



# Content-based image retrieval for big visual data using image quality assessment model

Baisakhi Sur Phadikar<sup>1</sup> · Subro S. Thakur<sup>1</sup> · Goutam Kumar Maity<sup>2</sup> · Amit Phadikar<sup>3</sup>

Received: 25 November 2016 / Accepted: 26 November 2016 / Published online: 21 December 2016  
© CSI Publications 2016

**Abstract** Modern technology made storing, sharing, organizing huge amounts of data simple through the Internet of things. Search engines and query-based retrieval databases made access to relevant data easy through ranking and indexing based on content stored. This paper presents a content-based image retrieval technique using image quality assessment (IQA) model. The scheme combines two images IQA model namely mean-structural-similarity-index measure (MSSIM) and feature-similarity-index measure (FSIM) to take the relative advantage of each other. The MSSIM algorithm believes that human visual system can greatly reorganize the structural information from an image signal. On the other hand, color information is extracted using FSIM model. A combination of four image features i.e. luminance, contrast, structure, and color are used for similarity matching. Extensive experiments are carried out and assessment results reveal the outperforming result of the proposed technique with other related scheme.

**Keywords** Content-based image retrieval · Internet of things · Big visual data · Image quality assessment model

## 1 Introduction

With the seamless integration between devices and Internet in Internet of things (IoT), millions of images are accessed through social networking. The huge amount of data due to digital media, smart phone, and likewise result in ‘Big data’. Since these data are continuously uploaded, traditional “compute and storage” method does not able to handle the massive amount of data. In addition, most of the data are unstructured, however, although accompanied by explicit structured (e.g. geo-location and timestamps). This tendency leads large number of peoples to access large image database. A major part of these visual data becomes publicly available from blogs, social networking websites, image and video sharing websites, wikis, often guided by explicit structured and unstructured text annotations.

Furthermore, crowdsensing is the concept of a group of individuals with mobile devices collectively monitoring, sharing, and utilizing sensory information of their environment with common interests [1]. The huge amount of mobile users and the pervasive distribution of humans significantly expand the potential monitoring area to a global scale. However, since crowdsensing participators vary frequently, the sensory information is exhibited in various forms (images, audio, video, etc.), and the data accuracy is different. Extracting appropriate information from a huge amount of information are the main challenges here. Then how one can locate specific data such as image from a sea of images? This tough issue can be solved by content-based image retrieval (CBIR) scheme. CBIR is a demanding research area. Most popular implementation of

---

✉ Baisakhi Sur Phadikar  
baisakhi.sp@gmail.com

Subro S. Thakur  
subroto\_thakur@yahoo.com

Goutam Kumar Maity  
goutam123\_2005@yahoo.co.in

Amit Phadikar  
amitphadikar@rediffmail.com

<sup>1</sup> Department of Computer Science and Engineering, MCKV Institute of Engineering, Liluah, India

<sup>2</sup> Department of Electronics and Communication Engineering, Netaji Subhash Engineering College, Garia, India

<sup>3</sup> Department of Information Technology, MCKV Institute of Engineering, Liluah, India

this content-based image retrieval is Google image search with input as image instead of a query.

The requirement of CBIR has received extensive attention and huge numbers of solutions have been given in [2–8]. Malik and Baharudin [2] presented a technique that uses statistical texture information. This information is taken out from discrete cosine transform (DCT) blocks based on direct current (DC) and first three alternating current (AC) coefficients. For effective matching, various distance measurement techniques are used to calculate the similarities. Mohamed et al. [3] used texture and color information to extract related images with an average precision of 65%. Thawari and Janwe [4] explored color histogram, HSV and texture moments and achieved 53% precision. Duan et al. [5] presented a technique which combines shape, color and texture features for retrieval process. Singh et al. [6] proposed a technique that select the most suitable attribute (feature) to evaluate the newly received images to increase the retrieval efficiency and accuracy. For similarity assessment between query and database images, the scheme calculates the feature vectors after segmentation. To design approximate similarity retrieval Tang et al. [7] presented a neighborhood-discriminant-hashing (NDH) technique. The scheme is based on NDH by using local discriminative information. Li et al. [8] described an algorithm called subspace learning, for image retrieval. The scheme picks up an appropriate representation of data based on image feature learning and understanding by an integrated framework. The subspace learning is used to decrease the semantic differences between the low-level and high-level semantics.

