Appearance-Based Person Re-identification by Intra-Camera Discriminative Models and Rank Aggregation

Raphael Felipe de Carvalho Prates, William Robson Schwartz
Department of Computer Science, Universidade Federal de Minas Gerais, Brazil
pratesufop@gmail.com, william@dcc.ufmg.br

Abstract

The main challenges in person re-identification are related to different camera acquisition conditions and high inter-class similarities. These aspects motivated us to handle such problems by learning intra-camera discriminative models, based on training samples, to discover representative individuals for a given sample (probe or gallery samples), referred to as prototypes. These prototypes are used to weight the features according to their discriminative power by using the Partial Least Square (PLS) method. We also exploit models built from the gallery and probe samples to generate re-identification results that will be combined in a single ranking using ranking aggregation techniques. According to the experiments, the proposed method achieves state-of-the-art results. They also demonstrate that aggregating the results achieved by our method with results achieved by a distance metric learning method, outperforms the state-of-the-art, e.g., the top-1 rank is increased in almost 10 percent points for VIPeR and PRID 450S data sets.

1. Introduction

Responsible for automatically tracking individuals in a camera network, the person re-identification (Re-ID) has attracted the attention of many researches in the past few years [1]. The high intra-class variance caused by different camera acquisition conditions such as viewpoints, illumination, poses and background interference makes it a challenge problem. In addition, due to the lack of temporal information (the individuals cannot be tracked continuously by only one camera) and the low reliability on traditional biometrics cues such as face and iris recognition because of low resolution images [31], the ranking generated by re-identification need to be based on the appearance alone [7].

The Re-ID is highly related with image retrieval [27]. Considering that we have a probe sample with unknown identity, the goal is to rank a gallery of known subjects. In Re-ID, the probe and gallery images are usually captured by different cameras, from which training samples can be employed to learn a similarity model. In fact, the most common scenario evaluated by state-of-the-art methods for Re-ID considers closed-set Re-ID, single-shot, two cameras and short period [1]. In other words, there are two nearby cameras, c_1 and c_2, with just one image of each subject per camera and the identity being presented in the probe sample is always in the gallery. Figure 1 illustrates this scenario.

Figure 1. Examples from VIPeR dataset [6]. Images are split in two subsets according to the camera (blue borders for camera c_1 and red for camera c_2). The left column shows the same person in the probe (top row) and in the gallery (bottom row). In the center are individuals similar to the samples in the left. The right column shows their image pairs in the other camera (prototypes).

Different approaches for Re-ID can be roughly categorized as distance metric learning and appearance-based. Distance metric learning methods estimate affine transformations [10] that respect a pairwise constraint, keeping

pairs of the same person closer than pairs of different persons regardless of the representation of choice. These methods usually have the drawback of overfitting and presenting high computational costs due to the complex optimization problems [31]. On the other hand, appearance-based methods assume that it is possible to extract features that are simultaneously robust to different camera acquisition conditions and discriminative, which is in practice unfeasible. These methods classify the pairs in same or not-same using some standard distances [31] or a classifiers.

In Schwartz and Davis [25], the obtained results indicated a superior performance of Partial Least Square (PLS) method, in Re-ID problem, when compared to PCA and different setups of SVM. Then, a PLS one-against-all classification scheme is employed to model appearance in our work. In addition, inspired by the work of Chunxiao *et al*. [15], our models are learned using prototypes - subjects with similar appearances - obtained from a training set. To avoid inter-camera hurdles, our prototypes are constructed using intra-camera similarity models, such as the one-shot similarity (OSS) [29], and are restricted to individuals captured by the same camera. Finally, the matches are ranked according to the responses given by the PLS models and when multiple rankings are available for a given probe image, they are aggregated to obtain a final answer regarding its identity. To the best of our knowledge, this is the first work that evaluates and aggregates different techniques to construct a unified ranking.

The main contribution of this work is that we tackle the person re-identification problem without dealing directly with different camera conditions (a probe image captured by a camera c_1 is not compared against gallery samples captured by a camera c_2 directly). As illustrated in Figure 1, for a given gallery image captured by a camera c_2, we use an intra-camera discriminative model to discover a set of similar individuals in training set, which were also captured by camera c_2. Thus, the transition from c_2 to c_1 in training set is seamless, since pair of images in both cameras are available. Then, instead of comparing the probe image with gallery samples directly, we compare it to the similar individuals in training set and at the same camera as the probe (our prototypes). This strategy avoids camera transformations, such as brightness transfer functions [9] and explores the inter-class similarity in our favour.

Experimental results demonstrate that our proposed appearance-based method is successful for person re-identification, reaching results that only have been achieved by approaches based on distance metric learning in two important data sets, the VIPeR and the PRID 450S.

