Person Re-Identification based on Weighted Indexing Structures

Cristianne R. S. Dutra, Matheus Castro Rocha, and William Robson Schwartz

Department of Computer Science, Universidade Federal de Minas Gerais
Belo Horizonte, Minas Gerais, Brazil, 31270-901
cristianne.dutra@gmail.com, rocha_matheus@hotmail.com,
william@dcc.ufmg.br

Abstract. Surveillance cameras are present almost everywhere, indicating an increasing interest regarding people safety. 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 which aims at maintaining the identity of agents that pass through the monitored environment, despite the occurrence of significant gaps in time and space. Many approaches have been proposed to person re-identification. However, there are still problems to be solved, such as illumination changes, pose variation, occlusions, appearance modeling and the management of the large number of people being monitored. This work approaches the last problem with the employment of multiple indexing structures associated with a weighting strategy to maintain the scalability and improve the accuracy. Experimental results demonstrate that the proposed approach is able to improve results based only on a single indexing structure.

Keywords: Person re-identification, weighting strategies, visual dictionaries, predominance filter, inverted lists.

1 Introduction

Due to security concerns, surveillance cameras have become widely spread in public locations. As a result, vast amounts of visual data have to be manually screened and interpreted, which is prone to misinterpretations due to the expected lack of attention of human operators [1]. Therefore, the employment of automatic approaches, including pedestrian detection, person re-identification, face recognition, person tracking and action recognition, is essential to aid the analysis of such data so that one can, for instance, make inferences regarding suspicious activities in a monitored area.

Applied in applications such as surveillance [2], sporting events [3] and traffic monitoring [4], the person re-identification 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 [5]. Specifically, given the large number of images captured by a camera network and a probe

Fig. 1: Illustration of the person re-identification process.

sample, the goal of person re-identification is to find matching candidates to the sample, as illustrated in Figure 1. An extended description and discussion regarding this topic can be found in [5], [6].

Due to the unavailability of facial information due to low quality images captured from surveillance cameras, the majority of the approaches in the literature solve the re-identification problem employing different visual resources to model the subject's appearance based on information regarding color, texture and shape separately [7, 8] as on the combination of such sources of information to maximize the discriminability among different appearances [9].

Context information has also been considered in person re-identification [10, 11]. Leng *et al.* [11] combine the results returned in the search for the probe sample and a number of its k-nearest neighbors. The method proposed by Bialkowski *et al.* [10] exploits group information to improve the recognition rates assuming that groups tend to stay stable (with the same people) over time.

Recent methods have been based on PTZ cameras, capable of focusing on specific parts of the human body [12] and RGB-D cameras, which recover the scene depth [13]. Another approach is to formulate re-identification as a problem relative to distance comparison, which aims at minimizing the distance between a pair of images belonging to the same subject and maximizing the distance between pairs of images of different people [14]. Other works focus on data scalability of transference learning, dealing with the ability of learning a model obtained from an initial pair of cameras to adapt to a new pair of target cameras [15, 16].

In this work, we propose methods to improve results obtained in [17] with the purpose of making the process of searching a person scalable and accurate such that the search does not depend on the number of distinct people being considered. To improve results published in the literature, this work proposes the use of multiple indexing structures based on visual dictionaries and the employment of different weighting strategies to emphasize the more discriminative regions. Experimental results demonstrate improvements achieved by employing a weighting strategy.

2 Methodology

The re-identification process can be performed in two steps. First, feature extraction is performed for test and gallery samples. Then, distances between pairs of potential correspondences are computed to determine which sample from the

