CBRA: COLOR-BASED RANKING AGGREGATION FOR PERSON RE-IDENTIFICATION

Raphael Felipe de Carvalho Prates, William Robson Schwartz

Department of Computer Science, Universidade Federal de Minas Gerais, Brazil

ABSTRACT

The problem of automatically tracking a pedestrian within camera networks with non-overlapping field-of-view, known as person re-identification, is a challenging task with still suboptimal results. Different features have been proposed in the literature, specially colors which achieved the best results when fused in a unique feature representation. Despite being better than considering individually, the fusion still does not explores all the feature discriminative power. Therefore, we propose the use of rank aggregation to improve the results. In this paper, we address the person re-identification problem using a Color-based Ranking Aggregation (CBRA) method, which explores different feature representations to obtain complementary ranking lists and combine them using the Stuart ranking aggregation method. The obtained experimental results demonstrate a great improvement in state-of-the-art, reaching top-1 rank recognition rates of 50.0% and 56.9% in the ViPER and PRID450S data sets, respectively.

Index Terms— person re-identification, ranking aggregation, color features, smart surveillance.

1. INTRODUCTION

Camera networks have been employed in public and private locations, usually with non-overlapping field-of-view to monitor wider areas at a reduced cost. They are helpful to security personnel to monitor pedestrian behaviors. Person re-identification (Re-ID) uses appearance models alone (visual cues) to automatically tracking persons in camera networks and has attracted many researches in the past years [1]. It is a challenging task because traditional biometric cues such as face and gait are unreliable in low-resolution images and the appearance changes according to different camera viewpoints, illuminations, poses and partial occlusions [2].

In person re-identification, there are gallery images captured by a camera $c_1$ which must be ranked according to the similarity to a probe image captured by a camera $c_2$. Training samples from cameras $c_1$ and $c_2$ are available to compute a similarity model, usually based on color and texture characteristics present in clothing [1]. Different color and texture features have been employed in the person Re-ID problem, such as color histograms [3, 4, 5, 6], color names [5, 7], Local Binary Patterns (LBP) [8] and Gabor and Schmidt filters [3]. Color features are regarded as the most important cue in Re-ID problem [7, 3]. Yang et al. [9] demonstrated that different color models can compensate each other when fused (concatenated) in a unique high-dimensional feature vector. Color names are explored in Yang et al. [7] using Saliency Color Names based Color Descriptor (SCNCD), which consists in a probability distribution over 16 colors names.

Machine learning techniques are employed to learn the similarity models from training samples. In general, a high-dimensional vector is constructed by the fusion of distinct features and the feature importance is computed using AdaBoost [3], PLS [6], RankBoost [5] and RankSVM [10]. Distance metric learning presents great success in Re-ID problem, it learns an affine-transformation that respects a pairwise constraint, keeping closer pairs of the same person [11, 12, 13]. For instance, KISSME [4] is a simple but effective distance metric learning method based on statistical inference which has been widely adopted. Differently, in Efficient Imposter-based Metric Learning (EIML) [14], an efficient approximation for metric learning explores impostors informations. However, these approaches consider that feature importance as a dataset property, which, according to [15, 16], is in fact dependent of the pairs of images being analyzed.

In this paper, we consider each extracted feature (e.g., a color channel histogram) as a judge that produces a ranking list based on its preferences. Then, the person Re-ID can be seen as a ranking aggregation problem [17] in which the goal is to compute a unique rank based on several judges preferences, approach widely explored in meta-search engine for web [17] and more recently in content-based image retrieval [18].

Kuo et al. [5] considered each feature as a different ranker (weak ranker) which are combined to form a strong ranker using their proposed RankBoost algorithm. The similarity measures returned by eight color features are also considered by Liu et al. [19] to combine the ranking scores in a unique ranking list using Ensemble of Color Models (ECM). Liu et al. [20] proposed a post-rank optimization method which has the drawback of needing the user feedback. A ranking aggregation and re-ranking algorithm (RRA) was proposed by Ye et al. [21] to handle different ranking lists generated by local and global features. Differently, Leng et al. [22] proposed a bi-directional re-ranking using visual content and contextual information extracted from top rank positions. While they explore only the top-k ranking list similarity, we, however, consider the full ranking list and employed a widely used rank aggregation method.

