Efficient texture retrieval using multiscale local extrema descriptors and covariance embedding

Minh-Tan Pham[0000−0003−0266−767X]

IRISA - University of Southern Brittany, Vannes 56000, France
minh-tan.pham@irisa.fr

Abstract. We present an efficient method for texture retrieval using multiscale feature extraction and embedding based on the local extrema keypoints. The idea is to first represent each texture image by its local maximum and local minimum pixels. The image is then divided into regular overlapping blocks and each one is characterized by a feature vector constructed from the radiometric, geometric and structural information of its local extrema. All feature vectors are finally embedded into a covariance matrix which will be exploited for dissimilarity measurement within retrieval task. Thanks to the method's simplicity, multiscale scheme can be easily implemented to improve its scale-space representation capacity. We argue that our handcrafted features are easy to implement, fast to run but can provide very competitive performance compared to handcrafted and CNN-based learned descriptors from the literature. In particular, the proposed framework provides highly competitive retrieval rate for several texture databases including 94.95% for MIT Vistex, 79.87% for Stex, 76.15% for Outex TC-00013 and 89.74% for USPtex.

Keywords: Texture retrieval · Handcrafted features · Local extrema · Feature covariance matrix

1 Introduction

Content-based image retrieval (CBIR) has been always drawing attention from researchers working on image analysis and pattern recognition within computer vision field. Texture, i.e. a powerful image feature involving repeated patterns which can be recognized by human vision, plays a significant role in most of CBIR systems. Constructing efficient texture descriptors to characterize the image becomes one of the key components which have been focused in most research works related to texture image retrieval [41, 35, 4].

From the literature, a great number of multiscale texture analysis methods using probabilistic approach have been developed to tackle retrieval task. In [8], the authors proposed to model the spatial dependence of pyramidal discrete wavelet transform (DWT) coefficients using the generalized Gaussian distributions (GGD) and the dissimilarity measure between images was derived based on the Kullback-Leibler divergences (KLD) between GGD models. Sharing the similar principle, multiscale coefficients yielded by the discrete cosine transform

(DCT), the dual-tree complex wavelet transform (DT-CWT), the Gabor Wavelet (GW), etc. were modeled by different statistical models such as GGD, the multivariate Gaussian mixture models (MGMM), Gaussian copula (GC), Student-t copula (StC), or other distributions like Gamma, Rayleigh, Weibull, Laplace, etc. to perform texture-based image retrieval [18, 43, 5, 17, 19, 21, 47]. However, one of the main drawbacks of these techniques is the their expensive computational time which has been observed an discussed in several papers [17, 19, 21].

Other systems which have provided effective CBIR performance include the local pattern-based framework and the block truncation coding (BTC)-based approach. The local binary patterns (LBP) were first embedded in a multiresolution and rotation invariant scheme for texture classification in [25]. Inspired from this work, many studies have been developed for texture retrieval such as the local maximum edge binary patterns (LMEBP) [39], local ternary patterns (LTP) [40], local tetra patterns (LTrP) [22], local tri-directional patterns (LTriDP) [44], local neighborhood difference pattern (LNDP) [45], etc. These descriptors, in particular LTrP and LTRiDP, can provide good retrieval rate. However, due to the fact that they work on grayscale images, their performance on natural textures is limited without using color information. To overcome this issue, recent schemes have proposed to incorporate these local patterns with color features. Some techniques can be mentioned here are the joint histogram of color and local extrema patterns (LEP+colorhist) [23], the local oppugnant color texture pattern (LOCTP) [14], the local extrema co-occurrence pattern (LECoP) [46], LBPC for color images [38]. Beside that, many studies have also developed different BTC-based frameworks, e.g. the ordered-dither BTC (ODBTC) [12, 10], the error diffusion BTC (EDBTC) [11] and the dot-diffused BTC (DDBTC) [13], which have provided competitive retrieval performance. Within these approaches, an image is divided into multiple non-overlapping blocks and each one is compressed into the so-called color quantizer and bitmap image. Then, a feature descriptor is constructed using the color histogram and color co-occurrence features combined with the bit pattern feature (including edge, shape, texture information) of the image. These features are extracted from the above color quantizer and bitmap image to tackle CBIR task.

Last but not least, not focusing on developing handcrafted descriptors as all above systems, learned descriptors extracted from convolution neural networks (CNNs) have been recently applied to image retrieval task [36, 42]. An end-to-end CNN framework can learn and extract multilevel discriminative image features which are extremely effective for various computer vision tasks including recognition and retrieval [36]. In practice, instead of defining and training their own CNNs from scratch, people tend to exploit pre-trained CNNs (on a very large dataset such as ImageNet [7]) as feature extractors. Recent studies have shown the effective performance of CNN learned features w.r.t. traditional handcrafted descriptors applied to image classification and retrieval [6, 24].