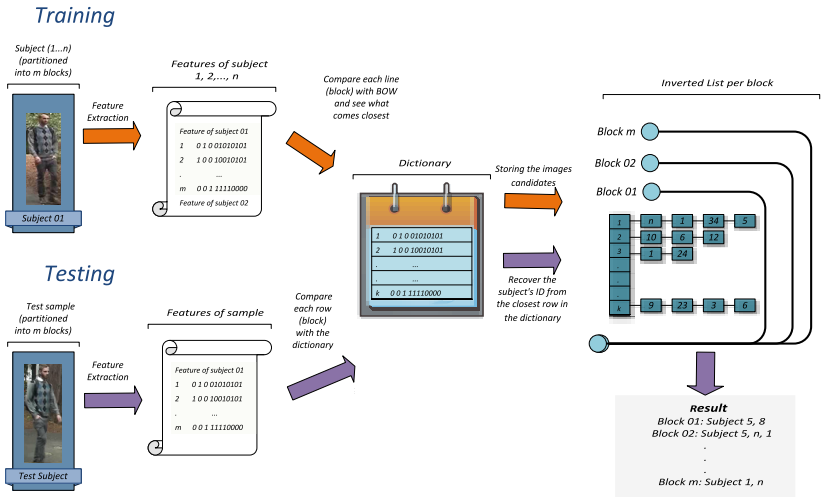


Fig. 2: Training and testing of the person re-identification approach.

gallery corresponds to each testing probe sample, in which the largest similarity indicates a pair of corresponding samples. Optionally, a training stage can be performed to optimize the model parameters and to emphasize the discriminability among individuals.

In this work, we perform improvements on the indexation structure proposed in [17], which is based on inverted lists [18] and visual dictionaries computed with bag-of-words [19]. This structure is capable of reducing the number of candidates when a test sample is presented to the algorithm. In addition, we also propose the inclusion of a weighting strategy to improve the discrimination among subjects.

2.1 Indexing Structure

Figure 2 illustrates the general re-identification process as proposed in [17]. First, feature extraction is performed using features descriptors such as color histogram and predominance filter in RGB space. Then, an indexing structure based on inverted lists [18] and visual dictionary [19] is created. Note that such structure enables a fast search since it does not depend on the number of subjects enrolled in the gallery but only on the number of the codewords in the dictionary.

Dictionary Creation. As the inverted list is indexed by the attributes of the object, a dictionary based on *bag-of-words* [19] is learned to extract these attributes from the samples through feature descriptors. First, each sample of the n known people in the gallery is divided into m non-overlapping blocks, from which descriptors are extracted and stored in features vectors. The goal of dividing the samples into blocks is to locate discriminative characteristics captured from the spatial arrangements. Different feature extraction methods and blocks configurations are evaluated in the experiments.

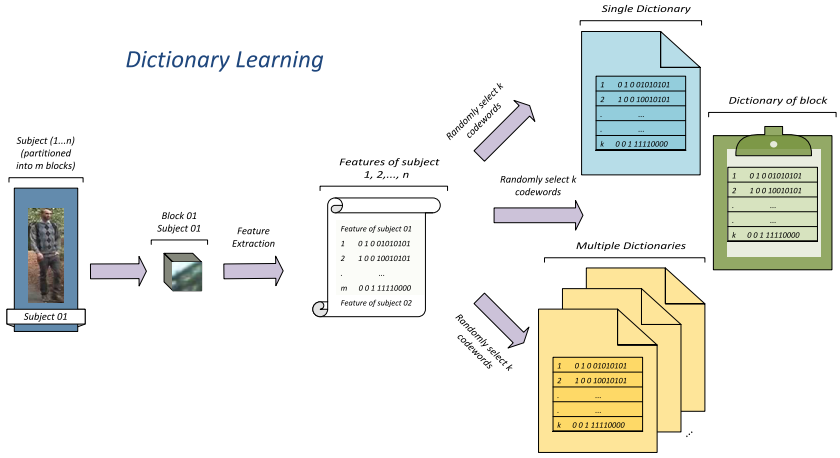


Fig. 3: Approaches for the indexing structure based on dictionaries: single global dictionary, local dictionary for each block and multiple global dictionaries. The k codewords used in the dictionary are selected randomly from feature vectors extracted from the training samples.

With the purpose of increasing the accuracy of the re-identification achieved by Dutra *et al.* [17], which employed a single global dictionary as indexing structure, this work proposes two new approaches (Figure 3). While the first creates one dictionary for each of the m blocks aiming at capturing local information directly in the dictionary, the second creates multiple global dictionaries to being able to have more than one "vote" when a probe samples is being match.

