

Local Neighborhood Based Robust Colour Occurrence Descriptor for Colour Image Retrieval

Shiv Ram Dubey, Satish Kumar Singh, Rajat Kumar Singh

Indian Institute of Information Technology, Allahabad, India
shivram1987@gmail.com, sk.singh@iiita.ac.in, rajatsingh@iiita.ac.in

This paper is copyrighted by Institution of Engineering and Technology and published by IET Image Processing at <http://dx.doi.org/10.1049/iet-ipr.2014.0769>
Please cite as: S.R. Dubey, S.K. Singh, R.K. Singh, "Local neighbourhood-based robust colour occurrence descriptor for colour image retrieval," *IET Image Processing*, vol. 9, no. 7, pp. 578-586, 2015.

Abstract:

Content-based image retrieval is demanding accurate with efficient retrieval approaches to index and retrieve the most similar images from the huge image databases. This paper introduces a novel local neighbourhood based robust colour occurrence descriptor (LCOD) to encode the colour information present in the local structure of the image. The colour information is processed in two steps: first, the number of colour is reduced into a less number of shades by quantizing the RGB colour space; second, the reduced colour shade information of the local neighbourhood is used to compute the descriptor. A local colour occurrence binary pattern is generated for each pixel of the image by representing each reduced colour shade occurrence in its local neighbourhood using a binary pattern. The descriptor is constructed by summing the local colour occurrence binary patterns of all the pixels in the image. LCOD is tested over the natural and colour texture databases for content based image retrieval and experimental results suggest that LCOD outperforms other state-of-the-art descriptors. The performance of the proposed descriptor is promising in the case of illumination difference, rotation and scaling also and it can be effectively used for accurate image retrieval under various image transformations.

Keywords:

Feature description, Colour quantization, Image retrieval, Colour occurrence binary pattern, Rotation invariant

1. Introduction

The demand of efficient image retrieval is rapidly increasing for the purpose of retrieving most visually appropriate similar images. Broadly, image retrieval is categorized into three types according to its usability: text-based, content-based, and semantic-based. In the early days, text based approaches were being used for image retrieval but the scope of such methods got reduced upon the existence of Content Based Image Retrieval (CBIR) because retrieving images from its content are more visually accurate [1]. The main aim of CBIR is to facilitate efficient searching, browsing and matching over large databases either offline or online opening an active research area in the field of computer vision for more than decades. The efficiency of the any CBIR system primarily depends upon the discriminating ability present in their image feature descriptions. To describe the semantic concepts and to be able to get the results close to human perception, semantic image retrieval algorithms are used by some researchers [2-4]. Semantic gap is also considered in CBIR to enhance the retrieval performance close to semantic retrieval [5]. Image retrieval is also being used for medical diagnostic purposes [6]. MPEG-7 software is designed by Chang et al. to provide a research friendly platform [7]. In order to extract the features for the classification, Zhang et al. optimized the matrix mapping with data dependent kernel [8]. Representation-based nearest feature plane is proposed in [9] to reduce the computational complexity nearest feature plane for pattern recognition.

Several methods to represent the image in the form of feature description are proposed for image retrieval and classification [10-12]. These methods used low level pixel details of the image such as colour, texture, shape, gradient, orientation etc. in the form of pattern to represent the whole image. Visual features can be computed in two ways: globally and locally. Several approaches have been designed to extract the global [13] and local [14-15] feature descriptions of the image.

Colour is a very important visual cue and widely used in CBIR systems. Chen et al. [16] used the colour information to represent the image features by using the image colour distributions. Colour Difference Histogram (CDH) is designed for colour image analysis [17]. CDH is based on the colour information and spatial layout. CDH represents the image using the colour difference of two pixels in the image for each colour and edge orientation. The use of edge information limits the scale and rotation invariance of the CDH. Colour features is also being used in image hashing, e.g. very recently a robust hashing method is introduced by Tang et al. [18] which combines the colour vector angles with discrete wavelet transform.

Texture feature is important visual information and widely adopted to represent the images. Local Binary Pattern (LBP) is the most popular texture feature and it is widely used in many applications [19]. Various LBP based approaches [14-15, 20-21] have been also reported and became more popular because of their highly accurate retrieval performance and simplicity. LBP and LBP based descriptors can be used efficiently to match the images having some geometric and photometric transformations. Most of these methods have not considered the colour information of the image which is

most important in the CBIR systems and the performance of such methods can be boosted effectively using colour information in their framework.

Images are also represented by different types of structures present in the image [22-24]. Liu et al. [22] represented the co-occurrence matrix properties using histogram and proposed a multi-texton histogram as a feature descriptor for image retrieval. Liu et al. [23] have introduced microstructure descriptors but it does not consider the rotation of the objects. Structure Element Histogram (SEH) is presented in [24] which integrate texture with colour feature. First, the image is quantized in HSV colour space to reduce the number of colours and descriptor is constructed over that quantized image. However, the dimension of the SEH feature descriptor is high and also it is not fully rotation invariant. These structures based methods shown promising results in image retrieval but their performance degrades under illumination, rotation and scaling. Recently, we introduced an illumination compensation mechanism using an intensity reduction to cope with the illumination problem [25]. An interleaved intensity order based descriptor is also proposed recently for rotation and illumination invariant image matching [26].

It is hard to represent an image using only one type of features such as colour or texture. Therefore, it becomes highly desirable to merge these features in such a way that dimensionality should not increase too much. To overcome the drawbacks of the above mentioned descriptor, a novel colour based descriptor is proposed in this paper. The colour descriptor is computed in two steps: 1) number of colours is reduced by quantizing the RGB colour space; 2) local colour occurrence binary patterns are generated for each pixel; and finally each binary pattern is aggregated to find a single pattern. Most of the low-level features based on the colour, texture, shape and spatial location cues are reviewed in [27-28] and point out that image retrieval now uses the feedback mechanism to improve the performance which is obviously a complex process. In this paper, we achieved the best performance by using only low-level feature i.e. colour and occurrences. The novelty of proposed descriptor is with the binary coding of the local colour occurrences (i.e. colour occurrence binary pattern extraction). Basically most of the descriptors such as LBP [19] consider only the boundary pixels of the local regions but we have considered all the pixels of the local region. Other methods such as CDH [17] use the colour differences to encode the descriptor which loses significant information whereas we used the local colour occurrences to construct the descriptor. SEH [24] uses a limited number of structure elements (i.e. 5 structure elements) to avoid the high dimensionality which restricts its discriminating power. Proposed descriptor also differs from the widely used co-occurrence matrices because it does not depend upon the occurrence of a pixel value w.r.t. other pixel value at a particular offset [29]. The proposed feature descriptor is tested in the image retrieval framework over various natural and colour textural databases and found promising results and also exhibit rotation and scale invariance property.

