. Such considerations are also substantiated by the results shown in Fig.15 where we have analyzed the behavior of KCCA+DCIA by limiting the number of true matches included in the content set generated during the training.

To conclude, it is worth noticing that, for every considered case, the ranking improvement is higher than its degradation, thus showing that initial results are always improved.

A visual example of improved and deteriorated rankings is shown in Fig. 18. Results have been computed for CUHK02 dataset. For each probe, only the images in the content set are shown. In particular, first row corresponds to the initial ranking computed using the KCCA model

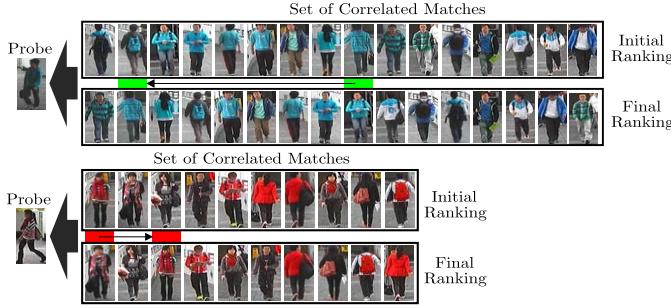


Fig. 18. Examples of improved (green box) and deteriorated (red box) rankings obtained using the proposed algorithm. Results have been computed for CUHK02 dataset. The location of the rectangle indicates the true match. (*Best viewed in color*).

while the second one shows the final ranking obtained by DCIA+KCCA. DCIA improves the true match location if there exists discriminant information in the content set.

To summarize, results achieved by the proposed DCIA framework (applied on four baseline methods) have demonstrated that remarkable recognition improvements can be obtained for first ranks. In particular, for every considered dataset and each baseline method, the first rank recognition performances are always improved. This shows that the DCIA framework can be generally applied to an existing method to improve its results, thus supporting our assumption that every initial ranking carries valuable information that can be exploited to improve the initial performance.

F. Comparison With State-of-the-Art Methods

In this section we compare the results achieved by the DCIA framework with state-of-the-arts methods on the VIPeR, PRID and CUHK02 datasets.

In Table IV, comparisons with 24 state-of-the-art methods on VIPeR dataset are given. Results are shown in terms of recognition percentage for a relevant subset of all possible ranks. These demonstrate that both LADF+DCIA as well as KCCA+DCIA outperform the existing methods. In particular, KCCA+DCIA improves the best rank 1 results obtained by almost 20%.

Similarly, in Table V, results on PRID dataset are given in terms of recognition percentages for six most representative rank scores. Reported performances demonstrate that our solution outperforms existing approaches. In particular, KCCA+DCIA and LADF+DCIA achieve a correct recognition rate of 32.5% at rank 1, thus increasing the previous state-of-the-art performance by 4.5%.

In Table VI a comparison between our solution and state-of-the-art methods the on CUHK02 dataset is given. Results, provided for six relevant ranks, show that KCCA+DCIA improves the rank 1 previous state-of-the-art performance by more than 23%. More interestingly, all other DCIA solutions perform poorly than the KCCA baseline model alone. Despite this the DCIA improvement at rank 1 is never lower than 3%.

G. Comparison With Post-Ranking Methods

In Table VII, we report on the benefits of the DCIA framework with respect to other post-ranking approaches.

TABLE IV
COMPARISON WITH STATE-OF-THE-ART METHODS ON THE VIPeR DATASET. RESULTS ARE SHOWN AS CORRECT RECOGNITION RATES (IN PERCENTAGE) FOR A SUBSET OF RELEVANT RANKS. BEST RESULTS ARE IN BOLDFACE FONT

