Person Re-Identification Ranking Optimisation by Discriminant Context Information Analysis

Jorge García[1], Niki Martinel[2], Christian Micheloni[2] and Alfredo Gardel[1]
[1]Department of Electronics, University of Alcala, Alcalá de Henares, Spain
[2]Department of Mathematics and Computer Science, University of Udine, Udine, Italy
jorge.garcia@depeca.uah.es, niki.martinel@uniud.it, christian.micheloni@uniud.it,
alfredo@depeca.uah.es

Abstract

Person re-identification is an open and challenging problem in computer vision. Existing re-identification approaches focus on optimal methods for features matching (e.g., metric learning approaches) or study the inter-camera transformations of such features. These methods hardly ever pay attention to the problem of visual ambiguities shared between the first ranks. In this paper, we focus on such a problem and introduce an unsupervised ranking optimization approach based on discriminant context information analysis. The proposed approach refines a given initial ranking by removing the visual ambiguities common to first ranks. This is achieved by analyzing their content and context information. Extensive experiments on three publicly available benchmark datasets and different baseline methods have been conducted. Results demonstrate a remarkable improvement in the first positions of the ranking. Regardless of the selected dataset, state-of-the-art methods are strongly outperformed by our method.

1. Introduction

Person re-identification is the problem of re-associating a same person moving between the disjoint Fields-of-View of a wide area camera network. Due to the inherent challenges present in a multi-camera setting, the person re-identification is still an open problem. In particular, when a person is sensed by the different viewpoints of disjoint cameras, his/her appearance undergoes significant illumination and color variations as well as pose changes. The non-rigid shape of the human body, as well as background clutter, introduce additional challenges.

In the recent past, the research community endeavored to overcome the aforementioned issues by proposing different methods based on: (i) discriminative signatures exploiting multiple local and global features [39, 22, 25, 24] to compute the persons representations. These representations have been combined with reference sets [3], patch matching strategies [42, 28], saliency learning [36] and joint attributes [16]; (ii) feature transformations addressing the re-identification problem by finding the transformation functions that affect the visual features acquired by disjoint cameras [44, 27]. In [14], a unique brightness transfer function (BTF) computed between features was used to match persons across camera pairs. Recent works [26, 44, 7] also considered that the transformation is not unique and it depends on several factors; (iii) metric learning where approaches still rely on particular features but also advantage of a training phase to learn distances used to compute the match in a different feature space [26, 35]. In [6], 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 [12], by consider-

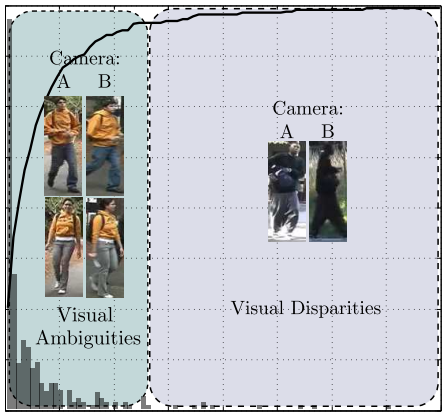


Figure 1: Typical cumulative matching characteristics (CMC) curve achieved by re-identification approaches. In the background, a bar-chart indicates the number of true matches for each rank. In the first ranks the matched persons share visual ambiguities, while higher ones have visual disparities.

ing multiple metrics [29] in a transfer learning set up [19], or by relying on equivalence constraints [35].

Despite all such efforts, the currently achieved performance are not satisfactory and sufficient to provide systems able to autonomously solve the re-identification problem. Indeed, the re-identification problem is usually cast as a ranking problem whose results need the final judgment of the end user. 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 believe that such a ranking is not the optimal one for the task and it is just a first step to remove the majority of the possible mismatches. Thus, additional inspections on the ranking can be applied to refine the output. The current work is based on the idea that any ranking can carry useful information to increase the position of the true match.

In Figure 1 first ranks share images with visual ambiguities, while higher ones have visual disparities [23]. The visual disparities, introduced by variations in viewpoints, pose, illumination changes, etc., induce current methods to assign a high rank to true match. When the visual disparities are not significantly affecting the visual appearance of the true match, this is usually located in the first ranks. However, it is often the case that persons in such first ranks share a similar visual appearance (i.e., visual ambiguities) and existing methods have not collected enough ability to precisely locate the true match among these. This motivates a study of the visual ambiguities occurring at first ranks so as discriminative information can be used to improve the true match rank.

The proposed discriminant context information analysis builds upon such motivation and introduces an unsupervised post-ranking framework able to increase the true matches in the first ranks. Since the approach is specifically designed to focus on visual ambiguities, it is assumed that the true match is located in the first ranks. The main goal is to find the visual ambiguities in a ranking and remove them. For such a purpose, the concepts of content and context information carried by the initial ranking are taken from [17]. In our formulation, the content information is given by the features belonging to the gallery persons that have low dissimilarity with respect to the probe (i.e., the correlated matches). While, the context information is given by the features extracted from gallery persons that have low dissimilarity with both the probe and a correlated match. In this way, content and context information lead to extract the global appearance shared by the probe and the correlated matches, thus the visual ambiguities. Then, this is removed before re-ranking. We named such a framework discriminant context information analysis (DCIA).

2. Related Work

Post-ranking methods for person re-identification is a relatively unexplored area. Earliest works following the post-ranking approach exploited ranking SVMs [34], boosting techniques for feature selection [10] or additional cues coming from soft biometrics [2]. Ranked lists computed for multiple probe persons were exploited to refine a single probe ranking [30]. Therefore, the approach works only if additional rankings (minimum 3 or 4) besides the one obtained for the current probe are available. Bidirectional ranking [17] and a saliency-based matching scheme [4] 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, not between galleries themselves. 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 both similar and dissimilar samples [1, 37], to provide relative feedback [31], or to select a single strong negative feedback to refine the ranking [23] in the deployment stage. In contrast to all such methods, we propose a single-shot approach that does not require human intervention.

A slightly different approach was recently introduced in [22], where an iterative extension to sparse discriminative classifiers was adopted to ensure that the best candidates are ranked at each iteration. However, such method did not directly consider the content and the context similarities of ranked individuals. It cast the problem by analyzing the reconstruction error and by partially ranking the gallery in terms of similarity to the probe.