In this work, we continue the traditional approach of handcrafted feature designing by introducing a powerful retrieval framework using multiscale local extrema descriptors and covariance embedding. Here, we inherit the idea of using

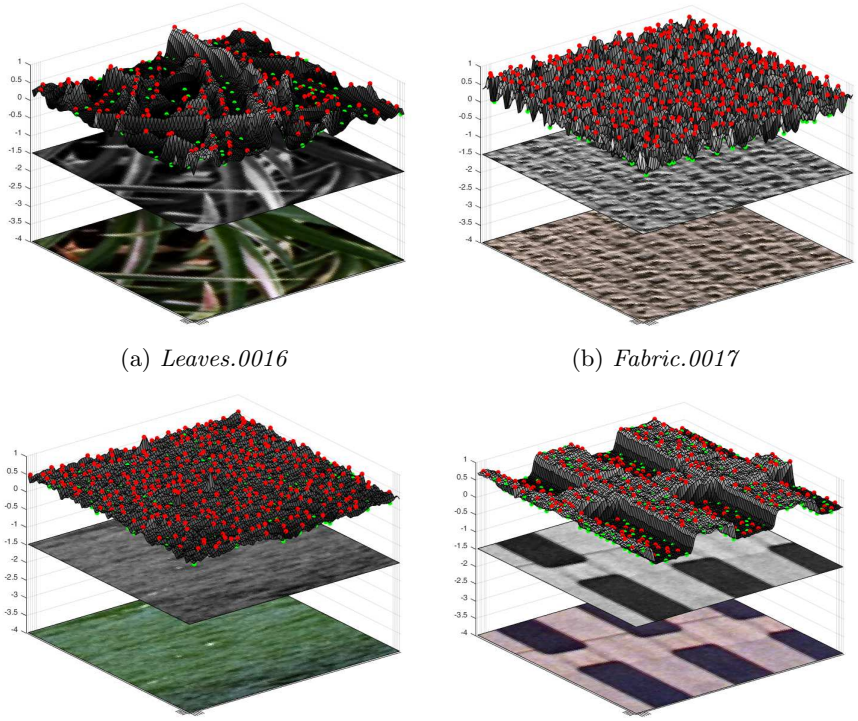
local extrema pixels for texture description and retrieval from [28] but provide a simpler and faster feature extraction algorithm which can be easily integrated into a multiscale scheme. Due to the fact that natural images usually involve a variety of local textures and structures which do not appear homogeneous within the entire image, an approach taking into account local features becomes relevant. Also, a multiscale approach could help to provide a better scale-space representation capacity to deal with complex textures. Within our approach, a set of local maximum and local minimum pixels (in terms of intensity) is first detected to represent each texture image. Then, to extract local descriptors, the image is divided into regular overlapping blocks of equal size and each block is characterized by a feature vector constructed using the radiometric (i.e. color), geometric ans structural information of its local extrema. The descriptor of each block is named SLED, i.e. simple local extrema descriptor. Thus, the input image is encoded by a set of SLED vectors which are then embedded into a feature covariance matrix. Moreover, thanks to the simplicity of the approach, we propose to upsample and downsample each image to perform the algorithm at different scales. Finally, we exploit the geometric-based riemannian distance [9] between covariance matrices for dissimilarity measurement within retrieval task. Our experiments show that the proposed framework can provide highly competitive performance for several popular texture databases compared against both state-of-the-art handcrafted and learned descriptors. In the rest of this paper, Section 2 describes the proposed retrieval framework including the details of SLED feature extraction, covariance embedding and multiscale scheme. We then present our experiments conducted on four popular texture databases in Section 3 and Section 4 finally concludes the paper with some potential developments.

2 Proposed texture retrieval framework

2.1 Texture representation using local extrema pixels

The idea of using the local extrema (i.e. local max and local min pixels) for texture analysis was introduced in [26, 30, 27] for texture segmentation in very high resolution images and also exploited in [28, 34, 29] for texture image retrieval. Regarding to this point of view, a texture is formed by a certain spatial distribution of pixels holding some illumination (i.e. intensity) variations. Hence, different textures are reflected by different types of pixel's spatial arrangements and radiometric variations. These meaningful properties can be approximately captured by the local extrema detected from the image. Hence, these local keypoints are relevant for texture representation and description [31–33].

The detection of local extrema from a grayscale image is quite simple and fast. Using a sliding window, the center pixel (at each sliding position) is supposed to be a local maximum (resp. local minimum) if it has the highest (resp. lowest) intensity value. Hence, by only fixing a $w \times w$ window size, the local extrema are detected by scanning the image only once. To deal with color images, there are different ways to detect local extrema (i.e. detecting from the grayscale version, using the union or intersection of extrema subsets detected from each color



(a) *Leaves.0016* (b) *Fabric.0017*

(c) *Water.0005* (d) *Buildings.0009*

Fig. 1. Illustration: spatial distribution and arrangement of local max pixels (red) and local min pixels (green) within 4 different textures from the MIT Vistex database [3]. These local extrema are detected using a 5×5 search window.

channel, etc.). For simplicity, we propose to detect local extrema from grayscale version of color images in this paper. To illustrate the capacity of the local extrema of representing and characterizing different texture contents, Fig. 1 shows their spatial apperance within 4 different textures of the MIT Vistex database [3]. Each 128×128 color texture (at the bottom) is first converted to a grayscale image (in the middle). On the top, we display for each one a 3-D surface model using the grayscale image intensity as the surface height. The local maxima (in red) and local minima (in green) are detected using a 5×5 sliding window. Some green points may be unseen since they are obscured by the surface. We observe that these extrema contain rich information that represent each texture content. Therefore, extracting and encoding their radiometric (or color) and geometric features could provide a promising texture description tool.

2.2 Simple local extrema descriptor (SLED)

Given an input texture image, after detecting the local extrema using a $w \times w$ sliding window, the next step is to divide the image into N overlapping blocks of size $W \times W$ and then extract the simple local extrema descriptor (SLED) feature vector from each block. The generation of SLED vector is summarized in Fig. 2. From each image block $B_i, i = 1 \dots N$, we first separate the local maxima set $S_i^{\max}$ and the local minima set $S_i^{\min}$, and then extract the color, spatial and gradient features of local keypoints to form their description vectors.

