

Rotation and Scale Invariant Hybrid Image Descriptor and 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 has been published by Computers and Electrical Engineering, Elsevier.

The final publication is available at: <http://dx.doi.org/10.1016/j.compeleceng.2015.04.011>.

This paper is covered with a Creative Commons Attribution Non-Commercial No Derivatives License.

Citation: Dubey, S. R., Singh, S. K., & Singh, R. K. (2015). Rotation and scale invariant hybrid image descriptor and retrieval. Computers & Electrical Engineering, vol. 46, pp. 288-302.

Abstract

Accurate image retrieval is required to index and retrieve large number of images from huge databases. In this paper, an efficient approach is presented to encode the color and textural features of images from the local neighborhood of each pixel. The color features are extracted by quantizing the RGB color space into a single channel with reduced number of shades. The texture information is encoded with structuring patterns generated from the locally structured elements chosen as a basis. Color and textural features are fused together to construct the inherently rotation and scale-invariant hybrid image descriptor (RSHD). This fusion is carried out by extracting textural cues over each shade independently. RSHD has been tested on the Corel dataset and experimental results suggest that RSHD outperforms state-of-the-art descriptors. The performance of the RSHD is promising under rotation and scaling. It can also be effectively used under more complex image transformations.

Keywords

Structure element, Rotation-invariant descriptor, Scale-invariant descriptor, Texture analysis, Content-based image retrieval, Color quantization

1. Introduction

1.1. Motivation

The demand for efficient image retrieval is rapidly increasing. In the early days, text based approaches were being used for image retrieval, but since the scope of such methods got reduced upon the existence of content-based image retrieval (CBIR), because retrieving images from its content is more visually accurate. In the published literature, content-based approaches describe more objectively and effectively, than text based approaches [1]. The main aim of CBIR is to facilitate efficient searching, browsing and matching over large datasets either offline or online opens an active research area in the field of computer vision and image processing from more than decades. Some recently reported typical applications of image retrieval are computerized facial diagnosis and retrieving human actions from realistic video databases [2]. The efficiency of the any CBIR system primarily depends upon the discriminating power present in their image feature descriptions. A CBIR system must be able to retrieve the images having details oriented randomly in nature. Recently, semantic based approaches became popular for image retrieval because it copes with the problem of CBIR [3, 4]. To describe the semantic concepts and to be able to get the results close to human perception, relevance feedback algorithms are used by some researchers [5] in semantic image retrieval. They considered the priority provided by the users and bridge the semantic gap, but drawback of such methods is that it is not fully automated and it requires user intervention to provide the feedback. Several methods to encode the information present in the images are proposed through published literature for image retrieval [6, 7, 8, 9, 10, 11]. These methods used pixel level details (i.e. low level feature descriptions) of the image including color, texture, shape, gradient, orientation, etc. in the form of pattern to represent the whole image and images matched with their pixel level details. These features have been used efficiently in each type of CBIR systems such as based on global feature, region based feature, local feature, and structuring feature. The main shortcoming of these methods is with the utilization of different more robust features and lack of multiple type of information because by using just one type of features the method may not be able to characterize the image more accurately.

1.2. Related works

Global feature-based CBIR [12, 13, 14] does not divide the image into multiple regions. Chen et al. [12] used only color information to represent the image features by using the image color distributions. The color distributions are preserved up to the third moment in their method. The results are better than other color based features, but it should be noted that they have not considered any features other than color while the proposed method extracts structuring patterns over quantized shades distributions which increases the distinctiveness of proposed descriptor. Color difference histogram (CDH) is designed for color image analysis [13]. Color, texture and shape features are used for global image representation effectively by Wang et al. [14]. They have used only dominant set of color information with steerable filter decomposition and pseudo-Zernike moments in their descriptor construction. These methods have not taken care of local neighboring structures required to encode the relationship among the neighboring pixels.

Region based approaches [15, 16] also used by several researchers to integrate the spatial location with the feature description. Hsiao et al. [15] partitioned images into five regions with fixed absolute locations. Similar to the case of semantic retrieval, their approach also needs user intervention in the middle of the retrieval process. On the other hand, proposed method considers only local neighborhoods of a given pixel which boost it with local discriminative power. In order to represent the image's spatial and color arrangements, Lin et al. [16] introduced three kinds of feature descriptors. To extract these features, they used K-means clustering approach to partition the whole image into different groups (i.e. clusters) using its intensity values. These regions based approaches have shown promising results with the expense of large dimensional descriptions and high computation.

Over the year's local binary pattern (LBP) [17] and various LBP based approaches [18, 19, 20, 21, 22, 23] have been reported and became more popular because of their highly accurate performance and simplicity. LBP is constructed by comparing the each pixel in the image with its neighboring pixels and according to the sign of comparison the LBP is generated [17]. To reduce the size of LBP, only orthogonal pairs are compared in the center-symmetric local binary pattern (CS-LBP) [18]. Based on the multi-resolution, Zhu and Wang [19] have used multiple local patterns to encode the textural feature. Their method is also promising in the case of rotation and scaling, but lack of color information restricts their use in image retrieval. The complete local binary pattern is used for the fruit disease recognition [20]. To achieve invariance towards monotonic intensity change an order based descriptor local intensity order pattern (LIOP) is proposed [21]. Promising results using LIOP have been reported in the case of a monotonic intensity change, but this type of methods fails if there is a change in the objects of the image because using orders only does not consider actual scenarios. Dubey et al. [22] have extended the concept of LIOP over interleaved neighboring sets and designed the interleaved intensity order based local descriptor (IOLD) for image matching. They also proposed an illumination compensation mechanism to cope with the varying illumination for brightness-invariant image retrieval [23]. 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 fine structures of the images which can be seen as a basic building block of the image and the performance of such methods can be boosted effectively using fine structures in their framework.

Images are also represented by different types of structures present in the image [24, 25, 26]. Liu et al. [24] represented the co-occurrence matrix properties using the histogram to compound the advantages of histogram with co-occurrence matrix and proposed a multi-texton histogram (MTH) as a feature descriptor for image retrieval. Liu et al. [25] have introduced microstructure descriptors (MSD). MSD integrates color, texture, shape and spatial layout properties of the image for efficient content-based image retrieval. An efficient Structure Element Histogram (SEH) is presented by Xingyuan et al. [26] which integrates texture with color feature. These structures based methods shown promising results in image retrieval, but their performance degrades under rotation and scaling. In the image retrieval and image classification problems, it is not possible to encode the exact information contained by an image using only one type of features such as color or texture. Therefore, it becomes highly desirable to merge these features in such a way that dimensionality should not increase too much.

Color and texture information are used by Wang et al. [27] to design a CBIR system. They used Zernike chromaticity distribution moments to capture the color features from the opponent chromaticity space which is a rotation and flip-invariant. They also used the Contourlet transform to encode the texture feature which is a rotation and scale-invariant. In [27], the color and texture features are first encoded separately and then combined to form the final feature vector, whereas we formed the descriptor by simultaneous encoding of color and texture in a hybrid manner. The main difference between the approach in [27] and proposed is that the color and texture features are processed independently in [27], whereas we processed the texture feature in conjunction with the color feature. Curvelet transform is also adopted by Youssef and integrated with enhanced dominant colors for texture analysis in CBIR [28]. The HSV color space is quantized to encode the color information. The Curvelet captures accurate texture information as well as directional features by tuning to different orientations, but also increases the dimensionality of the descriptor. Another major difference is the approach in [28] is region based which results in the increased dimension of the descriptor, whereas in our case the dimension is limited and based on the number of structuring element.

To overcome the drawbacks of the above-mentioned descriptors, a rotation and scale-invariant hybrid descriptor (RSHD) is proposed in this paper. The proposed approach considers the whole image as a single region and constructs the descriptor over it. The RSHD is the fusion of color and textural cues present in the image in an efficient manner.

The rest of the paper is organized as follows: Section 2 is dedicated to the RGB color space quantization; Section 3 presents two descriptors intended for benchmarking and the RSHD descriptor; Section 4 explores some distance measure concepts used in this work to compute the similarity score between two images and evaluation criteria; Section 5 employs special image retrieval databases to show the promising experimental results from the RSHD method, including its scale invariance property, its efficiency and how it comes up as a solution to the problems identified in the construction of discriminative image feature descriptors; and, finally, Section 6 concludes this article.

2. Quantization of RGB color space into shades

RGB color images are being frequently considered for the extraction of color features. RGB color images contain three channels representing Red (R), Green (G) and Blue (B) colors respectively. According to the intensities of these three colors, the original color is defined. The benefit with the RGB color space is its similarity with the actual color of the natural scenes. It is highly desirable to handle the colors in an efficient manner because color is a key factor in the image retrieval for efficient feature description. According to the RGB color space, the range of shades is $[0, l-1]$, where l is the number of distinguished shades in each channel. The number of different colors possible in this color space is l^3 , which is a large number. Even at a particular time, perceiving large number of colors is not feasible for the human eyes. Considering all the colors for feature description is not feasible because the dimension of the descriptor should be as minimal as possible. In order to facilitate the extraction of small dimensions but efficient features, for example color, texture, shape, etc., we quantize RGB color space such as it becomes a single channel with reduced number of different shades. We refer the image after RGB quantization as quantized image. For simultaneous encoding of texture with color feature, we extract the texture information for each quantized shade separately and finally combine them to find a single pattern. In this way, only useful color information is combined with the texture information which can extract the image features more efficiently and performs better in CBIR.