Two main differences between the proposed approach and all such existing works can be highlighted: (i) there is no human neither in the training nor in the deployment loops; (ii) most importantly, the proposed approach is the only one studying the visual ambiguities shared between first ranks to improve re-identification performance, thus re-ranking is performed on a subset of the gallery.

3. Our Approach

3.1. Overview

The proposed re-identification architecture is shown in Figure 2. It consists of three main modules: ranking computation, re-ranking training and re-ranking computation. The ranking computation module resembles common re-identification pipelines and defines the basis for our approach. Let $\mathcal{T}$ be the set of training image pairs $(\mathbf{I}^A_{\text{Tr}}, \mathbf{I}^B_{\text{Tr}})$ acquired by disjoint cameras A and B. To model the appearance of each image a feature vector $\mathbf{x} \in \mathbb{R}^d$ is extracted. The set of corresponding pairs of feature vectors, here denoted as $\{\mathbf{x}^A_{\text{Tr}}, \mathbf{x}^B_{\text{Tr}}\}$, is used to learn the model parameters $\mathcal{L}_{\mathbf{x}^B\mathbf{x}^A}$ of a classifier/metric that distinguishes between pos-

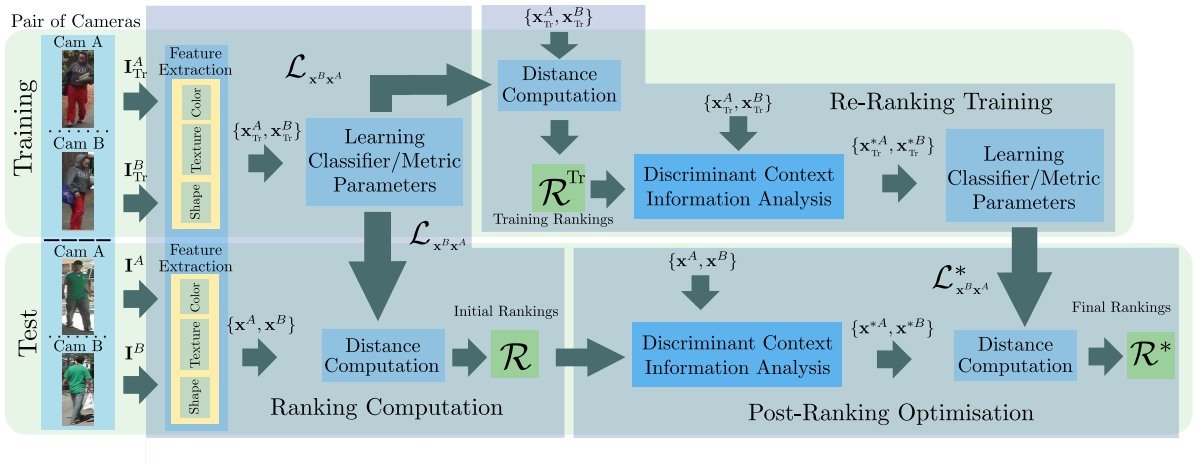


Figure 2: Overview of the proposed person re-identification system consisting of three modules: ranking computation, re-ranking training and post-ranking optimisation.

itive and negative pairs. Then, these are used to compute the distance between a test probe $\mathbf{x}^A$ and every gallery feature vector $\mathbf{x}^B$. This yields to the initial ranking $\boldsymbol{\mathcal{R}}$ for the set $\{\mathbf{x}^A, \mathbf{x}^B\}$.

The purpose of the re-ranking training module is to learn the parameters of a classifier/metric by exploiting the discriminative persons' details identified by analyzing the given training ranking. The first step is to compute the training rankings $\boldsymbol{\mathcal{R}}^{\text{Tr}}$. This is accomplished by computing the distance between each pair in the training set feature vectors $\{\mathbf{x}^A_{\text{Tr}}, \mathbf{x}^B_{\text{Tr}}\}$ using the classifier/metric parameters $\mathcal{L}_{\mathbf{x}^B\mathbf{x}^A}$. Then, the information carried by $\boldsymbol{\mathcal{R}}^{\text{Tr}}$ is exploited by the discriminant context information analysis to transform the feature vector set $\{\mathbf{x}^A_{\text{Tr}}, \mathbf{x}^B_{\text{Tr}}\}$. The obtained transformed feature vector set $\{\mathbf{x}^{*A}_{\text{Tr}}, \mathbf{x}^{*B}_{\text{Tr}}\}$ contains the discriminative persons' details. Finally, the classifier/metric parameters $\mathcal{L}^*_{\mathbf{x}^B\mathbf{x}^A}$ are learned to compute the distance for the re-ranking of the set $\{\mathbf{x}^{*A}_{\text{Tr}}, \mathbf{x}^{*B}_{\text{Tr}}\}$.

In the test phase, the re-ranking module exploits the model parameters $\mathcal{L}^*_{\mathbf{x}^B\mathbf{x}^A}$ to compute the final ranking $\boldsymbol{\mathcal{R}}^*$. Precisely, given the initial ranking $\boldsymbol{\mathcal{R}}$ produced by the ranking computation module on the feature vectors set $\{\mathbf{x}^A, \mathbf{x}^B\}$, discriminant context information analysis is applied. Then, the final ranking is obtained by computing the distance between each pair in the transformed feature vector set $\{\mathbf{x}^{*A}, \mathbf{x}^{*B}\}$ with the learned classifier/metric model parameters $\mathcal{L}^*_{\mathbf{x}^B\mathbf{x}^A}$.

