HAL
open science

Color Texture Image Retrieval Based on Local Extrema Features and Riemannian Distance

Minh-Tan Pham, Grégoire Mercier, Lionel Bombrun

To cite this version:

Minh-Tan Pham, Grégoire Mercier, Lionel Bombrun. Color Texture Image Retrieval Based on Local Extrema Features and Riemannian Distance. Journal of Imaging, 2017, 3 (4), pp.43. 10.3390/jimaging3040043 . hal-01629218

HAL Id: hal-01629218

https://hal.science/hal-01629218v1

Submitted on 6 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Journal of Imaging

MDPI

Article

Color Texture Image Retrieval Based on Local Extrema Features and Riemannian Distance

Minh-Tan Pham [1,*], Grégoire Mercier [2] and Lionel Bombrun [3]

1 IRISA-Université Bretagne Sud, Campus de Tohannic, Rue Yves Mainguy, 56000 Vannes, France
2 TELECOM Bretagne, UMR CNRS 6285 Lab-STICC/CID, 29238 Brest CEDEX 3, France; gregoire.mercier@telecom-bretagne.eu
3 CNRS-IMS Lab. (UMR 5218), University of Bordeaux, 33402 Talence CEDEX, France; lionel.bombrun@u-bordeaux.fr
* Correspondence: minh-tan.pham@irisa.fr

Received: 28 August 2017 ; Accepted: 5 October 2017; Published: 10 October 2017

Abstract: A novel efficient method for content-based image retrieval (CBIR) is developed in this paper using both texture and color features. Our motivation is to represent and characterize an input image by a set of local descriptors extracted from characteristic points (i.e., keypoints) within the image. Then, dissimilarity measure between images is calculated based on the geometric distance between the topological feature spaces (i.e., manifolds) formed by the sets of local descriptors generated from each image of the database. In this work, we propose to extract and use the local extrema pixels as our feature points. Then, the so-called local extrema-based descriptor (LED) is generated for each keypoint by integrating all color, spatial as well as gradient information captured by its nearest local extrema. Hence, each image is encoded by an LED feature point cloud and Riemannian distances between these point clouds enable us to tackle CBIR. Experiments performed on several color texture databases including Vistex, STex, color Brodazt, USPtex and Outex TC-00013 using the proposed approach provide very efficient and competitive results compared to the state-of-the-art methods.

Keywords: content-based image retrieval (CBIR); pointwise approach; local extrema features; color textures; Riemannian distance (RD)

1. Introduction

Texture and color, together with other features such as shape, edge, surface, etc., play significant roles in content-based image retrieval (CBIR) systems thanks to their capacity of providing important parameters and characteristics not only for human vision but also for automatic computer-based visual recognition. Texture has appeared in most CBIR frameworks while color has been more and more exploited to improve retrieval performance, in particular for natural images.

Some recent comprehensive surveys on the CBIR field can be found in [1–3]. Among state-of-the-art propositions, a great number of multiscale texture representation and analysis methods using probabilistic approach have been developed within the past two decades. In [4], Do and Vetterli proposed modeling 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) or its rotated version (DT-RCWT), 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 [5–10]. Then, by taking into account color information within these probabilistic approaches, several studies have provided significant improvement for color

image retrieval [11,12]. However, one of the main drawbacks of these techniques is their expensive computational time, which has been observed an discussed in several papers [7,9,12].

The second family of methods that has drawn the attention of many researchers and has provided quite effective CBIR performance is the local pattern-based framework. The local binary patterns (LBP) descriptor, which compares neighboring pixels to the center pixel and affects them by 0 and 1 to form a binary number, was first embedded in a multiresolution and rotation invariant scheme for texture classification in [13]. Inspired from this work, many propositions have been developed for texture retrieval and classification such as the local derivative patterns (LDP) [14], the local maximum edge binary patterns (LMEBP) [15], the local ternary patterns (LTP) [16], the local tetra patterns (LTrP) [17], local tri-directional patterns (LTriDP) [18], etc. These descriptors, particularly the LTrP, provide quite efficient texture retrieval performance. However, due to the fact that they are applied to gray-scale images, their performance on natural images is limited without exploiting color information. To overcome this issue, several recent strategies have been proposed to incorporate these local patterns with color features. Some techniques that can be stated here are the joint histogram of color and local extrema patterns (LEP+colorhist) [19], the local oppugnant color texture pattern (LOCTP) [20] and the local extrema co-occurrence pattern (LECoP) [21].

Another CBIR system that has offered very competitive results within the past few years relies on an image compression approach called the block truncation coding (BTC). The first BTC-based retrieval scheme for color images was proposed in [22] followed by some improvements a few years later [23,24]. Until very recently, one has been witnessing the evolution of BTC-based retrieval frameworks such as the ordered-dither BTC (ODBTC) [25,26], the error diffusion BTC (EDBTC) [27] and the dot-diffused BTC (DDBTC) [28]. Within these approaches, an image is divided into multiple non-overlapping blocks and one of the BTC-based systems compresses each block into the so-called color quantizer and bitmap image. Then, to characterize the image content, a feature descriptor is constructed using the color histogram feature (CHF) or the color co-occurrence feature (CCF) combined with the bit pattern feature (BPF), which describes edge, shape and texture information of the image. These features are extracted from the above color quantizer and bitmap image to tackle CBIR.

In this work, we would like to develop a powerful color texture image retrieval strategy based on the capacity of characteristic points to capture significant information from the image content. Due to the fact that natural images usually involve a variety of local textures and structures that do not appear homogeneous within the entire image, an approach taking into account local features could become relevant. This may be the reason why most local feature-based CBIR schemes (e.g., LTrP [17], LOCTP [20], LECoP [21]) or BTC-based approaches [25–28] (i.e., which, in fact, sub-divide each query image into multiple blocks) have achieved better retrieval performance than probabilistic methods, which model the entire image using different statistical distributions [4–9,11,12,29]. We will provide later their performance for a comparison within the experimental study. Going back to this paper, by taking into consideration the above-mentioned remark, our motivation here is to represent and characterize an input image by a set of local descriptors generated only for characteristic points detected from the image. Recent study [30–35] addressing texture-based image segmentation and classification using very high resolution (VHR) remote sensing images has proved the capacity of the local extrema (i.e., local maximum and local minimum pixels in terms of intensity) to capture and represent textural features from an image. In this work, the idea of using the local extrema pixels is continued and improved to tackle CBIR. By embedding both color and geometric information, as well as gradient features, captured by these feature points, we propose the local extrema-based descriptor (LED) for texture and color description. As a result, an input image can be encoded by a set of LEDs, which is considered as an LED feature point cloud. Then, we propose to exploit a geometric-based distance measure, i.e., the Riemannian distance (RD) [36] between the feature covariance matrices of these point clouds, for dissimilarity measurement. This distance takes into account the topological structure of each point cloud within the LED feature space. Therefore, it becomes an effective measure of dissimilarity in our CBIR scheme.

The remainder of this paper is organized as follows. Section 2 presents the pointwise approach for local texture representation and description using local extrema features and the construction of LED descriptors. The proposed CBIR framework is described in details in Section 3. Section 4 carries out our experimental study and provides comparative results of the proposed algorithm against existing methods. Conclusions and some perspectives of our work are discussed in Section 5.

2. Local Texture Representation and Description Using Local Extrema Features

2.1. Approach

The idea of using the local maximum and local minimum pixels for texture representation and characterization has been introduced in [30–32] in the scope of VHR optical satellite images. Regarding this point of view, an image texture is formed by a certain spatial arrangement of pixels holding some variations of intensity. Hence, different textures are reflected by different types of pixel's spatial arrangements and intensity variations. These meaningful properties can be approximated by the local maximum and local minimum pixels extracted from the image. These local extrema have been proved to be able to capture the important geometric and radiometric information of the image content, and hence is relevant for texture analysis and description. In this work, we exploit and improve the capacity of local extrema pixels to characterize both texture and color features within CBIR context. Let us first recall their definition and extraction.