The rest of the paper is organized as follows: section 2 is dedicated to the RGB colour space quantization; section 3 introduces the proposed concept; section 4 exhibits the distance measure and evaluation criteria; section 5 presents the experimental results; and finally section 6 concludes the article.

2. Colour quantization

RGB colour images contain three channels representing Red (R), Green (G) and Blue (B) colours respectively. According to the RGB colour space, the range of shades is $[0, l-1]$, where l is the number of distinguished shades in each channel. The number of different colours possible in this colour space is l^3 , which is a large number. Considering all the colours for feature description is not feasible because the dimension of the descriptor should be minimal as possible. We quantize RGB colour space such as it becomes a single channel with reduced number of shades. To reduce the complexity of the computation, RGB colour space is quantized into $q \times q \times q = q^3$ bins with each colour is quantized into q bins, where $q \ll l$. To retain equal weighting of each colour, all colour components are quantized into equal number of bins. The steps involved in the quantization are as follows:

(1) Divide each Red, Green and Blue component of image I into q shades from l shades respectively. The reduced colour components (i.e. R_{red} , G_{red} and B_{red}) are computed as,

$$R_{red} = \left\lceil \frac{R + 1}{stp} \right\rceil \quad (1)$$

$$G_{red} = \left\lceil \frac{G + 1}{stp} \right\rceil \quad (2)$$

$$B_{red} = \left\lceil \frac{B + 1}{stp} \right\rceil \quad (3)$$

where, $\lceil \cdot \rceil$ is the ceiling operator and $stp = l/q$.

(2) Combine all three components R_{red} , G_{red} and B_{red} into a one-dimension to construct the reduced colour image I_{red} as follows,

$$I_{red} = q^2(R_{red} - 1) + q(G_{red} - 1) + B_{red} \quad (4)$$

We quantize each colour of the RGB image into equal number of shades which retains the symmetric information. Liu et al. [22] also quantized RGB colour space into 64 shades whereas [23] and [24] quantized HSV colour space into

72 shades and [17] quantized $L^*a^*b^*$ colour space into 90 shades. In this paper, the value of q is chosen as 4 which lead to the 64 number of different shades after quantization.

3. Descriptor construction

In this section, we describe the construction process of proposed Local Neighbourhood Based Robust Colour Occurrence Descriptor (LCOD). The descriptor will be computed over quantized image obtained in the previous section. Here, we focus on constructing the descriptor only on the basis of colour features of the image. The whole construction consists of two steps. In the first step, a local colour occurrence binary pattern is generated for each pixel in the image in its local neighbourhood. In the second step, the binary patterns of all pixels of the image are aggregated to find a single pattern.

3.1. Local neighbourhood based colour occurrence binary pattern

The performance of most of the descriptors is restricted under geometric and photometric transformation conditions because they use spatial information. To overcome this problem, we considered the local occurrences of individual reduced colour shades over a local neighbouring region. A local colour occurrence binary pattern is generated for each pixel of the image. First, we find the number of occurrences for each shade in local neighbourhoods and represent it in binary form. Finally, the binary pattern for each shade is concatenated to obtain a single pattern for that pixel. The binary pattern generated for each pixel is used to construct the descriptor.

Let I be an image of size $m \times n$ and I_{red} is the image obtained after quantization of I . We represent the number of occurrences of shade c within D distance local neighbourhood of pixel $P(x, y)$ of image I_{red} by $\mathbb{N}_{(x,y)}^{c,D}$, where $c \in [1, q^3]$, $x \in [D + 1, m - D]$, and $y \in [D + 1, n - D]$. The D distance local neighbourhood of pixel $P(x, y)$ consists of all those pixels $P(x', y')$ which fulfils the following criteria,

$$\max(|x - x'|, |y - y'|) \leq D \quad (5)$$

where $|\cdot|$ is the operator to find the absolute value and $\max(x, y)$ is the operator to find the maximum value among x and y . The maximum possible value of D is $\lfloor \frac{\min(m,n)-1}{2} \rfloor$, where $\min(x, y)$ is the operator to find the minimum value among x and y and $\lfloor \alpha \rfloor$ is the operator to find the largest integer not greater than α .

The binary pattern $\mathcal{B}P$ of length k for shade c within the D distance local neighbourhood of pixel $P(x, y)$ of image I_{red} is defined as,

$$\mathcal{B}P_{(x,y)}^{c,D} = [\mathcal{B}P_{(x,y)}^{c,D}(1), \mathcal{B}P_{(x,y)}^{c,D}(2), \dots, \mathcal{B}P_{(x,y)}^{c,D}(\gamma), \dots, \mathcal{B}P_{(x,y)}^{c,D}(k)] \quad (6)$$

$$\mathcal{B}P_{(x,y)}^{c,D}(\gamma) = \zeta_{\gamma}(\mathbb{N}_{(x,y)}^{c,D}, k) \quad (7)$$

where $\zeta_{\gamma}(a, b)$ finds the γ^{th} element in the binary representation of length b for decimal value a where $\gamma \in [1, k]$ and $\mathbb{N}_{(x,y)}^{c,D} \in [0, (2D + 1)^2]$. We used the operator $[\]$ to denote the concatenation operation.

Note that, the length k of binary pattern of each shade should be sufficiently large to avoid the under length problem in the decimal to binary conversion. The value of k is decided according to the maximum possible value of $\mathbb{N}_{(x,y)}^{c,D}$ which is $(2D + 1)^2$. Mathematically, k is given as,

$$k = \lceil \log_2(2D + 1)^2 \rceil \quad (8)$$

where the operator $\lceil \alpha \rceil$ finds the smallest integer not less than α .

The final binary pattern $\mathfrak{F}P_{(x,y)}^D$ for pixel $P(x, y)$ over its D distance neighbourhood is obtained by concatenating the patterns $\mathcal{B}P_{(x,y)}^{c,D}$ for each shade c and defined as,

$$\mathfrak{F}P_{(x,y)}^D = [\mathfrak{F}P_{(x,y)}^D(1), \mathfrak{F}P_{(x,y)}^D(2), \dots, \mathfrak{F}P_{(x,y)}^D(z), \dots, \mathfrak{F}P_{(x,y)}^D(k \times q^3)] \quad (9)$$

where $\mathfrak{F}P_{(x,y)}^D(z)$ is the z^{th} element of the pattern $\mathfrak{F}P_{(x,y)}^D$ and computed as follows,

$$\mathfrak{F}P_{(x,y)}^D(z) = \mathcal{B}P_{(x,y)}^{\lfloor \frac{z}{k} \rfloor, D}(f(z)) \quad (10)$$

where f is a function defined as follows,

$$f(z) = \begin{cases} k, & \text{if } \text{mod}\left(\frac{z}{k}\right) = 0 \\ \text{mod}\left(\frac{z}{k}\right), & \text{otherwise} \end{cases} \quad (11)$$

where the operator ‘ $\text{mod} \left(\frac{a}{b} \right)$ ’ finds the remainder of the division of a by b .

