

## Implementation of early and late fusion methods for content-based image retrieval



Ali Ahmed <sup>1,\*</sup>, Sara Mohamed <sup>2</sup>

<sup>1</sup>Faculty of Computer Science and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

<sup>2</sup>Faculty of Computer Science and Information Technology, Sudan University for Science and Technology, Khartoum, Sudan

### ARTICLE INFO

#### Article history:

Received 6 January 2021

Received in revised form

31 March 2021

Accepted 17 April 2021

#### Keywords:

Content-based image retrieval

Feature extraction

Fusion method

### ABSTRACT

Content-Based Image Retrieval (CBIR) systems retrieve images from the image repository or database in which they are visually similar to the query image. CBIR plays an important role in various fields such as medical diagnosis, crime prevention, web-based searching, and architecture. CBIR consists mainly of two stages: The first is the extraction of features and the second is the matching of similarities. There are several ways to improve the efficiency and performance of CBIR, such as segmentation, relevance feedback, expansion of queries, and fusion-based methods. The literature has suggested several methods for combining and fusing various image descriptors. In general, fusion strategies are typically divided into two groups, namely early and late fusion strategies. Early fusion is the combination of image features from more than one descriptor into a single vector before the similarity computation, while late fusion refers either to the combination of outputs produced by various retrieval systems or to the combination of different rankings of similarity. In this study, a group of color and texture features is proposed to be used for both methods of fusion strategies. Firstly, an early combination of eighteen color features and twelve texture features are combined into a single vector representation and secondly, the late fusion of three of the most common distance measures are used in the late fusion stage. Our experimental results on two common image datasets show that our proposed method has good performance retrieval results compared to the traditional way of using single features descriptor and also has an acceptable retrieval performance compared to some of the state-of-the-art methods. The overall accuracy of our proposed method is 60.6% and 39.07% for Corel-1K and GHIM-10K datasets, respectively.

© 2021 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### 1. Introduction

There are two possible solutions for searching large image databases; the first, old and insufficient solution is text-based image retrieval (TBIR) (Sumathi et al., 2011), in which images are manually annotated with suitable names; this meta-data is attached to images, and the end-user later applies keywords to search for and obtain the necessary images. This approach has two major disadvantages; the first problem is obvious, as a significant amount of time is required for the manual annotation process, while the second disadvantage is the

amount of responsibility placed on the end user to form their own queries. An alternative and effective method is content-based image retrieval (CBIR); this method has many advantages that solve the major problems highlighted by the previous method. Also, CBIR became the main method and an area of much research in the last two decades due to the rapid advancement of multimedia data through modern sources such as the Internet, smart devices, internet of things IoT devices and sources, social networks, and medical image sources (Seetharaman and Kamarasan, 2014; Alsmadi, 2020). Most of the earlier CBIR studies were based on individual or single feature descriptors extracted from color, textures, or shape contents/descriptors (Mistry, 2020); however, single feature spaces are not always sufficient to provide the best retrieval results. Therefore, most of the recent research has combined more than one descriptor from the above examples. The early survey of numerous and diverse application areas of CBIR was reported by Veltkamp

\* Corresponding Author.

Email Address: [aabdelrahim@kau.edu.sa](mailto:aabdelrahim@kau.edu.sa) (A. Ahmed)

<https://doi.org/10.21833/ijaas.2021.07.012>

Corresponding author's ORCID profile:

<https://orcid.org/0000-0002-8944-8922>

2313-626X/© 2021 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

and Tanase (2002), where the authors provided an early review of CBIR at the end of the early years. Researchers could refer to [Latif et al. \(2019\)](#) for the latest comprehensive review of CBIR.

In this paper, we proposed a retrieval method that exploits the use of color and textual information, and implements and compares the two common types of fusion, after which it suggests the use of the best approach to retrieve similar images. The main contributions of our work are as follows:

1. A fully automatic CBIR method is proposed for retrieving similar images based on the fusion of visual and textual characteristics
2. The proposed approach has good retrieval results when good representative extracted fractures of both color and textual images are combined into a single vector. The more accurate the features used by the system, the higher the performance retrieval results obtained by the system.
3. The method also supports the idea of merging more than one result with similar measures. The user can use any group of similarity measures that may present a good final fused value of precision better than individual similarity values.

The rest of the paper is organized as follows. The related work is defined in Section 2. In Section 3, the proposed method is described in detail; feature extraction methods from the color and textual information of images are described. In Section 4, the experimental setup and image datasets are discussed along with implementation details and an evaluation of the results. Section 5 concludes the proposed work.

## 2. Related works

Over the years, numerous approaches have been suggested in the literature to combine and fuse various image descriptors ([Bhowmik et al., 2014](#)). Fusion approaches are usually categorized into two classes: Early and late fusion ([Snoek et al., 2005](#); [Atrey et al., 2010](#)). There are different fusion strategies and approaches that have been developed, these strategies could be classified into five groups: traditional early fusion, features reweighting for early fusion, representation by multi-feature spaces for late fusion, relevance feedback approaches, and multimodal retrieval. In the following, we will give a short definition and recent and some examples of recent studies that applied these ideas. In normal or traditional early fusion, multiple image representations (such as color, texture, and shape) are combined into a single feature vector. This approach is straightforward and very simple and normally uses an equal weight value scheme assigned to any of the feature spaces. Several studies have followed this method, such as that in [Singh and Hemachandran \(2012\)](#) and [Alsmadi \(2020\)](#). Feature weighting for early fusion is suitable when the retrieval process is considered as a classification task in which pools of images are assigned to a set of