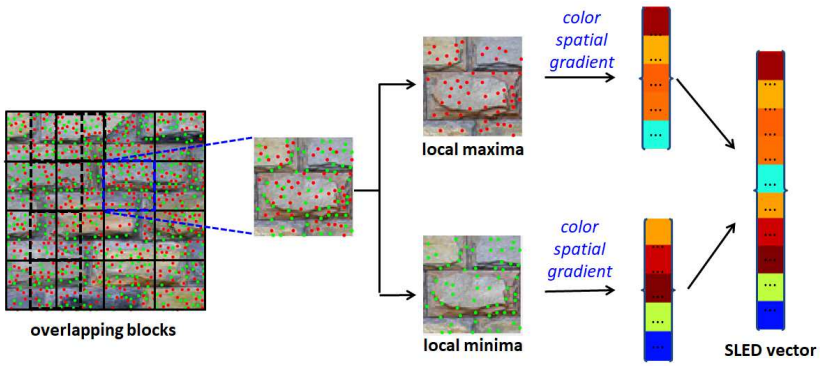


Fig. 2. Generation of SLED feature vector for each image block.

In details, below are the features extracted from $S_i^{\max}$, the feature generation for $S_i^{\min}$ is similar.

+ Mean and variance of each color channel:

$$\mu_{\text{i,color}}^{\max} = \frac{1}{|S_i^{\max}|} \sum_{(x,y) \in S_i^{\max}} I_{\text{color}}(x,y), \tag{1}$$

$$\sigma_{\text{i,color}}^{2\,\max} = \frac{1}{|S_i^{\max}|} \sum_{(x,y) \in S_i^{\max}} (I_{\text{color}}(x,y) - \mu_{\text{i,color}}^{\max})^2, \tag{2}$$

  where color $\in \{\text{red}, \text{green}, \text{blue}\}$ represents each of the 3 color components; (x, y) is the pixel position on the image grid and $|S_i^{\max}|$ is the cardinality of the set $S_i^{\max}$.
+ Mean and variance of spatial distances from each local maximum to the center of B_i:

$$\mu_{\text{i,spatial}}^{\max} = \frac{1}{|S_i^{\max}|} \sum_{(x,y) \in S_i^{\max}} d_i(x,y), \tag{3}$$

$$\sigma_{\text{i,spatial}}^{2\,\max} = \frac{1}{|S_i^{\max}|} \sum_{(x,y) \in S_i^{\max}} (d_i(x,y) - \mu_{\text{i,spatial}}^{\max})^2, \tag{4}$$

where $d_i(x, y)$ is the spatial distance from the pixel (x, y) to the center of block B_i on the image plane.

\+ Mean and variance of gradient magnitudes:

$$\mu_{\text{i,grad}}^{\max} = \frac{1}{|S_i^{\max}|} \sum_{(x,y)\in S_i^{\max}} \nabla I(x, y), \tag{5}$$

$$\sigma_{\text{i,grad}}^{2\,\max} = \frac{1}{|S_i^{\max}|} \sum_{(x,y)\in S_i^{\max}} (\nabla I(x, y) - \mu_{\text{i,grad}}^{\max})^2, \tag{6}$$

where ∇I is the gradient magnitude image obtained by applying the Sobel filter on the gray-scale version of the image.

All of these features are integrated into the feature vector $f_i^{\max} \in \mathbb{R}^{10}$, which encodes the color (i.e. three channels), spatial and structural features of the local maxima inside the block B_i:

$$f_i^{\max} = \left[\mu_{\text{i,color}}^{\max}, \sigma_{\text{i,color}}^{2\,\max}, \mu_{\text{i,spatial}}^{\max}, \sigma_{\text{i,spatial}}^{2\,\max}, \mu_{\text{i,grad}}^{\max}, \sigma_{\text{i,grad}}^{2\,\max}\right] \in \mathbb{R}^{10}. \tag{7}$$

The generation of $f_i^{\min}$ from the local min set $S_i^{\min}$ is similar. Now, let f_i^{SLED} be the SLED feature vector generated for block B_i, we finally define:

$$f_i^{\text{SLED}} = \left[f_i^{\max}, f_i^{\min}\right] \in \mathbb{R}^{20}. \tag{8}$$

The proposed feature vector f_i^{SLED} enables us to characterize the local textures of each image block B_i by understanding how local maxima and local minima are distributed and arranged, and how they capture color information as well as structural properties (given by gradient features). The extraction of our handcrafted SLED is quite simple and fast. We observe that it is also feasible to add other features to create more complex feature vector as proposed in [28]. However, we argue that by using covariance embedding and performing multiscale framework (described in the next section), the simple and fast SLED already provides very competitive retrieval performance.

2.3 Covariance embedding and multiscale framework

The previous section has described the generation of SLED vector for each block of the image. Once all feature vectors are extracted to characterize all image blocks, they are embedded into a covariance matrix as shown in Fig. 3. Given a set of N SLED feature vectors $f_i^{\text{SLED}}, i = 1 \dots N$, the embedded covariance matrix is estimated as follow:

$$C^{\text{SLED}} = \frac{1}{N} \sum_{i=1}^{N} (f_i^{\text{SLED}} - \mu^{\text{SLED}})(f_i^{\text{SLED}} - \mu^{\text{SLED}})^T, \tag{9}$$

where $\mu^{\text{SLED}} = \frac{1}{N} \sum_{i=1}^{N} f_i^{\text{SLED}}$ is the estimated mean feature vector.

