Local Log-Euclidean Covariance Matrix (L^2ECM) for Image Representation and Its Applications

Peihua Li and Qilong Wang

Heilongjiang University, School of CS, School of EE, China
peihualj@hotmail.com

Abstract. This paper presents Local Log-Euclidean Covariance Matrix (L^2ECM) to represent neighboring image properties by capturing correlation of various image cues. Our work is inspired by the structure tensor which computes the second-order moment of image gradients for representing local image properties, and the Diffusion Tensor Imaging which produces tensor-valued image characterizing the local tissue structure. Our approach begins with extraction of raw features consisting of multiple image cues. For each pixel we compute a covariance matrix in its neighboring region, producing a tensor-valued image. The covariance matrices are symmetric and positive-definite (SPD) which forms a Riemannian manifold. In the Log-Euclidean framework, the SPD matrices form a Lie group equipped with Euclidean space structure, which enables common Euclidean operations in the logarithm domain. Hence, we compute the covariance matrix logarithm, obtaining the pixel-wise symmetric matrix. After half-vectorization we obtain the vector-valued L^2ECM image, which can be flexibly handled with Euclidean operations while preserving the geometric structure of SPD matrices. The L^2ECM features can be used in diverse image or vision tasks. We demonstrate some applications of its statistical modeling by simple second-order central moment and achieve promising performance.

1 Introduction

Characterizing local image properties is of great research interest in recent years [1]. The local descriptors can be either used sparsely for describing region of interest (ROI) extracted by the region detectors [2], or be used at dense grid points for image representation [3, 4]. They play a fundamental role for success of many middle-level or high-level vision tasks, e.g., image segmentation, scene or texture classification, image or video retrieval and person recognition. It is challenging to present feature descriptors which are highly distinctive and robust to photometric or geometrical transformations.

The motivation of this paper is to present a general kind of image descriptors for representing local image properties by fusing multiple cues. We are inspired by the structure tensor method [5, 6] and Diffusion Tensor Imaging (DTI) [7], both concerning tensor-valued (matrix-valued) images and enjoying important applications in image or medical image processing. The former computes second-order moment of image gradients at every pixel, while the latter associates to

each voxel a 3-D symmetric matrix describing the molecular mobility along three directions and correlations between these directions. Our methodology consists in the pixel-wise covariance matrix for representing the local image correlation of multiple image cues. This leads to the model of Local Log-Euclidean Covariance matrix, L^{2}ECM. It can fuse multiple image cues, e.g., intensity, lower- or higher-order gradients, spatial coordinates, texture, etc. In addition, it is insensitive to scale and rotation changes as well as illumination variation. Comparison of structure tensor, DTI and L^{2}ECM is presented in Table 1. Our model can be seen as a generalization of the structure tensor method; it can also be interpreted as a kind of "imaging" method, in which each pixel point is associated with logarithm of a covariance matrix, describing the local image properties through the correlation of multiple image cues.

Table 1. Comparision of Tensor-valued (Matrix-valued) images

Structure Tensor	DTI	L^{2}ECM
$\begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{bmatrix}$	$\begin{bmatrix} D_{11} & D_{12} & D_{13} \\ D_{12} & D_{22} & D_{23} \\ D_{13} & D_{23} & D_{33} \end{bmatrix}$	$\mathbf{C} = \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \vdots & \vdots \\ C_{n1} & \cdots & C_{nn} \end{bmatrix} \xrightarrow[\text{vlog}]{} \begin{bmatrix} \text{vlog}\mathbf{C}_1 & \cdots & \text{vlog}\mathbf{C}_{m-n+1} \\ & \ddots & \vdots \\ & & \text{vlog}\mathbf{C}_m \end{bmatrix}$
2^{nd}-order moment of partial derivatives of image I w.r.t x, y.	3×3 symmetric matrix describing molecules diffusion	Logarithm of $n\times n$ (n=2∼5) covariance matrix $\mathbf{C}$ of raw features followed by half-vectorization ($m = (n^2+n)/2$ due to symmetry)

Fig. 1(a) shows the outline of the modeling methodology of L^{2}ECM. Given an image $I(x, y)$, we extract raw features for all pixels, obtaining a feature map $\mathbf{f}(x, y)$ which may contain intensity, lower- or higher-order derivatives, spatial coordinates, texture, etc. Then for every pixel point, one covariance matrix is computed using the raw features in the neighboring region around this pixel. We thus obtain a tensor-valued (matrix-valued) image $\mathbf{C}(x, y)$. By computing the covariance matrix logarithm followed by half-vectorization, we obtain the vector-valued L^{2}ECM image which can be handled with Euclidean operations. For illustration purpose, 3-dimensional (3-D) raw features are used which comprising intensity and the absolute values of partial derivatives of image I w.r.t x, y; the resulting L^{2}ECM features are 5-D and the corresponding slices are shown at the bottom-right in Fig. 1(a). Refer to section 3 for details on L^{2}ECM features.