Finding the best ranking is very expensive, a NP-hard problem,
and the proposed algorithms in literature are approximations and heuristics [23]. To achieve the aggregated ranking list we chose the Stuart [24] algorithm which, based on order statistics, finds a relevant output even when there are irrelevant and noise inputs. In addition, since color features outperform shape and texture [3], different features based on color channel histograms and salient color names [7] are employed in similarity models learned using KISSME [4]. Therefore, our method is a Color-based Ranking Aggregation (CBRA) approach that employs a distance metric learning to compute the similarity models. Experimental results demonstrate great improvement in state-of-the-art, reaching top-1 rank recognition rates of 50% for VIPER and 56.9% for PRID450S, two widely adopted and challenging data sets.

2. PROPOSED METHOD

In this section, we describe our method, schematically represented in Figure 2. For each sample image, our method extracts two types of color features: color histograms and SCNC with part-based models [25] which captures different body information, such as head, upper torso, legs. Three different feature representations (Section 2.1) are employed to compute ranking lists regarded as complementary. When correctly combined, these ranking lists are expected to better explore the strength of each color feature. Finally, different strategies for ranking lists combination are analyzed (Section 2.2).

2.1. Feature Representation

Using color histograms and SCNC feature types, three levels of feature representation are built: individual, type-based and fusion-based. In the individual representation, each extracted feature composing a feature type is considered alone, while in the type-based, all features that belong to the same feature type, e.g., color histograms, create a unique feature representation (one for each feature type). Finally, in the fusion-based representation, all extracted features are concatenated resulting in a high-dimensional feature vector that mixes different feature types.

We assume that \( F = [D_1, D_2] \) is the fusion-based representation constructed by concatenating two types of features, where \( D_1 = [d_{11}, d_{12}, \ldots, d_{1k}] \) and \( D_2 = [d_{21}, d_{22}, \ldots, d_{2l}] \) are these feature types formed by the fusion of \( k \) and \( l \) different extracted features \( \langle d_{ij} \rangle \), respectively. For instance, if \( D_1 \) is the color histogram feature type, we can have that \( d_{11} \) and \( d_{12} \) are two extracted features using RGB and HSV color models, respectively.

The three feature representations \( F = \{F\}, D = \{D_1, D_2\} \) and \( d = \{d_{11}, \ldots, d_{1k}, d_{21}, \ldots, d_{2l}\} \) are the fusion-based, type-based and individual, respectively. They are employed to compute different ranking lists of the gallery images, which can be aggregated using the Stuart method to improve the results obtained individually. The next section describes different ranking aggregation strategies.

2.2. Aggregation Strategies

Let \( F', D' \) and \( d' \) be the feature representation with dimensionality already reduced using PCA for fusion-based, type-based and individual representations, respectively. We assume that \( R \) is the set of all possible ranking lists of the gallery images (permutations), and that \( R_{F'}, R_{D'} \) and \( R_d \) are subsets of \( R \) computed using KISSME algorithm and the feature representation \( F', D' \) and \( d' \), respectively. We also denote by \( \oplus \) the aggregation operation, for instance if \( r_n = r_1 \oplus r_2 \oplus \cdots \oplus r_{n-1} \), then \( r_n \) is a ranking list computed by the aggregation of ranking lists from \( r_1 \) to \( r_{n-1} \).

In this paper, we are interested in computing the element \( r \in R \) using the aggregation of the ranking lists present in \( R_{F'}, R_{D'} \) and \( R_d \). Three strategies are considered: feature-specific, cascade and global, defined as follows.

**Feature-Specific.** In feature-specific strategy, the feature types are considered isolated using \( D' \) and \( d' \) features representations. Assuming that the ranking lists \( R_{D'} = \bigcup R_{D'_i} \) and that \( R_d = \{R_{d_{11}}, R_{d_{12}}, \ldots, R_{d_{1k}}\} \cup \{R_{d_{21}}, R_{d_{22}}, \ldots, R_{d_{2l}}\} \), the final ranking list is computed by \( r_i = R_{d_{11}} \oplus R_{d_{12}} \oplus \cdots \oplus R_{d_{1k}} \oplus R_{d_{21}} \oplus R_{d_{22}} \oplus \cdots \oplus R_{d_{2l}} \), where \( n = k \) if \( i = 1 \) and \( n = l \), otherwise. Since we have two feature types, two final ranking lists \( (r_1 \text{ and } r_2) \) are computed.

