

An Appraisal of Content-Based Image Retrieval (CBIR) Methods

Olaleke J.O.^{1*}, Adetunmbi A.O.², Ojokoh, B.A.³, Iroju Olaronke⁴
^{1,4}Department of Computer Science, Adeyemi College of Education, Ondo, Nigeria
^{2,3}Department of Computer Science, Federal University of Akure, Nigeria

MINI REVIEW

ABSTRACT

Background: Content Based Image Retrieval (CBIR) is an aspect of computer vision and image processing that finds images that are similar to a given query image in a large scale database using the visual contents of images such as colour, texture, shape, and spatial arrangement of regions of interest (ROIs) rather than manually annotated textual keywords. A CBIR system represents an image as a feature vector and measures the similarity between the image and other images in the database for the purpose of retrieving similar images with minimal human intervention. The CBIR system has been deployed in several fields such as fingerprint identification, biodiversity information systems, digital libraries, Architectural and Engineering design, crime prevention, historical research and medicine. There are several steps involved in the development of CBIR systems. Typical examples of these steps include feature extraction and selection, indexing and similarity measurement.

Problem: However, each of these steps has its own method. Nevertheless, there is no universally acceptable method for retrieving similar images in CBIR.

Aim: Hence, this study examines the diverse methods used in CBIR systems. This is with the aim of revealing the strengths and weakness of each of these methods.

Methodology: Literatures that are related to the subject matter were sought in three scientific electronic databases namely CiteseerX, Science Direct and Google scholar. The Google search engine was used to search for documents and WebPages that are appropriate to the study

Results: The result of the study revealed that three main features are usually extracted during CBIR. These features include colour, shape and text. The study also revealed that diverse methods that can be used for extracting each of the features in CBIR. For instance, colour space, colour histogram, colour moments, geometric moment as well as colour correlogram can be used for extracting colour features. The commonly used methods for texture feature extraction include statistical, model-based, and transform-based methods while the edge method, Fourier transform and Zernike methods can be used for extracting shape features.

Contributions: The paper highlights the benefits and challenges of diverse methods used in CBIR. This is with the aim of revealing the methods that are more efficient for CBIR.

Conclusion: Each of the CBIR methods has their own advantages and disadvantages. However, there is a need for a further work that will validate the reliability and efficiency of each of the method.

Keywords: CBIR, features, feature extraction, Image

17 **1. INTRODUCTION**

18

19 CBIR is a term that was first introduced by Kato in 1992 [1]. Content-based image retrieval
20 (CBIR) is also known as query by image content (QBIC) and content-based visual
21 information retrieval (CBVIR). The major aim of a CBIR is to find images of interest from a
22 large image database using the visual content of the images. In addition, CBIR reduces the
23 semantic gap issue that improves the performance of image retrieval [2].

24 CBIR is however entirely different from other classical information retrieval systems because
25 they are highly unstructured. This is because digitized images consist purely of arrays of
26 pixel intensities, with no inherent meaning [3]. They are however cheap, fast and efficient
27 when compared with the text based image search method [4]. CBIR draws its methods from
28 the field of image processing and computer vision. Generally, a Content based image
29 retrieval (CBIR) is a term that is used to describe a retrieval technique which involves the
30 use of visual information or contents called low level features to search and retrieve images
31 from a large scale image database according to the requests of the user which is provided in
32 the form of a query image. Nevertheless, image content may include semantic content [5].
33 The visual information is usually in form of colours, textures, shapes and spatial
34 arrangement of region of interest. A CBIR retrieves relevant images by comparing the
35 features of the images in the database with a given query image as well as finding the
36 images that are similar to the queried image [5]. However, the retrieval of images in an
37 image database using visual attributes is a challenging task due to the close visual
38 appearance among the visual attributes of these images [6]. Thus, a CBIR can be viewed as
39 an image search technique that is intended to search images that are almost similar in terms
40 of colour, shape and text to a given query. Hence, the principal goal of a CBIR is to
41 represent each image as a feature vector and to measure the similarity between the queried
42 image and the images in database and also to retrieve similar images based on the features
43 and not on textual annotations [7]. The general architecture of a CBIR system is as shown in
44 figure 1.

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

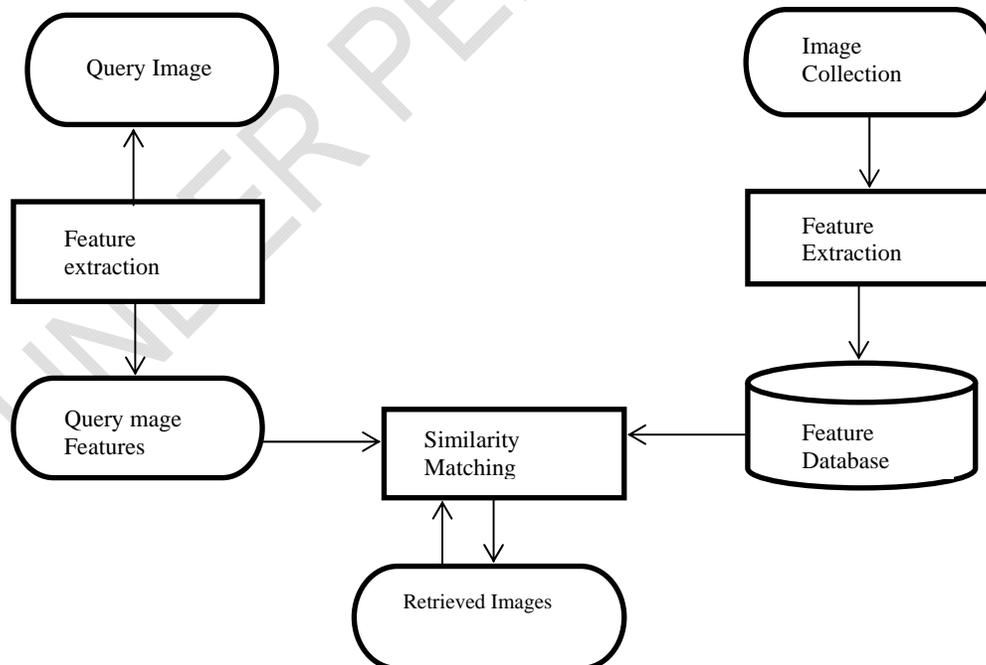


Figure 1: The General Architecture of a CBIR [3]