Rank →	1	5	10	20
Euclidean	10.02	27.27	38.75	53.00
Euclidean+DCIA	11.07	27.37	42.24	55.53
LF [64]	24.20	52.20	67.10	82.30
RS-KISS [47]	24.50	-	66.60	81.70
WFS [13]	25.81	-	69.56	83.67
eSDC.knn [25]	26.31	-	58.86	72.77
eSDC.ocsvm [25]	26.74	-	62.37	76.36
PatMatch [26]	26.90	-	62.34	75.63
MtMCML [15]	28.83	-	75.82	88.51
LMF [34]	29.10	52.90	66.30	81.00
LAFT [12]	29.60	-	69.30	81.34
RCCA [24]	30.00	-	75.00	87.00
SalMatch [26]	30.16	-	65.54	79.15
QALF [65]	30.17	51.60	62.44	73.81
kBiCov [22]	31.11	-	70.71	82.44
SWF [42]	32.97	-	75.63	86.87
KISSME [46]	33.22	63.60	76.74	89.08
IDLA [35]	34.81	-	-	-
PKFM [66]	36.80	-	83.70	91.70
KISSME+DCIA	37.18	63.76	76.74	89.08
LADF [11]	39.71	71.67	83.54	91.13
LOMO+XQDA [67]	40.00	-	80.51	91.08
KCCA [63]	42.08	72.62	85.75	94.77
KEPLER [68]	42.41	-	82.37	90.70
LMF [34]+LADF [11]	43.40	73.10	85.00	94.40
LTR-ME [44]	45.90	77.50	88.90	95.80
LADF+DCIA	46.67	71.99	83.54	91.13
KCCA+DCIA	64.78	76.85	86.88	94.77

TABLE V
COMPARISON WITH STATE-OF-THE-ART METHODS ON THE PRID DATASET (INCLUDING DISTRACTORS). RESULTS ARE SHOWN AS CORRECT RECOGNITION RATES (IN PERCENTAGE) FOR A SUBSET OF RELEVANT RANKS. BEST RESULTS ARE IN BOLDFACE FONT

Rank →	1	5	10	20	50	100
Euclidean	5.5	18.5	25.6	38.0	56.0	68.0
DDC [52]	4.0	18.0	24.0	37.0	56.0	70.0
KISSME [46]	7.0	21.0	31.0	45.5	68.5	84.0
DAM [69]	8.0	-	30.0	41.0	59.0	75.0
Euclidean+DCIA	9.5	19.0	26.5	38.0	56.0	68.0
KISSME+DCIA	14.0	23.0	31.5	46.0	68.5	84.0
EIML [70]	15.0	30.0	38.0	50.0	67.0	80.0
RPLM [41]	15.0	31.0	42.0	54.0	70.0	80.0
KCCA [63]	15.0	33.0	42.0	57.0	74.5	85.0
LADF [11]	28.0	37.0	44.0	56.0	74.5	84.0
LADF+DCIA	32.5	38.0	44.5	56.0	74.5	84.0
KCCA+DCIA	32.5	39.0	43.0	57.0	74.5	86.0

Recognition performances achieved on the VIPeR dataset are shown for five relevant ranks. Since the majority of the re-ranking methods provide results on such dataset only, we have not considered all the other ones used so far. Notice that, RIRO [59], IRT [57] an POP [17] post-ranking techniques have an iterative operation in which the user participates by manually selecting some samples from the initial ranking. Thus, to make a fair comparison, we considered the

TABLE VI

COMPARISON WITH STATE-OF-THE-ART METHODS ON THE CUHK02 DATASET. RESULTS ARE SHOWN AS CORRECT RECOGNITION RATES (IN PERCENTAGE) FOR A SUBSET OF RELEVANT RANKS. BEST RESULTS ARE IN BOLDFACE FONT

Rank →	1	5	10	20	50	100
CCA [71]	5.2	15.7	22.0	29.9	47.2	63.1
ITML [72]	9.6	22.2	28.0	41.6	62.5	82.5
LDM [73]	11.3	30.5	38.6	52.5	70.1	82.2
mLMNN [74]	14.2	31.3	38.4	52.0	72.1	87.7
Euclidean	16.8	29.8	37.9	47.9	63.7	77.8
Euclidean+DCIA	19.9	31.1	38.4	48.0	63.7	77.8
KISSME [46]	21.0	50.4	61.1	71.4	86.6	95.0
LADF [11]	25.3	55.9	67.4	80.0	90.3	96.2
LAFTV [12]	25.8	53.6	64.9	78.1	92.8	98.7
KISSME+DCIA	26.6	50.4	61.2	73.0	86.6	95.0
SWF [42]	31.0	57.3	68.7	79.5	91.2	96.9
LADF+DCIA	31.7	57.0	67.4	80.0	90.3	96.2
KCCA [63]	36.5	63.0	71.5	81.9	91.7	96.7
KCCA+DCIA	60.0	76.7	82.4	84.3	91.7	96.7