Image quality estimation shows a significant role in the area of image processing. Image quality is the property of any image that is usually compared with an ideal or perfect image. The process like compression and acquisition etc may affect the quality of an image. Therefore, in many image-based applications precise evaluation of image fidelity is a significant step. The objective image fidelity evaluation is essential in multimedia applications since they eliminate or decrease the requirement for extensive subjective assessment. In the present paper, we use mean-structural-similarity-index measure (MSSIM) and feature-similarity-index measure (FSIM) as both of the techniques are based on HVS, are also widely used & cited in the literature and also have low computation cost [9].

In this paper, a content-based image retrieval technique is presented using full reference objective image quality metric. The major findings of this paper are described as follows:

1. Selection of proper IQA model that has low computational cost, easy to compute, and low execution time.
2. Combine the models that results better value of precision and recall for CBIR.

To the best of our information this feature of IQA has not yet been discussed by researcher for CBIR applications.

The organization of the article is as follows: Sect. 2 presents MSSIM for IQA and Sect. 3 discusses FSIM for IQA. Section 4 illustrates how MSSIM and FSIM are used jointly in CBIR for efficient feature selection. Section 5 presents the experimental results and discussion. Finally, conclusions and scope of future works are depicted in Sect. 6.

## 2 Preliminary on MSSIM for IQA

The MSSIM algorithm [10] believes that HVS can greatly reorganize the structural information from an image signal. Therefore, the system attempts to form the structural information of an image signal. The scheme splits the job of similarity calculation into three assessments i.e. luminance, contrast, and structure. Let,  $\mathbf{a} = \{a_j | j = 1, 2, \dots, M\}$  and  $\mathbf{b} = \{b_j | j = 1, 2, \dots, M\}$  be the un-distorted and distorted image signal, respectively. First, the luminance of each image signal (original and distorted) is evaluated. This is anticipated as the average intensity i.e.

$$\mu_a = \frac{1}{M} \sum_{j=1}^M a_j. \quad (1)$$

So, the luminance similarity measurement is a function of  $\mu_a$  and  $\mu_b$ . It is represented by  $L(\mathbf{a}, \mathbf{b})$ . Then the technique eliminates the average intensity from the image signal. The resultant image signal  $(\mathbf{a} - \mu_a)$  matches with the projection of vector ' $\mathbf{a}$ ' onto the hyper-plane. It is described as:

$$\sum_{j=1}^M a_j = 0. \quad (2)$$

The algorithm calculates signal contrast based on the standard deviation. The discrete form becomes:

$$\sigma_a = \left( \frac{1}{M-1} \sum_{j=1}^M (a_j - \mu_a)^2 \right)^{1/2}. \quad (3)$$

The contrast estimation is the similarity measurement of  $\sigma_a$  and  $\sigma_b$ . It is represented by  $C(\mathbf{a}, \mathbf{b})$ . Third, the image signal is divided with its own standard deviation. The structural similarity measurement  $S(\mathbf{a}, \mathbf{b})$  is calculated on these normalized image signals  $(\mathbf{a} - \mu_a)/\sigma_a$  and  $(\mathbf{b} - \mu_b)/\sigma_b$ . Finally, those three elements are pooled to represent an overall similarity assessment,

$$\mathbf{Sf}(\mathbf{a}, \mathbf{b}) = F(L(\mathbf{a}, \mathbf{b}), C(\mathbf{a}, \mathbf{b}), S(\mathbf{a}, \mathbf{b})). \quad (4)$$