2. Related Works

This section briefly reviews the main works focused on person re-identification. Detailed discussions can be found in [1] and [28]. We describe works based on two main approaches: (1) distance metric learning and (2) appearance-based. While the former is based on finding global and linear transformations to emphasized relevant dimensions reinforcing a pairwise constraint, the latter focuses on the image representation seeking for more robust and discriminative features. Experimental results show that improved results can be achieved when both approaches are combined, indicating that they provide complementary information.

2.1. Methods Based on Distance Metric Learning

This line of research aims at learning a distance metric that gives smaller distances between pairs of the same person and larger otherwise, usually based on the Mahalanobis distance. Koerstinger *et al*. [10] proposed KISSME, a simple yet effective strategy based on the likelihood-ratio test that is orders of magnitude faster than comparable methods. However, when employed for small training sets, it needs a smoothing and regularization step as pointed out by Tao *et al*. [27]. In [31], the authors addressed the problem employing probability in which the objective function maximizes the probability of similar pairs being closer. Differently, a relaxed pairwise metric is learned considering pairs at different camera conditions [8]. While the majority of the methods use the feature space, in [14] a common subspace is employed to jointly transform image features and learn local optimal metrics. The scalability issue is addressed in [17], where multiple camera pairs are used to learn a multi-task distance metric, called MtMCML.

2.2. Appearance-Based Methods

Appearance-based methods seek for features that are stable to changes in resolution, pose, background and illumination. These features in low resolution images are highly based on the clothing information, mainly in color and texture, since shape is related to the articulated human body and depends on the pose and the viewpoint [1]. Even though texture descriptors as Local Binary Patterns (LBP) [8] and filters, such as Gabor and Schmidt, have been employed [7], they are in usually outperformed by color descriptors such as color names [12] and histogram of color channels [7].

While in [18] the authors use pixels directly, the majority of the works use more complex features and some machine learning techniques such as AdaBoost [7], PLS [25], RankSVM [21], RankBoost [12], PCA [30] and RCCA [2]. Furthermore, some works regard feature information as complementary and concatenate them given weights accordingly to their importance [25]. However, what is the most important feature? That depends on the most discriminative characteristics present in probe and gallery images [15, 16]. In [16], for instance, the importance of each feature is estimated in subsets of similar individuals (prototypes) using random forests.

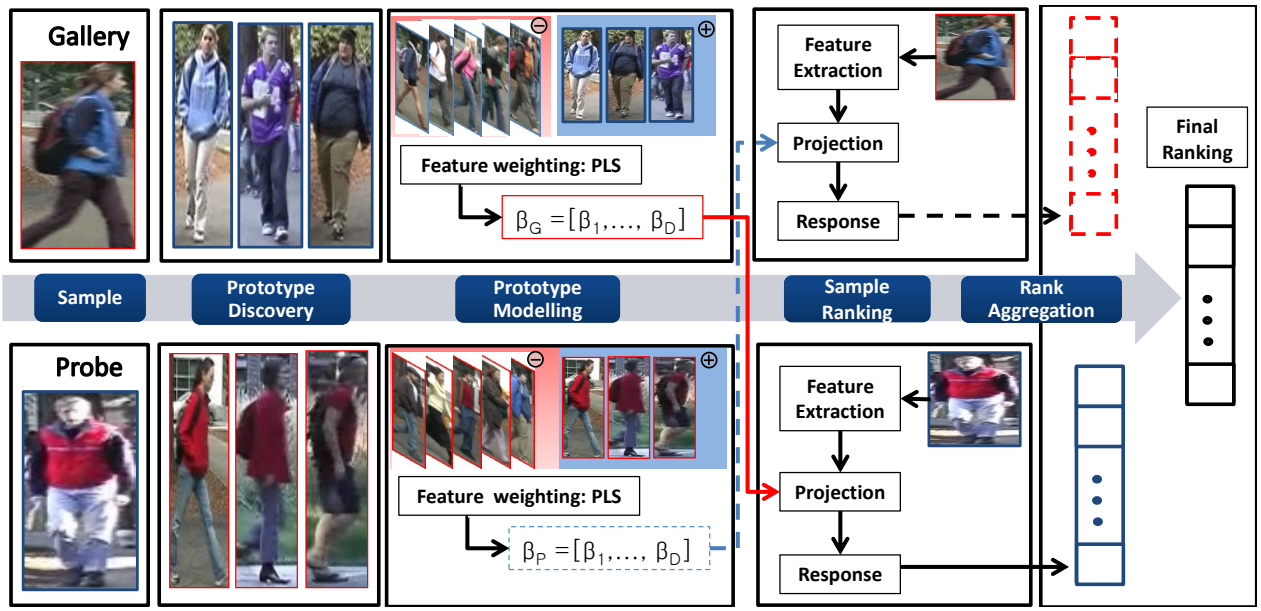


Figure 2. Overview of the proposed method. For a given image sample (probe or gallery), first we find the prototypes using the intra-camera similarity models based on training samples (Section 3.1), then the prototypes are used to model the image sample (Section 3.2) and finally the both rankings are aggregated to obtain the final answer regarding the identity of the sample (Section 3.3).

