Brownian descriptor: a Rich Meta-Feature for Appearance Matching

Sławomir Bąk   Ratnesh Kumar   François Brémond

INRIA Sophia Antipolis, STARS group

2004, route des Lucioles, BP93

06902 Sophia Antipolis Cedex - France

`firstname.surname@inria.fr`

Abstract

This paper introduces an image region descriptor and applies it to the problem of appearance matching. The proposed descriptor can be seen as a natural extension of covariance. Driven by recent studies in mathematical statistics related to Brownian motion, we design the Brownian descriptor. In contrast to the classical covariance descriptor, which measures the degree of linear relationship between features, our novel descriptor measures the degree of all kinds of possible relationships between features. We argue that the proposed covariance is a richer descriptor than the classical covariance, especially when fusing non-linearly dependent features. We evaluate our approach on tracking related applications, demonstrating that the Brownian descriptor outperforms the classical covariance in terms of matching accuracy and efficiency.

1. Introduction

The appearance matching is one of the most desired bricks in computer vision applications. Recognition, detection, tracking and classification are common examples of vision tasks, which require robust features for matching. Good features should be distinctive and invariant to illumination and scale changes [6, 15]. Moreover during the last few years, there is of great interest to look for features which are also invariant to camera changes. This is the most challenging case of appearance matching that consists in determining whether two image regions registered by non-overlapping cameras correspond to the same object. This problem is particularly hard due to significant appearance changes caused by variations in view angle, lighting conditions and different object pose.

Recent studies have shown that the covariance descriptor [22], can be successfully applied to appearance matching [1, 2, 3, 4, 16, 19, 23, 24]. This descriptor encodes information on feature variances inside an image region, their linear correlations with each other and their spatial layout. The

performance of the covariance features is found to be superior to other methods, as rotation and illumination changes are absorbed by the covariance matrix.

Although it has been established that covariance is an effective feature, we revisit its fundamentals and highlight its limitations. In this work we point out that the covariance descriptor measures only information on linear dependence between features. This makes the descriptor incomplete for characterizing an image region. In other words, having a diagonal covariance matrix, we may imply feature independence, which is not a sufficient condition for statistical independence in general case. Nonlinear or non-monotone dependence may exist (section 3).

We overcome this issue by introducing a new image region descriptor that is a natural extension of covariance (section 4). The proposed descriptor is referred to as *Brownian descriptor* due to its analogy to the Brownian covariance. The classical covariance measures the degree of linear relationship between features, whereas Brownian covariance measures the degree of *all kinds of possible relationships* between features [20].

The paper illustrates advantages of the new descriptor over the covariance matrix using theoretical data as well as real images. Moreover, we provide an efficient algorithm for extracting the descriptor, while using *integral images*.

Theoretical deliberations are confirmed by experimental results on real-data (section 5). We employ the Brownian descriptor for two vision tasks, *i.e.* visual tracking and person re-identification. We demonstrate that the Brownian descriptor outperforms the classical covariance descriptor in terms of both efficiency and accuracy.

2. Related work

In [23] covariance matrix is used for designing a robust **human detection** algorithm. Human appearance is modeled by a dense set of covariance features extracted inside a detection window. Covariance descriptor is computed from sub-windows with different sizes sampled from different lo-

cations. Then, a boosting mechanism selects the best regions for characterizing a human silhouette.

Covariance matrix has also been successfully applied to **object tracking**. In [19] object deformations and appearance changes were handled by a model update algorithm using the Lie group structure of the positive definite matrices.

The first approach which employs the covariance descriptor for **person re-identification** (appearance matching across non-overlapping cameras) is [2]. In this work a HOG-based detector establishes the correspondence between body parts, which are matched using a spatial pyramid of covariance descriptors.

In [16] we can find biologically inspired features combined with the similarity measure of covariance descriptors. The new descriptor is not represented by the covariance matrix but by a distance vector computed using the similarity measure between covariances extracted at different resolution bands. This method shows promising results not only for person re-identification but also for face verification.

Matching groups of people by covariance descriptor is the main topic of [8]. It is shown that contextual cues coming from group of people around a person of interest can significantly improve the re-identification performance. This contextual information is also kept by the covariance matrix.

In [3] the authors use *one-against-all* learning scheme to enhance distinctive characteristic of covariances for a specific individual. As covariances do not lie in Euclidean space, binary classification is performed on a Riemannian manifold. Tangent planes extracted from positive training data points are used as a classification space for a boosting algorithm. Similarly, in [13] discriminative models are learned by a boosting scheme. However, covariance matrices are transformed into *Sigma Points* to avoid learning on a manifold, which often produces an over-fitted classifier.