A significant point is that the three elements are relatively independent. Luminance comparison is calculated as:

$$L(\mathbf{a}, \mathbf{b}) = \frac{2\mu_a\mu_b + k_1}{\mu_a^2 + \mu_b^2 + k_1}, \tag{5}$$

Constant ‘ $k_1$ ’ is incorporated to avoid instability in the above equation. This condition is occurred when  $\mu_a^2 + \mu_b^2$  is very close to zero. The contrast similarity is calculated as:

$$C(\mathbf{a}, \mathbf{b}) = \frac{2\sigma_a\sigma_b + k_2}{\sigma_a^2 + \sigma_b^2 + k_2}. \tag{6}$$

Structural similarity is conducted after luminance subtraction and variance normalization. Structural similarity is represented as:

$$S(\mathbf{a}, \mathbf{b}) = \frac{\sigma_{ab} + k_3}{\sigma_a\sigma_b + k_3}. \tag{7}$$

The component  $\sigma_{ab}$  is represented as:

$$\sigma_{ab} = \frac{1}{M-1} \sum_{j=1}^M (a_j - \mu_a)(b_j - \mu_b). \tag{8}$$

Finally, the above three relationship of (5), (6), and (7) are pooled to estimate SSIM index between image signals  $\mathbf{a}$  and  $\mathbf{b}$ .

$$SSIM(\mathbf{a}, \mathbf{b}) = [L(\mathbf{a}, \mathbf{b})]^{p1} \cdot [C(\mathbf{a}, \mathbf{b})]^{p2} [S(\mathbf{a}, \mathbf{b})]^{p3}, \tag{9}$$

where  $p1 > 0, p2 > 0, p3 > 0$ . Those factors are used to regulate the relative significance of the three components. In order to simplify the term, we set  $p1 = p2 = p3 = 1$  and  $k_3 = k_2/2$ . This results in an explicit form of SSIM as:

$$SSIM(\mathbf{a}, \mathbf{b}) = \frac{(2\mu_a\mu_b + k_1)(2\sigma_{ab} + k_2)}{(\mu_a^2 + \mu_b^2 + k_1)(\sigma_a^2 + \sigma_b^2 + k_2)}. \tag{10}$$

Finally, MSSIM is given as:

$$MSSIM(\mathbf{a}, \mathbf{b}) = \frac{1}{B'} \sum_{i=1}^{B'} SSIM(\mathbf{a}_i, \mathbf{b}_i), \tag{11}$$

The symbols  $\mathbf{a}_i$  and  $\mathbf{b}_i$  are the image information at the  $i$ th local patch, i.e. the block. The symbol  $B'$  is the number of samples in the image.

### 3 Preliminary on FSIM for IQA

The FSIM algorithm depends on the assumption that HVS identifies an image signal based on its low level image information. In this scheme, phase-congruency (PC) is incorporated as main information. On the other hand, gradient-magnitude (GM) is incorporated as secondary information. The calculation of FSIM index consists of two phases [11]. The local comparison map is calculated first. Then the algorithm groups the similarity map into a single similarity index.

Let, the original and the distorted image signal are represented by  $\mathbf{a}$  and  $\mathbf{b}$ , respectively. The PC maps calculated from  $\mathbf{a}$  and  $\mathbf{b}$  are  $PC_a$  and  $PC_b$ , respectively. Similarly, the symbols  $G_a$  and  $G_b$  are the GM maps calculated from them. The algorithm then divides the feature similarity calculation between  $\mathbf{a}(\mathbf{x})$  and  $\mathbf{b}(\mathbf{x})$  into two elements, i.e. for PC or GM. The similarity calculation (for a given location  $\mathbf{x}$ ) for  $PC_a(\mathbf{x})$  and  $PC_b(\mathbf{x})$  is represented as:

$$Sim_{PC}(\mathbf{x}) = \frac{2PC_a(\mathbf{x}) \cdot PC_b(\mathbf{x}) + c_1}{PC_a^2(\mathbf{x}) \cdot PC_b^2(\mathbf{x}) + c_1}, \tag{12}$$

