

Article

Local Ternary Cross Structure Pattern: A Color LBP Feature Extraction with Applications in CBIR

Qinghe Feng ¹, Ying Wei ^{1,*}, Yugen Yi ², Qiaohong Hao ³ and Jiangyan Dai ^{4,*}

¹ College of Information Science and Engineering, Northeastern University, Shenyang 110004, China; 1510377@stu.neu.edu.cn

² School of Software, Jiangxi Normal University, Nanchang 330022, China; yiyg510@jxnu.edu.cn

³ College of Intelligence and Computing, Tianjin University, Tianjin 300350, China; qiaohonghao@gmail.com

⁴ School of Computer Engineering, Weifang University, Weifang 261061, China

* Correspondence: weiyi@ise.neu.edu.cn (Y.W.); daijyan@wfu.edu.cn (J.D.); Tel.: +86-024-83688326 (Y.W.)

Received: 8 April 2019; Accepted: 24 May 2019; Published: 29 May 2019



Abstract: With the advent of medical endoscopes, earth observation satellites and personal phones, content-based image retrieval (CBIR) has attracted considerable attention, triggered by its wide applications, e.g., medical image analytics, remote sensing, and person re-identification. However, constructing effective feature extraction is still recognized as a challenging problem. To tackle this problem, we first propose the five-level color quantizer (FLCQ) to acquire a color quantization map (CQM). Secondly, according to the anatomical structure of the human visual system, the color quantization map (CQM) is amalgamated with a local binary pattern (LBP) map to construct a local ternary cross structure pattern (LTCSP). Third, the LTCSP is further converted into the uniform local ternary cross structure pattern (LTCSP_{uni}) and the rotation-invariant local ternary cross structure pattern (LTCSP_{ri}) in order to cut down the computational cost and improve the robustness, respectively. Finally, through quantitative and qualitative evaluations on face, objects, landmark, textural and natural scene datasets, the experimental results illustrate that the proposed descriptors are effective, robust and practical in terms of CBIR application. In addition, the computational complexity is further evaluated to produce an in-depth analysis.

Keywords: content-based image retrieval; feature extraction; five-level color quantizer; local ternary cross structure pattern

1. Introduction

Along with the development of imaging equipment, a larger number of images have been extensively collected from various fields [1–3]. Meanwhile, CBIR technology has gradually become a hot research field, due to its applications in place recognition [4], image classification [5], and remote sensing [6]. Therefore, the problem of extracting effective, robust and practical features has attracted an increasing number of researchers. Thanks to these pioneers' breakthroughs, many approaches [7–26] have been continuously proposed and extended for the task of feature extraction.

In the early days, a family of local binary pattern (LBP)-based methods [7–17] have been sequentially reported for the grayscale-based feature extraction. As a milestone, the LBP definition was initially authored by Ojala et al. [7], in which the referenced pixel and its nearest pixels were encoded as a binary string. Hereafter, Zhang et al. [8] extended the LBP to the local derivative pattern (LDP) descriptor for refining the magnitude difference. Subsequently, Guo et al. [9] designed a variant of the LBP named the completed LBP, and it is used for improving the robustness to rotation. After that, the LBP variance was developed by Guo et al. [10] for addressing the drawback of the global information loss. Further, Tan et al. [11] introduced an improvement of the LBP named the

local ternary pattern (LTP), and it was integrated with the kernel principal component analysis to improve the robustness to illumination. After, Murala et al. [12] modified the LTP to the local tetra pattern (LTrP), in which the referenced pixel and its surrounding pixels were computed by comparing the vertical and horizontal direction differences. Afterwards, Subrahmanyam et al. [13] converted the LTP to the local maximum edge binary patterns (LMEBP), and the LMEBP was combined with the Gabor transform for CBIR and object tracking applications. In order to extend the LBP to the dynamic-texture application, Zhao et al. [14] studied the LBP histogram Fourier (LBP-HF) for video sequence recognition. Further, in order to improve the robustness to noise, Ren et al. [15] proposed the noise-resistant local binary pattern (NRLBP) to extract the local structure information. In order to retrieve the magnetic and resonance and computer tomography images, Subrahmanyam et al. [16] studied the local ternary co-occurrence patterns (LTCoP) encoding. Motivated by the concatenation strategy, Verma et al. [17] used the LBP feature map and the local neighborhood difference pattern (LNDP) to integrate the binary information into the local intensity difference for the natural scene and texture retrieval applications. Nevertheless, since all the above approaches are limited to the grayscale-based feature extraction, the color information unavoidably is lost.

In recent years, a family of color LBP descriptors [18–28] have been continuously explored for the color-based feature extraction. Among them, inspired by inter- and intra- channel encoding mechanisms, Mäenpää et al. [18] constructed the opponent color local binary patterns (OCLBP) for color textural classification. After that, Bianconi et al. [19] proposed an extension of OCLBP named the improved opponent color local binary patterns (IOCLBP), in which point-to-average thresholding replaced point-to-point thresholding. Further, the grayscale-based LTrP was extended to the local oppugnant color texture pattern (LOCTP) by Jeena Jacob et al. [20], and the LOCTP was extracted from the RGB, YCbCr, and HSV color models respectively. Inspired by the pair-based strategy, Qi et al. [21] presented the pairwise rotation-invariant co-occurrence LBP (PRICoLBP) by incorporating the R, G and B components. Similarly, Hao et al. [22] proposed the pairwise cross pattern (PCP), in which the color and LBP information were combined in pairwise and cross manners. With the help of the encoding-decoding technology, Dubey et al. [23] designed the multi-channel adder LBP (maLBP) and the multi-channel decoder LBP (mdLBP) to extract the LBP feature maps from the R, G, and B channels. In 2017, inspired by the vector quantization (VQ) strategy, Guo et al. [24] proposed the max and min color quantizer in order to extract the color information feature (CIF) in the RGB color model, and the CIF feature was combined with the LBP-based feature for the image classification and retrieval applications. In 2018, Somasekar et al. [25] integrated the Fuzzy C-Mean color clustering in the RGB color model with the LBP feature for the side-scan-sonar image enhancement. In the same year, the CIELAB color model was quantized by Singh et al. [26] to extract the color histogram (CH), and the CH feature was linearly combined with the orthogonal combination of LBP (OC-LBP) for the color image retrieval applications. More recently, Feng et al. [27] introduced the local parallel cross pattern (LPCP) which integrated the color and LBP information in parallel and cross manners. In order to capture the cross-channel information, Agarwal et al. [28] studied the multi-channel local ternary pattern (MCLTP) to extract the correlation from the H-V, S-V and V-V channels in a cross manner.

In this paper, we present the main following contributions:

1. We construct a five-level color quantizer, and it is applied to quantize the a^* and b^* components for the color quantization map extraction.
2. We integrate the color quantization map into the LBP feature map to extract a local ternary cross structure pattern (LTCSP).
3. We further extend the local ternary cross structure pattern to the uniform local ternary cross structure pattern and the rotation-invariant local ternary cross structure pattern for reducing the computational cost and improving the robustness.
4. We benchmark a series of experiments on face, landmark, object and textural datasets, and extensive experimental results demonstrate the effectiveness, robustness, and practicability of the proposed descriptor.

The rest of this paper is organized as follows. In Section 2, the local binary pattern definition and the color prior knowledge are briefly reviewed. Section 3 concretely details the feature extraction. Similarity measure and retrieval system are introduced in Section 4. Section 5 presents the experiments and discussion. Section 6 concludes this paper, and points out the possible future directions.

2. Related Work

2.1. Local Binary Pattern Definition

Firstly, the local binary pattern (LBP) definition was initially authored by Ojala et al. [7] for the gray-scale feature extraction. Given a referenced pixel $P(s, t)$ and its surrounding pixels $P_k(s, t)$, the LBP encoding at $P(s, t)$ is formulated as follows:

$$LBP_{r,n}(s, t) = \sum_{k=0}^{n-1} \mu(P(s, t) - P_k(s, t)) \times 2^k, \tag{1}$$

$$\mu(m) = \begin{cases} 1, & m \geq 0 \\ 0, & m < 0 \end{cases} \tag{2}$$

where r is the radius of a circle, and n is the number of surrounding pixels in the circle with radius r .

Then, depending on the realistic application, over 90% of image micro-structures are encoded by only 23% of the LBP patterns. To cut down the computational cost, the LBP is simplified to the “uniform local binary pattern”, which is formulated as follows:

$$LBP_{r,n}^{uni}(s, t) = U\left\{\sum_{k=1}^n \left| \mu(P(s, t) - P_k(s, t)) - \mu(P(s, t) - P_{k-1}(s, t)) \right|\right\}, \tag{3}$$

where $U\{\cdot\}$ represents a measure operator, and $U\{\cdot\} \leq 2$. $P_0(s, t)$ is equivalent to $P_n(s, t)$.

Finally, to improve robustness, the LBP further is converted into the rotation invariant local binary pattern $LBP_{r,n}^{ri}(s, t)$, which is expressed as follows [21]:

$$LBP_{r,n}^{ri}(s, t) = \min\{ROR(LBP_{r,n}(s, t), k) \mid k \in 0, 1, \dots, n - 1\}, \tag{4}$$

where $ROR(LBP_{r,n}(s, t), k)$ is a circular bit-wise right shift for x times on n -bit number $LBP_{r,n}(s, t)$.

Referring to [21,29], r and n are defined as 1 and 8 respectively. In the following, the LBP feature map, the uniform LBP feature map, and the rotation invariant LBP feature map are abbreviated as LBP , LBP_{uni} , and LBP_{ri} respectively.

2.2. Color Quantization Scheme

Currently, the equal-interval color quantizer (EICQ) is considered as the most common used quantizer [26,30,31]. Among them, Singh et al. [26] proposed the color histogram (CH)-based scheme in which the number of quantization blocks for L^* , a^* and b^* components was fixed at 3, 2 and 2. Similarly, Reta et al. [30] proposed the Lab color coherence vector (Lab-CVV) to quantize the L^* , a^* and b^* components to 5, 4 and 4 respectively. Considering the HSV color model, Wan and Kuo [31] developed the multiresolution histogram representation (MHR)-based quantizer in which a variety of combinations of 2, 4, 8 and 16 blocks were designed for the H, S and V components quantization. Motive by the flexible strategy, Liu et al. [32] proposed the flexible micro structure descriptor histogram (MSD)-based quantizer to quantize the L^* , a^* and b^* component to 20, 3 and 3 blocks respectively. Similarly to the MSD-based quantizer, Liu et al. [33] presented the adaptable color difference histogram (CDH)-based quantizer in which the optimal number of quantization in the L^* component was set to 10, and the number of quantization in the a^* and b^* components was fixed as 3. With help of the analysis of color distribution, Wan et al. [34] proposed standard vector quantization, a method of partitioning the

vector space by minimizing the mean squared error. Motivated by unsupervised clustering techniques, Duda and Hart [35] designed the criterion-based schemes and the heuristic schemes. Inspired by the color clusters distribution, Xia and Kuo [36] proposed the octree with pruning color quantization which calculated the color information of each image respectively.

2.3. Color Prior Knowledge in the CIELAB Color Model

The CIELAB color model consists of three components, namely the white-black component of L^* (ranging from 0 to 100), the yellow-blue component of a^* (ranging from -128 to $+127$), and the red-green component of b^* (ranging from -128 to $+127$) [37]. The CIELAB color model is not only an excellent splitter between color (represented by the a^* and b^* components) and intensity (represented by the L^* component), but also is a perceptually uniform color space; that is to say there are the same amount of numerical changes between the CIELAB model and the visual perception in human color vision [38]. On the basis of the CIELAB color model, a color prior knowledge was originally introduced by Feng et al. [5], in which the frequency of pixels in the a^* and b^* components were explored and analyzed. Firstly, in Figure 1a,b the frequency of pixels is mostly distributed in the center of the a^* and b^* components on the Caltech-256 [39] dataset. Secondly, to verify the consistency, thousands of image datasets are calculated, and extensive experimental results demonstrate the consistency of this prior. Obviously, it can be summarized that most pixels focus on the middle third of the range. Thirdly, to verify the stability, a series of additional experiments are performed, which illustrate that the prior knowledge is stable even if the image dataset is changed. For instance, the probability distribution of the Caltech-256 dataset (see Figure 1a,b) is extremely approximate to 10% of the Caltech-256 dataset (see Figure 1c,d).

