PREDOMINANT COLOR NAME INDEXING STRUCTURE FOR PERSON RE-IDENTIFICATION

Raphael Prates, Cristianne R. S. Dutra, William Robson Schwartz

Smart Surveillance Interest Group, Computer Science Department
Universidade Federal de Minas Gerais, Minas Gerais, Brazil

ABSTRACT

The automation of surveillance systems is important to allow real-time analysis of critical events, crime investigation and prevention. A crucial step in the surveillance systems is the person re-identification (Re-ID) which aims at maintaining the identity of agents in non-overlapping camera networks. Most of the works in literature compare a test sample against the entire gallery, restricting the scalability. We address this problem employing multiple indexing lists obtained by color name descriptors extracted from part-based models using our proposed Predominant Color Name (PCN) indexing structure. PCN is a flexible indexing structure that relates features to gallery images without the need of labelled training images and can be integrated with existing supervised and unsupervised person Re-ID frameworks. Experimental results demonstrate that the proposed approach outperforms indexation based on unsupervised clustering methods such as *k-means* and *c-means*. Furthermore, PCN reduces the computational efforts with a minimum performance degradation. For instance, when indexing 50% and 75% of the gallery images, we observed a reduction in AUC curve of 0.01 and 0.08, respectively, when compared to indexing the entire gallery.

***Index Terms*—** Person re-identification, color names, inverted lists, visual dictionaries, surveillance scalability

1 Introduction

Applied in surveillance [1], sporting events [2] and traffic monitoring [3], the person re-identification (Re-ID) is responsible for maintaining the identity of a large number of people being monitored in a camera network, in which the cameras not necessarily have superposed viewpoints [4]. Specifically, given the large number of images captured by a camera c_1 (gallery samples) and a probe image captured by camera c_2, the goal of person re-identification is to find matching candidates to the probe image. In supervised Re-ID, image pairs captured by distinct cameras are employed to learn machine learning models and improve the matching results. An extended description and discussion regarding this topic can be found in [4, 5].

Due to the low-resolution images captured by surveillance cameras, biometric cues such as face and iris are unavailable and the re-identification problem is addressed using individuals appearance models obtained by color, texture and shape information, which can be employed separately [6, 7] or as a combination to maximize the discriminability [8]. In addition, contextual information [9, 10] can be applied with appearance models to improve the results.

Color information is regard as the most important cue in the Re-ID problem [11, 12] and is extracted usually by either color histograms [12, 13, 8] or color names [14, 11] from part-based models, such as horizontal stripes [15]. For instance, Yang $et\ al.$ [11] improved the state-of-the-art results employing Salient Color Names based Color Descriptor (SCNCD), which probabilistically relates RGB color values with 16 color names[1].

In general, after a high-dimensional feature vector is extracted, the feature importance is computed using some machine learning technique such as AdaBoost [12], PLS [8], RankBoost [14] and RankSVM [16]. Distance metric learning based methods have achieved great success in Re-ID problem, since they learn an affine-transformation that respects a pairwise constraint, keeping closer pairs belonging to the same person [17, 18, 13].

Once the results obtained in state-of-the-art have considerably improved over the years, the efficiency needs to be considered for the Re-ID problem [19, 20]. On this regard, most of the works in literature require the comparison of each probe image with all possible candidates in gallery, restricting the system scalability. For instance, in a public monitored environment, the gallery set may contain thousands of individuals in a unique day. Camera geometry approaches limit the gallery size using complex probabilistic transition models learned using spatio-temporal relationship between camera pairs [21, 22, 23]. The main drawback of these methods is the huge amount of labelled training image from camera pairs. Differently, Dutra $et\ al.$ [19] addressed this problem using inverted lists that relate appearance descriptors to objects and does not require labelled training images. These lists are learned from gallery samples and are populated with subsets of gallery ids. Thus, when a probe sample is presented, just a subset of the gallery images is considered.