70 CBIR is performed usually in two steps. These include indexing and searching. During
71 indexing the contents or features of both the queried image and the images in the image
72 database are extracted and stored in the form of a feature vector in a feature database. This
73 process is called the feature extraction. There are several methods that are used for
74 extracting features in CBIR. Examples of features that can be extracted in CBIR include
75 colour, texture and shape. Each of these characteristic features has diverse extraction
76 methods. For colour extraction methods, colour space, colour histogram and colour
77 moments are usually deployed. The commonly used methods for textural feature extraction
78 are described by Manjunath and Ma [8] as statistical, model-based, and transform-based
79 methods. One of the most widely used shape feature extraction method is the Edge method.
80 In the searching step, a user query image feature vector is constructed and compared with
81 all feature vectors in the database for similarity in order to retrieve the most similar images to
82 the query image from the database [9, 10]. This process is referred to as similarity
83 measurement. Again, there are diverse methods for computing similarity between a queried
84 image and the images in the database. Typical examples of the methods used for similarity
85 measurement include Sum of Absolute Difference (SAD), Sum of squared absolute
86 Difference (SSAD), City Block Distance Canberra Distance and Euclidean Distance.
87 Nevertheless, there is no universally acceptable method for extracting features and
88 retrieving similar images in CBIR. Hence, this study examines the diverse methods used in
89 CBIR systems. This is with the aim of revealing the strengths and weakness of each of these
90 methods.

91 The paper is as organized as follows: section 2 is the methodology, section 3
92 reviews the features in CBIR, section 4 examines feature extraction methods in CBIR while
93 section 5 examines the methods for computing similarity between a queried image and the
94 images in a database. **Section 6 discusses the paper while the paper concludes in section 7.**
95

96 97 **2. METHODOLOGY**

98
99 Literatures that are related to the subject matter were sought in three scientific electronic
100 databases namely CiteseerX, Science Direct and Google scholar. The Google search
101 engine was used to search for documents and WebPages that are appropriate to the study
102

103 **3. FEATURES IN CBIR**

104
105 Features are observable patterns in the image that contain relevant information of an image.
106 Pradeep et al. [11] also viewed a feature as a piece of information that is relevant for solving
107 a computational task related to a certain application. Features describe and define the
108 content of an image. They are described as the characteristics or the properties of the
109 image. Features are usually used in image processing for searching, retrieval, and storage in
110 order to achieve a high classification rate. The main goal of feature extraction is to obtain
111 the most relevant information from the original data and represent that information in a lower
112 dimensionality space .A good feature set contains discriminating information, which can
113 distinguish one object from other objects [12, 13]. One of the easiest ways of comparing
114 images is through their features. This is because the direct method of comparing images by
115 their pixels is not feasible and it is also time consuming for thousands of images stored in
116 databases. Examples of features that can be extracted from images include colour, texture
117 and shape. However, no particular feature is most suitable for retrieving all types of images.
118

119 **3.1 Colour Features**

120 Colour is the sensation caused by the light as it interacts with the human eyes and brain.
121 Colours simplify objects identification. Color is one of the most widely used low-level visual
122 features used in CBIR. However, different images can have the same colour distribution.

123 Unfortunately, the retrieval of images with colour features only does not give accurate result
124 because in many cases, images with similar colors do not have similar content.

125
126

127 **3.2 Textural Features**

128 There is no standard definition for texture. Haralick et al. [13] define texture as a
129 characteristic of an image that provides a higher-order description of the image and includes
130 information about the spatial distribution of tonal variations or gray tones. Texture according
131 to Hiremath and Pujari [8] is an innate property of virtually all surfaces, including clouds,
132 trees, bricks, hair and fabric. Texture contains important information about the structural
133 arrangement of surfaces and their relationship to the surrounding environment [13]. Texture-
134 based features are capable of finding the spatial variations between intensity values and
135 surface attributes of an object within an image [14]. Texture can also be defined as the
136 pattern of information or arrangement of the structure of an image. An image can have more
137 than one texture.

138

139 **3.3 Shape Features**

140 The shape of an object can be defined as the characteristic surface configuration of the
141 object as defined by the outline or contour [4]. Shape features provide information for image
142 retrieval, because humans can recognize objects solely from their shapes and shape carries
143 semantic information about an object. Shape is highly significant in CBIR because it
144 corresponds to region of interests in the images. There are two types of shape features.
145 These include boundary-based and region-based features. The boundary based features
146 extracts features based on the outer boundary of a region while the region-based features
147 extracts features based on the entire region of an object [15].

148
149

150 **4. FEATURE EXTRACTION IN CBIR**

151

152 According to Nithya and Santhi [16], feature extraction in CBIR is a method of capturing the
153 visual content of an image. Feature extraction can also be described as the process of
154 extracting information that is semantically meaningful from images. The objective of feature
155 extraction is to represent a raw image in a reduced form in order to facilitate decision making
156 process. Hence, Kayode [17] views feature extraction as a special form of dimensionality
157 reduction which takes place when the input data to an algorithm is too large to be processed
158 and it is suspected to be redundant.

159
160

161 **4.1 Colour Extraction Methods**

162 There are diverse methods for extracting colour features in an image. These include colour
163 space, colour histogram and colour moments.