Unfortunately, using covariance matrices, we also influence significantly computational complexity. This issue has been addressed in [24]. The covariance matrices of feature subsets rather than the full feature vector, provide similar performance while significantly reducing the computation load. Similarly, in [1] the authors do not define *a priori* feature vector for extracting covariance, but learn which features are the most appropriate for matching. Depending on feature localization in a model, the covariance may be build using different feature combination, characterizing the appearance of a given object class in a more efficient way.

State of the art approaches either investigate strategies for combining covariance descriptors into more robust representation or look for particular features which could be embedded into the covariance descriptor. However no systematic study has been undertaken to examine a design and limitations of the covariance descriptor.

In contrast to these approaches, we present new insights into the covariance descriptor, raising fundamental limitations of covariance as a dependence measure. We design a new descriptor driven by recent advances in mathematical statistics related to Brownian motion [5, 20].

The new descriptor not only keeps more information on feature dependence than the classical covariance, but also can be treated as a point on an Euclidean space. This makes the descriptor computationally efficient and useful for real-time systems.

3. Covariance descriptor and its limitations

In [22] covariance of *n*-features has been proposed to characterize an image region. Covariance matrix can be computed from any type of image such as one dimensional intensity image, three channel color image or even other types of images, *e.g.* infrared. Let $\mathcal{I}$ be an image and F be a n-dimensional feature image extracted from $\mathcal{I}$

$$F(x, y) = \phi(\mathcal{I}, x, y), \quad (1)$$

where function ϕ can be any mapping, such as color, intensity, gradients, filter responses, *etc*. For a given rectangular region $Reg \subset F$, let $\{f_k\}_{k=1\ldots p}$ be the n-dimensional feature points inside Reg. Each feature point f_k is characterized by function ϕ. We represent region Reg by the $n \times n$ covariance matrix of the feature points

$$C_{Reg} = \frac{1}{p-1} \sum_{k=1}^{p} (f_k - \mu)(f_k - \mu)^T, \quad (2)$$

where μ is the mean of the points. In the result, (i, j)-th element of the covariance matrix can be expressed as

$$C_{Reg}(i, j) = \frac{1}{p-1} \sum_{k=1}^{p} (f_k(i) - \mu(i))(f_k(j) - \mu(j)). \quad (3)$$

Standardization Covariance values are very often normalized by the product of corresponding standard deviations

$$\rho = \hat{C}_{Reg}(i, j) = \frac{C_{Reg}(i, j)}{\sqrt{C_{Reg}(i, i)C_{Reg}(j, j)}}, \quad (4)$$

and are referred to as *the Pearson Product-Moment Correlation Coefficients* (ρ).

Covariance limitations ρ measures a linear correlation between two variables (the strength of linear dependence). However, as it is computed w.r.t. the mean of the variable (see equation 3), it is not able to measure nonlinear or non-monotone dependence (see section 4.3 for elaboration).

4. Brownian descriptor

This section introduces the Brownian descriptor, discussing its advantages over the classical covariance. Before elaborating the descriptor, let us introduce the *distance*

covariance statistics $\mathcal{V}^2$ that measures dependence between random vectors in arbitrary dimension. The mathematical notations and formulas provided in the next section are in accordance with [20].

4.1. Distance covariance $\mathcal{V}^2$

Distance covariance is a new class of multivariate dependence measure applicable to random vectors of arbitrary and not necessarily equal dimensions. Distance covariance is analogous to the classical covariance measure, but with an important property characterizing independence, *i.e.* distance covariance $\mathcal{V}^2$ is zero if and only if the random vectors are independent. The same condition is not sufficient for $\rho = 0$, as nonlinear or non-monotone dependence may exist.