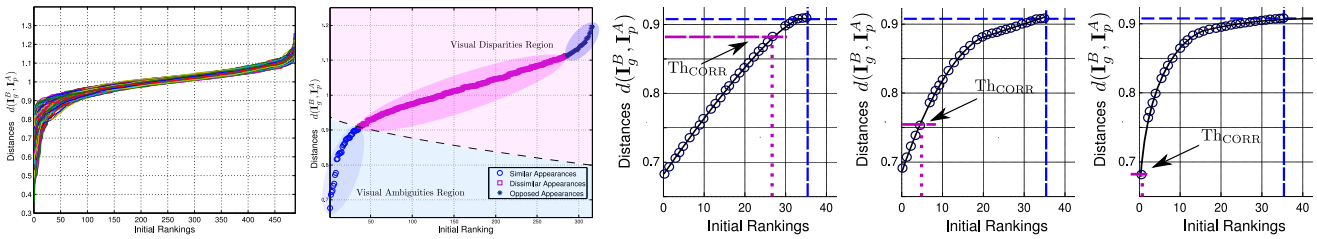
3.2. Preliminaries and Definitions

Let $\mathcal{A} = \{\mathbf{I}^A_p\}_{p=1}^N$ be the set of N probe images and $\mathcal{B} = \{\mathbf{I}^B_g\}_{g=1}^M$ be the set of M gallery images. Given a probe image $\mathbf{I}^A_p$ its initial ranking is defined as $\mathcal{R}_p = \{\mathbf{I}^B_i\}_{i=1}^M$ where the gallery images $\mathbf{I}^B_i$ are sorted depending on the dissimilarity to the probe. In other words, $d(\mathbf{I}^B_i, \mathbf{I}^A_p) < d(\mathbf{I}^B_{i+1}, \mathbf{I}^A_p)$, where $d(\cdot,\cdot)$ is a suitable dissimilarity measure [1] and i goes from 1 to $M-1$. $\boldsymbol{\mathcal{R}} = \{\mathcal{R}_p\}_{p=1}^N$ denotes the set of such initial rankings computed for the N probes.

Figure 3: Selection of the correlated matches. Gallery images in the first ranks share visual ambiguities with the probe. The content information threshold Th_{CORR} determines which gallery images should be included in the correlated matches (orange rectangle).

Our aim is to improve the rank of the true match in $\mathcal{R}_p$. Towards this objective we first select the content information for the probe image. The *content information* is defined as the set of features extracted from the correlated matches, i.e., a subset of gallery images $\mathcal{B}^{\text{cn}} \subseteq \mathcal{B}$ present in the first ranks and which are likely to share visual ambiguities with the probe. Then, the context information is computed. The *context information* consists of those features extracted from gallery persons that share visual ambiguities with both the probe and any correlated match. Content and context information are exploited to remove the visual ambiguities encoded in the original feature vectors, thus to obtain the discriminant feature vectors. These are used to compute the final ranking $\mathcal{R}^*_p$.

[1] In the following presentation, the cosine distance applied to the output of KCCA has been considered.



(a) (b) (c) (d) (e)

Figure 4: Selection of the correlated matches threshold. In (a), ranks and corresponding distances obtained for all the probe images are shown. In (b), an example of the results obtained by applying K-means clustering algorithm to the distances between a probe and all the gallery images. In (c), (d), (e) three common behaviors of the distances in the visual ambiguities region are shown. Blue dashed lines represent the limits of the similar appearances cluster. The correlated matches thresholds ($\mathrm{Th}_{\mathrm{CORR}}$) are drawn in magenta.

3.3. Content Information

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 Figure 3, the visual ambiguities bring false matches in the first ranks, often before the true match. To study the discriminative information shared among the first ranked images, we define the content information for a given probe $\mathbf{I}_p^A$. Before computing the content information, the set of correlated matches $\mathcal{B}^{\mathrm{cn}}$ have to be selected. Elements in such a set are selected from the top m positions in the initial ranking $\mathcal{R}_p$ which have a matching distance less than $\mathrm{Th}_{\mathrm{CORR}}$.

To select such m correlated matches, we propose a dynamic method that does not require $\mathrm{Th}_{\mathrm{CORR}}$ to be fixed *a priori* but let it vary for every probe. Such dynamic method requires two steps: definition of the visual ambiguities region and analysis of the distances distribution.

The solution to the first step is inspired by the shape of the distances vs ranks plots depicted in Figure 4(a). Indeed, Figure 4(a) shows that there exists a significant trend among all distance vectors highlighting that: (i) at first ranks, distances with the probe image increases abruptly, then flatten (first elbow); (ii) from the first elbow, distances grow linearly till reaching high ranks, where they finally start increasing significantly. According to such trend we have identified three classes of gallery images (see Figure 4(b)): (i) similar appearance class (C_{sa}) which correspond to gallery images with distances located before the first elbow; (ii) dissimilar appearance class (C_{da}) corresponding to gallery images having distances located between the two elbows and (iii) opposed appearance class (C_{oa}) which correspond to all the other galleries.

As shown in Figure 4(b), C_{sa} represents gallery images lying in the visual ambiguities region (first positions of the ranking). To identify such a cluster, we propose to use the K-means clustering algorithm as follows. Let $\mathcal{D} = \{d(\mathbf{I}_1^B, \mathbf{I}_p^A), \ldots, d(\mathbf{I}_M^B, \mathbf{I}_p^A)\}$ be the set of distances used to generate the ranking $\mathcal{R}_p = \{\mathbf{I}_i^B\}_{i=1}^M$. Then, the objective is to divide $\mathcal{D}$ in the three clusters C_{sa}, C_{da} and C_{oa}. This task is accomplished by minimizing:

$$\arg\min_{C} \sum_{i=1}^{K} \sum_{d(\mathbf{I}_j^B, \mathbf{I}_p^A) \in C_i} \|d(\mathbf{I}_j^B, \mathbf{I}_p^A) - \mu_i\|^2, j = 1, \ldots, M \tag{1}$$

where μ_i is the mean of the distances within cluster $C_i \in C$ and $K = 3$ corresponds to the number of clusters. Once the minimization is concluded, C_{sa} is defined by $\{d(\mathbf{I}_1^B, \mathbf{I}_p^A), \ldots, d(\mathbf{I}_k^B, \mathbf{I}_p^A)\}$, where distances are sorted in ascending order. Thus, k represents the index of the largest distance in C_{sa}.

Once the similar appearance images (i.e., the visual ambiguities region) are detected, the m correlated matches are selected by analyzing the distribution of the distances in C_{sa}. Such a process is carried out considering that not all the gallery images corresponding to distances within C_{sa} are likely to share visual ambiguities with the probe. Indeed, as shown in Figure 4(c), (d) and (e) it may happen that, due to the appearance of the probe image or the ability of the baseline model, the ranked distances are close to the centroid but not to each other. As a result large differences between consecutive rank distances can appear.

We hypothesize that only the gallery images corresponding to distances occurring before the largest gap are relevant to identify the visual ambiguities. These define a subspace where visual ambiguities are present, therefore removing them may help in distinguish the true match from the other gallery images with similar appearance. Following this idea, three cases can be identified: (i) the most of gallery images are considered as correlated matches since the gap among distances is practically uniform (Figure 4(c)); (ii) a few gallery images occurring before the largest gap are selected (Figure 4(d)); (iii) only the first gallery image is selected to form the correlated matches (Figure 4(e)). In such a case, the gallery generally corresponds to the true match.