The covariance matrices are symmetric and positive-definite (SPD), the spaces of which is not a Euclidean space but a smooth Riemannian manifold. In the Log-Euclidean framework [8], the SPD matrices form a commutative Lie group which is equipped with a Euclidean structure. This framework enables us to compute the logarithms of SPD matrices, which can then be flexibly and efficiently handled with common Euclidean operations. Our technique avoids complicated and computationally expensive operations such as the geodesic distance, intrinsic mean computation or statistical modeling *directly* in Riemannian manifold. Instead, we can conveniently operate in the Euclidean space.

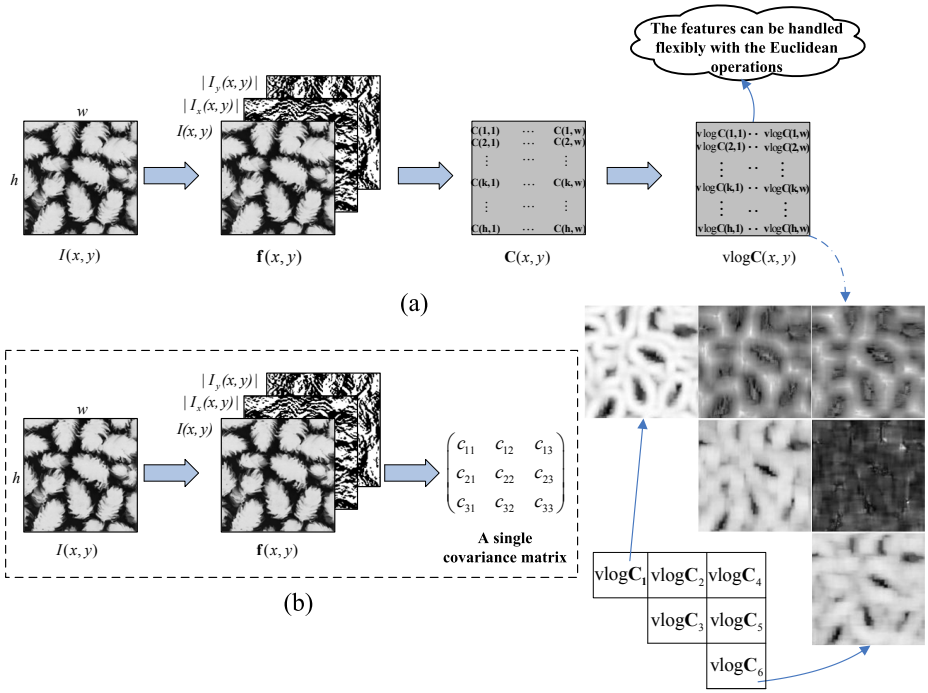

Fig. 1. Overview of L^{2}ECM (3-D raw features are used for illustration). (a) shows the modeling methodology of L^{2}ECM. Given an image $I(x, y)$, the raw feature image $\mathbf{f}(x, y)$ is first extracted; then the tensor-valued image $\mathbf{C}(x, y)$ is obtained by computing the covariance matrix for every pixel; after the logarithm of $\mathbf{C}(x, y)$, the symmetric matrix $\log \mathbf{C}(x, y)$ is vectorized to get the 6-D vector-valued image denoted by $\mathrm{vlog}\mathbf{C}(x, y)$, slices of which are shown at the bottom-right. Refer to Section 3 for details on L^{2}ECM. (b) shows the modeling methodology of Tuzel et al. [9]–only one *global* covariance matrix is computed for the overall image of interest.

The covariance matrix as a region descriptor was first proposed by Tuzel et al. [9], the modeling methodology of which was illustrated in Fig. 1(b). The main difference is their *global* modeling vs our *local* one: the former computes one covariance matrix for the overall image region of interest; while the latter computes a covariance matrix for every pixel point. Above all, Tuzel et al. employed affine-Riemannian metric [10] which is known to have high computational cost in computations of geodesic distance and statistics, e.g., the intrinsic mean, the second-order moment, mixture models or Principal Component Analysis (PCA), in the Riemannian space [11]. This may make the modeling of pixel-wise covariance matrices prohibitively inefficient. By contrast, we employ Log-Euclidean metric [8] which transforms the operations from the Riemannian space into Euclidean one so that the above-mentioned limitations are overcome.

The remainder of this paper is organized as follows. Section 2 reviews the papers related to our work. Section 3 describes in detail the L^{2}ECM features

and the underlying theory. Applications of statistics modeling of L^2ECM by the second-order central moment are presented in section 4. The conclusion is given in section 5.

2 Related Work

The natural choice for region representation is to vectorize the intensities or gradients [12] of pixels in the region. The most commonly used techniques are modeling image gradient distributions through histograms. The Scale Invariant Feature Transform (SIFT) descriptor [13] uniformly partitions an image region into small cells and in each one an orientation histogram is computed, which results in a 3D histogram of image gradient position and angles. The Gradient Location and Orientation Histogram (GLOH) [1] is an extension of the SIFT which quantizes the patch into log-polar cells instead of cartesian ones. Similar techniques are also used in Histogram of Orientation Gradient (HOG) that are particularly suitable for human detection [14]. SURF [15] and DAISY [16] descriptors retain the strengths of SIFT and GLOH and can be computed quickly at every pixel point. Other approaches may be based on spatial-frequency analysis, image moments, etc. For an overview of local descriptors and their performance evaluation, one may refer to [1]