Sample **distance covariance** $\mathcal{V}_n^2$ between random vectors X and Y is defined as

$$\mathcal{V}_n^2(X, Y) = \frac{1}{n^2} \sum_{k,l=1}^{n} A_{kl} B_{kl}, \qquad (5)$$

where A_{kl} and B_{kl} are simple linear functions of the pairwise distances between sample elements. These functions are defined in the following. For a random sample $(\mathbf{X}, \mathbf{Y}) = \{(X_k, Y_k) : k = 1 \ldots n\}$ of n i.i.d random vectors (X, Y) from their joint distribution , compute the Euclidean distance matrices $(a_{kl}) = (|X_k - X_l|_p)$ and $(b_{kl}) = (|Y_k - Y_l|_q)$. Define

$$A_{kl} = a_{kl} - \bar{a}_{k.} - \bar{a}_{.l} + \bar{a}_{..}, \quad k, l = 1, \ldots, n, \quad (6)$$

where

$$\bar{a}_{k.} = \frac{1}{n} \sum_{l=1}^{n} a_{kl}, \ \ \bar{a}_{.l} = \frac{1}{n} \sum_{k=1}^{n} a_{kl}, \ \ \bar{a}_{..} = \frac{1}{n^2} \sum_{k,l=1}^{n} a_{kl}. \qquad (7)$$

Similarly, we define $B_{kl} = b_{kl} - \bar{b}_{k.} - \bar{b}_{.l} + \bar{b}_{..}$.

In [20] it has been shown that the definitions of the new dependence coefficients have theoretical foundations based on characteristic functions and on the new concept of covariance w.r.t. Brownian motion. Surprising coincidence is that the properties of distance covariance are the same properties established for Brownian covariance [5]. Further, we can see that standardized Brownian covariance measures the degree of *all kinds of possible relationships* between two real-valued random variables, while the standardized product-moment covariance coefficients (ρ) measures the degree of linear relationship between two real-valued variables [20]. In the result $\mathcal{R}^2(X, Y) = 0$ implies that there is no dependence between variables. It is relevant to mention that this is the main advantages of $\mathcal{R}^2(X, Y)$ over ρ. $\rho = 0$ *means that there is no linear correlation between variables, while non-linear or non-monotone dependence may exist.*
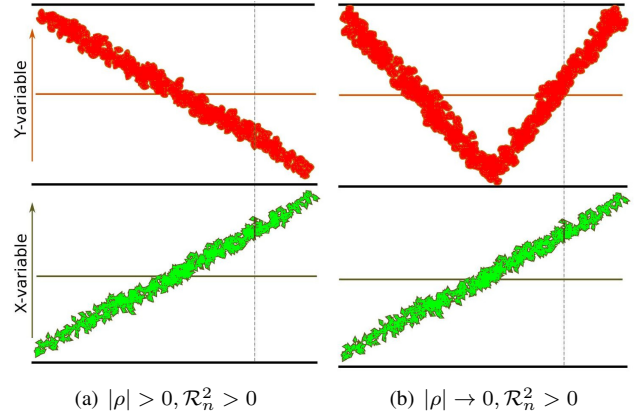

(a) $|\rho| > 0, \mathcal{R}_n^2 > 0$ (b) $|\rho| \rightarrow 0, \mathcal{R}_n^2 > 0$

Figure 1. Comparison of ρ *vs.* $\mathcal{R}_n^2$. Values of variables X and Y are presented as dotted lines. In the first case (a) ρ is positive and relatively high; in the second case (b) ρ will be close to zero due to mean-dependent computation, while actually variables X and Y are far from being independent. $\mathcal{R}_n^2$ in both cases has high value.

Standardization Similarly to covariance which has its standardized counterpart ρ, $\mathcal{V}_n^2$ has its standardized version referred to as *distance correlation* $\mathcal{R}_n^2$ and defined by

$$\mathcal{R}_n^2(X, Y) = \begin{cases} \frac{\mathcal{V}_n^2(X,Y)}{\sqrt{\mathcal{V}_n^2(X)\mathcal{V}_n^2(Y)}}, & \mathcal{V}_n^2(X)\mathcal{V}_n^2(Y) > 0; \\ 0, & \mathcal{V}_n^2(X)\mathcal{V}_n^2(Y) = 0, \end{cases} \qquad (8)$$

where

$$\mathcal{V}_n^2(X) = \mathcal{V}_n^2(X, X) = \frac{1}{n^2} \sum_{k,l=1}^{n} A_{kl}^2. \qquad (9)$$

4.2. Image region descriptor

Let $\mathcal{I}$ be an image and $\mathcal{L} = \{L_1, L_2, \ldots, L_n\}$ be a set of feature layers ($\mathcal{L}$ is defined by mapping ϕ). The Brownian descriptor is obtained by computing $\mathcal{R}_n^2(\mathcal{L}, \mathcal{L})$, while keeping distance coefficients in the form of matrix. Similarly to the classical covariance matrix, the Brownian descriptor is represented by a positive definite and symmetric matrix and it provides a natural way of fusing multiple features. This descriptor does not contain any information regarding the ordering and the number of points (pixels). This implies a certain scale and rotation invariance over the image regions in different images as long as layers L_i are invariant (similarly to [22]).