Training. Once the dictionary (or multiple dictionaries) has been learned, the training is executed to fill the m inverted lists that are created – one list per block. In this stage, the samples of the n subjects in the training data set are partitioned into m non-overlapping blocks, from which feature descriptors are extracted. Then, using the dictionary created in the previous stage, the feature vector extracted from each block is compared to all *codewords* in the dictionary and the subject's identifier is added to the position corresponding to the closest codeword in the inverted list. More specifically, if the feature vector extracted from the j-th block of the m-th subject is closer to the i-th *codeword* in the dictionary, the identifier of the m-th person is added in the i-th position of the inverted list associated with block j.

At the end of the training stage, the j-th entry of the inverted list, associated with the i-th block, will have the set $s_{i,j}$, containing the subject's identifiers. The subjects in $s_{i,j}$ share similar characteristics since the feature vectors extracted from their i-th block are closest to the j codeword in the dictionary. Note that when multiple global dictionaries are considered, the number of inverted lists will be increased since one inverted list is created for each block and indexed according to similarity to the dictionary codewords.

Testing. At the testing stage, when a probe sample is presented to be matched, it is split into m non-overlapping blocks and features are extracted from each block. Then, the feature vector of the i-th block is compared to all the codewords of the dictionary and the index j of the closest codeword is used to retrieve the set $s_{i,j}$ from the i-th inverted list, containing a list of candidates more likely to match the probe sample according to the i-th block.

Once the above matching process has been performed for all m blocks, the more likely candidates in the gallery to match the probe sample are ranked according to their frequency of occurrence in the sets $s_{i,j}$, where i denotes the block index and j the index of the closest codeword for the i-th block. We refer to the step that ranks candidates as voting since the set $s_{i,j}$ will receive a weight proportional to the importance of the i-th block (e.g., blocks containing mostly background regions should receive a smaller weight due to their lack of discriminability). Differently from [17], in which all blocks receive equal weights.

2.2 Block Weighting

To increase the accuracy for person re-identification, some approaches are devised to give higher weights to more discriminative regions (blocks) of a sample. With such approaches, it is possible to attenuate the influence of background regions and increase the significance of regions belonging to the person to be identified during a voting stage.

In the first proposed approach, after partitioning the sample image into blocks, a Gaussian function is used to weigh each block. Using a Gaussian function allow us to add more weight to blocks closer to the center and less weight to blocks far from the center. This is possible since the data sets assume that people have been properly detected and are at the center of the image.

The second approach consists in storing and using the distances calculated in the training stage when the feature vectors are compared to the codewords from the dictionary. The calculated Euclidean distance is used to weigh the blocks in such way that blocks with smaller average distances receive higher weights since present feature descriptors more similar to the ones in the dictionary.

The last approach uses learning to adapt the dictionaries according to the gallery. First, the dictionary learning and the training are performed and then a validation set is used to perform a test using individual blocks. The recognition rate obtained by each block is used in the weighting following the intuition that a more discriminative block should have higher weights. This way, the block with the highest recognition rate receives the highest weight and so on, until all of the blocks from the samples receive a specific weight, which will be used in the voting.

3 Experimental Results

In this section, we describe the experiments executed to validate the proposed method using the widely employed VIPeR Dataset [20]. The results will be shown

using the recognition rate as a function of the rank, common approach to validate person re-identification approaches [6]. The experiments test aspects such as the setup of blocks, feature descriptors, number of visual dictionaries as well as the influence of local and global dictionaries, and the employment of different weighting strategies.

Block setup. In this experiment, we evaluate how the recognition rates vary according to the number of horizontal blocks for sample partition. Even though we tested vertical and squared blocks, the recognition rates with these block types are significantly lower than the ones achieved with horizontal blocks. This might be justified by the way people dress. People tend to wear homogeneous clothing when seen as horizontal stripes, which is not common in the vertical (e.g., one color for the shirt and a different color for the pants would be in the same block), becoming hard to capture with color descriptors due to their bimodal distribution. According to the results shown in Figure 4(a), the best result is achieved when the sample is partitioned into 32 non-overlapping horizontal blocks.