To locate the largest gap in the visual ambiguities region, hence to obtain the threshold for correlated matches selection $\mathrm{Th}_{\mathrm{CORR}}$ we proceed as follows. Given the set of

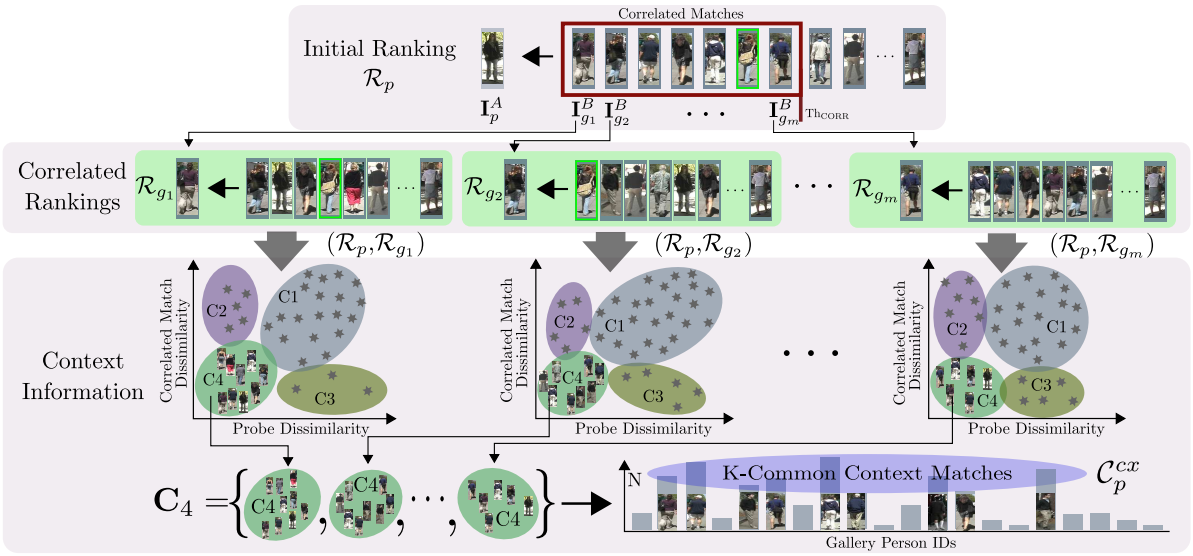


Figure 5: Representation of content and global context information for a probe image. Gallery images inside the red rectangle represent the content information set. Gallery images in the blue ellipse represent the global context information set.

ordered distances $C_{sa} = \{d(\mathbf{I}_1^B, \mathbf{I}_p^A), \ldots, d(\mathbf{I}_k^B, \mathbf{I}_p^A)\}$, the correlated matches threshold Th_{CORR} is computed as

$$\arg\max_{d(\mathbf{I}_i^B, \mathbf{I}_p^A)} \{d(\mathbf{I}_i^B, \mathbf{I}_p^A) - d(\mathbf{I}_{i+1}^B, \mathbf{I}_p^A)\}, i = 1, \ldots, k-1 \quad (2)$$

The set of m correlated matches equals $\mathcal{B}^{\text{cn}} = \{\mathbf{I}_i^B | d(\mathbf{I}_i^B, \mathbf{I}_p^A) \leq \text{Th}_{\text{CORR}}\}$. Therefore, the content information set $\mathcal{C}_p^{\text{cn}} = \{\mathbf{x}_1^{\text{cn}}, \ldots, \mathbf{x}_m^{\text{cn}}\}$ contains the m feature vectors extracted from the correlated matches in $\mathcal{B}^{\text{cn}}$. Notice that, only persons in the correlated matches are re-ranked. This is compliant with the assumption that the true match is located in the first rank positions that share visual ambiguities.

3.4. Context Information

Context information can be defined as the object frequency appearance in a particular domain. In image retrieval, the context information is the set of images containing the target object. We provide a similar definition for the person re-identification problem: the context information is given by the K-common nearest neighbors of the probe and a correlated match.

Figure 5 shows the steps to obtain the context information for a probe $\mathbf{I}_p^A$ using its correlated matches. First, we compute the initial ranking list $\mathcal{R}_g$ for each correlated matching image $\mathbf{I}_g^B \in \mathcal{B}^{\text{cn}}$. In such a case, the ranking is computed by evaluating its similarity with images in the gallery set $\mathcal{B}^* = (\mathcal{B} \backslash \mathbf{I}_g^B) \cup \mathbf{I}_p^A$ using the model parameters $\mathcal{L}_{\mathbf{x}^B \mathbf{x}^A}$ and distance $d(\cdot, \cdot)$.

Then, given $\mathcal{R}_p$ and $\mathcal{R}_g$, we define four clusters using the distance measure $d(\cdot, \cdot)$ and two thresholds $\text{Th}_{\text{CORR}}^{\mathcal{R}_p}$ and $\text{Th}_{\text{CORR}}^{\mathcal{R}_g}$ computed using the method proposed in section 3.3. The clustering conditions are given in Table 1.

Table 1: Proposed context clustering conditions. Each $\mathbf{I}_i^{BB^*} \in \mathcal{B} \cap \mathcal{B}^*$ is assigned to the cluster that satisfies both conditions.

Distances	C_1	C_2	C_3	C_4
$d(\mathbf{I}_i^{BB^*}, \mathbf{I}_p^A), \text{Th}_{\text{CORR}}^{\mathcal{R}_p}$	$>$	$<$	$>$	$<$
$d(\mathbf{I}_i^{BB^*}, \mathbf{I}_g^B), \text{Th}_{\text{CORR}}^{\mathcal{R}_g}$	$>$	$>$	$<$	$<$

C_4 is the context relevant cluster. The clustering process is carried for each correlated match, thus generating m clusters denoted as C_4^g, for $g = 1, \ldots, m$. The elements in each C_4^g represent the images that have high similarity with both the probe and the correlated match $\mathbf{I}_g^B$ itself (i.e., the nearest neighbors). The context information is extracted from the K common context matches. These are the gallery images having the K highest frequencies in the set $\boldsymbol{C_4} = \{C_4^g\}_{g=1}^m$ (see Figure 5). The feature vectors extracted from such images form the context information set $\mathcal{C}_p^{\text{cx}} = \{\mathbf{x}_1^{\text{cx}}, \ldots, \mathbf{x}_n^{\text{cx}}\}$, where $n = \max(|\cup_{g=1}^m C_4^g|, K)$.