The dimension d of final binary pattern $\mathfrak{FB}_{(x,y)}^D$ is given as,

$$d(\mathfrak{FB}_{(x,y)}^D) = d(\mathcal{BP}_{(x,y)}^{c,D}) \times q^3 \tag{12}$$

where $d(\mathcal{BP}_{(x,y)}^{c,D}) = k$ and k is defined in (8).

To illustrate the methodology to compute the final binary pattern \mathfrak{FB} for a given pixel, we have considered an example in Fig. 1. In this example, we find the \mathfrak{FB} pattern for the middle pixel (i.e. $\mathfrak{FB}_{(3,3)}$) having shade value 3 (highlighted in green in Fig. 1) by considering the value of D as 2 and 1.

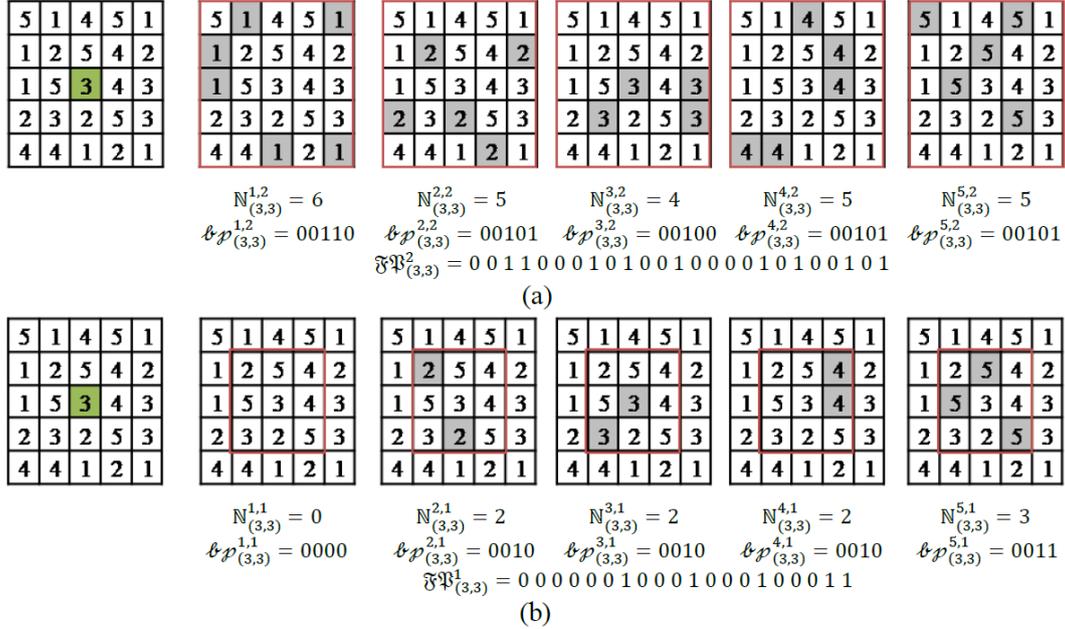


Fig. 1. An illustration to compute the local colour occurrence binary pattern for (a) $D = 2$, and (b) $D = 1$. The number of shades is considered as 5 in this example.

Only 5 different shades are considered in this example (i.e. $c = [1, 5]$). In Fig. 1(a), the value of D is considered as 2 such that the maximum possible value of $N_{(3,3)}^{c,2}$ becomes 25 (i.e., $(2D + 1)^2$). The number of occurrences of shade c (i.e. $N_{(3,3)}^{c,2}$) is 6, 5, 4, 5, and 5 for $c = 1, 2, 3, 4$, and 5 respectively. The value of k will be 5 in the example of Fig. 1(a) such that the maximum possible number of occurrence 25 can be represented in the k -bit binary form. $\mathfrak{FB}_{(3,3)}^2$ is computed by concatenating $\mathcal{BP}_{(3,3)}^{c,2}$ for $c = [1, 5]$ which is binary representation of $N_{(3,3)}^{c,2}$. Similarly, $\mathfrak{FB}_{(3,3)}^1$ is also computed in the Fig. 1(b) for the same example of Fig. 1(a). The value of k is 4 in Fig. 1(b) because the maximum possible number of occurrence is 9 when the value of D is considered as 1. The length of \mathfrak{FB} is 25 in Fig. 1(a), whereas it is 20 in Fig. 1(b).

3.2. Local neighbourhood based robust colour occurrence descriptor

A local colour occurrence binary pattern is generated for each pixel of the image over its local neighbourhood of maximum distance D . Now, we have $m \times n$ (i.e. size of the image) number of binary patterns each of length $k \times q^3$ and our aim is to combine these binary patterns to form a single pattern. If we simply concatenate each binary pattern to obtain a single pattern then the dimension of the resulting pattern will be $m \times n \times k \times q^3$ which is too high to facilitate the efficient storing and matching. To obtain the low dimensional and efficient descriptor, we adopted bitwise addition of each binary pattern. The descriptor des obtained after bin-wise addition is given as,

$$des(z) = \sum_{i=D+1}^{m-D} \sum_{j=D+1}^{n-D} \mathfrak{FB}_{(i,j)}^D(z) \quad \forall z = [1, \quad k \times q^3] \tag{13}$$

The construction process of des depends upon the size of the image (i.e. the values of m and n). The values of the elements of des will be larger for the higher resolution images and vice-versa, i.e.,

$$des(z) \propto (m \times n) \tag{14}$$

To overcome the effect of image size over the descriptor, we normalize the *des* such that it becomes invariant to the scale of the image. Normalization is carried out in such a way that the sum of resultant descriptor becomes 1. The final local colour occurrence descriptor (LCOD) is given as,

$$LCOD(z) = \frac{des(z)}{\sum_{t=1}^{k \times q^3} des(t)} \quad \forall z = [1, \quad k \times q^3] \quad (15)$$

The pseudo code for the proposed LCOD descriptor is summarized in the Algorithm-1. The computation complexity for LCOD is given by $O(iter)$ (i.e. only for the second loop) because D is a constant so first loop will not considered in time complexity. In other words, we can say that the computation time complexity of LCOD descriptor is of the order of $O(n)$.