In this work, we propose a novel Predominant Color Name indexing structure, referred to as PCN, that similarly to the works presented in [19, 20] is computed using part-based models and multiple inverted lists. However, our indexing structure explores the SCNCD, taking advantage of its direct relation with color names. In addition, we extend the work of Dutra $et\ al.$ [19] by indexing multiple inverted lists for each image block, instead of a single list. As a result, it increases the number of lists returned for a given probe image. These inverted lists are efficiently combined through a ranking aggregation method and the trade-off between recognition rate and efficiency is analyzed using a state-of-the-art Re-ID algorithm proposed in Yang $et\ al.$ [11]. Experimental results demonstrate improved results when compared to more complex indexing approaches.

2 Proposed Method

Indexing gallery samples by inverted lists aims at attributing gallery subjects with similar characteristics to the same inverted list. For

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

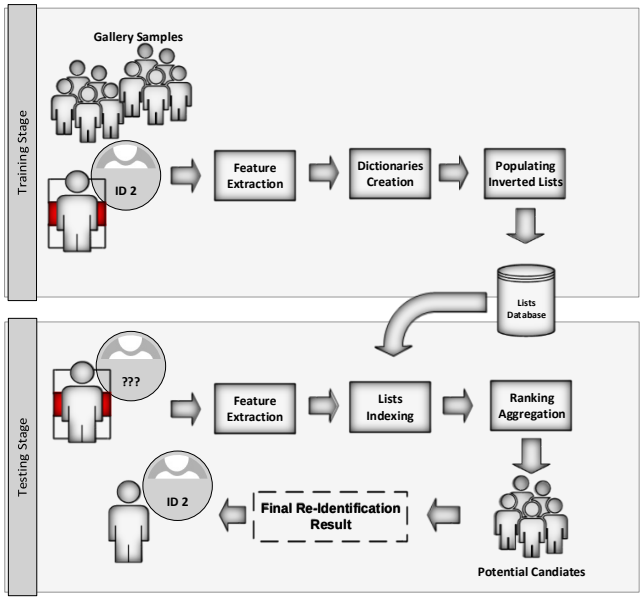

Fig. 1. Indexing structure overview. In the training stage, the features are extracted and employed in the dictionary creation process. Then, the similarity between gallery images and inverted list (codewords) is used to populate the inverted list with gallery ids, which are stored in a database for the testing stage. In testing stage, the similarity between a probe image and the codewords is used to index the inverted list and obtain the potential candidates. The next step is the assignment of a id between the potential matches to the probe sample using a more complex methods (this step, represent by a dashed-line block, is outside of our scope).

instance, we expect that two gallery persons wearing black shirts belong to a same list. This direct relation between semantic information, such as "black shirt", and an inverted list index motivates us to use SCNCD extracted from part-based models. While the former obtains the color information ("black"), the latter captures its location in the body ("shirt").

Our indexing approach is composed of training and testing, as illustrated in Figure 1. In both stages, the images are divided into m horizontal non-overlapping blocks that are considered independently. Initially, the features are extracted and a dictionary (a set of codewords) is created (or learned) for each block. Each codeword results in an inverted list, which are populated using gallery image ids. These lists are computed offline and stored in a database.

In the testing stage, the database provides inverted lists for each image block based on the similarity between feature descriptor and codewords, these lists contain the local candidates. Multiple local lists are combined into a unique global list containing the potential candidates to match the probe sample. Global list has gallery ids sorted by a ranking aggregation strategy, the Borda's method [24].

In the method [24], the Borda score for an element c in an inverted list i, represented by $B_i(c)$, is given by the number of elements above c and the total Borda score $B(c)$ is defined as $B(c) = \sum_{i=1}^{n} B_i(c)$, where n is the number of inverted lists returned for a given probe sample. Considering the gallery samples in increasing order of Borda score, we emphasize the best ranked subjects.