labels. In this case, images with the same labels are similar to each other; a technique such as principal component analysis (PCA) could be used for the extraction of a new feature space rather than capturing the similarities ([Piras and Giacinto, 2017](#)). Several studies that address the different weighting estimation approaches and their related problems can be found in [Rui et al. \(1997\)](#), and reweighting schemes for similar retrieval problems can be found in [Abdelrahim \(2013\)](#) and [Ahmed et al. \(2013; 2014; 2017; 2019\)](#). Over the past years, multi-feature spaces for late fusion were enhanced; authors in the pattern recognition area of research have suggested a number of solutions for fusing the information based on combinations of the outputs of different classifiers that use different feature spaces, as shown in [Kuncheva \(2014\)](#). The most popular and effective techniques for merging similar outputs are based on late fusion methods such as the mean, maximum and minimum rules ([Alhassan and Alfaki, 2017](#); [Ibrahim et al., 2018](#); [Ahmed and Malebary, 2019](#)). Relevance feedback has an obvious role in enhancing the retrieval task, where an image query is reformulated according to the top rank image retrieved ([Karamti et al., 2018](#); [Ahmed, 2020](#)). Similarly, the query expansion process has significantly enhanced the retrieval performance ([Houle et al., 2017](#); [Ahmed and Malebary, 2020](#)). The main idea of multimodal retrieval is to combine keywords with low-level features in order to use a combined input space. This suggestion was followed by some works such as that described by [Zhou and Huang \(2002\)](#), in which the authors use the word association via relevance feedback (WARF) method to learn the keyword similarity matrix during user interaction ([Sclaroff et al., 1999](#)); several other researchers have addressed this problem from different points of view ([Barnard and Forsyth, 2001](#)).

## 3. Proposed method

In this paper, the normal early fusion method with no weighting scheme and multi-feature space with multiple similarity measures for late fusion is proposed. As we will discuss in detail, those two proposed fusion methods are considered a simple approach that will enhance the retrieval performance. The overall CBIR framework is shown in [Fig. 1](#).

### 3.1. Feature extraction and early fusion

As shown in the proposed framework, a group of color and texture feature extraction functions is first applied to extract the most important representative image features. After successful extraction, these features will be combined into a single vector; this process will be performed for both queries as well as every image in the database. For color feature descriptors, a total of 18 features are extracted using most of the common color moment functions.

First, each image in the database, as well as the query images, are converted from the RGB color

space to the HSV color space (Erkut et al., 2019), then each of the “H”, “S” and “V” channels are used as

an input for the color moment functions and six values are generated.

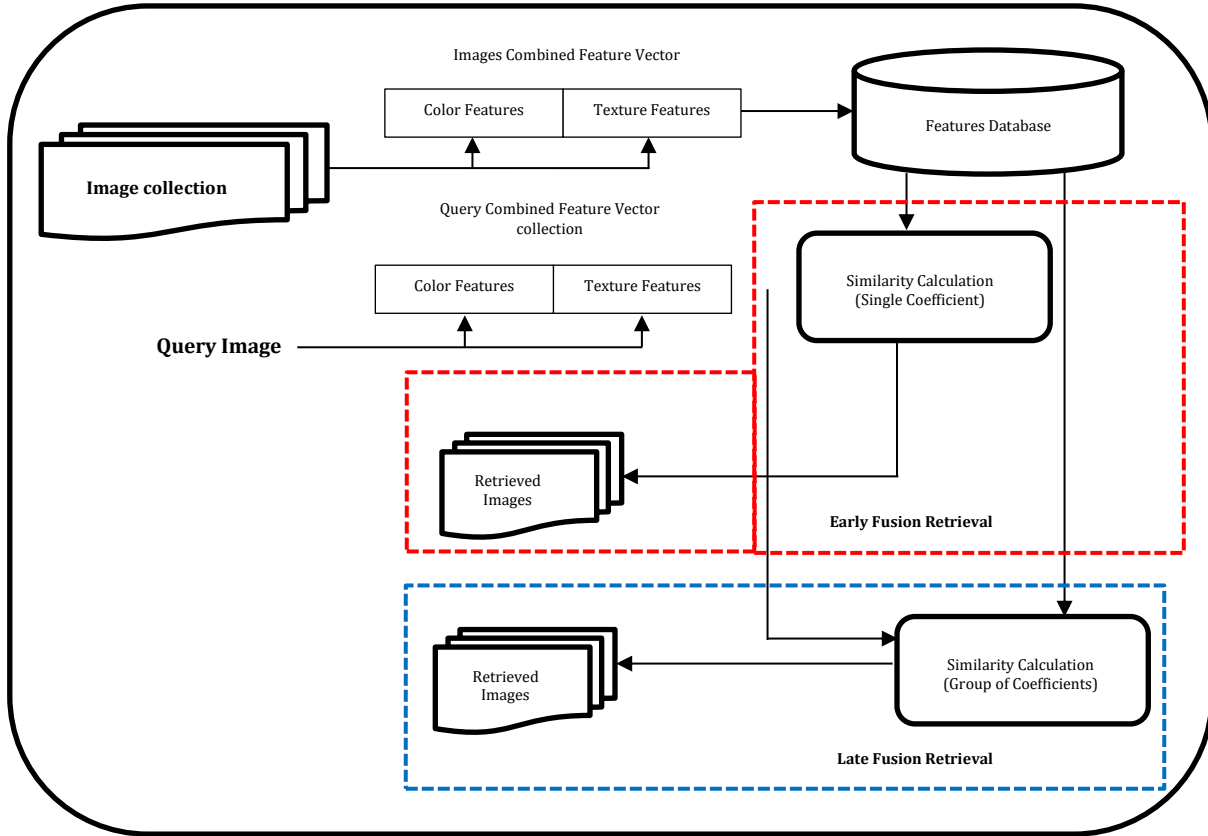


Fig. 1: Workflow of our proposed CBIR method

All color features for this descriptor are combined into a single vector. The group of six-moment functions used here is: mean, variance, skew, kurtos, smoothness, and contrast. The following equations (1 to 6) describe the formula for each of these functions; these were proposed by Maheshwary and Srivastava (2009) and were successfully used in recent studies (Zenggang et al., 2019). Here, the value of pixels at the  $i^{th}$  row and  $j^{th}$  column become  $v_{ij}$  and the dimension of the image is (M, N) pixels; therefore, the six color moment functions are defined as follows:

$$Mean = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N v_{ij} \quad (1)$$

$$Variance = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (v_{ij} - m)^2 \quad (2)$$

$$skew = \left( \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (v_{ij} - m)^3 \right)^{1/3} \quad (3)$$

$$kurtos = \left( \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (v_{ij} - m)^4 \right)^{1/4} \quad (4)$$

$$Smoothness = 1 - \frac{1}{(1+V)} \quad (5)$$

$$Contrast = \left( \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (v_{ij} - m)^2 \right)^{\frac{1}{2}} \quad (6)$$

Similarly, a total of 12 texture features were extracted using the grey-level co-occurrence matrix (GLCM); this method was introduced by Weszka et al. (1976). As in color features, after converting each RGB image to the HSV color space, then every single channel is used as an input for the GLCM method; then, from the output matrix of co-occurrence, the

following functions were used for texture feature extractions:

$$uniformity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \{p(i, j)\}^2 \quad (7)$$

$$correlation = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{(i-\mu)(j-\mu)p(i, j)}{\sigma^2} \quad (8)$$

$$contrast = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j)(i - j)^2 \quad (9)$$

$$entropy = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \log\{p(i, j)\} \quad (10)$$

$p(i, j) = (p(i, j)) / R$  is the  $(i, j)^{th}$  entry in the normalized matrix,  $N_g$  is the number of unique or distinct grey levels, and  $\mu$  and  $\sigma$  are the mean and standard deviation for image intensity, respectively.

### 3.2. Similarity measures and late fusion

After constructing features extracted using color and texture descriptors, the next stage is to perform late fusion based on three common similarity measures; the measures used for late fusion here are Bray, Cityblock, and Euclidean. These distance similarity metrics have shown good performance results (Shirkhorshidi et al., 2015). The following equations describe these distance measures. For the late fusion strategy here, the minimum distance (maximum similarity) is applied, and then the combined values will be used for further ranking processes, as will be discussed in the following section.

$$Bray = 2 \frac{C_{ij}}{S_i + S_j} \quad (11)$$

where,  $C_{ij}$  is the sum of the lesser values for only those species in common for both vectors, and  $S_i$  and  $S_j$  are the total number of specimens counted at both vectors.

$$CityBlock = \sum_i^n |x_i - y_i| \tag{12}$$

$$Euclidean = [\sum_i^n (x_i - y_i)^2]^{\frac{1}{2}} \tag{13}$$

where,  $n$  is the number of dimensions of the  $X$  and  $Y$  vectors.

#### 4. Experimental results and discussion

##### 4.1. Image datasets

In this study, the Corel-1K (Li and Wang, 2008) and GHIM-10K (Liu et al., 2015) datasets were used. These two image datasets have been recently used in Varish et al. (2020). The images in the first dataset were divided into ten different categories: Africans, Beaches, Buildings, Buses, Dinosaurs, Elephants, Flowers, Horses, Mountains, and Food. Each class or category consists of a hundred images with the resolution 256×384 or 384×256 pixels. The second dataset used in this study is GHIM-10K which has a total of 10K images (10 times larger than Corel-1K) and is considered to be more challenging and diverse and contains various objects. It consists of 10,000 images divided into 20 groups/categories where each group or class has 500 images with a resolution of 300×400 pixels or 400×300 pixels; the semantic names of the images in each group are sunsets, bikes, forts, ships, flies, cars, etc. Figs. 2 and 3 show the sample images of each dataset where a single image from each class has been taken.



Fig. 2: Samples images from the Corel-1K dataset



Fig. 3: Samples of images from GHIM-10K

##### 4.2. Performance evaluation

For evaluation measures, recall and precision were used; these are the most common performance measures for most retrieval methods. Here, both

measures are calculated as the top ten retrieved images using the following formula:

$$Recall = \frac{\text{number of relevant images retrieved}}{\text{total number of retrieved images}} \tag{14}$$

$$Precision = \frac{\text{number of relevant images retrieved}}{\text{total number of relevant images}} \tag{15}$$

#### 4.3. Results and discussion

In this study, three different experiments were conducted: Retrieval was based on individual feature descriptors of color and textures only, with the implementation of early fusion for both feature descriptors and late fusion using combined vectors of feature descriptors and three similarity measures described in our methodology. For the three experiments, the same ten random image queries selected from each class of both datasets were used and recall and precision were calculated as the top ten images retrieved as performed in many related studies. The following subsections give more detail about our experiment results. The first experiment used eighteen color features and twelve texture features for both datasets images separately with the three individual similarities (distances) measures. Average recall and precision for each class were computed as the top ten retrieved images, as shown in Tables 1 and 2 for both datasets; the results in these two tables represent the base results that our proposed methods will target to enhance and increase their performance, as we will see in the next scenarios. From the two tables, we can see that color features have better retrieval results or performances compared to texture, and Euclidean distances provide the best distance measures of the three distance coefficients.

The concept of the early fusion method is implemented in the second experiment, in which the simple merging of feature vectors of both color and texture are combined horizontally, resulting in a single feature vector with thirty features (eighteen color features and twelve texture features). The same random image queries were used for each class and the average recall and precision were calculated. In the late fusion scenario, the combined vector of thirty features was used again and similarity values for the three similarity measures are fused or grouped together; then, the minimum for these three values was taken. Finally, the average recall and precision were taken from the fused similarity values; the results of early and late fusion experiments are shown in Tables 3 and 4 for the two datasets, respectively. Also, Figs. 4 and 5 show the combined results of the three retrieval methods for two datasets which make them easy to compare and conclude some important findings from their recall and precision values.