To reduce the complexity of the computation, RGB color space is quantized into $q \times q \times q = q^3$ bins with each color, is quantized into q bins, where $q \ll l$. To retain equal weighting of each color, all color components are quantized into equal number of bins. After quantizing into q^3 number of bins, we still have enough number of distinguishing color information. The steps involved in the quantization are as follows:

- (1) Divide each Red, Green and Blue component into q shades from l shades respectively using the following equation:

$$\begin{aligned} R &= t \quad \forall R \in [(t-1) \times l / q, t \times l / q), t = \{1, 2, \dots, q\} \\ G &= t \quad \forall G \in [(t-1) \times l / q, t \times l / q), t = \{1, 2, \dots, q\} \\ B &= t \quad \forall B \in [(t-1) \times l / q, t \times l / q), t = \{1, 2, \dots, q\} \end{aligned} \quad (1)$$

where, t is the shade number in which any pixel belongs after quantization for a particular channel.

- (2) Combine all three components R , G and B into a one-dimension to construct the image color feature as follows:

$$C = q^2(R-1) + q(G-1) + B \quad (2)$$

- (3) Extract the point set P_C which is a set of points having a quantized color C by following equation:

$$P_C = \{(x, y) \mid (x, y) \in I, I(x, y) = C, C = [1, q^3]\} \quad (3)$$

where, $I(x, y)$ represent the shade value obtained after quantization at pixel coordinate (x, y) .

The quantization may cut the homogeneous region into two or more parts, but we will see that if the number of quantization level is sufficient then its effect can be minimized. The benefits in RGB color space quantization over HSV and L*a*b* color space quantization is that we can quantize each color of the RGB image into equal number of bins which retains the symmetric information. Liu et al. [24] also quantized RGB color space into 64 bins whereas [25] and [26] quantized HSV color space into 72 bins and [13] quantized L*a*b* color space into 90 bins. In this paper, the value of q is chosen to 4 which lead to the 64 number of different shades after quantization. Texture feature for each pixel in the image will be constructed over each quantized shade using local neighboring structure pattern.

3. Overview of Some Benchmark Descriptors and the RSHD

This section introduces the reader to the two descriptors used for the sake of comparison with the new descriptor: the structure element histogram (SEH) and the color difference histogram (CDH). Next, it describes the proposed rotation and scale-invariant hybrid descriptor (RSHD), which has five rotation-invariant structure elements used to encode the texture information. We have used SEH and CDH as the benchmark descriptors for three reasons: 1) these methods also utilized the concept of color quantization for the color

data encoding, but in other color spaces, 2) these methods also integrate the texture with color features, and 3) these methods are also proposed for the image retrieval problem.

3.1 The Structure Element Histogram Method

An efficient structure element histogram (SEH) is presented by Xingyuan et al. [26] which integrates texture with color feature. They have quantized the images in HSV color space to reduce the number of colors and descriptor is constructed over that quantized image. Their descriptor exhibits the color and texture spatial correlation in the form of structure elements. They have used five types of structure elements to encode the textural information as shown in the Fig. 1. These structure elements are not rotation-invariant and also have not used the concept symmetry with center pixel, whereas, we will use the rotation-invariant structure elements having the symmetry with center pixel. Moreover, the dimension of the SEH feature descriptor is high and also it is not fully rotation-invariant because their basic structure elements are partial invariant to the rotation.



Fig. 1. Five structure elements of SEH, 1's are represented by the shaded square [26].

3.2 The Color Difference Histogram (CDH) Method

Color difference histogram (CDH) is designed for color image analysis [13]. CDH is based on the color information and spatial layout. CDH represents the image using a color difference of two pixels in the image for each color and edge orientation. Basically, CDH counts the perceptually uniform color difference between two points in $L^*a^*b^*$ color space under different backgrounds with regard to colors and edge orientations. They quantized $L^*a^*b^*$ color space to encode the color information and used the local gradients to encode the textural information. Unlike SEH and proposed method, this method captures the color and textural data separately and concatenates it to obtain the single feature vector. The use of highly correlated neighboring information limits the scale and rotation invariance property of the CDH.

3.3. Rotation and Scale Invariant Hybrid Descriptor (RSHD)

This sub-section describes the proposed rotation and scale-invariant hybrid descriptor (RSHD). First, we present the five rotation-invariant structure elements used in this paper to encode the texture information and then we describe the descriptor construction process in detail.

3.3.1. Rotation-invariant structure element

Color, texture and shape are the key information that present in the natural images. We can think of the image as an arrangement of the regions with different color, texture and shapes. The local neighborhood contains much texture and shape information and plays an important role in the human visual system to perceive the local structures. Structure element has a strong influence on texture representation and recently it is used by some researchers to encode the texture information [24, 25, 26]. But these structure elements are partially invariant to the rotation in some extent. In the present work, we have further enhanced the concept of structure element by considering rotation-invariant structure element and the descriptor is constructed by considering local neighborhood textural information with the color information. The local neighboring structures of the images from the same class are quite similar and if such information can be encoded in a rotation-invariant manner, it may become an important description in image matching. Five structure elements are considered in this paper to match the local neighborhood of each pixel in the image with the structure element. Proposed structure elements basically encode the similarity information among the local neighborhood by considering 4 neighbors.

The proposed approach considers the whole image as a single region and constructs the descriptor over it. The RSHD is the fusion of color and textural cues present in the image in an efficient manner. To encode the color information, we quantized the RGB color space into single channel having less number of shades. The textural feature is computed in the form of binary structuring pattern. Structuring pattern is generated by considering local neighboring structure element. In order to fuse the color and texture, we extracted the structuring pattern for each quantized shade independently. In this way, we are able to encode the color and texture information simultaneously. The local neighboring structure elements facilitate proposed descriptor to boost with rotation and scale-invariant property. We incorporated the basic local structures into the framework of neighboring patterns.

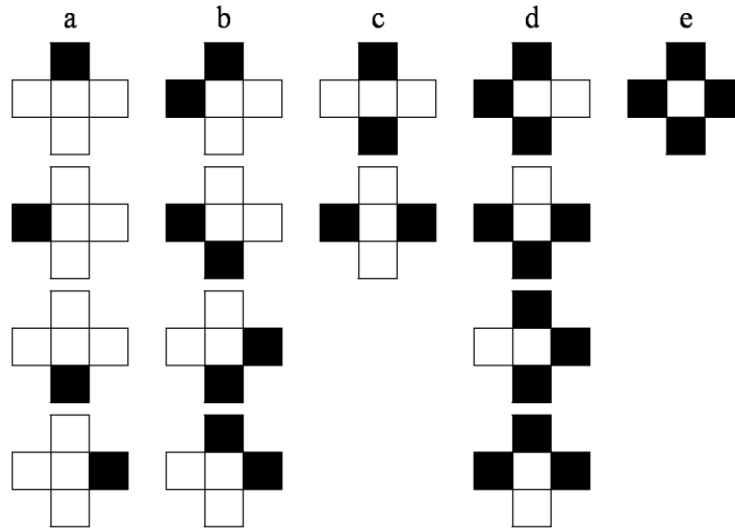


Fig. 2. Five structure element containing (a) only one, (b) two consecutive, (c) two non-consecutive, (d) three consecutive, and (e) four consecutive active elements.

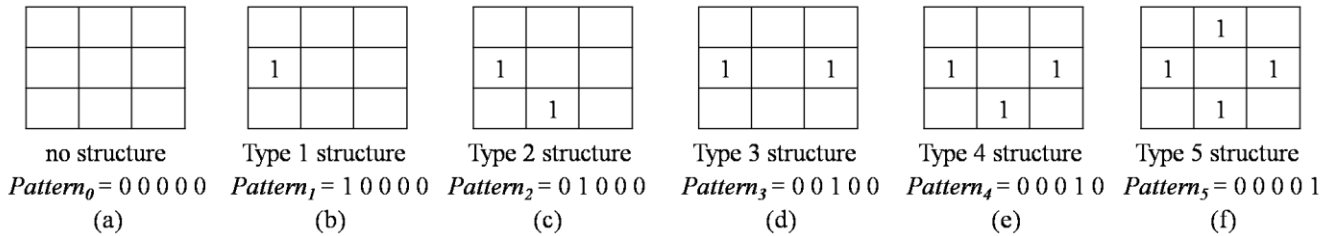


Fig. 3. Six patterns derived from the five structure elements, representing (a) no structure, (b) type 1 structure, (c) type 2 structure, (d) type 3 structure, (e) type 4 structure, and (e) type 5 structure.