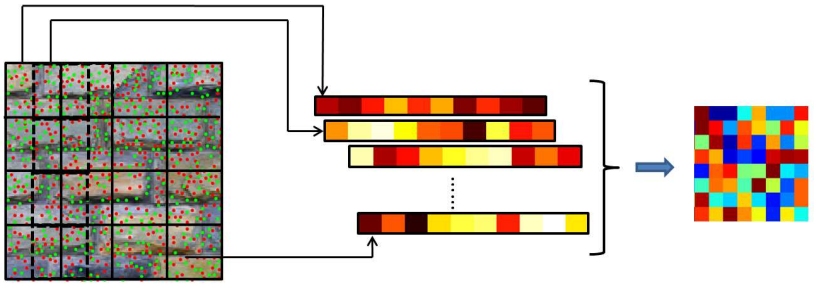


Fig. 3. Proposed method to extract SLED feature vectors from all image overlapping blocks and embed them into a covariance matrix.

Last but not least, thanks to the simplicity of the proposed SLED extraction and embedding strategy, we also propose a multi-scale framework as in Fig. 4. Here, each input image will be upsampled and downsampled with the scale factor of 3/2 and 2/3, respectively, using the bicubic interpolation approach. Then, the proposed scheme in Fig. 3 is applied to these two rescaled images and the original one to generate three covariance matrices. It should be noted that the number of rescaled images as well as scaling factors can be chosen differently. Here, without loss of genarality, we fix the number of scales to 3 and scale factors to 2/3, 1 and 3/2 for all implementations in the paper. To this end, due to the fact that covariance matrices possess a postitive semi-definite structure and do not lie on the Euclidean space, we finally exploit the geometric-based riemannian distance for dissimilarity measurement within retrieval task. This metric has been proved to be relevant and effective for covariance descriptors in the literature [9].

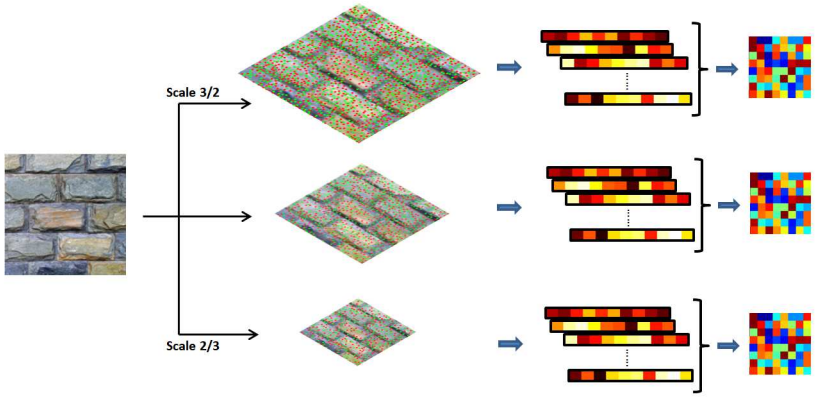
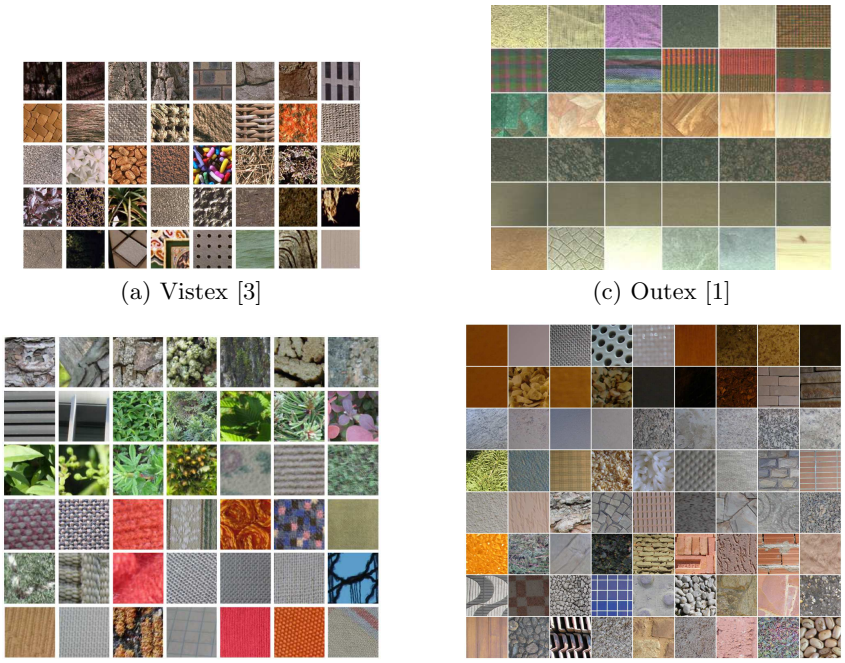


Fig. 4. Proposed multi-scale framework.

3 Experimental study

3.1 Texture databases

Four popular texture databases including the MIT Vistex [3], the Salzburg Texture (Stex) [16], the Outex TC-00013 [1] and the USPtex [2] were exploited to conduct our experiments. Vistex is one of the most widely used texture databases for performance evalution and comparative study in the CBIR field. It consists of 640 texture images (i.e. 40 classes $\times$ 16 images per class) of size 128×128 pixels. Being much larger, the Stex database is a collection of 476 texture classes captured in the area around Salzburg, Austria under real-word conditions. As for Vistex, each class includes 16 images of 128×128 pixels, hence the total number of images from the database is 7616. The third dataset, the Outex TC-00013 [1], is a collection of heterogeneous materials such as paper, fabric, wool, stone, etc. It comprises 68 texture classes and each one includes 20 image samples of 128×128 pixels. Finally, the USPtex database [2] includes 191 classes of both natural scene (road, vegetation, cloud, etc.) and materials (seeds, rice, tissues, etc.). Each class consists of 12 image samples of 128×128 pixels. Fig. 5 shows some examples of each texture database and Table 1 provides a summary of their information.



