



Content-based Image Retrieval Considering Colour Difference Histogram of Image Texture and Edge Orientation

M. M. AlyanNezhadi^{*a}, H. Qazanfari^a, A. Ajam^b, Z. Amiri^a

^a Image Processing and Data Mining Lab, Shahrood University of Technology, Shahrood, Iran

^b Department of Computer Engineering, Shahrood Non-profit and Non-government Higher Edu. Inst, Shahrood, Iran

PAPER INFO

Paper history:

Received 09 February 2020

Received in revised form 06 April 2020

Accepted 06 April 2020

Keywords:

Content-based Image Retrieval

Colour Difference

LBP Texture

Edge Orientation

Colour Difference Histogram

Data Mining

Entropy

ABSTRACT

Content-based image retrieval is one of the interesting subjects in image processing and machine vision. In image retrieval systems, the query image is compared with images in the database to retrieve images containing similar content. Image comparison is done using features extracted from the query and database images. In this paper, the features are extracted based on the human visual system. Since the human visual system considers the texture and the edge orientation in images for comparison, the colour difference histogram associated with the image's texture and edge orientation is extracted as a feature. In this paper, the features are selected using the Shannon entropy criterion. The proposed method is tested using the Corel-5K and Corel-10K databases. The precision and recall criteria were used to evaluate the proposed system. The experimental results show the ability of the proposed system for more accurate retrieval rather than recently content-based image retrieval systems.

doi: 10.5829/ije.2020.33.05b.28

1. INTRODUCTION

In recent years, due to the growth of technology and extensive use of media, the number of digital images has been growing at a tremendous speed. This led to serious difficulties in their storage and retrieval. Digital image retrieval is the task of finding images similar to the query based on perceptual and visual features [1].

Earlier image retrieval methods use a text-based management system. However, modern methods use low-level features, such as colour and texture instead of textual information, labels or tags. Since low-level features cannot satisfactorily capture the proper image content, systems based on these features are not able to retrieve the similar images with high accuracy [1, 2]. This problem is related to the concept of semantic gap. Semantic gap is a serious challenge, defined by the difference between the perception of machine and human from the image. Researchers in this field are trying to reduce this gap in order to increase the

compatibility between the perception of machine and human.

Figure 1 illustrates the semantic gap problem and several ways to reduce it. Typically, methods for reducing the semantic gap are in two types: methods with combination of low-level features, and methods with machine learning techniques such as object recognition [3]. Both of these two types have their own advantages and disadvantages. As seen in Figure 1, using low-level features is not enough to capture the visual and perceptual content of the image. On the other hand, methods with low-level features have low time complexity.

A combination of these two types of methods is also used to narrow the semantic gap. This combination uses relevance feedback along with the features so that an interface is provided between the user and the system.

Each time the system retrieves images, the user marks the correct ones. Then, machine learning methods are employed to learn and improve the retrieval results. The process is repeated, thus the results provided by the system are greatly improved [4,5].

*Corresponding Author Email: Alyan.nezhadi@gmail.com (M. M. AlyanNezhadi)

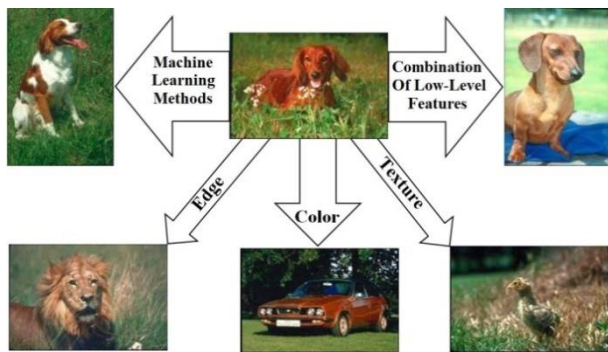


Figure 1. Results of various image retrieval methods on a query image

Methods equipped with a machine learning technique detect and extract high-level features such as objects. The most importance disadvantage of these methods is time complexity in detecting high-level features.

Various studies shown that the human visual system is very sensitive to the colour and texture of the images [3,6]. Hence, in this paper the semantic features are extracted from the texture, colour, and edge orientation of the images. The idea of using these features was taken from the Colour Difference Histogram (CDH) method proposed in [7]. The well-known algorithm for extracting texture features is local binary patterns (LBP) [8]. LBP has two great advantages, non-parametricity and simplicity of calculation.

In the proposed approach, the texture is extracted from the image and the edge orientation of each pixel in the image. Between two adjacent pixels with the same texture and edge orientation, the colour difference is calculated. Histogram of this difference includes the visual and perceptual content of the image with significant details. The proximity of the HSV colour space to the human vision system allows extracting more descriptive features in relation to the image content.

In order to increase accuracy of the retrieval, as well as reducing the computation time, the superior features are selected using the Shannon entropy criterion.

The proposed method is compared with the existing Content based Image Retrieval (CBIR) methods such as HCSF[3], CDH[7], MSD[9], and the methods based on colour and texture [10], as well as LMDEP [11] and CDHT[12]. The results show ability of the proposed method compared to the other existing CBIR methods.

2. RELATED WORKS

As aforementioned, methods for reducing the semantic gap between the difference in the perception of images by a machine and a person are divided into two

categories. The first category uses machine learning methods while the second one achieves the semantic content of images using low-level features. The advantages of the second category compared to the first include the lack of segmentation, object recognition and training processes. The methods in the second category use low-level features in such a way that the extracted information is suggested to be close to the one which is extracted from the human visual system.