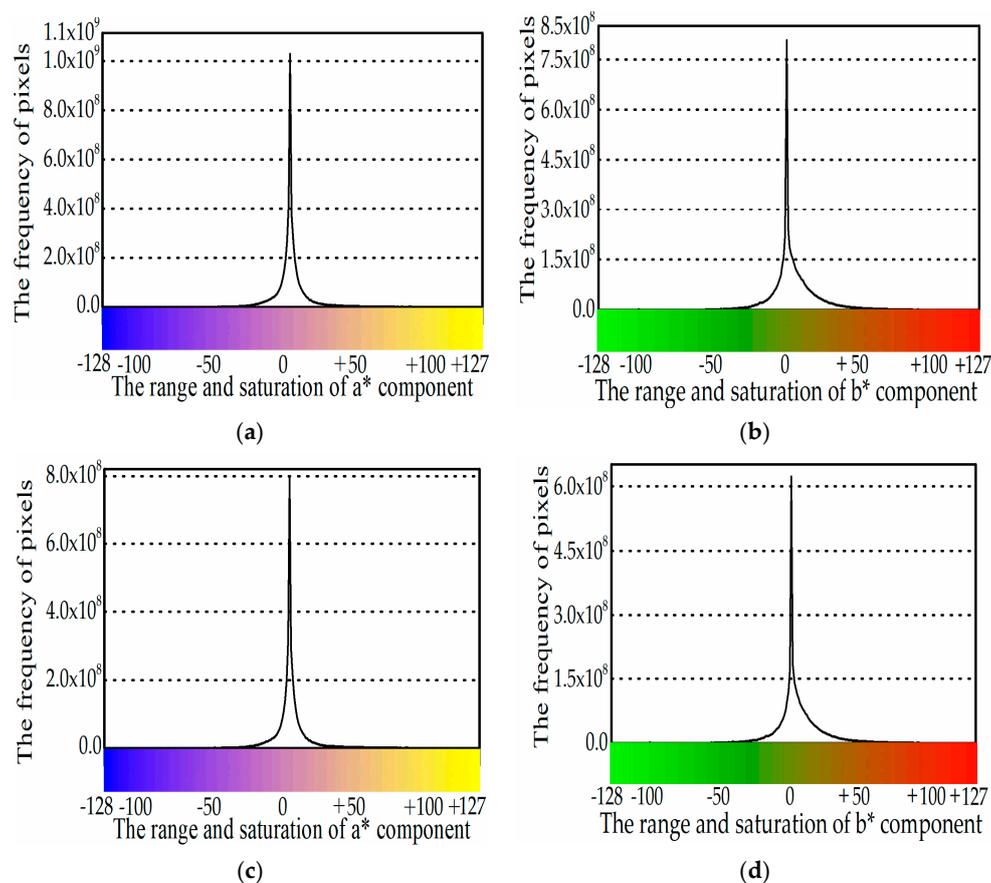


Figure 1. The frequency of pixels on: (a,b) Caltech-256 and (c,d) 10% of Caltech-256.

3. Feature Extraction

3.1. Five-Level Color Quantizer

Inspired by the color prior knowledge in the CIELAB color model, a novel five-level color quantizer is designed for the color quantization map (CQM) extraction. For convenience, the original $[-128, +127]$ is mapped as 2^8 . In our scheme, 2^8 is first subdivided into four blocks at Level 1, in which two blocks $2^8/3$ are on both sides, and two refined blocks $2^7/3$ are in the middle because most pixels focus on the middle third of the range. The corresponding indices are sequentially flagged as 0, 1, 2, and 3 at Level 1. Then, two refined blocks $2^7/3$ are subdivided into two refined $2^7/3^2$ in the middle and two $2^8/3^2$ on both sides from Level 1 to 2. In this manner, the pixels in the middle range can be further refined. The remaining blocks are duplicated from Level 1 to 2. Finally, two operators of “Duplicate” and “Subdivide” are sequentially repeated until the two middle blocks $2^7/3^5$ at Level 5. We combine Level 1 to 5 to construct the five-level color quantizer. For clarity, the process is displayed in Figure 2, in which each level contains a group of blocks and indices. Mathematically, the quantization levels in the a^* and b^* components are flagged as A_{a^*} and A_{b^*} , where $A_{a^*} \in \{1, 2, \dots, 5\}$ and $A_{b^*} \in \{1, 2, \dots, 5\}$. Meanwhile, the corresponding indices in the a^* and b^* components are flagged as \hat{A}_{a^*} and \hat{A}_{b^*} , where $\hat{A}_{a^*} \in \{0, 1, \dots, 2(A_{a^*} + 1) - 1\}$ and $\hat{A}_{b^*} \in \{0, 1, \dots, 2(A_{b^*} + 1) - 1\}$ respectively.

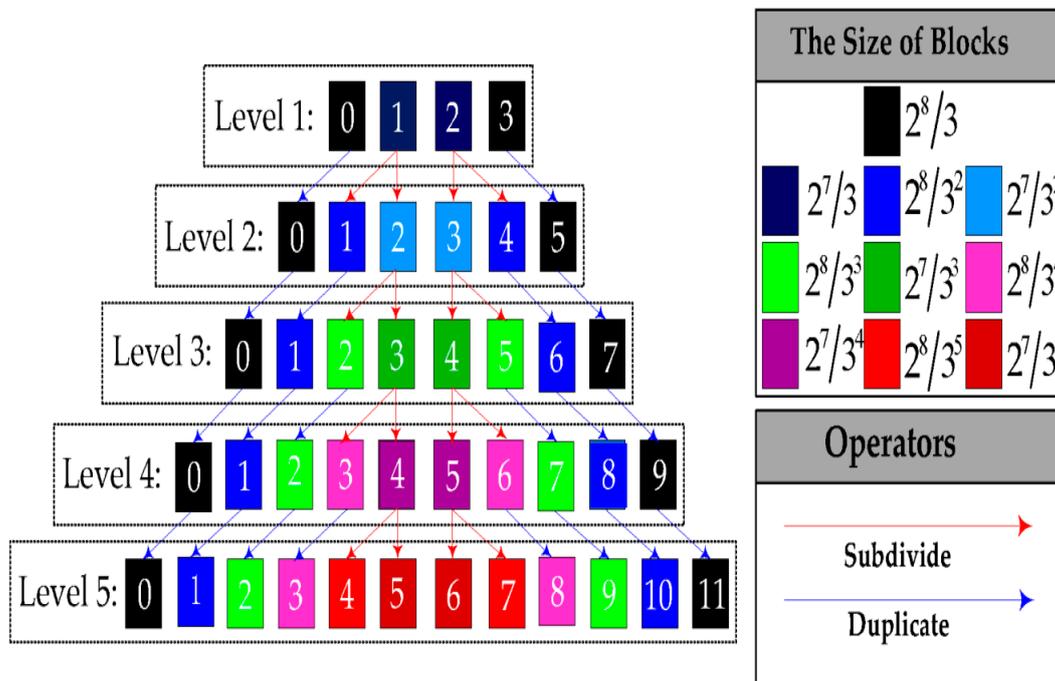


Figure 2. The details of the five-level color quantizer.

Moreover, in terms of the humans’ eye intensity perception mechanism [40], the original scope $[0, +100]$ in the L^* component is quantized into 3 blocks, e.g., $[0, +25]$, $[+26, +75]$ and $[+76, +100]$. Similarly, the quantization level in the L^* component is flagged as A_{L^*} , where $A_{L^*} = 1$ and the index is defined as $\hat{A}_{L^*} \in \{0, 1, 2\}$.

An example of the proposed color quantization scheme is shown in Figure 3. From Figure 3a, the values in the CIELAB color model are set to $L^* = +87$ in the L^* component, $a^* = +84$ in the a^* component, and $b^* = -28$ in the b^* component. Correspondingly, the quantization levels are set to $A_{L^*} = 1$, $A_{a^*} = 2$ and $A_{b^*} = 2$ respectively. From Figure 3b, according to $A_{L^*} = 1$, the L^* component is firstly quantized to 3 blocks, e.g., $[0, +25]$, $[+26, +75]$ and $[+76, +100]$, and then the index of $L^* = +87$ can be encoded as $\hat{A}_{L^*} = 2$. From Figure 3c,d, considering $A_{a^*} = 2$ and $A_{b^*} = 2$ in the FLCQ quantizer, the a^* and b^* components are firstly quantized to six blocks, e.g., $2^8/3$, $2^8/3^2$, $2^7/3^2$, $2^7/3^2$, $2^8/3^2$ and

$2^8/3$, and then the indices of $a^* = +84$ and $b^* = -82$ is encoded as $\hat{A}_{a^*} = 5$ and $\hat{A}_{b^*} = 1$ respectively. We combine the indices of \hat{A}_{L^*} , \hat{A}_{a^*} , and \hat{A}_{b^*} to acquire the color quantization map (CQM), and the index of CQM is flagged as E , $E \in \{0, 1, \dots, 3 \times 2(A_{a^*} + 1) \times 2(A_{b^*} + 1) - 1\}$.

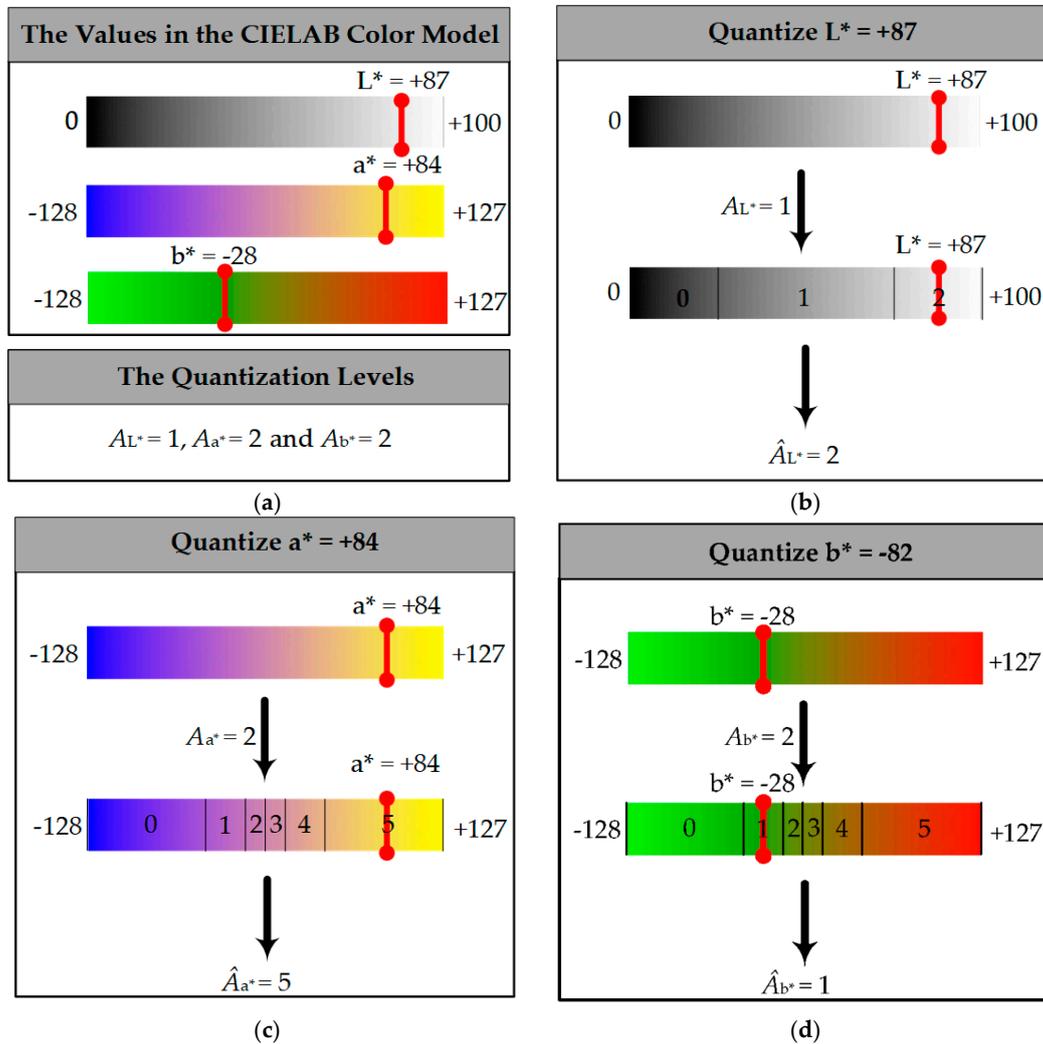


Figure 3. An example of the proposed quantization scheme: (a) The values and its quantization levels in the CIELAB color model; (b) The quantization in the L^* component; (c) The quantization in the a^* component; and (d) The quantization in the b^* component.