Figs. 6 and 7 show all of the recall and precision values for three retrieval methods for the two datasets, respectively. The enhanced results from Tables 3 and 4 and Figs. 6 and 7 show that both early and late fusion were significant and simply enhanced

the retrieval performance in terms of recall and precision values.

**Table 1: Recall and precision for Corel-1k dataset using color and texture features**

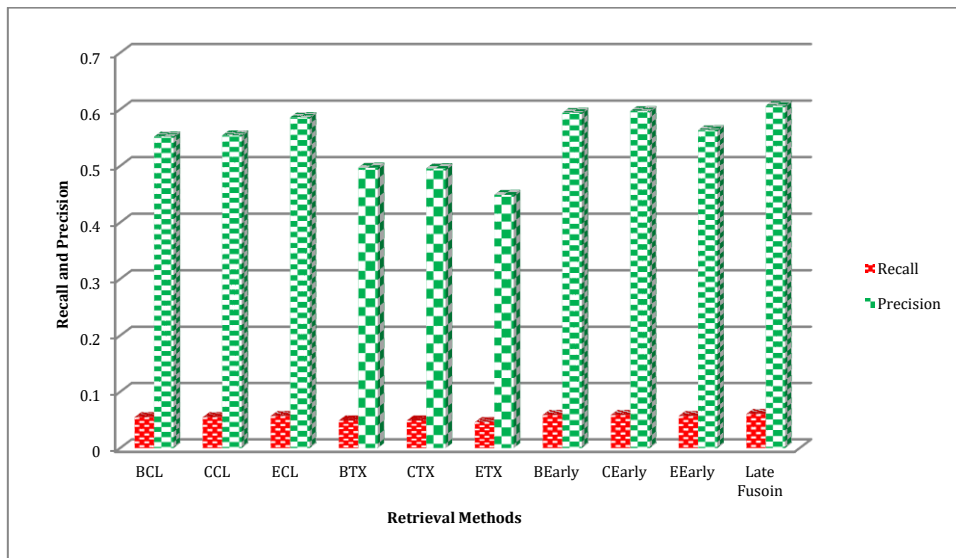
Class	Based on Colour Features						Based on Texture Features					
	Bray		Cityblock		Euclidean		Bray		Cityblock		Euclidean	
	R	P	R	P	R	P	R	P	R	P	R	P
Africa	0.03	0.3	0.03	0.3	0.03	0.34	0.04	0.42	0.04	0.42	0.04	0.38
Beach	0.03	0.31	0.03	0.31	0.03	0.32	0.03	0.29	0.03	0.3	0.03	0.26
Building	0.05	0.45	0.05	0.47	0.05	0.48	0.04	0.45	0.04	0.44	0.04	0.39
Buses	0.07	0.68	0.07	0.69	0.07	0.74	0.07	0.68	0.07	0.68	0.06	0.59
Dinosaurs	0.1	1	0.1	1	0.1	1	0.1	1	0.1	0.99	0.1	1
Elephants	0.05	0.53	0.05	0.52	0.06	0.6	0.03	0.31	0.03	0.31	0.02	0.25
Flowers	0.05	0.53	0.05	0.53	0.05	0.53	0.08	0.83	0.08	0.83	0.07	0.72
Horses	0.09	0.93	0.09	0.94	0.09	0.95	0.05	0.5	0.05	0.53	0.05	0.45
Mountains	0.04	0.41	0.04	0.4	0.04	0.43	0.02	0.2	0.02	0.19	0.02	0.2
Food	0.04	0.38	0.04	0.38	0.05	0.47	0.03	0.29	0.03	0.27	0.03	0.25
Mean	0.055	0.552	0.055	0.554	0.057	<b>0.586</b>	0.049	<b>0.497</b>	0.049	0.496	0.046	0.449

**Table 2: Recall and precision for GHIM-10k dataset using color and texture features**

Class	Based on Color Features						Based on Texture Features					
	Bray		Cityblock		Euclidean		Bray		Cityblock		Euclidean	
	R	P	R	P	R	P	R	P	R	P	R	P
Fireworks	0.013	0.63	0.013	0.64	0.014	0.68	0.01	0.5	0.01	0.49	0.009	0.44
Buildings	0.005	0.25	0.004	0.22	0.005	0.23	0.007	0.36	0.007	0.35	0.007	0.37
Walls	0.004	0.18	0.004	0.18	0.005	0.25	0.002	0.12	0.002	0.12	0.002	0.09
Cars	0.003	0.16	0.003	0.15	0.004	0.19	0.004	0.18	0.003	0.17	0.003	0.15
Flies	0.005	0.24	0.005	0.23	0.006	0.29	0.006	0.31	0.006	0.31	0.005	0.26
Mountains	0.003	0.16	0.003	0.15	0.003	0.15	0.002	0.12	0.003	0.13	0.002	0.09
Flowers	0.003	0.17	0.003	0.17	0.004	0.2	0.007	0.36	0.007	0.36	0.006	0.31
Trees	0.009	0.44	0.009	0.45	0.01	0.48	0.008	0.4	0.008	0.41	0.008	0.39
Green Grounds	0.008	0.4	0.008	0.39	0.009	0.45	0.009	0.43	0.009	0.43	0.007	0.35
Beaches	0.005	0.23	0.005	0.23	0.005	0.23	0.003	0.15	0.003	0.15	0.003	0.16
Aeroplanes	0.004	0.2	0.004	0.19	0.005	0.26	0.006	0.28	0.005	0.27	0.005	0.26
Butterflies	0.007	0.34	0.006	0.32	0.007	0.35	0.004	0.22	0.004	0.22	0.003	0.17
Forts	0.002	0.12	0.002	0.11	0.002	0.12	0.008	0.39	0.007	0.37	0.006	0.32
Sunsets	0.007	0.36	0.007	0.36	0.009	0.46	0.012	0.6	0.012	0.6	0.01	0.5
Bikes	0.006	0.31	0.006	0.31	0.007	0.34	0.005	0.26	0.005	0.26	0.005	0.25
Boats	0.006	0.31	0.006	0.3	0.006	0.31	0.004	0.21	0.004	0.2	0.004	0.21
Ships	0.004	0.19	0.004	0.19	0.003	0.16	0.005	0.26	0.005	0.26	0.005	0.24
Chickens	0.006	0.32	0.006	0.31	0.006	0.3	0.005	0.23	0.005	0.25	0.004	0.19
Insects	0.006	0.28	0.006	0.31	0.006	0.32	0.003	0.15	0.003	0.15	0.002	0.12
Horses	0.002	0.11	0.002	0.12	0.003	0.14	0.004	0.18	0.004	0.18	0.003	0.15
Mean	0.0054	0.27	0.005	0.266	0.0059	<b>0.296</b>	0.0057	<b>0.286</b>	0.005	0.284	0.005	0.251