A pixel in a grayscale image is supposed to be a local maximum (resp. local minimum) if it holds the highest (resp. lowest) intensity value within a neighborhood window centered at it. Let $S_\omega^{\max}(I)$ and $S_\omega^{\min}(I)$ denote the local maximum and local minimum sets extracted from a grayscale image I using the $\omega \times \omega$ search window. Let $p = (x_p, y_p)$ be a pixel located at position (x_p, y_p) on the image plane having its intensity value $I(p)$, we have:

$$p \in S_\omega^{\max}(I) \Leftrightarrow \left\{ I(p) = \max_{q \in \mathcal{N}_{\omega\times\omega}(p)} I(q) \right\}, \tag{1}$$

$$p \in S_\omega^{\min}(I) \Leftrightarrow \left\{ I(p) = \min_{q \in \mathcal{N}_{\omega\times\omega}(p)} I(q) \right\}, \tag{2}$$

where $\mathcal{N}_{\omega\times\omega}(p)$ represents a set of pixels inside the $\omega \times \omega$ neighborhood window of p. We note that there are many ways to extend the above definition for color images (i.e., detecting on the grayscale version, using the union or intersection of subsets detected on each color channel, etc.). In this work, the local extrema pixels are extracted from the grayscale version of color image since this produced superior performance within most of our experiments.

Figure 1 illustrates the capacity of the local max and min pixels to represent, characterize and discriminate different textures that are extracted from the Vistex image database [37]. Each image patch consists of 100×100 pixels. We display for each one a 3D surface model using the image intensity as the surface height. The local max pixels (in red) and local min pixels (in green) are extracted following Equations (1) and (2) by a 5×5 search window. Some green points may be unseen since they are obscured by the surface. We observe from the figures how these local extrema appear within each texture. Their spatial information (relative distance, direction, density, etc.) and intensities can be encoded to describe and discriminate each texture, which constitutes our motivation in this paper.

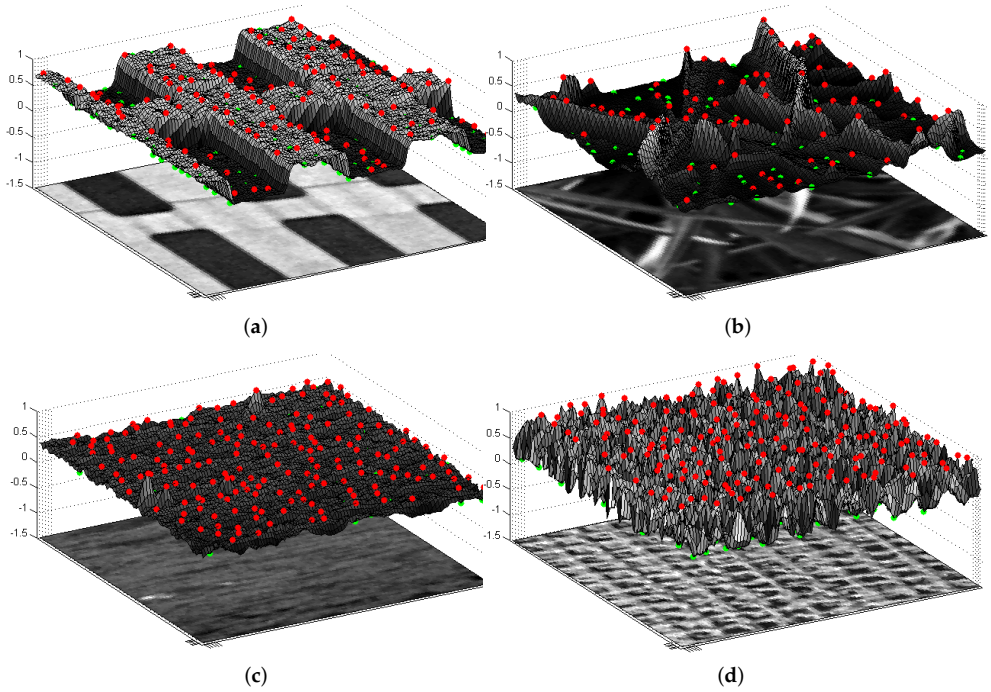


Figure 1. Illustration: spatial distribution and arrangement of local maximum pixels (in red) and local minimum pixels (in green) within four different textures (size of 100 × 100 pixels) extracted from the Vistex database [37]: *Buildings.0009*, *Leaves.0016*, *Water.0005* and *Fabric.0017*. The local extrema are detected using a 5 × 5 search window. The figure is better visualized in colors. (**a**) *Buildings.0009*; (**b**) *Leaves.0016*; (**c**) *Water.0005*; (**d**) *Fabric.0017*.

2.2. Generation of Local Extrema-Based Descriptor (LED)

As previously discussed, our work not only considers the local extrema pixels as characteristic points to represent the image but also exploits them to construct texture and color descriptors. The local extrema-based descriptor (LED) is generated for each keypoint by integrating the color, spatial and gradient features captured by its nearest local maxima and local minima on the image plane. Two strategies can be considered for the research of nearest local extrema around each keypoint:

1. fix the number (N) of nearest local maxima and nearest local minima for each one;
2. or, fix a window size $W \times W$ around each keypoint; then, all local maxima and minima inside that window are considered.

The first strategy takes into account the local properties around keypoints since the implicit neighborhood size considered for each one varies depending on the density of local extrema around it (i.e., the search process goes further to look for enough extrema when their density is sparse). On the other hand, by fixing the window size, the second approach considers equivalent contributions of neighboring environments for all keypoints and hence better deals with outlier points. Moreover, its implementation is less costly than the first one, especially for large-size images (i.e., the search of N points is more costly within large-size image). Nevertheless, our experimentation shows that both strategies provide similar performance for the studied data sets. In this paper, we describe the approach considering a fixed window size for all keypoints, but it is worth noting that a similar principle can be applied to build the proposed descriptor using the first approach.

The generation of LED descriptors is described as follows. Let $\mathcal{N}_W^{\max}(p)$ and $\mathcal{N}_W^{\min}(p)$ be the two set of local maximum and local minimum pixels inside the $W \times W$ window around the understudied keypoint p, and the following features involving color, spatial and gradient information are extracted for each set. For better explanation, we display in Figure 2 the geometric and gradient

features derived from each local max or local min within these sets. Below are the features extracted from $\mathcal{N}_W^{\max}(p)$, the feature generation for $\mathcal{N}_W^{\min}(p)$ is similar.

1. Mean and variance of three color channels:

$$\mu_c^{\max}(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} I_c(q), \tag{3}$$

$$\sigma_c^{2\max}(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} (I_c(q) - \mu_c^{\max}(p))^2, \tag{4}$$

where $c \in \{\text{red}, \text{green}, \text{blue}\}$ represents the three color components and $|\mathcal{N}|$ is the cardinality of the set $\mathcal{N}$.

2. Mean and variance of spatial distances from each local maximum to point p:

$$\mu_d^{\max}(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} d(p,q), \tag{5}$$

$$\sigma_d^{2\max}(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} (d(p,q) - \mu_d^{\max}(p))^2, \tag{6}$$

where $d(p,q) = \sqrt{(x_p - x_q)^2 + (y_p - y_q)^2}$ is the spatial distance between two points p and q on the image plane. We remind readers that $p = (x_p, y_p)$ and $q = (x_q, y_q)$.

3. Circular variance [38] of angles of geometric vectors formed by each local maximum and point p:

$$\sigma_{\text{cir},\alpha}^{2\max}(p) = 1 - \sqrt{\bar{c}_\alpha(p)^2 + \bar{s}_\alpha(p)^2}, \tag{7}$$

where

$$\bar{c}_\alpha(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} \cos\alpha(p,q),$$

$$\bar{s}_\alpha(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} \sin\alpha(p,q),$$

$$\alpha(p,q) = \arctan\left(\frac{y_q - y_p}{x_q - x_p}\right),\ \alpha(p,q) \in [-\pi, \pi], \forall p, q.$$

4. Mean and variance of gradient magnitudes:

$$\mu_g^{\max}(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} \nabla I(q), \tag{8}$$

$$\sigma_g^{2\max} = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} \left(\nabla I(q) - \mu_g^{\max}(p)\right)^2, \tag{9}$$

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

5. Circular variance [38] of gradient orientations:

$$\sigma_{\text{cir},\theta}^{2\max}(p) = 1 - \sqrt{\bar{c}_\theta(p)^2 + \bar{s}_\theta(p)^2}, \tag{10}$$

where

$$\bar{c}_\theta(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} \cos\theta(q),$$

$$\bar{s}_\theta(p) = \frac{1}{|\mathcal{N}_W^{\max}(p)|} \sum_{q \in \mathcal{N}_W^{\max}(p)} \sin\theta(q).$$

θ is the gradient orientation image obtained together with ∇I by Sobel filtering.