TABLE VII

COMPARISON WITH RE-RANKING METHODS ON THE VIPeR DATASET. RESULTS ARE SHOWN AS RECOGNITION PERCENTAGES FOR SOME RELEVANT RANKS. BEST RESULTS ARE IN BOLDFACE FONT

Rank →	1	5	10	25	50
Euclidean+DCIA	11.07	27.37	42.24	57.59	72.94
DDC [52]	19	-	52	69	80
ISR [10]	27	50	61	76	86
RIRO [59] (1 Iteration)	28	30	34	51	64
PRRS [56]	33.29	-	78.35	-	97.53
KISSME+SB [53]	34.01	64.10	78.97	92.16	98.16
KISSME+CCRR [55]	36.25	65.25	80.04	93.01	98.40
KISSME+DCIA	37.18	63.76	76.74	91.93	97.62
IRT [57] (1 Iteration)	43	45	46	53	61
LADF+DCIA	46.67	72.00	83.54	93.03	98.41
POP [17] (1 Iteration)	59.05	60.95	63.10	72.20	-
KCCA+DCIA	64.78	76.85	86.88	96.36	99.05

performance obtained using a single iteration. Results show that KCCA+DCIA outperforms all existing methods. In particular, it achieves a rank 1 recognition percentage of 64.78%, thus improving the best result (59.05%) obtained by POP [17] (1 iteration) by more than 5%.

SB [53] and CCRR [55] provide the performance achieved using the KISSME baseline model. However, in their work they used a different feature representation. To provide a fair comparison, we have implemented the SB and CCRR re-ranking methods and applied them on the same KISSME baseline model used by our DCIA framework. Results show that, starting from the same baseline method with the same features, SB and CCRR improve the rank 1 performance from 33.22% to 34.01% and from 33.22% to 36.25%, respectively. For the same rank 1, DCIA improves the performance from 33.22% up to 37.18%, thus obtaining a gain of about 4%. This shows that DCIA is better than existing methods when applied to the same baseline model.

H. Computational Performance Analysis

To conclude the experimental section, we have included an analysis of the time complexity of the proposed DCIA

framework when applied to the VIPeR dataset. The following results have been obtained by running an optimized code on a High-Performance Computing Cluster.

Results presented in Table VIII show that exploitation of the DCIA framework requires considerable additional training time. Although this appears to be a demanding approach, we can notice that the majority of the time is used to compute the visual expansion. Considering the results in Fig.11, which show that visual expansion improves the performance by less than 1%, we can neglect this step in favor of a much faster solution.

Despite this, the re-identification can be completed in less than 35ms for all the baseline methods.

While the proposed framework requires additional training time, the rank 1 improvement is considerable.

I. Discussion and Future Directions

The proposed DCIA framework represents one of the first efforts in improving re-identification performance by means of re-ranking. More specifically, our solution aims to identify and remove the global information shared between the first ranks such that the true match location can be improved.

With an in-depth evaluation conducted on three datasets, we have demonstrated that, whenever the re-identification task relies on a metric computed in the feature space, DCIA applied on the top of any baseline method shows relevant performance improvement. Since this characteristic is independent from the specific task the proposed approach can be further investigated and applied to other application domains. These could be historical ones like biometrics (e.g., face/finger/iris recognition), object recognition, information retrieval and new domains like food recognition.

While the DCIA advantages have been deeply discussed, it still suffers from few limitations.

- 1) A wrong selection of the two algorithm parameters, namely the context set size K and the number of principal components k , yields to limited re-identification improvements.
- 2) To identify the global information we have introduced the concepts of content and context sets. The content set is defined by those images which have low dissimilarity values with a probe and determines the person identities which are finally re-ranked. Due to this, DCIA can improve initial ranking performance only if the baseline method assigns the true match to the first ranks.

These limitations open to interesting future works to:

- 1) Select the two algorithm parameters relying on an optimization problem;
- 2) Improve re-identification performance when the true match is not included in the content set.