3.2. Human Visual System

As documented in Gray’s Anatomy, the anatomical structure of the human visual system consists of the eyeball, optic nerve, lateral geniculate nucleus of thalamus, optic radiation and visual cortex [41]. Remarkably, it should be noted that the optic chiasm is a critical anatomical structure, in which a part of the visual cues between the left and right cerebral hemispheres are exchanged. According to the human visual system, the left and right eye balls first extract the low-level visual cues. Second, the extracted low-level visual cues are encoded and transmitted to the left and right optic nerves. Third, a part of the encoding visual information is crossed at the optic chiasm. Fourth, the crossed visual information is reconstructed at the left and right lateral geniculate nucleus of thalamus. Fifth, the reconstructed visual information is radiated at the left and right optic radiations. Finally, the radiated visual information is re-aggregated at the left and right visual cortices to form the high-level semantics perception. For more details, please refer to [41].

3.3. Local Ternary Cross Structure Pattern

According to the anatomical structure of the human visual system, a novel local ternary cross structure pattern (LTCSP) is proposed to integrate the color information into the LBP information as a whole. For an original map $M(i, j)$, the reference point is flagged as $M(i_0, j_0)$ and the eight nearest points are flagged as $M(i_p, j_p)$, where $p \in \{1, 2, \dots, 8\}$.

Firstly, the LBP feature map is computed as $LBP(i, j)$ in Section 2.1, and the color quantization map is computed as $CQM(i, j)$ in Section 3.1. Secondly, with the help of a thresholded polynomial selectivity indicator $\text{sign}(\cdot)$, the color quantization map and the LBP feature map are encoded into the color ternary map and the LBP ternary map respectively. Mathematically, the thresholded polynomial selectivity indicator $\text{sign}(\cdot)$ is defined as follows:

$$\text{sign}(Y(i_0, j_0), Y(i_p, j_p)) = \begin{cases} 1 & \text{if } Y(i_0, j_0) > Y(i_p, j_p) \\ 0 & \text{if } Y(i_0, j_0) = Y(i_p, j_p) \\ -1 & \text{if } Y(i_0, j_0) < Y(i_p, j_p) \end{cases}, \quad (5)$$

where $p \in \{1, 2, \dots, 8\}$, and $Y(i, j)$ represents a feature map. For the LBP feature map, $Y(i, j)$ is considered as $LBP(i, j)$; for the color quantization map, $Y(i, j)$ is considered as $CQM(i, j)$. Thirdly, motivated by the optic chiasm, in which a part of the visual cues between the left and right cerebral hemispheres are exchanged, the eight nearest points in the color and LBP ternary maps are correspondingly crossed to extract the color and LBP cross maps. Fourthly, with the help of a counter $\mathcal{G}\{\cdot\}$ that computes the occurring numbers of ternary in the eight nearest points of the color and LBP cross maps, the maximum numbers are retained to construct the color and LBP structure maps respectively. Inspired by the max-pooling strategy, when there exist two maximum occurring numbers, the structure with the larger value is retained. Finally, the retained points in the color and LBP structure maps are computed as the feature vectors, and the values of the reference points are correspondingly calculated as the indices of the feature vectors. Mathematically, we define the LTCSP as follows:

$$LTCSP_{LBP}(LBP(i_0, j_0)) = \max_{p \in \{1, 2, \dots, 8\}} \mathcal{G}\{\text{sign}(CQM(i_0, j_0), CQM(i_p, j_p))\}, \quad (6)$$

$$LTCSP_{CQM}(CQM(i_0, j_0)) = \max_{p \in \{1, 2, \dots, 8\}} \mathcal{G}\{\text{sign}(LBP(i_0, j_0), LBP(i_p, j_p))\}, \quad (7)$$

For clarity, the schematic diagram of the LTCSP is illustrated in Figure 4, where the LTCSP is computed as $[LTCSP_{CQM}(250) = 6, LTCSP_{LBP}(235) = 4]$. Experimentally, the feature dimensionality of $LTCSP_{CQM}$ and $LTCSP_{LBP}$ are $3 \times 2(A_{a^*} + 1) \times 2(A_{b^*} + 1)$ and 256 respectively.

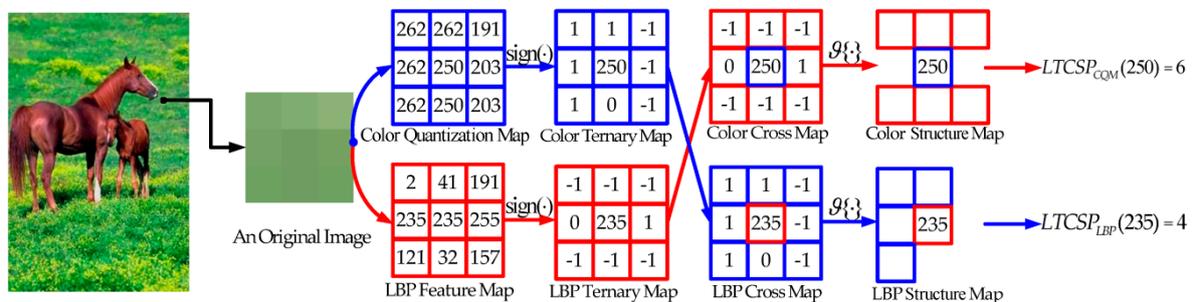


Figure 4. Schematic diagram of the local ternary cross structure pattern (LTCSP).

In order to select the optimal color quantization levels (A_{a^*}, A_{b^*}), we need to compare the average precision rates (APR) of 25 possible combinations according to different image datasets in the offline stage. Mathematically, given an image dataset D , the maximization APR is defined as follows:

$$\operatorname{argmax}_{A_{a^*}, A_{b^*}} \operatorname{APR}(D|A_{a^*}, A_{b^*}), \quad (8)$$

where $\operatorname{APR}(D|A_{a^*}, A_{b^*})$ represents the APR score, where $A_{a^*} \in \{1, 2, \dots, 5\}$ and $A_{b^*} \in \{1, 2, \dots, 5\}$. In Section 5.4., we provide the optimal color quantization levels (A_{a^*}, A_{b^*}).

Moreover, in order to reduce the computational cost, the LBP feature map $LBP(i, j)$ in $LTCSP$ can be replaced by the LBP_{uni} feature map $LBP_{uni}(i, j)$ to construct the uniform local ternary cross structure pattern ($LTCSP_{uni}$). Similarly, in order to improve the robustness to rotation, the LBP feature map $LBP(i, j)$ in $LTCSP$ can be replaced by the LBP_{ri} feature map $LBP_{ri}(i, j)$ to construct the rotation-invariant local ternary cross structure pattern ($LTCSP_{ri}$).

4. Similarity Measure and Retrieval System

4.1. Similarity Measure

Given a query image provided by a user, the query and dataset images are first encoded as the query and dataset feature vectors respectively. Then, the similarity measure between the query and dataset feature vectors is performed. Finally, based on the sorting results of the similarity measure, the top similar images are returned to the user. Referring to [6,22,33,42], extended Canberra distance (ECD) is utilized in this paper, and the ECD is defined as follows:

$$ECD(f^d, f^q) = \sum_{\tau=1}^{\delta} \frac{|f_{\tau}^d - f_{\tau}^q|}{|f_{\tau}^d + v_d| + |f_{\tau}^q + v_q|}, \quad (9)$$

where $ECD(\cdot)$ denotes the result of extended Canberra distance. f_{τ}^d and f_{τ}^q are the feature vectors of the query and database images, and τ represents the dimensionality of the feature vector. v_d and v_q are defined as $\sum_{\tau=1}^{\delta} f_{\tau}^d / \delta$ and $\sum_{\tau=1}^{\delta} f_{\tau}^q / \delta$ respectively.

4.2. Retrieval System

User, query image, dataset images, feature extraction, query and dataset feature vectors, similarity measure, and returned images are totally combined to construct the proposed retrieval system. The schematic diagram of the proposed retrieval system is shown in Figure 5. From Figure 5, the proposed retrieval system can be divided into the offline stage and the online stage. In the offline stage, based on the optimal quantization levels (A_{a^*}, A_{b^*}) in Equation (8), all dataset images are sent to the feature extraction block to extract the dataset feature vectors only once. In the online stage, the user firstly inputs the query image. Secondly, the query image is sent to the feature extraction block to extract the query feature vector. Thirdly, the similarity measure block is performed between the query and dataset feature vectors. Finally, according to the similarity measure scores, the top-n similar images are considered as returned images to the user. If there exist several datasets, the query image would be adaptively encoded into different query feature vectors.

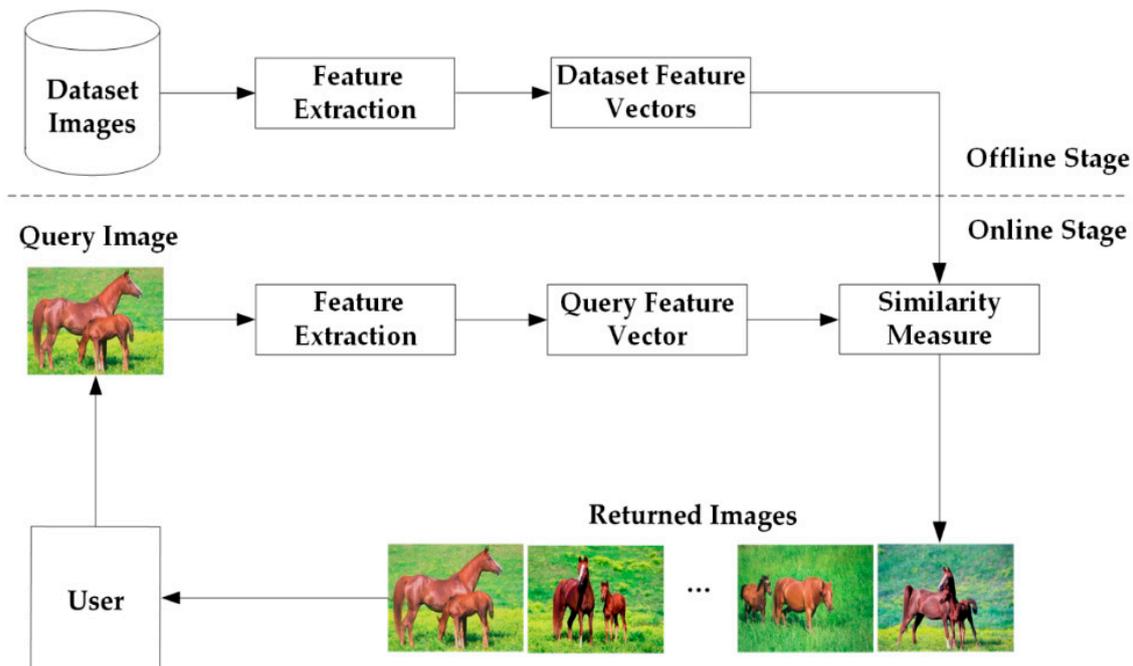


Figure 5. Schematic diagram of the proposed retrieval system.