Some image retrieval methods in each category are presented in the following. The most famous methods from the first category include classification of different images with supervised learning methods [3,13], image segmentation [14], deep learning methods to detect objects from different areas of the image [15], and neural networks for pattern learning from samples [16]. Deep learning refers to machine learning methods that have a hierarchical architecture through which high-level features are made from low-level features for decision-making or modelling. It is worth mentioning that the results of the employing of the convolutional neural network (CNN) in recent works have shown its effectiveness in image retrieval. The obtained features using the CNN can be effective in many visual recognition tasks [17-19].

In [20], local image descriptors are employed to find a set of the best k-similar images. To do this, an open access library called fast library for close neighbours was used [20]. The most suitable method for an approximate search for the nearest neighbour is an approach based the product quantization introduced in [21], which provides an accuracy fast image retrieval method.

To enhance the retrieval performance, Zhang et al. [22] proposed a method with merging the local and global features, which leads to a significant improvement in their results. They have proposed a graph-based query-specific fusion approach where multiple retrieval sets are merged and re-ranked by conducting a link analysis on the fused graph.

In [23], a content-based deep learning neural network was applied. This method uses a pre-trained neural network that compares a class-based histogram between database and query images. In [24], a Bayesian structure is used to classify each image.

In the following, a brief description of methods based on a combination of low-level features is presented. In [25], the authors used a combination of colour and texture features, while in [26,27] in addition to these two types of features, shape features are also employed. In [28], a colour image is encoded into a sequence of letters, which is used as a feature vector for image retrieval.

Liu et al. in [9] used a histogram called a microstructure descriptor (MSD) to extract texture information. This histogram is generated based on the

colour of adjacent pixels with a similar edge orientation in the HSV colour space. Colour values corresponding to the microstructure of the image serve as feature of histogram in order to retrieve the image.

For a long period, global features formed the basis of CBIR systems, but recently local features have begun to take their place, giving greater efficiency in describing image content. Unfortunately, some factors negatively affect global descriptors, such as factors associated with occlusion, viewpoint change, illumination change and local characteristics of image shape. At the same time, these factors do not affect local signs as strongly as global signs. The most effective local features are DAISY [29], Rank-SIFT [30], BRIEF [31], ORB [32].

Most prominent local features use SIFT as a basis. Such features can effectively work even at the presence of perspective distortions, occlusion, and geometric transformation.

Recently, global descriptors with combination of local features have been developed, and they have shown effectiveness in image retrieval. The most widely used known methods are LBP, CNN, and the bag-of-visual-words (BoVW). The BoVW methods were actually designed to retrieve text which use key points and salient patches [33,34]. Methods based on BoVW, which were obtained from local features, such as key points and salient patches, were proposed for object recognition and scene categorization [33,35-37].

Most of the global descriptors are defined based on texture and shape. The most visited and widespread feature is LBP [8] and its variants, which can be found in [38].

In [39] the authors used Multi Texton Histogram (MTH) descriptor to simulate the human visual system. MTH uses the first- and second-order statistics to analyse the texton, and therefore, the texture extraction is more accurate. MTH can present the spatial correlation of colours and edges using textons analysis. MTH uses the spatial correlation of pixels in the same neighbourhood based on four special texton types. In this method, four models of the texton move on the matrix of the quantized colour values. If a texton is matched on a part of color values, its colour values will be stored as a texton matrix. The edge orientation is also calculated per pixel and stored in the edge matrix. Finally, given the two matrices of the texton and the edge orientation, if the value of the pixel P in the texton matrix is equal to its neighbours, the value of the edge orientation of the pixel P is increased by one. In addition, if the edge values of these two pixels are equal, the value of the pixel P in the texton matrix will be increased by one.

In [10], researchers proposed a CBIR approach that uses a combination of colour and texture attributes. The textural features are block difference of inverse

probabilities (BDIP) and block variation of local correlation coefficients (BVLC). In addition, as colour attributes, features are extracted from the colour histogram in the HSV space. The BDIP extracts texture features by applying a 2x2 pixel window to the brightness channel, obtaining the edges and boundaries of the regions to create features set. In fact, the BDIP uses the difference in local brightness of the image to measure a texture feature. BVLC is one of the fast and accurate CBIR approaches that measures the smoothness level of local textures as texture features.

Texture descriptors are often used to retrieve images. For example, authors in [11] proposed a descriptor called the local mean differential excitation pattern (LMDeP). The underlying idea of this descriptor is to illuminate differential excitation using the average of the points for each angular and radial neighbouring points. Thus, the obtained features will be more reliable, since the noise effect of adjacent pixels is reduced.

The edge features and colour information together can create an effective vector of features, as suggested in [7], where these attributes are extracted in the $L^*a^*b^*$ colour space. The content of the image can be determined using the colour difference between two adjacent pixels. The best-known descriptor of this kind is called CDH, which considers the colour difference between two adjacent pixels with the same properties. Thus, each pixel is compared with adjacent pixels to produce a CDH.

In [6], the authors proposed a CBIR approach that uses a combination of colour and edge features. This paper demonstrated the closest colour space to the human visual system is HSV colour space. The colour histogram as colour features is extracted from HSV colour space. The edge features also are extracted using the Sobel operator in the HSV colour space.

Ajam et al. in [12] used a data mining method along with low-level features for image retrieval. In this paper, the researchers proposed a CBIR system that extracts colour difference histogram in the texture image. First, texture features are extracted by the LBP method. Then, by texture map for any image, the colour difference between two neighbour pixels is calculated. Finally, these values make the colour difference histogram for any image.