164

165 **4.1.1 Colour space**

166 Colours are usually defined in three-dimensional color space so as to facilitate the
167 specification of colours in an acceptable way. The RGB colour space is the most widely used
168 color space. RGB stands for Red, Green, and Blue. RGB colour space combines the three
169 colors in different ratio to create other colors. One of the major disadvantages of RGB
170 colour space as emphasized by Mikhraq [4] that the RGB colour space is not uniform. The
171 HSx color space is commonly used in digital image processing to convert the color space of
172 an image from RGB color space to one of the HSx color spaces. HSx color comes in diverse
173 forms. These include the HSI, HSV and HSB color spaces. The H and S in these colour
174 space represents Hue and Saturation while the I, V, and B stand for Intensity, Value, and

175 Brightness respectively. HSV color space is however the most commonly used colour space
176 [18].

177

178

179 **4.1.2 Color histograms**

180 A color histogram according to Sivakamasundari and Seenivasagam [19] is a type of bar
181 graph, where the height of each bar represents an amount of particular colour of the colour
182 space being used in the image. The bars in a colour histogram are referred to as bins and
183 they represent the x-axis and the number of bins relies on the number of colours in the
184 image. The number of pixels in each bin is represented by the y-axis of the bar graph. There
185 are two basic methods of obtaining a colour histogram. These include the global color
186 histogram (GCH) and the local color histogram (LCH). GCH method takes the histogram of
187 the image and computes the distance between two images by measuring the distance
188 between their colour histograms. The drawback of the GCH as emphasized by Mikhraq [4] is
189 that this method does not include information about all image regions. An LCH on the other
190 hand divides an image into fixed blocks or regions, and takes the colour histogram of each of
191 those blocks individually [4]. The similarity between two images is compared using LCH by
192 computing the distance between the blocks of the images in the same location. The
193 advantage of the LCH over the GCH is that the LCH is more efficient for image retrieval.
194 However Mikhraq [4] stated that the LCH is computationally expensive and it does not work
195 well when images are translated or rotated.

196

197

197 **4.1.3 Colour moments**

198 Color moments are measures that are used to compute the similarity of images based on
199 their colour features. In color moments, the colour distribution of an image is seen as a
200 probability distribution which is characterized by unique moments which include mean,
201 standard deviation and skewness. The mean is defined as the average colour value in the
202 image; the standard deviation is the square root of the variance of the distribution while the
203 skewness is a measure of the degree of asymmetry in the distribution [4].

204

205

205 **4.1.4 Geometric moment**

206 This feature uses one value for the feature vector, thus, when the size of the image becomes
207 relatively large, the computation of the feature vector will require a lot of time. The advantage
208 of this method is that it produces a better result when combined with other feature extraction
209 methods [20]. The drawback of geometric moments is that higher order moments are difficult
210 to construct.

211

212

212 **4.1.5 Colour correlogram**

213 Colour correlogram is used for encoding the color information of an image [20]. A colour
214 correlogram is a three-dimensional table indexed by colour and distance between pixels
215 which expresses how the spatial correlation of colour changes with distance in a stored
216 image. The colour correlogram may be used to distinguish an image from other images in a
217 database. To create a colour correlogram, the colours in the image are quantized into m
218 colour values $c_1 \dots c_m$. The advantage of this method is that it can be used to describe the
219 global distribution of local spatial correlation of colours. It is also simple to compute.

220

221

221 **4.1.6 Average RGB**

222 The color average is described by Sharma and Signh [21] in the RGB color space by X, as
223 shown in equation 1.

224

225

$$X = (R(avg), G(avg), B(avh)) ^t \quad (1)$$

226

227

where R(avg), G(avg), and B(avg) are red, green and blue images average value

228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250

Sharma and Signh [21] emphasized that this feature is used to filter out images with larger distance at first stage when multiple feature queries are involved.

4.1.7 Dominant Colour Descriptor (DCD)

This method is based on colour histogram. DCD chooses a small number of colors from the highest bins of a histogram. The number of bins chosen depends on the threshold value of the bin height [21].

4.1.8 Colour Coherence Vector

The colour coherence method is also based on the colour histogram. According to Sharma and Signh [21], the colour coherence divides a histogram into two components namely coherent and non-coherent components. In coherent component, the pixels are spatially connected while in non-coherent component the pixels are isolated.

The advantages and disadvantages of the colour based extraction methods are summarized in Table 1.

Table 1. Pros and Cons of Different Colour Extraction Methods

Methods	Advantages	Disadvantages
Colour Space	RGB is easy to implement. HSV represents a wealth of similar colours; they are also device dependent	RGB colour space is not uniform.
Colour Histogram	comparison of histogram features is faster and more efficient than other methods	It can be potentially identical for two images with different colour content
Colour Moments	There is no need to store the complete colour distribution; hence image retrieval is faster with this technique. It also involves the comparison of less features	They cannot handle occlusion successfully
Geometric Moments	produces a better result when combined with other feature extraction methods	higher order moments are difficult to construct, thus they are difficult to compute
Colour Correlogram	It is simple to compute and may be used to distinguish an image from images in a database. It takes into cognizance the local colour spatial correlation. Effective for CBIR from a large image database.	they have high computational complexity and low retrieval accuracy

Average RGB	Its computation cost is not high	it is less accurate if not combined with other feature extraction methods
DCD	It is scalable and accurate if compact	it does not give spatial information of the image
Colour Coherence Vector	it gives spatial information about the image	it has high computational cost

251
252

253 **4.1 Texture Extraction Techniques**

254 The commonly used methods for texture feature extraction are described by Manjunath and
255 Ma [8] as statistical, model-based, and transform-based methods.

256

257 **4.2.1 Statistical Approaches**

258 In statistical methods, the spatial distribution of grey values is computed by finding the local
259 features at each point in the image, and deriving a set of statistics from the distribution of the
260 local features. Typical examples of statistical approaches include gray level co-occurrence
261 matrix (GLCM) and Tamura features.

262

263 *4.2.1.1 Gray Level Co-Occurrence Matrix (GLCM)*

264 Gray Level Co-Occurrence Matrix (GLCM) is usually computed to discriminate different
265 textures. The GLCM is a function of an angular relationship between two pixels with
266 corresponding gray level, i and j and a function of the distance between them which
267 characterizes the spatial distribution of gray levels between them. An element in the GLCM,
268 $P_{d,\theta}(i,j)$, represents the frequency of occurrence of the pair of gray levels (i,j) , separated by
269 a distance d at a direction θ . When each entry in the matrix is divided by the total number of
270 neighbouring pixels R , a normalized GLCM is obtained, and the sum of its elements is equal
271 to 1.