Based on the *global* covariance descriptor, Tuzel et. al presented approaches for human detection, texture classification or object tracking [9, 17, 18]. The underlying theory of their approaches is the computation of intrinsic distance and mean of covariance matrices by the affine-invariant Riemannian metric. In [19], a novel online subspace learning method is presented for adapting to appearance changes during tracking. They employed Log-Euclidean metric rather than affine-invariant Riemannian one so that the mean and PCA can be computed efficiently in the Euclidean space. By augmenting the lower triangular matrix of Cholesky factorization of the covariance matrix and the mean vector, the Shape of Gaussian (SOG) descriptor is introduced and its Lie group structure is analyzed [20]. Nakayama et al. [21] used both covariance matrix and the mean vector via Gaussian distribution to express statistics of dense features such as SIFT [13] or SURF [15]; they presented some kernel metrics based on the theory of information geometry for scene categorization. A novel method was proposed in [22] for sparse decomposition of covariance matrices. The above papers employed the covariance matrix as a *global* descriptor in the sense that one covariance matrix is computed for representing the overall object region.

The structure tensor has a long history [5] and enjoys successful applications ranging from image segmentation, motion estimation to corner detection. Because DTI is able to provide at the microscopic level the information on tissue structure, it has been applied in many studies on neuroscience, neurodevelopment, neurodegenerative studies etc. Review of the two techniques are beyond the scope of this paper. One may refer to [6] for recent advances and development on structure tensor and DTI.

3 Local Log-Euclidean Covariance Matrix ($\mathbf{L^2}$ECM)

The space of SPD matrices is not a vector space but a Riemannian manifold (an open convex half-cone). Hence, the conventional Euclidean operations, e.g., the Euclidean distance, mean or the statistics do not apply. Two class of Riemannian framework have been presented for dealing with SPD matrices: the affine-invariant Riemannian framework [10, 23] and the Log-Euclidean Riemannian framework [8]. The latter has almost the same good theoretical properties as the former, and in the meantime enjoys a drastic reduction in computational cost. In the following we first introduce briefly the Log-Euclidean Framework (refer to [8] for details) and then present the proposed L^2ECM features.

3.1 Log-Euclidean Framework on SPD Matrices

Matrix exponential and logarithm The matrix exponential and logarithm are fundamental to the Log-Euclidean framework. Let $\mathcal{SPD}(n)$ and $\mathcal{S}(n)$ denote the space of $n \times n$ real SPD matrices and $n \times n$ real symmetric matrices, respectively. Any matrix $\mathbf{S} \in \mathcal{S}(n)$ has the eigen-decomposition of the form $\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, where $\mathbf{U}$ is an orthonormal matrix and $\mathbf{\Lambda} = \mathrm{Diag}(\lambda_1, \cdots, \lambda_n)$ is a diagonal matrix composed of the eigenvalues λ_i of $\mathbf{S}$. Furthermore, if $\mathbf{S}$ is positive-definite, i.e., $\mathbf{S} \in \mathcal{SPD}(n)$, then $\lambda_i > 0$ for $i = 1, \ldots, n$. The exponential map, $\exp\colon \mathcal{S}(n) \mapsto \mathcal{SPD}(n)$, is bijective, i.e., one to one and onto. By eigen-decomposition, the exponential of a $\mathbf{S} \in \mathcal{S}(n)$ can be computed as

$$\exp(\mathbf{S}) = \mathbf{U} \cdot \mathrm{Diag}(\exp(\lambda_1), \ldots, \exp(\lambda_n)) \cdot \mathbf{U}^T \tag{1}$$

For any SPD matrix $\mathbf{S} \in \mathcal{SPD}(n)$, it has a unique logarithm $\log(\mathbf{S})$ in $\mathcal{S}(n)$:

$$\log(\mathbf{S}) = \mathbf{U} \cdot \mathrm{Diag}(\log(\lambda_1), \ldots, \log(\lambda_n)) \cdot \mathbf{U}^T \tag{2}$$

The exponential map, $\exp\colon \mathcal{S}(n) \mapsto \mathcal{SPD}(n)$, is diffeomorphism, i.e., the bijective map exp and its inverse log are both differentiable.

Lie group structure on $\mathcal{SPD}(n)$ For any two matrices $\mathbf{S}_1, \mathbf{S}_2 \in SPD(n)$, define the logarithmic multiplication $\mathbf{S}_1 \odot \mathbf{S}_2$ as

$$\mathbf{S}_1 \odot \mathbf{S}_2 \triangleq \exp(\log(\mathbf{S}_1) + \log(\mathbf{S}_2)) \tag{3}$$