Algorithm 1. The LCOD descriptor construction

```

1  Input:  Quantized Image ( $I_{red}$ )
           Quantization bin ( $q$ ) for each channel
2  Output: Descriptor (LCOD)
3  Initial:  $t \leftarrow 0, des \leftarrow []$ 
4
5  Begin
6   $[s1, s2] \leftarrow size(I_{red})$ 
7   $D \leftarrow 1$  // Set local neighbourhood size to 1
8   $k \leftarrow \lceil \log_2((2D + 1)^2) \rceil$ 
9  For  $i \leftarrow -D : D$ 
10  For  $j \leftarrow -D : D$ 
11   $t = t + 1$ 
12   $tmp(1:s1, 1:s2, t) = I_{red}(D + i + 1:s1 - D + i, D + j + 1:s2 - D + j)$ 
13  end
14  end
15   $iter \leftarrow q^3$ 
16  For  $c \leftarrow 1 : iter$ 
17   $tmp1 \leftarrow (tmp == c)$  // Set to 1 if true
18   $tmp2 \leftarrow sum(tmp1, 3)$  // Summation in 3rd dimension
19   $tmp3 \leftarrow de2bi(tmp2, k)$  // Decimal to binary conversion of  $tmp2$  into  $k$  bits
20   $tmp4 \leftarrow sum(tmp3)$  // Summation in 1st dimension
21   $des \leftarrow [des tmp4]$  // Concatenation
22  end
23   $LCOD \leftarrow des/sum(des)$ 
24  return (LCOD)
25 end
```

4. Distance measure and Evaluation criteria

4.1. Distance measure

In this paper, we use two distance measures used in SEH [24] and CDH [17] respectively. Let $F = \{f_1, f_2, \dots, f_{dim}\}$ be the feature vector for the images in the database and $T = \{t_1, t_2, \dots, t_{dim}\}$ be the feature vector of the query image where dim is the dimension of the descriptor. The distance metric used in [23] is defined as,

$$d_{SEH}(T, F) = \sum_{i=1}^{dim} \frac{|f_i - t_i|}{1 + f_i + t_i} \quad (16)$$

and the distance metric used in [17] is defined as,

$$d_{CDH}(T, F) = \sum_{i=1}^{dim} \frac{|f_i - t_i|}{|f_i + f_\mu| + |t_i + t_\mu|} \quad (17)$$

where $f_\mu = \sum_{i=1}^{dim} f_i / dim$ and $t_\mu = \sum_{i=1}^{dim} t_i / dim$. We used these two distance measures for fair comparison because we compared the proposed method with SEH and CDH in the result section. Basically both distance measures find the binwise displacement with respect to the total possible displacement. The main difference between both distances is that d_{CDH} uses the mean of both feature vectors to avoid the same displacement for different set of feature points.

4.2. Evaluation criteria

In content based image retrieval, the main task is to find most similar images of a query image in the whole database. We adopted *Precision* and *Recall* curves to represent the effectiveness of proposed descriptor. For a particular database, we define the average retrieval precision (ARP) and average retrieval rate (ARR) from the average precision and the average recall of each category of that database.

$$ARP = \frac{\sum_{i=1}^{N_C} AP(i)}{N_C} \quad (18)$$

$$ARR = \frac{\sum_{i=1}^{N_C} AR(i)}{N_C} \quad (19)$$

Where N_C is the total number of categories in that database, AP and AR are the average precision and average recall respectively for a particular category of that database and defined as,

$$AP(j) = \frac{\sum_{i=1}^{n_j} P(i)}{n_j} \quad \forall j = [1, N_C] \quad (20)$$

$$AR(j) = \frac{\sum_{i=1}^{n_j} R(i)}{n_j} \quad \forall j = [1, N_C] \quad (21)$$

Where n_j is the number of images in the j^{th} category of that database, P and R are the precision and recall for a query image and defined by the following equation,

$$P(k) = \frac{NS}{NR} \quad \forall k = [1, n_j] \quad (22)$$

$$R(k) = \frac{NS}{ND} \quad \forall k = [1, n_j] \quad (23)$$

where NS is the number of similar images retrieved, NR is the number of images retrieved, and ND is the number of similar images in the whole database. We have considered 10 values of the NR from 10 to 100 with an interval of 10.

5. Experiments and results

In this section, we test the robustness and discriminating power of proposed local colour occurrence descriptor (LCOD) under rotation, different scale and illumination change conditions. We also compare proposed descriptor with state-of-the-art descriptors in CBIR systems over various databases.

5.1. Databases

5.1.1. Corel-10k database

In order to evaluate the proposed descriptor, a standard Corel Database for Content based Image Retrieval [30] is used in this paper. A large number of images having different details are present in the Corel image database ranging from natural scenarios and outdoor sports to animals. We refer this database as the *Corel-10k* database. The *Corel-10k* database consists of the 80 categories having nearly more than 100 images in each category totaling 10800 images in the database. The images in a particular group are category-homogeneous. The resolutions of the images are either 120×80 or 80×120 pixels.

5.1.2. Corel-1k database

We also evaluate proposed descriptor on the *Corel-1k* database which is a subset of the *Corel-10k* database. In the *Corel-1k* database, we have considered 10 categories of *Corel-10k* database namely dinosaur, fitness, bus, ship, flower, mountain, butterfly, elephant, fish and horse. In *Corel-1k*, each category contains 100 images; totaling 1000 images.

5.1.3. Corel-rotated database

In order to emphasize the performance of proposed descriptor under rotation, we synthesized a *Corel-rotated* database by rotating all images of *Corel-1k* with angle 0, 90, 180, and 270 degrees. *Corel-rotated* database contains same 10 categories as of *Corel-1k* but the number of images in each category becomes 400 including rotated images; totaling 4000 images in *Corel-rotated* database.

5.1.4. Corel-scale database

We synthesized a *Corel-scale* database also by scaling all the images of *Corel-1k* at the scales of 0.5, 0.75, 1, 1.25, and 1.5. *Corel-scale* database contains 500 images for each category with 5000 total number of images.

5.1.5. Corel-illumination database

To test the performance of the descriptor under monotonic intensity change, we synthesized *Corel-illumination* database by adding -60, -30, 0, 30, and 60 in the all channels (i.e. Red, Green and Blue) of the images of the *Corel-1k* database. Thus, the *Corel-illumination* database consists of the 5000 images with 500 images per categories.

5.1.6. MIT-VisTex database

In this experiment, we used MIT-VisTex database consisting of 40 different images. The images are having 512×512 dimension and collected from the [31]. We created a database of 640 images (i.e. 40×16) by partitioning each 512×512 image into sixteen 128×128 non overlapping sub-images.

5.1.7. STex Database

We also used STex-512-splitted database from [32] for retrieval experiment. This database consists of the 7616 numbers of colour images of dimension 128×128 from 26 categories.