272 The notations in equations (2)-(5) are used to describe the various textual features in GLCM.

$$273 \quad p(i,j) = \text{the } (i,j)^{\text{th}} \text{ entry in a normalized GLCM given by } P_{d,\theta}(i,j)/R \quad (2)$$

274

$$275 \quad N_g = \text{the number of distinct gray levels in quantized image} \quad (3)$$

276

$$277 \quad p_x(i) = \sum_{j=1}^{N_g} p(i,j), \text{ the } i^{\text{th}} \text{ entry in the marginal probability matrix} \quad (4)$$

278

obtained by summing the rows of $p(i,j)$ and $p_x(i)$ is the i^{th} entry of row i

279

$$280 \quad p_y(j) = \sum_{i=1}^{N_g} p(i,j) \text{ the } j^{\text{th}} \text{ entry in the marginal probability matrix} \quad (5)$$

281

obtained by summing the rows of $p(i,j)$ and $p_y(j)$ is the j^{th} entry of column j

282

283 Hence, the general equations for the Haralick features are given in equations (6) and (7)
284 respectively.

$$285 \quad p_x + y(k) = \sum_{i=1}^{N_g} (i,j) \sum_{i+j=k}^{j=1} p(i,j), \quad k = 2,3, \dots, 2N_g \quad (6)$$

$$286 \quad p_x + y(k) = \sum_{i=1}^{N_g} (i,j) \sum_{|i-j|=k}^{j=1} p(i,j), \quad k = 0,1, \dots, N_g - 1 \quad (7)$$

287

288 Based on the notations above, Haralicks et al. [13] proposed 13 common statistical features
289 known as the Haralicks textual features. The 13 Haralick features are given in equations (8)
290 to (25)

291

292 i. Energy: This is also known as the angular second moment. It measures the textual
293 uniformity of an image. Energy is as given in equation (2.20).

294
$$energy = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (p(i, j))^2 \quad (8)$$

295 ii. Contrast: This is a measure of intensity or gray-level variations between the reference
296 pixel and its neighbor. Contrast is as given in equation (9).

297
$$contrast = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i - j|^2 p(i - j) \quad (9)$$

298

299 iii. Correlation: This presents how a reference pixel is related to its neighbour. Correlation is
300 expressed in equation (10) as follows:

301
$$correlation = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i, j) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (10)$$

302

303 Where μ_x, μ_y, σ_x and σ_y are the means and standard deviations of $p(x)$ and $p(y)$ respectively.

304

305 iv. Homogeneity: This is also known as Inverse Difference Moment. It measures image
306 homogeneity. Heterogeneity is as given in equation (11).

307
$$Homogeneity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{1}{1+(i-j)^2} p(i, j) \quad (11)$$

308

309 v. Entropy: This measures the disorder or complexity of an image. Entropy is as given in
310 equation (12).

311
$$Entropy = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \log(p(i, j)) \quad (12)$$

312

313 vi. Variance: This is also referred to as the sum of squares. It measures the dispersion of
314 the difference between the reference and the neighbour pixel in a window. Variance is
315 as given in equation (13).

316
$$Variance = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 p(i, j) \quad (13)$$

317 Where μ is the mean gray level of the image

318

319 vii. Sum Average: The sum average is given as shown in equation (14).

320
$$Sum\ Average\ (SA) = \sum_{k=2}^{2N_g} k \cdot p_{x+y}(k) \quad (14)$$

321

322 viii. Sum Entropy: Sum Entropy is given as shown in equation (15).

323
$$Sum\ Entropy, SE = - \sum_{k=2}^{2N_g} p_{x+y}(k) \log(p_{x+y}(k)) \quad (15)$$

324

325 ix. Sum Variance: This is given in equation (16).

326
$$Sum\ Variance\ (SV) = \sum_{k=2}^{2N_g} (k - SE)^2 p_{x+y}(k) \quad (16)$$

327
328 x. Difference Variance: This is expressed in equation (17).

$$329 \quad \text{Difference Variance (DV)} = \sum_{k=0}^{N_g-1} k^2 p_{x-y}(k) \quad (17)$$

330
331 xi. Difference Entropy: The difference entropy is as given below in equation (18)

$$332 \quad \text{Difference Entropy (DE)} = -\sum_{k=0}^{N_g-1} p_{x-y}(k) \log(p_{x-y}(k)) \quad (18)$$

333
334 xii. Information Entropy of Correlation (IEC): This is expressed in equation (19).
335

$$336 \quad IEC = \frac{HXY - HXY1}{\max(HX, HY)} \quad (19)$$

337 Where

$$338 \quad HXY = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \log(p(i, j)) \quad (20)$$

$$339 \quad HXY1 = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \log(p_x(i) p_y(j)) \quad (21)$$

340
341
342 HX and HXY1 are the entropies of p_x and p_y respectively
343

344 xiii. Information Measure of Correlation (IMC): This is as expressed in equations (22)
345 and (23) respectively.
346

$$347 \quad IMC = (1 - \exp(-2(HXY2 - HXY)))^{1/2} \quad (22)$$

348 Where

$$349 \quad HXY2 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_x(i) p_y(j) \log(p_x(i) p_y(j)) \quad (23)$$

350

351 4.2.1.2 Tamura Features Extraction Method

352 Tamura et al. [22] also proposed texture representations that were based on
353 psychological studies of human perception, and these representations consists of six
354 statistical features, including coarseness, contrast, directionality, regularity, line-likeness,
355 roughness to describe various texture properties.
356

357 i. Contrast: Contrast measures the distribution of gray levels that varies in an image and to
358 what extent its distribution is biased to black or white. The second order and normalized
359 fourth-order central moments of the gray levels are used to define the contrast. Contrast is
360 given as in equations (24) and (25) respectively.

$$361 \quad \text{Contrast} = \sigma / (\alpha 4) \quad (24)$$

$$362 \quad \alpha 4 = \mu 4 / \sigma 4 \quad (25)$$

363 $\mu 4$ is the fourth moment about the mean and σ is the variance.
364

365 ii. Directionality: Directionality of an image is measured by the frequency distribution of
366 oriented local edges against their directional angles. This texture feature given by Tamura
367 does not differentiate between orientations or patterns but measures the total degree of
368 directionality in an image. Directionality is expressed as depicted in equation (26).

$$369 \quad \text{Directionality} = 1 - \text{rnpeaks} \sum_{p=1}^{npeaks} \sum a \in w_p (a - \text{ap})^2 H_{\text{directionality}}(a) \quad (26)$$

370

371 where n_{peaks} is the number of peaks, a_p , is the position of the peak, w_p is the range of the
 372 angles attributed to the P^{th} peak, r denotes a normalizing factor related to quantizing levels
 373 of the angles a , and a denotes quantized directional angle, $HDirectionality$, is the histogram
 374 of quantized direction values, a is constructed by counting number of the edge pixels with
 375 the corresponding directional angles.