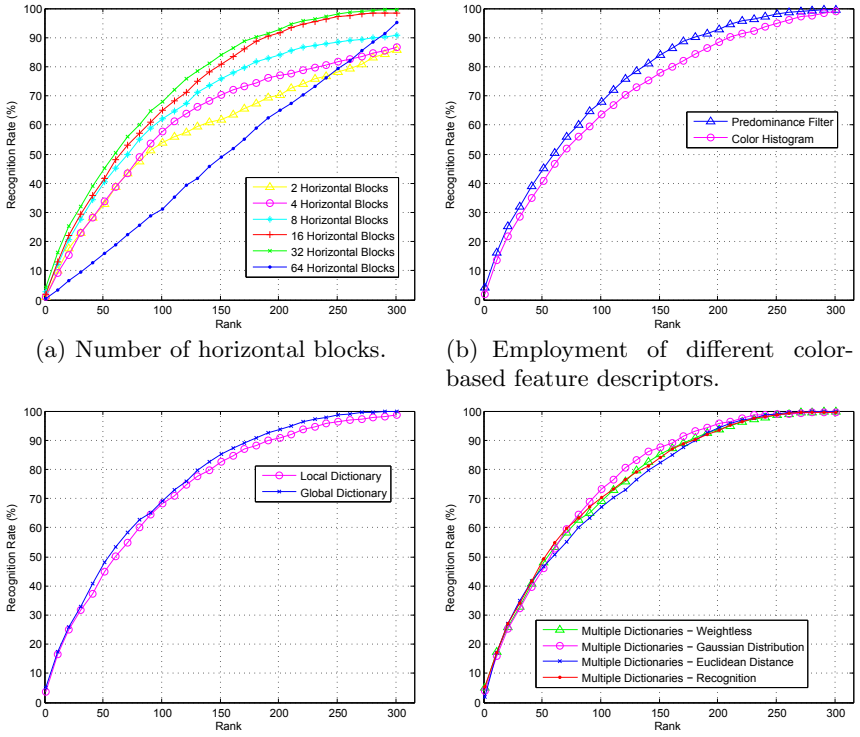
Feature descriptors. This experiment evaluates the recognition rates obtained using two feature descriptors: predominance filter [17] and simple color histogram. The results shown in Figure 4(b) indicate that the predominance filter is more discriminative achieving better recognition rates.

Visual dictionaries. To verify the contribution of a different number of global dictionaries, we evaluate the recognition rates with he use of 1, 2, 3, 5, 10 and 20 dictionaries. The best results were obtained with 5 dictionaries, 85.1% of recognition rate at rank 150. For instance, the employment of a single dictionary achieved recognition rate of 83.5% at the same rank.

Considering the use of global (dictionaries for all blocks) or local dictionaries (one dictionary per block), Figure 4(c) show that the best result was achieved using global dictionaries, with a recognition rate of 85.1% at rank 150 (against 83.5% for a single global dictionary as seen earlier)and 82.4% using a dictionary per block. The local dictionaries achieved the worst results, which might be due to the addition of more weight to some less significant blocks, reducing the accuracy of the method and adding noise during the voting. On the other hand, using five global dictionaries increased the number of selected characteristics from the samples, making the method more discriminative.

Block weighting. In this experiment, we evaluate the influence of different weighting strategies used for blocks. For this test, configurations with best results in previous tests were used. According to Figure 4(d), the best recognition rate was obtained when using the combination of multiple dictionaries and block weighting with a Gaussian distribution, getting a 87.5% rate at rank 150.

Comparisons. When comparing the results obtained by the improvements proposed in this work to the results obtained in [17], we achieved improved recognition rates. The average recognition rate obtained in [17] was 83.5% and after incorporating the improvements, we obtained 87.5%, a significant increase of 4 percentage points at rank 150, an important improvement for such challenging data set as the VIPeR.



(a) Number of horizontal blocks.

(b) Employment of different color-based feature descriptors.

(c) Employment of different dictionary configurations.

(d) Different block weighting strategies.

Fig. 4: Experimental Results.