Obvious outcomes when considering and using fusion methods as an enhancement tool or approach are clearly shown in Figs. 6 and 7, which highlight the fact that the good retrieval performance (in a red

color curve) is better than when using traditional methods that are based on applying either color or texture feature descriptors individually.



**Fig. 4: Recalls and precisions for Corel-1K dataset**

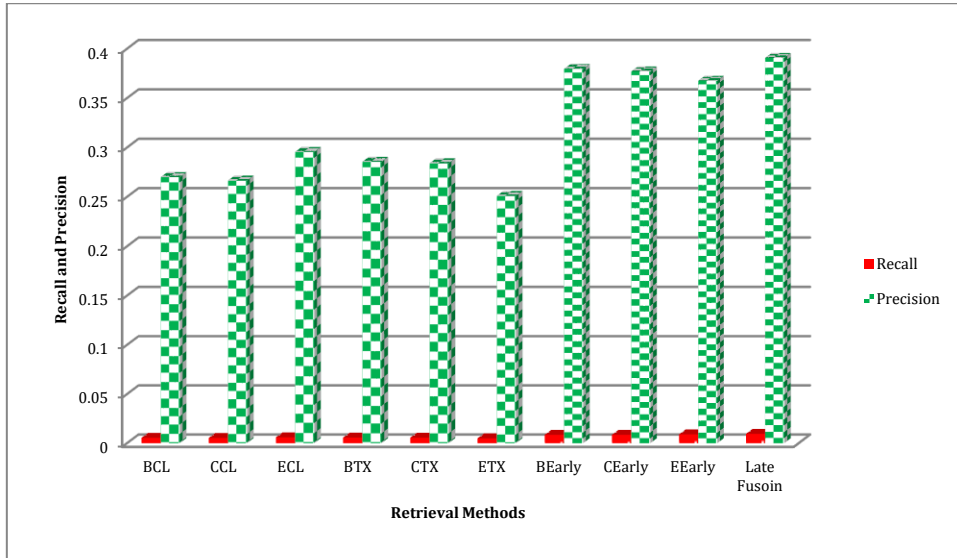


Fig. 5: Recalls and precisions for GHIM-10K dataset

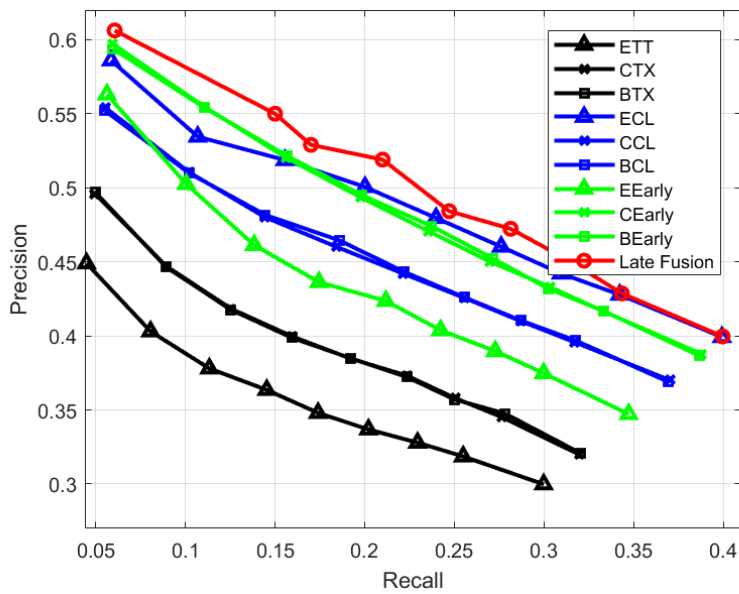


Fig. 6: Recalls and precisions graph for Corel-1K dataset

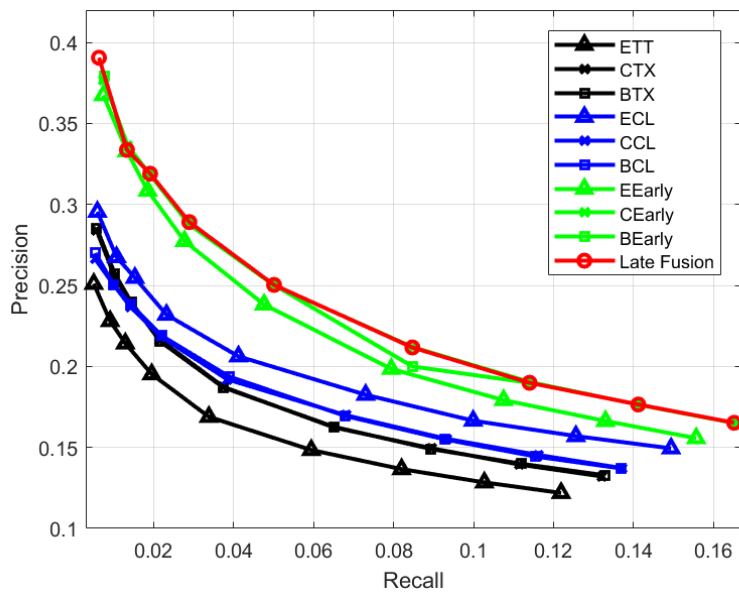


Fig. 7: Recalls and precisions graph for GHIM-10K dataset

**Table 3:** Recall and precision for Corel-1k dataset based on early and late fusion

Class	Based on Early Fusion						Based on Late Fusion		Yue et al. (2011)
	Bray		Cityblock		Euclidean		R	P	
	R	P	R	P	R	P			
Africa	0.04	0.4	0.04	0.41	0.04	0.38	0.0412	0.412	0.5875
Beach	0.03	0.33	0.03	0.32	0.04	0.37	0.0338	0.338	0.4119
Building	0.05	0.53	0.05	0.54	0.05	0.51	0.0575	0.575	0.4235
Buses	0.08	0.8	0.08	0.81	0.07	0.72	0.0825	0.825	0.7169
Dinosaurs	0.1	1	0.1	1	0.1	1	0.1	1	0.7453
Elephants	0.05	0.46	0.05	0.49	0.04	0.42	0.0438	0.438	0.6508
Flowers	0.09	0.9	0.09	0.88	0.09	0.86	0.0912	0.912	0.8324
Horses	0.08	0.79	0.08	0.81	0.07	0.71	0.0825	0.825	0.693
Mountains	0.04	0.38	0.04	0.37	0.04	0.37	0.0362	0.362	0.4486
Food	0.03	0.35	0.03	0.34	0.03	0.29	0.0375	0.375	0.4454
Mean	0.059	0.594	0.059	0.597	0.057	0.5630	0.06062	<b>0.6060</b>	<b>0.5960</b>

**Table 4:** Recall and precision for GHIM-10k dataset based on early and late fusion

Class	Based on Early Fusion						Based on Late Fusion	
	Bray		Cityblock		Euclidean		R	P
	R	P	R	P	R	P		
Fireworks	0.01	0.65	0.01	0.59	0.01	0.62	0.01	0.637
Buildings	0.01	0.37	0.01	0.35	0.01	0.32	0.01	0.375
Walls	0.01	0.29	0.01	0.28	0.01	0.26	0.007	0.313
Cars	0.01	0.26	0.01	0.27	0.01	0.26	0.01	0.263
Flies	0.01	0.38	0.01	0.42	0.01	0.37	0.01	0.4
Mountains	0.001	0.23	0.002	0.23	0.001	0.21	0.006	0.25
Flowers	0.01	0.37	0.01	0.39	0.01	0.36	0.01	0.375
Trees	0.01	0.51	0.01	0.48	0.01	0.49	0.01	0.512
Green Grounds	0.01	0.63	0.01	0.62	0.01	0.63	0.01	0.65
Beaches	0.01	0.29	0.01	0.28	0.01	0.26	0.008	0.313
Aeroplanes	0.01	0.37	0.01	0.37	0.01	0.37	0.01	0.375
Butterflies	0.01	0.39	0.01	0.39	0.01	0.35	0.01	0.4