5. Experiments and Discussion

5.1. Evaluation Criteria

The precision rate, recall rate, average precision rate, and average recall rate are commonly adopted as the most used evaluation criteria. Among them, the precision and recall rates are used to evaluate the retrieval performance of a single query image, and they are defined as follows [3,5]:

$$Precision = \frac{\text{Number of similar images returned}}{\text{Total number of images returned}}, \quad (10)$$

$$Recall = \frac{\text{Number of similar images returned}}{\text{Number of all relevant images}}, \quad (11)$$

Further, the average precision rate (APR) and average recall rate (ARR) are applied to evaluate the retrieval performance of the total number of query images, and they are computed as follows:

$$APR = \frac{\sum_{t=1}^T Precision(t)}{T}, \quad (12)$$

$$ARR = \frac{\sum_{t=1}^T Recall(t)}{T}, \quad (13)$$

where t represents the t -th query image, and T is the total number of query images.

5.2. Image Datasets

Six benchmark datasets, consisting of one face image dataset (Face-95 [43]), one object image dataset (ETHZ [44]), one landmark image dataset (ZuBuD [45]), two color textural image datasets (KTH-2a [46] and VisTex [47]), and one natural scene image dataset (Corel-100 [48]), are summarized in Table 1 to evaluate the effectiveness, robustness, and practicability of the proposed descriptors.

Table 1. Summary of six image datasets.

Number	Name	Image Size	Class	Images in Each Class	Images Total	Format
1	Face-95	180 × 200	72	20	1440	JPG
2	ETHZ	320 × 240	53	5	265	BNG
3	ZuBuD	640 × 480	201	5	1005	JPG
4	VisTex	128 × 128	40	16	640	PPM
5	KTH-2a	200 × 200	11	396/432	4608	BNG
6	Corel-1000	384 × 256 or 256 × 384	10	100	1000	JPG

The Face-95 (No. 1) is a face image dataset, and it was captured by a S-VHS camcorder. The Face-95 consists of 1440 images in 72 persons. Each person contains 20 images in JPG format with size of 180 × 200. Note that all images are with face expressions, illumination changes, head scales and head turns. In Figure 6a, there are some samples from Face-95.

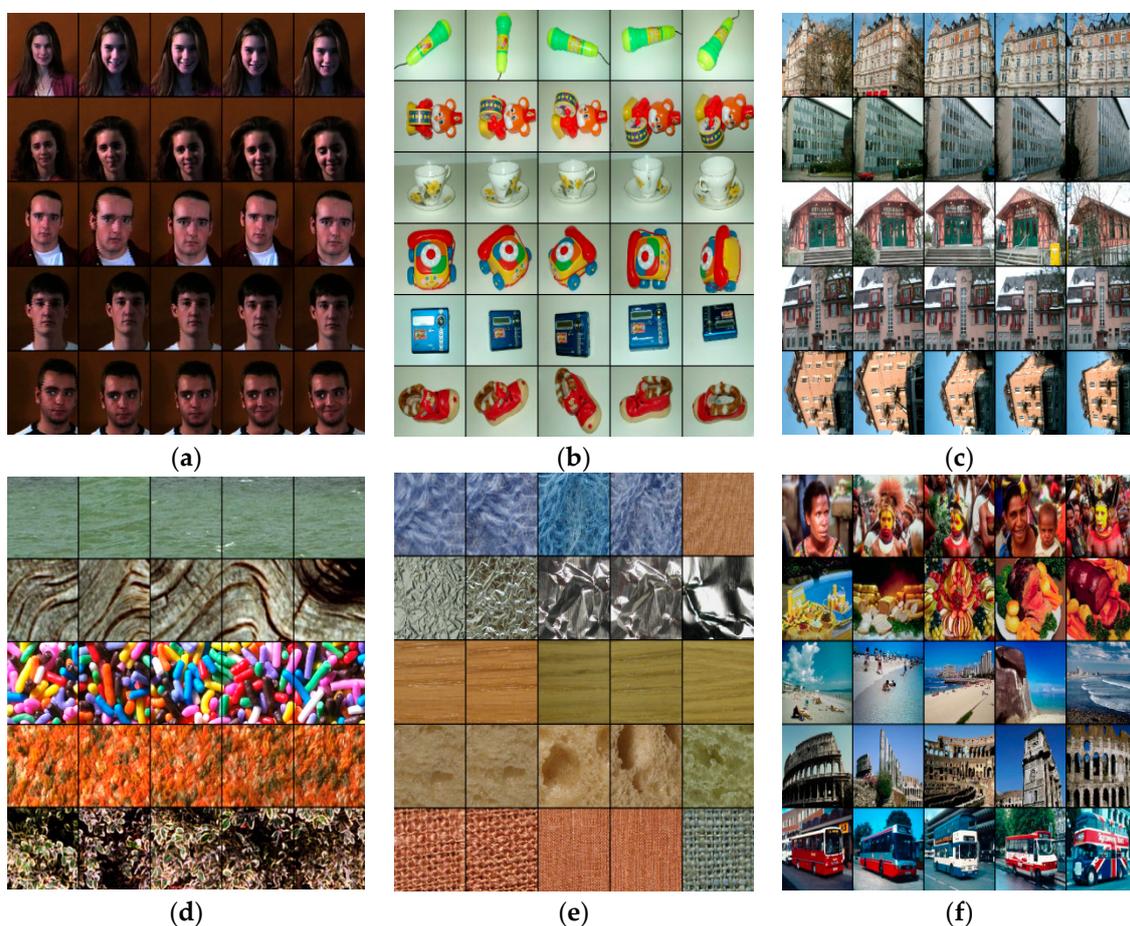


Figure 6. Some samples from six datasets: (a) Face-95; (b) ETHZ; (c) ZuBuD; (d) VisTex; (e) KTH-2a; and (f) Corel-1000.

The ETHZ (No. 2) is an object image dataset, and it was taken with a Sony XC-77P camera. The ETHZ has 265 images in 53 natural objects. For each object, there are 72 images in BNG format with size of 320 × 240. Specially, each object is rotated by an arbitrary degree, so the ETHZ also can be used to evaluate the robustness of rotation. In Figure 6b, some samples from ETHZ are displayed.

The ZuBuD (No. 3) is a landmark image dataset, and it was produced by the Panasonic-NV-MX300 and Pentax-Opti430 cameras. The ZuBuD has 1005 images in 201 landmarks. Each landmark contains

five images in JPG format with size of 640×480 . Note that not only some occlusions (e.g., tree and car) are purposely included in images, but also all images are captured under different views points, weather conditions and seasons. Thus, the ZuBuD can be used for evaluating the effectiveness under the complex environment. In Figure 6c, some samples from ZuBuD are showned.

The VisTex (No. 4) and KTH-2a (No. 5) are two color textural image datasets. On one hand, the VisTex (No. 4) was acquired by collecting images from videos and photographs. The VisTex is having 640 images in 40 categories (e.g., fabric, flower, matel, terrain, bark, sand, leave, stone, and so on). Each category includes 16 images in PPM format with size of 128×128 . In Figure 6d, there are some samples from VisTex. On the other hand, the KTH-2a (No. 5) was taken with an Olympus C-3030ZOOM camera, and it consists of 4608 images in 11 categories (e.g., cotton, brown bread, wood, wool, white bread, corduroy, linen, cracker, cork, aluminium foil and lettuce leaf). Each category contains 396/432 images in BNG format with size of 200×200 . Specially, the KTH-2a also can be applied to evaluate whether the proposed descriptors are robust against rotation, scaling and illumination. In Figure 6e, there are some samples from KTH-2a.

The Corel-1000 (No. 6) is a natural scene image dataset, and it was collected by the SIMPLIcity system. The Corel-1000 has 1000 images in 10 natural scenes. Each natural scene contains 100 images in JPG format with size of 384×256 or 256×384 . In Figure 6f, some samples from Corel-1000 are presented.

5.3. Experimental Details

All experiments were performed on a personal computer with an Intel Core i7-7700k CPU@4.20 Ghz, a 16 GB DDR4 RAM@2400 MHz, and a 6 GB NVIDIA GTX1070ti. All six benchmark datasets in our experiments can be freely downloaded from [43–48]. Note that all images need to be transformed from the RGB to the grayscale space and the CIELAB color model for extracting the LBP feature map and the color quantization map respectively. For details, these transformations are referred to [38]. In addition, the open source implementation of the LBP feature map [9,10] can be downloaded from [http://www.ee.oulu.fi/~gyzhao/](http://www.ee.oulu.fi/~gyzha/). Referring to [6,22,23,27,40,49,50], all images in the dataset are considered as the query images to guarantee the accurateness and reproducibility. Referring to [6,22,23,27,40,49,50], the number of returned images on VisTex, KTH-2a, Face-95 and Corel-1000 is defined as 10, and the number of returned images on ETHZ and ZuBuD is defined as five because there are only five images in each class.

5.4. Evaluation of Color Quantization Levels

Table 2 reports the APR rates of LTCSP, LTCSP_{uni} and LTCSP_{ri} with the optimal levels (A_{a^*} , A_{b^*}) over six datasets. For clarity, the best APR values with the optimal level (A_{a^*} , A_{b^*}) are documented in bold. Firstly, in Table 2, it can easily draw the conclusions that a single quantization level cannot satisfy the need of all image datasets. For example, LTCSP yields the highest APR rate on Face-95 when ($A_{a^*} = 5$, $A_{b^*} = 5$); (2) LTCSP acquires the top APR rate on ETHZ when ($A_{a^*} = 3$, $A_{b^*} = 4$); and (3) LTCSP products the highest APR rate on Corel-1000 when ($A_{a^*} = 3$, $A_{b^*} = 3$). Secondly, it can be noted that a relatively simple level also products the highest APR rate. For instance, when ($A_{a^*} = 3$, $A_{b^*} = 2$), LTCSP_{uni} acquires the top APR rate on VisTex, and LTCSP_{uni} and LTCSP_{ri} product the highest APR rate on Corel-1000. Depending upon all proven observations, it is necessary to optimally choose (A_{a^*} , A_{b^*}) from the five-level color quantizer. Additionally, we provide the APR values of about 25 possible combinations on six datasets in the supplementary file. In the following experiments, the optimal levels of A_{a^*} and A_{b^*} are adaptively adopted in the proposed descriptors according to different datasets.

Table 2. The best APR (%) values of LTCSP, LTCSP_{uni} and LTCSP_{ri} with the optimal levels (A_{a^*} , A_{b^*}) over six datasets.