Colors descriptors have advanced from the early works that used brightness transfer functions to handle different camera illuminations [9, 20] to more stable features which do not require constant calibration [18, 12, 30, 13]. In [13], the authors captured stable aspects of the multi-modal color distributions using shape-context descriptors and log-chromacity color space. On the other hand, in [12], semantic color names (*e.g.* red, black, gray) learned from the internet are employed to relate RGB values with probability distributions over color names. Instead of using a pallet of colors names captured externally, Cai and Pietik [3] implicitly discovered it using k-means and soft-assignment using the bag-of-words scheme. Yang *et al.* [30] presented the SCNCD, which computes for each RGB value a probabilistic distribution over color names. They used 16-color palette[1] in the RGB color space. The SCNCD employs an index to assign multiple RGB values to the same descriptor, increasing its robustness. The saliency is then obtained attributing non-zero probabilities only to the nearest color names with respect to the RGB value.

3. Proposed Method

In this section we describe our novel method, which is schematically represented in Figure 2. For each sample s_i (either in the gallery or a probe), our approach computes a subset containing the most similar individuals from a training set (note that there is no intersection between individuals in the gallery and in the training set). These individual are considered to be representative of sample s_i. Since these individuals are in the training set, their samples in the other camera, say c_2, are available. We call such subset of samples in camera c_2 prototypes. The use of prototypes indirectly handles different camera conditions.

When using the gallery samples, we have a set of n individuals, denoted by $G = \{g_1, ..., g_n\}$ and a set of prototypes $\hat{P} = \{\hat{P}_1, ..., \hat{P}_n\}$, where $\hat{P}_i$ are the prototypes for the subject g_i.

Once the prototypes are available, the PLS one-against-all is employed to find discriminative models (Section 3.1). PLS finds latent variables that model the linear relations between observed variables considering response variables (class labels). It is a class aware dimensionality reduction algorithm which iteratively constructs a low dimensional subspace given higher weights to more discriminative features [22]. These weights are proportional to the regression coefficients (β) values [24]. For each image sample, a different PLS is learned considering its prototype as positive and all remaining images in the training sample as negative, which has the advantage of adaptively weighting the features importance (Section 3.2). By projecting the feature vector of the sample onto the multiple models, it is possible to build a rank of similarities, the higher the response, more similar the sample and the prototypes are (Section 3.3). By executing the described process, two rankings can be obtained, one when the gallery sample is considered and an-

[1] 16-color palette color names are fuchsia, blue, aqua, lime, yellow, red, purple, navy, teal, green, olive, maroon, black, gray, silver and white.

other when the probe is considered.

3.1. Prototype Discovery

In this section, we describe the method employed to discover prototypes. The search for prototypes is performed using only samples from the training set captured by the same camera as the sample image being considered. For instance, given a person captured by camera c_1, just the training images at camera c_1 will be considered in the similarity models. This strategy avoids inter-camera hurdles.

Our approach was inspired in [16], with some important improvements. First, as consequence of high inter-class similarities, the prototypes here are not considered as disjoint subsets, which avoids similar subjects assigned to distinct prototypes and further modeled in different training classes. For instance, a person dressing black shirt and white pants can belong to these two subsets black shirt and white pants, while in [16], it must be assigned to one of them. In addition, in [16], the prototype discover is an unsupervised method which considers all images as captured by a unique camera. Differently, our method is supervised and restricted to images captured by the same camera, which means that each prototype is a representative appearance model for a sample image and can be computed by a intra-camera similarity model.

The procedure to discover prototypes, illustrated in Figure 3, works as follows. Given a person p in the testing set (probe or gallery), captured by a known camera c_1, and a set of training samples, denoted by T, captured by two cameras c_1 and c_2. It is possible to obtain training images from the same camera from which p was captured (c_1), the subset $T_{c_1} \subset T$, represented as $T_{c_1} = (t_1, ..., t_M)$, where M is the number of individuals on the training set. The goal is to compute a list of similarities between the person p and the individuals in T_{c_1}, denoted as $L_p = (l_{p1}, ..., l_{pM})$, in which the higher the value of l_{pi}, the more similar to p it is. Once the L_p is computed, the top k elements in L_p, denoted by $T^p_{c_1} = \{t_{p1}, ..., t_{pk}\}$, are selected as the most similar individuals to the person p. Since person p was captured by camera c_1 and we are interested in finding his/her identity in camera c_2, we use the corresponding images of $T^p_{c_1}$ in camera c_2 as our prototypes ($T^p_{c_2} = \{\hat{t}_{p1}, ..., \hat{t}_{pk}\}$) to avoid the artifacts caused by different camera conditions.

