Re-identifying People based on Indexing Structure and Manifold Appearance Modeling

Cristianne R. S. Dutra, William Robson Schwartz
Department of Computer Science
Universidade Federal de Minas Gerais, UFMG
Belo Horizonte-MG, Brazil, 31270-901

Tiago Souza, Raul Alves, Luciano Oliveira
Dept. of Computer Science, Intelligent Vision Research Lab
Federal University of Bahia, UFBA
Salvador, Brazil
http://www.ivisionlab.eng.ufba.br

***Abstract*—The role of person re-identification has increased in the recent years due to the large camera networks employed in surveillance systems. The goal in this case is to identify individuals that have been previously identified in a different camera. Even though several approaches have been proposed, there are still challenges to be addressed, such as illumination changes, pose variation, low acquisition quality, appearance modeling and the management of the large number of subjects being monitored by the surveillance system. The present work tackles the last problem by developing an indexing structure based on inverted lists and a predominance filter descriptor with the aim of ranking candidates with more probability of being the target search person. With this initial ranking, a more strong classification is done by means of a mean Riemann covariance method, which is based on a appearance strategy. Experimental results show that the proposed indexing structure returns an accurate short-list containing the most likely candidates, and that manifold appearance model is able to set the correct candidate among the initial ranks in the identification process. The proposed method is comparable to other state-of-the-art approaches.**

***Keywords*-Person re-identification; bag-of-features; predominance filter; inverted lists; mean Riemann covariance;**

I. Introduction

In recent years, person re-identification has played an important role on visual surveillance due to the need of managing activities in a large area covered by a surveillance system. The amount of data to be processed in surveillance systems has increased due to the large availability of camera networks. Therefore efficient approaches have to be employed to solve some intrinsic problems, mainly involving person monitoring. Among these intrinsic problems, person re-identification is responsible for maintaining a broad identity of subjects, in a camera network, wherein the cameras not necessarily present intersection of field of view. This problem has been considered for several applications, such as surveillance and monitoring [1], and sport events [2].

There are two main concerns when developing approaches to perform person re-identification. First, since facial information cannot be used in all situations due to the small size of persons in the acquired videos, the subject's appearance has to be modeled accurately. Second, due to the possibly large number of subjects in the scene, and the need to constantly save person information, scalable and efficient approaches for matching samples is also necessary. Regard to the first aforementioned problem, some approaches have been proposed based either on single shot or multiple shots of the subjects [3], [4], [5]. While the former approach considers a single image to model the person appearance, the latter employs a set of images.

Particularly speaking of each category of the previous problems, some approaches, considering the problem as a classification, are based on a single image can be found in [6], [7], [8]. A one-against-all strategy to model the appearances is described in [7], then extended to a more scalable one-against-some solution in [6]. A method based on spatial color histograms is proposed by Hirzer [8]. Considering multiple shots, some methods are preferably based on face detection instead of full body. In [9], a mean Riemannian covariance grid (MRCG) is introduced as a descriptor to obtain a very high discriminative human body signature from multiple images of a person. In [10], the automatic labeling of faces is performed employing tracking and appearance modeling. Regarding re-identification by appearance modeling, recent people re-identification methods have considered feature descriptors to model the appearance of a person [11], detection of interest points to identify previously known subjects, considering a KD-tree structure [12] and the use of auto-similarity images [13]. However, such appearance methods suffer with the change of look of individuals captured by different cameras, low quality of the samples gathered due to variations in illumination or shadows, and the large amount of data been captured.

The re-identification problem can be treated simultaneously as an indexing of feature descriptors extracted from the subjects and an appearance modeling. In this sense, we propose a fast and scalable indexing scheme based on inverted lists and bag-of-words. By indexing person detections, it is possible to build a short list of known individuals to be matching candidates, when a target sample is sought. For the final matching and, consequently, estimate decision whether a person is in the database or not, we applied a mean Riemann covariance (MRC) following the formulation of [14], but differing from [9] in the way of computing the descriptors.

The main contributions of this paper are the following. First, we address the problem of searching a previously detected image person in a database by a two-stage strategy, that is, instead of matching the target individual to all individuals in the database, we build an inverted list from a bag-of-words approach, looking only at the k most relevant hy-

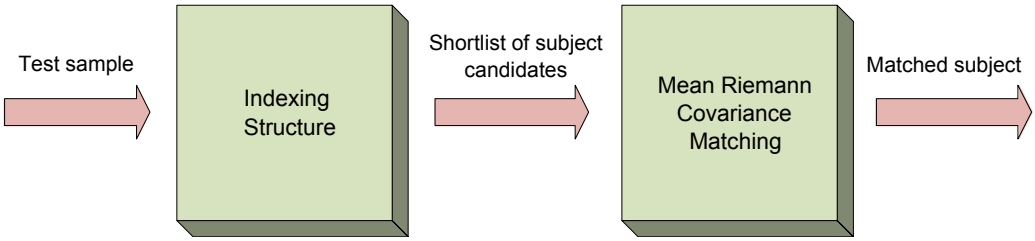

Fig. 1. Modules employed to perform person re-identification.

potheses. Second, the predominant RGB filter is introduced here as a novel discriminative color descriptor for person re-identification. Third, we take the local MRC to compute the final matching, resulting in an accurate appearance-based recognition. Finally, thorough evaluations were accomplished on two datasets (ETHZ and Viper), showing promising results of the proposed method.