Such a last point may have interesting side effects in detecting if the true match is not present in the gallery at all. In this case, we could rely on the analysis of the DCIA behavior when the true match is included or not in the content set. If a different

TABLE VIII

ANALYSIS OF THE COMPUTATIONAL PERFORMANCE ACHIEVED BY APPLYING THE DCIA FRAMEWORK ON THE FOUR CONSIDERED BASELINE METHODS. TRAINING TIME IN BRACKETS CORRESPONDS TO VISUAL EXPANSION (VE) COMPUTATION. RESULTS HAVE BEEN COMPUTED FOR THE VIPeR DATASET

Baseline				Baseline+DCIA		
Model	Training Time	Test Time to re-identify a single person	Rank-1	Training Time Without (With) VE	Test Time to re-identify a single person	Rank-1
KCCA	38.9s	3.3ms	42.08%	788s (3.42h)	22.17ms	64.78%
KISSME	5.8s	3.2ms	33.22%	339s (3.42h)	29.57ms	37.18%
LADF	139.7s	3.4ms	39.71%	1223s (3.42h)	33.74ms	46.67%
Euclidean	—	3.5ms	10.02%	—	7.03ms	11.07%

behavior exists, then it can be exploited for the aforementioned purpose.

VI. CONCLUSION

In this work we have proposed a novel post-ranking framework for person re-identification. We have introduced an unsupervised approach that exploits the visual ambiguities shared between first ranked persons. Two sets of gallery images are extracted from the initial ranking to model the content and context information. This is exploited by the discriminant information analysis which transforms the original feature vectors by removing the common information, thus defining the discriminant feature vector space. This is later exploited by a learning algorithm to discriminate between positive and negative image pairs. First rank matches are then re-ranked on the basis of the output of such a learning algorithm. Extensive evaluation performance have been conducted using three public benchmark datasets. Results demonstrated that, for every considered dataset, baseline models performance are always improved. This strongly supports our belief, i.e., that the initial ranking includes relevant information that can be used to improve first rank performance. Finally, comparisons with state-of-the-art methods have shown that rank 1 performance have been improved by more than 22% on two very challenging datasets.

REFERENCES

- [1] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multicamera people tracking with a probabilistic occupancy map," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 267–282, Feb. 2008.
- [2] S. M. Khan and M. Shah, "Tracking multiple occluding people by localizing on multiple scene planes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 505–519, Mar. 2009.
- [3] C. Alcaraz and J. Lopez, "Wide-area situational awareness for critical infrastructure protection," *Computer*, vol. 46, no. 4, pp. 30–37, Apr. 2013.
- [4] A. K. Roy-Chowdhury, M. Kankanhalli, J. Konrad, C. Micheloni, and P. Varshney, "Introduction to the issue on signal processing for situational awareness from networked sensors and social media," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 2, pp. 201–203, Mar. 2015.
- [5] E. Ricci, G. Zen, N. Sebe, and S. Messelodi, "A prototype learning framework using EMD: Application to complex scenes analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 513–526, Mar. 2012.
- [6] B. Solmaz, B. E. Moore, and M. Shah, "Identifying behaviors in crowd scenes using stability analysis for dynamical systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 2064–2070, Oct. 2012.
- [7] S. Kwak, B. Han, and J. H. Han, "On-line video event detection by constraint flow," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1174–1186, Jun. 2014.
- [8] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," in *Proc. CVPR*, Jun. 2014, pp. 152–159.
- [9] Z. Wu, Y. Li, and R. J. Radke, "Viewpoint invariant human re-identification in camera networks using pose priors and subject-discriminative features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 5, pp. 1095–1108, May 2015.
- [10] G. Lisanti, I. Masi, A. D. Bagdanov, and A. Del Bimbo, "Person re-identification by iterative re-weighted sparse ranking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1629–1642, Aug. 2014.
- [11] Z. Li, S. Chang, F. Liang, T. Huang, L. Cao, and J. R. Smith, "Learning locally-adaptive decision functions for person verification," in *Proc. CVPR*, Jun. 2013, pp. 3610–3617.
- [12] W. Li and X. Wang, "Locally aligned feature transforms across views," in *Proc. CVPR*, Jun. 2013, pp. 3594–3601.
- [13] N. Martinel, A. Das, C. Micheloni, and A. K. Roy-Chowdhury, "Re-identification in the function space of feature warps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1656–1669, Aug. 2005.
- [14] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 653–668, Mar. 2013.
- [15] L. Ma, X. Yang, and D. Tao, "Person re-identification over camera networks using multi-task distance metric learning," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3656–3670, Aug. 2014.
- [16] D. Tao, L. Jin, Y. Wang, and X. Li, "Person reidentification by minimum classification error-based KISS metric learning," *IEEE Trans. Cybern.*, vol. 45, no. 2, pp. 242–252, Feb. 2014.
- [17] C. Liu, C. C. Loy, S. Gong, and G. Wang, "POP: Person re-identification post-rank optimisation," in *Proc. ICCV*, Dec. 2013, pp. 441–448.
- [18] R. Vezzani, D. Baltieri, and R. Cucchiara, "People reidentification in surveillance and forensics: A survey," *ACM Comput. Surveys*, vol. 46, no. 2, Nov. 2014, Art. no. 29.
- [19] A. Bedagkar-Gala and S. K. Shah, "A survey of approaches and trends in person re-identification," *Image Vis. Comput.*, vol. 32, no. 4, pp. 270–286, Apr. 2014.
- [20] L. Bazzani, M. Cristani, and V. Murino, "Symmetry-driven accumulation of local features for human characterization and re-identification," *Comput. Vis. Image Understand.*, vol. 117, no. 2, pp. 130–144, 2013.
- [21] S. Bak, E. Corvée, F. Brémond, and M. Thonnat, "Boosted human re-identification using Riemannian manifolds," *Image Vis. Comput.*, vol. 30, nos. 6–7, pp. 443–452, Jun. 2012.
- [22] B. Ma, Y. Su, and F. Jurie, "Covariance descriptor based on bio-inspired features for person re-identification and face verification," *Image Vis. Comput.*, vol. 32, nos. 6–7, pp. 379–390, 2014.
- [23] B. Ma, Q. Li, and H. Chang, "Gaussian descriptor based on local features for person re-identification," in *Proc. ACCV*, 2014, pp. 1–14.
- [24] L. An, M. Kafai, S. Yang, and B. Bhanu, "Reference-based person re-identification," in *Proc. AVSS*, Aug. 2013, pp. 244–249.
- [25] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proc. CVPR*, 2013, pp. 3586–3593.
- [26] R. Zhao, W. Ouyang, and X. Wang, "Person re-identification by salience matching," in *Proc. ICCV*, Dec. 2013, pp. 2528–2535.
- [27] H. Wang, S. Gong, and T. Xiang, "Unsupervised learning of generative topic saliency for person re-identification," in *Proc. BMVC*, 2014, pp. 1–11.
- [28] J. Roth and X. Liu, "On the exploration of joint attribute learning for person re-identification," in *Proc. ACCV*, 2014, pp. 673–688.