The similarity models analysed in this work are: (1) Euclidean distance, (2) PLS one-against-all model [24] and (3) One-shot Similarity (OSS) [29].

Euclidean distance. In this approach all features are equally important and the similarity between the person p and the training image $t_i \in T_{c_1}$ is calculated as $l_{pi} = ((t_i - p)(t_i - p)^T)^{-1}$.

PLS one-against-all. Let $I_{c_1} = \{i_1, i_2, ..., i_N\}$ be a set of N individuals captured by the camera c_1. We intend to compute the similarity between $i_j \in I_{c_1}$ and person p captured by camera c_1. This process can be divided in appearance modeling and similarity computation. To model each individual, a PLS model for $i_j \in I_{c_1}$ is learned using its feature descriptors as the positive class and the remaining $I_{c_1} \setminus i_j$ as negative class. The result of the PLS model is the regression coefficients $\beta_j = \{\beta_1, \beta_2..., \beta_D\}$, in which D is the size of the feature vector. The idea is that these coefficients will focus on attributes that will differentiate the subject i_j from the other subjects in I_{c_1}. Finally, to compute the similarity between the individual $i_j \in I_{c_1}$ and p, we need project the features of p onto β_j, regression coefficients estimated for individual i_j.

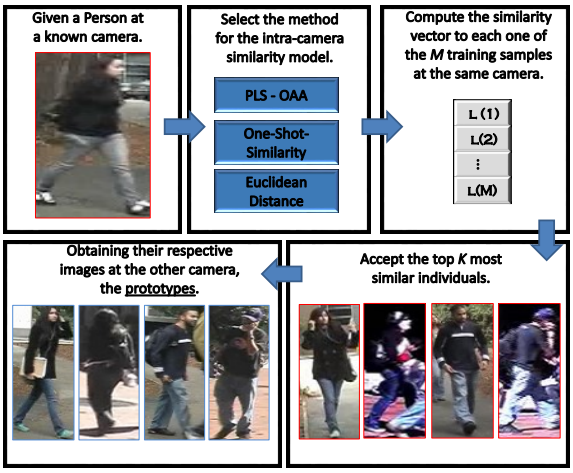


Figure 3. Prototype discovery. For a given person p captured by camera c_1, the similarity to training image captured by c_1 can be computed using an intra-camera similarity model. The k most similar individuals (k is a parameter), are selected and their respective image pairs captured by camera c_2 are used as the prototype of p.

One-shot similarity. In the OSS model, the similarity between a given person and a training sample captured by the same camera c_1, in our case p and $t_i \in T_{c_1}$, can be computed by training models using a fixed set of negative samples, denoted as background B, and then applying each one at the opposite model [29]. The set B must be disjoint w.r.t. the person p and the training samples T_{c_1}, being selected as the training samples $T \setminus T_{c_1}$. Two PLS models are learned, the person model uses p as positive class and B as negative, while in the training sample model, the individual t_i is employed as positive. The similarity between p and the $t_i \in T_{c_1}$ is computed as $l_{pi} = (p\beta^T_t + t_i\beta^T_p)/2$, where β_t and β_p are the learned regression coefficients by the PLS model of training sample and person, respectively.

3.2. Prototype Modeling

Once the prototypes have been computed for a person p captured by the camera c_1, they are used to construct rep-

resentative models. The motivation of using the elements in $T^p_{c_2}$ instead of p directly is two fold: (1) the images in $T^p_{c_2}$ indirectly deals with the transfer of features at different camera conditions and background interference; (2) similar subjects are assumed to be in $T^p_{c_2}$ and are labeled at the same class, which together with the large number of positive samples constructs better PLS models.

Considering that $T_{c_2} \setminus T^p_{c_2}$ denotes the remaining elements in training set captured by the camera c_2, a PLS model is learned considering $T^p_{c_2}$ as the positive class and $T_{c_2} \setminus T^p_{c_2}$ as negative.

3.3. Sample Ranking

In the previous sections, we considered how to model subjects using their prototypes. These models are then used to construct similarity ranks. In this section, on the other hand, we will consider two different types of rankings: one based on the gallery (modeled in the training) and one based on the probe (modeled during the testing).