The five structure elements used in this paper are shown in the Fig. 2 and it is represented by the number and orientation of the active elements (i.e. highlighted pixels). We refer these structures of Fig. 2 (a-e) as type 1, type 2, type 3, type 4, and type 5 structures respectively. Note that structure element in Fig. 2. (a-e) have (4, 4, 2, 4, 1) degrees of freedom respectively towards rotation. This freedom of rotation and closest neighboring information allows the construction of rotation and scale-invariant hybrid descriptor using these five structure elements. If the number of active elements is more it means that the local neighborhood of a particular pixel is dense with a particular shade. We quantized the original RGB color space into 64 different shades. Five structure elements shown in Fig. 2 serve as the base templates to extract the texture information over each quantized shade for each pixel in the image. To illustrate the extraction of pattern using these structuring elements, consider only two quantized shades in the image. We define six patterns, $Pattern_i$ for $i=0$ to 5 from five structuring elements of 5 bins. The six patterns derived from the five structure element are illustrated in Fig. 3. $Pattern_0$ corresponds to no structure for a particular quantized shade and the values of all five bins are set to zero. $Pattern_i$ for $i=1$ to 5 correspond to type i structure element and only the i^{th} bin is set to 1 in the corresponding pattern (all other bins are set to zero).

3.3.2. Descriptor construction using structuring pattern

Structuring pattern is the textural information on the basis of six patterns derived from the five structure elements for each pixel in the image over all quantized shades. To illustrate the concept efficiently, suppose we have only four quantized shades form 1 to 4. Fig. 4 illustrates the extraction of local neighborhood maps for each quantized shade and according to the structure of the map its type is defined. Fig. 5 shows three examples to compute the structuring pattern SP for any pixel from its four neighboring pixels.

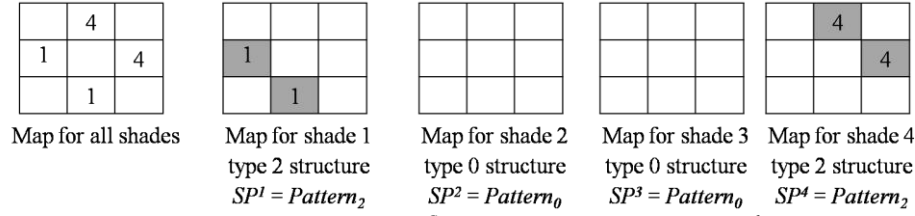


Fig. 4. Extraction of structure map for each quantized shade; sp^C represents the pattern over C^{th} quantized shade for a particular pixel. In this example the value of q is set to four.

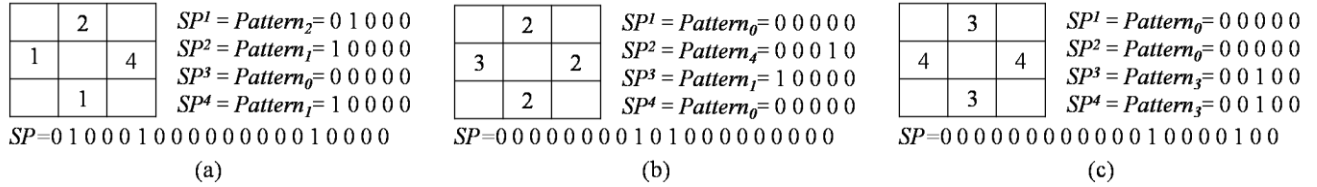


Fig. 5. Three examples to illustrate the computation of q patterns over each quantized shade for a particular pixel; in this example also the value of q is set to four.

The structuring pattern over each quantized shade SP^C is computed first and then combined it in order to find a single structuring pattern SP for any given pixel. In Fig. 5 (a), SP^1 is the pattern for the type 2 structure over shade 1; SP^2 is the pattern for the type 1 structure over shade 2; SP^3 is the pattern for the no structure over shade 3; SP^4 is the pattern of the type 1 structure over shade 4; and SP is the concatenation of the SP^1 , SP^2 , SP^3 , and SP^4 which represents the final structuring pattern obtained for middle pixel. The SP for the middle pixel in Fig. 5 (b-c) is also generated similarly. The length of SP becomes 20 for 4 quantized shades considering 5 structure elements. Generally speaking, the dimension of the structuring pattern SP for any pixel (x, y) in the image M of size $m \times n$ over Q quantized color considering S structure element is given by $Q \times S$ where $Q = q^3$. Mathematically SP for each pixel (x, y) of image M is defined by the following equation (i.e. concatenating structuring pattern over each shade for pixel (x, y)):

$$SP(x, y) = (SP^1(x, y), SP^2(x, y), \dots, SP^C(x, y), \dots, SP^Q(x, y)) \quad (4)$$

where, $SP^C(x, y)$ is the structuring pattern of pixel (x, y) over shade C in its local neighborhood and $C = [1, Q]$. $SP^C(x, y)$ over shade C for any pixel (x, y) is defined as:

$$SP^C(x, y) = \begin{cases} 00000, & \text{if no structure for shade } C \\ 10000, & \text{if type 1 structure for shade } C \\ 01000, & \text{if type 2 structure for shade } C \\ 00100, & \text{if type 3 structure for shade } C \\ 00010, & \text{if type 4 structure for shade } C \\ 00001, & \text{if type 5 structure for shade } C \end{cases} \quad (5)$$

Finally, we have the structuring pattern SP of dimension $Q \times S$ for each pixel of the image, now our goal is to combine these SP_s in an efficient manner in order to construct the final descriptor. We combine these SP_s for all pixels in the image by summing them for each bin individually; it means that the value at the j^{th} bin in the final descriptor is the number of 1's at the j^{th} bin in all SP_s (i.e. the sum of j^{th} bin of all SP_s). In this way, the size of the final descriptor remains $Q \times S$. The combined descriptor des for image M is mathematically defined as,

$$des = \sum_{\forall (x, y) \in M} SP(x, y) \quad (6)$$

The proposed descriptor construction process is quite different from other descriptors such as [17, 18, 24, 25, 26]. Note that the construction of proposed descriptor heavily depends upon the resolution of the image because it computes SP for each pixel and then add SP_s of all pixels in the image to find the final descriptor. The values of the pattern will be more for high resolution images and less for low resolution images. To cope with the problem of sensitivity to the image resolution and scale, we normalize the descriptor such that the sum of all elements of the descriptor becomes one. The final normalized rotation and scale-invariant hybrid descriptor (RSHD) is given as,

$$RSHD(Z) = \frac{des(z)}{\sum_{t=1}^{Q \times S} des(t)} \quad \forall z = [1, \quad Q \times S] \quad (7)$$

The final RSHD descriptor is invariant to the rotation because we constructed it by using rotation-invariant structure elements which makes it discriminative under rotation. Note that, the use of local neighboring information makes RSHD scale-invariant, too because the information in just the local neighborhood of a pixel does not change with the change in scale of the image. In this paper, the value of Q and S is set to 64 and 5 respectively, which lead to the size of RSHD descriptor is $64 \times 5 = 320$. The proposed image feature descriptor RSHD is an improvement over SEH [26]. SEH doesn't consider the neighboring information around a center pixel which we incorporated in the RSHD descriptor. In methodology also our descriptor is based on the binary representation of the structuring pattern, whereas SEH is based on the decimal representation of the local structures. The structure used in our method is rotation-invariant whereas it is not in the case of SEH.

4. Performance measures and Evaluation criteria

4.1. Performance measures

The RSHD descriptor can be used in those problems, where image description is needed. We evaluate proposed descriptor in content-based image retrieval, where the main goal is to find the most similar images of an image from a database of images. The similar images are extracted on the basis of the similarity score between the descriptor of query image and database images. The main problem in the evaluation is the computation of similarity score between two descriptors. We refer $F = \{f_1, f_2, \dots, f_{dim}\}$ as the feature descriptor for the images in the database and $T = \{t_1, t_2, \dots, t_{dim}\}$ as the feature descriptor of the query image where dim is the dimension of the feature. Various performance measure matrices produce different similarity score for the same set of image descriptors. Various performance measure matrices produce different similarity score for the same set of image descriptors. In this paper, we use two distance measures used in SEH [26] and CDH [13] respectively. The distance metric used in [26] is defined as,

$$d_{SEH}(T, F) = \sum_{i=1}^{dim} \frac{|f_i - t_i|}{1 + f_i + t_i} \quad (8)$$

The distance metric used in [13] is defined as,

$$d_{CDH}(T, F) = \sum_{i=1}^{dim} \frac{|f_i - t_i|}{|f_i + f_\mu| + |t_i + t_\mu|} \quad (9)$$

where $f_\mu = \sum_{i=1}^{dim} f_i / dim$ and $t_\mu = \sum_{i=1}^{dim} t_i / dim$. We will compare the performances using these two distance measures using proposed descriptor and accordingly decide which distance measure should be used in this paper.