In this section, three indexing structures are considered: Predominant Color Name (PCN), *k-means* and *c-means*. The first is a novel indexing structure which explores the semantic color information in SCNCD (Section 2.1), while *k-means* and *c-means* [25] are based on *bag-of-words* [26] and are just modifications of the method proposed in [20] (Section 2.2). The main difference is in how the lists are created. While *k-means* and *c-means* employ unsupervised clustering algorithms, PCN determines them directly from the color names. It is important to emphasize that the focus of the indexing structure is to reduce the number of potential candidates needed with a minimum degradation of the performance. Thus, the assignment of unique id for the probe image can be done using the potential candidates and a more complex state-of-the-art Re-ID algorithm.

2.1 Predominant Color Names

In this indexing structure, the inverted lists are populated based on the p predominant color names, where p is a parameter. The main idea is that clothing characteristics are captured in the p larger values of SCNCD, while the remaining values are related to noise and background information. For instance, a person wearing blue pants probably presents the largest value at the histogram position related to color name "blue" in the block associated with "legs".

In SCNCD, each pixel is related to a probability distribution over 16 salient color names. These distributions are employed to compute a histogram of salient color names for each image block. Thus, for a given image i and block j, this histogram is represented by $D_{ij} = [d_1, d_2, ..., d_{16}]$. To increase the indexing robustness due to different camera conditions, instead of using the values in D_{ij} directly, these values are sorted in descending order and stored in S_{ij}, while the salient color name positions are represented by F_{ij}. In S_{ij} and F_{ij} representations, the p predominant color names values and indexes are located in the initial p positions, respectively.

Figure 2 presents an overview of the proposed PCN. In the training stage, feature vectors are extracted from image blocks and the p predominant color names are indexes to inverted lists where the gallery sample id is inserted. Since dictionary creation is costless, as we will discuss in the forthcoming paragraphs, it is omitted in Figure 2. These lists are stored for future queries at the testing stage.

For a given probe image, the feature vectors are extracted and the p predominant color names are used to query the inverted lists database, obtaining local candidates to match the probe. These candidates are used as input for the ranking aggregation algorithm, returning a unique global list of potential candidates. The following paragraphs explain how representations S_{ij} and F_{ij} are used to fill the inverted lists. The problem of obtaining a unique global list from multiple local lists is also addressed.

Training. In the training stage, we address the problem of constructing dictionaries of k codewords for each block and then mapping each gallery sample to one or multiple lists based on its feature descriptors. Since the inverted lists are ordered sequence of gallery ids, the position where the id is included in each list is computed based on some similarity criterion and using binary search algorithm.

We explore the direct association of feature descriptors and color names to consider each color name as a codeword, resulting in a block dictionary of k color names. In Figure 2, for instance, the first codeword (k_1) in all m block dictionaries is yellow, since yellow is the first color name. Each of the m dictionaries and k codewords will index subsets of gallery individuals, resulting in $m \times k$ inverted lists. The remaining question is how to relate the feature descriptors extracted from a block i with one or multiple codewords?

For a given gallery sample i, the feature descriptor of the j-th block, represented by D_{ij}, is extracted and the corresponding repre-

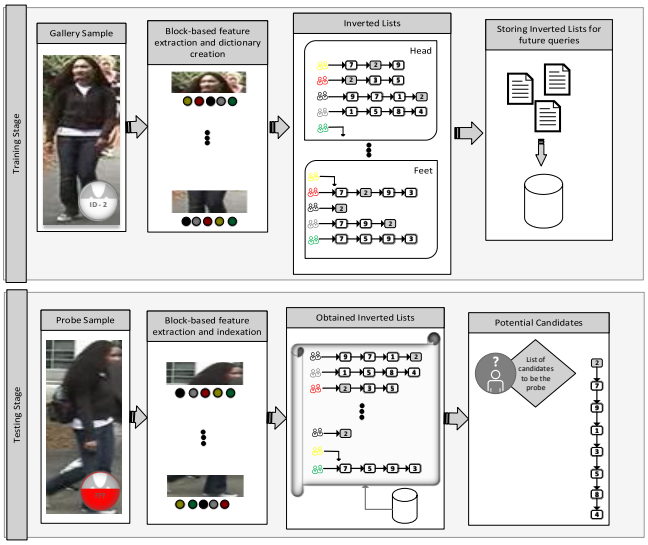

Fig. 2. Overview of the proposed Predominant Color Name (PCN) Indexing Structure. Initially, the gallery images are divided into blocks, which have their features are extracted and sorted in descending order. The p most predominant color names are then used to index and fill inverted lists with gallery ids. Inverted lists stored in a database are accessed in the testing stage based on the p predominant color names of each block of a probe sample. The local multiple lists are aggregated in a global list using Borda's method. In this Figure, we illustrate the case of $p = 3$.