(a) Vistex [3] (c) Outex [1]

(b) Stex [16] (d) USPtex [2]

Fig. 5. Four image databases used in the experimental study.

Table 1. Number of total images (N_t), number of classes (N_c) and number of relevant images (N_R) per class within the 4 experimental texture databases.

	Vistex	**Stex**	**Outex**	**USPtex**
N_t	640	7616	1380	2292
N_c	40	476	68	191
N_R	16	16	20	12

3.2 Experimental setup

To perform our retrieval framework, the local extrema keypoints were detected using a 3×3 search window ($w = 3$). We recommend this small window size to ensure a dense distribution of local extrema for different texture scenes. Next, each texture image is divided into overlapping blocks of size 32×32 pixels and two consecutive blocks are 50% overlapped. Thus, for each 128×128 image, the number of blocks is $N = 64$. For multiscale framework as in Fig.4, we set 3 scales of 2/3, 1 and 3/2 as previously mentioned. There are no other parameters to be set, which confirms the simplicity of our proposed method.

For comparative evaluation, we compare our results to several state-of-the-art methods in the literature including:

+ probabilistic approaches in [8, 18, 43, 5, 17, 19],[20];
+ handcrafted local pattern-based descriptors such as LMEBP [39], LtrP [22], LEP+colorhist [23], LECoP [46], ODII [12];
+ handcrafted BTC-based frameworks including DDBTC [13], ODBTC [10] and EDBTC [11];
+ learned descriptors based on pre-trained CNNs [6, 24]. For these, we exploited the AlexNet [15], VGG-16 and VGG-19 [37] pre-trained on ImageNet database [7] as feature extractors. We used the 4096-D feature vector from the FC7 layer (also followed by a ReLu layer) and the L1 distance for dissimilarity measure as recommended in [24].
+ the LED framework proposed [28] by setting equivalent parameters to our algorithm. In details, we set the 3 window sizes for keypoint extraction (ω_1), local extrema detection (ω_2) and LED generation (W) to 9×9, 3×3 and 36×36, respectively.

For evaluation criteria, the average retrieval rate (ARR) is adopted. Let N_t, N_R be the total number of images in the database and the number of relevant images for each query, and for each query image q, let $n_q(K)$ be the number of correctly retrieved images among the K retrieved ones (i.e. K best matches). ARR in terms of number of retrieved images (K) is given by:

$$\mathrm{ARR}(K) = \frac{1}{N_t \times N_R} \sum_{q=1}^{N_t} n_q(K) \Big|_{K \geq N_R} \tag{10}$$

We note that K is generally set to be greater than or equal to N_R. By setting K equal to N_R, ARR becomes the primary benchmark considered by

most studies to evaluate and compare the performance of different CBIR systems. All of ARR results shown in this paper were produced by setting $K = N_R$.

3.3 Results and discussion

Tables 2 and 3 show the ARR performance of the proposed SLED and mutilscale SLED (MS-SLED) on our four texutre databases compared to reference methods. The first observation is that, most local feature-based CBIR schemes (e.g. LtrP [22], LEP+colorhist [23], LECoP [46]) or BTC-based techniques [10, 11, 13] have achieved better retrieval performance than probabilistic methods which model the entire image using different statistical distributions [8, 18, 43, 5, 17, 19]. Also, learned descriptors based on pre-trained CNNs have yielded very competitive retrieval rate compared to handcrafted features which prove the potential of CNN feature extractor applied to retrieval task [6, 24]. Then, more importantly, our proposed SLED and MS-SLED frameworks have outperformed all reference methods for all datasets. We now discuss the results of each database to validate the effectiveness of the proposed strategy.

Table 2. Average retrieval rate (%) on the **Vistex** and **Stex** databases yielded by the proposed method compared to reference methods.

Method	**Vistex**	**Stex**
GT+GGD+KLD [8]	76.57	49.30
MGG+Gaussian+KLD [43]	87.40	-
MGG+Laplace+GD [43]	91.70	71.30
Gaussian Copula+Gamma+ML [17]	89.10	69.40
Gaussian Copula+Weibull+ML [17]	89.50	70.60
Student-t Copula+GG+ML [17]	88.90	65.60
Gaussian Copula+Gabor Wavelet [21]	92.40	76.40
LMEBP [39]	87.77	-
LtrP [22]	90.02	-
LEP+colorhist [23]	82.65	59.90
DDBTC [13]	92.65	44.79
ODBTC [10]	90.67	-
EDBTC [11]	92.55	-
LECoP [46]	92.99	74.15
ODII [12]	93.23	-
CNN-AlexNet [24]	91.34	68.84
CNN-VGG16 [24]	92.97	74.92
CNN-VGG19 [24]	93.04	73.93
LED [28]	94.13	76.71
Proposed SLED	**94.31**	**77.78**
Proposed MS-SLED	**94.95**	**79.87**

The best ARR of 94.95% and 79.87% was produced for Vistex and Stex by our MS-SLED algorithm. A gain of 0.82% and 3.16% was achieved compared to

Table 3. Average retrieval rate (%) on the **Outex** and **USPtex** databases yielded by the proposed method compared to reference methods.