II. Outline of the Proposed Method

In the proposed method, we focus in two main aspects. First, we develop an indexing scheme based on inverted lists being able to reduce the number of candidates when a test sample is presented to the algorithm (Section II-A). Then, a MRC is applied to perform the actual identification of the sample (Section II-B). Figure 1 illustrates the main building blocks of the proposed method, which will all be addresses in the next sections.

A. Indexing Structure

With the aim at obtaining a shortlist of subjects being candidates to match the identity of a target sample, we propose the usage of inverted lists. Inverted lists are index data structures proposed by [15] that allow the mapping from attributes to objects, instead of objects to their attributes. Applied to our domain, that means to map features extracted from subjects (bag-of-words approach [16], in our case) to their identifiers, which, differently from [7], [6], allows us to create an indexing structure with size independent from the number of subjects in the gallery (candidates to match the identity of a target sample), making it faster, as it will be shown in the experiments. The process is divided into three steps, as depicted in Fig. 2: *dictionary creation*, *learning* and *candidate selection*.

1) Dictionary Creation: Since the inverted list is indexed by attributes of the object, we employ a dictionary based on bags-of-words [16] to extract such attributes from the image samples. First, the sample images for the n subjects with known identity are split into m non-overlapping blocks, from which feature descriptors are extracted and stored in feature vectors. A dictionary is created by selecting randomly k feature vectors as codewords (experiments have been performed by using the *k-means* clustering algorithm to select the codewords, but the results are very similar to those achieved by the random selection).

The reason for using a dictionary is to find the codewords that better represent the image blocks so that their positions in the dictionary can be used as indexes in the inverted lists. For instance, if the i-th codeword is the closest to a set of subjects, the subject's identifiers will be added in the index i of the inverted list. Our hypothesis is that if a test sample has a feature vector similar to the i-th codeword in the dictionary, it is likely that its identity belongs to one of the subjects in the index i of the inverted list.

To create the dictionary, instead of having a raw information of color, which would be dependent on lighting change, and not discriminative in practice, we consider two feature descriptors: the histogram of oriented gradients (HOG) [17], based on shape information, and a RGB predominance filter, as a novel color descriptor. In the experiments, we evaluate the accuracy of both descriptors.

HOG captures edge or gradient structures, which are characteristics of local shape with a controllable degree of invariance to local geometric transformations. In this work, HOG descriptor was used in each one of the m blocks of the images, considering 8 bins and 1×1 cell in each block, that is, only one histogram is computed for each block.

In addition to the shape-based feature descriptor, a predominance filter is introduced here. For that, the goal is to highlight the color channel which is predominant in each pixel. Given an RGB image, $I = (C_R, C_G, C_B)$, where C_R, C_G and C_B represents a value between 0 and 1 for each channel, the input color space is divided into 8 regions well defined by a unique parameter T, as follows

$$P_T(C_R, C_G, C_B) = \begin{cases} (\mu_R, 0, 0) & \text{if } C_R - \max(C_G, C_B) > T \\ (0, \mu_G, 0) & \text{if } C_G - \max(C_R, C_B) > T \\ (0, 0, \mu_B) & \text{if } C_B - \max(C_G, C_B) > T \\ (0, \mu_G, \mu_B) & \text{if } C_G \approx C_B > C_R + T \\ (\mu_R, 0, \mu_B) & \text{if } C_R \approx C_B > C_G + T \\ (\mu_R, \mu_G, 0) & \text{if } C_R \approx C_G > C_B + T \\ (\mu_R, \mu_G, \mu_B) & \text{if } C_R \approx C_G \approx C_B > 0.5 \\ (0, 0, 0) & \text{if } C_R \approx C_B \approx C_B < 0.5 \end{cases}$$

where $\mu_{\{R,G,B\}}$ denotes the mean of the channel that satisfies the conditions of the equation for a image region, and T is in the interval $[0; 1]$; "$\approx$" means $|C_i - C_j| < T$, with $i, j \in \{R, G, B\}$. The predominance filter has the physical effect of separating the colors of RGB according to Fig. 3.

The idea with the combination of the two descriptors – shape and color – is to build a reliable dictionary capturing the gist of a person in the images.

2) Learning: As described earlier, once the dictionary has been created, in an unsupervised fashion, the learning proce-