sentations S_{ij} and F_{ij} are obtained. The p predominant color names are the initial p positions of F_{ij}, represented by $l = (f_1, f_2, ..., f_p)$. These p indexes in the vector l indicate in which inverted lists of the block j the $id\,i$ will be included. The position where the $id\,i$ will be included will depend on the corresponding value in S_{ij}. Considering the example in Figure 2, the gallery sample with $id = 2$ is inserted at the yellow inverted list (τ_{yellow}), since yellow is the predominant color name (f_1 = yellow) of the head's block. The id 2 is inserted at a position that maintains the yellow inverted list sorted. Notice that despite each insertion requires the update of Borda scores, it is just adding one in the pushed back individuals and the position value for the inserted individual.

Testing. For a given probe sample i, we first extract the feature descriptor D_{ij} for each block j and obtain the corresponding representation F_{ij} and S_{ij}. As in the training stage, the initial p positions in F_{ij} are used as inverted lists indexes, so that $l = (f_1, f_2, ..., f_p)$, and the elements present in the lists, denoted by τ_1, τ_2, ..., τ_p, are regarded as local candidates to match the probe image. In Figure 2, for instance, in block head of the testing stage, we have $f_3 = red$ and the respective local candidates ids are $\tau_3 = (2, 3, 5)$.

Since we have m blocks in the image, we obtain $m \times p$ lists which are aggregated using Borda's method into a global candidate list τ. Depending on the values of m and p, the list τ can contain all the gallery samples. However, since the elements in τ are sorted using Borda score, it is possible to establish a threshold t, considering as candidates just the initial t individuals. As we will demonstrate in the experiments, t determines the trade-off between efficiency (smaller t) and recognition rate (larger t).

2.2 Unsupervised Clustering

In this section, we will describe the unsupervised clustering algorithms employed to obtain the codewords from the feature descriptors: *k-means* and *c-means*.

Training. Inspired by the work of Dutra et al. [20], we also learn a dictionary for each block based on *bag-of-words* [26]. First, a given gallery sample i is divided into m horizontal blocks and, for each block j, a feature descriptor is extracted using SCNCD, represented by D_{ij}. All feature vectors extracted for the same block j are denoted by $\hat{D}_j = [D_{1j}, D_{2j}, ..., D_{nj}]$, where n is the number of gallery samples. For each block j, a dictionary of k codewords is computed using the *k-means* or *c-means* algorithm and as input the feature vectors extracted specifically from block j ($\hat{D}_j$). A set of $m \times k$ inverted lists are obtained considering the m blocks and k codewords. Depending on the chosen algorithm, the codewords will be different due the distinct objective functions. A drawback of these algorithms is that initialization conditions lead to different local optimal values (codewords).

Differently from Dutra et al. [20], for each gallery image i, we include the index i in all inverted lists of the block j, represented by τ_1, τ_2, ..., τ_k but the position in which the index i is inserted in each of the k lists will depend on the similarity between the feature descriptor D_{ij} and the respective codeword that generates the list. In *c-means*, this similarity is computed based on a membership function, while in *k-means* we employed the Euclidean distance.

Testing. For a given probe sample i, the feature vector of a block j (D_{ij}) is extracted and closest index list (codeword) is computed using Euclidean distance. The same process is repeated for all m blocks, and the closest lists, represented by $\tau = (\tau_1, \tau_2, ..., \tau_m)$, are aggregated using Borda's method.