Gallery-based ranking. In this approach, the similarity between the gallery images and the probe, denoted by L_p, is calculated using the set of gallery images to find the prototypes. For each $g_j \in G$ captured by a camera c_1, the prototypes $T^{g_j}_{c_2}$ are computed as described in Section 3.1 and the PLS model is learned as detailed in Section 3.2. It is important to emphasize that these models are learned offline, during training. Then, given a probe image p, represented by a feature vector $f_p \in \mathbb{R}^{1\times D}$, the similarity between p and a gallery image g_j, denoted by $l_{pj} \in L_p$, is obtained by $l_{pj} = p\beta^T_j$, where β_j are the regression coefficients learned using the PLS model. The final ranking is then obtained sorting the elements in L_p in descend order (the larger the value of l_{pj}, the higher is the similarity between g_j and p).

Probe-based ranking. A similar process is used here. However, instead of learning the prototypes from the gallery images, the problem is addressed as discovering the prototypes and learning the PLS model for the probe. That is, given a probe image p, the prototype is computed and a single PLS model is learned using the samples in the prototype as positive class. The similarity between each gallery $g_j \in G$ and p, denoted as $l_{pj} \in L_p$, is computed by $l_{pj} = g_j\beta^T$, where β is the regression coefficients learned using PLS.

Rank aggregation. Rank aggregation is widely employed in literature to integrate different rank results in an unbiased manner. This is a problem shared for diverse scientific areas such as web [5] and content-based image retrieval [19]. To the best of our knowledge, this is the first time that person re-identification is formulated as a rank aggregation problem. Our assumption is that the PLS models obtained from gallery and probe images are complementary and can be used to construct a unique rank of similarities. In addition, we also apply rank aggregation to integrate results achieved by the proposed method with results achieved by the method based on distance metric learning proposed in [30]. According to the results, such aggregation achieves the best results up to now in the literature, in two widely used Re-ID data sets.

Different rank aggregation methods will be analyzed in the experiments: Max, Mean, Robust Rank Aggregation (RRA) [11] and Stuart [26]. In Max, the unified similarity list L is computed as $l_j = \max_{1\leq j\leq n}\{l_{gj}, l_{pj}\}$. Then, these values are used to construct a final ranking, where the higher value in L is the first element and so on. Differently, in Mean, the similarity is given by $l_j = (l_{gj} + l_{pj})/2$. RRA and Stuart methods were proposed to aggregation of noisy list of genes and have as the main characteristic the robustness to noisy information and scalability. Therefore, they have the necessary requisites for aggregation in the Re-ID problem.

4. Experimental Results

This section evaluates the proposed approach and compares it to other state-of-the-art methods using VIPeR and PRID 450S data sets. First, we describe the data sets and the experimental setup used. Then, we evaluate the proposed prototype discovery (Section 4.1) and the approaches to rank the results (Section 4.2). Finally, in Section 4.3, we compare our approach to other methods in literature and demonstrate results achieved by considering the rank aggregation of different methods.

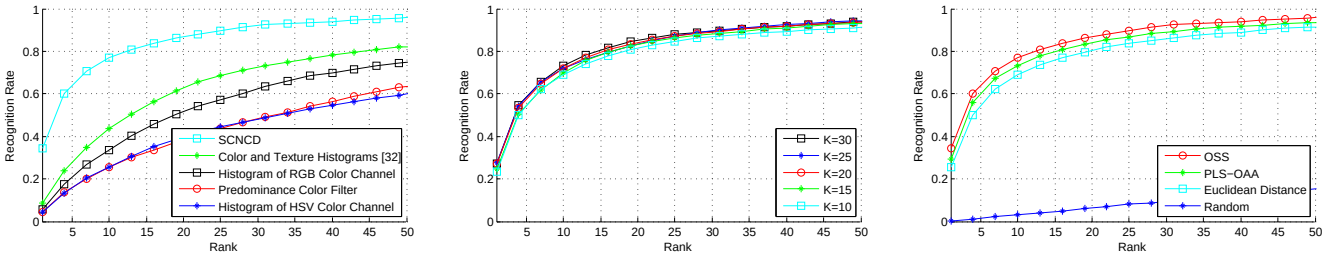
VIPeR Dataset [6]. This is a challenge dataset for Viewpoint Invariant Pedestrian Recognition (VIPeR)[2]. It contains 632 image pairs captured by two different outdoor cameras located in an academic environment, in which each subject appears once in each camera. Some examples of VIPeR dataset are shown in Figure 1. The variations are mostly caused by viewpoint changes, illumination and image quality. We select this dataset because it covers scenarios that can be found in real world applications and it provides pairs of labeled images from two non-overlapping cameras.