Colour space plays very important role in feature extraction step. In this study, it is shown that the extraction of the proposed features from the HSV colour space represents the semantics of the image in more detail than in the $L^*a^*b^*$ colour space. An overview on image retrieval methods can be found in [40].

3. THE PROPOSED METHOD

In the following, details of the proposed method for the content-based image retrieval system are provided. The

general scheme of the method is illustrated in Figure 2. Initially, the information about the texture and edges orientation of the query image and those that are in the database is extracted using the LBP algorithm and Sobel operator, respectively.

Then, for adjacent pixels that have the same texture and edge orientation, the colour difference is calculated. A histogram of this difference is computed for each image. As aforementioned, superior features are selected by applying the Shannon entropy. The selected features are used in a vector. In order to calculate the similarity between the query image and each of the database images, the similarity between their feature vectors is calculated using similarity criterion. Each steps of the proposed method is discussed in more details in the following.

3. 1. Image Texture Using LBP

LBP is one of the widely used and known methods for extracting texture features [8]. This operator generates a binary number with respect to the 3x3 values of the neighbouring pixels of the target pixel. The binary value of each neighbouring pixel is obtained by comparing its value with the central pixel of the window and applying a threshold. If the neighbouring pixel has a greater or equal value compared to the central pixel, then the binary value is indicated by one, otherwise it is set to zero. These binary values of each neighbouring pixel are arranged in a circle and an eight-bit binary number is formed. An example of the binary number computation can be seen in Figure 3. Information about the structure of the image is extracted by a moving window throughout the image. After the texture features are obtained, their value is discriminated on 32 levels. The reason for choosing the discrimination level is depicted in the experimental results in subsections 4.3.

3. 2. Edge Orientation Using Sobel Operator

As aforementioned, the human visual system is very sensitive to the edge orientation of the image. One of the conventional methods to extract the edge orientation from an image is the Sobel operator. In this operator, unit vectors a, b, and c are located along the axes A, B, and C in the colour space to effectively detect edges caused by colour changes in the colour space of perception. Then Equations (1) and (2) are calculated for a colour image $f(x, y)$.

$$u = \frac{\partial A}{\partial x} a + \frac{\partial B}{\partial x} b + \frac{\partial C}{\partial x} c \tag{1}$$

$$v = \frac{\partial A}{\partial y} a + \frac{\partial B}{\partial y} b + \frac{\partial C}{\partial y} c \tag{2}$$

In the following equations g_{xx} , g_{yy} and g_{xy} are defined based on the internal products of the previous vectors:

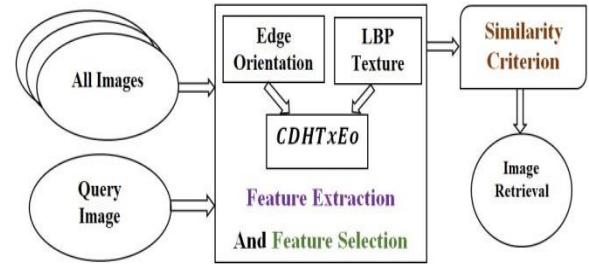


Figure 2. The block diagram of the proposed CBIR system

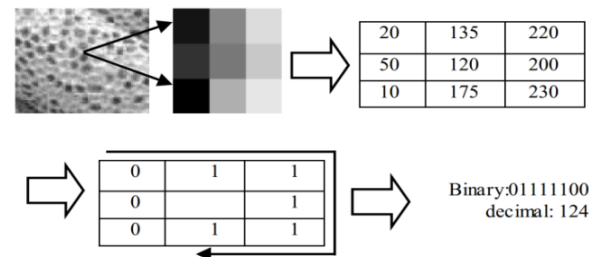


Figure 3. An example of LBP operation for texture analysis [12]

$$g_{xx} = u^T u \tag{3}$$

$$g_{yy} = v^T v \tag{4}$$

$$g_{xy} = u^T v \tag{5}$$

The necessary partial derivatives for the implementation of Equations (3)-(5) can be calculated using the Sobel operator. If the pixel $I(x, y)$ is in the HSV colour space, then the change orientation of $I(x, y)$ based on the previous notation is as follows:

$$\varphi(x, y) = \frac{1}{2} \tan^{-1} \frac{2g_{xy}}{g_{xx} - g_{yy}} \tag{6}$$

After the computation of the edge orientation $\varphi(x, y)$ for each pixel, their values are uniformly discretized into m bins. The best value for m in the HSV colour space is 16 based on our experiments (details of the experiments are given in Section 4.3).

3. 3. Perceptually Uniform Colour Difference Histogram

To extract features, a histogram of the difference in texture, colour and image edge orientation is employed. The perceptually uniform colour difference between image colours, textures, and edge orientation is a wide range of visual information for analysing image content. Since the HSV colour space is more uniform in perception and closer to the human vision system, it was chosen with the aim of calculating the difference in colour between the two pixels. Channels H, S, and V were calculated using the transformation from the space of the RGB to the HSV

[7]. To enhance expression of the differences in colour, the channels of this space were discriminated into several levels. The range of the channel H is uniformly divided into 15 levels, while channels C and V are divided into five levels. Thus, the entire space has $15 \times 5 \times 5 = 375$ different values. The quantization level of each channel was assigned experimentally (see Section 4.3).

In this paper, three types of feature sets are considered. 1) Color related feature set, 2) Edge orientation related feature set, and 3) Texture related feature set. Equations (7)-(10) are used to calculate the color, edge orientation and texture feature sets. As it's shown in Equation (7), to calculate the color-related features only pixels with the same texture are considered. The texture features of pixels with the same color and same edge orientation can be calculated using Equations (8) and (9), respectively. Finally, the edge orientation features of pixels with the same texture is calculated using Equation (10).