Method	**Outex**	**UPStex**
DDBTC (L_1) [13]	61.97	63.19
DDBTC (L_2) [13]	57.51	55.38
DDBTC (χ^2) [13]	65.54	73.41
DDBTC (Canberra) [13]	66.82	74.97
CNN-AlexNet [24]	69.87	83.57
CNN-VGG16 [24]	72.91	85.03
CNN-VGG19 [24]	73.20	84.22
LED [28]	75.14	87.50
Proposed SLED	**75.96**	**88.60**
Proposed MS-SLED	**76.15**	**89.74**

the second-best method with original LED features in [28]. Within the proposed strategy, the multi-scale scheme has considerable improved the ARR from the single-scale SLED (i.e. 0.62% for Vistex and 2.09% for Stex), which confirms the efficiency of performing multiscale feature extraction and embedding for better texture description, as our motivation in this work. Next, another important remark is that most of the texture classes with strong structures and local features such as buildings, fabric categories, man-made object's surfaces, etc. were perfectly retrieved. Table 4 shows the per-class retrieval rate for each class of the Vistex database. As observed, half of the classes (20/40 classes) were retrieved with 100% accuracy (marked in bold). These classes generally consist of many local textures and structures. Similar behavior was also remarked for Stex data. This issue is encouraging since our motivation is to continue developing hand-designed descriptors which represent and characterize local features better than both handcrafted and learned descriptors from the literature.

Similarly, the proposed MS-SLED framework also provided the best ARR for both Outex (76.15%) and USPtex data (89.74%) (with a gain of 1.01% and 2.24%, respectively), as observed in Table 3. Compared to learned descriptors based on pretrained AlexNet, VGG-16 and VGG-19, an improvement of 2.95% and 4.71% was adopted, which confirms the superior performance of our method over the CNN-based counterparts. To this end, the efficiency of the proposed framework is validated for all tested databases.

Last but not least, Table 5 provides the comparison of descriptor dimensions within different methods. We note that our SLED involves a 20×20 covariance matrix estimated as (9). Since the matrix is symmetrical, it is only necessary to store its upper or lower triangular entries. Thus, the SLED feature dimension is calculated as $20 \times (20+1) = 210$. For MS-SLED, we multiply this to the number of scales and hence the length becomes 630 in our implementation. Other feature lengths from the table are illustrated from their related papers. We observe that the proposed SLED has lower dimension than the standard LED in [28] (i.e. 210 compared to 561) but can provide faster and better retrieval performance. To support this remark, we show in Table 6 a comparison of computational time for

Table 4. Per-class retrieval rate (%) on the **Vistex-640** database using the proposed LED+RD method

Class	Rate	Class	Rate	Class	Rate
Bark.0000	75.00	Fabric.0015	**100.00**	Metal.0002	**100.00**
Bark.0006	94.14	Fabric.0017	96.88	Misc.0002	**100.00**
Bark.0008	84.38	Fabric.0018	**100.00**	Sand.0000	**100.00**
Bark.0009	77.73	Flowers.0005	**100.00**	Stone.0001	85.55
Brick.0001	99.61	Food.0000	**100.00**	Stone.0004	93.75
Brick.0004	97.27	Food.0005	99.61	Terrain.0010	94.14
Brick.0005	**100.00**	Food.0008	**100.00**	Tile.0001	90.23
Buildings.0009	**100.00**	Grass.0001	94.53	Tile.0004	**100.00**
Fabric.0000	**100.00**	Leaves.0008	**100.00**	Tile.0007	**100.00**
Fabric.0004	78.13	Leaves.0010	**100.00**	Water.0005	**100.00**
Fabric.0007	99.61	Leaves.0011	**100.00**	Wood.0001	98.44
Fabric.0009	**100.00**	Leaves.0012	60.93	Wood.0002	88.67
Fabric.0011	**100.00**	Leaves.0016	90.23	**ARR**	**94.95**
Fabric.0014	**100.00**	Metal.0000	99.21		

Table 5. Comparison of feature vector length of different methods.

Method	Feature dimension
DT-CWT [8]	$(3 \times 6 + 2) \times 2$=40
DT-CWT+DT-RCWT [8]	$2 \times (3 \times 6 + 2) \times 2$=80
LBP [25]	256
LTP [40]	$2 \times 256 = 512$
LMEBP [39]	$8 \times 512 = 4096$
Gabor LMEBP [39]	$3 \times 4 \times 512 = 6144$
LEP+colorhist [23]	$16 \times 8 \times 8 \times 8 = 8192$
LECoP($H_{18}S_{10}V_{256}$) [46]	$18 + 10 + 256 = 284$
LECoP($H_{36}S_{20}V_{256}$) [46]	$36 + 20 + 256 = 312$
LECoP($H_{72}S_{20}V_{256}$) [46]	$72 + 20 + 256 = 348$
ODII [12]	128+128 = 256
CNN-AlexNet [24]	4096
CNN-VGG16 [24]	4096
CNN-VGG19 [24]	4096
LED [28]	$33 \times (33 + 1)/2 = 561$
Proposed SLED	$\mathbf{20 \times (20 + 1)/2 = 210}$
Proposed MS-SLED	$\mathbf{3 \times 210 = 630}$

feature extraction and dissimilarity measurement of LED and SLED. In short, a total amount of 95.17 seconds is required by our SLED to run on the Vistex data, thus 0.148 second per image, which is very fast. All implementations were carried out using MATLAB 2017a on computer of 3.5GHz/16GB RAM.