- [29] S. Khamis, C.-H. Kuo, V. K. Singh, V. D. Shet, and L. S. Davis, "Joint learning for attribute-consistent person re-identification," in *Proc. ECCVW*, 2014, pp. 134–146.
- [30] R. Layne, T. M. Hospedales, and S. Gong, "Re-id: Hunting attributes in the wild," in *Proc. BMVC*, 2014, pp. 1.1–1.12.
- [31] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. ICCV*, Dec. 2015, pp. 1116–1124.
- [32] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, and J. Bu, "Semi-supervised coupled dictionary learning for person re-identification," in *Proc. CVPR*, 2014, pp. 3550–3557.
- [33] Y. Wu, M. Minoh, M. Mukunoki, W. Li, and S. Lao, "Collaborative sparse approximation for multiple-shot across-camera person re-identification," in *Proc. AVSS*, Sep. 2012, pp. 209–214.
- [34] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," in *Proc. CVPR*, Jun. 2014, pp. 144–151.
- [35] E. Ahmed, M. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in *Proc. CVPR*, Jun. 2015, pp. 3908–3916.
- [36] T. Wang, S. Gong, X. Zhu, and S. Wang, "Person re-identification by video ranking," in *Proc. ECCV*, vol. 8692, 2014, pp. 688–703.
- [37] 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. 2, pp. 146–162, Feb. 2008.
- [38] J. García, N. Martinel, A. Gardel, I. Bravo, G. L. Foresti, and C. Micheloni, "Modeling feature distances by orientation driven classifiers for person re-identification" *J. Vis. Commun. Image Represent.*, vol. 38, pp. 115–129, Jul. 2016.
- [39] A. J. Ma, P. C. Yuen, and J. Li, "Domain transfer support vector ranking for person re-identification without target camera label information," in *Proc. ICCV*, Dec. 2013, pp. 3567–3574.
- [40] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja, "Pedestrian recognition with a learned metric," in *Proc. ACCV*, 2010, pp. 501–512.
- [41] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof, "Relaxed pairwise learned metric for person re-identification," in *Proc. ECCV*, 2012, pp. 780–793.
- [42] N. Martinel, C. Micheloni, and G. L. Foresti, "Saliency weighted features for person re-identification," in *Proc. ECCVW*, 2014, pp. 191–208.
- [43] F. Xiong, M. Gou, O. Camps, and M. Sznaier, "Person re-identification using kernel-based metric learning methods," in *Proc. ECCV*, 2014, pp. 1–16.
- [44] S. Paisitkriangkrai, C. Shen, and A. van den Hengel, "Learning to rank in person re-identification with metric ensembles," in *Proc. CVPR*, 2015, pp. 1846–1855.
- [45] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in *Proc. ACCV*, 2012, pp. 31–44.
- [46] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. CVPR*, Jun. 2012, pp. 2288–2295.
- [47] D. Tao, L. Jin, Y. Wang, Y. Yuan, and X. Li, "Person re-identification by regularized smoothing KISS metric learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 10, pp. 1675–1685, Oct. 2013.
- [48] G. Zhang, Y. Wang, J. Kato, T. Marutani, and K. Mase, "Local distance comparison for multiple-shot people re-identification," in *Proc. ACCV*, 2013, pp. 677–690.
- [49] A. Mignon and F. Jurie, "PCCA: A new approach for distance learning from sparse pairwise constraints," in *Proc. CVPR*, Jun. 2012, pp. 2666–2672.
- [50] N. Martinel, A. Das, C. Micheloni, and A. K. Roy-Chowdhury, "Temporal model adaptation for person re-identification," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 858–877.
- [51] N. Martinel, G. L. Foresti, and C. Micheloni, "Person reidentification in a distributed camera network framework," *IEEE Trans. Cybern.*, to be published.
- [52] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Proc. 17th Scandinavian Conf. Image Analysis*, vol. 6688, 2011, pp. 91–102.