4.2. Evaluation criteria

Most of the algorithms of image matching and image retrieval in the literature used *Precision* and *Recall* to measure the accuracy of the system. In content-based image retrieval, the main task is to find most similar images of a query image in the whole database. We also adopted *Precision vs Recall* curves to represent the effectiveness of proposed descriptor under image retrieval. The Precision and Recall for any given query image is defined by the following equation:

$$\begin{aligned} Precision &= \frac{\# \text{ num_similar}}{\# \text{ num_retrieved}} \\ Recall &= \frac{\# \text{ num_similar}}{\# \text{ num_database}} \end{aligned} \quad (10)$$

where, $\# \text{ num_similar}$ is the number of similar images retrieved, $\# \text{ num_retrieved}$ is the number of images retrieved, and $\# \text{ num_database}$ is the number of similar images in the whole database. We have considered 10 values of the $\# \text{ num_retrieved}$ from 5 to 50 on an interval of 5. $\# \text{ num_database}$ depends upon the dataset we are using. It is 100, 400 and 500 for each category in *Corel-1000*, *Corel-rotated* and *Corel-scale* dataset respectively. For *Corel-cbir* dataset $\# \text{ num_database}$ is not fixed, it is different for each category. For a particular data set, we select 10 images randomly from each class of the dataset and obtain *Average Precision* and *Recall* by taking the average of *Precision* and *Recall* obtained for each query image.

5. Experimental observation, result and discussion

This section presents the result obtained by applying various descriptors for content-based image retrieval. We test the robustness and discriminative power of proposed rotation and scale-invariant hybrid descriptor (RSHD) under rotation and scale transformations.

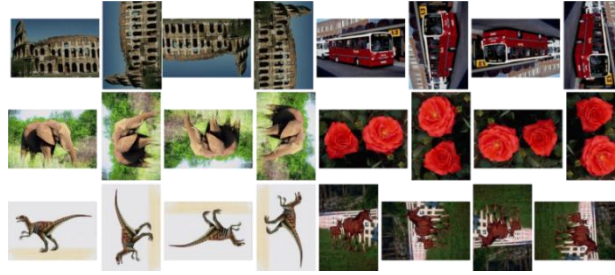


Fig. 6. Some example images of the synthesized *Corel-rotated* dataset containing original images and rotated images at 90, 180 and 270 degrees.

We also compare proposed descriptor with state-of-the-art descriptors for image matching in CBIR system. In this section, first we discuss about the Corel image matching dataset which is used in this paper for evaluation, and then we finally present the experimental result with discussion.

5.1. Data Sets

In order to evaluate the proposed descriptor, standard Corel Database for Content-based Image Retrieval [29] 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 dataset as *Corel-cbir* dataset. *Corel-cbir* dataset consists of the 80 categories having nearly more than 100 images in each category totaling 10800 images in the dataset. The images in a particular group are category-homogeneous with resolution either 120×80 or 80×120 . We also evaluate proposed descriptor on the *Corel-1000* dataset containing 1000 images from 10 categories having 100 images each category taken from [30]. The resolution of images is either 384×256 or 256×384 in *Corel-1000* dataset. The 10 categories of *Corel-1000* dataset are building, bus, dragon, elephant, flower, food, horse, human beings, landscapes and mountain. In order to emphasize the performance of proposed descriptor under rotation, we synthesized a *Corel-rotated* dataset by rotating all images of *Corel-1000* with angle 0, 90, 180, and 270 degrees. *Corel-rotated* dataset contains same 10 categories as of *Corel-1000* but the number of images in each category becomes 400 including rotated images; totaling 4000 images in *Corel-rotated* dataset. Fig. 6 shows some example images of the *Corel-rotated* dataset. We synthesized a *Corel-scale* dataset also by scaling all the images of *Corel-1000* at the scales of 0.5, 0.75, 1, 1.25, and 1.5. *Corel-scale* dataset contains 500 images for each category with 5000 total number of images.

5.2. Experimental Result

Color and local structures are the key components in the efficient description of the image features. RSHD uses both the information in an effective manner. We have shown the discriminating power of RSHD in Fig.7 for two sets of similar images of bus and flower type. The second and fourth columns represent the RSHD patterns for the images in the first and fifth columns respectively. The third column shows the difference of patterns in second and fourth columns. It is observed that the patterns for both similar images are also same and similarity is higher in the case of flower. The patterns of buses are also similar irrespective of the change in the color information in the middle of the buses.

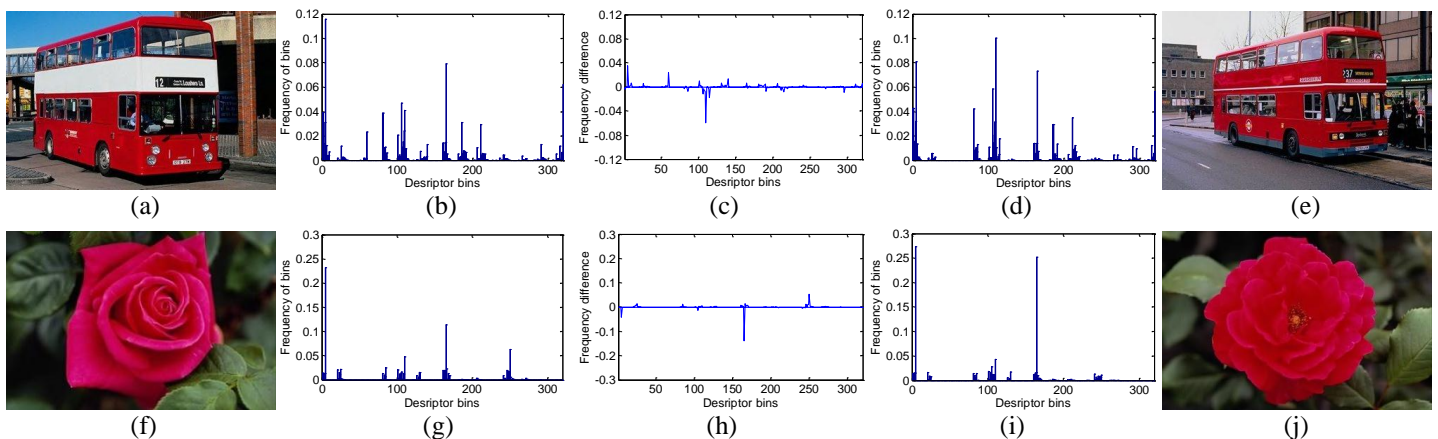


Fig. 7. Comparison of patterns obtained using RSHD descriptor for two sets of similar images. Plots (b), (d), (g) and (i) represent the RSHD patterns for the images in (a), (e), (f) and (j) respectively. Plots in (c) and (h) are the difference of (b-d) and (g-i) respectively.

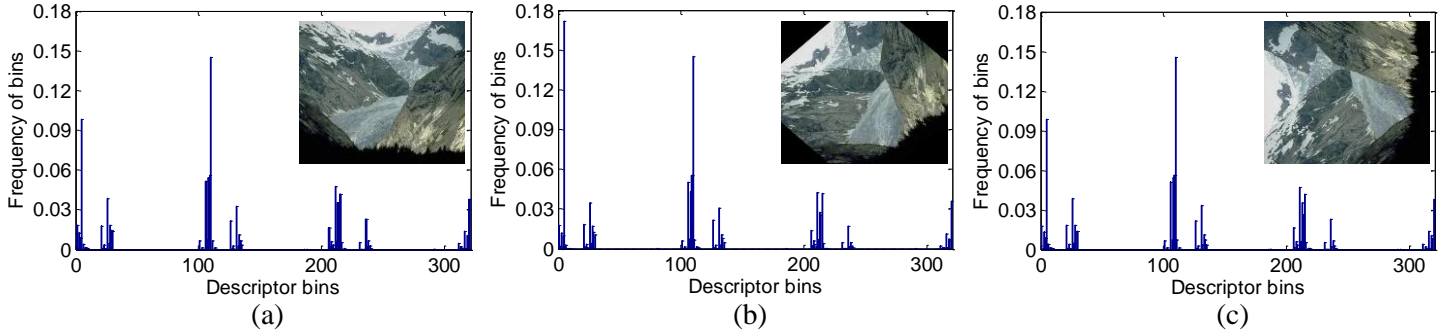


Fig. 8. Comparison of patterns obtained using the proposed descriptor under rotation, for (a) original image, (b) image rotated by 45° , and (c) image rotated by 90° .