where the symbol  $c_1$  is a positive constant. In the same way,  $G_a(\mathbf{x})$  and  $G_b(\mathbf{x})$  are evaluated. The similarity measurement is calculated as:

$$Sim_G(\mathbf{x}) = \frac{2G_a(\mathbf{x}) \cdot G_b(\mathbf{x}) + c_2}{G_a^2(\mathbf{x}) \cdot G_b^2(\mathbf{x}) + c_2}, \tag{13}$$

where the symbol  $c_2$  is a positive constant. Then,  $Sim_{PC}(\mathbf{x})$  and  $Sim_G(\mathbf{x})$  are pooled together to obtain the similarity  $Sim_L(\mathbf{x})$  as:

$$Sim_L(\mathbf{x}) = [Sim_{PC}(\mathbf{x})]^\alpha [Sim_G(\mathbf{x})]^\beta, \tag{14}$$

where the symbols  $\alpha$  and  $\beta$  are used to regulate the relative significance of PC and GM features, respectively. For a given location  $\mathbf{x}$ , if  $\mathbf{a}(\mathbf{x})$  or  $\mathbf{b}(\mathbf{x})$  has a major PC value, it represents that this location has higher impact on HVS. Therefore,  $PC_{max}(\mathbf{x})$  is calculated as  $PC_{max}(\mathbf{x}) = \max(PC_a(\mathbf{x}), PC_b(\mathbf{x}))$  to power the significance of  $Sim_L(\mathbf{x})$  in the overall comparison between  $\mathbf{a}$  and  $\mathbf{b}$ . Finally, the FSIM index is calculated as:

$$FSIM = \frac{\sum_{\mathbf{x} \in \Omega} Sim_L(\mathbf{x}) PC_{max}(\mathbf{x})}{\sum_{\mathbf{x} \in \Omega} PC_{max}(\mathbf{x})}, \tag{15}$$

where the symbol  $\Omega$  represents the spatial domain of image signal.

### 3.1 Color image quality assessment

The FSIM can be applied to color image also [11]. At first, RGB (red, green, blue) color images are transformed into YIQ color model. In YIQ color model,  $Y$  corresponds to the luminance feature and  $I$  and  $Q$  represent the chrominance feature. Let, the symbols  $I_a$  ( $I_b$ ) and  $Q_a$  ( $Q_b$ ) are the  $I$  and  $Q$  chromatic information of the images  $\mathbf{a}(\mathbf{b})$ , respectively. The similarities between chromatic features are defined as:

$$Sim_I(\mathbf{x}) = \frac{2I_a(\mathbf{x}) \cdot I_b(\mathbf{x}) + c_3}{I_a^2(\mathbf{x}) + I_b^2(\mathbf{x}) + c_3}, \tag{16}$$

$$Sim_Q(\mathbf{x}) = \frac{2Q_a(\mathbf{x}) \cdot Q_b(\mathbf{x}) + c_4}{Q_a^2(\mathbf{x}) + Q_b^2(\mathbf{x}) + c_4}, \tag{17}$$

where the symbols  $c_3$  and  $c_4$  are positive constants. Then

$Sim_I(\mathbf{x})$  and  $Sim_Q(\mathbf{x})$  is pooled to obtain the chrominance comparison measure. It is expressed by:

$$Sim_c(\mathbf{x}) = Sim_I(\mathbf{x}) \cdot Sim_Q(\mathbf{x}). \quad (18)$$

Finally, the FSIM is extended to  $FSIM_c$  by integrating the chromatic information as:

$$FSIM_c = \frac{\sum_{x \in \Omega} Sim_L(x) \cdot [Sim_c(x)]^\lambda \cdot PC_{max}(x)}{\sum_{x \in \Omega} PC_{max}(x)}, \quad (19)$$

where the symbol  $\lambda > 0$  is used to adjust the significance of the chromatic components.