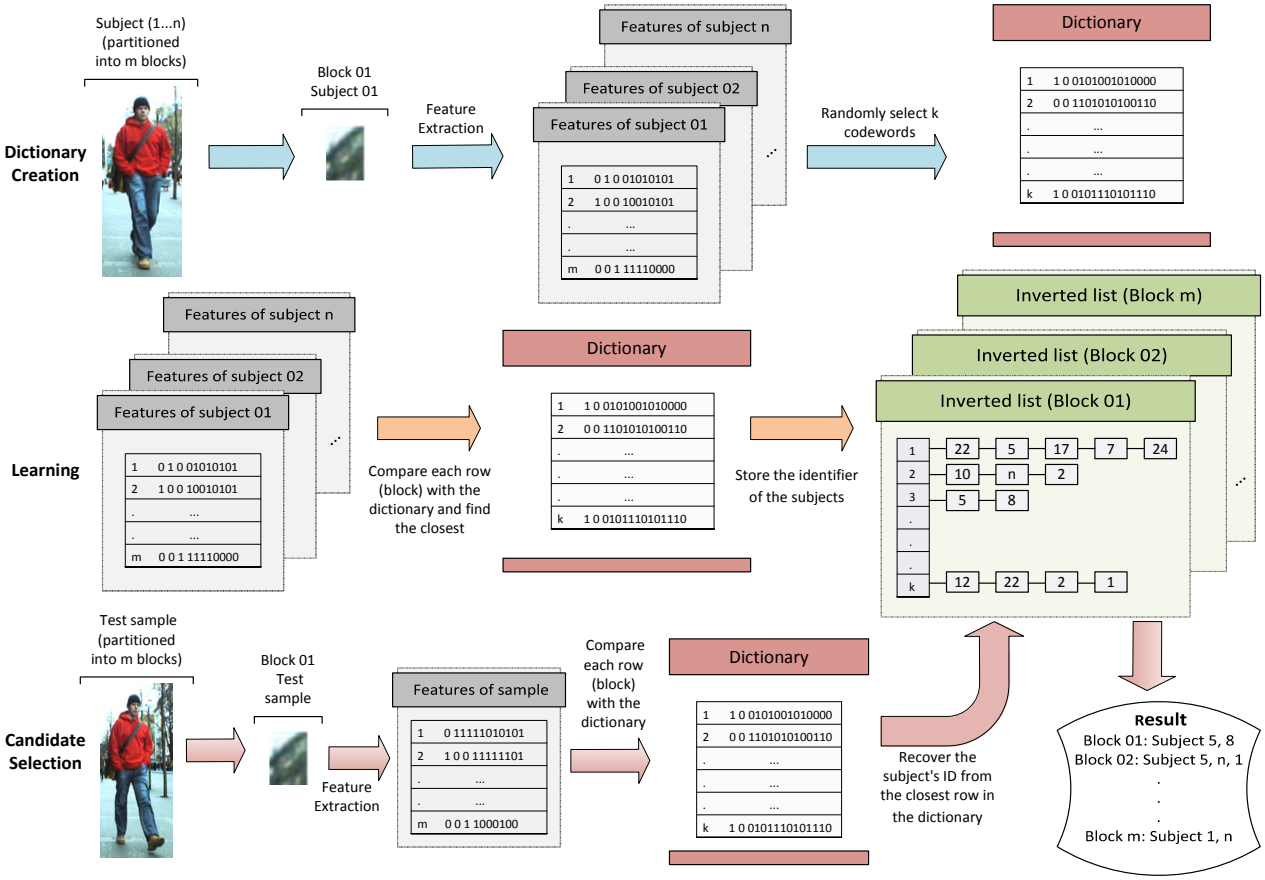

Fig. 2. Indexing structure based on bag-of-words and inverted lists to obtain a shortlist of candidates in order to identify a test sample.

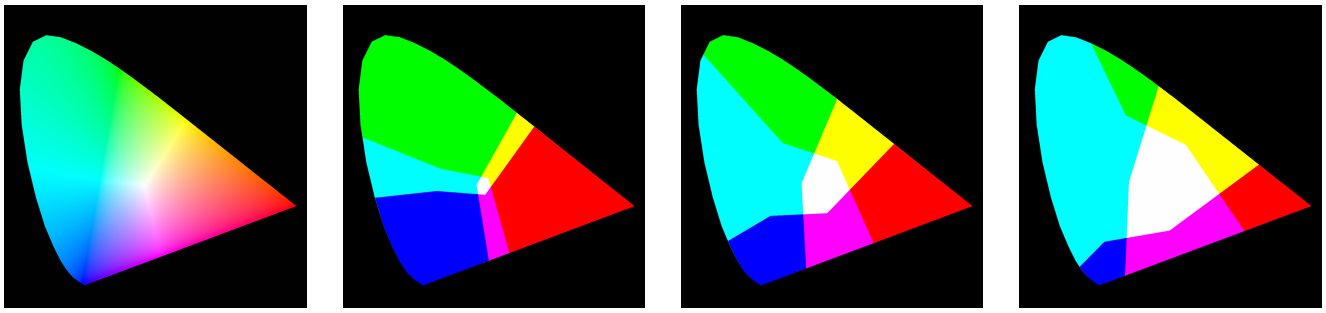

Fig. 3. Original RGB space and effect of the predominance filter with $T = 0.1, 0.3$ and 0.5, respectively.

dure is executed to populate the inverted lists, and m lists are created – one per image block. The idea of creating multiple inverted lists is to allow voting to select the most likely subjects to identify a target sample, without being affected by noise incurred in some image regions, which might interfere in the extracted feature descriptors.

In the learning stage, the feature descriptors are extracted from the m blocks of each subject sample. Next, using the dictionary created in the previous step, the feature vector, extracted in each block, is compared to all codewords in the dictionary, and the closest index – the i-th codeword – is kept. In addition to the comparisons, we create m inverted lists, each one with k indexes, one per codeword. Finally, if the feature vector extracted from the j-th block of the l subject is closest to the i-th codeword in the dictionary, the subject's identifier l will be added in the i-th position of the j-th inverted list. This way, during test, if a test sample presents a feature vector similar to this, it will index the same subject identity. Note that the i-th position of the j-th inverted list might have more than one subject's identifier when more than one subject presents

similar feature vectors for the j-th block, as illustrated in the second row of Figure 2,

3) Candidate Selection: The goal of this final step, in the indexing structure, is to select a subset of subjects that is more likely to contain a target sample given to the method. This subset of candidates will be used to perform the mean Riemann covariance matching.