5.2. Experimental Results

Local colour occurrence descriptor (LCOD) uses the colour information of the image in an effective manner to describe the local features. Until specified, RGB colour space is quantized into 64 numbers of bins (i.e. $q = 4$) and the value of D is considered as 1 (i.e. $k = 4$) to compute the LCOD descriptor which leads to a descriptor of dimension 256-bin. Image retrieval results are represented by ARP and ARR curves by presenting ARP and ARR at y-axis and a number of retrieved images at x-axis. We evaluated proposed descriptor in terms of the ARP by changing distance metric, colour space, distance of the local neighbourhood, number of quantization, levels of quantization, and step between two adjacent centre pixels over Corel-1k database (Fig. 2).

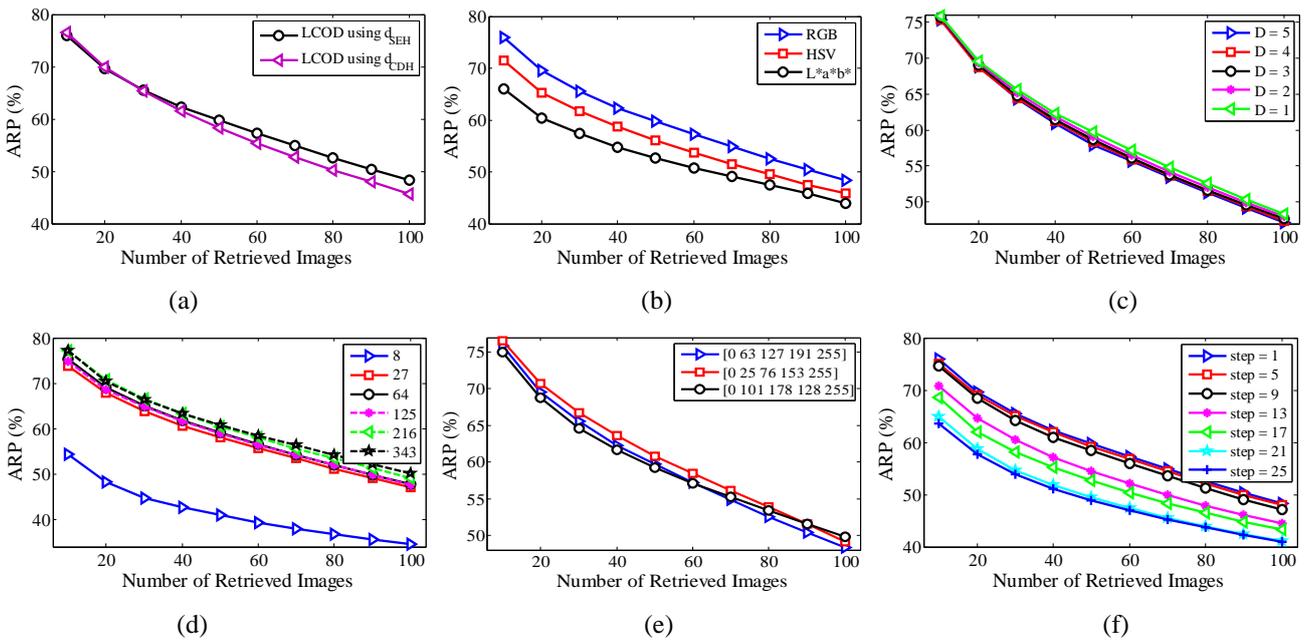


Fig. 2. ARP curves using LCOD descriptor (a) comparison among the distance metric used in [24] and [17], (b) comparison among colour spaces RGB, HSV, and L*a*b*, (c) comparison between different values of D (i.e. area of local region), (d) comparison among different number of RGB quantization, (e) comparison among different levels of RGB quantization, and (f) comparison among the different step size between centre pixels in both x - and y -directions over Corel-1k database.

We evaluated the LCOD descriptor using two distance metric d_{SEH} and d_{CDH} introduced in [24] and [17] respectively (see Fig. 2(a)). The performance of our descriptor in terms of ARP using d_{SEH} is better over *Corel-1k* database. In the rest of the paper, we use d_{SEH} distance to find the distance. Proposed descriptor is designed by uniform quantization of RGB colour space into 64 bins, whereas, SEH quantized HSV colour space into 72 bins by giving more weight to the Hue channel (i.e. H channel is divided into 8 bins and S and V channels are divided into 3 bins each) and CDH quantized L*a*b* colour space into 90 bins giving more weight to the L* channel (i.e. L* channel is divided into 10 bins and a* and b* channels are divided into 3 bins each). We also used the quantization technique used by SEH and CDH with our descriptor and compared it with original LCOD (i.e. using RGB colour quantization) on the Corel-1k

database in Fig. 2(b) and found that RGB colour space quantization is better suited for LCOD descriptor. In the rest of the paper LCOD descriptor is designed by RGB colour quantization. The efficiency of LCOD descriptor also depends upon the size of local neighbourhood under consideration (i.e. D). We tested it over Corel-1k database and observed that increasing the distance of local neighbourhood lowers the performance of the descriptor as illustrated in Fig. 2(c). In the rest of the paper the value of D is considered as 1. We also observed that changing the number of quantization levels does not guidance any benefits as depicted in Fig. 2(d). We tested LCOD descriptor constructed by considering q as 2, 3, 4, 5, 6 and 7 (i.e. number of quantization level is 8, 27, 64, 125, 216 and 343) over the Corel-1k database. The dimension of the LCOD descriptor is 32, 98, 256, 500, 864 and 1372 when q is set 2, 3, 4, 5, 6 and 7 respectively. It is observed across the plot that for $q=2$ the results are not good and for $q=5$, $q=6$ and $q=7$ the results are not too improved as the dimension of the descriptor is increased. Between $q=3$ and $q=4$, the results for $q=4$ is better and the descriptor dimension is also reasonable. In the rest of the paper the value of q is considered as 4 which lead to the 64 distinct quantization levels. The dimension of the LCOD descriptor also depends upon the value of the D as depicted in Table 1. The dimension of the descriptor is same for $D=4$ and $D=5$ because the maximum possible values of $\mathbb{N}_{(x,y)}^{c,4}$ and $\mathbb{N}_{(x,y)}^{c,5}$ are 81 and 121 respectively and both require at least 7 bits to represent it in binary form. The neighbouring points close to the centre pixel provide more relevant information in the construction of the descriptor. The comparison also made between the linear and nonlinear RGB quantization in Fig. 2(e). We considered the levels of quantization at linear and nonlinear intervals (i.e. both increasing and decreasing steps) and found that the performance is better for nonlinear quantization with increasing step size (i.e. step is less for darker pixels and more for brighter pixels). Although, we have used the linear quantization in this paper but it is suggested to use nonlinear quantization with increasing step size. All the pixels of the image are considered in the process of descriptor construction (i.e. step between two adjacent pixels in either x and y direction is simply one). Here, we experimented by changing the step in both directions and presented the result in Fig. 2(f) and observed that the performance is decreasing with an increase in the step between two adjacent centre pixel.