4.3. Brownian *vs.* covariance ($\mathcal{R}_n^2$ *vs.* ρ)

Designing the Brownian descriptor, we propose to replace ρ by $\mathcal{R}_n^2$ for measuring dependence between image features. We claim that this keeps more information on dependence between features included in mapping ϕ.

In fig. 1 we present sample distribution of two variables X and Y, which are obviously highly correlated. We can notice that $\mathcal{R}_n^2$ can keep information on dependence between variables even when they are in non-monotone correlation. ρ ignores that correlation due to mean-dependent computation (see equation 3). This is a fundamental problem of covariance, in which ρ *may go very close to zero even if the two variables are highly correlated.*

For illustrating actual correlations between features in real data, we use the famous image of Lena's eye and we apply mapping commonly used in state of the art [22, 23, 24] (*i.e.* $\phi = \left[x, y, I_{xy}, \nabla_{xy}^I, \theta_{xy}^I\right]$), where x and y are pixel location, I_{xy} is an intensity, ∇_{xy} and θ_{xy} correspond to gradient magnitude and orientation, respectively. Then, we plot values of these layers together to show dependence between each two layers. From fig. 2 it is evident that correlations between feature layers have non-linear characteristics. In the result it is obvious that $\mathcal{R}_n^2$ keeps more information than ρ. Note that *e.g.* dependence between I_{xy} and θ_{xy}^I ($row = 5, column = 3$) is held by $\mathcal{R}_{53}^2 = \mathcal{R}_{35}^2 = 0.55$, while $\rho_{53} = -0.01$ almost does not detect mutual dependence.

5. Experimental results

In the previous section, we showed advantages of $\mathcal{R}_n^2$ over ρ either on a theoretical sample (fig. 1) or on a sample image (fig. 2). This section focuses on evaluating the Brownian descriptor in real world scenarios. We compare our descriptor with the classical covariance descriptor, while testing their performance on two vision tasks: (1) visual tracking and (2) person re-identification.

5.1. Visual tracking

Sample tracking results are given in figures 3,4 and 5. In each case, a tracker is initialized manually by selecting a target window (green rectangle). From this target window we extract either the covariance descriptor or the Brownian descriptor using state of the art feature mapping ($\phi = \left[x, y, I_{xy}, \nabla_{xy}^I, \theta_{xy}^I\right]$) [22, 23, 24]. The candidate region in the next frame is selected by a dense scanning with 2×2 pixel step and by minimizing *geodesic distance* [10] between the target and the candidate matrix (Brownian or covariance). We follow the scheme of [19], in which the target is updated by averaging the descriptor on a Riemannian manifold. We show the tracking results for both, covariance and Brownian descriptor, using data related to different applications. In figures red boxes indicate the misses.

Car tracking In fig. 3 we can observe a drift of both descriptors due to texture changes (shadows on the car modify the appearance). The results show that the Brownian descriptor is able to track the vehicle through the whole sequence while the covariance descriptor loses the object in the middle of the sequence.
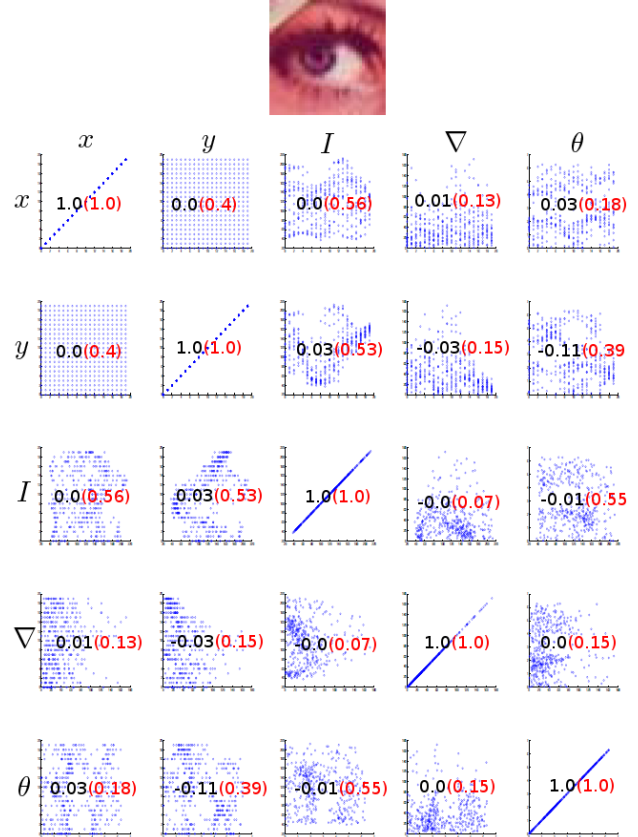