Method	Performance	Data Set					
		Face-95	ETHZ	ZuBuD	VisTex	KTH-2a	Corel-1000
LTCSP	(A_{a^*} , A_{b^*})	(5, 5)	(3, 4)	(5, 5)	(4, 2)	(5, 3)	(3, 3)
	APR (%)	94.31	90.57	85.63	98.56	98.96	83.94
LTCSP _{uni}	(A_{a^*} , A_{b^*})	(5, 5)	(5, 3)	(5, 4)	(3, 2)	(4, 3)	(3, 2)
	APR (%)	97.19	94.04	85.97	97.81	99.15	82.83
LTCSP _{ri}	(A_{a^*} , A_{b^*})	(5, 5)	(5, 3)	(5, 5)	(3, 3)	(4, 5)	(3, 2)
	APR (%)	97.39	94.72	86.11	97.53	99.19	82.33

5.5. Comparison with Other Hierarchical Quantization Schemes

Referring to [51–53], the mean square error (MSE) values obtained by the hierarchical quantization schemes of CH [26], Lab-CVV [30], CDH [32], MSD [33], LTCSP, LTCSP_{uni}, and LTCSP_{ri} on six datasets are reported in Table 3. First, it can be clearly observed that the MSE values by the proposed LTCSP, LTCSP_{uni}, and LTCSP_{ri} methods are obviously lower than all other descriptors on six datasets. These phenomena illustrate the hierarchical quantization schemes of the proposed LTCSP, LTCSP_{uni}, and LTCSP_{ri} are superior to all other descriptors. Secondly, it can be concluded that the mean square errors by the proposed LTCSP, LTCSP_{uni}, and LTCSP_{ri} methods are extremely close to each other on six datasets. These results demonstrate that the hierarchical quantization schemes of the proposed descriptors are stable and consistent between six datasets. Thirdly, it can be summarized that there exist obvious differences among different datasets. These phenomena demonstrate it is necessary to adaptively select different color quantization levels according to different datasets. In addition, we provide the MSE values of 25 color quantization levels (A_{a^*} , A_{b^*}) between the FLCQ and EICQ quantizers in the CIELAB color model on all six datasets in Appendix A.

Table 3. The mean square errors obtained by the hierarchical quantization schemes of different descriptors on six datasets.

Method	Data Set					
	Face-95	ETHZ	ZuBuD	VisTex	KTH-2a	Corel-1000
CH	6085.91	6714.11	7483.35	6716.29	5858.41	6386.27
Lab-CVV	1024.93	1430.97	1622.65	1358.55	1090.41	1276.06
CDH	391.83	205.99	90.02	255.97	487.37	370.97
MSD	385.29	200.00	83.31	249.44	481.69	365.06
LTCSP	48.28	64.15	17.75	57.97	95.03	92.12
LTCSP _{uni}	48.28	64.22	17.78	58.77	95.07	102.21
LTCSP _{ri}	48.28	64.22	17.75	58.77	94.88	102.21

5.6. Comparison with LBP-Based Methods

Table 4 details the evaluations of the APR and ARR rates resulting from the LBP, LBP_{uni}, and LBP_{ri} methods and the proposed LTCSP, LTCSP_{uni}, and LTCSP_{ri} methods. The best {APR, ARR} values are highlighted in bold. First, it can be clearly observed that the proposed LTCSP, LTCSP_{uni}, and LTCSP_{ri} methods achieve remarkable enhancements as compared to the LBP, LBP_{uni}, and LBP_{ri} methods on all six datasets. The foremost reason is that the proposed methods are beneficial to integrate the color information and the LBP information. Secondly, it is noted that LTCSP achieves the highest APR rate on VisTex and Corel-1000, LTCSP_{ri} generates the top APR rates on Face-95, ETHZ, ZuBuD, and KTH-2a datasets respectively. The possible reasons are summarized as follows: (1) there is no difference VisTex and Corel-1000; and there exist rotation differences on Face-95, ETHZ, ZuBuD, and KTH-2a. According

to these encouraging results, it can be deduced that the proposed LTCSP, LTCSP_{uni}, and LTCSP_{ri} methods are superior to the LBP-based methods on all six datasets.

Table 4. The evaluations the APR and ARR rates resulting from the proposed methods and the LBP-based methods on six datasets.

Method	Performance	Date Set					
		Face-95	ETHZ	ZuBuD	VisTex	KTH-2a	Corel-1000
LBP	APR (%)	63.45	49.28	61.45	93.37	91.56	71.86
	ARR (%)	31.73	49.28	61.45	58.36	2.19	7.19
LBP _{uni}	APR (%)	58.25	44.38	54.63	90.83	88.56	68.94
	ARR (%)	29.12	44.38	54.63	56.77	2.11	6.89
LBP _{ri}	APR (%)	59.78	45.96	53.07	89.75	85.52	66.73
	ARR (%)	29.89	45.96	53.07	56.09	2.04	6.67
LTCSP	APR (%)	94.31	90.57	85.63	98.56	98.96	83.94
	ARR (%)	47.15	90.57	85.63	61.60	2.36	8.39
LTCSP _{uni}	APR (%)	97.19	94.04	85.97	97.81	99.15	82.83
	ARR (%)	48.60	94.04	85.97	61.13	2.37	8.28
LTCSP _{ri}	APR (%)	97.39	94.72	86.11	97.53	99.19	82.33
	ARR (%)	48.69	94.72	86.11	60.96	2.37	8.23

5.7. Comparison with Other Color LBP Descriptor

Table 5 reports the evaluations of the APR and ARR rates obtained by the proposed descriptors and a series of state-of-the-art color LBP descriptors including OCLBP [18], IOCLBP [19], maLBP [23] mdLBP [23], OC-LBP + CH [26], LPCP [27] on all six datasets. The best values are highlighted in bold. For LTCSP, it not only reaches the higher results than all other previous color LBP descriptors on Face-95, ETHZ, ZuBuD and KTH-2a datasets, but also yields the highest {APR = 98.56%, ARR = 61.60%} on VisTex and {APR = 83.94%, ARR = 8.39%} on Corel-1000. In the case of "LTCSP_{uni}", it achieves a better performance than OCLBP, IOCLBP, mdLBP, maLBP, and OC-LBP + CH on all six datasets. For LTCSP_{ri}, it acquires the highest {APR = 98.39%, ARR = 48.69%} on Face-95, {APR = 94.72%, ARR = 94.72%} on ETHZ, {APR = 86.11%, ARR = 86.11%} on ZuBuD, and {APR = 99.19%, ARR = 2.37%} on KTH-2a respectively. But we also note that the {APR, ARR} rates of LTCSP_{uni} and LTCSP_{ri} are slightly inferior to the LPCP on VisTex and Corel-1000. The main reason is that LPCP has a higher feature dimension. However, the issue can be addressed by adding more useful feature vectors. Based on these considerable results, the effectiveness of the proposed descriptors is demonstrated by comparing with six state-of-the-art color LBP descriptors. Furthermore, there exists illumination differences, head scale and head turn on Face-95, rotation differences on ETHZ, point view differences on ZuBuD, rotation, scaling and illumination differences on KTH-2a. Therefore, the proposed descriptors are also robust against rotation, scaling and illumination to some extent.

Table 5. The evaluations of APR and ARR rates obtained by the proposed descriptors and six state-of-the-art color LBP descriptors on six datasets.

Method	Performance	Date Set					
		Face-95	ETHZ	ZuBuD	VisTex	KTH-2a	Corel-1000
OCLBP	APR (%)	64.40	42.57	56.42	92.42	90.62	68.86
	ARR (%)	32.20	42.57	56.42	57.76	2.16	6.89
IOCLBP	APR (%)	66.47	45.51	61.05	95.59	94.26	73.01
	ARR (%)	33.24	45.51	61.05	59.75	2.25	7.30
maLBP	APR (%)	67.94	55.17	59.46	95.80	92.25	74.45
	ARR (%)	33.97	55.17	59.46	59.87	2.20	7.45
mdLBP	APR (%)	72.97	61.43	61.85	97.05	94.88	76.02
	ARR (%)	36.49	61.43	61.85	60.65	2.26	7.60
OC-LBP + CH	APR (%)	80.50	78.04	63.98	92.20	95.31	74.94
	ARR (%)	40.25	78.04	63.98	57.63	2.27	7.49
LPCP	APR (%)	92.33	88.15	84.82	98.33	98.77	82.85
	ARR (%)	46.16	88.15	84.82	61.46	2.36	8.29
LTCSP	APR (%)	94.31	90.57	85.63	98.56	98.96	83.94
	ARR (%)	47.15	90.57	85.63	61.60	2.36	8.39
LTCSP _{uni}	APR (%)	97.19	94.04	85.97	97.81	99.15	82.83
	ARR (%)	48.60	94.04	85.97	61.13	2.37	8.28
LTCSP _{ri}	APR (%)	97.39	94.72	86.11	97.53	99.19	82.33
	ARR (%)	48.69	94.72	86.11	60.96	2.37	8.23

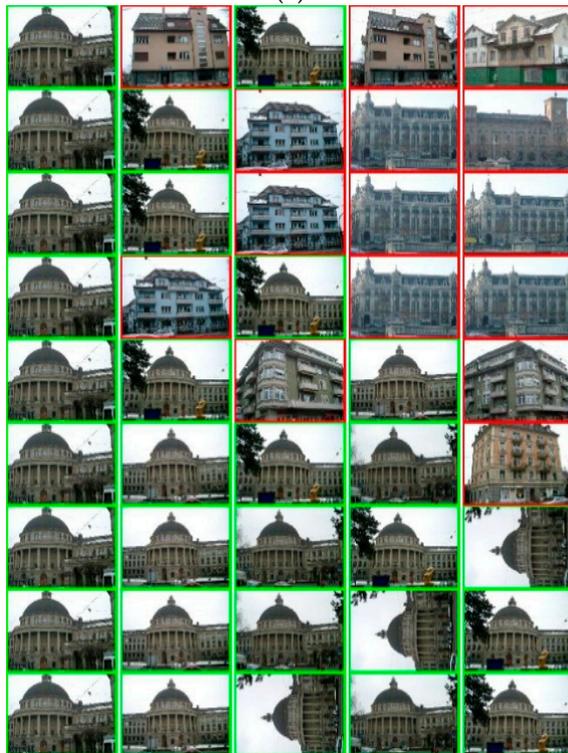
Figure 7 depicts the comparisons of the top-10 returned images acquired by the proposed methods and six previous color LBP methods on six benchmark datasets. Based on the similarity measure score, the top-10 returned images are sorted in descending order. Among them, the leftmost image in each row not only is the most similar image, but also is the query image. Clearly, when a returned image is the relational image, it is tagged in a green box; else it is tagged in a red box. In Figure 7a, the APR rate is 10% using OCLBP, 10% using IOCLBP, 10% using maLBP, 20% using mdLBP, 40% using mdLBP, 60% using OC-LBP + CH, 100% using LTCSP, 100% using LTCSP_{uni} and 100% using LTCSP_{ri} respectively. From this figure, we can deduce that the proposed LTCSP, LTCSP_{uni} and LTCSP_{ri} descriptors not only are efficient for face-based image retrieval applications, but also are insensitive to head scales and head turns. In Figure 7b, the {APR, ARR} rate of relational images using OCLBP, IOCLBP, maLBP, mdLBP, OC-LBP + CH, LPCP, LTCSP, LTCSP_{uni} and LTCSP_{ri} are {60%, 60%}, {60%, 60%}, {60%, 60%}, {60%, 60%}, {60%, 60%}, {80%, 80%}, {100%, 100%}, {100%, 100%} and {100%, 100%} respectively. From this comparison, it can be observed that LTCSP, LTCSP_{uni} and LTCSP_{ri} are robust for object-based image retrieval applications when the object of “toy plane” is rotated arbitrarily. In Figure 7c, the number of relational images using OCLBP, IOCLBP, maLBP, mdLBP, OC-LBP + CH, LPCP, LTCSP, LTCSP_{uni} and LTCSP_{ri} are 6, 6, 6, 8, 7, 6, 6, 6, 10, 10 and 10. From Figure 7c, we summarize that the proposed descriptors are effective and robust for landmark-based image retrieval applications even if some occlusions (e.g., tree and car) are purposefully included in images. As expected, from Figure 7d–f, LTCSP, LTCSP_{uni} and LTCSP_{ri} still bring about higher APR rates than existing color LBP descriptors, apart from LPCP in Figure 7d. However, by comparing with the 10th returned images using LPCP, LTCSP, LTCSP_{uni} and LTCSP_{ri}, it can be clearly seen that the proposed descriptors are more semantically similar with the leftmost query image. From these observations and analyses, the practicability and usability of LTCSP, LTCSP_{uni} and LTCSP_{ri} are illustrated.