All of these features are then integrated into the feature vector $\delta^{\max}(p)$, which encodes the local max set around p:

$$\begin{aligned}\delta^{\max}(p) = \Big[&\mu_{\text{red}}^{\max}(p), \sigma_{\text{red}}^{2\max}(p), \mu_{\text{green}}^{\max}(p), \sigma_{\text{green}}^{2\max}(p),\\ &\mu_{\text{blue}}^{\max}(p), \sigma_{\text{blue}}^{2\max}(p), \mu_{d}^{\max}(p), \sigma_{d}^{2\max}(p),\\ &\sigma_{\text{cir},\alpha}^{2\max}(p), \mu_{g}^{\max}(p), \sigma_{g}^{2\max}(p), \sigma_{\text{cir},\theta}^{2\max}(p)\Big].\end{aligned} \tag{11}$$

The generation of $\delta^{\min}(p)$ is similar. Now, let $\delta^{\text{LED}}(p)$ be the LED feature vector extracted for p, we have:

$$\delta^{\text{LED}}(p) = \left[I_{\text{red}}(p), I_{\text{green}}(p), I_{\text{blue}}(p), \delta^{\max}(p), \delta^{\min}(p)\right]. \tag{12}$$

The proposed descriptor $\delta^{\text{LED}}(p)$ enables us to characterize the local environment around p by understanding how local maxima and local minima are distributed and arranged, and also how they capture structural properties (given by gradient features) as well as color information. We note that LED descriptors are invariant to rotation. As observed from the feature computation process, for the two directional features including angles α and gradient orientations θ, only their circular variance [38] is taken into account, and their circular mean is not involved to ensure the rotation-invariant property.

The feature dimensionality of δ^{LED} in Equation (12) is equal to 27. One may realize that this descriptor can be feasibly improved by modifying or adding other features into the vector. One of the improvements can be considered here is to involve other filtering approaches for gradient computation. For example, besides the gradient images $(\nabla I, \theta)$, it is interesting to add $(\nabla I_\sigma, \theta_\sigma)$, which are generated by smoothing the image by a lowpass Gaussian filter ($I_\sigma = G_\sigma * I$) before applying the Sobel operator. Then, similar gradient features are extracted as in Equations (8)–(10). By inserting these features (computed for two extrema sets) into $\delta^{\text{LED}}(p)$, an enhanced descriptor of dimension 33 is built. Both 27D and 33D versions of LED descriptors are going to be experimented later in the paper.

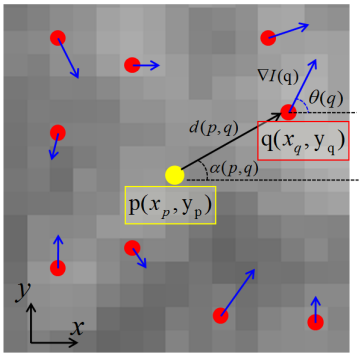


Figure 2. Geometric and gradient information from a local maximum (resp. local minimum) pixel $q = (x_q, y_q)$ within $\mathcal{N}_W^{\max}(p)$ (resp. $\mathcal{N}_W^{\min}(p)$) considered for the calculation of LED descriptor at the studied keypoint $p = (x_p, y_p)$. Here, $d(p,q)$ is the distance between p and q; $\alpha(p,q)$ is the angle of the vector yielded from p to q. We have $d(p,q) = d(q,p)$ but $\alpha(p,q) \neq \alpha(q,p)$. Then, $\nabla I(q)$, $\theta(q)$ are the gradient magnitude and gradient orientation at q.

3. Proposed Framework for Texture and Color Image Retrieval

The proposed retrieval algorithm consists of two primary stages: the extraction of LED texture and color descriptors to characterize each query image from the image database and the computation

of distance measure for retrieval process. Each of them is now described in detail. Then, the complete framework is presented.

3.1. LED Feature Extraction

The previous section has proved the capacity of the local extrema pixels to represent and characterize both local texture and color information from the image content. In this section, each query image will be encoded by a set of LED descriptors generated at characteristic points (i.e., keypoints) extracted from the image. In this work, we propose using the local maxima as our keypoints. It should be noted that other keypoint types can be exploited such as Harris keypoints, Scale Invariant Feature Transform (SIFT) [39], Speeded Up Robust Features (SURF) [40], etc. (see a survey in [41]). However, we choose local maxima as keypoints since they can be extracted within any kind of textures (i.e., with variation of intensity). Meanwhile, such keypoints like Harris, SIFT or SURF are normally focused on corners, edges, and salient features so that they may not probably be detected from some quite homogeneous textures within natural scenes such as sand, water, early grass fields, etc.

We propose that our keypoints are the local max pixels detected by a search window of size $\omega_1 \times \omega_1$ and LED descriptors are generated for each keypoint with the support of the local max and local min sets extracted by another search window of size $\omega_2 \times \omega_2$. Here, we separate the two window sizes since ω_2 should be small enough to ensure sufficient max and min pixels around each keypoint (i.e., dense enough) for descriptor computation. On the other hand, the density of keypoints within the image can be lower than or equal to the density of local extrema by setting $\omega_1 \geq \omega_2$. It is revealed that a coarser density of keypoints (by increasing ω_1) still produces good retrieval results but reduces quite a lot of computation time. The sensitivity of the proposed method to these window sizes will be analyzed in Section 4.3.3.

Subsequently, a set of LED descriptors is constructed for the keypoint set to characterize the image. Given a neighborhood window size $W \times W$, the computation of LED feature vector for each keypoint is carried out as described in Section 3.1 (see Equations (3)–(12)). Similarly, the algorithm sensitivity to parameter W is discussed in Section 4.3.3.

3.2. Dissimilarity Measure for Retrieval

Each query image from the database is now represented and characterized by a set of LED descriptors that can be considered as an LED feature point cloud. The dissimilarity measure between two images becomes the distance between two LED feature point clouds. Several metrics can be performed. One may consider that these sets of feature vectors follow a multivariate normal distribution and thus employ the simplified Mahalanobis distance or the symmetric Kullback–Leibler distance for dissimilarity measurement as follows:

$$d_{\text{Mahal}} = (\mu_1 - \mu_2)^T \left(C_1^{-1} + C_2^{-1}\right)(\mu_1 - \mu_2), \tag{13}$$

$$d_{\text{KL}} = \text{trace}\left(C_1 C_2^{-1} + C_2 C_1^{-1}\right) + (\mu_1 - \mu_2)^T \left(C_1^{-1} + C_2^{-1}\right)(\mu_1 - \mu_2), \tag{14}$$

where μ_1, C_1, μ_2, C_2 are the estimated means and covariance matrices of the normal distributions for the two point clouds.

Others would like to only account for the space of the feature covariance matrices by supposing that these point clouds have the same mean feature vector. Hence, their dissimilarity measure is calculated based on the distance between their covariance matrices. Our proposition in this paper is to utilize a geometric-based distance between the feature covariance matrices estimated from each set of LED descriptors. Thanks to the fact that these feature covariance matrices possess a positive semi-definite structure, their Riemannian distance (which is also called geodesic distance in some studies) is proposed to become the distance measure in our retrieval scheme. The Riemannian distance between two covariance matrices is defined in [36] as follows:

$$d_{RD}(C_1, C_2) = \sqrt{\sum_{\ell=1}^{d} \log^2 \lambda_\ell}, \tag{15}$$

where λ_ℓ is the ℓth generalized eigenvalue satisfying $\lambda_\ell C_1 \chi_\ell - C_2 \chi_\ell = 0, \ell = 1 \ldots d$. χ_ℓ is the corresponding eigenvector to λ_ℓ and d is the feature dimensionality of LED feature vectors (i.e., 27D or 33D as described in Section 2.2).

Experiments show that the Riemannian distance (RD) outperforms the Mahalanobis and the symmetric Kullback–Leibler distances in terms of both retrieval accuracy and computational time. Moreover, by using the RD, only one feature covariance matrix needs to be computed and stocked for each query image. On the other hand, the distances in Equations (13) and (14) also involve the mean feature vector for their computation, and hence require more computer storage memory.