It can be shown that $\mathcal{SPD}(n)$ with the associated logarithmic multiplication $\odot$ satisfies the following axioms: (1) $\mathcal{SPD}(n)$ is closed, i.e., $\mathbf{S}_1 \in \mathcal{SPD}(n)$ and $\mathbf{S}_2 \in \mathcal{SPD}(n)$ implies $\mathbf{S}_1 \odot \mathbf{S}_2 \in \mathcal{SPD}(n)$. (2) $\odot$ is associative: $(\mathbf{S}_1 \odot \mathbf{S}_2) \odot \mathbf{S}_3 = \mathbf{S}_1 \odot (\mathbf{S}_2 \odot \mathbf{S}_3)$. (3) For the $n \times n$ identity matrix $\mathbf{I}$, $\mathbf{S} \odot \mathbf{I} = \mathbf{I} \odot \mathbf{S} = \mathbf{S}$. (4) For matrix $\mathbf{S}$ and its inverse $\mathbf{S}^{-1}$, $\mathbf{S} \odot \mathbf{S}^{-1} = \mathbf{S}^{-1} \odot \mathbf{S} = \mathbf{I}$. (5) It is commutative, that is, $\mathbf{S}_1 \odot \mathbf{S}_2 = \mathbf{S}_2 \odot \mathbf{S}_1$. In viewing of the above, $\mathcal{SPD}(n)$ is a commutative Lie group. Actually, $(\mathcal{SPD}(n), \odot, \mathbf{I})$ is isomorphic to its Lie algebra $\mathfrak{spd}(n) = \mathcal{S}(n)$.

Vector space structure on $\mathcal{SPD}(n)$ The commutative Lie group $\mathcal{SPD}(n)$ admits a bi-invariant Riemannian metric and the distance between two matrices $\mathbf{S}_1, \mathbf{S}_2$ is

$$d(\mathbf{S}_1, \mathbf{S}_2) = \big\|\log(\mathbf{S}_1) - \log(\mathbf{S}_2)\big\| \tag{4}$$

where $||\cdot||$ is the Euclidean norm in the vector space $\mathcal{S}(n)$. This bi-invariant metric is called Log-Euclidean metrics, which is invariant under similarity transformation. For a real number λ, define the logarithmic scalar multiplication between λ and a SPD matrix $\mathbf{S}$

$$\lambda \circledast \mathbf{S} \triangleq \exp(\lambda \log(\mathbf{S})) = \mathbf{S}^{\lambda} \tag{5}$$

Provided with logarithmic multiplication (3) and logarithmic scalar multiplication (5), the $\mathcal{SPD}(n)$ is equipped with a vector space structure.

The desirable property of such a vector space structure of $\mathcal{SPD}(n)$ is that, by matrix logarithm, the Rimannian manifold of SPD matrices is mapped to the Euclidean space. As such, in the logarithmic domain, the SPD matrices can be handled with simple Euclidean operations and, if necessary, the results can be mapped back to the Riemannian space via the matrix exponential. This greatly facilitates the statistical analysis of SPD matrices, for example, the geometric mean of N SPD matrices $\mathbf{S}_1, \cdots, \mathbf{S}_N$ is the algebraic mean (the zeroth-order moment) of their logarithms mapped back to $\mathcal{SPD}(n)$, which has a closed form $\mathbb{E}_{\text{LE}}(\mathbf{S}_1, \cdots, \mathbf{S}_N) = \exp\left(\frac{1}{N}\sum_{i=1}^{N} \log(\mathbf{S}_i)\right)$.

3.2 L^{2}ECM Feature Image

We describe the feature descriptor of L^{2}ECM with the gray-level image. Given an image of interest, $I(x,y), (x,y) \in \Omega$, we can extract the raw feature vector $\mathbf{f}(x,y)$ which, for example, has the following form:

$$\mathbf{f}(x,y) = \left[I(x,y)\ |I_x(x,y)|\ |I_y(x,y)|\ |I_{xx}(x,y)|\ |I_{yy}(x,y)|\right]^T \tag{6}$$

where $|\cdot|$ denotes the absolute value, I_x (resp. I_{xx}) and I_y (resp. I_{yy}) denote the first (resp. second)-order partial derivative with respect to x and y, respectively. Other image cues, e.g., spatial coordinates, gradient orientation or texture feature can also be included in the raw features. For a color image, the gray-levels of three channels may be combined as well.

Provided with the raw feature vectors, we can obtain a tensor-valued image by computing the covariance matrix function $\mathbf{C}(x,y)$ at every pixel

$$\mathbf{C}(x,y) = \frac{1}{N_{S_r} - 1} \sum_{(x',y') \in S_r(x,y)} (\mathbf{f}(x',y') - \bar{\mathbf{f}}(x,y))(\mathbf{f}(x',y') - \bar{\mathbf{f}}(x,y))^T \tag{7}$$

$$\bar{\mathbf{f}}(x,y) = \frac{1}{N_{S_r}} \sum_{(x',y') \in S_r(x,y)} \mathbf{f}(x',y')$$

where $S_r(x,y) = \{(x',y') || x - x'| \leq r/2, |x - x'| \leq r/2, (x,y) \in \Omega\}$ and N_{s_r} denotes the number of points inside S_r. In practice, there may exist covariance matrices which are not positive-definite and which we should take care in computations. From our experience with computation of a huge number of covariance matrices, involved in human detection and texture classification in real images,

this violation is really a rare event. With increase of r, the local image structure in larger scales will be captured. Small r may be helpful in capturing fine local structure, which may however result in the singularity of the covariance matrix due to insufficient number of samples for estimation. In our experiments $r \geq 16$ is appropriate for most applications. The covariance matrix is robust to noise and insensitive to changes of scale, rotation and illumination [9].