**Global.** In the global strategy, all the ranking lists are aggregated to
produce a unique final ranking list \( r' = R_{a} \oplus R_{D} \oplus R_{p} \).

**Cascade.** The cascade strategy executes the aggregation in a cascade, using the final ranking lists computed in feature-specific and global approaches to compute a new ranking list \( r_c = r_1 \oplus r_2 \oplus r_3. \)

### 3. EXPERIMENTAL RESULTS

In this Section, we evaluate the proposed CBRA method and compare 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 employed. Then, we evaluate the different color features employed (Section 3.1) and combination strategies of ranking lists are analyzed (Section 3.2). Finally, in Section 3.3, we compare our approach to other methods in literature.

**VIPeR Dataset [3].** This is a challenge dataset for Viewpoint Invariant Pedestrian Recognition (VIPeR). 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.

**PRID 450S Dataset [26].** PRID 450S \(^2\) consists in 450 image pairs of pedestrians (between 100 and 200 pixels tall). This dataset has been recently released and is regarded as more realistic. 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 1.

**Experimental Setup.** We used a common procedure in literature to achieve more stable results in VIPeR and PRID450S which consists in randomly partitioning the dataset in training and testing subsets of equal sizes. 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. When computing histograms of color feature types, we set the number of bins to 32. The salient color names are computed using the same parameters configuration defined by Yang *et al.* [7] in image-only representation, which attributes probabilities to only the nearest 5 color names. Furthermore, when reporting the Area Under the Curve (AUC), we consider the rankings in range [1,30].

---

1Available at: http://vision.soe.ucsc.edu/%7e/q=node/178  
2Available at: https://irr.ics.tugraz.at/download.php

### 3.1. Color Feature Types

Now, we will discuss the results obtained when considered the colors feature types (color histograms and salient colors names) alone.

#### 3.1.1. Color Histograms

The different color models considered in this work for color histogram computation include the original RGB, normalized rgb, \( l_1 l_2 l_3 \) [27], HSV and YUV. The results are presented for the individual and type-based (fusion of all color models histograms) feature representations. Furthermore, the feature-specific rank aggregation strategy using color histograms is included since it also considers each feature type alone. Feature dimension of individual and type-based representations are reduced to 34 and 70, respectively, using PCA (same done in Yang *et al.* [7]).

Figures 3(a) and 4(a) show different CMC curves for the color histogram feature in VIPeR and PRID450S datasets. According to the results, the HSV color model presents superior performance when considered only individual feature representation, but the type-based representation outperforms any individual feature representation. However, the most interesting result is obtained using the feature-specific ranking aggregation strategy, which in both datasets represents a significant improvement in the AUC (0.17 in VIPER and 0.36 in PRID450S).

#### 3.1.2. Salient Color Name Based Color Descriptor (SCND)

In the SCND, we considered the original RGB, normalized rgb, \( l_1 l_2 l_3 \) [27] and HSV color models. These models are employed to obtain the individual and type-based representations which jointly with feature-specific ranking aggregation strategy produce the results related to only SCND feature type. The parameters for PCA are employed as in Section 3.1.1.

The CMC curves for VIPeR and PRID450S datasets are presented in Figure 3(b) and 4(b). Considering the individual feature representation, HSV and RGB achieved the best results for VIPeR and PRID450S datasets, respectively. However, in the overall comparison, the feature-specific ranking aggregation strategy outperformed the other approaches in both datasets, with significant improvement in AUC (0.15 in VIPER and 0.29 in PRID450S). It is interesting to notice that the SCND feature type presented superior performance when compared with color histograms. In feature-specific ranking aggregation strategy, for instance, the improvement in AUC is 0.02 and 0.07 for VIPER and PRID450S, respectively.
3.2. Ranking Aggregation Strategies

A common strategy in person Re-ID literature is to combine features to improve results. In this paper, we evaluate different approaches for performing such combination, which includes the ranking aggregation strategies described in Section 2.2 (feature-specific, global and cascade) and the classical fusion-based feature representation (concatenation of all features in a unique high-dimensional feature vector).