Last but not least, for better confirming and validating our choice of Riemannian distance, we will provide in Section 4.3.4 a detailed comparison (i.e., in terms of retrieval performance and computation time) of this metric not only with the mentioned Mahalanobis and Kullback–Leibler metrics but also against some other distance measures of covariance matrices such as the log-Euclidean, the Wishart-like and the Bartlett distances [42,43].

3.3. Proposed Retrieval Framework

The full proposed framework for texture and color-based image retrieval using LED features and Riemannian distance is outlined in Figure 3. The algorithm can be highlighted as follows:

1. Load the query color image I_{color}.
2. Convert the image to grayscale I.
3. Compute gradient images from I:
   + $(\nabla I, \theta)$ for the 27D version,
   + $(\nabla I, \nabla I_\sigma, \theta, \theta_\sigma)$ for the enhanced 33D version.
4. Extract the keypoint set and the two local extrema sets from I:
   + keypoint set: $S(I) = S^{\max}_{\omega_1}(I)$,
   + extrema sets: $S^{\max}_{\omega_2}(I)$ and $S^{\min}_{\omega_2}(I)$.
5. Generate LED descriptors for all keypoints:
   + $\forall p \in S(I)$, extract $\delta^{\text{LED}}(p)$ as (12).
6. Estimate the feature covariance matrice for these LED descriptors.
7. Compute the Riemannian distance (15) between the query and the other images from the database.
8. Sort these distance measures and produce the best matches as the final retrieval result for the query.

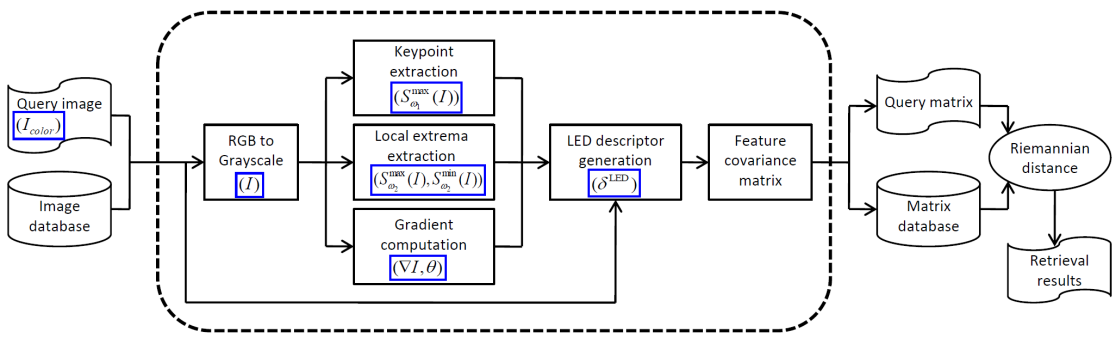


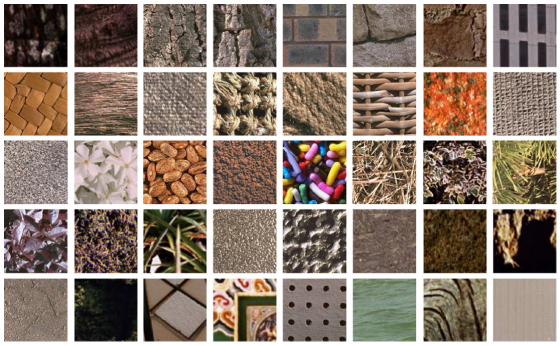
Figure 3. Proposed framework for color texture image retrieval using the local extrema-based descriptors and the Riemannian distance.

4. Experimental Study

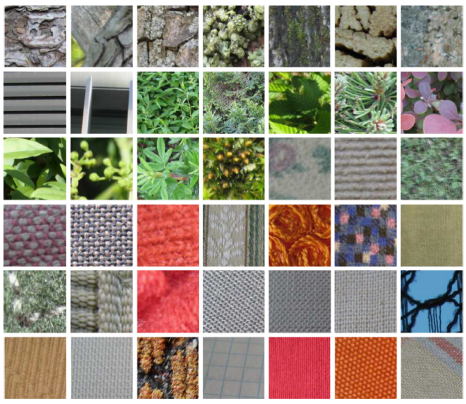
4.1. Image Databases

Five popular color texture databases including the MIT Vision Texture database (Vistex) [37], the Salzburg Texture database (Stex) [44], the Colored Brodatz Texture database (CBT) [45], the USPtex [46] and the Outex TC-00013 [47] are exploited to perform our experimental study.

Vistex is one of the most widely used texture databases for performance assessment and comparative study in the CBIR field. It consists of 40 texture images (i.e., 40 texture classes) of size 512 × 512 pixels. As in most literature studies, each texture image is divided into 16 non-overlapping subimages of size 128 × 128 pixels in order to create a database of 640 subimages (i.e., 40 classes × 16 subimages/class). Being much larger, the Stex database is a collection of 476 color texture images captured in the area around Salzburg, Austria under real-word conditions. Like for Vistex, each 512 × 512 texture image is divided into 16 non-overlapping patches to build a total of 7616 images. Next, the Colored Brodatz Texture (CBT) database is an extension of the gray-scale Brodatz texture database [45]. This database both preserves the rich textural content of the original Brodatz and possesses a wide variety of color content. Hence, it becomes relevant for the evaluation of texture and color-based CBIR algorithms. The CBT database consists of 112 textures of size 640 × 640 pixels. Each one is divided into 25 non-overlapping subimages of size 128 × 128 pixels, thus creating 2800 images in total (i.e., 112 classes × 25 images/class). The USPtex database [46] 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. Finally, the Outex TC-00013 [47] 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.



(a)



(b)

(c)

Figure 4. *Cont.*

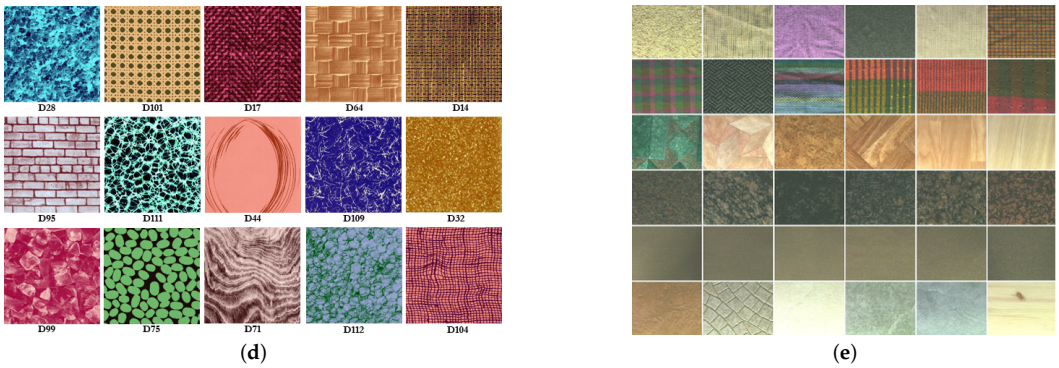


Figure 4. Five image databases used in the experimental study. (**a**) Vistex-640 [37]; (**b**) Stex-7616 [44]; (**c**) USPtex-2292 [46]; (**d**) CBT-2800 [45]; (**e**) Outex-1360 [47].

For our convenience in this work, we name the above five texture databases as Vistex-640, Stex-7616, CBT-2800, USPtex-2292 and Outex-1360 where the related number represents the total number of samples in each database. Figure 4 shows some examples of each database and Table 1 provides a summary of their information.

Table 1. Summary of texture databases used in the experimental study.

Database	Number of Classes	Number of Images/Class	Total Number
Vistex-640	40	16	640
Stex-7616	476	16	7616
CBT-2800	112	25	2800
USPtex-2292	191	12	2292
Outex-1380	68	20	1380

4.2. Evaluation Criteria

The main criterion used to assess the performance of the proposed retrieval framework compared to state-of-the-art methods is the average retrieval rate (ARR). 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:

$$\mathrm{ARR}(K) = \frac{1}{N_t \times N_R} \sum_{q=1}^{N_t} n_q(K) \Big|_{K \geq N_R}. \tag{16}$$

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 are adopted by setting $K = N_R$.