We wish to exploit these covariance matrices as fundamental features for vision applications. It is known that the Affine-Riemannian framework involves intensive computations of matrix square root, matrix inverse, matrix exponential and logarithm. Hence, we utilize the competent Log-Euclidean framework: $\mathbf{C}(x, y)$ in the commutative Lie group $\mathcal{SPD}(n)$ is mapped by matrix logarithm to $\log \mathbf{C}(x, y)$ in the vector space of $\mathcal{S}(n)$. $\log \mathbf{C}(x, y)$ can then be handled with the Euclidean operations and the intensive computations involved in Affine-Riemannian framework are avoided. It also facilitates greatly further analysis or modeling of the SPD matrices.

From the tensor-valued image $\mathbf{C}(x, y) \in \mathcal{SPD}(n)$, we compute the logarithm of the covariance matrix $\mathbf{C}(x, y)$ according to Eq. (2). $\log \mathbf{C}(x, y)$ is a symmetric matrix of Euclidean space, i.e., $\log \mathbf{C}(x, y) \in \mathcal{S}(n)$. Because of its symmetry, we perform half-vectorization of $\log \mathbf{C}(x, y)$, denoted by $\mathrm{vlog}\mathbf{C}(x, y)$, i.e., we pack into a vector in the column order the upper triangular part of $\log \mathbf{C}(x, y)$. The final L^2ECM feature descriptor can thus be represented as

$$\mathrm{vlog}\mathbf{C}(x, y) = \left[\mathrm{vlog}\mathbf{C}_1(x, y)\ \mathrm{vlog}\mathbf{C}_2(x, y)\ \ldots\ \mathrm{vlog}\mathbf{C}_{n(n+1)/2}(x, y)\right] \tag{8}$$

The covariance matrices can be computed efficiently via the Integral Images [9]. The computational complexity of constructing the integral images is $O(|\Omega|n^2)$, where $|\Omega|$ is the image size. The complexity of eigen-decomposition of all covariance matrices is $O(|\Omega|n^3)$. In practice, n=2∼5 may suffice for most problems. In particular, when n=2 or n=3, the eigen-decomposition of covariance matrices can be obtained analytically; hence, the L^2ECM features, which are 3-D (n=2) or 6-D (n=3), can be computed fast through integral images and closed-form of eigen-decomposition.

The L^2ECM may be used in a number of ways:

- It may be seen as "imaging" technology by which various novel multi-channel images are produced. When $n = 2$, by combinations of varying raw features, e.g., I_x and I_y, I_{xx} and I_{yy}, or I and $\sqrt{I_x^2 + I_y^2}$, we obtain different 3-D "color" images that may be suitable for a wide variety of image or vision tasks. Considering its desirable properties, it is particularly interesting to use the 3-D L^2ECM images instead of the original ones for object tracking [24, 25]. In the cases of larger n, we achieve compact L^2ECM features which can be densely sampled and packed as region descriptors.
- It enables a wide variety of statistical techniques applicable to covariance matrices. For example, it is straightforward to model L^2ECM features by probabilistic mixture models, e.g, Gaussian mixture model (GMM), principal component analysis, etc. This way, the geometric structure of covariance

matrices is preserved while avoiding computationally expensive algorithms in Riemannian space [26, 23].
– We can describe the region statistics simply by the first-order (mean) or second-order moments (covariance matrix) of L^2ECM features. Note that they actually represent the lower-order statistics of SPD matrices in the logarithm domain.

In the following, we demonstrate applications of statistical modeling of L^2ECM features by the second-order moment (covariance matrix).

4 Applications of L^2ECM Features

In this section, applications of L^2ECM features are exhibited to human detection, texture classification and object tracking.

4.1 Human Detection

For performance evaluation, we exploit the INRIA person dataset [14], a challenging benchmark dataset containing large changes in the pose and appearance, partial occlusions, illumination variation and cluttered background. It includes 2416 positive, normalized images and 1218 person-free images for training, together with 288 images of humans and 453 person-free images for testing.

The normalized image is of 96×160 pixels and the centered 64×128 pixels window is used. This way, the boundary effects are eliminated. For a normalized image, we first compute the L^2ECM feature image ($r = 16$). Then we divide the vector-valued image into 12 overlapping, 32×32 blocks with a stride of 16 pixels. We compute for each block the second-order moment (covariance matrix) which is again subject to matrix logarithm and half-vectorization. The resulting feature for the whole, normalized image is a 1440-dimensional vector. We exploit the linear SVM [27] with default parameters for classification. Our training process is similar to that of Dalal and Tiggs [14].

Fig. 2 shows the *Detection Error Tradeoff* (DET) curves on a log-log scale of the proposed method and those of [14] and [17]. The DET curves of other methods are produced from their respective papers. It is clear that our method is superior to the methods that uses the linear or kernel SVM proposed in [14]. Our method is also better than the method of classification on Riemannian manifold [17] when FFPW$\geq7\times10^{-4}$, but the method of [17] outperforms when FFPW$<7\times10^{-4}$. At 10^{-4} FFPW, our method has the lowest miss rate of 5.7% while [17] has the second lowest of 6.8%.