Since in the proposed method, texture information plays an important role in an accurate image retrieval, another texture feature set is also considered where it is calculated using the same formula but when the image is quantized to different levels. Therefore, we have four feature sets.

The colour value of the pixel with coordinates (x,y) is denoted by $C(x,y)$, which belongs to the set $M = \{0, \dots, m-1\}$. The value of the texture of pixel (x,y) that was obtained through the LBP operator is denoted by $T_p(x,y)$ and $T_q(x,y)$ and it is an integer number from the set $N_c = \{0, \dots, N_c - 1\}$ and $N_o = \{0, \dots, N_o - 1\}$. The $T_p(x,y)$ and $T_q(x,y)$ are texture values of a point (x,y) with different quantization levels.

For edge orientation, the value of a pixel located in the coordinate (x,y) is calculated using the Sobel operator and denoted by $O(x,y)$, which can take a value from the set $O = \{0, \dots, o-1\}$.

Suppose that distance between two adjacent pixels is denoted by D , then the equations for calculating difference in texture colour and edge orientation between two adjacent pixels are presented as follows:

$$H_{c-tx}(C(x,y)) = \begin{cases} \sum \sum \sqrt{(\Delta H)^2 + (\Delta S)^2 + (\Delta V)^2} \\ \text{where } T_p(x,y) = T_p(x',y'); \\ \max(|x - x'|, |y - y'|) = D \end{cases} \quad (7)$$

$$H_{tx-c}(T_p(x,y)) = \begin{cases} \sum \sum \sqrt{(\Delta H)^2 + (\Delta S)^2 + (\Delta V)^2} \\ \text{where } C(x,y) = C(x',y'); \\ \max(|x - x'|, |y - y'|) = D \end{cases} \quad (8)$$

$$H_{tx-o}(T_q(x,y)) = \begin{cases} \sum \sum \sqrt{(\Delta H)^2 + (\Delta S)^2 + (\Delta V)^2} \\ \text{where } O(x,y) = O(x',y'); \\ \max(|x - x'|, |y - y'|) = D \end{cases} \quad (9)$$

$$H_{o-tx}(O(x,y)) = \begin{cases} \sum \sum \sqrt{(\Delta H)^2 + (\Delta S)^2 + (\Delta V)^2} \\ \text{where } T_q(x,y) = T_q(x',y'); \\ \max(|x - x'|, |y - y'|) = D \end{cases} \quad (10)$$

where the difference between two points in channels H, S and V are denoted by ΔH , ΔS and ΔV , respectively. In the above equations, (x',y') is each of eight neighbouring pixels for pixel (x,y) . The first formula represents that if two adjacent pixels located at a distance D and have the same texture ($T_p(x,y)$), their colour difference, that is $C(x,y)$, can be calculated based on Equation (7). Therefore, the H_{c-tx} component is a vector with 375 features.

Similarly, the colour difference of two adjacent pixels with distance D , which have the same colour value, can be calculated using Equation (8) and is denoted by $T_p(x,y)$. Thus, H_{tx-c} is a vector with 32 features. Another situation occurs when two adjacent pixels with a distance D have the same edge orientation value and it is necessary to calculate the colour difference for them. Equation (9) provides this opportunity and the colour difference is indicated by $T_q(x,y)$. Therefore, H_{tx-o} is a feature vector with 36 values.

When two adjacent pixels with distance D have the same texture value ($T_q(x,y)$), the colour difference between them is calculated using Equation (10). This difference is denoted by $O(x,y)$. It should be mentioned that H_{o-tx} is a vector with 16 attributes.

Equation (7) is different from Equation (10). In fact, in Equation (7), the texture of the image is quantized to 32 levels, while in Equation (10), it is quantized to 36 levels.

With the combination of the obtained feature vectors H_{c-tx} , H_{tx-c} , H_{tx-o} , and H_{o-tx} , the final feature vector is created, which has $375 + 32 + 36 + 16 = 459$ features, denoted by $H_{CDHTxEo}$ (Equation (11)).

$$H_{CDHTxEo} = \begin{bmatrix} H_{\text{color}}(0), H_{\text{color}}(1), \dots, H_{\text{color}}(M-1), \\ H_{\text{tx}_c}(0), H_{\text{tx}_c}(1), \dots, H_{\text{tx}_c}(N_c-1), \\ H_{\text{tx}_o}(0), H_{\text{tx}_o}(1), \dots, H_{\text{tx}_o}(N_o-1), \\ H_o(0), H_o(1), \dots, H_o(O-1) \end{bmatrix} \quad (11)$$

By conducting computational experiments with different values of the quantization level for a better image retrieval, parameters D , M , N_c , N_o and O were experimentally assigned as 1, 375, 32, 36, and 16, respectively.

4. EXPERIMENTS AND EVALUATION

This section provides details about the experiments, as well as criterion to make decision on similarity between images. We will also consider the method of feature

selection and choosing the best values for quantizing colour, texture and edge orientation. The developed image retrieval system was implemented in Visual Studio 2015 (with C# language), which is available at <https://github.com/hamedmit/Content-based-Image-Retrieval.git> and www.smartcbir.nph-co.ir.

4. 1. Image Datasets and Similarity Criterion

In most image retrieval researches, Corel-5K and Corel-10K datasets were used, which have five and ten thousand images, respectively. The datasets are available in [41]. The classes of these datasets contain a variety of images with relatively complex content, which make them an excellent choice for this experiment.