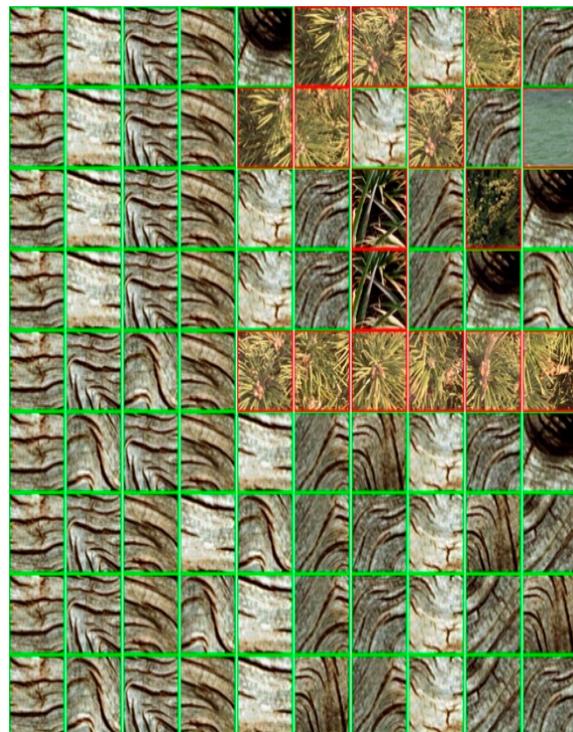
(a)



(b)



(c)



(d)

Figure 7. Cont.

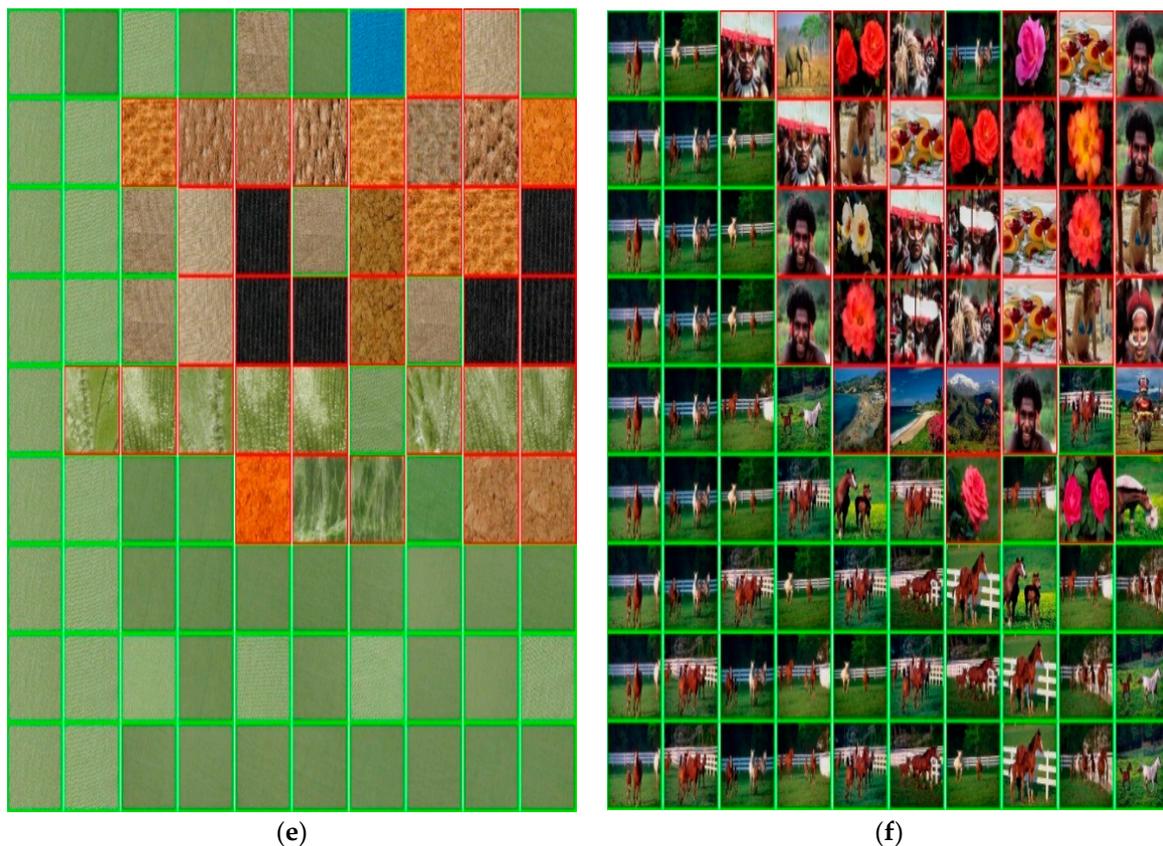


Figure 7. The top-10 returned images using nine methods (the 1st row using OCLBP, the 2nd row using IOCLBP, the 3rd row using maLBP, the 4th row using mdLBP, the 5th row using OC-LBP + CH, the 6th row using LPCP, the 7th row using LTCSP, the 8th row using LTCSP_{uni}, and the 9th row using LTCSP_{ri}) on: (a) Face-95; (b) ETHZ; (c) ZuBuD; (d) VisTex; (e) KTH-2a; and (f) Corel-1000 respectively.

Table 6 reports the comparison of the feature dimensionality (d) and the memory cost (kB) comparison among the proposed descriptors and six former color LBP methods. Similar to LPCP, the items of 688/496/688/436/544/448 (d) and 5.38/3.88/5.38/3.41/4.25/3.50 (kB) represent LTCSP with 688 d and 5.38 kB conducts experiments on Face-95, 496 d and 3.88 kB on ETHZ, 688 d and 5.38 kB on ZuBuD, 436 d and 3.41 kB on VisTex, 544 d and 4.25 kB on KTH-2a, as well as 448 d and 3.50 kB on Corel-1000 respectively. As documented in Table 6, the feature dimension and memory cost of the proposed descriptors are obviously lower than OCLBP, IOCLBP, maLBP, mdLBP and LPCP on all six datasets (apart from LPCP on Corel-1000), yet LTCSP, LTCSP_{uni}, and LTCSP_{ri} are also higher than OC-LBP + CH respectively. However, the superiorities of LTCSP, LTCSP_{uni}, and LTCSP_{ri} are still summarized as follows:

1. The additional feature dimensionality and memory cost effectively improve the accuracy by a large marginal.
2. The LTCSP, LTCSP_{uni}, and LTCSP_{ri} achieve the highest score on all six datasets.
3. The proposed methods achieve a trade-off compromise: adaptive feature dimensionality and acceptable memory cost, and competitive candidate in the real-world CBIR applications.

Table 6. Feature dimensionality (d) and memory cost (kB) among the proposed descriptors and six former descriptors.

Method	Feature Dimensionality (d)	Memory Cost (kB)
OCLBP	1535	11.99
IOCLBP	3072	24.00
maLBP	1024	8.00
mdLBP	2048	16.00
OC-LBP + CH	108	0.84
LPCP	844/760/844/616/592/424	6.59/5.94/6.59/4.81/4.63/3.31
LTCSP	688/496/688/436/544/448	5.38/3.88/5.38/3.41/4.25/3.50
LTCSP _{uni}	491/347/419/203/299/203	3.84/2.71/3.27/1.59/2.34/1.56
LTCSP _{ri}	468/324/468/288/396/180	3.66/2.53/3.66/2.25/3.09/1.41

5.8. Comparison with Deep Learning (DL)-Based Models

Additionally, the proposed LTCSP, LTCSP_{uni}, and LTCSP_{ri} descriptors are further compared with the emerging deep learning (DL)-based models including ALEX [54], GoogleNet [55], VGGm128 [56], VGGm1024 [56], VGGm2048 [56], and VGGm4096 [56]. Referring to [57,58], firstly, the last full-connected layers in the pre-trained models are converted into the corresponding feature vector. Secondly, the converted feature vectors are sent to perform the L2 normalization. Thirdly, the normalized feature vectors are used for computing the similarity measure score.

Figure 8 presents the comparisons among the proposed descriptors and the DL-based models. For LTCSP, LTCSP_{uni}, and LTCSP_{ri}, it can be observed that the proposed descriptors product higher APRs rates than all DL-based models on VisTex, KTH-2a, Face-95, ETHZ and ZuBuD datasets. For Corel-1000, LTCSP, LTCSP_{uni}, and LTCSP_{ri} product lower APR rates than all DL-based models. There are two main reasons for this phenomenon: (1) Corel-1000 is a natural scene image dataset that include more complex scene semantic information; and (2) all DL-based models are pre-trained on the ImageNet that is the natural scene dataset. Comparing with the DL-based models, the superiorities of LTCSP, LTCSP_{uni}, and LTCSP_{ri} are summarized as follows:

1. The DL-based models rely heavily on expensive hardware configurations (e.g., RAM and GPU), yet the proposed descriptors can be easily embedded into cheap hardware devices (e.g., chip and microcontroller).
2. The DL-based models are sensitive to rotation, scaling and illumination differences, while the proposed descriptors are robust against rotation, scaling, and illumination differences to some extent.
3. The DL-based models need to be pre-trained on large-scale and annotated datasets (e.g., ImageNet), which seriously limits its applications.
4. LTCSP, LTCSP_{uni}, and LTCSP_{ri} are superior to the DL-based models on five datasets out of six.

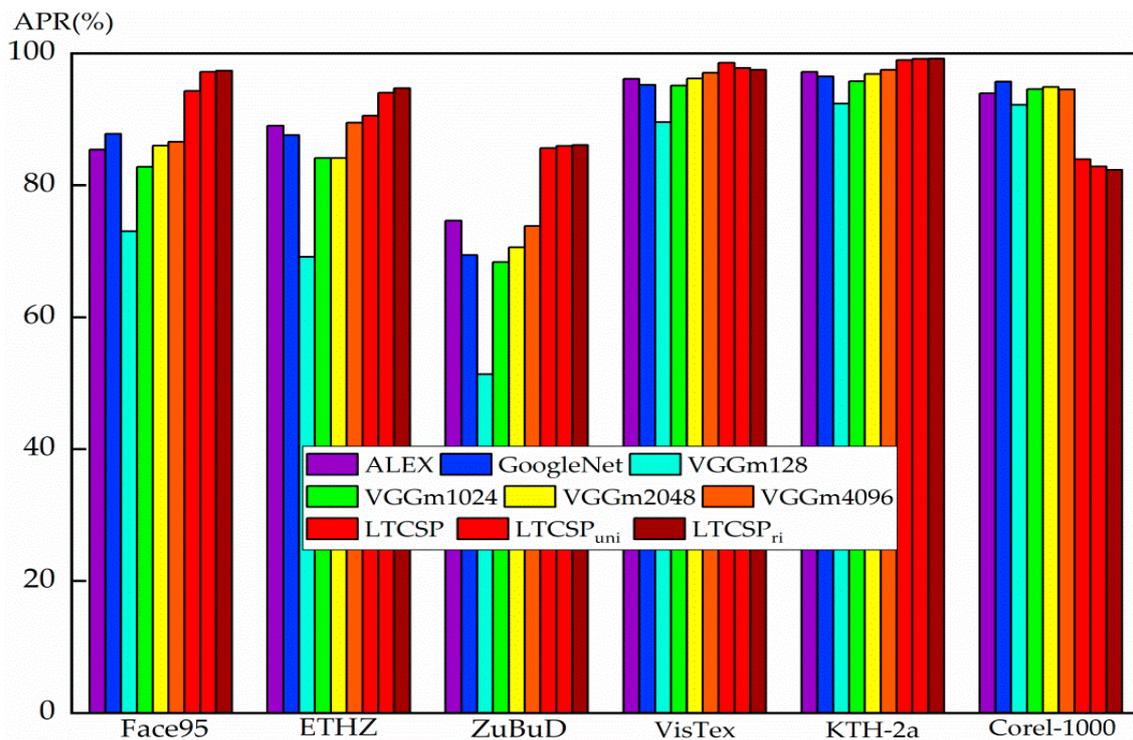


Figure 8. The comparisons of the APR rates among the proposed LTCSP, LTCSP_{uni} and LTCSP_{ri} descriptors and the emerging deep learning (DL)-based models on six datasets.