- [53] L. An, X. Chen, M. Kafai, S. Yang, and B. Bhanu, "Improving person re-identification by soft biometrics based reranking," in *Proc. ICDSC*, Oct./Nov. 2013, pp. 1–6.
- [54] V.-H. Nguyen, T. D. Ngo, K. M. T. T. Nguyen, D. A. Duong, K. Nguyen, and D.-D. Le, "Re-ranking for person re-identification," in *Proc. SoCPaR*, 2013, pp. 304–308.
- [55] Q. Leng, R. Hu, C. Liang, Y. Wang, and J. Chen, "Person re-identification with content and context re-ranking," *Multimedia Tools Appl.*, vol. 74, no. 17, pp. 6989–7014, Sep. 2015.
- [56] L. An, M. Kafai, S. Yang, and B. Bhanu, "Person reidentification with reference descriptor," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 4, pp. 776–787, Apr. 2016.
- [57] S. Ali, O. Javed, N. Haering, and T. Kanade, "Interactive retrieval of targets for wide area surveillance," in *Proc. 18th ACM Int. Conf. Multimedia (MM)*, 2010, pp. 895–898.
- [58] A. Parkash and D. Parikh, "Attributes for classifier feedback," in *Proc. ECCV*, vol. 7574, 2012, pp. 354–368.
- [59] Z. Wang, R. Hu, C. Liang, Q. Leng, and K. Sun, "Region-based interactive ranking optimization for person re-identification," in *Proc. Adv. Multimedia Inf. Process.*, 2014, pp. 1–10.
- [60] J. García, N. Martinel, C. Micheloni, and A. Gardel, "Person re-identification ranking optimisation by discriminant context information analysis," in *Proc. ICCV*, Dec. 2015, pp. 1305–1313.
- [61] D. C. G. Pedronette and R. da S. Torres, "Image re-ranking and rank aggregation based on similarity of ranked lists," *Pattern Recognit.*, vol. 46, no. 8, pp. 2350–2360, Aug. 2013.
- [62] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. ECCV*, 2008, pp. 262–275.
- [63] G. Lisanti, I. Masi, and A. Del Bimbo, "Matching people across camera views using kernel canonical correlation analysis," in *Proc. ICDSC*, 2014, Art. no. 10.
- [64] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, "Local Fisher discriminant analysis for pedestrian re-identification," in *Proc. CVPR*, Jun. 2013, pp. 3318–3325.
- [65] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian, "Query-adaptive late fusion for image search and person re-identification," in *Proc. CVPR*, 2015, pp. 1741–1750.
- [66] D. Chen, Z. Yuan, G. Hua, N. Zheng, and J. Wang, "Similarity learning on an explicit polynomial kernel feature map for person re-identification," in *Proc. CVPR*, Jun. 2015, pp. 1565–1573.
- [67] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. CVPR*, Jun. 2015, pp. 2197–2206.
- [68] N. Martinel, C. Micheloni, and G. L. Foresti, "Kernelized saliency-based person re-identification through multiple metric learning," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5645–5658, Dec. 2015.
- [69] M. Hirzer, C. Beleznai, M. Köstinger, P. M. Roth, and H. Bischof, "Dense appearance modeling and efficient learning of camera transitions for person re-identification," in *Proc. ICIP*, Sep. 2012, pp. 1617–1620.
- [70] M. Hirzer, P. M. Roth, and H. Bischof, "Person re-identification by efficient impostor-based metric learning," in *Proc. AVSS*, Sep. 2012, pp. 203–208.
- [71] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, nos. 3–4, pp. 321–377, 1936.
- [72] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *Proc. ICML*, 2007, pp. 209–216.
- [73] L. Yang, R. Jin, R. Sukthankar, and Y. Liu, "An efficient algorithm for local distance metric learning," in *Proc. AAAI*, 2006, pp. 543–548.
- [74] K. Q. Weinberger and L. K. Saul, "Fast solvers and efficient implementations for distance metric learning," in *Proc. ICML*, 2008, pp. 1160–1167.