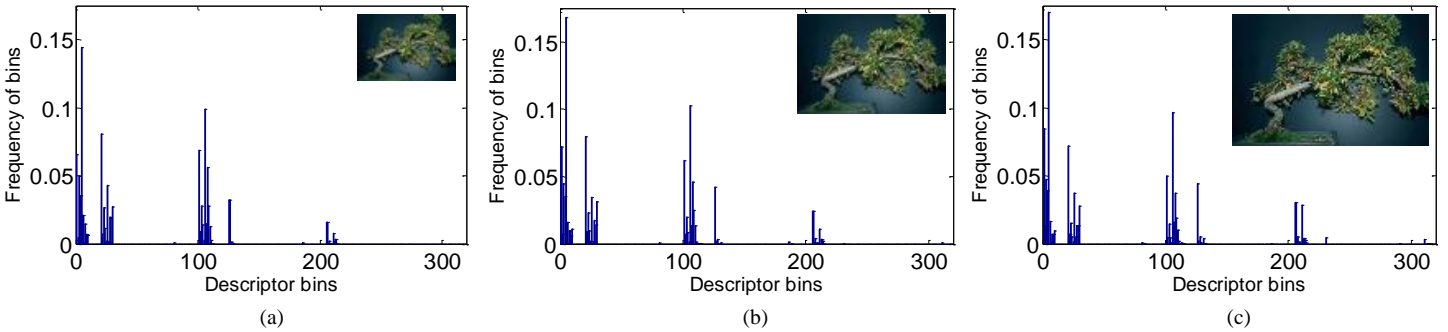


Fig. 9. Comparison of patterns obtained using RSHD under scaling condition, for (a) scaled at 0.5, (b) scaled at 0.75, and (c) original image.

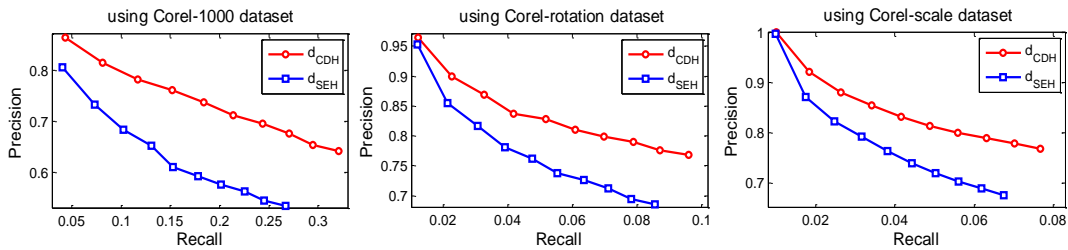


Fig. 10. Performance comparison of RSHD using different distance measure

RSHD descriptor is designed in such a way that it becomes rotation-invariant inherently. Fig. 8 depicts the pattern generated by the proposed method for an image and also for the rotated version of that image by 45 and 90 degrees. The patterns for original and 90° rotated image are exactly same while the patterns for original and 45° rotated image are not exactly same, but quite similar and the difference arises due to the inclusion of some black pixels in rotated version in the process of interpolation. We also depicted the patterns generated by RSHD for different scaled images in Fig. 8. The original image in Fig. 9(c) is scaled at 0.5 and 0.75 in Fig. 9(a-b) respectively. Due to use of neighboring information RSHD becomes scale-invariant as supported by Fig. 9. Image retrieval results are represented by Precision vs Recall pairs in all the plots by presenting Precision at y-axis and Recall at x-axis. We compared the results using proposed descriptor for distance measures d_{SEH} and d_{CDH} used in [26] and [13] respectively, over *Corel-1000*, *Corel-rotated* and *Corel-scale* datasets (see Fig. 10). The performance using d_{CDH} distance is better over each data set.

We have also evaluated proposed descriptor by changing some parameters and color space (see Fig. 11). Proposed descriptor is designed by uniform quantization of RGB color space into 64 bins, whereas, SEH [26] quantized HSV color space into 72 bins by giving more weightage to the Hue channel and CDH [13] quantized $L^*a^*b^*$ color space into 90 bins by giving more weightage to the L^* channel. We also used the quantization technique used by SEH and CDH and compared it with original RSHD on *Corel-1000*, *Corel-rotated* and *Corel-scale* data sets in Fig. 11 (a) and found that proposed descriptor is better in RGB color space quantization. The quantization technique used in SEH loses important details of the image (see Fig.12). We also observed that changing the number of quantization levels does not guidance any benefits as depicted in Fig. 11(b). If the number of quantized colors is reduced to 27 colors from 64 colors, the performance degrades too much. Whereas, the results for 125 quantized colors is better with the results of 64 quantized colors, but the dimension of descriptor with 125 quantized colors is too large. The RGB color quantization into 64 bins is used in the rest of the paper. The efficiency of RSHD also depends upon the size of structure element (i.e. radius of

neighboring pixels), and we have observed that increasing the radius of structure element reduces the efficiency of the descriptor except in the case of scaling as illustrated in Fig. 11(c). In the case of scaling, the performance of RSHD is better when the radius of structure element is 2. The just neighbors of a pixel provide more relevant local information in the neighborhood of that pixel. In the rest of the paper, the radius of the neighboring structure is considered as 1 to construct the RSHD descriptor.

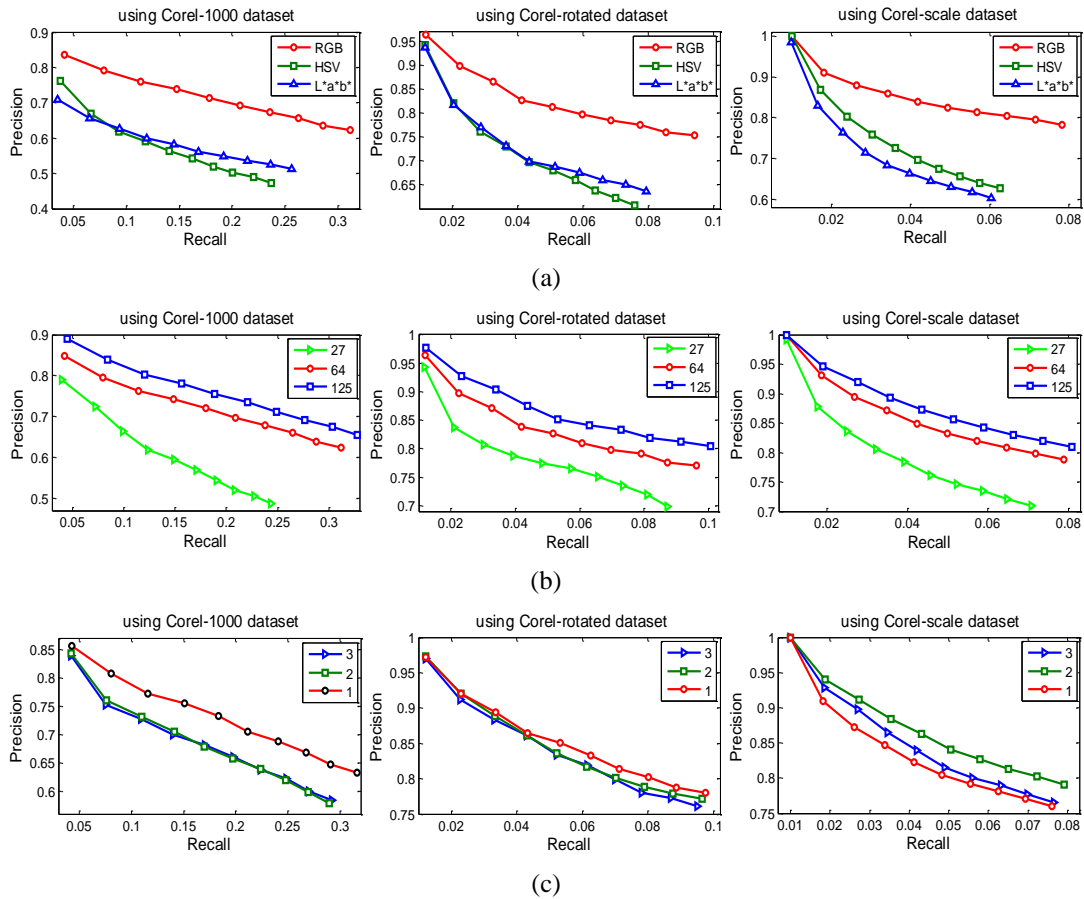


Fig. 11. Plots showing the effects of (a) color space, (b) number of quantization, and (c) radius of structure element.



Fig. 12. Visualization of RGB (64 bins) and HSV (72 bins) color quantization. The first two columns shows RGB images and RGB quantized images and last two columns shows HSV images with HSV quantized images. It can be observed that last two images in the second row lost various details of the original image as compared to the first two images in the second row.