4.3. Results and Discussion

4.3.1. Performance in Retrieval Accuracy

The proposed CBIR system was applied to the five studied databases. For all cases, the local maximum and local minimum sets (i.e., $S_{\omega_2}^{\max}$ and $S_{\omega_2}^{\min}$) were extracted using a 3×3 search window ($\omega_2 = 3$). The local max keypoints were detected by the 5×5 window ($\omega_1 = 5$) and the neighborhood size for LED descriptor construction (W) was set to 30×30 pixels. Here, we set $\omega_1 > \omega_2$ to accelerate the computational time. Meanwhile, as discussed in Section 3.1, a denser or coarser density of

keypoints (by varying ω_1) could produce stable performance. Our experiments showed that ω_1 set from 3 to 9 could yield very close ARR. Similarly, a value of W from 15 to 40 could provide quite equivalent retrieval results. We deeply discuss the sensitivity of our algorithm to these parameters later in Section 4.3.3.

Tables 2–6 show the average retrieval rate (ARR) of the proposed framework performed on our five databases compared to several state-of-the-art methods in the literature. Then, Table 7 provides the descriptor size (i.e., feature dimension) of some of these methods. We note that some learned features extracted from pre-trained convolutional neural networks (CNNs) [48,49] that have been recently proposed by [50,51] were also investigated to enrich our comparative study. From these tables, the first observation is that recent CBIR propositions have taken into consideration more and more color information in order to improve their retrieval performance. This trend is reasonable since color components play a significant role in natural images. The second remark is that, as previously mentioned in the introduction of our article, the local patterns-based schemes (such as LtrP [17], LECoP [21], etc.) and the OTB-based systems [26–28] as well as the recently proposed learned descriptors based on pre-trained CNNs [50,51] generally provide higher ARR than the wavelet-based probabilistic approaches [4–9,11,12,29]. Then, more importantly, our LED+RD framework (both the 27D version and the improved 33D version) has outperformed all reference methods for all the three databases. We now discuss the results for each database to validate the effectiveness of the proposed strategy.

An ARR improvement of 1.47% (i.e., from 93.23% to 94.70%) and 3.68% (i.e., from 76.40% to 80.08%) is observed for the Vistex-640 and Stex-7616 databases, respectively, using the proposed algorithm. Within the proposed strategy, the 33D improved LED gives a slightly higher ARR compared to the 27D version (i.e., 0.06% for Vistex-640 and 0.13% for Stex-7616). This means that our framework could be still enhanced by modifying or incorporating other features for LED construction. Next, an important remark is that during our experimentation, most of the texture classes with strong structures and local features such as buildings, fabric categories, man-made object's surfaces, etc. were retrieved with 100% accuracy. Table 3 shows the per-class retrieval rate for each image class in the Vistex-640 database. As observed from the table, half of all the classes (19/40 classes by 27D LED and 20/40 classes by 33D LED) were indexed with 100% accuracy. These perfectly-retrieved image classes in general consist of many local textures and structures. Similar behavior was also remarked for the Stex-7616 database. This issue is encouraging since our motivation and expectation in this work are to exploit and emphasize local features inside each image in order to tackle the retrieval problem.

Table 2. Average retrieval rate (ARR) on **Vistex-640** database by the proposed method compared to the state-of-the-art methods.

Method	Using Color	ARR (%)
GT+GGD+KLD [4]	-	76.57
DT-CWT [5]	-	80.78
DT-CWT+DT-RCWT [5]	-	82.34
MGG+Gaussian+KLD [11]	✓	87.40
MGG+Laplace+GD [11]	✓	91.70
DCT+MGMM [7]	-	84.94
Gaussian Copula+Gamma+ML [12]	✓	89.10
Gaussian Copula+Weibull+ML [12]	✓	89.50
Student-t Copula+GG+ML [12]	✓	88.90
LMEBP [15]	-	87.77
Gabor LMEBP [15]	-	87.93
LtrP [17]	-	90.02
Gabor LtrP [17]	-	90.16
LEP+colorhist [19]	✓	82.65
MCMCM+DBPSP [52]	✓	86.17
Gaussian Copula-MWbl [9]	-	84.41
ODBTC [26]	✓	90.67

Table 2. *Cont.*

Method	Using Color	ARR (%)
Gaussian Copula+Gabor Wavelet [10]	✓	92.40
EDBTC [27]	✓	92.55
DDBTC [28]	✓	92.65
LECoP [21]	✓	92.99
ODII [25]	✓	93.23
CNN-AlexNet [51]	✓	91.34
CNN-VGG16 [51]	✓	92.97
CNN-VGG19 [51]	✓	93.04
Proposed LED+RD (27D)	✓	**94.64**
Proposed LED+RD (33D)	✓	**94.70**

Table 3. Per-class retrieval rate (%) on the **Vistex-640** database using the proposed LED+RD method.

Class	27D	33D	Class	27D	33D
Bark.0000	76.95	75.00	Food.0008	100.00	100.00
Bark.0006	98.05	98.05	Grass.0001	94.53	93.36
Bark.0008	83.20	84.38	Leaves.0008	99.61	100.00
Bark.0009	77.13	78.13	Leaves.0010	100.00	100.00
Brick.0001	100.00	99.22	Leaves.0011	100.00	100.00
Brick.0004	97.66	98.05	Leaves.0012	55.86	56.25
Brick.0005	100.00	100.00	Leaves.0016	86.72	86.72
Buildings.0009	100.00	100.00	Metal.0000	98.83	99.61
Fabric.0000	100.00	100.00	Metal.0002	100.00	100.00
Fabric.0004	76.95	77.34	Misc.0002	100.00	100.00
Fabric.0007	99.61	99.61	Sand.0000	100.00	100.00
Fabric.0009	100.00	100.00	Stone.0001	82.42	82.42
Fabric.0011	100.00	100.00	Stone.0004	90.63	91.02
Fabric.0014	100.00	100.00	Terrain.0010	94.92	95.31
Fabric.0015	100.00	100.00	Tile.0001	91.80	89.45
Fabric.0017	96.48	97.66	Tile.0004	100.00	100.00
Fabric.0018	98.83	100.00	Tile.0007	100.00	100.00
Flowers.0005	100.00	100.00	Water.0005	100.00	100.00
Food.0000	100.00	100.00	Wood.0001	96.88	96.88
Food.0005	99.22	99.22	Wood.0002	88.28	90.23
			ARR	**94.64**	**94.70**

Table 4. Average retrieval rate (ARR) on **Stex-7616** database by the proposed method compared to the state-of-the-art methods.

Method	Using Color	ARR (%)
GT+GGD+KLD [4]	-	* 49.30
DT-CWT+Weibull+KLD [6]	-	* 58.80
MGG+Laplace+GD [11]	✓	* 71.30
DWT+Gamma+KLD [8]	-	* 52.90
Gaussian Copula+Gamma+ML [12]	✓	69.40
Gaussian Copula+Weibull+ML [12]	✓	70.60
Student-t Copula+GG+ML [12]	✓	65.60
LEP+colorhist [19]	✓	59.90
DDBTC [28]	✓	44.79
LECoP [21]	✓	74.15
Gaussian Copula+Gabor Wavelet [10]	✓	76.40
CNN-AlexNet [51]	✓	68.84
CNN-VGG16 [51]	✓	74.92
CNN-VGG19 [51]	✓	73.93
Proposed LED+RD (27D)	✓	**79.95**
Proposed LED+RD (33D)	✓	**80.08**

(*) These results are extracted from [12], not from the original papers.

Table 5. Average retrieval rate (ARR) on **CBT-2800** database by the proposed method compared to the state-of-the-art methods.

Method	Using Color	ARR (%)
LBP [13]	-	* 81.75
LtrP [17]	-	* 82.05
LOCTP-YCbCr [20]	✓	84.46
LOCTP-HSV [20]	✓	88.60
LOCTP-LAB [20]	✓	88.90
LOCTP-RGB [20]	✓	93.89
CNN-AlexNet [51]	✓	90.72
CNN-VGG16 [51]	✓	91.64
CNN-VGG19 [51]	✓	90.36
Proposed LED+RD (27D)	✓	**99.06**
Proposed LED+RD (33D)	✓	**98.79**

(*) These results are extracted from [20], not from the original papers.