4.2 Texture Classification

In this section, we apply the L^2ECM features to texture classification. The Brodatz database and KTH-TIPS database [28] are used for performance evaluation.

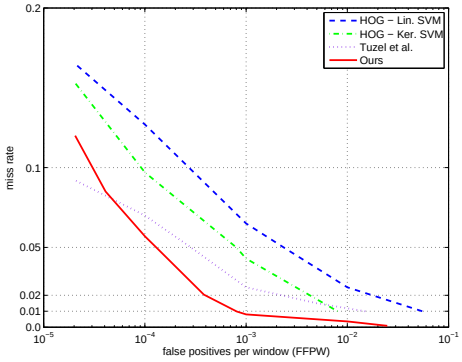

Fig. 2. DET curves of human detection in the INRIA person dataset.

Brodatz Dataset The Brodatz dataset contains 111 textures (texture D14 is missing); each texture is represented by one 640×640 image. Note that though the dataset does not integrate scale, viewpoint or illumination changes, it includes non-homogenous textures which pose difficulty for classification.

For each texture, the corresponding 640×640 image is divided into 49 overlapping blocks, each of which is of 160×160 pixles with a block stride of 80. Twenty four images of each class are randomly selected for training and the remaining ones for testing. For each image, we first compute the L^2ECM feature image; it is then divided into four 80×80 patches, the covariance matrices of which are computed; the resultant four covariance matrices are subject to logarithmic operation and half-vectorization. When testing, each covariance matrix of the testing image is compared to all covariance matrices of the training set and its label is determined by KNN algorithm ($k = 5$). The maximum votes of the four matrices associated with this testing image determine its classification. To avoid bias, the experiments are repeated twenty times and the result is computed as the average value and standard deviation over the twenty runs.

Fig. 3(a) shows the classification accuracy of our method and the following six texture classification approaches: Harris detector+Laplacian detector+SIFT descriptor+SPIN descriptor (HLSS) [29], Lazebnik's method [30], VZ-joint [31], and Jalba's method [32], Hayman's method [28], and Tuzel et al.'s method [9]. The data of other methods are duplicated from their respective papers. Our method obtains the best classification accuracy; the random covariances method of Tuzel et al. obtains the second best accuracy which, however, does not report standard deviation. These two are far better than the remaining ones.

KTH-TIPS Database[28] This dataset is challenging in the sense that it contains varying scale, illumination and pose. There are 10 texture classes each of which is represented by 81 image samples. The size of samples is 200×200 pixels.

The classification method is similar to that in the Brodatz database. For each sample, the L^2ECM feature image is computed ($r = 32$). Every feature image is

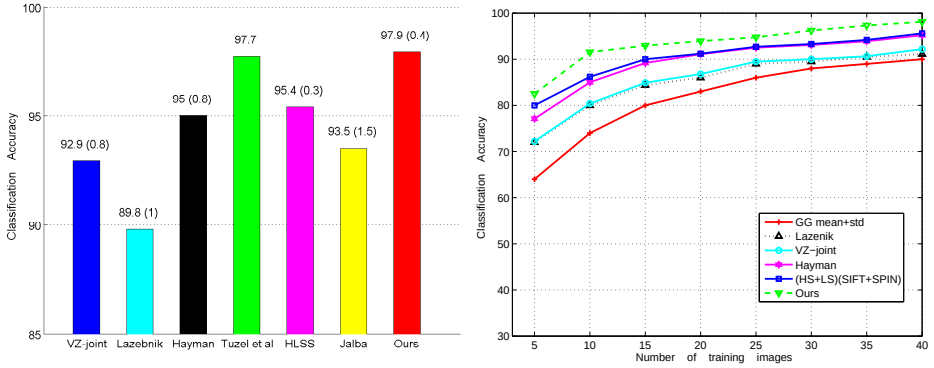

Fig. 3. Texture classification in the Brodatz (a) and KTH-TIPS (b) databases.

uniformly divided into four blocks and a global covariance matrix is computed on each block. The logarithms of these covariance matrices are subject to half-vectorization.

Fig. 3(b) shows the classification accuracy against the number of training images. For comparison, the classification results of the following methods are also shown [29]: Lazebnik's method [30], VZ-joint [31], Hayman's method [28], global Gabor Filters(GG) [33], and Harris detector+Laplacian detector+SIFT descriptor+SPIN descriptor((HS+LS)(SIFT+SPIN))[29]. It can be seen that the classification accuracy of our method is much better than all the others.

4.3 Object Tracking

Based on the covariance matrix as a global region descriptor, we compare our method based on L^2ECM (L^2ECM tracker for short) and that of Tuzel et al. [18] (Tuzel tracker for short). Fig. 4 shows different covariance matrix computation in these two methods. In L^2ECM tracker, 15×15 covariance matrices are computed from the L^2ECM image ($r = 16$, 5-D raw features (6) and 15-D L^2ECM feature); in Tuzel tracker, 9×9 covariance matrices are computed from the raw feature image, in which every 9-D feature vector comprises x- and y-coordinates, intensity, the first- and second-order partial derivatives w.r.t x and y, gradient magnitude and orientation. Initialization of two trackers are both by hand in the first frame and three image sequences are used for comparison. The average tracking time of L^2ECM tracker is approximately one half of that of Tuzel tracker. The most time-consuming procedure of Tuzel tracker is the model update [18, Sec. 2.4], which involves the mean computation via the affine-invariant Riemannian metrics known to be iterative and very expensive [10, 11].