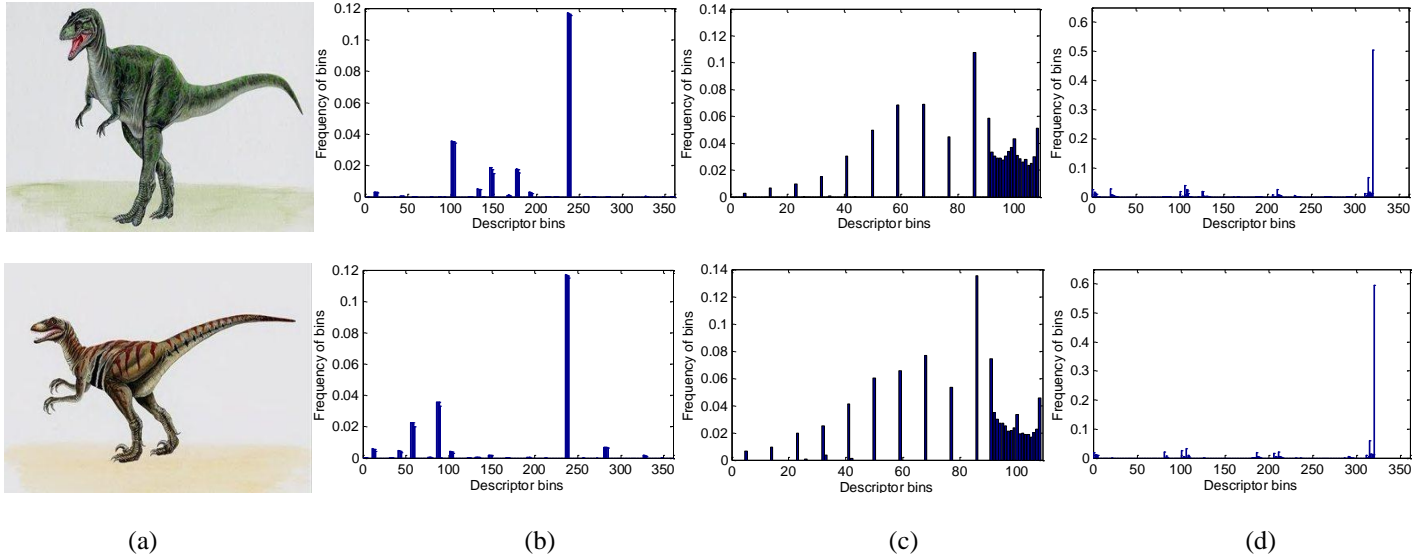


Fig. 13. Representing (a) two similar images, with (b) SEH, (c) CDH, and (d) RSHD descriptor

5.3. Comparison and Discussion

We compared the patterns extracted by SEH [26] and CDH [13] with RSHD for two similar dinosaur images in Fig. 13. The dimensions of SEH, CDH and RSHD descriptors are 360, 108 and 320 bins respectively. The difference between both images is mainly due to the color and orientation information. Due to the lack of uniform color quantization and rotation invariance, some patterns of first image got flipped and shifted in the second image using the SEH descriptor (see Fig. 13(b)). Whereas, proposed descriptor is able to produce nearly the same pattern for both the images (see Fig. 13(d)). Proposed descriptor also exhibits large value for some patterns only; it means that only some colors are dominating after color quantization, which makes proposed descriptor partial color-invariant for images of the same group, whereas in SEH patterns some more colors are also dominating. The patterns of CDH descriptor for both images are also similar (see Fig. 13(c)). Note that all the descriptors used in this paper are normalized for fair comparison among descriptors in such a way that the sum of the values of the descriptor is always 1.

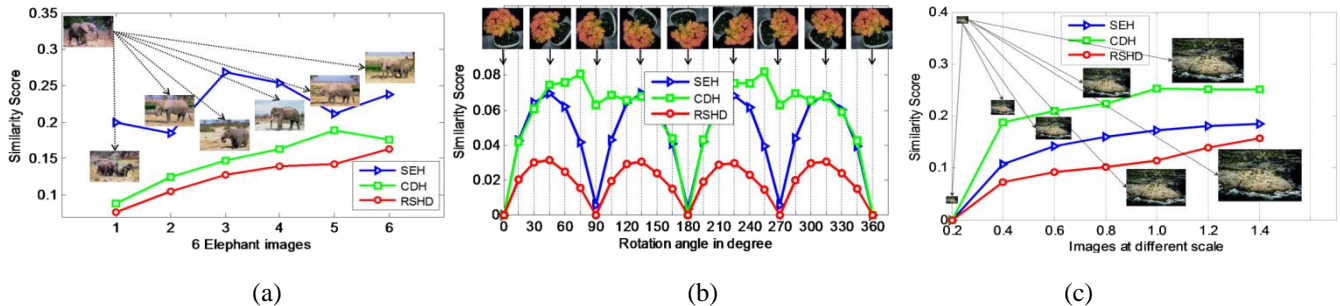


Fig. 14. Comparison of RSHD descriptor with SEH and CDH descriptors (a) over 7 elephant images, (b) under rotation, and (c) under scale change conditions.

Fig. 14(a) shows the similarity score for an elephant image with another 6 elephant image using SEH, CDH and RSHD. The similarity score decreases between more similar images and increases between less similar images. The similarity score obtained using RSHD is always less than SEH and CDH. The similarity score between the input image and third image increases using SEH due to the change in orientation and scale of objects within the image, whereas proposed method is able to maintain the low similarity score. Across the plots of Fig. 13 and Fig. 14(a), it is observed that RSHD is partial color and orientation-invariant. The partial color invariance is achieved due to the symmetric RGB color space quantization. We also compared RSHD with SEH and CDH considering rotation in the image as shown in Fig. 14(b). We compared with SEH and CDH because these are the descriptors proposed originally for the image retrieval recently. Due to partial rotation-invariant construction of SEH, it shows poor result if the amount of interpolation is increasing in rotation such as for angles 45° , 135° , 225° , and 315° . Whereas, proposed descriptor is able to cope with the problem caused by rotation very efficiently and performs better than other descriptors due to the inherent rotation-invariant construction of proposed descriptor. CDH also have not taken rotation into consideration and it's performs is not good under rotation.

Proposed method performs better because we have considered neighboring structuring pattern also to encode the textural information. To observe the performance of RSHD and SEH descriptor, we considered 7 images at different scale from 0.2 to 1.4 at an interval 0.2 as illustrated in Fig. 14(c). The image at scale 0.2 is compared with all images using SEH, CDH and RSHD descriptors. It is observed across the plot that proposed RSHD descriptor is having more similar patterns for scaled images compared to the patterns of the SEH and CDH descriptor. RSHD is better because the neighboring structures are less influenced by the scaling of the whole image.

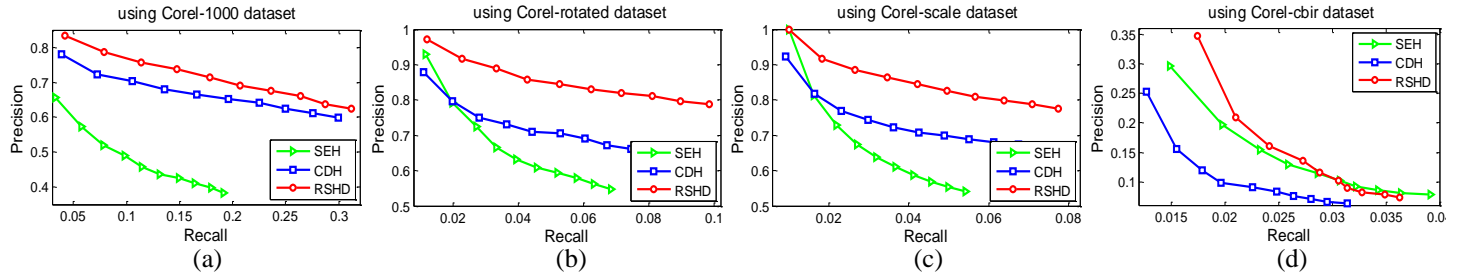


Fig. 15. Performance comparison of RSHD descriptor with other prominent feature descriptors such as SEH and CDH over (a) *Corel-1000*, (b) *Corel-rotated*, (c) *Corel-scale*, and (d) *Corel-cbir* datasets