Table 6. Average retrieval rate (%) on **USPtex-2292** and **Outex-1360** databases by the proposed method compared to some reference methods.

Method	UPStex-2292	Outex-1360
DDBTC (L_1) [28]	63.19	61.97
DDBTC (L_2) [28]	55.38	57.51
DDBTC (χ^2) [28]	73.41	65.54
DDBTC (Canberra) [28]	74.97	66.82
CNN-AlexNet [51]	83.57	69.87
CNN-VGG16 [51]	85.03	72.91
CNN-VGG19 [51]	84.22	73.20
Proposed LED+RD (27D)	**90.22**	**76.54**
Proposed LED+RD (33D)	**90.50**	**76.67**

Table 7. Comparison of feature vector length of different methods.

Method	Feature Dimension
DT-CWT [4]	$(3 \times 6 + 2) \times 2 = 40$
DT-CWT+DT-RCWT [4]	$2 \times (3 \times 6 + 2) \times 2 = 80$
LBP [13]	256
LTP [16]	$2 \times 256 = 512$
LMEBP [15]	$8 \times 512 = 4096$
Gabor LMEBP [15]	$3 \times 4 \times 512 = 6144$
LEP+colorhist [19]	$16 \times 8 \times 8 \times 8 = 8192$
LECoP($H_{18}S_{10}V_{256}$) [21]	$18 + 10 + 256 = 284$
LECoP($H_{36}S_{20}V_{256}$) [21]	$36 + 20 + 256 = 312$
LECoP($H_{72}S_{20}V_{256}$) [21]	$72 + 20 + 256 = 348$
ODII [25]	128 + 128 = 256
CNN-AlexNet [51]	4096
CNN-VGG16 [51]	4096
CNN-VGG19 [51]	4096
Proposed LED+RD (27D)	27
Proposed LED+RD (33D)	33

For the CBT-2800 database, a nearly perfect retrieval performance was produced by the proposed method. In Table 5, the 27D LED version yielded slightly better performance (99.06%) than the 33D version (98.79%). Compared to the LOCTP [20], an enhancement of 5.17% in ARR (i.e., 99.06% compared to 93.89%) is achieved. Since the CBT database has been recently created [45], there are not many studies working on it. In [20], the authors only considered 110 texture classes to create a database of 2750 sub-images, instead of 2800 sub-images from 112 classes as in our work (Section 4.1).

However, since our database is more complete and ARR is supposed not to change too much within 110 or 112 classes, we still consider the retrieval performance in [20] as reference results for our comparative study in Table 5. Finally, Table 6 shows that the proposed method also provided the best ARR for both USPtex-2292 (90.50%) and Outex-1360 (76.67%) using the 33D version of LED. The 27D version produced slightly lower ARR but still very effective (90.22% for USPtex-2292 and 76.54% for Outex-1360). Compared to the pre-trained CNN-based feature extraction technique [51], an improvement of 5.47% and 3.47% has been gained, respectively. To this end, in terms of retrieval accuracy, the efficiency of the proposed CBIR framework compared to reference methods is confirmed and validated for all tested databases.

4.3.2. Computation Time

Table 8 displays the computational cost required by the proposed strategy in terms of feature extraction (FE) time and dissimilarity measurement (DM) time performed on the Vistex-640 database. All implementations were carried out using MATLAB 2012 on a laptop machine *Core i7-3740QM 2.7GHz, 16GB RAM*. In short, a total time of 445.1 s is required by the 27D version to yield an ARR of 94.64% for the Vistex-640 database. For the 33D version, it needs 504.7 s to deliver an accuracy of 94.70%. One remark is that the DM stage only requires 0.035 s (resp. 0.044 s) to measure the distance from each query image to 640 candidates from the database. In fact, to compute the Riemannian distance in Equation (15), only a generalized eigenvalue problem with matrix size of 27×27 (resp. 33×33) needs to be solved. Our experiments proved that this distance is less costly than other metrics (i.e., except the log-Euclidean) mentioned in Section 3.2. A detailed comparison will be given later in Section 4.3.4.

Table 8. Performance of the proposed method in terms of feature extraction (FE) time and dissimilarity measurement (DM) time. Experiments were conducted on the Vistex-640 database.

Version	FE Time (s)		DM Time (s)		Total Time (s)		ARR (%)
	t_{data}	t_{image}	t_{data}	t_{image}	t_{data}	t_{image}	
27D	422.8	0.661	22.3	0.035	445.1	0.695	94.64
33D	476.6	0.745	28.1	0.044	504.7	0.789	94.70

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

Some other experiments were also conducted to compare the processing time for LED extraction with other popular descriptors including SIFT [39], SURF [40], Binary robust invariant scalable keypoints (BRISK) [53] and Binary robust independent elementary features (BRIEF) [54] . In order to perform a fair comparison, we propose that all descriptors will be constructed from the same number of 100 SURF keypoints extracted from a texture image of size 512×512 pixels. Comparative results are reported in Table 9. From the table, we observe that the proposed LED needs more computational time than the other descriptors. However, since the computation time is not the main favor of our work, this issue is not so significant and does not neglect the contribution of the proposed strategy. In addition, we note that our implementation of LED was done in a very standard way without any code optimization. Further improvements in terms of optimizing codes may possibly reduce LED computation time.

Table 9. Comparison of feature description time of different methods.

Method	Feature Dimension	Extraction Time (ms)
SIFT [39]	128	538.6
SURF [40]	64	162.2
BRISK [53]	64	8.2
BRIEF [54]	32	3.2
Proposed LED (27D)	27	1298.7
Proposed LED (33D)	33	1476.3

4.3.3. Sensitivity Analysis

This subsection aims at studying the sensitivity of the proposed method to its parameters. As observed from the full algorithm in Section 3.3, only three parameters are required to perform our retrieval system: the window size to detect keypoints (ω_1), the window size to extract local maximum and local minimum pixels (ω_2) and the window size to generate LED descriptors (W). Since ω_2 needs to be small enough to ensure a sufficiently good density of local extrema to support the computation of LED descriptor. It is fixed to 3×3 pixels in our work. We now investigate the sensitivity to the other two parameters, ω_1 and W.

Figure 5a shows the performance of the LED+RD algorithm (version 27D) obtained by fixing $\omega_2 = 3$, $W = 30$ and varying ω_1 from 3 to 11. Experiments were conducted on the Vistex-640 data. We observe a stable performance in ARR (varying from 94.18% to 94.64%) for $3 \leq \omega_1 \leq 9$. Then, a decrease of 1.45% in ARR results when it switches from 9 to 11. In terms of computational cost, the feature extraction time is significantly reduced when ω_1 increased. Although the best performance in ARR (94.64%) is obtained by setting $\omega_1 = 5$, one may prefer setting it to 7 or 9 to speed up the computation time. Indeed, by using $\omega_1 = 7$, we gain about 28% of time (i.e., reduced from 0.661 s to 0.473 s per image), but only a reduction of 0.16% in ARR (i.e., from 94.64% to 94.48%) results. Similarly, an ω_1 set to 9 could save 44% of time (i.e., reduced from 0.661 s to 0.373 s per image) and yields 94.18% retrieval accuracy (i.e., reducing only 0.46%).