Given a target sample, its image will be partitioned in m blocks in the same way as in the stages of creation and learning of the dictionary, and features will be extracted for each one of the blocks. Then, the feature vector for the j-th block will be compared to all the codewords contained in the dictionary. Finally, the index of the closest codeword, the i-th, will be used to index the set s_j containing the subject's identifiers at the i-th position of the j-th inverted list.

After the set $S = \{s_1, s_2, \ldots, s_m\}$ has been obtained with the process described, the following strategy to find the p most likely subjects to identify a test sample was adopted: The most likely subject is the one that appears in the most number of times in S; the second one, is the one that appears the second most number of times, and so forth, until obtaining the p subjects that appear the most in S. The rationale is that a subject that appears multiple times, presents feature vectors similar to those extracted from the test sample.

B. Mean Riemann Covariance Matching

After building a ranked list for each person in a dictionary, by using the inverted lists, the next step was to use only the k-top persons of that list in order to match them to the target individual (the value k is analyzed in Section III). In this stage, a mean Riemann covariance was preferred following the formulation of [14].

Given an input image and a set of covariance matrices, $\{\Gamma_i\}_1^Z$, with each γ_i computed on overlapped image blocks, in the same way as in [18], with the exception of the coordinates of the pixel. In other words, the following feature vector was used to compute the covariances: $[Ix, Iy, Ixx, Iyy, |G|, \theta, R, G, B]$, where Ix, Iy are the first derivatives with respect to x and y of the pixel, Ixx, Iyy are the second derivatives with respect to x and y of the pixel, $|G|, \theta$ are the magnitude and angle of the pixel gradient, and R, G, B are the values of each RGB channel of the pixel. It is noteworthy that since those matrices rely in a manifold space, $\mathcal{M}$, a distance between two matrices is computed following [19], and is given by

$$d(\Gamma_i, \Gamma_j) = \sqrt{\sum_{l=1}^{m} ln^2 \lambda_l(\Gamma_i, \Gamma_j)} \tag{1}$$

where $\lambda_l(.)$ are the generalized eigenvalues determined by $|\lambda\Gamma_i - \Gamma_j| = 0$.

Since it is generally not possible to integrate manifold-valued functions, the mean, μ, in the manifold is not unique, and should be achieved by a minimization procedure with the following objective function

$$\mu = arg \min_{\Gamma \in \mathcal{M}} \sum_{i=1}^{Z} d^2(\Gamma, \Gamma_i) \tag{2}$$

where μ is known as Fréchet mean.

In order to find the μ from (2), we utilized the method of Gauss-Newton gradient descend proposed by [20], until it converges with an error of 0.01. It is noteworthy that here it is computed locally in the detection window, rather than temporally as in [9].

Fig. 4. Samples of the ETHZ dataset.

III. Experimental Evaluation

In this section, we evaluate the proposed method focusing on the indexing scheme (Section III-B), and on the appearance modeling based on the mean covariance (Section III-C). The full sequence of the method is also evaluated by employing first the indexing and the appearance modeling (Section III-D). The datasets used to perform the evaluation were the VIPer and the ETHZ, described in more detail in the next section.

A. Evaluation Datasets

Two data sets were used to evaluate the proposed method: The ETHZ person re-identification dataset [7] and the VIPeR Dataset [21]. The former (Fig. 4 illustrates some samples) presents characteristics such as changes in illumination, pose variation, changes in sample size and low acquisition quality, which impose challenges to the re-identification problem, while the latter dataset has a large number of subjects (632), having only two samples per subjects (one used for learning and one for testing), captured from different viewpoints of two cameras. Figure 5 illustrates some examples of VIPer dataset.

The ETHZ dataset is composed of three video sequences, the first with 1000 frames and 83 subjects, the second with

Fig. 5. Samples of the VIPeR dataset.

451 frames and 35 subjects and the third sequence with 354 frames and 28 subjects. We chosen the sequence #3 to set the parameters, since it presents the lowest number of subjects, and remaining sequences to evaluate the proposed approach. The VIPeR dataset is also used for testing also using the same parameters estimated from sequence #3 of the ETHZ dataset. Table I summarizes information about the datasets. All samples are rescaled to the same size.

TABLE I
DATASETS USED IN THE EXPERIMENTS.