Figure 2. Lena's eye and correlation plots. Plots illustrate pairwise correlations of features included in ϕ. The values correspond to elements in output 5×5 matrix; $\rho_{kl} = \hat{C}_{Reg}(k, l)$ *vs.* $\mathcal{R}_{kl}^2$ ($\mathcal{R}_{kl}^2$ in brackets).

Head tracking For evaluation we select the action recognition dataset [18] (fig. 4). The results indicate that the Brownian descriptor is able to adapt to the undergoing changes. Note that although the appearance of the head changed significantly (front and back of the head), the descriptor was still able to successfully update the model.

Tennis player tracking [7] This is a widely accepted motion segmentation dataset, which contains moving and stationary camera recordings. In fig. 5 we can see that the Browning descriptor has only a few misses, while the covariance failed in most part of the sequence. From the results it is apparent that the Brownian descriptor consistently achieves better tracking performance than the classical covariance.

This *template drift problem* occurs due to accumulation of small errors while successfully matching the template in time. The prime cause of this drift problem is the change in appearance of object while moving across frames [14, 17]. The results demonstrate that we can track the template for longer time duration with Brownian as opposed to the classical covariance matrix. This is due to the fact that match-
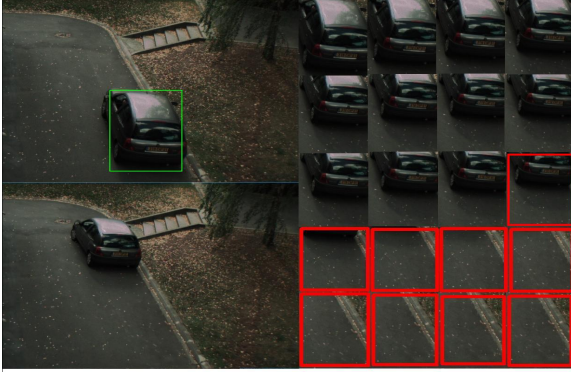
Figure 3. Car tracking results (every $5th$ frame) using: the covariance descriptor (upper image) and the Brownian descriptor (bottom image).

Figure 4. Head tracking results (every $3rd$ frame) using: the covariance descriptor (upper image) and the Brownian descriptor (bottom image).

ing in successive frames is more reliable with Brownian and the template position is better updated w.r.t. time. Therefore we obtain better drift immunity as compared to covariance-based tracking.

5.2. Person re-identification

For illustrating quantitative results, we select the person re-identification problem. During the past few years this problem has been the focus of intense research bringing proper metrics and datasets for evaluation. Moreover, as this is a challenging case of appearance matching, we believe that it is a good choice for evaluating the Brownian descriptor.

We carry out experiments on 3 i-LIDS datasets[1]: i-LIDS [25], i-LIDS-MA and i-LIDS-AA [3]. These datasets have been extracted from the 2008 i-LIDS Multiple-Camera Tracking Scenario (MCTS) dataset with multiple non-overlapping camera views. The results are analyzed in terms of recognition rate, using the *cumulative matching characteristic* (CMC) curve [12]. The CMC curve represents the expectation of finding the correct match in the top n matches. Additionally, we report a quantitative scalar of

CMC curve obtained by the normalized area under CMC curve (nAUC), which is reported in brackets.

As a baseline we select MRCG [4] descriptor. This descriptor is represented by a dense grid of covariance matrices which are averaged on Riemannian manifold. Using the same appearance representation we replace the classical covariance descriptor with our Brownian descriptor. We refer to this modified descriptor as MRCG+B. For a fair comparison, we evaluate both descriptors without applying discriminative analysis proposed in [4].

i-LIDS-MA [3] This dataset consists of 40 individuals extracted from two non-overlapping camera views. For each individual a set of 46 images is given. Hence, we have in total $40 \times 2 \times 46 = 3680$ images. Figure 6 illustrates sample images from i-LIDS-MA.