3 Experimental Results

In this section, we describe the experiments executed to validate the proposed method using the widely employed VIPeR Dataset [27]. The results will be shown using the recognition rate as a function of the rank, a common approach to validate person re-identification approaches [5]. We also report the results using the Area Under the Curve (AUC), that despite being unusual, it is a good way of comparing different results using a unique measure. The AUC values are computed in the range [1:160] and are normalized to [0,1].

The experiments are divided in two parts. First, we analyze the indexing scheme using the training set considering different aspects, such as the indexing structure, the influence of aggregation strategy (Borda's method) and the choice of feature descriptors. Then, we apply the indexing scheme in the testing set and present a quantitative analysis of the performance degradation considering a state-of-the-art Re-ID algorithm proposed by Yang *et al.* [11].

Indexing Structure. Figure 3 presents the evaluation of different indexing structures employed: Predominant Color Name (PCN), *c-means*, *k-means* and random. The random strategy just selects randomly the codewords and is included in the experiments due the superior results reported in Dutra *et al.* [20]. The results reported with *c-means* and *k-means* use 10 and 15 clusters, respectively, which were the best estimated parameters. For the PCN, we present three curves, each considering different number of predominant color names, represented by the parameter p. According to the results, we conclude that the PCN with p equals 3 presented the best performance, retaining 90% of probe samples when 140 potential

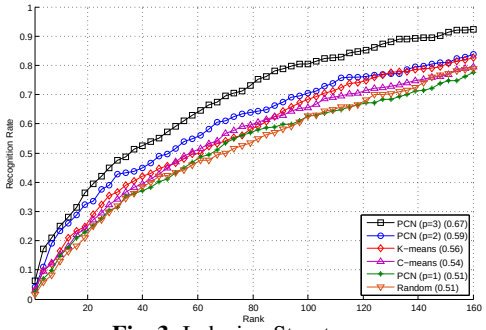

Fig. 3. Indexing Structures.

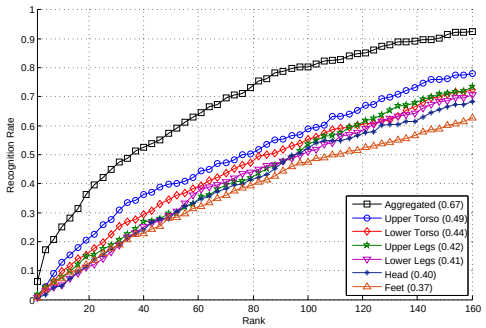

Fig. 4. Aggregation Strategy.

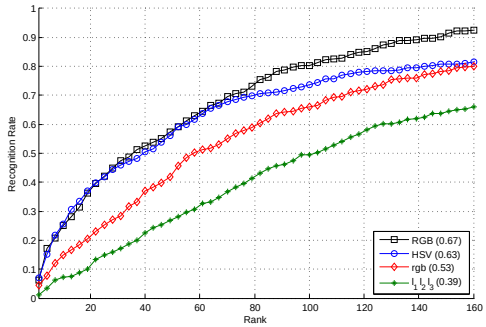

Fig. 5. Color Models.

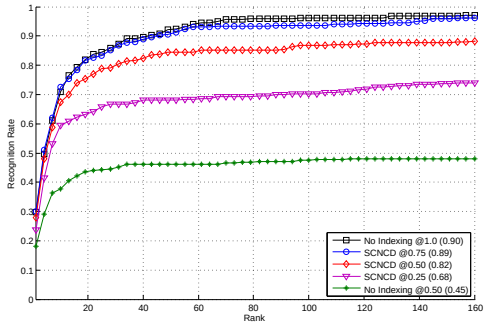

Fig. 6. Performance Degradation.

candidates are returned. Therefore, the remaining experiments will apply this indexing structure.