376
 377 iii. Line-Likeness: Line-Likeness in an image is the average coincidence of direction of edges
 378 that co-occurred in the pairs of pixels separated by a distance along the edge direction in
 379 every pixel.

380
 381 iv. Regularity: Regularity measures a regular or similar pattern that occurred in an image.
 382 Regularity is defined in equation (27) as follows:

$$383 \quad \text{Regularity} = 1 - r(\text{Scrs} + \text{Scon} + \text{Sdir} + \text{Slin}) \quad (27)$$

384 Where $Scrs$, $Scon$, $Sdir$ and $Slin$ are similar coarseness, contrast, directionality and line-
 385 likeness in an image respectively.

386
 387 v. Roughness: Roughness is the summation of contrast and coarseness measures.
 388 Roughness is as shown in equation (28).

$$389 \quad \text{Roughness} = \text{Contrast} + \text{Coarseness} \quad (28)$$

390
 391 vi. Coarseness: Coarseness basically relates to the distance in gray levels of spatial
 392 variations, which is implicitly related to the size of primitive elements forming the texture.
 393 It has the direct relationship to scale and repetition rates and most fundamental texture
 394 features. Coarseness is expressed as shown in equation (29).

$$395 \quad A_k(x, y) = \frac{\sum_{i=x-2k-1}^{x+2k-1} \sum_{j=y-2k-1}^{y+2k-1} f(i, j)}{2^{2k}} \quad (29)$$

396 Where 2^{2k} size is the average of neighborhood.

397
 398 In most cases, coarseness, contrast and directionality are commonly used for CBIR systems
 399 because they capture high-level perceptual attributes of a texture and are also useful for
 400 browsing of images [22].

401 402 **4.2.2 Model Based Approaches**

403 Model-based texture methods are used to compute the process that generated the texture.
 404 The model based approach is generated by computing a random field as stated by Mikhraq
 405 [4] as follows:

406
 407 Assuming an image is modeled as a function $f(r)$, where r is the position vector
 408 representing the pixel location in the 2-D space and ω is a random parameter. For a given
 409 value of r , $f(r)$ is a random variable because ω is a random variable. Once a specific
 410 texture ω is selected, $f(r)$ is an image, which is a function over the two-dimensional grid
 411 indexed by r . Function $f(r)$ is called a random field. A typical example of the model based
 412 approach is the Markov random fields.

413 414 **4.2.3 Transform-Based Methods**

415 Typical examples of transform based methods include Fourier transform and wavelet
 416 analysis.

417 418 **4.2.3.1 Fourier transform**

419 In Fourier transform, the image signal is broken into sine waves of various frequencies. A
 420 variant of the Fourier transform is the Fast Fourier Transform (FFT). The FFT according to
 421 Shukla and Vania [20] refers to a class of algorithms for efficiently computing the Discrete
 422 Fourier Transform (DFT). Hence, Shukla and Vania [20] emphasized that FFT is not an
 423 approximation of the DFT, but rather it is the DFT with a reduced
 424 number of computations. One of the disadvantages of the FT is that it does not capture the
 425 objects locations in an image [19].
 426

427 *4.2.3.2 Discrete wavelet transform*

428 Discrete Wavelet Transform involves the decomposition of an image into basic functions
 429 obtained through translation and dilation of a special function. The Discrete Wavelet
 430 Transform is very effective in image analysis and compression [19].
 431

432 *4.2.3.3 Ranklet transform*

433 The Ranklet Transform belongs to a family of non-parametric, orientation-selective, and
 434 multi-resolution features. This method has three main properties. First, it is nonparametric
 435 because it deals with the relative order of pixels instead of their intensity values. Second, it is
 436 orientation selective because it is modeled on Haar wavelets. Lastly, it is multi-resolution.
 437 This implies that the Ranklet Transform can be calculated at different resolutions using Haar
 438 wavelet supports. The Ranklet Transform performs better than the pixel-based and wavelet-
 439 based image representations.
 440

441 *4.2.3.4 Steerable pyramid*

442 This technique generates a multi-scale, multidirectional representation of the image [20]. It
 443 involves the decomposition of the image into low-pass sub-band and high-pass sub-band.
 444 However, the decomposition is iterated in the low-pass sub-band [22].
 445

446 The advantages and disadvantages of the textural based extraction methods are
 447 summarized in Table 2.
 448
 449

450 **Table 2. Pros and Cons of Different Textural Extraction Methods**

Methods	Class	Advantages	Disadvantages
GLCM	Statistical	GLCM produces better results than other texture discrimination methods. It also enhances the details of an image and gives the interpretation. It reduces image compression time. It is a good discriminator when studying images. Hence, it is a widely used textural extraction method	Characterized by high development cost and time
Tamura	Statistical	It motivates human visual perception	works only on homogenous texture images and performs poor on generic images
Fourier Transform	Transform based	It improves signal to noise ratio	it does not capture the objects locations in an image
Discrete Transform	Transform	Discrete Wavelet Transform is very	It has poor directionality.