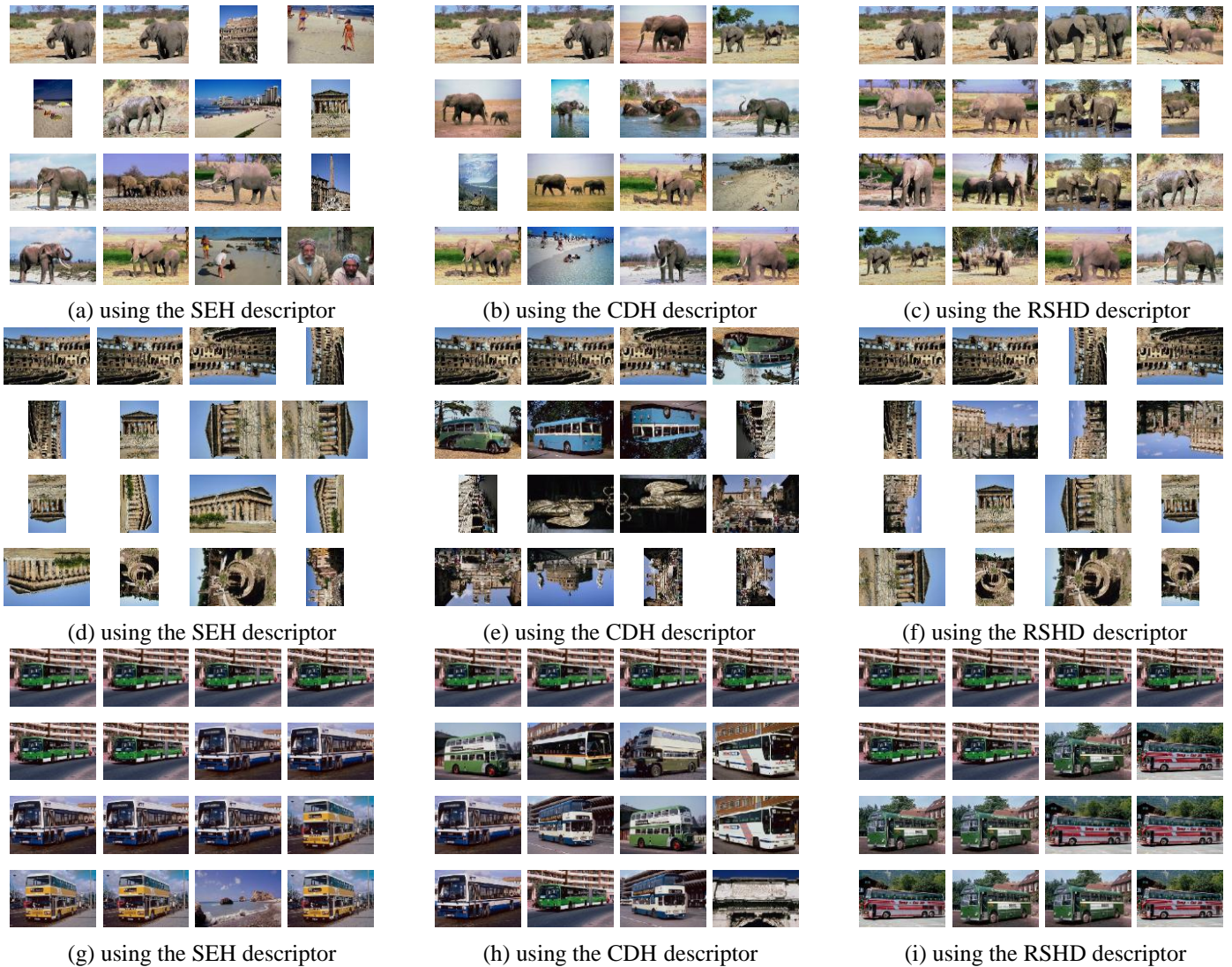


Fig. 16. Retrieval results using SEH, CDH and RSHD descriptors from (a-c) *Corel-1000*, (d-f) *Corel-rotated* and (g-i) *Corel-scale* datasets

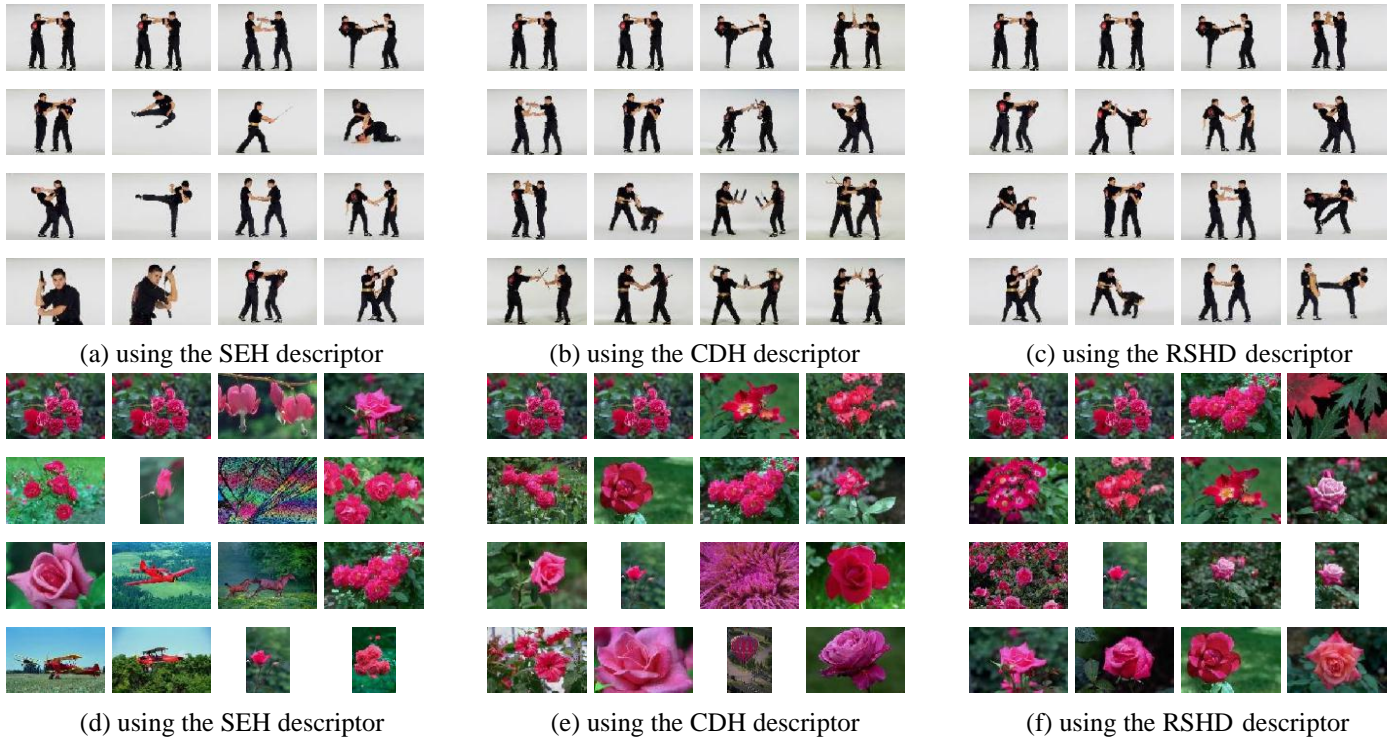


Fig. 17. Image retrieval results using SEH, CDH and RSHD descriptors from *Corel-cbir* dataset.

We also evaluated the descriptors on whole datasets (i.e. *Corel-1000*, *Corel-rotated*, *Corel-scale* and *Corel-cbir*), and results are presented in Fig. 15. CDH performs better than SEH if the number of categories in the database is less (i.e. for *Corel-1000*, *Corel-rotated* and *Corel-scale* dataset), but the efficiency of CDH degrades with increase in the number of image groups (i.e. for *Corel-cbir* dataset), whereas, proposed descriptor is able to maintain better retrieval performance than both descriptors over each data set. The use of neighboring structure elements on a rotation-invariant framework over quantized color boosts proposed RSHD descriptor and increases its discriminating power under rotation, scaling and orientation change.

Fig. 16 depicts three image retrieval results, one from *Corel-1000*, *Corel-rotated* and *Corel-scale* datasets using SEH (column 1), CDH (column 2) and RSHD (column 3) descriptors. We have retrieved 15 most similar images to the query image from each data set. The 1st image in each result of Fig.(16-17) is the query image and next 15 images are the retrieved images. The images in Fig. 16 (a-c) are retrieved from *Corel-1000* dataset corresponding to the query image using SEH, CDH and RSHD descriptors respectively. All the 15 images retrieved by RSHD are relevant to the query image (i.e. 100% precision), whereas, only 7 (precision 46.67%) and 12 (precision 80%) images retrieved by SEH and CDH respectively are relevant to the query image. It means that RSHD is benefited with the relevant color and textural information and represents the image more semantically. We also retrieved images from the *Corel-rotated* dataset using SEH, CDH and RSHD descriptors in order to show the efficiency under rotation transformation. Fig. 16 (d-f) present the images retrieved by each descriptor respectively from the *Corel-rotated* dataset. It is interesting to see that the similarity between the original image and its rotated images are more using RSHD descriptor, whereas, it is less similar in the case of SEH and CDH (see Fig. 16(d), 4th image of 4th row should be the rotated version of 2nd image 4th row). Considering this fact, we can say that the RSHD descriptor is robust to the orientation and rotation of the objects within the image. Fig.16 (g-i) depicts the retrieval results using each descriptor in the case of scaling using *Corel-scale* dataset. The output images are resized to fit and low scale images got blurred. Under scaling, for the considered example RSHD retrieved all the relevant images with precision 100%, while SEH and CDH are able to achieve only 93.33% precision by retrieval 1 irrelevant images. RSHD is able to retrieve all the scaled images (five bus images at different scales) whereas SEH and CDH failed to retrieve all the scaled images (see the second last retrieved image by SEH descriptor in Fig. 16(g)). The orientation of bus in all retrieved images using RSHD descriptor is same as of query image, whereas it is not same using CDH descriptor.