Table 1. Dimension of LCOD vs size of local neighbourhood

Size of neighbourhood (D)	1	2	3	4	5	6	7	8	9	10	11
Number of bits required (k)	4	5	6	7	7	8	8	9	9	9	10
Dimension of LCOD	256	320	384	448	448	512	512	576	576	576	640

5.3. Comparison and discussion

We considered SEH [24] and CDH [17] for comparison with the proposed descriptor LCOD. SEH and CDH are the state-of-art descriptors proposed recently for CBIR system. The dimension of SHE, CDH and LCOD are 360, 108 and 256 respectively. To test the discriminating power of LCOD, we compared it with SEH and CDH using ARP and ARR in Fig. 3 over (a) Corel-1k, (b) Corel-10k, (c) MIT-VisTex, and (d) STex databases, where Corel-1k and Corel-10k are the natural databases and MIT-VisTex and STex are the colour texture databases. LCOD performs better than other descriptors over each database (i.e. for both natural and textural databases). Moreover, it is much better when the number of categories in the database is more (i.e. Corel-10k and STex databases).

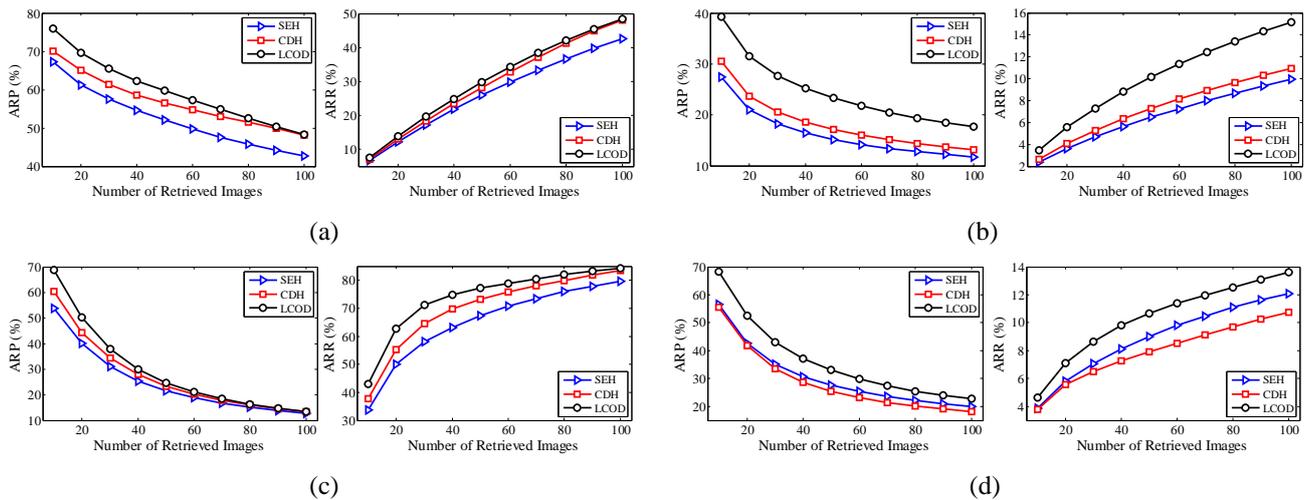


Fig. 3. Comparison of proposed LCOD descriptor with SEH and CDH over (a) Corel-1k, (b) Corel-10k, (c) MIT-VisTex and (d) STex databases using ARP and ARR curves.

We also compared each descriptor over Corel-rotated, Corel-scale and Corel-illumination databases to test the robustness of descriptors for rotation, scaling and lighting conditions respectively and illustrated in Fig. 4 using the ARP values in (a) and Average Precision per category in (b). The performance of LCOD descriptor is better than other

descriptors in the case of both geometric (rotation and scaling) and photometric (illumination) transformations (see Fig. 4(a)). The improvement using LCOD over SEH and CDH is more in the case of geometric transformations. The better ARP is achieved when images are retrieved using the proposed descriptor as compared to the other descriptors under rotation, scaling and different lighting conditions. The performance of proposed descriptor is also better in most of the categories of the databases as compared to the other approaches (see Fig. 4(b)). In the case of rotation (i.e. for Corel-rotated database), LCOD performs better in 80% categories in terms of average precision than SEH and CDH both. It is observed that ‘ship’ is the category (i.e. category no. 4) in which LCOD is not dominating in any scenario because of the smoother regions in the images of ‘ship’ category (i.e. lack of good features). It is evident from Fig. 3-4 that the LCOD descriptor is more discriminating and robust than the SEH and CDH descriptors.

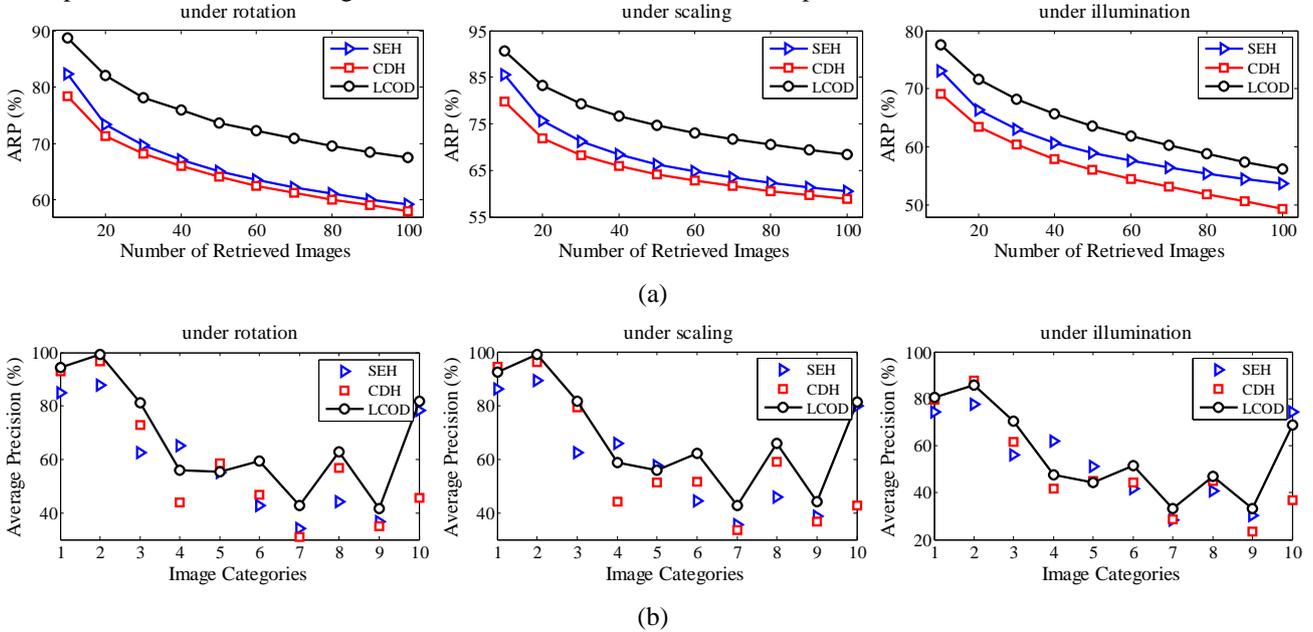


Fig. 4. Comparison of proposed LCOD descriptor with SEH and CDH over *Corel-rotated*, *Corel-scale* and *Corel-illumination* databases in terms of the (a) ARP and (b) Average Precision per category.