Jorge García received the B.S. degree in telecommunications engineering, the M.Sc. degree in electronics system engineering the University of Alcalá, Madrid, Spain, in 2009 and 2011, where he is currently pursuing the Ph.D. degree with the Electronics Department. He is currently with the Electronics Department, University of Alcalá, since 2009. His research interests include computer vision, video surveillance applications, scene understanding, and embedded FPGA systems.



Member of the IAPR.

Niki Martinel (S'12) received the M.Sc. (Hons.) in multimedia communications from the University of Udine, Italy, where he is currently pursuing the Ph.D. degree in information engineering with the AVIRES Laboratory, Department of Mathematics and Computer Science. He is currently a member of the AVIRES Laboratory, Department of Mathematics and Computer Science, University of Udine. His research interests include machine (deep) learning, wide area scene analysis, feature transformations, and human-computer interaction. He is Student



Gian Luca Foresti (SM'01) received the Laurea degree (*cum laude*) in electronic engineering and the Ph.D. degree in computer science from the University of Genoa, Italy, in 1990 and 1994, respectively. Since 1998, he has been a Professor of Computer Science with the Department of Mathematics and Computer Science, University of Udine, and the Director of the Artificial Vision and Real-Time System (AVIRES) Laboratory. He is a fellow member of the IAPR.



Alfredo Gardel received degrees in telecommunication engineering from the Polytechnic University of Madrid, Spain, in 1999, and a Ph.D. in telecommunication from the University of Alcalá, Spain, in 2004. Since 1997, he has been a Lecturer with the Electronics Department, University of Alcalá. His main areas of research comprise infrared and computer vision, monocular metrology, robotics sensorial systems, and design of advanced digital systems.



Ignacio Bravo received the B.Sc. degree in telecommunications engineering, the M.Sc. degree in Electronics engineering, and the Ph.D. degree in electronics from the University of Alcalá, Madrid, Spain, in 1997, 2000, and 2007, respectively. Since 2002, he has been a Lecturer with the Electronics Department, University of Alcalá, where he is currently an Associate Professor. His areas of research are reconfigurable hardware, vision architectures based on FPGAs, and Electronic Design.



Christian Micheloni received the M.Sc. degree in 2002 and the Ph.D. degree in 2006. He is currently an Associate Professor with the University of Udine. His main interests involve active vision for the wide area scene understanding and resource aware camera networks. He is also interested in pattern recognition and machine learning. His current research projects include camera network self-reconfiguration and person reidentification. He is member of the IAPR.