#### 4 Image feature extraction and similarity measure for the proposed CBIR technique

It is already mentioned that MSSIM algorithm imagines the fact that HVS is highly recognize for extracting structural information of a scene. But, this scheme does not deal with the color information of the image. However, the color information is one of the most significant visual properties to express an image. In the present CBIR scheme, both the structural and color features are incorporated for image retrieval. The color feature comparison is measured by

$$S_{Color}(\mathbf{x}) = [Sim_I(\mathbf{x}) \cdot Sim_Q(\mathbf{x})]^\lambda. \quad (20)$$

In the present scheme, the value of  $\lambda = 0.03$ . Finally, the similarity of two images are measured as:

$$F_{Sim} = \alpha \times L(\mathbf{a}, \mathbf{b}) + \beta \times C(\mathbf{a}, \mathbf{b}) + \gamma \times S(\mathbf{a}, \mathbf{b}) + \delta \times S_{Color}, \quad (21)$$

where the symbols  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  are the positive load factors. The load factors mean the amount of significance of individual feature in CBIR. In other words, if all loads are evenly significant, all load values must be 1/4 (i.e.  $\alpha + \beta + \gamma + \delta = 1.0$  must hold).

The proposed technique consists of two modules namely, image features extraction and similarity matching of images. In the first stage, all the images are acquired, one after another from the image database and then feature extraction are done. After that feature vector is formed by the extracted image features which are stored in a database called feature database. From the 'Y' channel of the image, the luminance, contrast and structural information are extracted, while color information is extracted from 'I' and 'Q' components as described in Sects. 2 and 3, respectively. In second phase, user enters the search image to extract related images from the image database. The feature vector of the search image is then calculated and compared with the feature vectors of the database by calculating the equivalence as described in (21). According to the search image, the most related images are then displayed.

## 5 Results and discussion

In this section, we highlight large number of test to prove the effectiveness of the proposed technique for retrieval of images. We perform experiments on widely available benchmark image database, i.e., Corel dataset [12]. In the present scheme  $\alpha = 0.148$ ,  $\beta = 0.148$ ,  $\gamma = 0.038$ , and  $\delta = 0.666$ . The weight factors are obtained experimentally carried out over huge number of images with diverse characteristics(features) of image in the database. All the images in the database are in form of red–green–blue (RGB) color format as depicted in Fig. 1.

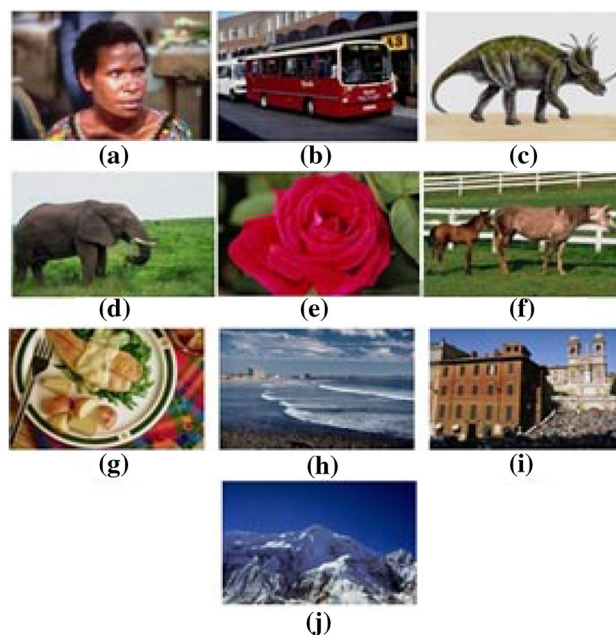
In image retrieval technique, the precession (Pr) value is defined as the amount of extracted appropriate images for the query image to the sum of extracted images. The symbol Pr is expressed by:

$$Pr = \frac{M}{N}, \quad (22)$$

where the symbol 'M' represents the related extracted images and the symbol 'N' is the sum of extracted images from the image database. The recall (Re) in CBIR is expressed as the amount of retrieved related images to the sum of images in the image database. The symbol Re is represented as:

$$Re = \frac{M}{O}, \quad (23)$$

where the symbol 'M' signifies the associated extracted images and the symbol 'O' is the sum of associated images in the image database. The 'Pr' and 'Re' values calculate