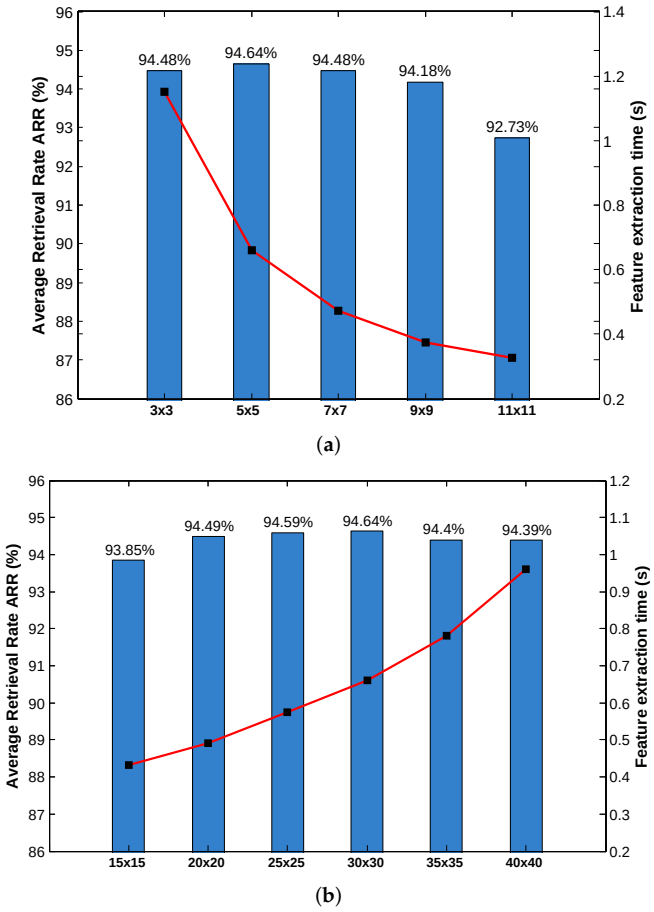


Figure 5. Sensitivity of the proposed method to its parameters in terms of average retrieval rate (%) and feature extraction time (s). Experiments were conducted on Vistex-640 data set using the 27D LED+RD. (**a**) sensitivity to window size ω_1 for keypoint extraction; (**b**) sensitivity to window size W for descriptor generation.

Similarly, the algorithm sensitivity to W (i.e., window size used for the generation of LED descriptors) can be found in Figure 5b. Again, a stable performance can be observed when W varies from 15 to 40. Here, we set ω_1 to 5 and ω_2 always to 3. The highest ARR is adopted by setting W to 30. However, a smaller window size $W = 20$ may be more interesting since it reduces only 0.15% of ARR (i.e., reduced from 94.64% to 94.49%) while speeding up by about 21% the feature extraction time (i.e., from 0.661 s to 0.523 s per image). As a result, a total time of only 349.2 seconds is necessary for our framework to produce 94.49% retrieval accuracy for Vistex-640 database. This issue makes the proposed strategy become very effective and competitive in both retrieval performance and computational cost. Furthermore, Figure 5a,b show that the proposed LED method is not very sensitive to its parameters. A stable performance can be adopted with a wide range of parameters: $\omega_1 \in [3,\ 9]$ and $W \in [15,\ 40]$.

4.3.4. Sensitivity to Distance Measure

It has been stated in Section 3.2 that the Riemannian distance (RD) in Equation (15) outperformed other metrics that we implemented during the experimentation. In this subsection, we study and compare the performance of the proposed scheme by investigating different distance measures. The experiments were carried out on the Vistex-640 database using the 27D LED descriptors. Comparative results are shown in Table 10 in terms of computational time and retrieval rate. From the table, the Riemannian distance produces the highest ARR (i.e., 94.64%) and the second-lowest computational cost (i.e., 0.0349 s, only greater than the log-Euclidean metric with 0.0316 s) to measure the distance from each query image to 640 candidates. The simplified Mahalanobis and the symmetric Kullback–Leibler distances (see Equations (13) and (14)) consider that each point cloud follows a multivariate normal distribution with mean vector μ and covariance matrix C. They both yield interesting retrieval results (i.e., 90.48% and 91.82%, respectively). On the other hand, by not accounting for the mean feature vectors of those point clouds and considering the topological space of the feature covariance matrices (Riemannian manifolds), some geometric-based metrics such as Wishart-like and the recommended Riemannian could be more relevant. The Wishart-like distance in the second-last row gives a good performance in terms of ARR (92.39%) and requires 0.059 s of computation time per image. However, the recommended Riemannian distance shows better behaviors in both time consumption and retrieval rate. Besides that, two other metrics including the log-Euclidean and the Bartlett distances seem to be not suitable with very low retrieval rates. Therefore, our choice of Riemannian distance for dissimilarity measurement task is confirmed and validated.

Table 10. Sensitivity to distance measures in terms of dissimilarity measurement time and average retrieval rate (ARR). Experiments were conducted on the **Vistex-640** database using the 27D LED descriptors.

Distance Measure	Formula	t_{data} (s)	t_{image} (ms)	ARR (%)
Taking into account mean feature vectors				
Simplified Mahalanobis	$(\mu_1 - \mu_2)^T \left(C_1^{-1} + C_2^{-1}\right)(\mu_1 - \mu_2)$	40.51	63.30	90.48
Symmetric Kullback–Leibler	$\text{trace}\left(C_1 C_2^{-1} + C_2 C_1^{-1}\right) + (\mu_1 - \mu_2)^T \left(C_1^{-1} + C_2^{-1}\right)(\mu_1 - \mu_2)$	47.10	73.59	91.82
Not accounting for mean feature vectors				
Log-euclidean	$\|\|\log(C_1) - \log(C_2)\|\|$	**20.21**	**31.58**	72.65
Bartlett	$\log \frac{\|C_1+C_2\|^2}{\|C_1\|\|C_2\|}$	27.03	42.23	76.51
Wishart-like	$\text{trace}\left(C_1 C_2^{-1} + C_2 C_1^{-1}\right)$	37.79	59.90	92.39
Riemannian	$\sqrt{\sum_{\ell=1}^{d} \log^2 \lambda_\ell}$, where $\lambda_\ell C_1 \chi_\ell - C_2 \chi_\ell = 0, \ell = 1 \ldots d$	22.34	34.91	**94.64**

t_{data}: time for the entire database (640 images); t_{image}: time for each image; $||C||$: the Frobenius norm [42] of matrix C; $|C|$: the determinant of C.

5. Conclusions

A novel texture and color-based image retrieval framework has been proposed in this work by exploiting the local extrema-based features and Riemannian metric. In this article, we have demonstrated the capacity of the local extrema pixels to capture significant information from the

image content in order to encode color and local textural features. The proposed LED descriptor has proved its efficiency and relevance to tackle CBIR tasks. It is easy to implement, feasible to extend or improve, and not very sensitive to parameters. Then, by using such a Riemannian distance to measure the dissimilarity between their feature covariance matrices, the proposed strategy has produced very competitive results in terms of both computational cost and average retrieval rate. It is expected to become one of the state-of-the-art methods in this field.

Future work can improve the performance of LED descriptor by exploiting other features within its generation. One may be interested by a deeper study on LED feature space to propose some other efficient distance metrics. Others may want to study the generalization of the proposed approach by considering the local max and min pixels within a scale-space context [55] where the proposed LED can be regarded as a special case at one scale. For other perspectives, the proposed framework can be applied to other types of images for retrieval and classification tasks within other application fields such as remote sensing imagery, medical imaging, etc. In addition, LED features may be exploited to tackle feature matching tasks and compare them with existing powerful descriptors such as SIFT [39], SURF [40], SYBA [56] or BRIEF [54].

Author Contributions: Minh-Tan Pham proposed the algorithm and conducted the experiments under the supervision of Grégoire Mercier. Lionel Bombrun provided some prior concepts and the image data. Minh-Tan Pham wrote the paper. Grégoire Mercier and Lionel Bombrun revised and gave comments on it.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Alzu'bi, A.; Amira, A.; Ramzan, N. Semantic content-based image retrieval: A comprehensive study. *J. Vis. Commun. Image Represent.* **2015**, *32*, 20–54.
2. Dharani, T.; Aroquiaraj, I.L. A survey on content based image retrieval. In Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME), Salem, India, 21–22 February 2013; pp. 485–490.
3. Veltkamp, R.; Burkhardt, H.; Kriegel, H.P. *State-of-the-Art in Content-Based Image and Video Retrieval*; Springer Science & Business Media: New York, NY, USA, 2013; Volume 22.
4. Do, M.N.; Vetterli, M. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Trans. Image Process.* **2002**, *11*, 146–158.
5. Kokare, M.; Biswas, P.K.; Chatterji, B.N. Texture image retrieval using new rotated complex wavelet filters. *IEEE Trans. Syst. Man Cybern., Part B Cybern.* **2005**, *35*, 1168–1178.
6. Kwitt, R.; Uhl, A. Image similarity measurement by Kullback-Leibler divergences between complex wavelet subband statistics for texture retrieval. In Proceedings of the 15th IEEE International Conference on Image Processing, ICIP 2008, San Diego, CA, USA, 12–15 October 2008; pp. 933–936.
7. Kwitt, R.; Uhl, A. Lightweight probabilistic texture retrieval. *IEEE Trans. Image Process.* **2010**, *19*, 241–253.
8. 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.* **2010**, *19*, 281–289.
9. Lasmar, N.E.; Berthoumieu, Y. Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms. *IEEE Trans. Image Process.* **2014**, *23*, 2246–2261.
10. Li, C.; Huang, Y.; Zhu, L. Color texture image retrieval based on Gaussian copula models of Gabor wavelets. *Pattern Recognit.* **2017**, *64*, 118–129.
11. Verdoolaege, G.; De Backer, S.; Scheunders, P. Multiscale colour texture retrieval using the geodesic distance between multivariate generalized Gaussian models. In Proceedings of the 15th IEEE International Conference on Image Processing, ICIP 2008, San Diego, CA, USA, 12–15 October 2008; pp. 169–172.
12. Kwitt, R.; Meerwald, P.; Uhl, A. Efficient texture image retrieval using copulas in a Bayesian framework. *IEEE Trans. Image Process.* **2011**, *20*, 2063–2077.
13. Ojala, T.; Pietikäinen, M.; Mäenpää, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987.
14. Zhang, B.; Gao, Y.; Zhao, S.; Liu, J. Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor. *IEEE Trans. Image Process.* **2010**, *19*, 533–544.