The hard threshold K has been introduced to reject gallery images that are likely to have different visual appearance in the context information computation. Finally, $\mathcal{C}_p^{\text{cx}}$ is updated by removing feature vectors that are in duplicated in $\mathcal{C}_p^{\text{cn}}$, i.e. $\mathcal{C}_p^{\text{cx}} = \mathcal{C}_p^{\text{cx}} \backslash \mathcal{C}_p^{\text{cn}}$.

3.5. Discriminant Information Analysis

The content and context information provide a set of images with similar global appearance. The goal of this analysis is to detect, hence to remove the visual ambiguities from the corresponding feature vectors. This allows to focus on discriminant features that might help to correctly locate the true match within the correlated matches and increase its position in the initial ranking.

We hypothesize that visual ambiguities mainly correspond to the global appearance, hence to remove such information the principal components of such feature vectors should be considered. Thus, to remove the visual ambiguities, we use principal component analysis (PCA) as follows.

Given a probe image $\mathbf{I}_p^A$, let $\mathcal{D}_p = \{\mathbf{x}_p, \mathcal{C}_p^{\text{cn}}, \mathcal{C}_p^{\text{cx}}\}$ be the set composed of its feature vector and of feature vectors obtained in the content and context information. We redefine $\mathcal{D}_p$ as a feature matrix $\mathbf{D}_p \in \mathbb{R}^{d \times l}$ with zero mean, where $l = 1 + m + n$ is the number of vectors. Let $\mathbf{P} \in \mathbb{R}^{d \times k}$ be the first k principal components of $\mathbf{D}_p$ selected to represent the common appearance subspace. We project $\mathbf{D}_p$ to the subspace as $\mathbf{P}\mathbf{P}^T\mathbf{D}_p$. Thus, the discriminant information can be obtained as:

$$\mathbf{D}_p^* = \mathbf{D}_p - \mathbf{P}\mathbf{P}^T\mathbf{D}_p \tag{3}$$

where each column of $\mathbf{D}_p^*$ represents a discriminant feature vector $\mathbf{x}^*$.

Re-Ranking Training: The proposed discriminant context information analysis is applied to each ranking $\mathcal{R}_p^{\text{Tr}} \in \boldsymbol{\mathcal{R}}^{\text{Tr}}$. As result, for each probe p we get the discriminant feature vectors $\mathbf{x}_p^{*A}$ and $\mathbf{x}_g^{*B} \in \{\mathcal{C}_p^{\text{cn}}, \mathcal{C}_p^{\text{cx}}\}$. The resulting sets $\mathbf{x}_{\text{Tr}}^{*A} = \{\mathbf{x}_p^{*A}\}_{p=1}^{|T|}$ and $\mathbf{x}_{\text{Tr}}^{*B} = \left\{\mathbf{x}_g^{*B} | \mathbf{x}_g^{*B} \in \{\mathcal{C}_p^{\text{cn}}, \mathcal{C}_p^{\text{cx}}\}\right\}_{p=1}^{|T|}$ together with the pairwise labels are used to learn a classifier/metric parameters $\mathcal{L}^*_{\mathbf{x}^B\mathbf{x}^A}$.

Post-Ranking Optimisation: In a similar way, given a test ranking in $\boldsymbol{\mathcal{R}}$, the discriminant context information analysis is performed to obtain the discriminative test feature vectors $\mathbf{x}_p^{*A}$, $\mathbf{x}_g^{*B}$. Then, the set of such vectors $\{\mathbf{x}^{*A}, \mathbf{x}^{*B}\}$ is evaluated by the classifier/metric having learned parameters $\mathcal{L}^*_{\mathbf{x}^B\mathbf{x}^A}$. The so computed final distance is used to re-rank the correlated matches, hence to compute the final ranking $\boldsymbol{\mathcal{R}}^*$.

4. Experimental Results

In this section we on report the performance of our approach on three publicly available datasets. First, datasets details and settings are given. Second, we introduce the baseline models selected to obtain initial rankings and our implementation details. Then, an evaluation of our approach for each baseline model and discussions about the results obtained using our post-ranking optimisation are provided. Finally, comparisons with state-of-the-arts methods are shown.

4.1. Datasets

VIPeR Dataset[2]: The VIPeR dataset [8] is considered the most challenging person re-identification dataset. It contains images of 632 persons viewed by two non-overlapping cameras. The 1264 images have severe lighting variations, different viewpoints and background clutter. Following the commonly adopted procedure [8, 15, 21], the dataset has been split into two subsets of 316 persons each, for training and test respectively.

PRID Dataset[3]: The PRID dataset [10] 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 [12]: persons from camera A are used as probes and persons from camera B as gallery. Among the 200 persons appearing in both cameras, 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.

CUHK02 Dataset[4]: The CUHK02 dataset [18] contains 1816 persons and five camera pairs which have 971, 306, 107, 193 and 239 persons, respectively. Each person has two images in each camera view. It is a challenging dataset due to pose variations and lighting changes that occurs between camera pairs. To evaluate our approach we have used the same protocol as in [18], hence selected the camera pair having 971 persons. We have split into two sets, containing 485 (training) and 486 (test) persons, respectively.

4.2. Implementation Details

To model the person appearance we have used the representation in [21]. Images have been resized to 64×128 pixels. Feature vectors are represented by isotropic Gaussian weighted color histograms extracted from 8 horizontal stripes. From each stripe we extract 24-bins histograms from the Hue-Saturation (HS), RGB and Lab channels. 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, have been concatenated to form the final 4842-dimensional feature vector.

We have used four baseline models to evaluate the performance of the proposed discriminant information analysis algorithm, namely KCCA [21], KISSME [15], LADF [20] and the Euclidean distance. Following our notation, $\mathcal{L}_{\mathbf{x}^B\mathbf{x}^A}$ and $\mathcal{L}^*_{\mathbf{x}^B\mathbf{x}^A}$ represent the set of parameters learned by such algorithms trained on feature vector sets $\{\mathbf{x}_{\text{Tr}}^A, \mathbf{x}_{\text{Tr}}^B\}$ and $\{\mathbf{x}_{\text{Tr}}^{*A}, \mathbf{x}_{\text{Tr}}^{*B}\}$, respectively. For KCCA $d(\cdot,\cdot)$ is the cosine distance applied in the KCCA space while for KISSME and LADF it is the learned non-Euclidean metric.