**Fig. 1** Benchmark images. **a** People, **b** buses, **c** dinosaurs, **d** elephants, **e** roses, **f** horses, **g** foods, **h** beaches, **i** buildings, **j** mountains

**Table 1** Average precision (Pr) value for the proposed technique

Image category	MSSIM	FSIM	Proposed
People	37	45	87
Bus	33	32	82
Dinosaurs	100	100	100
Elephants	28	73	100
Roses	91	91	100
Horses	21	28	100
Food	28	27	72
Beaches	55	64	91
Buildings	50	26	82
Mountains	32	50	78
Average	48	54	89

**Table 3** Average F-score (F) value for the proposed technique

Image category	MSSIM	FSIM	Proposed
People	38	46	88
Bus	30	29	79
Dinosaurs	100	100	100
Elephants	28	73	100
Roses	90	90	99
Horses	21	28	100
Food	32	31	77
Beaches	52	61	88
Buildings	51	27	83
Mountains	35	53	81
Average	48	54	90

**Table 2** Average recall (Re) factor for the proposed technique

Image category	MSSIM	FSIM	Proposed
People	40	48	90
Bus	28	27	77
Dinosaurs	100	100	100
Elephants	28	73	100
Roses	89	89	98
Horses	21	28	100
Food	38	37	82
Beaches	49	58	85
Buildings	52	28	84
Mountains	39	57	85
Average	48	55	90

the correctness and efficacy of image retrieval related to the database and query images. However, above two calculations cannot be considered individually as an absolute correctness for the efficient image retrieval. As a result, the above two measurements can be pooled to provide a single value that is called F-score, which is defined as:

$$F = 2 \times \frac{(Pr \times Re)}{(Pr + Re)} \tag{24}$$

In this paragraph, we explain the quantitative assessment for the benchmark images. Tables 1, 2 and 3 depict, the experimental outcome, in form of ‘Pr’, ‘Re’ and ‘F’ for all categories of images. From Table 1, 2 and 3, it is cleared that proposed technique offers attractive result using IQA. It is also cleared that when MSSIM and FSIM are used individually, offers poor results than the proposed scheme. The experimental results in Tables 1 and 2 are secured as



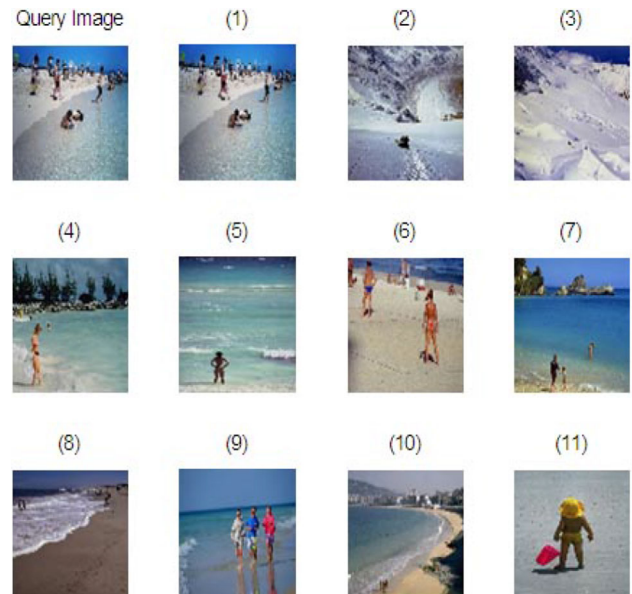
**Fig. 2** Experimental outcome of ‘Roses’ image

the mean value of hundred autonomous tests carried out on huge images in database. Figures 2, 3, 4, 5, 6 and 7 show the retrieval results for different images.