Fig. 5. Image retrieval results for a query image from (a) Corel-1k, (b) Corel-rotated, (c) Corel-scale and (d) Corel-illumination database using SEH (1st row), CDH (2nd row) and LCOD (3rd row) descriptors.

Fig. 5 depicts image retrieval results using SEH (1st row), CDH (2nd row) and LCOD (3rd row) descriptors for a query image taken from the (a) Corel-1k, (b) Corel-rotated, (c) Corel-scale and (d) Corel-illumination database. We retrieved 10 most similar images to the query image in each case. The first image in each row of each retrieval result of Fig. 5 is the query image as well as the most similar retrieved image and remaining images are the other retrieved images in the order of decreasing distance with the query image. The images in the Fig. 5(a) are retrieved from the Corel-1k database corresponding to the query image of type ‘horse’ using SEH, CDH and LCOD descriptors. Out of 10, 9 images are retrieved using LCOD which are relevant to the query image, whereas, only 7 and 4 images are retrieved using SEH and CDH respectively which are relevant to the query image. It means that LCOD is benefited by the relevant colour information and represents the image more semantically because the irrelevant images are retrieved due to similar colour information by SEH and CDH for different category images. Fig. 5(b) shows the image retrieval results from the Corel-

rotated database with a query image of type ‘ship’ using SEH, CDH and LCOD descriptors respectively. Out of 10 retrieved images, only 6 and 8 relevant images are retrieved by SEH and CDH descriptors respectively whereas proposed LCOD descriptor is able to retrieve all 10 relevant images to the query image.

Fig. 5(c) depicts the retrieval results using SEH, CDH and LCOD descriptors in the case of scaling using Corel-scale database. The output images are resized to fit such that the low scale images become blurred. Under scaling, for the considered example, LCOD retrieved 10 relevant images with a precision of 100% whereas SEH and CDH are able to achieve only 50% (5 relevant) and 90% (9 relevant) precision respectively. CDH fails to retrieve all the scaled images of the query image. From this retrieval result, it is deduced that LCOD retrieve more accurate images under scaling condition also as compared to the SEH and CDH. Fig. 5(d) present the retrieval results from the Corel-illumination database (i.e. under monotonic intensity change) using SEH, CDH and LCOD descriptors respectively. In this case, an image of category ‘dinosaur’ is taken as the query image. Each descriptor gains 100% precision in this case but LCOD and CDH fail to retrieve all the images at different illumination of the same image, whereas SEH are able to retrieve some images having illumination difference. Among the retrieved relevant images, SEH retrieved more semantically correct images than LCOD in the case of photometric transformation but LCOD is better in terms of precision. In this case, the performance of LCOD is better than CDH in both term precision and semantically more accurate.

We conducted the experiments using a system having Intel(R) Core(TM) i5 CPU 650@3.20 GHz processor, 4 GB RAM, and 32-bit Windows 7 Ultimate operating systems. Four cores are used in the computation of computation and execution times. Table 2 depicts the feature extraction time and total retrieval time in seconds for each descriptor over Corel-1k, Corel-10k, MIT-VisTex and STex databases. Both feature extraction and total retrieval time are high for Corel-10k and STex databases for each descriptor because these databases are having large number of images. In terms of the feature extraction time, SEH is faster by nearly 6.4 times than LCOD while LCOD is faster by nearly 1.15 times than CDH. The total retrieval time depends upon the dimension of the descriptor and the dimension of CDH. The retrieval using CDH is faster by nearly 2.2 times than LCOD while retrieval using LCOD is faster by nearly 1.6 times than SEH. From the feature extraction and total retrieval time, it is deduced that the extraction of proposed descriptor is more time efficient than CDH descriptor and the retrieval using proposed descriptor is more time efficient than SEH descriptor.

Table 2. The feature extraction time and total retrieval time in seconds using SEH, CDH and LCOD descriptors over Corel-1k, Corel-10k, MIT-VisTex and STex databases

Databases/Time	Feature Extraction Time (Seconds)			Total Retrieval Time (Seconds)		
	SEH	CDH	LCOD	SEH	CDH	LCOD
Corel-1k Database	8.28	61.98	53.11	2.98	1.03	2.23
Corel-10k Database	90.92	677.89	589.61	471.04	112.39	301.75
MIT-VisTex Database	10.44	72.88	66.37	1.39	0.55	1.02
STex Database	103.11	801.25	737.47	355	71.11	216.52

From the above experiments, it is deduced that the proposed LCOD descriptor is having a more discriminative ability as well as more robustness to geometric and photometric transformations. The proposed descriptor outperforms other state-of-art descriptors. LCOD has utilized the local colour information effectively to form a colour based descriptor. The computation cost of the LCOD is also comparable with other descriptors. The performance of the LCOD is not better in the categories where planar images are present such as ‘ship’ because the quantization step will fail in this case to produce the image with more information.

6. Conclusion

This paper presented an efficient colour based image feature description for CBIR. The proposed descriptor used the concept of finding the quantized colour occurrences into the local neighbourhood of any pixel to achieve the inherent rotation invariance. RGB colour space is quantized into 64 shades to represent the colour feature of the image and local colour occurrences is used to represent the colour feature most efficiently. We extracted the colour occurrences and represent it in the binary form to generate a local colour occurrence binary pattern for each quantized colour shade independently. In this way, proposed local neighbourhood based robust colour occurrence descriptor (LCOD) captures the most relevant local colour information of each quantized colour shade. Proposed descriptor is rotation invariant inherent and describes the image features more efficiently. Our experimental results on natural and colour textural databases including rotation, scale, and illumination cases suggest that the LCOD descriptor performs better than other descriptors and can be effectively applied in the CBIR system. LCOD is more robust towards scale and rotation and outperforms state-of-the-art colour and texture descriptors, especially in the case of geometric and photometric transformations.