Datasets	Frames	Persons
ETHZ Seq 1	1000	83
ETHZ Seq 2	451	35
ETHZ Seq 3	354	28
VIPer CamA	-	316
VIPer CamB	-	316

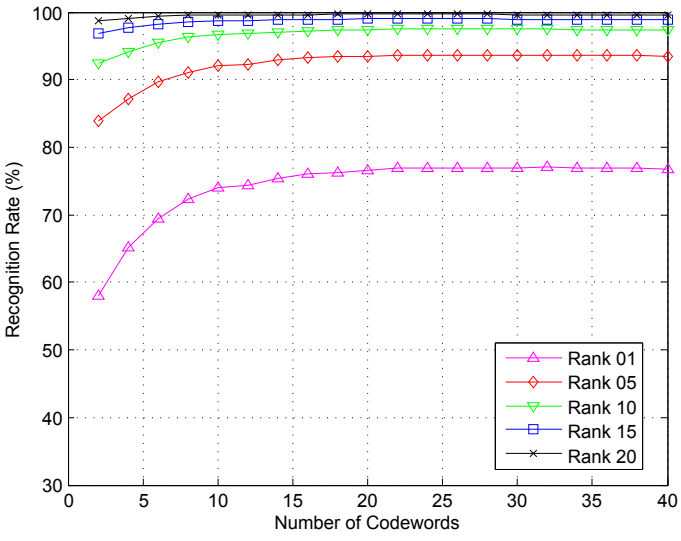

Fig. 6. Recognition rate by the proposed indexing structure as a function of the number of codewords to build the dictionary. Curves showing the recognition rates for several matching ranks.

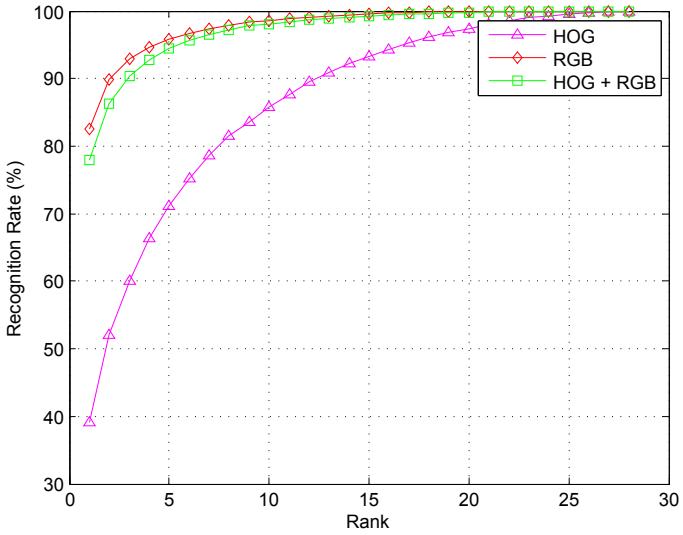

Fig. 7. Recognition rates achieved by the proposed indexing structure for different feature extraction methods as a function of the matching rank.

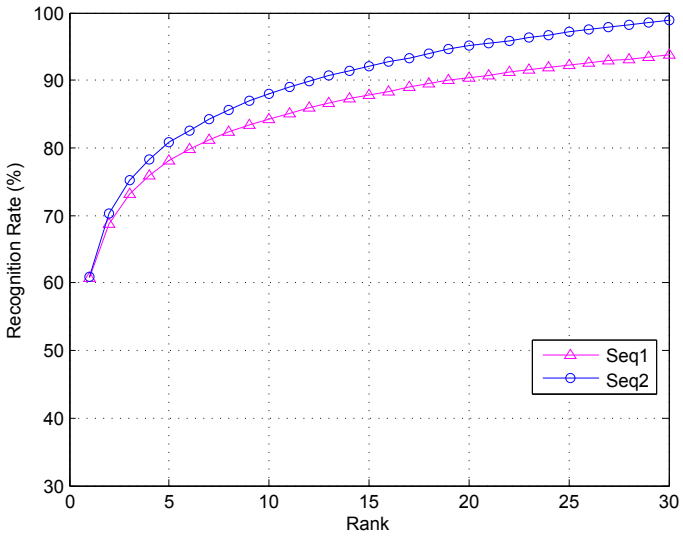

Fig. 8. Cumulative match characteristic curve achieved by the proposed indexing structure using the video sequences of the ETHZ dataset.

B. *Indexing Structure*

In this section we evaluate the indexing structure. First, we estimate the parameters using sequence #3 of the ETHZ dataset and then we evaluate the results achieved with the other sequences and over the VIPeR dataset.

Number of codewords. To evaluate the most suitable number of codewords to build the dictionary, we performed an experiment using the predominance filter and HOG as feature descriptors. The results are shown in Figure 6, in which several curves obtained for different matching ranks are displayed. According to the results, the recognition rate increases until the number of codewords reaches 10, after that, it becomes stable. Therefore, we have fixed the number of codewords to 10 for the remaining experiments.

Feature descriptors. To evaluate which feature descriptor

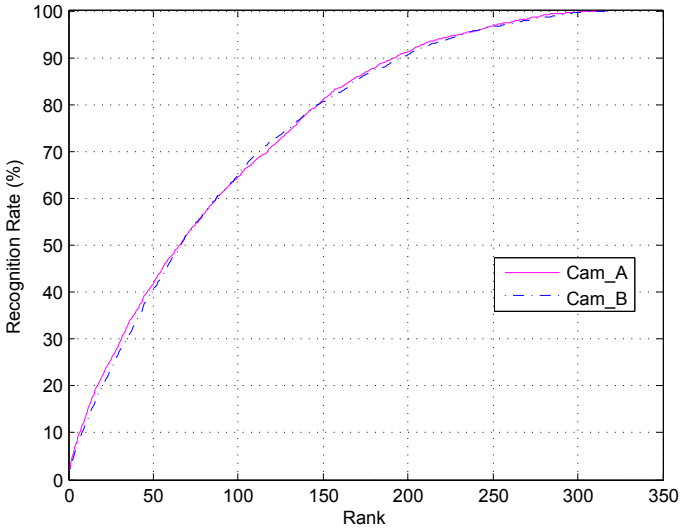

Fig. 9. Cumulative match characteristic curve achieved by the proposed indexing structure using the VIPer dataset.