6. Conclusions

In this study, a series of color LBP descriptors namely local ternary cross structure pattern (LTCSP), uniform local ternary cross structure pattern (LTCSP_{uni}) and rotation-invariant local ternary cross structure pattern (LTCSP_{ri}) is proposed for the CBIR applications. According to the experimental results, the effectiveness, robustness, and practicability of the proposed descriptors are evaluated and compared on face, landmark, object, natural scene and textural image datasets. Based on these considerable results, it can be concluded that the proposed methods achieve a trade-off/compromise among notable retrieval accuracy, adaptive feature dimensionality and acceptable memory cost, and they can be considered as a competitive candidate for real-world CBIR applications.

In the future, unsupervised feature selection [59] will be implemented to tackle the issue of the feature dimensionality and memory cost. In order to improve the robustness against illumination, the image normalization [60] will be exploited in the image pre-processing. In addition, manifold learning (ML) [61,62] and query expansion (QE) [63] will also be considered to further enhance the retrieval performance.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2076-3417/9/11/2211/s1>.

Author Contributions: Q.F. conceived the research idea. Q.F. and Q.H. performed the experiments. Q.F., Q.H. and J.D. wrote the paper. Q.F., Q.H., Y.Y., J.D. and Y.W. gave many suggestions and helped revise this manuscript.

Funding: This research was funded by the National Nature Science Foundation of China, grant number [61871106, 61370152], the Fundamental Research Grant Scheme for the Central Universities, grant number [130204003], the Project of Shandong Province Higher Educational Science and Technology Program, grant number [J16LN68], the Weifang Science and Technology Development Plan Project (Nos. 2017GX006, 2018GX009, 2018GX004) and the National Key Technology Research and Development Programme of the Ministry of Science and Technology of China, grant number [2014BAI17B02].

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

Appendix A

Table A1. The mean square error of the EICQ and FLCQ quantizers on Face-95.

Quantization Level A_{a^*}	Quantizer	Quantization Level A_{b^*}				
		$A_{b^*} = 1$	$A_{b^*} = 2$	$A_{b^*} = 3$	$A_{b^*} = 4$	$A_{b^*} = 5$
$A_{a^*} = 1$	EICQ	1150.65	877.69	825.03	812.65	792.54
	FLCQ	355.51	260.54	254.78	254.46	254.45
$A_{a^*} = 2$	EICQ	737.79	464.83	412.17	399.79	379.69
	FLCQ	156.92	61.95	56.20	55.88	55.86
$A_{a^*} = 3$	EICQ	615.66	342.70	290.04	277.66	257.56
	FLCQ	149.70	54.73	48.97	48.65	48.64
$A_{a^*} = 4$	EICQ	569.25	296.29	243.63	231.25	211.15
	FLCQ	149.35	54.38	48.63	48.31	48.29
$A_{a^*} = 5$	EICQ	549.35	276.39	223.73	211.35	191.25
	FLCQ	149.34	54.37	48.62	48.30	48.28

Table A2. The mean square error of the EICQ and FLCQ quantizers on ETHZ.

Quantization Level A_{a^*}	Quantizer	Quantization Level A_{b^*}				
		$A_{b^*} = 1$	$A_{b^*} = 2$	$A_{b^*} = 3$	$A_{b^*} = 4$	$A_{b^*} = 5$
$A_{a^*} = 1$	EICQ	1458.63	1018.17	874.3	811.39	778.25
	FLCQ	568.67	301.03	286.28	285.81	285.78
$A_{a^*} = 2$	EICQ	1033.79	593.33	449.46	386.55	353.41
	FLCQ	344.29	76.66	61.90	61.43	61.41
$A_{a^*} = 3$	EICQ	905.31	464.86	320.98	258.07	224.94
	FLCQ	347.00	79.37	64.61	64.15	64.12
$A_{a^*} = 4$	EICQ	851.94	411.49	267.61	204.70	171.56
	FLCQ	346.61	78.98	64.22	63.75	63.73
$A_{a^*} = 5$	EICQ	827.35	386.90	243.02	180.11	146.98
	FLCQ	346.61	78.97	64.22	63.75	63.72

Table A3. The mean square error of the EICQ and FLCQ quantizers on ZuBuD.

Quantization Level A_{a^*}	Quantizer	Quantization Level A_{b^*}				
		$A_{b^*} = 1$	$A_{b^*} = 2$	$A_{b^*} = 3$	$A_{b^*} = 4$	$A_{b^*} = 5$
$A_{a^*} = 1$	EICQ	1698.31	1247.15	1106.26	1048.22	1020.12
	FLCQ	609.53	359.00	344.97	344.27	344.24
$A_{a^*} = 2$	EICQ	1203.92	752.76	611.87	553.82	525.72
	FLCQ	302.63	52.11	38.07	37.37	37.34
$A_{a^*} = 3$	EICQ	1041.70	590.54	449.65	391.60	363.50
	FLCQ	283.95	33.42	19.38	18.68	18.66
$A_{a^*} = 4$	EICQ	971.45	520.28	379.40	321.35	293.25
	FLCQ	283.09	32.56	18.52	17.82	17.80
$A_{a^*} = 5$	EICQ	935.66	484.50	343.61	285.56	257.46
	FLCQ	283.04	32.52	18.48	17.78	17.75

Table A4. The mean square error of the EICQ and FLCQ quantizers on VisTex.

Quantization Level A_{a^*}	Quantizer	Quantization Level A_{b^*}				
		$A_{b^*} = 1$	$A_{b^*} = 2$	$A_{b^*} = 3$	$A_{b^*} = 4$	$A_{b^*} = 5$
$A_{a^*} = 1$	EICQ	1530.69	1167.55	1057.57	1011.01	987.40
	FLCQ	556.59	367.10	358.62	358.18	358.17
$A_{a^*} = 2$	EICQ	1053.23	690.09	580.11	533.55	509.94
	FLCQ	267.31	77.81	69.33	68.90	68.88
$A_{a^*} = 3$	EICQ	898.62	535.48	425.50	378.94	355.33
	FLCQ	248.27	58.77	50.29	49.85	49.84
$A_{a^*} = 4$	EICQ	832.16	469.02	359.04	312.48	288.87
	FLCQ	247.46	57.97	49.49	49.05	49.04
$A_{a^*} = 5$	EICQ	798.37	435.23	325.25	278.69	255.08
	FLCQ	247.42	57.92	49.44	49.01	48.99

Table A5. The mean square error of the EICQ and FLCQ quantizers on KTH-2a.

Quantization Level A_{a^*}	Quantizer	Quantization Level A_{b^*}				
		$A_{b^*} = 1$	$A_{b^*} = 2$	$A_{b^*} = 3$	$A_{b^*} = 4$	$A_{b^*} = 5$
$A_{a^*} = 1$	EICQ	1299.41	1090.02	1049.36	1031.49	1016.96
	FLCQ	438.11	379.63	373.66	373.47	373.47
$A_{a^*} = 2$	EICQ	838.96	629.56	588.91	571.04	556.51
	FLCQ	176.84	118.36	112.39	112.20	112.20
$A_{a^*} = 3$	EICQ	693.71	484.32	443.66	425.79	411.26
	FLCQ	160.37	101.89	95.92	95.74	95.73
$A_{a^*} = 4$	EICQ	633.55	424.16	383.50	365.64	351.10
	FLCQ	159.52	101.04	95.07	94.89	94.88
$A_{a^*} = 5$	EICQ	604.17	394.77	354.12	336.25	321.71
	FLCQ	159.49	101.01	95.03	94.85	94.84

Table A6. The mean square error of the EICQ and FLCQ quantizers on Corel-1000.

Quantization Level A_{a^*}	Quantizer	Quantization Level A_{b^*}				
		$A_{b^*} = 1$	$A_{b^*} = 2$	$A_{b^*} = 3$	$A_{b^*} = 4$	$A_{b^*} = 5$
$A_{a^*} = 1$	EICQ	1363.97	1026.78	928.64	887.69	865.82
	FLCQ	509.07	342.97	332.88	332.43	332.41
$A_{a^*} = 2$	EICQ	946.85	609.65	511.52	470.57	448.70
	FLCQ	283.30	117.20	107.12	106.67	106.65
$A_{a^*} = 3$	EICQ	822.90	485.70	387.57	346.62	324.75
	FLCQ	268.31	102.21	92.12	91.67	91.65
$A_{a^*} = 4$	EICQ	768.42	431.23	333.09	292.14	270.27
	FLCQ	267.57	101.47	91.39	90.94	90.92
$A_{a^*} = 5$	EICQ	741.21	404.02	305.88	264.93	243.06
	FLCQ	267.54	101.44	91.35	90.91	90.88

In order to illustrate the advantage of the proposed five-level color quantizer (FLCQ), the mean square error (MSE) values of the EICQ and FLCQ in the CIELAB color model are compared on all six datasets. To guarantee the experimental fairness, the same settings are set to the EICQ and FLCQ quantizers, apart from the number of quantization intervals. The lowest MSE values of the FLCQ and EICQ quantizers are highlighted in bold. Firstly, from Tables A1–A6, it can be summarized that the MSE values of the FLCQ quantizer are lower than EICQ on all six datasets. Secondly, it can be concluded

that along with the refinement of levels (A_{a^*} , A_{b^*}), where A_{a^*} , $A_{b^*} \in \{1, 2, \dots, 5\}$, the MSE values of the EICQ quantizers drops much more obviously, yet the MSE values of the FLCQ quantizers from $A_{a^*} = 4$ to 5 are decrease slightly. The results not only illustrate the stability of the FLCQ quantizers, but also demonstrate that it is suitable to stop the quantization level at $A_{a^*} = 5$. Thirdly, under the same quantization interval and level in both quantizers, the MSE values are different from one another among six datasets. These results show that there exist obvious color probability distribution differences among different datasets. So, it is reasonable to adopt the FLCQ quantizer. As a consequence, we can observe that the FLCQ quantizer produces the lower MSE than the EICQ quantizer.