PRID 450S Dataset [23]. PRID 450S[3] is a new dataset with 450 single-shot pairs of images captured by two spatially disjoint surveillance cameras capturing pedestrians (between 100 and 200 pixels tall). The main challenges are related to changes in viewpoint, pose, camera characteristic as well as significant differences in background and illumination. Some examples are depicted in Figure 5.

Experimental Setup. As in the majority of the works that use VIPeR and PRID 450S, we randomly partitioned the

[2]Available at: http://vision.soe.ucsc.edu/?q=node/178

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



(a) (b) (c)

Figure 4. Analysis of different aspects of the method. (a) employment of different feature descriptors; (b) number of individuals in prototype subset; (c) different methods used to discover the prototypes.

Figure 5. Examples from PRID 450S dataset. The top row shows images captured by camera c_1 and the bottom row shows their respective image pairs captured by camera c_2.

datasets in training and testing subsets, being half of the image pairs to each subset. In the testing subset, images from one camera are considered as gallery and images from the other camera are considered as probe. The results are reported using Cumulated Matching Characteristic (CMC) curves showing the average of results obtained from 10 trials - a common procedure to achieve more stable results.

4.1. Analysis of Prototype Discovery

Figure 4 shows different CMC curves for parameters that affect the prototype discovering method. First, Figure 4(a) compares different features considered in this work: SCNCD, described in Section 2, predominance color filter [4], the mixture of color and texture histograms[4][32] and the histogram of RGB and HSV color channels. According to the results, the SCNCD outperforms the other features in our approach. Therefore, we will employ this method in our experiments.

The second experiment, reported in Figure 4(b), shows that the proposed method seems to be robust to different sizes of prototypes. In addition, increasing the size beyond 25 individuals has not increased the performance. Therefore, we have selected the parameter $k = 25$ for the remaining of the experiments.

Finally, the results reported in Figure 4(c) compare different approaches employed for prototype discovery, as discussed in Section 3.1. As we can see, a random choice of the prototypes achieved the worst results, demonstrating the importance of creating representative prototypes. Among the other three methods considered, the OSS achieved the best results and will be used in the remaining experiments.

[4]Available at: http://sist.sysu.edu.cn/ zhwshi/ilids.html

4.2. Analysis of Ranking Methods

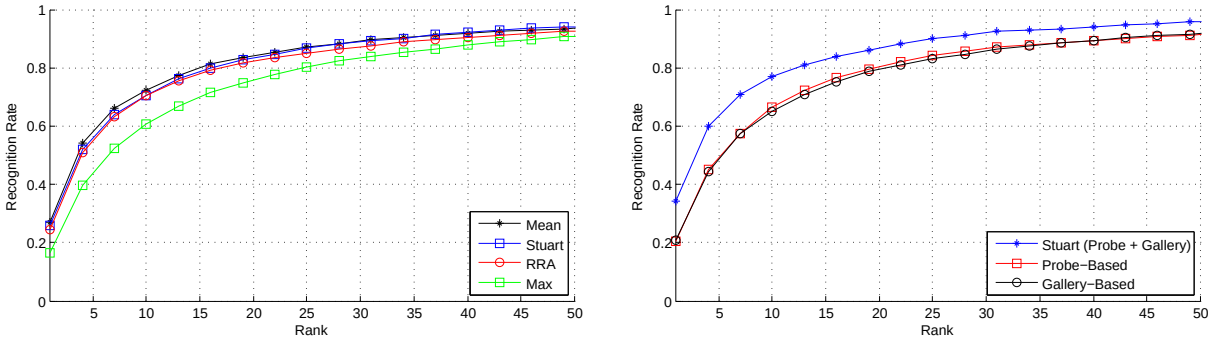
In Figure 6(a), we analyze different methods to addressing the rank aggregation problem. These methods receive the ranks as input (RRA and Stuart) or the models output (Mean and Max) and return a unified rank that better combines both predictions. According to Figure 6(a), the worst result was achieved using only the maximum value, while more complex methods such as RRA and Stuart performed slightly better than the Mean. In addition, Stuart method presented the best overall performance and, therefore, it will be employed in the remaining experiments.

As described in Section 3.3, we proposed the gallery-based and probe-based approaches and their aggregation. According to results shown in Figure 6(b), the probe- and gallery-based methods achieved similar results, indicating that models created using prototypes computed for each subject in the gallery and prototypes computed for a single probe sample achieve very similar results with the difference that learning process in the gallery-based ranking is offline, while probe-based must be addressed online.

The most interesting result according to Figure 6(b) was achieved by aggregating the ranks achieved by both approaches showing they are complementary and that using both rankings as input to construct an unified ranking significantly improves the results. As consequence of its superior performance, this aggregation of probe- and gallery-based rankings will be regard as our proposed method in the next section.