After the feature extraction, two factors impact on the decision about the similarity between the images of the query and the database: feature selection and similarity criterion. In fact, the criterion assesses the similarity between two images by measuring some characteristics between their two feature vectors. Even when extracted features are significant, descriptive and superior, choosing the wrong criterion will result in inefficient image retrieval, lower quality and rate of the retrieval.

In this paper, the similarity criterion was taken from [3], which is an improved version of the Canberra criteria and is defined as follows:

$$D(A, B) = \sum_{i=1}^Z \frac{|A_i - B_i|}{|A_i + M_A| + |B_i + M_B|} \quad (12)$$

where D is the similarity distance between the vectors A and B. The values of M_A and M_B represent the average values of the feature vector A and B, respectively. Z is the number of features. To compare and evaluate the proposed system with some recent methods, two standard criteria for precision and recall were used. The number of images retrieved at each stage was assigned 12. Precision (P) and recall (R) criteria are presented as follows:

$$P = \frac{K}{N} \quad (13)$$

$$R = \frac{K}{M} \quad (14)$$

where K is the number of correctly retrieved images, N is the total number of images retrieved at each stage and M is the total number of images belonging to the request image class. For the implementation of the comparison between the proposed method with other existing methods, the Corel-5k and Corel-10k databases with the parameters $M = 100$ and $N = 12$ were employed. The higher the average precision and recall values indicates the better the image retrieval performance.

4. 2. Features Selection using Shannon Entropy

This section depicts applying the Shannon entropy for

removing inefficient features that do not affect on the retrieval rate. Shannon entropy can be interpreted as the mean level of "information". When a feature has low entropy, it means that the feature does not have enough information. The great value of the entropy for a feature expresses the importance of information contained in the feature. Entropy for vector features X can be calculated as follows:

$$En(X) = - \sum_{i=1}^n P(x_i) \log_2 P(x_i) \quad (15)$$

where the value of the i-th feature is denoted by $P(x_i)$ and n is the number of features. It was previously mentioned that the feature vector includes 459 features. The obtained results by applying the Shannon entropy to the features vector of the Corel-10k dataset are illustrated in Figure 4. The figure shows that some features have zero entropy, that is, they have no effect on the image retrieval rate and are redundant. Therefore, their elimination from the feature vector will lead to a decrease in the calculation and complexity of the algorithm. The application of Shannon entropy divides the set of 459 features into two subsets. The first subset includes 217 features which are inefficient, hence they were removed to reduce the complexity of the algorithm. The second subset, consisting of the remaining 242 features, serves as the final feature vector for the proposed approach.

4. 3. Best Values for Colour, Texture and Edge Orientation Quantization

An important step of our proposed method for creating feature vectors is the quantization of three values: colour, texture, and edge orientation. In order to determine the best value of these parameters, computational experiments were carried out in which various values of the quantization level were considered with respect to the image retrieval rate. The results are presented in Tables 1 and 3. For each experiment, 20 percent of the Corel-10K dataset was

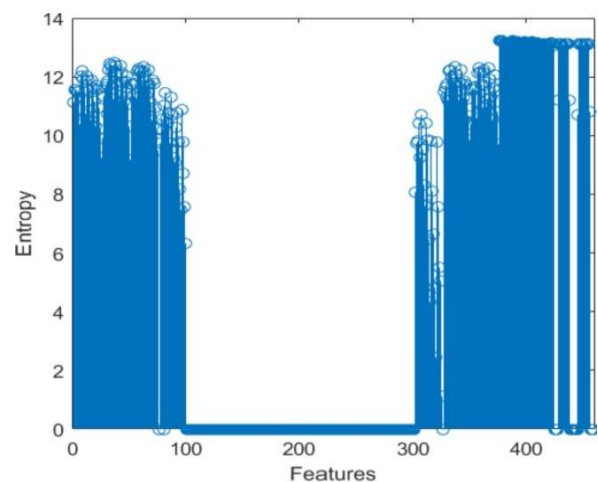


Figure 4. Shannon Entropy of all Proposed Features

randomly selected as validation set. The results in Table 1 indicate that the level of quantization for both texture and colour has a significant impact on image retrieval rate in HSV colour space. In parallel with the increase in colour quantization, the rate of image retrieval increases, but this increase is not continuous and has a limitation.

Once the colour quantization level reaches 5, 6, 6 for channels H, S and V, respectively, the increase becomes inefficient. Compared to the quantization of colour level, there is no clear-cut rule for increasing or decreasing the rate of image retrieval in relation to the texture quantization level.

From the general results of the experiments, we can conclude that the best level of discretization for colour is 15, 5, 5 for channels H, S, and V, respectively, and 32 for texture.

4. 4. Colour Space Selection One of the most effective factors on the performance of CBIR systems is the colour space that features are extracted from. One feature may have different effects on the rate of image

retrieval considering different colour spaces. This means that with the use of one specific feature, it is possible to extract different semantic content from the image in different colour spaces. From the existing colour spaces available in image processing, only two of them, L*a*b and HSV, have perceptually uniform properties [3]. In fact, a colour space is perceptually uniform if the distance between two colours is equal to the distance between the two colours in the human visual system. These two colour spaces are based on the human vision system.

In this section, the proposed features are extracted from these two colour spaces and compared based on their image retrieval rate. In Tables 1 and 3, the image retrieval rate of the proposed features in the HSV colour space is presented, while Tables 2 and 4 give the result of image retrieval in the L*a*b colour space. Comparing Tables 1 and 2, it is observed that the histogram features of the colour difference with the same texture and colour in the HSV colour space give much better semantic content than the L*a*b space.