Forts	0.01	0.27	0.01	0.27	0.01	0.29	0.01	0.275
Sunsets	0.01	0.67	0.01	0.67	0.01	0.69	0.011	0.675
Bikes	0.01	0.3	0.01	0.31	0.01	0.32	0.01	0.313
Boats	0.01	0.36	0.01	0.36	0.01	0.29	0.01	0.362
Ships	0.001	0.23	0.001	0.23	0.001	0.24	0.007	0.237
Chickens	0.01	0.42	0.01	0.42	0.01	0.37	0.01	0.45
Insects	0.01	0.37	0.01	0.37	0.01	0.34	0.01	0.388
Horses	0.002	0.23	0.001	0.24	0.01	0.31	0.01	0.25
Mean	0.0085	0.3795	0.0085	0.377	0.009	0.3675	0.00945	<b>0.3907</b>

More analytical results were shown using the Kendall W concordance test (Siegel and Castellan Jr., 1988). This test was found to evaluate the effectiveness of different retrieval approaches; it could also be used to measure the level of agreement between multiple sets of rankings for the same set of retrieved objects. This test was used in the current study, where the classes of images represent judges while the average precision objects were considered judges in the present context, and the precision rates of the three different methods were considered objects. The input for this test is the precision rates of the three different retrieval methods; it has two outputs: The Kendall coefficient and the associated level of significance. The first output parameter (W) ranges from 0 (which means there no agreement between a set of ranks) to 1 (which means complete agreement), while the second output parameter (P) indicates whether this coefficient value could have occurred by chance. Here, since all the retrieval performances were taken for the top 10 images, the results will be significant for this test if the value of the (P) parameter is less than or equal to 0.01 for the Corel-1K dataset and less than or equal to 0.001 for GHIM-10K, respectively. The results for this test are shown in Table 5, where the late fusion method has the agreement or confidence of 61.1% for the Corel-1K dataset and 52.4% for the GHIM-10K dataset and

the ranking of the different methods in the corresponding rows in Table 5 shows the good retrieval performance of the proposed methods. Finally, Figs. 8 and 9 show visual examples of the top retrieved images and queries used for the retrieval process; ten out of ten retrieved images were achieved for the Corel-1K dataset while eight out of the top ten images were retrieved in the GHIM-10K dataset.