Wavelet	based	effective in image analysis and compression	
Ranklet Transform	Transform based	It performs better than the pixel-based and wavelet-based image representation. They are robust in detecting outliers	It has high computational cost
Steerable Pyramid	Transform based	it allows the independent representation of scale and orientation of image structure	space-domain implementation is not perfect

451

452 **4.3 Shape Extraction Techniques**

453 Examples of shape extraction methods include the edge method, Fourier descriptor and
454 Zernike method. These methods are briefly described below.

455

456 **4.3.1 Edge method**

457 One of the most widely used shape feature extraction method is the Edge method. Edge is
458 used to capture the information about the shape of an object. A typical variance of edge is
459 the edge histogram. It is used to represent the relative frequency of occurrence of five types
460 of edges in each local area called a sub image or an image block. The sub-image is
461 obtained by dividing the image space into a 4x4 non-overlapping blocks. Thus, the image
462 partition always yields 16 equal-sized sub-images regardless of the size of the original image
463 [3].

464

465 **4.3.2 Fourier descriptors**

466 This technique involves the application of Fourier transform on the shape boundary of an
467 image. The Fourier transformed coefficients are usually referred to as the Fourier
468 descriptors (FD) of the shape. They are robust and easy to derive [21]. Fourier descriptors
469 are not affected by noise [20].

470

471 **4.3.3 Zernike moments**

472 This method allows independent moment invariants to be constructed to an arbitrarily high
473 order [19]. It is suitable for complex shape representation and does not need to know
474 boundary information of the image [20].

475

476

477 The advantages and disadvantages of the shape extraction methods are summarized in
478 Table 3.

479

480 **Table 3. Pros and Cons of Different Shape Extraction Methods**

Methods	Advantages	Disadvantages
Edge Method	captures the information about the shape of an object. It also locates sharp discontinuities in an image	The discontinuities abrupt changes in pixel intensity scene.
Fourier descriptors	They are robust and easy to derive	for Fourier descriptors to be accurate, their values are usually calculated and stored in float numbers
Zarnike Method	suitable for complex shape	Computational complexity is

representation and does not need high
to know boundary information of
the image

481
482
483
484

5. Similarity Measures in CBIR

486
487
488
489
490
491
492
493

The similarity between two images, represented by their features values, is defined by a similarity measure [4]. In similarity measurement, the query image is compared with the images in the database. Similarity measure is usually computed by finding the similarity between the query image and the database images. This is usually done by computing the difference between the query feature vector and the database feature vectors. Typical examples of the distance metrics used in CBIR include the following:

494
495
496
497
498
499

5.1 Sum of absolute difference (SAD)

The sum of absolute difference (SAD) is extensively used for computing the distance between the images in CBIR to get the similarity. In this metric, the sum of the differences of the absolute values of the two feature vectors, Q_i and D_i is calculated. This distance metric according to Selvarajah and Kodituwakku [24] can be calculated as shown in equation (30).

500
501

$$\Delta d = \sum_{i=1}^n (|Q_i| - |D_i|) \quad (30)$$

502
503
504
505

where n is the number of features, $i= 1, 2, \dots, n$. Both images are the same for $d= 0$ and the small value of Δd shows the relevant image to the query image.

SAD is simple when the query image and the image in the database are similar [23, 24]

506

5.3 Sum of squared absolute Difference (SSAD)

507
508
509
510

In this metric, the sum of the squared differences of absolute values of the two feature vectors is calculated. This distance metric according to Selvarajah and Kodiyuwakku [24] can be calculated as shown in equation (31).

511
512

$$\Delta d = \sum_{i=1}^n (|Q_i| - |D_i|)^2 \quad (31)$$

SSAD is more computationally complex than SAD.

513
514

5.4 Euclidean distance

515
516
517
518
519

This distance metric is the most commonly used for similarity measurement in image retrieval because of its efficiency and effectiveness [11, 25]. It measures the distance between two vectors of images by calculating the square root of the sum of the squared absolute differences and it can be calculated as shown in equation (32).

520
521
522

$$\Delta d = \sqrt{\sum_{i=1}^n (|Q_i| - |D_i|)^2} \quad (32)$$

523

5.5 City block distance

524
525
526
527

This distance metric is also called the Manhattan distance. The city block distance metric has robustness to outliers. This distance metric according to Szabolcs [26] is computed by the sum of absolute differences between two feature vectors of images and can be calculated as shown in equation (33).

528
529
530

$$\Delta d = \sum_{i=1}^n (|Q_i| - |D_i|) \quad (33)$$

531 The city block distance metric gives a large value for the two similar images which create
 532 dissimilarity between similar images.

533

534 **5.6 Canberra distance**

535 This metric is used for numerical measurement of the distance between the query and
 536 database feature vectors. The value of this method is arranged in ascending order such that
 537 the top most shows high similarity [27]. It has similarity with city block distance metric
 538 [26]. Canberra distance is computed as shown in equation (34).

539

540
$$\Delta d = \sum_{i=1}^n \frac{|Q_i - D_i|}{|Q_i + D_i|} \quad (34)$$

541 The advantages and disadvantages of the diverse similarity methods are summarized in
 542 Table 4.

543

544

545

546 **Table 4. Pros and cons of different methods for similarity measurements in**
 547 **CBIR**

Methods	Advantages	Disadvantages
Sum of Absolute Difference (SAD)	It is simple when the query image and the image in the database are similar	it is sensitive to background issues of images such as variations in size, color, illumination and direction of light
Sum of squared absolute Difference (SSAD)	It be used in both pixels and transformed domains but in the transformed domain	SSAD is more computationally complex than SAD
Euclidean Distance	it is the most efficient and effective similarity measure	It assumes that the sample points are distributed about the sample mean in a spherical manner
City Block Distance	has robustness to outliers	gives a large value for two similar images which create dissimilarity between similar images
Canberra Distance	It is good for data that are spread about the origin	It can only be used for positive values.