Brownian descriptor w.r.t. the number of shots For each pedestrian we create human signature using $N = 1, 3, 5, 10$ randomly selected images. Then, every signature is used as a query to the gallery set of signatures from different camera. The procedures were repeated 10 times and average CMC curves are displayed in fig. 7.

The results indicate that the larger number of frames, the better performance is achieved by both descriptors. Noticeably Brownian descriptor outperforms the classical covari-

---

[1]The Image Library for Intelligent Detection Systems (i-LIDS) is the UK government's benchmark for Video Analytics (VA) systems
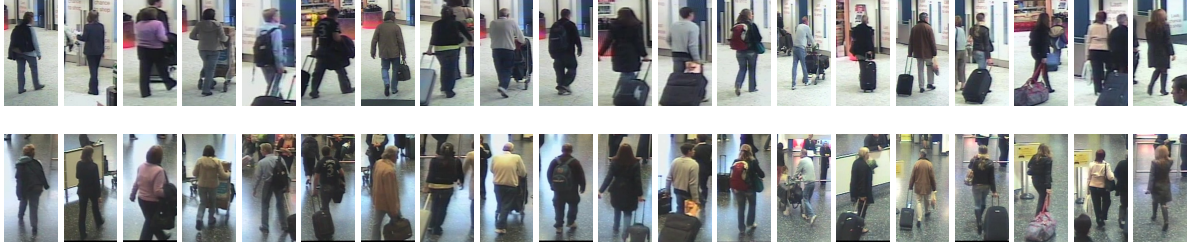
Figure 6. The sample images from i-LIDS-MA dataset. Top and bottom lines correspond to images from different cameras. Columns illustrate the same person.
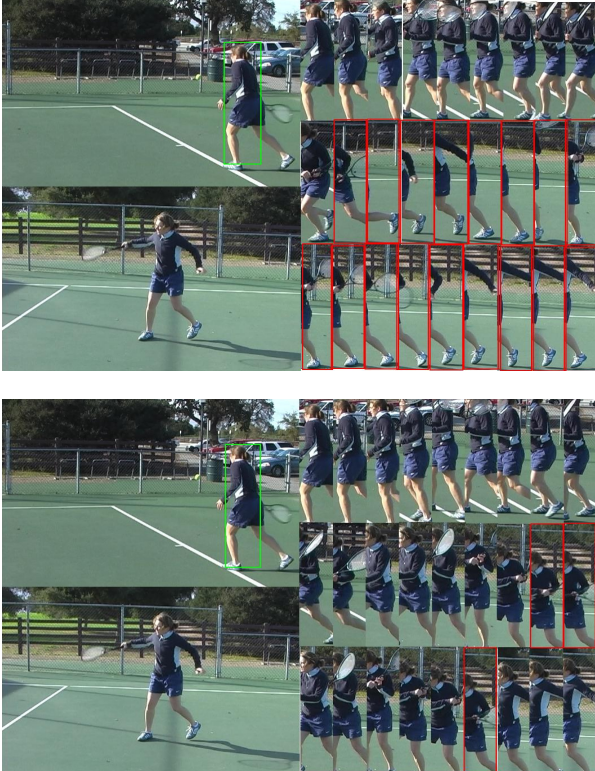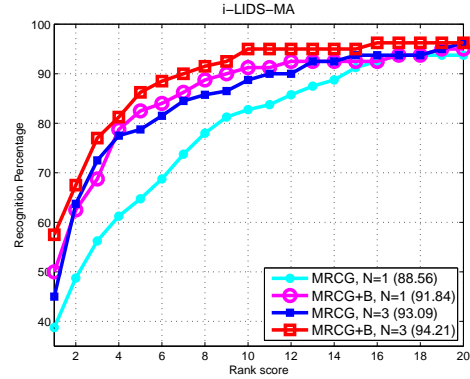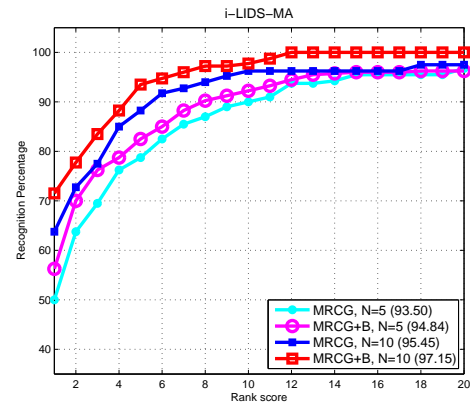

Figure 5. Tennis player tracking results (every $5th$ frame) using: the covariance descriptor (upper image) and the Brownian descriptor (bottom image).