Table 6. Computation time (in second) of LED and SLED feature extraction (FE) and dissimilarity measurement (DM). Experiments were conducted on the Vistex database.

Version	**FE time**		**DM time** (s)		**Total time**	
	t_{data}	t_{image}	t_{data}	t_{image}	t_{data}	t_{image}
LED [28]	193.77	0.308	21.39	0.033	215.16	0.336
SLED (ours)	86.35	0.135	8.82	0.013	95.17	0.148

t_{data}: time for the total database ; t_{image}: time per each image.

4 Conclusions

We have proposed a simple and fast texture image retrieval framework using multiscale local extrema feature extraction and covariance embedding. Without chasing the current trends of deep learning era, we continue the classical way of designing novel handcrafted features in order to achieve highly competitive retrieval performance compared to state-of-the-art methodologies. The detection of local extrema as well as the extraction of their color, spatial and gradient features are quite simple but they are effective for texture description and encoding. We argue that the proposed MS-SLED does not require many parameters for tuning. It is easy to implement, fast to run and feasible to extend or improve. The best retrieval rates obtained for four texture benchmarks shown in our experimental study have confirmed the effectiveness of the proposed strategy. Future work can improve the performance of MS-SLED by exploiting other textural features within its construction. Also, we are now interested in integrating SLED features into an auto-encoder framework in order to automatically learn and encode richer information for better texture representation.

References

1. Outex texture database. University of Oulu, Available online.
2. USPtex dataset (2012). Scientific Computing Group, Available online: http://fractal.ifsc.usp.br/dataset/USPtex.php.
3. Vision texture. MIT Vision and Modeling group, Available online: http://vismod.media.mit.edu/pub/VisTex/.
4. Alzubi, A., Amira, A., Ramzan, N.: Semantic content-based image retrieval: A comprehensive study. J. Visual Communi. Image Represent. **32**, 20–54 (2015)
5. Choy, S.K., Tong, C.S.: Statistical wavelet subband characterization based on generalized gamma density and its application in texture retrieval. IEEE Trans. Image Process. **19**(2), 281–289 (2010). https://doi.org/10.1109/TIP.2009.2033400
6. Cusano, C., Napoletano, P., Schettini, R.: Evaluating color texture descriptors under large variations of controlled lighting conditions. JOSA A **33**(1), 17–30 (2016)
7. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: CVPR. pp. 248–255. Ieee (2009)
8. Do, M.N., Vetterli, M.: Wavelet-based texture retrieval using generalized gaussian density and Kullback-Leibler distance. IEEE Trans. Image Process. **11**(2), 146–158 (2002). https://doi.org/10.1109/83.982822

9. Förstner, W., Moonen, B.: A metric for covariance matrices. In: Geodesy-The Challenge of the 3rd Millennium, pp. 299–309. Springer (2003). https://doi.org/10.1007/978-3-662-05296-9-31
10. Guo, J.M., Prasetyo, H.: Content-based image retrieval using features extracted from halftoning-based block truncation coding. IEEE Trans. Image Process. **24**(3), 1010–1024 (2015). https://doi.org/10.1109/TIP.2014.2372619
11. Guo, J.M., Prasetyo, H., Chen, J.H.: Content-based image retrieval using error diffusion block truncation coding features. IEEE Trans. Circuits Syst. Video Technol **25**(3), 466–481 (2015). https://doi.org/10.1109/TCSVT.2014.2358011
12. Guo, J.M., Prasetyo, H., Su, H.S.: Image indexing using the color and bit pattern feature fusion. J. Vis. Commun. Image Repres. **24**(8), 1360–1379 (2013). https://doi.org/10.1016/j.jvcir.2013.09.005
13. Guo, J.M., Prasetyo, H., Wang, N.J.: Effective image retrieval system using dot-diffused block truncation coding features. IEEE Trans. Multimedia **17**(9), 1576–1590 (2015). https://doi.org/10.1109/TMM.2015.2449234
14. Jacob, I.J., Srinivasagan, K., Jayapriya, K.: Local oppugnant color texture pattern for image retrieval system. Pattern Recogn. Letters **42**, 72–78 (2014). https://doi.org/10.1016/j.patrec.2014.01.017
15. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS. pp. 1097–1105 (2012)
16. Kwitt, R., Meerwald, P.: Salzburg texture image database. Available online: http://www.wavelab.at/sources/STex/.
17. Kwitt, R., Meerwald, P., Uhl, A.: Efficient texture image retrieval using copulas in a Bayesian framework. IEEE Trans. Image Process. **20**(7), 2063–2077 (2011). https://doi.org/10.1109/TIP.2011.2108663
18. Kwitt, R., Uhl, A.: Image similarity measurement by Kullback-Leibler divergences between complex wavelet subband statistics for texture retrieval. In: Proc. IEEE Int. Conf. Image Process. (ICIP). pp. 933–936 (2008). https://doi.org/10.1109/ICIP.2008.4711909
19. Lasmar, N.E., Berthoumieu, Y.: Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms. IEEE Trans. Image Process. **23**(5), 2246–2261 (2014). https://doi.org/10.1109/TIP.2014.2313232
20. Li, C., Duan, G., Zhong, F.: Rotation invariant texture retrieval considering the scale dependence of Gabor wavelet. IEEE Trans. Image Process. **24**(8), 2344–2354 (2015). https://doi.org/10.1109/TIP.2015.2422575
21. Li, C., Huang, Y., Zhu, L.: Color texture image retrieval based on gaussian copula models of gabor wavelets. Pattern Recognition **64**, 118–129 (2017). https://doi.org/10.1016/j.patcog.2016.10.030
22. Murala, S., Maheshwari, R., Balasubramanian, R.: Local tetra patterns: a new feature descriptor for content-based image retrieval. IEEE Trans. Image Process. **21**(5), 2874–2886 (2012). https://doi.org/10.1109/TIP.2012.2188809
23. Murala, S., Wu, Q.J., Balasubramanian, R., Maheshwari, R.: Joint histogram between color and local extrema patterns for object tracking. In: IS&T/SPIE Electronic Imaging. pp. 86630T–86630T–7. Int. Soc. Optics Photonics (2013). https://doi.org/10.1117/12.2002185
24. Napoletano, P.: Hand-crafted vs learned descriptors for color texture classification. In: Int. Wksh. Comput. Color Imaging. pp. 259–271. Springer (2017)
25. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Patt. Anal. Mach. Intell. **24**(7), 971–987 (2002). https://doi.org/10.1109/TPAMI.2002.1017623