4.3. Comparisons

We perform a comparison of our method with others state-of-the-art approaches in VIPeR and PRID 450S datasets. The method proposed in Schwartz and Davis [25] was employed as our baseline, since it was the first to



(a) (b)

Figure 6. Analysis of ranking methods. (a) different methods to address the rank aggregation; (b) different approaches to rank the data.

apply the PLS in the Re-ID problem. However, to be a fair comparison, instead of using their original features, we used their approach with the SCNCD. We also compared a widely used distance metric learning-based method, KISSME, with the features proposed by Koestinger *et al.* [10] and with the SCNCD [30].

In Figure 7(a), the PLS curve is our baseline and presents the results using the SCNCD features and addressing the problem as Schwartz and Davis [25]. The lower performance can be justified by the different cameras conditions and the higher inter-class similarities. The method in [10] was by much outperformed by [30] in our experiments. This difference can be credited to the features since both employed KISSME method, which reinforces the superior performance of the SCNCD. The proposed method achieved results very similar to [30] (to the best of our knowledge, [30] achieves the best results in the literature).

While the method proposed in [30] is based on distance metric learning, the method proposed here is an appearance-based method. Therefore, they can be regarded as different views of the same problem and their results can be aggregated in a unique rank, represented by the curve Stuart (Probe + Gallery) in Figure 7(a). It is possible to see that aggregating both methods increased the overall performance significantly, obtaining state-of-the-art results, e.g., achieving 90% of recognition rate at rank 16, which has not been achieved before rank 25 in previous works.

Figure 7(b) shows the same curves for PRID 450S data set. We did not perform the evaluation of the features proposed by Koestinger *et al.* [10], since they are available only for VIPeR. In addition, the evaluation of SCNCD for PRID 450S was conducted using the code provided by Yang *et al.* [30] to extract their features, which can explain any divergence with their reported results. Figure 7(b) confirms the superior performance obtained by the Stuart (Probe + Gallery), also increasing the top-1 rank in 10 percent points as in the VIPeR dataset.

5. Conclusions and Future Work

This paper proposed a novel method for person re-identification that indirectly handles different camera conditions and higher inter-class similarity constructing discriminative models using representative subsets of individuals, the prototypes. To obtain more stable representations, we used the powerful statistical method PLS to weight features and the OSS to construct intra-camera similarity models. Complementary ranking were constructed for gallery and probe image and unified achieving results similar to state-of-the-art based on distance metric learning methods. To the best of our knowledge, this is the first time that an appearance based model reaches this standard. Furthermore, experimental results demonstrated that, as each method can be seen as a different look at the same problem, they can be aggregated, increasing the results by as much as 10 percent points in the VIPeR and the PRID 405S data sets.

As future directions, we intend to analyze the proposed method in other datasets and generalize it to perform at systems with more than two cameras. In addition, since the scalability of the proposed method is a concern, due the similarity computation for the probe image, a scalable approach will be evaluated.

Acknowledgments

The authors would like to thank the Brazilian National Research Council – CNPq (Grant #487529/2013-8) and FAPEMIG.

References

[1] A survey of approaches and trends in person re-identification. *Elsevier IMAVIS*, 32(4):270.

[2] L. An, M. Kafai, S. Yang, and B. Bhanu. Reference-based person re-identification. In *IEEE AVSS*, pages 244–249, 2013.

[3] Y. Cai and M. Pietikäinen. Person re-identification based on global color context. pages 205–215, 2010.

[4] C. R. S. Dutra, T. Souza, R. Alves, W. R. Schwartz, and L. R. Oliveira. Re-identifying people based on indexing structure

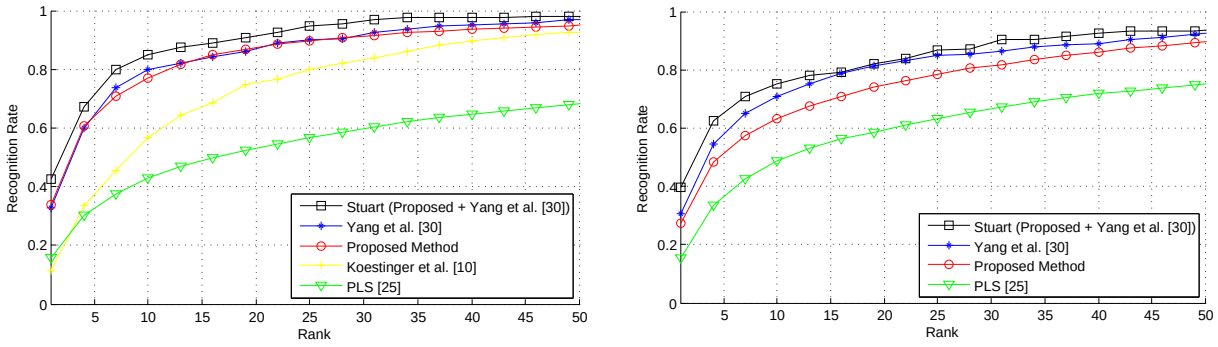


Figure 7. Analysis of different methods. (a) VIPer dataset; (b) PRID 450S dataset.

and manifold appearance modeling. In *SIBGRAPI*, pages 1–8, 2013.