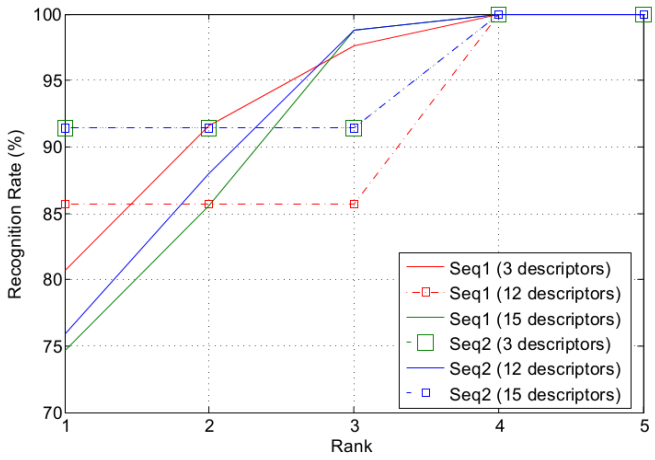

Fig. 10. Cumulative match characteristic curve achieved by MRC using ETHZ dataset, and considering 3, 12 and 15 covariance descriptors. For both sequences, the best performance was with 12 descriptors.

is more suitable to create the indexing, we performed an experiment comparing the HOG, the predominance filter and the combination of both (through concatenation of both descriptors). According to the results shown in Figure 7, the employment of the predominance filter achieves the best results, even when it is compared to its combination to the HOG descriptor. This can be explained by the nature of the problem, since the color is the most discriminative information and the person shape is mostly ambiguous, as we can see in the dataset samples shown in Figures 4 and 5. Therefore, the remaining experiments will be executed considering only the predominance filter as feature descriptor.

Cumulative match characteristic curves. According to the results shown in the evaluations, we have chosen 10 codewords to build the dictionary, with each feature vector comprised of the three channel of the predominance filter. Figure 8 shows the results achieved using sequences #1 and #2 of the ETHZ dataset and Figure 9 shows the results achieved using the VIPeR dataset. In Section III-D, we show the improvements achieved when the MRC is employed after the indexing structure.

C. Mean Riemann Covariance Matching

The MRC is an unsupervised method since it is not necessary the training of a model. However, a threshold must be found and the sequence #3 of EHTZ was used to define this. Since the images in the datasets have been resized to 50×150, the blocks to compute the covariances must be calculated in such a way that all the image regions are covered by the covariance descriptors. For that, we defined three types of computations: 3 descriptors, 12 descriptors, and 15 descriptors (in fact, the combination of 3 plus 12 descriptors). All descriptors were computed with overlapping of 50%.

In order to assess the performance of MRC, cumulative match characteristic curves were built over ETHZ and VIPer datasets, and they are depicted in Figures 10 and 11. Figure 12

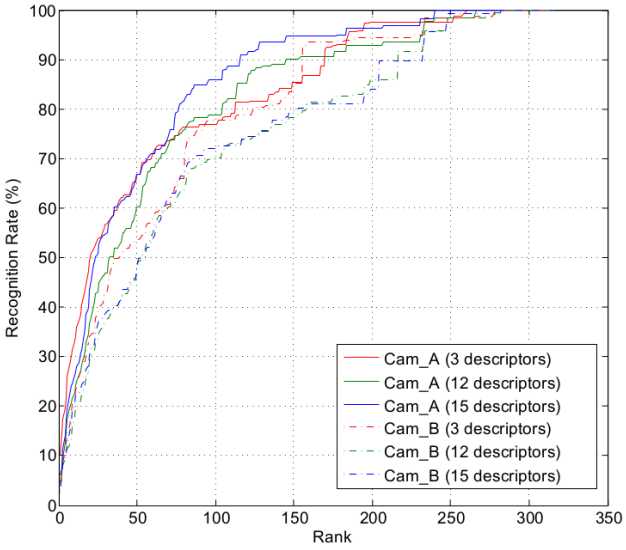

Fig. 11. Cumulative match characteristic curve achieved by MRC using VIPer dataset, and considering 3, 12 and 15 covariance descriptors. For both sequences, the best performances were with 15 and 3 descriptors over CamA and CamB sequences, respectively.

illustrates a zoom in Figure 11, in order to suitably visualize the plots. Observing the curves, it is noticeable that with 12 descriptors, the best result was achieved in ETHZ for all sequences, while 15 and 3 descriptors are more suitable for CamA and CamB sequences of VIPer datasets. There is no significant differences among the descriptors, and, in practice, 3 descriptors can perform well in most of the cases, alleviating the computational burn in the computation of covariance descriptors.

TABLE II
SUMMARY OF THE RESULTS FOUND IN FIGURE 13.