(a) $N = 1, 3$

(b) $N = 5, 10$

Figure 7. Performance comparison on i-LIDS-MA [3].

ance matrix in all experiments consistently. This confirms the expected increase in performance of the new descriptor and shows that Brownian descriptor is a richer meta-feature than the classical covariance for the appearance matching. Also notice a significant increase in performance for $N = 1$. This implies that the previously discussed theory has been successfully employed.

i-LIDS-AA [3] This dataset contains 100 individuals automatically detected and tracked in two cameras. Cropped images are noisy, which makes the dataset more challenging (*e.g.* detected bounding boxes are not accurately centered around the people, only part of the people is detected due to occlusion). For minimizing misalignment issues,

we employ *discriminatively trained deformable part models* [9, 11], which slightly improve detection accuracy.

Brownian descriptor w.r.t. the metric Covariance descriptor as a positive definite and symmetric matrix is usually assumed to lie on a Riemannian manifold. Hence, computing distance between two matrices, we need to solve the generalized eigenvalues problem (*geodesic distance* [10]), which brings a computational burden. The geodesic distance can be also approximated by mapping matrices to a tangent plane [21] and then computing the L_1 norm. Fi-
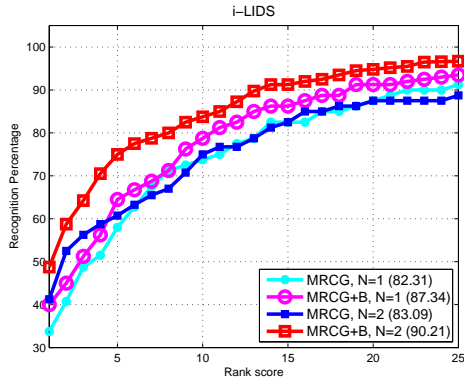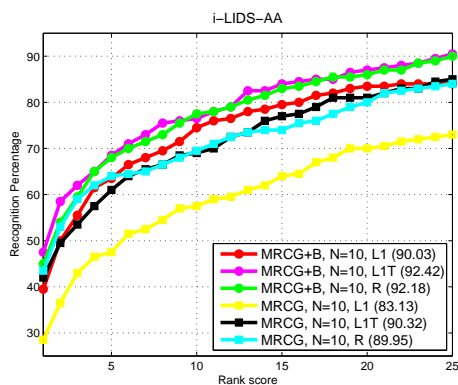
Figure 9. Performance comparison on i-LIDS [25].

Figure 8. Performance comparison on i-LIDS-AA [3] using different metrics: $L1$ - L_1 norm, $L1T$ - L_1 norm on a tangent plane, R - geodesic distance [10]

.

nally we can also treat the matrix as a simple one dimensional vector and apply directly the $L1$ norm.

Fig. 8 illustrates the performance w.r.t. the selected metric. Brownian descriptor not only outperforms the original covariance in every case, but also achieves the good performance even while using the L_1 norm. *This is the central point* which opens new directions for future research. Until now, the covariance descriptor was often avoided for employing to retrieval and to recognition tasks due to following limitations. First, the geodesic metric is computationally heavy. Second, learning on manifold is difficult and usually produces models which are over-fitted to training data. But now we can utilize this Brownian descriptor in combination with common classifiers employed in a Euclidean space.

i-LIDS [25] This evaluation dataset has been extracted automatically. It contains 476 images with 119 individuals registered by two different cameras. This dataset is very challenging due to many occlusions (often only the top part of the person is visible). We reproduce the same experimental settings as [4] and the results are shown in fig. 9. We can notice that our descriptor again outperforms the regular co-

variance descriptor. It clearly shows the advantage of our Brownian matrix over the classical covariance descriptor.

5.3. Descriptor complexity

Extraction complexity Time and memory complexity is the same as for computing the classical covariance matrix [22]. The computation complexity is $O(n^2 p)$. For fast covariance computation, similarly to [22], we can construct integral images (integral images need to be extracted for each $\bar{a}$ in equation 7).

Matching complexity Instead of using *geodesic distance*, we can employ directly Euclidean metric that makes the descriptor computationally efficient.