548

549

550

551

552

553

554

555

556

557

558

6. DISCUSSION

This study investigates features extracted in images during content based image retrieval. The study reveals that three main features can be extracted from images. These include colour, texture and shape. The study however revealed that there is no specific feature that is most suitable for retrieving all types of images. The study also examines the methods of capturing the visual content of an image. The study shows that colour can be extracted from an image using colour space, colour histogram and colour moments, geometric moment, colour correlogram and average RGB. The study showed that the colour space is easy to implement, however it is not uniform. The study also revealed that the colour histogram is faster and more efficient than other methods, but it potentially identical for two images with

559 different colour content. The colour moments involves the comparison of less features,
560 nonetheless, they cannot handle occlusion successfully. The study also showed that colour
561 correlogram is simple to compute but have high computational complexity and low retrieval
562 accuracy. The study also identifies the methods for extracting textural features; the methods
563 revealed by the study include GLCM, Tamura, Fourier Transform, discrete wavelet, Ranklet
564 transform and steerable pyramid. GLCM produces better results than other texture
565 discrimination methods, but it is characterized by high development cost and time. Tamura
566 feature extraction method motivates human visual perception but performs poor on generic
567 images. Fourier transfer improves signal to noise ratio but does not capture the objects
568 locations in an image. Ranklet transform are robust in detecting outliers. It however has high
569 computational cost. The study also examines shape extraction methods. The study revealed
570 that edge method, Fourier descriptors and Zarnike method can be used for extracting shape
571 features in images. The study however revealed that the edge method locates sharp
572 discontinuities in an image but the discontinuities abrupt changes in pixel intensity scene.
573 The Fourier descriptors are robust and easy to derive but their values are usually stored in
574 float numbers for them to be accurate. The Zarnike method does not need to know boundary
575 information of the image, however it has a high computational complexity. The study also
576 views different similarity measurements in CBIR. The study shows that Sum of Absolute
577 Difference (SAD), Sum Of Squared Absolute Difference (SSAD), Euclidean distance, city
578 block distance and canberra distance are some measures of similarity distance used in
579 CBIR. The study revealed that SAD is simple when the query image and the image in the
580 database are similar but sensitive to background issues of images such as variations in size,
581 color, illumination and direction of light. SSAD can be used in both pixels and transformed
582 domains but SSAD is more computationally complex than sad. The study showed that the
583 Euclidean distance it is the most efficient and effective similarity measure, however, it
584 assumes that the sample points are distributed about the sample mean in a spherical
585 manner.

586

587 The local features of an image such as shape, color, and texture are not sufficient for
588 effective CBIR [28]. Hence, Uzma et al. [28] emphasized that visual similarity is necessary
589 in CBIR. Hence to improve CBIR, Uzma et al. [28] proposed the use of Scale-Invariant
590 Feature Transform (SIFT) and Binary Robust Invariant Scalable Key points (BRISK)
591 descriptors. The sift descriptor detects objects robustly under cluttering due to its
592 invariance to scale, rotation, noise, and illumination variance but it does not perform well at
593 low illumination or poorly localized key points within an image. the brisk descriptor on the
594 other hand provides scale and rotation-invariant scale-space, high quality and adaptive
595 performance in classification based applications [28]. It also performs better at poorly
596 localized key points along the edges of an object within an image as compared to the sift
597 descriptor. Muhammed [29] also proposed the use of Local Intensity Order Pattern (LIOP)
598 Descriptors. The LIOP performs better than sift descriptor when the contrast and the
599 illumination of an image change [29].

600

601

602 7. CONCLUSION

603 CBIR is a fast-developing technology with considerable potential in digital libraries,
604 architectural and engineering design, crime prevention, historical research and medicine.
605 Nevertheless, the effectiveness of current CBIR systems is inherently limited because they
606 only operate at the primitive feature level. Furthermore, the technology still lacks maturity,
607 and is not widely used on a significant scale. Consequently, study examines different
608 techniques used in CBIR systems. The study reviewed diverse literatures that are related to
609 CBIR. The study found out that there are three basic features that can be extracted in CBIR.
610 These include colour, texture and shape. The study also revealed that each of these
611 features has different extraction methods. For instance, colour can be extracted in images

612 using colour histogram, geometric moments, colour sp[ace and colour moments. The study
613 revealed the strengths and weaknesses of each of these techniques. For instance, the
614 colour space method is easy to implement but it is not uniform while the colour histogram is
615 faster and more efficient than other colour extraction methods. It can however be identical
616 for two images with different colours. The study also reveals that the GLCM, Tamura,
617 Fourier transform, Ranklet transform and discrete wavelets are typical examples of textural
618 extraction methods. Similarly, the edge method, Fourier descriptors and Zernike method
619 were the shape extraction methods revealed in this study. Furthermore, the study
620 investigated the techniques for computing the similarity between a query image and the
621 images in the database. The result of the study showed that examples of similarity measures
622 used in CBIR include sum of absolute difference, sum of the squared differences of absolute
623 values and city block distance.

624
625 In recent times, there is no general breakthrough in CBIR in spite of the diverse methods
626 and tools developed to formulate and execute queries in large databases based on their
627 visual contents. Hence, future works should be tailored towards the development of CBIR
628 systems that will resolve the problem of semantic gap in CBIR.