**Table 5:** Ranking of two retrieval fusion methods based on Kendall W test for precision values at top term retrieved images

Dataset	W	P	Ranking
Corel-1K	0.611	0.0085 <0.01	Late Fusion>CEarly> BEarly> EEearly
GHIM-10K	0.524	0.0009 <0.001	Late Fusion>BEarly> CEarly> EEearly

### 5. Conclusion

This study proposed a retrieval method based on early and late fusion approaches. The proposed method is a simple process used to enhance the retrieval performance of image retrieval based on combining feature descriptors from color and texture spaces; also, combining similarity measure values for three known distance coefficients was implemented. The proposed method has two main

advantages: The automatic retrieval process and simplicity in early and late merging processes. Also, it has an acceptable level of enhancement. Further studies should be performed using more accurate and representative features; also, future studies could include some other descriptor features such as image shapes.

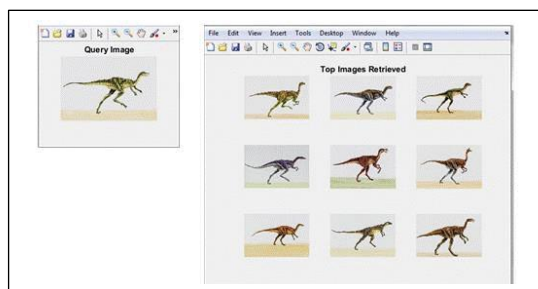


Fig. 8: Query and top retrieved images for Corel-1K

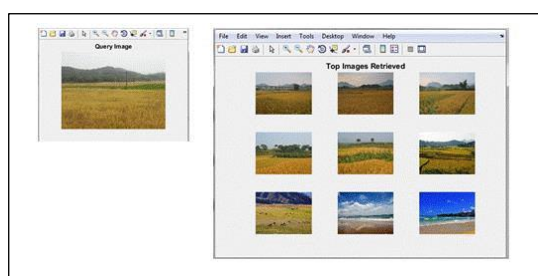


Fig. 9: Query and top retrieved images for GHIM-10K

## Acknowledgment

This research was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia. The authors, therefore, gratefully acknowledge the DSR for their technical and financial support.

## Compliance with ethical standards

## Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## References

Abdelrahim AAA (2013). Fragment reweighting in ligand-based virtual screening. Ph.D. Dissertation, Universiti Teknologi Malaysia, Johor Bahru, Malaysia.

Ahmed A (2020). Implementing relevance feedback for content-based medical image retrieval. *IEEE Access*, 8: 79969-79976. <https://doi.org/10.1109/ACCESS.2020.2990557>

Ahmed A and Malebary S (2019). Feature selection and the fusion-based method for enhancing the classification accuracy of SVM for breast cancer detection. *International Journal of Computer Science and Network Security*, 19(11): 55-60.

Ahmed A and Malebary SJ (2020). Query expansion based on top-ranked images for content-based medical image retrieval. *IEEE Access*, 8: 194541-194550. <https://doi.org/10.1109/ACCESS.2020.3033504>

Ahmed A, Alhaj Alobeed M, and Osman Ibrahim A (2017). Multi-classifier method based on voting technique for mammogram

image classification. *Journal of Software Engineering and Intelligent Systems*, 2(3): 280-285.

Ahmed A, Barukab OM, and Elsadig MA (2019). Heterogeneous multi-classifier method based on weighted voting for breast cancer detection. *International Journal of Advances in Science Engineering and Technology*, 7(4): 36-41.

Ahmed A, Saeed F, Salim N, and Abdo A (2014). Condorcet and borda count fusion method for ligand-based virtual screening. *Journal of Cheminformatics*, 6(1): 1-10. <https://doi.org/10.1186/1758-2946-6-19>  
PMid:24883114 PMCID:PMC4026830

Ahmed A, Salim N, and Abdo A (2013). Fragment reweighting in ligand-based virtual screening. *Advanced Science Letters*, 19(9): 2782-2786. <https://doi.org/10.1166/asl.2013.5012>

Alhassan AK and Alfaki AA (2017). Color and texture fusion-based method for content-based image retrieval. In the International Conference on Communication, Control, Computing and Electronics Engineering, IEEE, Khartoum, Sudan: 1-6. <https://doi.org/10.1109/ICCCCEE.2017.7867649>

Alsmadi MK (2020). Content-based image retrieval using color, shape and texture descriptors and features. *Arabian Journal for Science and Engineering*, 45(4): 3317-3330. <https://doi.org/10.1007/s13369-020-04384-y>

Atrey PK, Hossain MA, El Saddik A, and Kankanhalli MS (2010). Multimodal fusion for multimedia analysis: A survey. *Multimedia Systems*, 16(6): 345-379. <https://doi.org/10.1007/s00530-010-0182-0>

Barnard K and Forsyth D (2001). Learning the semantics of words and pictures. In the 8<sup>th</sup> IEEE International Conference on Computer Vision, IEEE, Vancouver, Canada, 2: 408-415. <https://doi.org/10.1109/ICCV.2001.937654>

Bhowmik N, González R, Gouet-Brunet V, Pedrini H, and Bloch G (2014). Efficient fusion of multidimensional descriptors for image retrieval. In the International Conference on Image Processing, IEEE, Paris, France: 5766-5770. <https://doi.org/10.1109/ICIP.2014.7026166>

Erkut U, Bostancıoğlu F, Erten M, Özbayoğlu AM, and Solak E (2019). HSV color histogram based image retrieval with background elimination. In the 1<sup>st</sup> International Informatics and Software Engineering Conference, IEEE, Ankara, Turkey: 1-5. <https://doi.org/10.1109/UBMYK48245.2019.8965513>

Houle ME, Ma X, Oria V, and Sun J (2017). Query expansion for content-based similarity search using local and global features. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 13(3): 1-23. <https://doi.org/10.1145/3063595>