6. Conclusion

We proposed Brownian descriptor for the problem of appearance matching. This new descriptor is based on *distance correlation* $\mathcal{R}_n^2$, which can measure dependence between features extracted from an image region. In contrast to the classical covariance which measures the degree of linear relationship between features, the distance correlation measures the degree of *all kinds of possible relationships* between features. Experiments demonstrate that $\mathcal{R}_n^2$ outperforms ρ, as an image descriptor. We also show that Brownian descriptor can be treated as a point on Euclidean space without relevant loss of the recognition accuracy. This makes the descriptor computationally efficient and useful for real-time systems. In future we plan to address various generalizations of this descriptor to different application domains including object detection and recognition.

References

[1] S. Bak, G. Charpiat, E. Corvee, F. Bremond, and M. Thonnat. Learning to match appearances by correlations in a covariance metric space. In *ECCV*, 2012.

[2] S. Bak, E. Corvee, F. Bremond, and M. Thonnat. Person re-identification using spatial covariance regions of human body parts. In *AVSS*, 2010.

[3] S. Bak, E. Corvee, F. Bremond, and M. Thonnat. Boosted human re-identification using riemannian manifolds. *Image and Vision Computing*, 2011.

[4] S. Bak, E. Corvee, F. Bremond, and M. Thonnat. Multiple-shot human re-identification by mean riemannian covariance grid. In *AVSS*, 2011.

[5] N. Bakirov and G. Székely. Brownian covariance and central limit theorem for stationary sequences. *Theory Probab. Appl*, 55(3):371 – 394, 2011.

[6] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *CVIU*, 2008.

[7] T. Brox and J. Malik. Object segmentation by long term analysis of point trajectories. In *ECCV*, 2012.

[8] Y. Cai, V. Takala, and M. Pietikainen. Matching groups of people by covariance descriptor. In *Proceedings of the 20th*

International Conference on Pattern Recognition, ICPR, pages 2744–2747. IEEE Computer Society, 2010.

[9] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, 2010.

[10] W. Förstner and B. Moonen. A metric for covariance matrices. In *Quo vadis geodesia ...?, Festschrift for Erik W. Grafarend on the occasion of his 60th birthday, TR Dept. of Geodesy and Geoinformatics, Stuttgart University*, 1999.

[11] R. B. Girshick, P. F. Felzenszwalb, and D. McAllester. Discriminatively trained deformable part models, release 5. http://people.cs.uchicago.edu/ rbg/latent-release5/.

[12] D. Gray, S. Brennan, and H. Tao. Evaluating Appearance Models for Recognition, Reacquisition, and Tracking. *PETS*, 2007.

[13] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof. Person re-identification by descriptive and discriminative classification. In *SCIA*, pages 91–102, 2011.

[14] T. Kaneko and O. Hori. Template update criterion for template matching of image sequences. In *CVPR*, volume 2, 2002.

[15] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004.

[16] B. Ma, Y. Su, and F. Jurie. Bicov: a novel image representation for person re-identification and face verification. In *Proceedings of the 23rd British Machine Vision Conference*, BMVC, 2012.

[17] I. Matthews, T. Ishikawa, and S. Baker. The template update problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:810–815, 2004.

[18] R. Messing, C. Pal, and H. Kautz. Activity recognition using the velocity histories of tracked keypoints. In *ICCV '09: Proceedings of the Twelfth IEEE International Conference on Computer Vision*, Washington, DC, USA, 2009. IEEE Computer Society.

[19] F. P. Oncel, F. Porikli, O. Tuzel, and P. Meer. Covariance tracking using model update based on lie algebra. In *CVPR*, 2006.

[20] G. J. Székely and M. L. Rizzo. Brownian distance covariance. *The Annals of Applied Statistics*, 3(4):1236–1265, 2009.

[21] D. Tosato, M. Farenzena, M. Spera, V. Murino, and M. Cristani. Multi-class classification on riemannian manifolds for video surveillance. In *ECCV*, 2010.

[22] O. Tuzel, F. Porikli, and P. Meer. Region covariance: A fast descriptor for detection and classification. In *ECCV*, pages 589–600, 2006.

[23] O. Tuzel, F. Porikli, and P. Meer. Pedestrian detection via classification on riemannian manifolds. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30:1713–1727, October 2008.

[24] J. Yao and J.-M. Odobez. Fast human detection from joint appearance and foreground feature subset covariances. *Computer Vision and Image Understanding*, 115(3):1414–1426, 2011.

[25] W.-S. Zheng, S. Gong, and T. Xiang. Associating groups of people. In *BMVC*, 2009.