[5] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *WWW*, pages 613–622, 2001.

[6] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *IEEE PETS*, 2007.

[7] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *ECCV*, pages 262–275, 2008.

[8] M. Hirzer, P. M. Roth, M. Koestinger, and H. Bischof. Relaxed pairwise learned metric for person re-identification. In *ECCV*, 2012.

[9] O. Javed. Appearance modeling for tracking in multiple non-overlapping cameras. In *IEEE CVPR*, pages 26–33, 2005.

[10] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *IEEE CVPR*, 2012.

[11] R. Kolde, S. Laur, P. Adler, and J. Vilo. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*, 28(4):573–580, 2012.

[12] C.-H. Kuo, S. Khamis, and V. Shet. Person re-identification using semantic color names and rankboost. In *Proc. WACV (2013)*, pages 281–287, Jan 2013.

[13] I. Kviatkovsky, A. Adam, and E. Rivlin. Color invariants for person reidentification. *IEEE TPAMI*, 35(7):1622–1634, 2013.

[14] W. Li and X. Wang. Locally aligned feature transforms across views. In *IEEE CVPR*, pages 3594–3601, 2013.

[15] C. Liu, S. Gong, and C. C. Loy. On-the-fly feature importance mining for person re-identification. *PR*, 47(4):1602–1615, 2014.

[16] C. Liu, S. Gong, C. C. Loy, and X. Lin. Person re-identification: What features are important? In *ECCVW*, 2012.

[17] L. Ma, X. Yang, and D. Tao. Person re-identification over camera networks using multi-task distance metric learning. pages 3656–3670, 2014.

[18] N. Martinel and C. Micheloni. Classification of local eigen-dissimilarities for person re-identification. *IEEE SPL*, 22(4):455–459, 2015.

[19] D. C. G. a. Pedronette and R. d. S. Torres. Image re-ranking and rank aggregation based on similarity of ranked lists. In *Springer CAIP*, pages 369–376, 2011.

[20] F. Porikli. Inter-camera color calibration by correlation model function. In *IEEE ICIP*, volume 2, pages II–133–6 vol.3, 2003.

[21] B. Prosser, W.-S. Zheng, S. Gong, T. Xiang, and Q. Mary. Person re-identification by support vector ranking. *BMVC*, 1(3), 2010.

[22] R. Rosipal and N. Krmer. Overview and recent advances in partial least squares. In *Springer SLSFS*, pages 34–51, 2006.

[23] P. Roth, M. Hirzer, M. Kstinger, C. Beleznai, and H. Bischof. Mahalanobis distance learning for person re-identification. In *Person Re-Identification*, Springer ACVPR, pages 247–267. 2014.

[24] W. Schwartz, H. Guo, J. Choi, and L. Davis. Face identification using large feature sets. *IEEE TIP*, 21(4):2245–2255, 2012.

[25] W. R. Schwartz and L. S. Davis. Learning discriminative appearance-based models using partial least squares. In *SIBGRAPI*, pages 322–329, 2009.

[26] J. M. Stuart, E. Segal, D. Koller, and S. K. Kim. A gene-coexpression network for global discovery of conserved genetic modules. *science*, 302(5643):249–255, 2003.

[27] D. Tao, L. Jin, Y. Wang, Y. Yuan, and X. Li. Person re-identification by regularized smoothing kiss metric learning. *IEEE TCSVT*, 23(10):1675–1685, 2013.

[28] R. Vezzani, D. Baltieri, and R. Cucchiara. People reidentification in surveillance and forensics: A survey. *ACM CSUR*, 46(2):29:1–29:37, 2013.

[29] L. Wolf, T. Hassner, and Y. Taigman. The one-shot similarity kernel. In *IEEE ICCV*, 2009.

[30] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li. Salient color names for person re-identification. In *ECCV*, pages 536–551. 2014.

[31] W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by probabilistic relative distance comparison. In *IEEE CVPR*, pages 649–656, 2011.

[32] W.-S. Zheng, S. Gong, and T. Xiang. Reidentification by relative distance comparison. *IEEE TPAMI*, 35(3):653–668, 2013.