TABLE 1. Image retrieval rate on the Corel-10K dataset for different values of colour and texture quantization in HSV colour space

The quantization number for texture	The quantization number for colour										
	Precision (%)										
	72	90	108	128	160	192	240	300	350	375	540
8	42.78	43.51	43.84	45.66	45.71	45.73	46.22	46.51	46.73	46.91	46.61
16	43.91	44.52	44.81	46.88	46.92	46.98	47.45	47.87	47.92	48.09	47.81
24	43.97	44.84	44.92	47.65	47.72	47.78	48.26	48.36	48.73	48.88	48.84
32	43.87	44.56	44.76	47.71	47.77	47.86	48.36	48.57	48.84	49.13	48.94
64	42.76	43.02	43.25	46.69	46.76	46.80	47.29	47.86	47.96	48.24	47.73
128	40.42	40.88	41.67	44.98	45.06	45.15	45.55	46.01	46.20	46.41	45.87
256	37.33	37.74	38.51	40.23	40.33	40.41	40.89	41.21	41.39	41.68	41.12

TABLE 2. Image retrieval rate on the Corel-10K dataset for different values of colour and texture quantization in La*b* colour space

The quantization number for texture	The quantization number for colour										
	Precision (%)										
	72	90	108	128	160	192	240	300	350	375	540
8	41.21	41.45	41.63	41.78	41.90	42.25	43.11	43.89	44.25	44.18	43.96
16	41.11	41.23	41.36	41.49	41.69	41.98	42.63	42.97	43.75	43.69	43.76
24	40.71	41.02	41.25	41.38	41.56	41.88	42.53	42.82	43.74	43.35	43.25
32	40.52	40.67	40.76	40.98	41.24	41.64	42.04	42.75	43.02	42.86	42.84
64	39.21	39.25	39.64	39.88	40.58	40.96	41.63	41.89	42.52	42.51	42.32
128	38.35	39.84	39.98	40.57	40.59	40.86	40.94	41.36	41.81	42.56	42.21
256	36.54	37.74	37.79	40.03	10.25	40.36	40.54	40.63	40.75	40.98	40.27

TABLE 3. Image retrieval rate on the Corel-10K dataset for different values of edge orientation and texture quantization in HSV colour space

The quantization number for texture	The quantization number for edge orientation					
	Precision (%)					
36	30	24	18	12	6	
45.73	45.71	45.66	43.84	43.51	42.78	8
46.98	46.92	46.88	44.81	44.52	43.91	16
47.78	47.72	47.65	44.92	44.84	43.97	24
47.86	47.77	47.71	44.76	44.56	43.87	32
46.80	46.76	46.69	43.25	43.02	42.76	64
45.15	45.06	44.98	41.67	40.88	40.42	128
40.41	40.33	40.23	38.51	37.74	37.33	256

TABLE 4. Image Retrieval Rate on the Corel-10K Dataset for Different Values of edge orientation and texture quantization in La*b* colour space

The quantization number for texture	The quantization number for edge orientation					
	Precision (%)					
	6	12	18	24	30	36
8	39.25	39.34	39.53	39.78	40.56	40.42
16	39.53	39.75	39.83	39.96	40.86	40.63
24	39.86	39.94	40.65	40.76	40.94	40.84
32	40.25	40.29	40.68	40.96	41.36	41.25
64	40.65	40.86	40.99	41.63	41.86	41.61
128	40.31	40.36	40.86	40.91	41.31	41.21
256	36.12	36.25	36.54	36.76	36.88	36.11

In addition, a comparison between Tables 3 and 4 shows that the histogram of the colour difference with the same texture and edge orientation for the proposed features in the L*a*b colour space works less efficiently compared to the HSV colour space with respect to

image retrieval rate. Hence, the HSV colour space is more suitable for the features suggested in this article.

4. 5. Evaluation of the Proposed Method

As already mentioned, two criteria for assessing precision and recall are employed to compare and evaluate the proposed method with some recent developed methods.

In this section, the proposed features are first grouped and then compared. In the following, the arrangement of features groups in the feature vector is discussed. In the 242 final selected features, features with indices 1 through 170 referred to the $H_{CDHTxEO}$ histogram, which refer to adjacent pixels with the same colour value. The second group represents the $H_{CDHTxEO}$ histogram for adjacent pixels with the same texture value and the features are indicated by indices from 171 to 202. The third group is the $H_{CDHTxEO}$ histogram for adjacent pixels with the same texture values and is indicated by indices 203-218.

Finally, the fourth group of features with indices 219 to 242 refers to the $H_{CDHTxEO}$ histogram of adjacent pixels with the same edge orientation value. In order to compare the performance of the approach of this study with recent analogues, the methods were evaluated in a random selection of a set of images, which consists of 20 percent of the images of each mentioned datasets as test set. Table 5 presents the information of this comparison with respect to the image retrieval rate. The results indicate that the proposed method has the best image retrieval rate in comparison with others.

According to the results of the study, similarly to the human vision system that uses the colour and texture differences of the image to determine the content, the suggested features have a good and detailed representation of the visual characteristics and perceptual content of the image.

5. CONCLUSION

Generally, in image retrieval system development, it is necessary to consider reducing the semantic gap between the perception of a computer system and of the

TABLE 5. Evaluation of the Proposed Approach and Comparison of Methods.