	Method	Rank1	Maximum recognition at
Seq1	IS	61%	Rank 30
	IS + MRC (12 descriptors)	81%	Rank 4
Seq2	IS	61%	Rank 30
	IS + MRC (12 and 15 descriptors)	91%	Rank 4
CamA	IS	1%	64% at rank 100
	MRC (15 descriptors)	10%	85% at rank 100
CamB	IS	1%	63% at rank 100
	MRC (15 descriptors)	1%	72% at rank 100

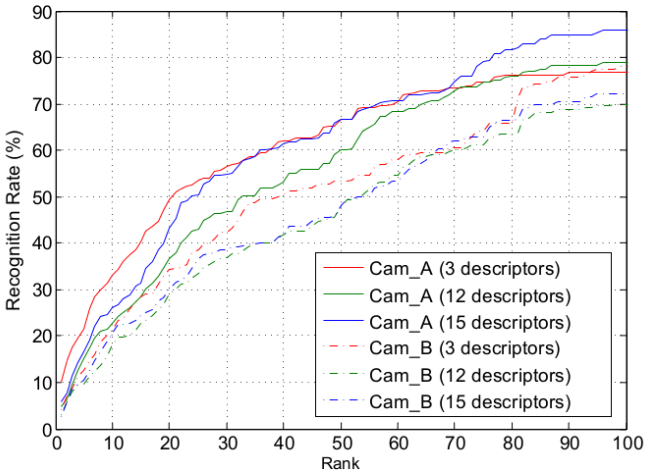

Fig. 12. Zoom of Figure 11.

TABLE III
COMPARATIVE RESULTS WITH OTHER STATE-OF-THE-ART METHODS. IS STANDS FOR INDEXING STRUCTURE AND MRC MEAN RIEMANN COVARIANCE.

	Method	Rank1
Seq1	One-against-all [6]	71.90%
	IS + MRC (12 descriptors)	81%
Seq2	One-against-all [6]	73.50%
	IS + MRC (12 and 15 descriptors)	91%
CamA	ELF 200 [21]	1%
	MRC (15 descriptors)	10%
CamB	ELF 200 [21]	1%
	MRC (15 descriptors)	1%

D. Comparative Evaluation

Considering the indexing structure along with the MRC, in this section, we present comparative results over ETHZ and VIPer datasets. First, we built a cumulative match characteristic curve with comparison in recognition performance with just the indexing structure and with the indexing structure along with MRC. Figure 13 depicts the results.

In order to better understand the results, Table II summarizes the results of Fig. 13. In the ETHZ dataset, the increase in performance using the MRC with 12 or 15 descriptors is clear. As a matter of fact, there is substantial gains between 20% and 30% considering Rank 1 (that is, the first person matched), or even in the top rank with maximum performance, which decreases from 30 to 4 in both sequences of the datasets. In the VIPer dataset, the gain was still high but not as high as in the ETHZ dataset, consisting in an increase between 10% and 20% with respect to the indexing structure. In fact, the VIPer dataset is a hard dataset, with difficult images even for a human to recognize, as illustrated in Figure 5.

Considering the proposed method comparatively with state-of-the-art approaches in person re-identification field, Table III illustrates the results. Taking into consideration the performance in Rank 1, the proposed method is superior either against [6] or [21], showing 10% of average gain in comparison to the others. Nevertheless, it is not completely true when using the VIPer datasets, since [21] achieves perfect results in Rank 200 (out of 316), while our method does not reach 100% of recognition rate. As a matter of fact, the proposed method achieves the maximum recognition rate of 85% and 72% over CamA and CamB sequences in VIPer dataset, respectively, keeping it constant after Rank 100.

IV. CONCLUSIONS

In this paper, we addressed the problem of person re-identification by using a indexing structure comprised of inverted lists and codewords, and a mean Riemann covariance matching, this latter applied locally in each person image. A thorough analysis was accomplished showing that the use of MRC improves significantly the indexing structure. Since the computational load to calculate covariance matrices is usually high, as well as their means, an MRC with 3 descriptors are best suitable in practice mainly to demonstrate similar performance in comparison with 12 or 15 descriptors, and to be fast to be calculated. In comparison with other state-of-the-art methods, IS+MRC showed superior performance over ETHZ datasets, but needs to be improved since presented lower performance in more difficult datasets as in VIPeR. Future works are driven to implement a temporal version of the system considering also the weighted mean rather than only the mean.

ACKNOWLEDGEMENTS

The work at UFBA was supported by Rede Nacional de Pesquisa, Edital Cidades Inteligentes, 2010, under the project SPACES-4D (Sistema Participativo de Gestão e Monitoramento de Cidades e Serviços Públicos Usando Rastreamento