Ibrahim AO, Ahmed A, Azizah AH, Lashari SA, Alobeed MA, Kasim S, and Ismail MA (2018). An enhancement of multi classifiers voting method for mammogram image based on image histogram equalization. *International Journal of Integrated Engineering*, 10(6): 209-215. <https://doi.org/10.30880/ijie.2018.10.06.030>

Karamti H, Tmar M, Visani M, Urruty T, and Gargouri F (2018). Vector space model adaptation and pseudo relevance feedback for content-based image retrieval. *Multimedia Tools and Applications*, 77(5): 5475-5501. <https://doi.org/10.1007/s11042-017-4463-x>

Kuncheva LI (2014). *Combining pattern classifiers: Methods and algorithms*. John Wiley and Sons, Hoboken, USA. <https://doi.org/10.1002/9781118914564>

Latif A, Rasheed A, Sajid U, Ahmed J, Ali N, Ratyal NI, and Khalil T (2019). Content-based image retrieval and feature extraction: A comprehensive review. *Mathematical Problems in Engineering*, 2019: 9658350. <https://doi.org/10.1155/2019/9658350>

Li J and Wang JZ (2008). Real-time computerized annotation of pictures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6): 985-1002.



- <https://doi.org/10.1109/TPAMI.2007.70847>  
**PMid:18421105**
- Liu GH, Yang JY, and Li Z (2015). Content-based image retrieval using computational visual attention model. *Pattern Recognition*, 48(8): 2554-2566.  
<https://doi.org/10.1016/j.patcog.2015.02.005>
- Maheshwary P and Srivastava N (2009). Prototype system for retrieval of remote sensing images based on color moment and gray level co-occurrence matrix. *International Journal of Computer Science Issues*, 3: 20-23.
- Mistry YD (2020). Textural and color descriptor fusion for efficient content-based image retrieval algorithm. *Iran Journal of Computer Science*, 3(3): 169-183.  
<https://doi.org/10.1007/s42044-020-00056-0>
- Piras L and Giacinto G (2017). Information fusion in content based image retrieval: A comprehensive overview. *Information Fusion*, 37: 50-60.  
<https://doi.org/10.1016/j.inffus.2017.01.003>
- Rui Y, Huang TS, and Mehrotra S (1997). Content-based image retrieval with relevance feedback in MARS. In the International Conference on Image Processing, IEEE, Santa Barbara, USA, 2: 815-818.  
<https://doi.org/10.1109/ICIP.1997.638621>
- Scaroff S, La Cascia M, Sethi S, and Taycher L (1999). Unifying textual and visual cues for content-based image retrieval on the world wide web. *Computer Vision and Image Understanding*, 75(1-2): 86-98.  
<https://doi.org/10.1006/cviu.1999.0765>
- Seetharaman K and Kamarasan M (2014). Statistical framework for image retrieval based on multiresolution features and similarity method. *Multimedia Tools and Applications*, 73(3): 1943-1962. <https://doi.org/10.1007/s11042-013-1637-z>
- Shirkhorshidi AS, Aghabozorgi S, and Wah TY (2015). A comparison study on similarity and dissimilarity measures in clustering continuous data. *PloS ONE*, 10(12): e0144059.  
<https://doi.org/10.1371/journal.pone.0144059>  
**PMid:26658987 PMCID:PMC4686108**
- Siegel S and Castellan Jr. NJ (1988). *Nonparametric statistics for the behavioral sciences*. McGraw-Hill, New York, USA.
- Singh SM and Hemachandran K (2012). Image retrieval based on the combination of color histogram and color moment. *International Journal of Computer Applications*, 58(3): 27-34.  
<https://doi.org/10.5120/9263-3441>
- Snoek CG, Worring M, Geusebroek JM, Koelma D, and Seinstra FJ (2005). On the surplus value of semantic video analysis beyond the key frame. In the International Conference on Multimedia and Expo, IEEE, Amsterdam, Netherlands: 399-402. <https://doi.org/10.1109/ICME.2005.1521441>
- Sumathi T, Devasena CL, and Hemalatha M (2011). An overview of automated image annotation approaches. *International Journal of Research and Reviews in Information Sciences*, 1(1): 1-5.
- Varish N, Pal AK, Hassan R, Hasan MK, Khan A, Parveen N, and Memon I (2020). Image retrieval scheme using quantized bins of color image components and adaptive tetrolet transform. *IEEE Access*, 8: 117639-117665.  
<https://doi.org/10.1109/ACCESS.2020.3003911>
- Veltkamp RC and Tanase M (2002). A survey of content-based image retrieval systems. In: Marques O and Furht B (Eds.), *Content-based image and video retrieval: 47-101*. Springer, Boston, USA. [https://doi.org/10.1007/978-1-4615-0987-5\\_5](https://doi.org/10.1007/978-1-4615-0987-5_5)
- Weszka JS, Dyer CR, and Rosenfeld A (1976). A comparative study of texture measures for terrain classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-6(4): 269-285. <https://doi.org/10.1109/TSMC.1976.5408777>
- Yue J, Li Z, Liu L, and Fu Z (2011). Content-based image retrieval using color and texture fused features. *Mathematical and Computer Modelling*, 54(3-4): 1121-1127.  
<https://doi.org/10.1016/j.mcm.2010.11.044>
- Zenggang X, Zhiwen T, Xiaowen C, Xue-min Z, Kaibin Z, and Conghuan Y (2019). Research on image retrieval algorithm based on combination of color and shape features. *Journal of Signal Processing Systems*, 93: 139-146.  
<https://doi.org/10.1007/s11265-019-01508-y>
- Zhou XS and Huang TS (2002). Unifying keywords and visual contents in image retrieval. *IEEE Multimedia*, 9(2): 23-33.  
<https://doi.org/10.1109/93.998050>