Dataset		Method						
		MSD[9]	CDH[7]	BBC[10]	LMDP[11]	CDHT[12]	HCSF[3]	Proposed Method
Corel-5K	Precision (%)	55.92	57.23	57.5	47.12	59.51	61.82	62.64
	Recall (%)	6.71	6.87	6.90	5.74	7.21	7.41	7.58
Corel-10K	Precision (%)	45.62	45.24	47.01	-	49.31	50.67	51.06
	Recall (%)	5.48	5.43	5.64	-	5.88	6.08	6.28

human visual system. Since the texture and colour of the image are two determining factors for the human visual system, the proposed method uses these factors, as well as the orientation of the edge to extract the contents of the image. Colour difference plays a critical role in our study. This difference can be considered from four points of view. The first point of view is when two adjacent pixels have the same texture value in colour, the second when they have the same colour value, the third when they have texture values on the orientation edge, and the fourth when they have the same orientation value. A histogram is generated from the colour difference derived from these four cases.

Since feature in different colour spaces may give different results, the proposed features of this article were investigated in both the L^*a^*b and HSV colour spaces. Our research shows that only these two colour spaces have the perceptually uniform property. The experimental results showed that the HSV colour space is more suitable for the proposed features, as it is closer to the human visual system.

An important step after feature extraction is identifying less-efficient features, removing them, and reducing feature space in order to increase the image retrieval rate. This stage was implemented using the Shannon entropy, which led to the selection of 242 features with a high impact on the image retrieval rate. A significant improvement in the image retrieval rate of the proposed method compared to others indicates that the proposed features are able to capture image content in more detail without any complex operations such as segmentation, training, and clustering.

6. REFERENCES

- Ghosh, N., Agrawal S., Motwani M. "A survey of feature extraction for content-based image retrieval system." *Proceedings of International Conference on Recent Advancement on Computer and Communication*. Springer, Singapore, (2018), 305-313.
- Haji, M. S., Alkawaz, M. H., Rehman, A., & Saba, T. "Content-based image retrieval: a deep look at features prospectus." *International Journal of Computational Vision and Robotics*, Vol. 9, No. 1, (2019), 14-38.
- Qazanfari, H., Hassanpour, H., & Qazanfari, K. "Content-Based Image Retrieval Using HSV Colour Space Features." *International Journal of Computer and Information Engineering*, Vol. 13, No. 10, (2019), 537-545.
- Keyvanpour, M., Tavoli, R., & Mozafari, S. "Document image retrieval based on keyword spotting using relevance feedback." *International Journal of Engineering*, Vol. 27, No. 1, (2014), 7-13.
- Qazanfari, H., Hassanpour, H., & Qazanfari, K. "A short-term learning framework based on relevance feedback for content-based image retrieval." 2017 3rd Iranian Conference on Intelligent Systems and Signal Processing (ICSPIS). IEEE, (2017), 136-140.
- Livingstone, M. S., & Hubel, D. H. "Anatomy and physiology of a colour system in the primate visual cortex." *Journal of Neuroscience*, Vol. 4, No. 1, (1984), 309-356.
- Liu, G. H., & Yang, J. Y. "Content-based image retrieval using colour difference histogram." *Pattern Recognition*, Vol. 46, No. 1, (2013), 188-198.
- Ojala, T., Pietikainen, M., & Maenpaa, T. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." *IEEE Transactions On Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, (2002), 971-987.
- Liu, G. H., Li, Z. Y., Zhang, L., & Xu, Y. "Image retrieval based on micro-structure descriptor." *Pattern Recognition*, Vol. 44, No. 9, (2011), 2123-2133.
- Singh, C., & Kaur, K. P. "A fast and efficient image retrieval system based on colour and texture features." *Journal of Visual Communication and Image Representation*, Vol. 41, (2016), 225-238.
- Kumar, G. S., & Mohan, P. K. "Local mean differential excitation pattern for content based image retrieval." *SN Applied Sciences*, Vol. 1, No. 1, (2019), 46.
- Ajam, A., AlyanNezhadi, M. M., Qazanfari, H., Forghani, M., & Amiri Z. "Content-based Image Retrieval Using Colour Difference Histogram in Image Textures." 2019 5rd Iranian Conference on Intelligent Systems and Signal Processing (ICSPIS). IEEE, 2019.
- Carneiro, G., Chan, A. B., Moreno, P. J., & Vasconcelos, N. "Supervised learning of semantic classes for image annotation and retrieval." *IEEE Transactions On Pattern Analysis and Machine Intelligence*, Vol. 29, No. 3, (2007), 394-410.
- Long, J., Shelhamer, E., & Darrell, T. "Fully convolutional networks for semantic segmentation." *Proceedings of the IEEE Conference On Computer Vision and Pattern Recognition*. (2015), 3431-3440.
- Meng, X., An, Y., He, J., Zhuo, Z., Wu, H., & Gao, X. "Similar image retrieval only using one image." *Optik*, Vol. 127, No. 1, (2016), 141-144.
- Xu, Y. Y. "Multiple-instance learning based decision neural networks for image retrieval and classification." *Neurocomputing*, Vol. 171, (2016), 826-836.
- Sharif Razavian, A., Azizpour, H., Sullivan, J., & Carlsson, S. "CNN features off-the-shelf: an astounding baseline for recognition." *Proceedings of the IEEE Conference On Computer Vision and Pattern Recognition Workshops*. (2014), 806-813.
- Khosravi, A., Tizhoosh, H. R., Babaie, M., Khatami, A., & Nahavandi, S. "A radon-based convolutional neural network for medical image retrieval." *International Journal of Engineering*, Vol. 31, No. 6, (2018), 910-915.
- Sezavar, A., Farsi, H., & Mohamadzadeh, S. "A Modified Grasshopper Optimization Algorithm Combined with CNN for Content Based Image Retrieval." *International Journal of Engineering*, Vol. 32, No. 7, (2019), 924-930.
- Muja, M., & Lowe, D. G. "Scalable nearest neighbor algorithms for high dimensional data." *IEEE Transactions On Pattern Analysis and Machine Intelligence*, Vol. 36, No. 11, (2014), 2227-2240.
- Jegou, H., Douze, M., & Schmid, C. "Product quantization for nearest neighbor search." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 1, (2010), 117-128.
- Zhang, S., Yang, M., Cour, T., Yu, K., & Metaxas, D. N. "Query specific fusion for image retrieval." *European Conference on Computer Vision*. Springer, Berlin, Heidelberg, (2012), 660-673.