References

- [1]. Penatti, O. A., Silva, F.B., Valle, E., Gouet-Brunet, V., Torres, R. D. S.: 'Visual word spatial arrangement for image retrieval and classification', *Pattern Recognition*, 2014, 47, (2), pp. 705-720
- [2]. Zagoris, K., Ergina, K., Papamarkos, N.: 'Image retrieval systems based on compact shape descriptor and relevance feedback information', *Journal of Visual Communication and Image Representation*, 2011, 22, (5), pp. 378-390
- [3]. Zhang, D., Islam, M.M., Lu, G.: 'A review on automatic image annotation techniques', *Pattern Recognition*, 2012, 45, (1), pp. 346-362
- [4]. Singh, N., Dubey, S.R., Dixit, P., Gupta, J.P.: 'Semantic Image Retrieval by Combining Color, Texture and Shape Features', *Proc. Int. Conf. Computing Sciences*, September 2012, pp. 116-120
- [5]. Bian, W., Tao, D.: 'Biased discriminant Euclidean embedding for content-based image retrieval', *IEEE Transactions on Image Processing*, 2010, 19, (2), pp. 545-554
- [6]. Chinnasamy, S.: 'Performance improvement of fuzzy-based algorithms for medical image retrieval', *IET Image Processing*, 2014, 8, (6), pp. 319-326
- [7]. Chang, F.C., Hang, H.M., Huang, H.C.: 'Research Friendly MPEG-7 Software Testbed', *Proc. IS&T/SPIE Symposium on Electronic Imaging Science and Technology*, 2003, pp. 890-901
- [8]. Zhang, D.L., Qiao, J., Li, J.B., Qiao, L.Y., Chu, S.C., Roddick, J.F.: 'Optimizing Matrix Mapping with Data Dependent Kernel for Image Classification', *Journal of Information Hiding and Multimedia Signal Processing*, 2014, 5, (1), pp. 72-79
- [9]. Feng, Q., Huang, C.T., Yan, L.: 'Representation-based Nearest Feature Plane for Pattern Recognition', *Journal of Information Hiding and Multimedia Signal Processing*, 2013, 4, (3), pp. 178-191
- [10]. Chen, Y., Li, X., Dick, A., Hill, R.: 'Ranking consistency for image matching and object retrieval', *Pattern Recognition*, 2014, 47, (3), pp. 1349-1360
- [11]. He, X.: 'Laplacian regularized D-optimal design for active learning and its application to image retrieval', *IEEE Transactions on Image Processing*, 2010, 19, (1), pp. 254-263
- [12]. Dubey, S.R., Jalal, A.S.: 'Detection and Classification of Apple Fruit Diseases Using Complete Local Binary Patterns', *Proc. Third IEEE International Conference on Computer and Communication Technology*, 2012, pp. 346-351
- [13]. Quéllec, G., Lamard, M., Cazuguel, G., Cochener, B., Roux, C.: 'Fast wavelet-based image characterization for highly adaptive image retrieval', *IEEE Transactions on Image Processing*, 2012, 21, (4), pp. 1613-1623
- [14]. Heikkilä, M., Pietikäinen, M., Schmid, C.: 'Description of interest regions with local binary patterns', *Pattern recognition*, 2009, 42, (3), pp. 425-436
- [15]. Wang, Z., Fan, B., Wu, F.: 'Local intensity order pattern for feature description', *Proc. Int. Conf. Computer Vision*, November 2011, pp. 603-610
- [16]. Chen, W.T., Liu, W.C., Chen, M.S.: 'Adaptive color feature extraction based on image color distributions', *IEEE Transactions on Image Processing*, 2010, 19, (8), pp. 2005-2016
- [17]. Liu, G.H., Yang, J.Y.: 'Content-based image retrieval using color difference histogram', *Pattern Recognition*, 2013, 46, (1), pp. 188-198
- [18]. Tang, Z., Dai, Y., Zhang, X., Huang, L., Yang, F.: 'Robust image hashing via colour vector angles and discrete wavelet transform', *IET Image Processing*, 2014, 8 (3), pp. 142-149
- [19]. Ojala, T., Pietikäinen, M., Maenpää, T.: 'Multiresolution gray-scale and rotation invariant texture classification with local binary patterns', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24 (7), pp. 971-987
- [20]. Murala, S., Maheshwari, R.P., Balasubramanian, R.: 'Local tetra patterns: a new feature descriptor for content-based image retrieval', *IEEE Transactions on Image Processing*, 2012, 21, (5), pp. 2874-2886
- [21]. Tan, X., Triggs, B.: 'Enhanced local texture feature sets for face recognition under difficult lighting conditions', *IEEE Transactions on Image Processing*, 2010, 19, (6), pp. 1635-1650
- [22]. Liu, G.H., Zhang, L., Hou, Y.K., Li, Z.Y., Yang, J.Y.: 'Image retrieval based on multi-texton histogram', *Pattern Recognition*, 2010, 43, (7), pp. 2380-2389
- [23]. Liu, G.H., Li, Z.Y., Zhang, L., Xu, Y.: 'Image retrieval based on micro-structure descriptor', *Pattern Recognition*, 2011, 44, (9), pp. 2123-2133
- [24]. Wang, X., Wang, Z.: 'A novel method for image retrieval based on structure elements' descriptor', *Journal of Visual Communication and Image Representation*, 2013, 24, (1), pp. 63-74
- [25]. Dubey, S.R., Singh, S.K., Singh, R.K.: 'A Multi-Channel based Illumination Compensation Mechanism for Brightness Invariant Image Retrieval', *Multimedia Tools and Applications*, 2014.
- [26]. Dubey, S.R., Singh, S.K., Singh, R.K.: 'Rotation and Illumination Invariant Interleaved Intensity Order Based Local Descriptor', *IEEE Transactions on Image Processing*, 2014, 23 (12), pp. 5323-5333
- [27]. Haralick, R.M., Shanmugam, K., Dinstein, I.H.: 'Textural features for image classification', *IEEE Transactions on Systems, Man and Cybernetics*, 1973, (6), pp. 610-621
- [28]. 'Corel Image Database', <https://sites.google.com/site/dctresearch/Home/content-based-image-retrieval>, accessed March 2014

- [29]. MIT Vision and Modeling Group, Cambridge, 'Vision texture', <http://vismod.media.mit.edu/pub/>, accessed April 2014
- [30]. Salzburg Texture Image Database, <http://www.wavelab.at/sources/STex/>, accessed April 2014
- [31]. Liu, Y., Zhang, D., Lu, G., Ma, W.Y.: 'A survey of content-based image retrieval with high-level semantics', *Pattern Recognition*, 2007, 40, (1), pp. 262-282.
- [32]. Rafiee, G., Dlay, S.S., Woo, W.L.: 'A review of content-based image retrieval', *Proc. 7th International Symposium on Communication Systems Networks and Digital Signal Processing*, 2010, pp. 775-779