4 Conclusions and Future Works

In this work, we approached person re-identification using an indexation structure composed by inverted lists and codewords. In addition, two approaches were employed to increase the accuracy: multiple dictionaries and block weighting. Experimental results demonstrated that the proposed improvements increase the recognition rates compared to the original method.

As future directions, we intend to investigate techniques to select feature vectors that will be the *codewords* in the dictionary to make the inverted list size more uniform to improve even more the scalability of the method.

Acknowledgments

The authors would like to thank the Brazilian National Research Council – CNPq (Grant #487529/2013-8) and the Minas Gerais Research Foundation - FAPEMIG (Grant APQ-01294-12).

References

1. H. Keval, "CCTV Control Room Collaboration and Communication: Does it Work?," in *HCTW*, pp. 1–4, 2006.
2. P. H. Tu, G. Doretto, N. O. Krahnstoever, a. A. Perera, F. W. Wheeler, X. Liu, J. Rittscher, T. B. Sebastian, T. Yu, and K. G. Harding, "An Intelligent Video Framework for Homeland Protection," in *Defence and Security Symposium*, 2007.
3. H. B. Shitrit, J. Berclaz, F. Fleuret, and P. Fua, "Tracking Multiple People Under Global Appearance Constraints," in *IEEE ICCV*, 2011.
4. C. Sun, G. Arr, R. Ramachandran, and S. Ritchie, "Vehicle reidentification using multidetector fusion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 3, pp. 155–164, 2004.
5. M. Song, D. Tao, and S. J. Maybank, "Sparse camera network for visual surveillance – a comprehensive survey," *CoRR*, vol. abs/1302.0446, 2013.
6. R. Vezzani, D. Baltieri, and R. Cucchiara, "People re-identification in surveillance and forensics: a survey," *ACM Computing Surveys*, dec 2013.
7. Y. Cai and M. Pietikäinen, "Person re-identification based on global color context," in *ACCV*, (Berlin, Heidelberg), pp. 205–215, 2011.
8. L. Bazzani, M. Cristani, and V. Murino, "Symmetry-driven accumulation of local features for human characterization and re-identification," *Elsevier CVIU*, vol. 117, no. 2, pp. 130 – 144, 2013.
9. W. Schwartz and L. Davis, "Learning discriminative appearance-based models using partial least squares," in *SIBGRAPI*, (Rio de Janeiro, Brazil), pp. 322–329, Oct. 2009.
10. A. Bialkowski, P. J. Lucey, X. Wei, and S. Sridharan, "Person re-identification using group information," in *Digital Image Computing : Techniques and Applications (DICTA)*, (Wrest Point Hotel, Hobart, TAS), November 2013.
11. Q. Leng, R. Hu, C. Liang, Y. Wang, and J. Chen, "Bidirectional ranking for person re-identification.," in *ICME*, pp. 1–6, 2013.
12. P. Salvagnini, M. Cristani, and V. Murino, "Person re-identification with a ptz camera: an introductory study," in *IEEE ICIP*, 2013.
13. J. Lorenzo-Navarro, M. Castrillón-Santana, and D. Hernández-Sosa, "On the use of simple geometric descriptors provided by rgb-d sensors for re-identification," *Sensors*, vol. 13, no. 7, pp. 8222–8238, 2013.
14. W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE TPAMI*, vol. 35, no. 3, pp. 653–668, 2013.
15. R. Layne, T. M. Hospedales, and S. Gong, "Domain transfer for person re-identification," in *ACM Multimedia International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams*, 2013.
16. Y. Wu, W. Li, M. Minoh, and M. Mukunoki, "Can feature-based inductive transfer learning help person re-identification?," in *IEEE ICIP*, 2013.
17. 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*, pp. 1–8, 2013.
18. D. Knuth, "Retrieval on secondary keys," in *The Art of Computer Programming*, 1997.
19. J. Sivic and A. Zisserman, "Video Google: A Text Retrieval Approach to Object Matching in Videos," in *IEEE ICCV*, pp. 1470–, 2003.
20. D. Gray, S. Brennan, and H. Tao, "Evaluating Appearance Models for Recognition, Reacquisition, and Tracking," in *10th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, 2007.