Aggregation Strategy. This experiment evaluates the Borda's method. Figure 4 presents the results obtained using six parts in our part-based models: head, upper and lower torso, upper and lower legs, and feet. As expected, the best performing parts are the upper and lower torso and the worst are the feet, which can be attributed to dressing styles and background influence. These different parts are aggregated in a unique list which increases the performance exploring the best characteristics of each part, as presented in curve Aggregated (Figure 4). Thus, in next experiments we will employ the Borda's method to aggregate the multiple inverted lists.

Color Models. The salient color names employed for SCNCD are defined as 16 RGB values. Then, to represent other color models using the SCNCD, first the color representation must be transformed into RGB values, which are then related to color names. In this experiment, we evaluate four different color models: RGB, normalized RGB (rgb), HSV and $l_1l_3l_3$ [28]. Figure 5 presents the obtained results and confirms the best performance of RGB color model, which was employed in our experiments.

Performance Degradation. We evaluate the influence of our PCN indexing structure using a state-of-the-art Re-ID method proposed in [11]. We use our implementation of their method as we could not obtain the entire implementation from the authors.

Figure 6 categorizes the results in two types: indexing and no indexing. The No Indexing @1.0 curve represents an upper-bound of performance, but a lower bound of scalability, since each probe sample is compared with all samples in the gallery, as proposed in [11]. Differently, in No indexing @0.5, only 50% of randomly selected gallery individuals are regarded as candidates, obtaining an better efficiency at the cost of a drastically drop in recognition rate. The results show improvements in the recognition rate when our indexing structure is employed, dropping a quarter of the gallery images, our results remain almost equal the upper bound. In addition, considering 50% of the candidates in the indexing scheme, shows an improvement of 0.37 in AUC when compared to randomly selected.

4 Conclusions and Future Works

In this work, we tackled the person re-identification problem using a novel Predominant Color Name indexing structure. The proposed method is composed by multiple inverted lists computed using part-based models and SCNCD descriptors and employs a classic ranking aggregation method (Borda's method) to combine multiple local lists in a unique global list. The obtained results demonstrated that our method needs to consider a small subset of potential candidates when compared to others common indexing scheme based on unsupervised clustering algorithms. In addition, a minimum performance degradation was achieved when only 75% and 50% of the images gallery were considered as potential candidates. As future directions, we intend to investigate techniques to reduce the numbers of potential candidates investigated and analyze how the information obtained in the indexing can be employed in the further stages to increase not just the efficiency, but also the recognition rate.

Acknowledgments

The authors would like to thank the Brazilian National Research Council – CNPq (Grant #477457/2013-4), the Minas Gerais Research Foundation – FAPEMIG (Grants APQ-00567-14 and PPM-00025-15) and the Coordination for the Improvement of Higher Education Personnel – CAPES (DeepEyes Project).

5 References

[1] Peter Henry Tu, Gianfranco Doretto, Nils O. Krahnstoever, a.G. Amitha Perera, Frederick W. Wheeler, Xiaoming Liu, Jens Rittscher, Thomas B. Sebastian, Ting Yu, and Kevin G. Harding, "An Intelligent Video Framework for Homeland Protection," in *Defence and Security Symposium*, 2007.

[2] Horesh Ben Shitrit, Jerome Berclaz, Francois Fleuret, and Pascal Fua, "Tracking Multiple People Under Global Appearance Constraints," in *IEEE ICCV*, 2011.

[3] C.C. Sun, G.S. Arr, R.P. Ramachandran, and S.G. Ritchie, "Vehicle reidentification using multidetector fusion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 3, pp. 155–164, 2004.

[4] Mingli Song, Dacheng Tao, and Stephen J. Maybank, "Sparse camera network for visual surveillance – a comprehensive survey," *CoRR*, vol. abs/1302.0446, 2013.

[5] Roberto Vezzani, Davide Baltieri, and Rita Cucchiara, "People re-identification in surveillance and forensics: a survey," *ACM Computing Surveys*, dec 2013.

[6] Yinghao Cai and Matti Pietikäinen, "Person re-identification based on global color context," in *ACCV*, Berlin, Heidelberg, 2011, pp. 205–215.