We have selected $K = 10$ as the maximum number of common context matches in our current framework. k principal components corresponding to the 55% of energy of the set of feature vectors, have been used to represent the

[2] Available at http://soe.ucsc.edu/~dgray/

[3] Available at http://lrs.icg.tugraz.at/download.php

[4] Available at http://www.ee.cuhk.edu.hk/~xgwang/CUHK_identification.html

common appearance subspace. To make a fair comparison, for each experiment, we have run 10 trials using random person IDs. The results averaged over these 10 trials are shown in terms of Cumulative Matching Characteristics (CMC) curve.

4.3. Analysis of the Proposed Approach with different baseline models

Figure 6 shows the performances of DCIA applied on baseline models, for VIPeR, PRID and CUHK02 datasets. On the right side of each sub-figure, a zoom of the 3 first ranks is shown to remark the improvement of DCIA with respect to the baseline models. Since only correlated matches are re-ranked, variations mainly occurs at first ranks. Thus, we show the results for the first 25 ranks. For the rest of ranks, we obtain comparable results to baseline models.

In Figure 6(a), comparisons with baseline models are given for the VIPeR dataset. KCCA obtains a recognition percentage of 42.09% for rank 1, whereas KCCA+DCIA achieves a 63.92%, thus improving the baseline model results by more than 20%. Similarly, for KISSME, LADF and Euclidean distance models, the rank 1 recognition percentage increases from 33.8% to 38.87%, 40.53% to 44.67% and from 12.97% to 16.29%, respectively. Though the DCIA boosts all baseline results, the most remarkable improvement is achieved using the KCCA baseline model. In such a case DCIA improves the first 18 ranks. The first 5, 4 and 5 ranks are improved when the DCIA is applied over KISMME, LADF and Euclidean distance.

As is shown in Figure 6(b), DCIA provides a remarkable performance gain over baseline models when the PRID dataset is considered. It improves the initial results up to the first 21 ranks for Euclidean distance, the first 15 ranks for KCCA, the first 14 ranks for KISSME and the first 17 ranks for LADF. In particular, the recognition percentage increases from 18.0% to 39.0% for rank 1 using KCCA as baseline model, from 14.0% to 23.5% using KISSME, from 29.5% to 36.5% using LADF and from 11% to 18% using the Euclidean distance. Notice that the performance boost affects more ranks than for the VIPeR. This is because the PRID dataset has several persons that looks very similar to each other. Hence, the content information set includes more images from which our method is able to extract the discriminative features.

Figure 6(c) shows the performance of the proposed approach on CUHK02 dataset. The best performance is obtained using KCCA. In such a case the rank 1 improvement is of about 24%. DCIA improves the initial results for the first 24 ranks using KCCA and Euclidean distance and for the first 8 and 10 ranks using KISMME and LADF.

Finally, to analyze the contribution of the content and context information we have computed the results in Table 2. Results are shown in terms of rank 1 recognition performance. Results demonstrate that by removing the content information (3^{rd} column) initial performances decrease. The opposite occurs when context information is removed (4^{th} column). Results show that for every dataset, the optimal performance are achieved when both of them are exploited (last column). To summarize, while content information is more important than the context one, both form the whole similar appearance space thus should be jointly used to obtain better results.

Table 2: Contribution of content and context information in DCIA. Rank 1 results obtained for each dataset are shown. Best performance are highlighted in boldface font.

Dataset	Baseline Model	DCIA		
		No Content	No Context	All
VIPeR	42.09	32.66	54.42	**63.92**
PRID	18.00	14.00	29.50	**39.00**
CUHK02	38.52	26.25	53.57	**61.67**

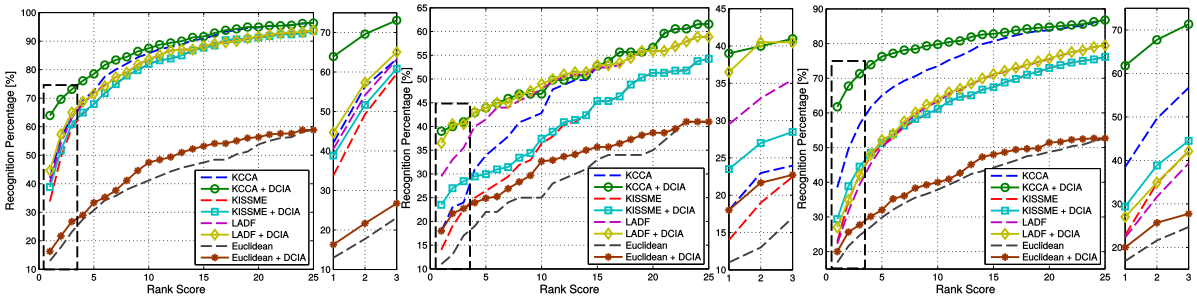
4.4. Comparison with state-of-the-art methods

In this section we compare the proposed algorithm with state-of-the-art methods on the three selected datasets. Figure 7 shows the CMC curves up to rank 25 for each method on every dataset. Similar to the previous section, we include a zoom for the 3 first ranks on the right side.

In Figure 7(a), performances of DCIA on VIPeR dataset are compared with existing methods for which either the full CMC curve or the code is available, namely KISSME [15], LF [33], LADF [20], LMF [43], LMF+LADF [43] and KCCA [21]. KCCA+DCIA outperforms existing ones especially for the most representative ranks. In particular, we achieve a correct recognition percentage of 63.92% for rank 1, while none of the other reach a recognition percentage higher than 45%. The performance gap decreases for higher ranks. This is due to the fact that we perform the post-ranking on correlated matches only.