630 **COMPETING INTERESTS**

631 Authors have declared that no competing interests exist.

634 **REFERENCES**

- 636 1. Lew M, Sebe N, Djeraba C, Jain R. Content-based multimedia information retrieval:
637 state of the art and challenges. *ACM Transactions on Multimedia Computing,*
638 *Communications and Applications.* 2006; 2(1): 1-19.
- 639 2. Zahid M, Fakhar A, Toqeer M, Muhammad AJ, Amjad R, Tabass N. Content-based
640 image retrieval based on visual words fusion versus features fusion of local and global
641 features. *Arabian Journal for Science and Engineering.* 2018; 43, (12), 7265–7284.
- 642 3. Sukhdeep K, Deepak A. Image content based retrieval system using cosine
643 similarity for skin disease images. *technique-survey. International Journal of Computer*
644 *Science and Information Technologies.* 2013; 5 (4):89-95.
- 645 4. Mikhraq AK Content-based image retrieval (cbir) system based on the clustering and
646 genetic algorithm. A Master's Thesis in Bioengineering Computer Engineering
647 Department, Faculty of Engineering Deanery of Higher Studies, Islamic University Gaza
648 Palestine. 2013.
- 649 5. Singh N, Dubey S, Dixit P, Gupta J. Semantic image retrieval by combining colour,
650 texture and shape features. *International Conference on Computing Sciences.* 2012.
- 651 6. Amna S, Zahid M, Tanzila S, Khurram AQ, Ahmed A, Habibullah J. A novel method for
652 content-based image retrieval to improve the effectiveness of the bag-of-words model
653 using a support vector machine. *Journal of Information Science.* 2018; 45(1): 117-135.
- 654 7. Hiremath P, Pujari J. Content based image retrieval using colour, texture and shape
655 features. 15th International Conference on Advanced Computing and Communications.
656 2008.
- 657 8. Manjunath B, Ma W. Texture features and learning similarity. *Proceedings of the IEEE*
658 *Conference on Computer Vision and Pattern Recognition (CVPR'96), San Francisco,*
659 *California.* 1998; 425–430.
- 660 9. Nezamabadi-pour, Hossein, Saryazdi, Saeid. Object-based image indexing and retrieval
661 in DCT domain using clustering techniques. *World Acad. Sci. Eng. Technol.* 2005; 3–6.
- 662 10. Veltkamp RC, Tanase M, Sent D. Content-based image retrieval systems: A survey.
663 Utrecht University, Department of Computing Science, Utrecht, The Netherlands, 2002;
664 97-124

- 665 11. Pradeep N, Girisha H, Sreepathi B, Karibasappa K. Feature extraction of
666 mammograms. *International Journal of Bioinformatics Research*. 2012; 4(1): 241-244.
- 667 12. Venkat GN, Vijay RV. *Content-Based Image Retrieval Systems*. IEEE Computer Society,
668 Institute of Electrical and Electronics Engineering. 1995;28(9):18-22.
- 669
- 670 13. Haralick RM, Shanmugan KS, Dunstein I. Textural features for image classification. *IEEE*
671 *Trans. Syst., Man, Cybern.* 1973; 3(6):610–621.
- 672 14. Zahid M, Naila G, Muhammad A, Toqeer M, Tanzila S, Amjad R, Muhammad TM. Scene
673 search based on the adapted triangular regions and soft clustering to improve the
674 effectiveness of the visual-bag-of-words model. *EURASIP Journal on Image and Video*
675 *Processing*.
- 676 15. Hiremath P, Pujari J. Content based image retrieval using colour, texture and shape
677 features. 15th International Conference on Advanced Computing and Communications.
678 2008.
- 679 16. Nithya R, Santhi B. Comparative study on feature extraction method for breast cancer
680 classification. *Journal of Theoretical and Applied Information Technology*. 2011; 33(2):
681 220-226.
- 682 17. Kayode AA. Development of a computational model for classification of mammographic
683 images. Ph.D Thesis, Computer Science Department, Obafemi Awolowo University Ile-Ife,
684 Osun State, Nigeria. 2014
- 685 18. Shih TK., Huang J, Wang C, Hung JC, Kao C. An intelligent content-based image retrieval
686 system based on colour, shape and spatial relations. *Proc. Natl.Sci.Counc. ROC(A)*. 2001;
687 25(4): 232-243.
- 688 19. Sivakamasundari G, Seenivasagam V. Different relevance feedback techniques in CBIR: a
689 survey and comparative study. *International Conference on Computing, Electronics and*
690 *Electrical Technologies*. 2012.
- 691 20. Shukla J, Vania J. A survey on CBIR features extraction techniques. *International Journal*
692 *Of Engineering And Computer Science*. 2014; 3(12): 9555-9559.
- 693 21. Sharma, R. Singh, N. Comparative study of different low level
694 feature extraction techniques. *International Journal of Engineering Research and*
695 *Technology*. 2014; 3(4):1454-1460.
- 696 22. Tamura H, Mori S, Yamawaki, T. Texture features corresponding to visual perception.
697 *IEEE Trans. on Systems, Man, and Cybernetics*. 1978; 8(6):460-473
- 698 23. Bagri N, Johari P. A comparative study on feature extraction using texture and shape for
699 content based image. *Proceedings of the SPIE Medical Imaging: PACS and Imaging*
700 *Informatics*. 2016: 5748: 134 -143.
- 701 24. Selvarajah S, Kodituwakku SR. Analysis and comparison of texture features for content
702 based image retrieval. *International Journal of Latest Trends in Computing*. 2011;2:108-
703 113.
- 704 25. Malik F, Baharudin, B. Analysis of distance metrics in content-based image retrieval using
705 statistical quantized features related to breast density patterns. *Journal of Digital Image*
706 *Processing*. 2013; 20(2): 172–190.
- 707 26. Szabolcs S. Colour Histogram Features Based Image Classification in Content Based
708 Image Retrieval System. A comprehensive review on recent relevance feedback
709 techniques in content based image retrieval. *International Journal of Computer Science*
710 *and Information Technologies*. 2008; 3(4); 4818-4822.
- 711 27. Schulz, J *Algorithms-Similarity*. 2007. Assessed 15 September 2017. Available:
712 [http://www.code10.info/index.php?option=com_content&view=article&id=49:article_camber](http://www.code10.info/index.php?option=com_content&view=article&id=49:article_camber_ra-distance&catid=38:cat_coding_algorithms_data-similarity&Itemid=5)
713 [ra-distance&catid=38:cat_coding_algorithms_data-similarity&Itemid=5](http://www.code10.info/index.php?option=com_content&view=article&id=49:article_camber_ra-distance&catid=38:cat_coding_algorithms_data-similarity&Itemid=5)
- 714 28. Uzma S, Zahid M, Toqeer M, Muhammad AJ, Amjad R, Tanzila S. Scene analysis and
715 search using local features and support vector machine for effective content-based
716 image retrieval. 2018; 1-25.

717 29. Muhammad Y, Zahid M, Hafiz AH, Toqeer M, Tanzila S, Amjad R, Muhammad R. A novel
718 technique based on visual words fusion analysis of sparse features for effective content-based image
719 retrieval Mathematical Problems in Engineering
720 2018.
721
722
723

UNDER PEER REVIEW