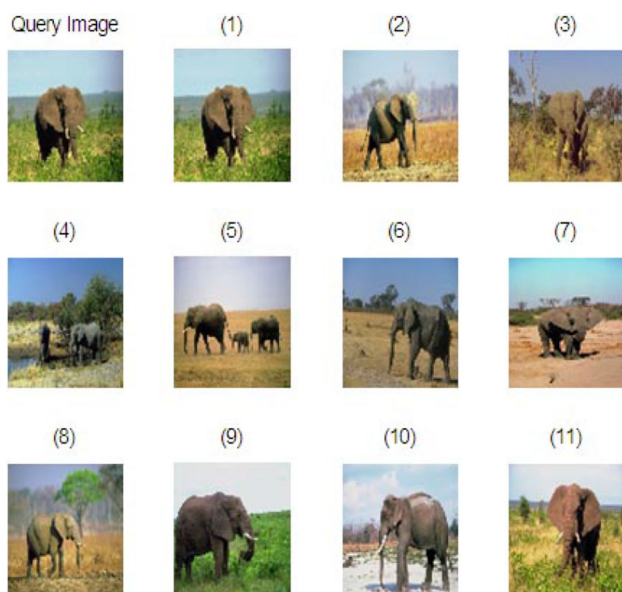
To express the comparative evaluation of the proposed technique for similarity measure, we evaluate it with the correlated techniques available in the literature. Table 4 depicts that the technique provides improved results compared to the related techniques. This is because, the present technique uses both structural and color information for the retrieval. All the above experiments are done in Pentium 4, 512 MB RAM, 2.80 GHz processor in MATLAB 7. The average execution times (in s) are 10.45, 116.99, and 14.14 for MSSIM, FSIM and proposed technique.



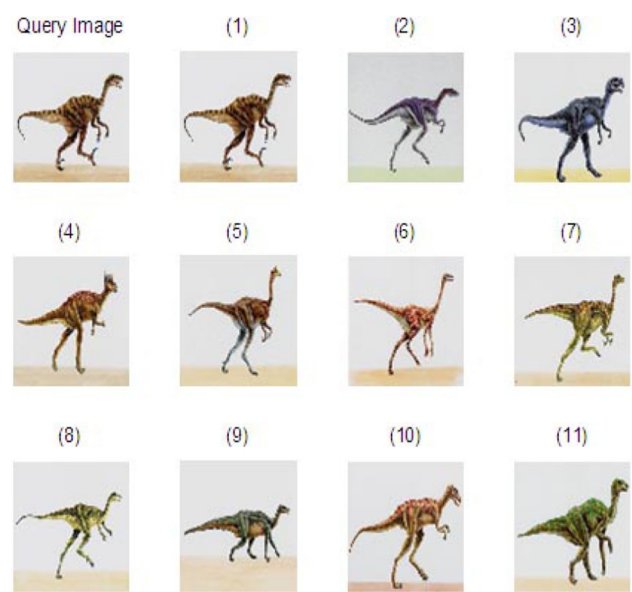
**Fig. 3** Experimental outcome of 'Horses' image