Figures 3(c) and 4(c) show CMC curves obtained comparing the different feature combination approaches from VIPeR and PRID450S datasets. According to these results, the fusion-based approach, despite of being widely adopted in the literature, presents the worst results in both datasets. It is outperformed even by the feature-specific ranking aggregation strategies that consider each feature type separately. The cascade strategy presents the best performance in both datasets, being slightly superior than the global, and will be regarded as our proposed method in the next section.

3.3. Comparisons

We perform a comparison of our method with other state-of-the-art approaches in VIPeR and PRID 450S datasets. The method proposed in Yang et al. [7] is employed as our baseline since the feature types (SCNCD and color histograms) are almost the same (we did not use the automatically extracted foreground masks, which is the main difference). Furthermore, we compare our approach with some ranking optimization methods (RRA [21], ECM [19] and Leng et al. [22]) and with other state-of-the-art approaches (Yang et al. [9], EIML [14] and KISSME [4]). All the methods reported results in both datasets, with the exception of the RRA and Leng et al. [22] that report only on VIPeR.

In Table 1, the CMC performance comparisons between the proposed CBRA and other methods are presented for VIPeR dataset. Based on these results it is possible to conclude that the proposed method greatly outperforms the conventional methods in all the ranking positions, mainly in the initial ranking reaching an expressive 50% at top-1 rank which is 11.1 percentage points more than the best reported result (ECM - 38.9%).

Table 2 shows the CMC performance comparisons for the PRID450S dataset. The obtained results with the proposed method outperformed the state-of-the-art for all ranking positions. For instance, in top-1 rank, we obtained 56.9%, which is 15.0 percentage points more than the best reported result in literature (ECM - 41.9%).

<table>
<thead>
<tr>
<th>Method</th>
<th>VIPeR (p=316)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RRA [21]</td>
<td>35.0</td>
</tr>
<tr>
<td>Leng et al. [22]</td>
<td>23.8</td>
</tr>
<tr>
<td>ECM [19]</td>
<td>38.9</td>
</tr>
<tr>
<td>Yang et al. [9]</td>
<td>37.6</td>
</tr>
<tr>
<td>SCNCD [7]</td>
<td>37.8</td>
</tr>
<tr>
<td>EIML [14]</td>
<td>22.0</td>
</tr>
<tr>
<td>KISSME [4]</td>
<td>27.0</td>
</tr>
<tr>
<td>Proposed (CBRA)</td>
<td>50.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>PRID 450S (p=225)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECM [19]</td>
<td>41.9</td>
</tr>
<tr>
<td>Yang et al. [9]</td>
<td>40.6</td>
</tr>
<tr>
<td>SCNCD [7]</td>
<td>37.8</td>
</tr>
<tr>
<td>EIML [14]</td>
<td>35.0</td>
</tr>
<tr>
<td>KISSME [4]</td>
<td>33.0</td>
</tr>
<tr>
<td>Proposed (CBRA)</td>
<td>56.9</td>
</tr>
</tbody>
</table>

4. CONCLUSION AND FUTURE WORKS

In this paper, we formulate person re-identification as ranking aggregation problem. Due its superior performance, different color features were explored to construct complementary ranking lists. Three ranking aggregation strategies were analyzed determining the best performing configuration of feature representation and ranking aggregation strategy that jointly with the Stuart ranking aggregation method constitute our proposed Color-based Ranking Aggregation (CBRA) approach.

Experimental results demonstrate that CBRA significantly improves the state-of-the-art for all rank ranges and mainly for the top-1 rank, with a increase of 11.1 and 15.9 percentage points for VIPeR and PRID450S, respectively. As future works, we intend to evaluate the integration in our approach of shape and texture features.

5. ACKNOWLEDGMENTS

The authors would like to thank the Brazilian National Research Council – CNPq (Grants #477457/2013-4 and #487529/2013-8), the Minas Gerais Research Foundation - FAPEMIG (Grant APQ-00567-14) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES (the DeepEyes Project).
6. REFERENCES