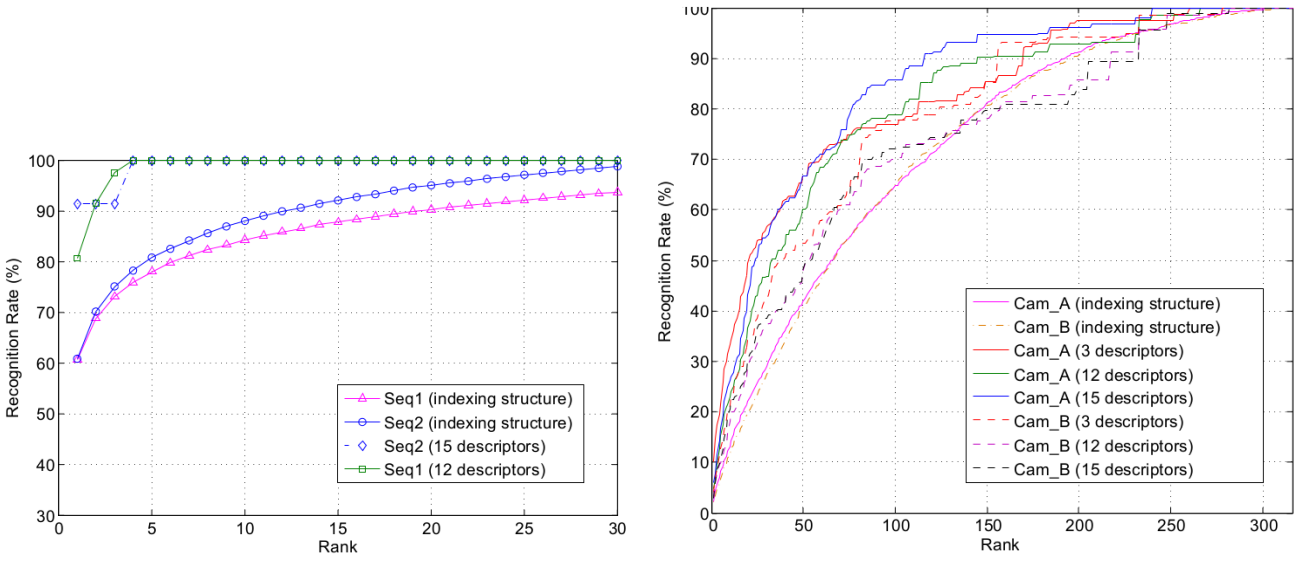

Fig. 13. Comparative results over ETHZ and VIPer datasets using indexing structure with and without MRC.

com Câmeras 4D). The authors also would like to thank CNPq, CAPES and FAPEMIG for the financial support.

Author Contributions

C.R.S.D. and W.R.S. contributed with the indexing structure; T.S., R.A., and L.O. contributed with the covariance matching; T.S. developed the RGB predominance filter.

References

[1] 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.

[2] H. Ben Shitrit, J. Berclaz, F. Fleuret, and P. Fua, "Tracking multiple people under global appearance constraints," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2011, pp. 137 –144.

[3] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher, "Appearance-based person reidentification in camera networks: Problem overview and current approaches," *Journal of Ambient Intelligence and Humanized Computing*, vol. 2, pp. 127–151, 2011.

[4] R. Mazzon, S. Tahir, and A. Cavallaro, "Person re-identification in crowd," *Pattern Recognition Letters*, vol. 33, no. 14, pp. 1828–1837, 2012.

[5] T. D'Orazio and G. Cicirelli, "People re-identification and tracking from multiple cameras: A review," in *IEEE International Conference on Image Processing*, 2012.

[6] W. R. Schwartz, "Scalable people re-identification based on a one-against-some classification scheme," in *IEEE International Conference on Image Processing*, 2012.

[7] W. R. Schwartz and L. Davis, "Learning discriminative appearance-based models using partial least squares," in *XXII Brazilian Symposium on Computer Graphics and Image Processing*, Rio de Janeiro, Brazil, Oct. 2009, pp. 322–329.

[8] M. Hirzer, C. Beleznai, M. Köstinger, P. Roth, and H. Bischof, "Dense appearance modeling and efficient learning of camera transitions for person re-identification," in *IEEE International Conference on Image Processing*, 2012.

[9] S. Bak, E. Corvee, F. Bremond, and M. Thonnat, "Multiple-shot human re-identification by mean riemannian covariance grid," in *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*, 2011, pp. 179–184.

[10] D. Ramanan, S. Baker, and S. Kakade, "Leveraging archival video for building face datasets," in *International Conference on Computer Vision*, 2007, pp. 1–8.

[11] N. Gheissari, T. Sebastian, and R. Hartley, "Person reidentification using spatiotemporal appearance," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2006, pp. 1528–1535.

[12] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux, "Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences," in *Second ACM/IEEE International Conference on Distributed Smart Cameras*, Stanford, CA, USA, 2008, pp. 1 –6.

[13] Y. Cai and M. Pietikainen, "Person re-identification based on global color context," in *Visual Surveillance*, 2010.

[14] X. Pennec, P. Fillard, and N. Ayache, "A riemannian framework for tensor computing," *International Journal of Computer Vision*, vol. 66, pp. 41–66, 2006.

[15] D. Knuth, "Retrieval on secondary keys," in *The Art of Computer Programming*. Reading, Massachusetts: Addison-Wesley, 1997.

[16] J. Sivic and A. Zisserman, "Video Google: A Text Retrieval Approach to Object Matching in Videos," in *International Conference on Computer Vision (ICCV)*, 2003, pp. 1470–.

[17] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, 2005, pp. 886–893.

[18] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: a fast descriptor for detection and classification," in *Proceedings of the 9th European conference on Computer Vision - Volume Part II*, ser. ECCV'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 589–600.

[19] W. Förstner and B. Moonen, "A metric for covariance matrices," 1999.

[20] X. Pennec, "Probabilities and statistics on riemannian manifolds: Basic tools for geometric measurements," in *Proc. of Nonlinear Signal and Image Processing*, ser. NSIP. IEEE, 1999, pp. 194–198.

[21] 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.