Fig. 17 shows the retrieval result in the situation of a large number of image types (i.e. 80 types in *Corel-cbir*) and shows the retrieved images for SEH, CDH and RSHD respectively for two query images of type fitness and flower. In the case of fitness the query image contains two person exercising in standing position and all the similar images returned by RSHD also contains two persons in nearly standing positions, whereas the images returned by the SEH contains some images having only one person and also with not in standing position and the images returned by the CDH contains some images having persons with some equipments which is not present in the query image. From this retrieval result, it is observed that RSHD can also match the number and orientation of the

objects within the image effectively. In the case of flower, RSHD is able to retrieve all 15 most similar images of type flower (i.e. precision = 1), whereas, SEH retrieved only 11 images of type flower (i.e. precision = 0.73) and CDH retrieved only 14 images of type flower (i.e. precision = 0.93). In this case SEH fails drastically, but RSHD is still performing very well because RSHD descriptor uses uniform RGB color quantization and rotation-invariant local structure pattern from the neighboring pixels of each pixel. From the above deductions, we can say that proposed rotation and scale-invariant hybrid image descriptor (RSHD) outperforms other state-of-the-art descriptors designed for image retrieval.

We also evaluated the average computation time required by each descriptor to extract the descriptor for an image from 100 randomly chosen images of *Corel-1000* (384×256 or 256×384 resolution images) and *Corel-cbir* (120×80 or 80×120 resolution images) data sets. The experiment is conducted 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 system. Table 1 presents the average time required by SEH, CDH and RSHD descriptors over *Corel-1000* and *Corel-cbir* datasets. It is evident that CDH requires more computation time than SEH and RSHD, whereas SEH is more computationally more efficient than RSHD. The proposed RSHD descriptor is nearly 100% slower than SEH, but it is nearly 400% faster than CDH.

Table 1: The average time required to extract the descriptor

Datasets	Descriptors		
	<i>SEH</i>	<i>CDH</i>	<i>RSHD</i>
<i>Corel-1000</i>	0.186	1.709	0.375
<i>Corel-cbir</i>	0.019	0.126	0.034

6. Conclusion

This paper presented an efficient image color and texture hybrid feature description for the content-based image retrieval. The proposed descriptor used the concept of structure element into local neighborhood of any pixel to achieve the inherent rotation invariance. RGB color space is quantized into 64 shades to represent the color feature of the image and local neighboring structure patterns are used to encode the textural information of the image. We extracted the structure pattern over each quantized shade separately to combine the color properties with texture information very effectively. In this way, proposed descriptor captures the more relevant relationship information among each color in the local neighborhood. Proposed descriptor is inherently rotation-invariant and describes the image features more efficiently. Our experimental results on Corel image matching datasets including rotated and scaled one suggest that proposed RSHD descriptor performs better than other descriptors and can be effectively applied in the content-based image retrieval. RSHD outperforms state-of-the-art descriptors and more robust, especially in the case of rotation and scaling.

References

- [1]. G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, C. Roux, Fast wavelet-based image characterization for highly adaptive image retrieval, *IEEE Transactions on Image Processing* 21 (4) (2012) 1613–1623.
- [2]. S. Jones, L. Shao, Content-based retrieval of human actions from realistic video databases, *Information Sciences* 236 (2013) 56–65.
- [3]. M.S. Zarchi, A. Monadjemi, K. Jamshidi, A semantic model for general purpose content-based image retrieval systems, *Computers & Electrical Engineering* 40 (7) (2014) 2062–2071.
- [4]. N. Singh, S.R. Dubey, P. Dixit, J.P. Gupta, Semantic Image Retrieval by Combining Color, Texture and Shape Features, In *International Conference on Computing Sciences* (2012) 116–120.
- [5]. D.S. Zhang, M.M. Islam, G.J. Lu, A review on automatic image annotation techniques, *Pattern Recognition* 45 (1) (2012) 346–362.
- [6]. R.D. Gupta, J.K. Dash, S. Mukhopadhyay, Rotation invariant textural feature extraction for image retrieval using eigen value analysis of intensity gradients and multi-resolution analysis, *Pattern Recognition* 46 (12) (2013) 3256–3267.
- [7]. S.R. Dubey, S.K. Singh, R.K. Singh, Local Diagonal Extrema Pattern: A New and Efficient Feature Descriptor for CT Image Retrieval, *IEEE Signal Processing Letters* 22 (9) (2015) 1215–1219.
- [8]. S.M. Mukane, S.R. Gengaje, D.S. Bormane, A novel scale and rotation invariant texture image retrieval method using fuzzy logic classifier, *Computers & Electrical Engineering* 40 (8) (2014) 154–162.
- [9]. S.K. Vipparthi, S.K. Nagar, Multi-joint histogram based modelling for image indexing and retrieval, *Computers & Electrical Engineering* 40 (8) (2014) 163–173.
- [10]. N. Shrivastava, V. Tyagi, An efficient technique for retrieval of color images in large databases, *Computers & Electrical Engineering* (2014).

- [11]. S.R. Dubey, S.K. Singh, R.K. Singh, Local neighbourhood-based robust colour occurrence descriptor for colour image retrieval, *IET Image Processing* (2015).
- [12]. W.T. Chen, W.C. Liu, M.S. Chen, Adaptive color feature extraction based on image color distributions, *IEEE Transactions on Image Processing* 19 (8) (2010) 2005–2016.
- [13]. G.H. Liu, J.Y. Yang, Content-based image retrieval using color difference histogram, *Pattern Recognition* 46 (1) (2013) 188–198.
- [14]. X.Y. Wang, Y.J. Yu, H.Y. Yang, An effective image retrieval scheme using color, texture & shape features, *Computer Standards and Interfaces* 33 (1) (2011) 59–68.
- [15]. M.J. Hsiao, Y.P. Huang, T. Tsai, T.W. Chiang, An efficient and flexible matching strategy for content-based image retrieval, *Life Science Journal* 7 (1) (2010) 99–106.
- [16]. C.H. Lin, D.C. Huang, Y.K. Chan, K.H. Chen, Y.J. Chang, Fast color-spatial feature based image retrieval methods, *Expert Systems with Applications* 38 (9) (2011) 11412–11420.
- [17]. T. Ojala, M. Pietikainen, T. Maenpa, Multi-resolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (7) (2002) 971–987.
- [18]. M. Heikkilä, M. Pietikäinen, C. Schmid, Description of interest regions with local binary patterns, *Pattern Recognition* 42 (3) (2009) 425–436.
- [19]. C. Zhu, R. Wang, Local multiple patterns based multi-resolution gray-scale and rotation invariant texture classification, *Information Sciences* 187 (2012) 93–108.
- [20]. S.R. Dubey, A.S. Jalal, Detection and Classification of Apple Fruit Diseases Using Complete Local Binary Patterns, In: *Third IEEE International Conference on Computer and Communication Technology* (2012) 346-351.
- [21]. Z. Wang, B. Fan, F. Wu, Local Intensity Order Pattern for Feature Description, in: *IEEE International Conference on Computer Vision* (2011) 603-610.
- [22]. S.R. Dubey, S.K. Singh, R.K. Singh, Rotation and Illumination Invariant Interleaved Intensity Order Based Local Descriptor, *IEEE Transactions on Image Processing*, 23 (12) (2014) 5323-5333.
- [23]. S.R. Dubey, S.K. Singh, R.K. Singh, A multi-channel based illumination compensation mechanism for brightness invariant image retrieval, *Multimedia Tools and Applications* (2014) 1-31.
- [24]. G.H. Liu, L. Zhang, Y.K. Hou, Z.Y. Li, J.Y. Yang, Image retrieval based on multi-texton histogram, *Pattern Recognition* 43 (7) (2010) 2380–2389.
- [25]. G.H. Liu, Z.Y. Li, L. Zhang, Y. Xu, Image retrieval based on micro-structure descriptor, *Pattern Recognition* 44 (9) (2011) 2123–2133.
- [26]. W. Xingyuan, W. Zongyu, A novel method for image retrieval based on structure elements descriptor, *Journal of Visual Communication and Image Representation* 24 (1) (2013) 63–74.
- [27]. X.Y. Wang, H.Y. Yang, D.M. Li, A new content-based image retrieval technique using color and texture information, *Computers & Electrical Engineering* 39 (3) (2013) 746-761.
- [28]. S.M. Youssef, ICTEDCT-CBIR: Integrating curvelet transform with enhanced dominant colors extraction and texture analysis for efficient content-based image retrieval, *Computers & Electrical Engineering* 38 (5) (2012) 1358-1376.
- [29]. <https://sites.google.com/site/dctresearch/Home/content-based-image-retrieval>.
- [30]. <http://wang.ist.psu.edu/docs/related/>.