23. Kundert-Gibbs, J. "Image-Based Content Retrieval via Class-Based Histogram Comparisons." *IT Convergence and Security* Springer, Singapore, (2018), 3-10.
24. Heller, K. A., & Ghahramani, Z. "A simple Bayesian framework for content-based image retrieval." 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). Vol. 2. IEEE, 2006.
25. Rao, M. B., Rao, B. P., & Govardhan, A. "Content based image retrieval using Dominant colour and Texture features." *International Journal of Computer Science and Information Security*, Vol. 9, (2011), 41-46.
26. Neelima, N., & Reddy, E. S. "An Efficient Multi Object Image Retrieval System Using Multiple Features and SVM." *Advances in Signal Processing and Intelligent Recognition Systems*. Springer, Cham, (2016), 257-265.
27. Singha, M., & Hemachandran, K. "Content based image retrieval using colour and texture." *Signal & Image Processing*, Vol. 3, No. 1, (2012), 39.
28. Jenni, K., Mandala, S., & Sunar, M. S. "Content based image retrieval using colour strings comparison." *Procedia Computer Science*, Vol. 50, (2015), 374-379.
29. Tola, E., Lepetit, V., & Fua, P. "Daisy: An efficient dense descriptor applied to wide-baseline stereo." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, NO. 5, (2009), 815-830.
30. Li, B., Xiao, R., Li, Z., Cai, R., Lu, B. L., & Zhang, L. "Rank-SIFT: Learning to rank repeatable local interest points." CVPR 2011. IEEE, (2011), 1737-1744.
31. Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., & Fua, P. "BRIEF: Computing a local binary descriptor very fast." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, NO. 7, (2011), 1281-1298.
32. Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. "ORB: An efficient alternative to SIFT or SURF." 2011 International Conference On Computer Vision. (2011), 2564-2571.
33. Veenman, C. J., Smeulders, A. W., & Geusebroek, J. M. "Visual word ambiguity." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, NO. 7, (2010), 1271-1283.
34. Jégou, H., Douze, M., & Schmid, C. "Improving bag-of-features for large scale image search." *International Journal of Computer Vision*, Vol. 87, NO. 3, (2010), 316-336.
35. Lazebnik, S., Schmid, C., & Ponce, J. "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories." 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). IEEE, (2006), 2169-2178.
36. Nowak, E., Jurie, F., & Triggs, B. "Sampling strategies for bag-of-features image classification." European Conference on Computer Vision. Springer, Berlin, Heidelberg, (2006), 490-503.
37. Philbin, J., Chum, O., Isard, M., Sivic, J., & Zisserman, A. "Lost in quantization: Improving particular object retrieval in large scale image databases." 2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, (2008), 1-8.
38. Zhu, C., Bichot, C. E., & Chen, L. "Image region description using orthogonal combination of local binary patterns enhanced with colour information." *Pattern Recognition*, Vol. 46, NO. 7, (2013), 1949-1963.
39. Liu, G. H., Zhang, L., Hou, Y. K., Li, Z. Y., & Yang, J. Y. "Image retrieval based on multi-texton histogram." *Pattern Recognition*, Vol. 43, NO. 7, (2010), 2380-2389.
40. Girija, O. K., & Sudheep E. M. "Overview of Image Retrieval Techniques." *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)*, Vol. 4, No. 1 (2015), 2319-5940.
41. Corel Dataset, <http://smartcbr.nph-co.ir/datasets.php>, last accessed 2020/02/01.

Persian Abstract

چکیده

بازیابی تصویر مبتنی بر محتوا یکی از مباحث مورد علاقه در علم پردازش تصویر و بینایی ماشین است. در یک سیستم بازیابی تصویر مبتنی بر محتوا، محتوای یک تصویر با تصاویر پایگاه داده مقایسه شده و تصاویر مشابه بازیابی می شوند. محتوای تصاویر توسط یکسری ویژگی‌ها از آنها استخراج می شوند. در این مقاله ویژگی‌های پیشنهادی بر اساس سیستم بینایی انسان استخراج می شوند. سیستم بینایی انسان به بافت و جهت لبه در تصویر بسیار حساس است. هیستوگرام اختلاف رنگ مبنی بر این دو ویژگی در فضای رنگی HSV ویژگی‌های نزدیک به سیستم بینایی انسان از تصویر استخراج می کند. هیستوگرام اختلاف رنگ در واقع اختلاف رنگ ادراکی بین دو پیکسل از تصویر است که جهت لبه برابر یا بافت برابر دارند. علاوه بر این، در روش پیشنهادی از روش های داده کاوی مانند اتروپی شانون جهت انتخاب ویژگی های کارآمد استفاده می شود. سیستم بازیابی تصویر پیشنهادی روی دو پایگاه داده استاندارد کورل 5k، کورل 10k ارزیابی شده است. نتایج آزمایشات ارائه شده در این مقاله توانایی بازیابی دقیق تر سیستم پیشنهادی را نسبت به سیستم های بازیابی تصویر اخیر نشان می دهد.