26. Pham, M.T., Mercier, G., Michel, J.: Pointwise graph-based local texture characterization for very high resolution multispectral image classification. IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens. **8**(5), 1962–1973 (2015)
27. Pham, M.T.: Pointwise approach for texture analysis and characterization from very high resolution remote sensing images. Ph.D. thesis, Télécom Bretagne (2016)
28. Pham, M.T., Mercier, G., Bombrun, L.: Color texture image retrieval based on local extrema features and riemannian distance. J. Imaging **3**(4), 43 (2017)
29. Pham, M.T., Mercier, G., Bombrun, L., Michel, J.: Texture and color-based image retrieval using the local extrema features and riemannian distance. arXiv preprint arXiv:1611.02102 (2016)
30. Pham, M.T., Mercier, G., Michel, J.: Textural features from wavelets on graphs for very high resolution panchromatic pléiades image classification. French Journal Photogram. Remote Sens. **208**, 131–136 (2014)
31. Pham, M.T., Mercier, G., Michel, J.: Change detection between SAR images using a pointwise approach and graph theory. IEEE Trans. Geosci. Remote Sens. **54**(4), 2020–2032 (2016)
32. Pham, M.T., Mercier, G., Michel, J.: PW-COG: an effective texture descriptor for VHR satellite imagery using a pointwise approach on covariance matrix of oriented gradients. IEEE Trans. Geosci. Remote Sens. **54**(6), 3345–3359 (2016)
33. Pham, M.T., Mercier, G., Regniers, O., Michel, J.: Texture retrieval from VHR optical remote sensed images using the local extrema descriptor with application to vineyard parcel detection. Remote Sensing **8**(5), 368 (2016)
34. Pham, M.T., Mercier, G., Regniers, O., Bombrun, L., Michel, J.: Texture retrieval from very high resolution remote sensing images using local extrema-based descriptors. In: Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS). pp. 1839–1842. IEEE (2016)
35. Raghuwanshi, G., Tyagi, V.: A survey on texture image retrieval. In: Proc. Second Int. Conf. Comput. Communi. Technol. pp. 427–435. Springer (2016)
36. Schmidhuber, J.: Deep learning in neural networks: An overview. Neural networks **61**, 85–117 (2015)
37. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
38. Singh, C., Walia, E., Kaur, K.P.: Color texture description with novel local binary patterns for effective image retrieval. Pattern Recognition **76**, 50–68 (2018)
39. Subrahmanyam, M., Maheshwari, R., Balasubramanian, R.: Local maximum edge binary patterns: a new descriptor for image retrieval and object tracking. Signal Process. **92**(6), 1467–1479 (2012). https://doi.org/10.1016/j.sigpro.2011.12.005
40. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans. Image Process. **19**(6), 1635–1650 (2010). https://doi.org/10.1109/TIP.2010.2042645
41. Tyagi, V.: Content-based image retrieval techniques: A review. In: Content-Based Image Retrieval, pp. 29–48. Springer (2017)
42. Tzelepi, M., Tefas, A.: Deep convolutional learning for content based image retrieval. Neurocomputing **275**, 2467–2478 (2018)
43. Verdoolaege, G., De Backer, S., Scheunders, P.: Multiscale colour texture retrieval using the geodesic distance between multivariate generalized Gaussian models. In: Proc. IEEE Int. Conf. Image Process. (ICIP). pp. 169–172 (2008). https://doi.org/10.1109/ICIP.2008.4711718
44. Verma, M., Raman, B.: Local tri-directional patterns: A new texture feature descriptor for image retrieval. Digital Signal Processing **51**, 62–72 (2016). https://doi.org/10.1016/j.dsp.2016.02.002

45. Verma, M., Raman, B.: Local neighborhood difference pattern: A new feature descriptor for natural and texture image retrieval. Multimedia Tools and Applications **77**(10), 11843–11866 (2018)
46. Verma, M., Raman, B., Murala, S.: Local extrema co-occurrence pattern for color and texture image retrieval. Neurocomputing **165**, 255–269 (2015). https://doi.org/10.1016/j.neucom.2015.03.015
47. Yang, H.y., Liang, L.l., Zhang, C., Wang, X.b., Niu, P.p., Wang, X.y.: Weibull statistical modeling for textured image retrieval using nonsubsampled contourlet transform. Soft Computing pp. 1–16 (2018)