Car sequence This sequence (190 frames, size: 384×288) concerns a surveillance scenario, available at http://www.cvg.cs.rdg.ac.uk/PETS2001/pets2001-

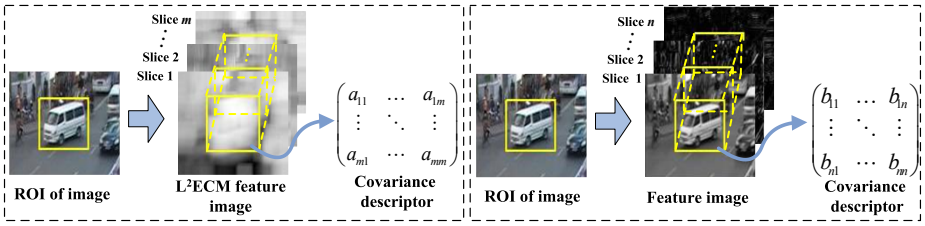

Fig. 4. Covariance matrix computation in the L^{2}ECM (left) and Tuzel (right) trackers.

dataset.html. The black car running towards the left is the object of interest. In the sequence, there exist pose variation of the object and cars nearby in the background which have similar appearances with the object. Both trackers succeed in following the object throughout the sequence. As seen in Table 2, the L^{2}ECM has smaller tracking errors than the Tuzel tracker. Some sample tracking results are shown in Fig. 5 (top pannel).

Face sequence In the face sequence (1480 frames, size: 320×240) the object undergoes large illumination changes [34]. We track the object every four frames (so there are 370 frames in total) because the object motion between consecutive frames is very small. Though both trackers can follow the object across the sequence, the L^{2}ECM tracker has much smaller tracking error than the Tuzel tracker. The average error is presented in Table 2. Fig. 5 (middle panel) shows some sample tracking results.

Mall sequence This sequence (190 frames, size: 360×288) is filmed by a moving camera in a mall [35]. The object (an adult along with a child) has large non-rigid pose variations and illumination changes. There also occur partial occlusions in the sequence. The Tuzel tracker loses the object around frame #116 despite the model update scheme, while the L^{2}ECM tracker successfully follows the object across the sequence. It can be seen from Table 2 that the L^{2}ECM tracker's error is much smaller than the Tuzel tracker. Some sample results are shown in Fig. 5 (bottom panel).

Table 2. Comparison of average tracking errors (mean±std) and number of successful frames vs total frames.

Image seq.	Method	Dist. err. (pixels)	Succ. frames
Car seq.	Tuzel	10.76± 5.72	190/190
	L^{2}ECM	7.55± 3.38	190/190
Face seq.	Tuzel	20.44±13.87	370/370
	L^{2}ECM	4.45± 3.87	370/370
Mall seq.	Tuzel	30.92±16.58	116/190
	L^{2}ECM	17.67±10.45	190/190

Fig. 5. Tracking results in the car sequence (top panel), face sequence (middle panel) and mall sequence (bottom panel). In each panel, the results of Tuzel tracker and L^2ECM tracker are shown in the first and second rows, respectively.

5 Conclusion

The L^2ECM is analogous to the structure tensor and Diffusion tensor, which can be seen as an "imaging" technology in producing novel multi-channel images that capture the local correlation of multiple cues in the original image. We compute a tensor-valued image from the original one in which each point is associated with a covariance matrix computed locally. Through matrix logarithm and half-vectorization we obtain the vector-valued, L^2ECM feature image. The theoretical foundation of L^2ECM is the Log-Euclidean framework, which endows the commutative lie group formed by the SPD matrices with a linear space structure. This enables the common Euclidean operations of covariance matrices in the logarithmic domain while preserving their geometric structure.

The paper demonstrates applications of simple statistical modeling of L^2ECM by the second-order moment in human detection, texture classification and object tracking. We achieve comparable performance with or superior performance over the state-of-the-art methods. In the future work, we are interested in statistical modeling of L^2ECM by mixture models and study of its applications in diverse image or vision tasks.

Acknowledgements

The work was supported by the National Natural Science Foundation of China (60973080, 61170149), Program for New Century Excellent Talents in University (NCET-10-0151), Key Project by Chinese Ministry of Education (210063), and High-level professionals (innovative teams) of Heilongjiang University (Hdtd2010-07). We thank Qi Sun for assistance in the experiment on covariance tracking.