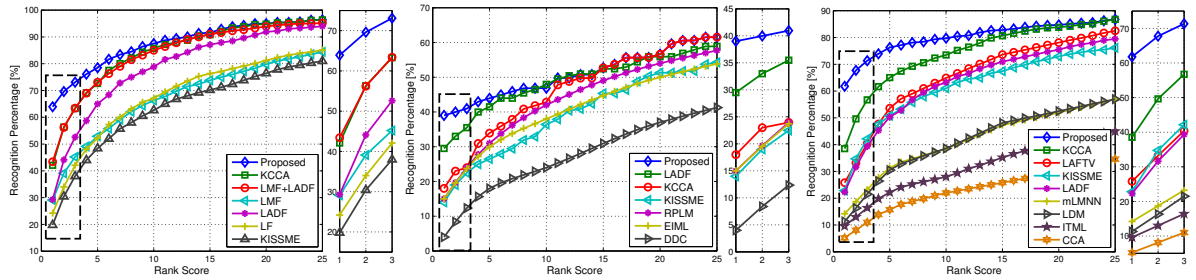
Figure 7(b) shows comparisons with those obtained by KISSME [15], DDC [10], EIML [11], RPLM [12], LADF [20] and KCCA [21] on PRID dataset. KCCA+DCIA and LADF+DCIA perform similarly but outperforms all states-of-the-art methods up to rank 9 where similar results to baseline models are achieved. In particular, KCCA+DCIA reaches a rank 1 recognition percentage of 39.0% thus improving the baseline performance by more than 20%.

In Figure 7(c), we report on the performance obtained by DCIA on the CUHK02 dataset and compare with those of CCA [13], ITML [5], LDM [41], mLMNN [38], LADF [20], KISSME [15], SWF [29], LAFTV [18] and KCCA [21] . As for previous scenarios KCCA+DCIA achieves a rank 1 recognition rate of 61.7% and outperforms existing approaches up to rank 20. For the same rank, recognition rates of 5.2%, 9.6%, 11.3%, 14.3%, 22.2% , 22.8%, $25,8\%$ and 38.5% are reached by the methods used for comparison.



(a) (b) (c)

Figure 6: Performance of the proposed algorithm using different baseline models on: (a) VIPeR dataset, (b) PRID dataset and (c) CUHK02 dataset. Results are shown as CMC curves and compared to the baseline approach.



(a) (b) (c)

Figure 7: Comparisons of the proposed algorithm with state-of-the-art methods: (a) on the VIPeR dataset, (b) on the PRID dataset (including distractors) and (c) on the CUHK02 dataset.

Table 3: Comparison with re-ranking methods on the VIPeR dataset. Recognition percentages for some relevant ranks are shown. Best results are in boldface font.

Rank $\rightarrow$	**1**	**5**	**10**	**25**	**50**
Euc. Dist.+ DCIA	16.29	33.38	47.46	58.86	72.78
DDC [10]	19	-	52	69	80
KISSME+SB [2]	19.3	50.7	63.3	78.2	90.6
KISSME+CCRR [17]	22	49	69	87	95
RIRO [37] (1 Iteration)	28	30	34	51	64
PRRS [4]	33.29	-	78.35	-	97.53
KISSME+ DCIA	38.87	67.96	82.01	93.62	98.36
IRT [1] (1 Iteration)	43	45	46	53	61
LADF+ DCIA	44.67	71.54	83.56	93.82	98.52
POP [23] (1 Iteration)	59.05	60.95	63.10	72.20	-
KCCA+ DCIA	**63.92**	**78.48**	**87.50**	**96.36**	**99.05**

4.5. Comparison with post-ranking methods

To demonstrate the benefits DCIA with respect to other post-ranking methods we have included Table 3. Since the majority of the re-ranking methods provide results using the VIPeR dataset only, we compare our performance on such a dataset. RIRO [37], IRT [1] an POP [23] have an iterative operation which requires users being in the loop. To make a fair comparison, we considered the performance obtained using a single iteration. Results show that KCCA+DCIA outperforms all existing methods. In particular, it improves the best performance (59.05%) obtained so far by POP [23] (1 iteration) by more than 4%. Concerning a comparison with post-ranking models for generic image search and retrieval like [40, 9], the proposed solution performs better since such methods, as shown in [23], achieve worse results than POP [23]. More interestingly, notice that SB [2] and CCRR [17] provide the performance achieved using the KISSME baseline model. Using such baseline model, SB and CCRR improve the rank 1 baseline performance by 0.7% and 2%, respectively. DCIA improves the rank 1 performance by 5%. Thus, it is more effective than existing methods when applied to the same baseline model.

5. Conclusion

In this paper we have proposed a novel unsupervised post-ranking approach to improve the first rank person re-identification performance. We have focused on the visual ambiguities share between first ranked persons. A discriminant information analysis, based on content and context information, has been proposed to remove common global appearance. The performance of our method has been compared with state-of-the-art methods using three public benchmark datasets. Results demonstrated that first rank performance improves. In particular, previously rank 1 performances have been improved by more than 20% on two datasets. This strongly support our believes,i.e., that the initial ranking includes relevant information that can be used to improve first rank performance.

As future works we will investigate different approaches to identify the persons sharing visual ambiguities, e.g. [32].

References

[1] S. Ali, O. Javed, N. Haering, and T. Kanade. Interactive retrieval of targets for wide area surveillance. *Proceedings of the international conference on Multimedia*, page 895, 2010. 2, 8

[2] L. An, X. Chen, M. Kafai, S. Yang, and B. Bhanu. Improving person re-identification by soft biometrics based reranking. In *ICDSC*, pages 1–6, Oct 2013. 2, 8

[3] L. An, M. Kafai, S. Yang, and B. Bhanu. Reference-Based Person Re-Identification. In *AVSS*, 2013. 1

[4] L. An, M. Kafai, S. Yang, and B. Bhanu. Person re-identification with reference descriptor. *IEEE TCSVT*, PP(99):1–1, 2015. 2, 8

[5] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In *ICML*, ICML, pages 209–216, New York, NY, USA, 2007. ACM. 7

[6] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja. Pedestrian Recognition with a Learned Metric. In *ACCV*, pages 501–512, 2010. 1

[7] J. Garcia, N. Martinel, G. L. Foresti, A. Gardel, and C. Micheloni. Person Orientation and Feature Distances Boost Re-identification. In *ICPR*, pages 4618–4623. IEEE, Aug. 2014. 1

[8] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In D. Forsyth, P. Torr, and A. Zisserman, editors, *ECCV*, volume 5302, pages 262–275. 2008. 6

[9] J. He, M. Li, Z. Li, H.-j. Zhang, H. Tong, and C. Zhang. Pseudo Relevance Feedback Based on Iterative Probabilistic One-Class SVMs in Web Image. In *PCM*, pages 213–220, 2004. 8

[10] M. Hirzer, C. Beleznai, P. Roth, and H. Bischof. Person re-identification by descriptive and discriminative classification. In *Image Analysis*, pages 91–102. 2011. 2, 6, 7, 8