[7] Loris Bazzani, Marco Cristani, and Vittorio Murino, "Symmetry-driven accumulation of local features for human characterization and re-identification," *Elsevier CVIU*, vol. 117, no. 2, pp. 130 – 144, 2013.

[8] W.R. Schwartz and L.S. Davis, "Learning discriminative appearance-based models using partial least squares," in *SIBGRAPI*, Oct. 2009, pp. 322–329.

[9] Alina Bialkowski, Patrick J. Lucey, Xinyu Wei, and Sridha Sridharan, "Person re-identification using group information," in *IEEE DICTA*, 2013.

[10] Qingming Leng, Ruimin Hu, Chao Liang, Yimin Wang, and Jun Chen, "Bidirectional ranking for person re-identification.," in *ICME*, 2013, pp. 1–6.

[11] Yang Yang, Jimei Yang, Junjie Yan, Shengcai Liao, Dong Yi, and Stan Z Li, "Salient color names for person re-identification," in *ECCV*, pp. 536–551. 2014.

[12] Doug Gray, Shane Brennan, and Hai Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *IEEE PETS*, 2007.

[13] Martin Koestinger, Martin Hirzer, Paul Wohlhart, Peter M. Roth, and Horst Bischof, "Large scale metric learning from equivalence constraints," in *IEEE CVPR*, 2012.

[14] Cheng-Hao Kuo, S. Khamis, and V. Shet, "Person re-identification using semantic color names and rankboost," in *Proc. WACV (2013)*, Jan 2013, pp. 281–287.

[15] Riccardo Satta, "Appearance descriptors for person re-identification: a comprehensive review," *CoRR*, vol. abs/1307.5748, 2013.

[16] Bryan Prosser, Wei-Shi Zheng, Shaogang Gong, Tao Xiang, and Q Mary, "Person re-identification by support vector ranking.," *BMVC*, vol. 1, no. 3, 2010.

[17] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang, "Person re-identification by probabilistic relative distance comparison," in *IEEE CVPR*, 2011, pp. 649–656.

[18] Lianyang Ma, Xiaokang Yang, and Dacheng Tao, "Person re-identification over camera networks using multi-task distance metric learning.," 2014, pp. 3656–3670.

[19] C. R. S. Dutra, T. Souza, R. Alves, W. R. Schwartz, and L. R. Oliveira, "Re-identifying People based on Indexing Structure and Manifold Appearance Modeling," in *SIBGRAPI*, 2013, pp. 1–8.

[20] Cristianne R. S. Dutra, Matheus Castro Rocha, and William Robson Schwartz, "Person re-identification based on weighted indexing structures," in *CIARP 2014*, 2014, pp. 359–366.

[21] Omar Javed, Khurram Shafique, Zeeshan Rasheed, and Mubarak Shah, "Modeling inter-camera spacetime and appearance relationships for tracking across non-overlapping views," *Computer Vision and Image Understanding*, vol. 109, no. 2, pp. 146 – 162, 2008.

[22] D. Makris, T. Ellis, and J. Black, "Bridging the gaps between cameras," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, June 2004, vol. 2, pp. II–205–II–210 Vol.2.

[23] Riccardo Mazzon, Syed Fahad Tahir, and Andrea Cavallaro, "Person re-identification in crowd," *Pattern Recognition Letters*, vol. 33, no. 14, pp. 1828–1837, 2012.

[24] J. C. Borda, "Memoire sur les elections au scrutin," *Histoire de l'Academie Royale des Sciences*, 1781.

[25] J. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191–203, 1984.

[26] Josef Sivic and Andrew Zisserman, "Video Google: A Text Retrieval Approach to Object Matching in Videos," in *IEEE ICCV*, 2003, pp. 1470–.

[27] Douglas Gray, S. Brennan, and H. Tao, "Evaluating Appearance Models for Recognition, Reacquisition, and Tracking," in *PETS*, 2007.

[28] Theo Gevers and Arnold W.M. Smeulders, "Color-based object recognition," *PR*, vol. 32, no. 3, pp. 453 – 464, 1999.