References

1. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. **27**(10) (2005) 1615–1630
2. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L.: A comparison of affine region detectors. Int. J. Comput. Vision **65**(1/2) (2005) 43–72
3. Liu, C., Yuen, J., Torralba, A.: SIFT flow: Dense correspondence across scenes and its applications. IEEE Trans. Pattern Anal. Mach. Intell. **33**(5) (2011) 978–994
4. Bosch, A., Zisserman, A., Munoz, X.: Scene classification via pLSA. In: Europ. Conf. on Comp. Vis. (2006) 517–530
5. Knutsson, H.: Representing local structure using tensors. In: Proc. 6th Scandinavian Conf. on Image Analysis. (1989) 244–251
6. Weichert, J., Hagen, H., eds.: Visualization and image processing of tensor fields. Berlin / Heidelberg: Springer (2006)
7. Bihan, D.L., Mangin, J., Poupon, C., Clark, C., Pappata, S., Molko, N.: Diffusion tensor imaging: Concepts and applications. J Magn. Reson. Imaging **66** (2001) 534–546
8. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Geometric means in a novel vector space structure on symmetric positive-definite matrices. SIAM J. Matrix Anal. Appl. (2006)
9. Tuzel, O., Porikli, F., Meer, P.: Region covariance: A fast descriptor for detection and classification. In: Europ. Conf. on Comp. Vis. (2006) 589–600
10. Pennec, X., Fillard, P., Ayache, N.: A Riemannian framework for tensor computing. Int. J. Comput. Vision (2006) 41–66
11. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Fast and simple calculus on tensors in the Log-Euclidean framework. In: Proc. of the 8th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention. (2005) 115–122
12. Yan, K., Sukthankar, R.: PCA-SIFT: a more distinctive representation for local image descriptors. In: Proc. Int. Conf. Comp. Vis. Patt. Recog. (2004) 506–513
13. Lowe, D.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision **60** (2004) 91–110

14. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proc. Int. Conf. Comp. Vis. Patt. Recog. (2005) 886–893
15. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (SURF). Comput. Vis. Image Underst. **110** (2008) 346–359
16. Tola, E., Lepetit, V., Fua, P.: DAISY: An efficient dense descriptor applied to wide-baseline stereo. IEEE Trans. Pattern Anal. Mach. Intell. **32**(5) (2010) 815–830
17. Tuzel, O., Porikli, F., Meer, P.: Human detection via classification on Riemannian manifolds. In: Proc. Int. Conf. Comp. Vis. Patt. Recog. (2007) 1–8
18. Porikli, F., Tuzel, O., Meer, P.: Covariance tracking using model update based on lie algebra. In: Proc. Int. Conf. Comp. Vis. Patt. Recog. (2006) 728–735
19. Li, X., Hu, W., Zhang, Z., Zhang, X., Zhu, M., Cheng, J.: Visual tracking via incremental Log-Euclidean Riemannian subspace learning. In: Proc. Int. Conf. Comp. Vis. Patt. Recog. (2008) 1–8
20. Gong, L., Wang, T., Liu, F.: Shape of gaussians as feature descriptors. In: Proc. Int. Conf. Comp. Vis. Patt. Recog. (2009) 2366–2371
21. Nakayama, H., Harada, T., Kuniyoshi, Y.: Global gaussian approach for scene categorization using information geometry. In: Proc. Int. Conf. Comp. Vis. Patt. Recog. (2010) 2336–2343
22. Sivalingam, R., Boley, D., Morellas, V., Papanikolopoulos, N.: Tensor sparse coding for region covariances. In: Europ. Conf. on Comp. Vis. (2010) 722–735
23. Fletcher, P.T., Joshi, S.: Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In: ECCV Workshops CVAMIA and MMBIA. (2004) 87–98
24. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based object tracking. IEEE Trans. Pattern Anal. Mach. Intell. **25**(5) (2003) 564–575
25. Wang, S., Lu, H., Yang, F., Yang, M.H.: Superpixel tracking. In: Int. Conf. on Comp. Vis. (2011) 1323–1330
26. de Luis García, R., Westin, C.F., Alberola-López, C.: Gaussian mixtures on tensor fields for segmentation: Applications to medical imaging. Comp. Med. Imag. and Graph. **35**(1) (2011) 16–30
27. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol. **2** (2011) 27:1–27:27
28. Hayman, E., Caputo, B., Fritz, M., Eklundh, J.: On the significance of real-world conditions for material classification. In: Europ. Conf. on Comp. Vis. (2004) 253–266
29. Zhang, J., Marszalek, M., Lzaebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: A comprehensive study. Int. J. Comput. Vision **73** (2007) 213–238
30. Lazebnik, S., Schmid, C., Ponce, J.: A sparse texture representation using local affine regions. IEEE Trans. Pattern Anal. Mach. Intell. **27** (2005) 1265–1278
31. Varma, M., Zisserman, A.: Texture classification: Are filter banks necessary? In: Proc. Int. Conf. Comp. Vis. Patt. Recog. (2003)
32. Jalba, A., Roerdink, J., Wilkinson, M.: Morphological hat-transform scale spaces and their use in pattern classification. Pattern Recognition **37** (2004) 901–915
33. Manjunath, B., Ma, W.: Texture features for browsing and retrieval of image data. IEEE Trans. Pattern Anal. Mach. Intell. (1996) 837–842
34. Ross, D.A., Lim, J., Yang, M.H.: Adaptive probabilistic visual tracking with incremental subspace update. In: Europ. Conf. on Comp. Vis. (2004) 470–482
35. Leichter, I., Lindenbaum, M., Rivlin, E.: Tracking by affine kernel transformations using color and boundary cues. IEEE Trans. on Pattern Anal. Mach. Intell. **31**(1) (2009) 164–171