15. Subrahmanyam, M.; Maheshwari, R.; Balasubramanian, R. Local maximum edge binary patterns: A new descriptor for image retrieval and object tracking. *Signal Process.* **2012**, *92*, 1467–1479.
16. Tan, X.; Triggs, B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. Image Process.* **2010**, *19*, 1635–1650.
17. Murala, S.; Maheshwari, R.; Balasubramanian, R. Local tetra patterns: A new feature descriptor for content-based image retrieval. *IEEE Trans. Image Process.* **2012**, *21*, 2874–2886.
18. Verma, M.; Raman, B. Local tri-directional patterns: A new texture feature descriptor for image retrieval. *Digit. Signal Process.* **2016**, *51*, 62–72.
19. Murala, S.; Wu, Q.J.; Balasubramanian, R.; Maheshwari, R. Joint histogram between color and local extrema patterns for object tracking. In Proceedings of the IS&T/SPIE Electronic Imaging, International Society of Optics and Photonics, Burlingame, CA, USA, 19 March 2013.
20. Jacob, I.J.; Srinivasagan, K.; Jayapriya, K. Local oppugnant color texture pattern for image retrieval system. *Pattern Recognit. Lett.* **2014**, *42*, 72–78.
21. Verma, M.; Raman, B.; Murala, S. Local extrema co-occurrence pattern for color and texture image retrieval. *Neurocomputing* **2015**, *165*, 255–269.
22. Qiu, G. Color image indexing using BTC. *IEEE Trans. Image Process.* **2003**, *12*, 93–101.
23. Gahroudi, M.R.; Sarshar, M.R. Image retrieval based on texture and color method in BTC-VQ compressed domain. In Proceedings of the 9th International Symposium on Signal Processing and Its Applications, ISSPA 2007, Sharjah, United Arab Emirates, 12–15 February 2007; pp. 1–4.
24. Yu, F.X.; Luo, H.; Lu, Z.M. Colour image retrieval using pattern co-occurrence matrices based on BTC and VQ. *Electron. Lett.* **2011**, *47*, 100–101.
25. Guo, J.M.; Prasetyo, H.; Su, H.S. Image indexing using the color and bit pattern feature fusion. *J. Vis. Commun. Image Represent.* **2013**, *24*, 1360–1379.
26. Guo, J.M.; Prasetyo, H. Content-based image retrieval using features extracted from halftoning-based block truncation coding. *IEEE Trans. Image Process.* **2015**, *24*, 1010–1024.
27. 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.* **2015**, *25*, 466–481.
28. Guo, J.M.; Prasetyo, H.; Wang, N.J. Effective Image Retrieval System Using Dot-Diffused Block Truncation Coding Features. *IEEE Trans. Multimedia* **2015**, *17*, 1576–1590.
29. Li, C.; Duan, G.; Zhong, F. Rotation Invariant Texture Retrieval Considering the Scale Dependence of Gabor Wavelet. *IEEE Trans. Image Process.* **2015**, *24*, 2344–2354.
30. Pham, M.T.; Mercier, G.; Michel, J. Pointwise graph-based local texture characterization for very high resolution multispectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 1962–1973.
31. 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.* **2016**, *54*, 3345–3359.
32. 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 Sens.* **2016**, *8*, 368.
33. Pham, M.T.; Mercier, G.; Michel, J. Textural features from wavelets on graphs for very high resolution panchromatic Pléiades image classification. *Revue française de photogrammétrie et de télédétection* **2014**, *208*, 131–136.
34. Pham, M.T.; Mercier, G.; Michel, J. Change detection between SAR images using a pointwise approach and graph theory. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2020–2032.
35. 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 Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; IEEE: Beijing, China, 2016; pp. 1839–1842.
36. Förstner, W.; Moonen, B. A metric for covariance matrices. In *Geodesy-The Challenge of the 3rd Millennium*; Springer: Berlin, Germany, 2003; pp. 299–309.
37. Vision Texture. MIT Vision and Modeling Group. Available online: http://vismod.media.mit.edu/pub/VisTex/ (accessed on 1 October 2017).
38. Mardia, K.V.; Jupp, P.E. *Directional Statistics*; John Wiley and Sons, Ltd.: Chichester, UK, 2000.
39. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.

40. Bay, H.; Tuytelaars, T.; Van Gool, L. SURF: Speeded up robust features. In Proceedings of the 9th European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; pp. 404–417.
41. Tuytelaars, T.; Mikolajczyk, K. Local invariant feature detectors: A survey. *Found. Trends Comput. Graph. Vis.* **2008**, *3*, 177–280.
42. Dryden, I.L.; Koloydenko, A.; Zhou, D. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.* **2009**, *3*, 1102–1123.
43. Frery, A.C.; Nascimento, A.D.; Cintra, R.J. Analytic expressions for stochastic distances between relaxed complex Wishart distributions. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 1213–1226.
44. Kwitt, R.; Meerwald, P. Salzburg Texture Image Database. Available online: http://www.wavelab.at/sources/STex/ (accessed on 1 October 2017).
45. Abdelmounaime, S.; Dong-Chen, H. New Brodatz-Based Image Databases for Grayscale Color and Multiband Texture Analysis. *ISRN Mach. Vis.* **2013**, *2013*, doi:10.1155/2013/876386.
46. USPTex dataset (2012). Scientific Computing Group. Available online: http://fractal.ifsc.usp.br/dataset/USPtex.php (accessed on 1 October 2017).
47. Outex Texture Database. University of Oulu. Available online: http://www.outex.oulu.fi/index.php?page=classification#Outex_TC_00013 (accessed on 1 October 2017).
48. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
49. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, Preprint arXiv:1409.1556.
50. Cusano, C.; Napoletano, P.; Schettini, R. Evaluating color texture descriptors under large variations of controlled lighting conditions. *JOSA A* **2016**, *33*, 17–30.
51. Napoletano, P. Hand-Crafted vs Learned Descriptors for Color Texture Classification. In Proceedings of the International Workshop on Computational Color Imaging, Milan, Italy, 29–31 March 2017; Springer: Berlin, Germany, 2017; pp. 259–271.
52. Subrahmanyam, M.; Wu, Q.J.; Maheshwari, R.; Balasubramanian, R. Modified color motif co-occurrence matrix for image indexing and retrieval. *Comput. Electr. Eng.* **2013**, *39*, 762–774.
53. Leutenegger, S.; Chli, M.; Siegwart, R.Y. BRISK: Binary robust invariant scalable keypoints. In Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 3–13 November 2011; pp. 2548–2555.
54. Calonder, M.; Lepetit, V.; Strecha, C.; Fua, P. BRIEF: Binary robust independent elementary features. In *Computer Vision–ECCV 2010*; Springer: Berlin, Germany, 2010; pp. 778–792.
55. Southam, P.; Harvey, R. Texture classification via morphological scale-space: Tex-Mex features. *J. Electron. Imag.* **2009**, *18*, 043007.
56. Desai, A.; Lee, D.J.; Ventura, D. Matching affine features with the SYBA feature descriptor. In *International Symposium on Visual Computing*; Springer: Berlin, Germany, 2014; pp. 448–457.

© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).