**Fig. 5** Experimental outcome of 'Beaches' image



**Fig. 4** Experimental outcome of 'Elephants' image

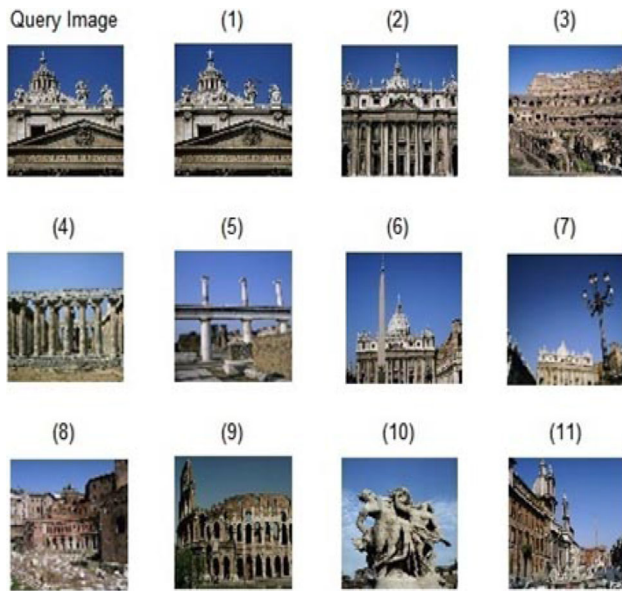


**Fig. 6** Experimental outcome of 'Dinosaurs' image

## 6 Conclusions and scope of future work

This paper, presents a CBIR scheme using IQA model. In this technique, a combination of four image features i.e. luminance, contrast, structure and color information are used for similarity measure. Performance assessment is

calculated on standard benchmarks like Pr, Re and F value. The proposed technique is only tested on image based query for similarity calculation. The future work includes the semantic annotation-based method to improve retrieval performance for text based query. Moreover, optimum selection of weight factors is another research issue.



**Fig. 7** Experimental outcome of ‘Buildings’ image

**Table 4** Comparative performance measure (Precision in %)

Images	A	B	C	D
People	76	88	50	87
Bus	52	86	50	82
Dinosaurs	100	100	90	100
Elephants	62	79	NA	100
Roses	80	96	NA	100
Horses	91	93	NA	100
Food	63	60	NA	72
Beaches	50	82	40	91
Buildings	47	77	35	82
Mountains	28	62	NA	78
Average	65	82	53	89

A Mohamed et al. [3], B Malik and Baharudin [2], C Thawari and Janwe [4], D Proposed, NA not applicable

**References**

1. Xu K, Qu Y, Yang K (2016) A tutorial on the internet of things: from a heterogeneous network integration perspective. *IEEE Netw* 30(2):102–108

2. Malik F, Baharudin B (2013) Analysis of distance metrics in content-based image retrieval using statistical quantized histogram texture features in the DCT domain. *J King Saud Univ Comput Inf Sci* 25:207–221

3. Mohamed A, Khellfi F, Weng Y, Jiang J, Ipson S (2009) An efficient image retrieval through DCT histogram quantization. In: *Proceedings of the international conference on Cyber Worlds, United Kingdom*, pp 237–240

4. Thawari PB, Janwe NJ (2011) CBIR based on color and texture. *Int J Inf Technol Knowl Manag* 4(1):129–132

5. Duan G, Yang J, Yang Y (2011) Content-based image retrieval research. In: *Proceedings of international conference on physics science and technology, Dubai*, pp 471–477

6. Singh N, Singh K, Sinha AK (2012) A novel approach for content based image retrieval. In: *Proceedings of 2nd international conference on computer, communication, control and information technology, Hooghly*, pp 245–250

7. Tang J, Li Z, Wang M, Zhao R (2015) Neighborhood discriminant hashing for large-scale image retrieval. *IEEE Trans Image Process* 24(9):2827–2840

8. Li Z, Liu J, Tang J, Lu H (2015) Robust structured subspace learning for data representation. *IEEE Trans Pattern Anal Mach Intell* 37(10):2085–2098

9. Xue W, Zhang L, Mou X, Bovik AC (2014) Gradient magnitude similarity deviation: a highly efficient perceptual image quality index. *IEEE Trans Image Process* 23(2):684–695

10. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 13(4):600–612

11. Zhang L, Zhang L, Mou X, Zhang D (2011) FSIM: a feature similarity index for image quality assessment. *IEEE Trans Image Process* 20(8):2378–2386

12. Wang JZ, Li J, Wiederhold G (2001) SIMPLcity: semantics-sensitive integrated matching for picture libraries. *IEEE Trans Pattern Anal Machine Intell* 23(9):947–963