[11] M. Hirzer, P. Roth, and H. Bischof. Person re-identification by efficient impostor-based metric learning. In *AVSS*, pages 203–208, Sept 2012. 7

[12] M. Hirzer, P. Roth, M. Kstinger, and H. Bischof. Relaxed pairwise learned metric for person re-identification. In *ECCV*, pages 780–793. 2012. 1, 6, 7

[13] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4):pp. 321–377, 1936. 7

[14] O. Javed, K. Shafique, Z. Rasheed, and M. Shah. Modeling inter-camera spacetime and appearance relationships for tracking across non-overlapping views. *CVIU*, 109(2):146–162, Feb. 2008. 1

[15] M. Kostinger, M. Hirzer, P. Wohlhart, P. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *CVPR*, pages 2288–2295, June 2012. 6, 7

[16] R. Layne, T. M. Hospedales, and S. Gong. Re-id : Hunting Attributes in the Wild. In *BMVC*, 2014. 1

[17] Q. Leng, R. Hu, C. Liang, Y. Wang, and J. Chen. Person re-identification with content and context re-ranking. *Multimedia Tools and Applications*, pages 1–26, 2014. 2, 8

[18] W. Li and X. Wang. Locally Aligned Feature Transforms across Views. In *CVPR*, pages 3594–3601. IEEE, June 2013. 6, 7

[19] W. Li, R. Zhao, and X. Wang. Human Reidentification with Transferred Metric Learning. In *ACCV*, pages 31–44, 2012. 2

[20] Z. Li, S. Chang, F. Liang, T. Huang, L. Cao, and J. Smith. Learning locally-adaptive decision functions for person verification. In *CVPR*, pages 3610–3617, June 2013. 6, 7

[21] G. Lisanti, M. L., and A. Del Bimbo. Matching people across camera views using kernel canonical correlation analysis. In *ICDSC*, pages 1–6, Nov 2014. 6, 7

[22] G. Lisanti, I. Masi, A. Bagdanov, and A. Del Bimbo. Person Re-identification by Iterative Re-weighted Sparse Ranking. *IEEE TPAMI*, pages 1–1, 2014. 1, 2

[23] C. Liu, C. Loy, S. Gong, and G. Wang. Pop: Person re-identification post-rank optimisation. In *ICCV*, pages 441–448, Dec 2013. 2, 8

[24] B. Ma, Q. Li, and H. Chang. Gaussian Descriptor based on Local Features for Person Re-Identification. In *ACCV*, pages 1–14, 2014. 1

[25] B. Ma, Y. Su, and F. Jurie. Covariance Descriptor based on Bio-inspired Features for Person Re-identification and Face Verification. *Image and Vision Computing*, 32:379–390, Apr. 2014. 1

[26] L. Ma, X. Yang, and D. Tao. Person Re-Identification Over Camera Networks Using Multi-Task Distance Metric Learning. *IEEE TIP*, 23(8):3656–3670, 2014. 1

[27] N. Martinel, A. Das, C. Micheloni, and A. K. Roy-Chowdhury. Re-Identification in the Function Space of Feature Warps. *IEEE TPAMI*, 37(8):1656–1669, Aug. 2015. 1

[28] N. Martinel and C. Micheloni. Sparse Matching of Random Patches for Person Re-Identification. In *ICDSC*, pages 1–6, 2014. 1

[29] N. Martinel, C. Micheloni, and G. L. Foresti. Saliency Weighted Features for Person Re-Identification. In *ECCV, Workshops and Demonstrations*, number i, pages 1–17, 2014. 2, 7

[30] V.-H. Nguyen, T. D. Ngo, K. M. Nguyen, D. A. Duong, K. Nguyen, and D.-D. Le. Re-ranking for person re-identification. In *Soft Computing and Pattern Recognition (SoCPaR), 2013 International Conference of*, pages 304–308, Dec 2013. 2

[31] A. Parkash and D. Parikh. Attributes for classifier feedback. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, editors, *ECCV*, volume 7574 of *Lecture Notes in Computer Science*, pages 354–368, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. 2

[32] M. Pavan and M. Pelillo. Dominant sets and pairwise clustering. *IEEE TPAMI*, 29(1):167–172, 2007. 8

[33] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian. Local fisher discriminant analysis for pedestrian re-identification. In *CVPR*, pages 3318–3325, June 2013. 7

[34] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by support vector ranking. In *BMVC*, pages 21.1–21.11, 2010. 2

[35] D. Tao, L. Jin, Y. Wang, and X. Li. Person Reidentification by Minimum Classification Error-Based KISS Metric Learning. *IEEE TOC*, pages 1–11, 2014. 1, 2

[36] H. Wang, S. Gong, and T. Xiang. Unsupervised Learning of Generative Topic Saliency for Person Re-identification. In *BMVC*, pages 1–11, 2014. 1

[37] Z. Wang, R. Hu, C. Liang, Q. Leng, and K. Sun. Region-based interactive ranking optimization for person re-identification. In *Advances in Multimedia Information Processing*, pages 1–10. 2014. 2, 8

[38] K. Weinberger and L. Saul. Fast solvers and efficient implementations for distance metric learning. In *ICML*, pages 1160–1167. ACM, 2008. 7

[39] Z. Wu, Y. Li, and R. Radke. Viewpoint Invariant Human Re-identification in Camera Networks Using Pose Priors and Subject-Discriminative Features. *IEEE TPAMI*, 8828:1–1, 2014. 1

[40] R. Yan, A. G. Hauptmann, and R. Jin. In *ACM MM*, page 343, New York, New York, USA, 2003. 8

[41] L. Yang, R. Jin, R. Sukthankar, and Y. Liu. An efficient algorithm for local distance metric learning. In *AAAI*, pages 543–548, 2006. 7

[42] R. Zhao, W. Ouyang, and X. Wang. Unsupervised Salience Learning for Person Re-identification. In *CVPR*, pages 3586–3593. IEEE, June 2013. 1

[43] R. Zhao, W. Ouyang, and X. Wang. Learning Mid-level Filters for Person Re-identification. *CVPR*, pages 144–151, June 2014. 7

[44] T. Zhou, M. Qi, J. Jiang, X. Wang, S. Hao, and Y. Jin. Person Re-identification based on nonlinear ranking with difference vectors. *Information Sciences*, (April), Apr. 2014. 1