References

1. Smeulders, A.W.M.; Worring, M.; Santini, S.; Gupta, A.; Jain, R. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *22*, 1349–1380. [[CrossRef](#)]
2. Zheng, L.; Yang, Y.; Tian, Q. SIFT meets CNN: A decade survey of instance retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 1224–1244. [[CrossRef](#)] [[PubMed](#)]
3. Irtaza, A.; Adnan, S.; Ahmed, K.; Jaffar, A.; Khan, A.; Javed, A.; Mahmood, M. An ensemble based evolutionary approach to the class imbalance problem with applications in CBIR. *Appl. Sci.* **2018**, *8*, 495. [[CrossRef](#)]
4. Zeng, Z.; Zhang, J.; Wang, X.; Chen, Y.; Zhu, C. Place recognition: An overview of vision perspective. *Appl. Sci.* **2018**, *8*, 2257. [[CrossRef](#)]
5. Zafar, B.; Ashraf, R.; Ali, N.; Lqbal, M.; Sajid, M.; Dar, S.; Ratyal, N. A novel discriminating and relative global spatial image representation with applications in CBIR. *Appl. Sci.* **2018**, *8*, 2242. [[CrossRef](#)]
6. Feng, Q.; Hao, Q.; Chen, Y.; Yi, Y.; Wei, Y.; Dai, J. Hybrid histogram descriptor: A fusion feature representation for image retrieval. *Sensors* **2018**, *18*, 1943. [[CrossRef](#)]
7. Ojala, T.; Pietikäinen, M.; Maenpää, T. Multi resolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987. [[CrossRef](#)]
8. 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. [[CrossRef](#)]
9. Guo, Z.; Zhang, L.; Zhang, D. A completed modeling of local binary pattern operator for texture classification. *IEEE Trans. Image Process.* **2010**, *19*, 1657–1663.
10. Guo, Z.; Zhang, L.; Zhang, D. Rotation invariant texture classification using LBP variance (LBPV) with global matching. *Pattern Recognit.* **2010**, *43*, 706–719. [[CrossRef](#)]
11. Tan, X.; Triggs, B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. Image Process.* **2010**, *19*, 1635–1650.
12. Murala, S.; Maheshwari, R.P.; Balasubramanian, R. Local tetra patterns: A new feature descriptor for content-based image retrieval. *IEEE Trans. Image Process.* **2012**, *21*, 2874–2886. [[CrossRef](#)] [[PubMed](#)]
13. Subrahmanyam, M.; Maheshwari, R.P.; Balasubramanian, R. Local maximum edge binary patterns: A new descriptor for image retrieval and object tracking. *Signal. Process.* **2012**, *92*, 1467–1479. [[CrossRef](#)]
14. Zhao, G.; Ahonen, T.; Matas, J.; Pietikäinen, M. Rotation-invariant image and video description with local binary pattern features. *IEEE Trans. Image Process.* **2012**, *21*, 1465–1477. [[CrossRef](#)]
15. Ren, J.; Jiang, X.; Yuan, J. Noise-resistant local binary pattern with an embedded error-correction mechanism. *IEEE Trans. Image Process.* **2013**, *22*, 4049–4060. [[CrossRef](#)] [[PubMed](#)]
16. Murala, S.; Wu, Q.M.J. Local ternary co-occurrence patterns: A new feature descriptor for MRI and CT image retrieval. *Neurocomputing* **2013**, *119*, 399–412. [[CrossRef](#)]
17. Verma, M.; Raman, B. Local neighborhood difference pattern: A new feature descriptor for natural and texture image retrieval. *Multimed. Tools Appl.* **2018**, *77*, 11843–11866. [[CrossRef](#)]
18. Mäenpää, T.; Pietikäinen, M. Texture analysis with local binary patterns. In *Handbook of Pattern Recognition and Computer Vision*; Word Scientific: Singapore, 2005; pp. 197–216.
19. Bianconi, F.; Bello-Cerezo, R.; Napoletano, P. Improved opponent color local binary patterns: An effective local image descriptor for color texture classification. *J. Electron. Imaging* **2017**, *27*, 011002. [[CrossRef](#)]
20. Jeena Jacob, I.; Srinivasagan, K.G.; Jayapriya, K. Local opponent color texture pattern for image retrieval system. *Pattern Recognit. Lett.* **2014**, *42*, 72–78. [[CrossRef](#)]
21. Qi, X.; Xiao, R.; Li, C.; Qiao, Y.; Guo, J.; Tang, X. Pairwise rotation invariant co-occurrence local binary pattern. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 2199–2213. [[CrossRef](#)] [[PubMed](#)]

22. Hao, Q.; Feng, Q.; Wei, Y.; Sbert, M.; Lu, W.; Xu, Q. Pairwise cross pattern: A color-LBP descriptor for content-based image retrieval. In Proceedings of the Nineteenth Pacific Rim Conference on Multimedia, Hefei, China, 21–22 September 2018; pp. 290–300.
23. Dubey, S.R.; Singh, S.K.; Singh, R.K. Multichannel decoded local binary patterns for content-based image retrieval. *IEEE Trans. Image Process.* **2016**, *25*, 4018–4032. [[CrossRef](#)] [[PubMed](#)]
24. Liu, P.; Guo, J.; Chamnongthai, K.; Prasetyo, H. Fusion of color histogram and LBP-based features for texture image retrieval and classification. *Inf. Sci.* **2017**, *390*, 95–111. [[CrossRef](#)]
25. Somasekar, M.; Sakthivel Murugan, S. Feature extraction of underwater images by combining Fuzzy C-Mean color clustering and LBP texture analysis algorithm with empirical mode decomposition. In Proceedings of the Fourth International in Ocean Engineering (ICOE2018), Chennai, India, 19 February 2018; pp. 453–464.
26. Singh, C.; Walia, E.; Kaur, K.P. Enhancing color image retrieval performance with feature fusion and non-linear support vector machine classifier. *Optik* **2018**, *158*, 127–141. [[CrossRef](#)]
27. Feng, Q.; Hao, Q.; Sbert, M.; Yi, Y.; Wei, Y.; Dai, J. Local parallel cross pattern: A color texture descriptor for image retrieval. *Sensors* **2019**, *19*, 315. [[CrossRef](#)] [[PubMed](#)]
28. Agarwal, M.; Singhal, A.; Lall, B. Multi-channel local ternary pattern for content-based image retrieval. *Pattern Anal. Appl.* **2019**, *22*, 1–12. [[CrossRef](#)]
29. Bianconi, F.; González, E. Counting local n-ary patterns. *Pattern Recognit. Lett.* **2018**, *177*, 24–29. [[CrossRef](#)]
30. Reta, C.; Cantoral-Ceballos, J.A.; Solis-Moreno, I.; Gonzalez, J.A.; Alvarez-Vargas, R.; Delgadillo-Checa, N. Color uniformity descriptor: An efficient contextual color representation for image indexing and retrieval. *J. Vis. Commun. Image Represent.* **2018**, *54*, 39–50. [[CrossRef](#)]
31. Wan, X.; Kuo, C.C. Content-based image retrieval using multiresolution histogram representation. In Proceedings of the SPIE: Digital Image Storage and Archiving Systems, Philadelphia, PA, USA, 21 November 1995; pp. 312–324.
32. Liu, G.H.; Li, Z.Y.; Zhang, L.; Xu, Y. Image retrieval based on micro-structure descriptor. *Pattern Recognit.* **2011**, *44*, 2123–2133. [[CrossRef](#)]
33. Liu, G.H.; Yang, J.Y. Content-based image retrieval using color difference histogram. *Pattern Recognit.* **2013**, *46*, 188–198. [[CrossRef](#)]
34. Wan, X.; Kuo, C.C. Color distribution analysis and quantization for image retrieval. In Proceedings of the SPIE: Storage and Retrieval for Still Image and Video Databases IV, San Jose, CA, USA, 13 March 1996; pp. 8–17.
35. Duda, R.O.; Hart, P.E. *Pattern Classification and Scene Analysis*; Wiley: New York, NY, USA, 1973; pp. 37–43.
36. Wan, X.; Kuo, C.C. A new approach to image retrieval with hierarchical color clustering. *IEEE Trans. Circ. Syst. Vid.* **1998**, *8*, 628–643. [[CrossRef](#)]
37. Hurvich, L.M.; Jameson, D. An opponent-process theory of color vision. *Psychol. Rev.* **1957**, *64*, 384–404. [[CrossRef](#)]
38. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 3rd ed.; Publishing House of Electronics Industry: Beijing, China, 2010; pp. 455–456. ISBN 9787121102073.
39. Caltech-256 Image Set. Available online: http://www.vision.caltech.edu/Image_Datasets/Caltech256/ (accessed on 8 August 2017).
40. Zhang, M.; Zhang, K.; Feng, Q.; Wang, J.; Kong, J. A novel image retrieval method based on hybrid information descriptors. *J. Vis. Commun. Image Represent.* **2014**, *25*, 1574–1587. [[CrossRef](#)]
41. Standring, S. *Gray's Anatomy: The Anatomical Basis of Clinical Practice*, 41st ed.; Elsevier Limited: New York, NY, USA, 2016; pp. 686–708. ISBN 9780702068515.
42. Guo, J.; Prasetyo, H.; Wang, N. Effective image retrieval system using dot-diffused block truncation coding features. *IEEE Trans. Multimed.* **2015**, *17*, 1576–1590. [[CrossRef](#)]
43. Libor Spacek's Facial Image Databases "Face 95 Image Database". Available online: <https://cswwww.essex.ac.uk/mv/allfaces/faces95.html> (accessed on 8 August 2014).
44. ETH Zurich. Available online: <http://www.vision.ee.ethz.ch/en/datasets/> (accessed on 8 August 2014).
45. Zurich Buildings Database. Available online: <http://www.vision.ee.ethz.ch/en/datasets/> (accessed on 8 August 2014).
46. MIT Vision and Modeling Group. Available online: <http://vismod.media.mit.edu/pub/> (accessed on 12 August 2014).
47. KTH-TIPs2 Image Database. Available online: <http://www.nada.kth.se/cvap/databases/kth-tips/download.html> (accessed on 12 August 2014).
48. Corel 1000 Image Database. Available online: <http://wang.ist.psu.edu/docs/related/> (accessed on 12 August 2014).

49. Guo, J.; Prasetyo, H. Content-based image retrieval using features extracted from halftoning-based block truncation coding. *IEEE Trans. Image Process.* **2015**, *24*, 1010–1024. [[PubMed](#)]
50. Guo, J.; Prasetyo, H.; Su, H. Image indexing using the color and bit pattern feature fusion. *J. Vis. Commun. Image Represent.* **2013**, *24*, 1360–1379. [[CrossRef](#)]
51. Orchard, M.T.; Bouman, C.A. Color quantization of Images. *IEEE Trans. Signal. Process.* **1991**, *39*, 2677–2690. [[CrossRef](#)]
52. Kolesnikov, A.; Trichina, E.; Kauranne, T. Estimating the number of clusters in a numerical data set via quantization error modeling. *Pattern Recognit.* **2015**, *48*, 941–952. [[CrossRef](#)]
53. Chen, Y.; Chang, C.; Lin, C.; Hsu, C. Content-based color image retrieval using block truncation coding based on binary ant colony optimization. *Symmetry* **2019**, *11*, 21. [[CrossRef](#)]
54. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the Advances Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
55. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutionals. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
56. Chatfield, K.; Simonyan, K.; Vedaldi, A.; Zisserman, A. Return of the devil in the details: Delving deep into convolutional nets. In Proceedings of the British Machine Vision Conference 2014, Nottinghamshire, UK, 1–5 September 2014.
57. 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; pp. 259–271.
58. Napoletano, P. Visual descriptors for content-based retrieval of remote-sensing images. *Int. J. Remote Sens.* **2018**, *39*, 1043–1376. [[CrossRef](#)]
59. Yi, Y.; Zhou, W.; Liu, Q.; Luo, G.; Wang, J.; Fang, Y.; Zheng, C. Ordinal preserving matrix factorization for unsupervised feature selection. *Signal. Process. Image Commun.* **2018**, *67*, 118–131. [[CrossRef](#)]
60. Cernadas, E.; Fernández-Delgado, M.; González-Rufino, E. Influence of normalization and color space to color texture classification. *Pattern Recognit.* **2017**, *61*, 120–138. [[CrossRef](#)]
61. Yi, Y.; Wang, J.; Zhou, W.; Zheng, C.; Kong, J.; Qiao, S. Non-Negative matrix factorization with locality constrained adaptive graph. *IEEE Trans. Circ. Syst. Vid.* **2019**. [[CrossRef](#)]
62. Liu, S.; Wu, J.; Feng, L.; Qiao, H.; Liu, Y.; Lou, W.; Wang, W. Perceptual uniform descriptor and ranking on manifold for image retrieval. *Inf. Sci.* **2017**, *424*, 235–249. [[CrossRef](#)]
63. Chum, O.; Mikulik, M.; Perdoch, M.; Matas, J. Total recall II: Query expansion revisited. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; pp. 889–896.

