IJASRM

International Journal of Advanced Scientific Research and Management, Volume 4 Issue 4, April 2019

www.ijasrm.com

ISSN 2455-6378

Color Image Retrieval Based on Adaptive Statistical Distance Measure with Local Features

Seetharaman. K[1], and Chembian. W. T.[2]

[1]Department of Computer and Information Sciences, Annamalai University, Annamalainagar 608 002, Tamilnadu, India.

[2]Department of Computer Science and Engineering, Jaya Engineering College, Chennai 600 052, Tamilnadu, India.

.

Abstract

In an automatic image retrieval system, it is difficult to understand the nature of an image data. If an automatic image retrieval system has prior knowledge about the image data, in the context of content-based image retrieval, it will give better results. The proposed method addresses this problem by applying an adaptive distance measure, Chernoff distance. The Chernoff distance adapts itself according to nature of the image data. First, the proposed method tests whether the image is a color or grayscale. If it is a color, which is converted to HSV color space; otherwise, it is treated as a grayscale. The query image is tested whether it is a structure or texture. If it is a texture, the whole image is subjected to experiment. If the image is a structure, which is segregated into various homogeneous regions, and features are extracted region-wise; and forms a feature vector (FV) region-wise. The FV of the query image is compared with the FVs in the feature vector database (FV^{db}) by the Chernoff distance. If the Chernoff distance is less than or equal to the critical value of the Chi-square table, then it is inferred that the query and target images are same or similar; otherwise, the images differ. The performance of the proposed method is evaluated, based on the precision and recall, and the ANMRR scores. The obtained precision and recall, and the ANMRR scores were compared with the state-of-the-art methods, which reveal that the proposed method yields better results.

Keyword*: Chernoff distance, target image, query image, feature-vector, Gaussian process.*

1. Introduction

The importance and awareness of the Internet and the World Wide Web have swiftly increased the number of Internet users, who access the data repository and multimedia data like movies, games, advertisements, etc. through online. Though, at present scenario, a number of techniques have been developed, they do not fulfill the users' requirements up to the satisfaction in terms of precise and speedy access to the data.

The literature reveals that a number of works has been developed, based on the contents of the images, that is, low-level visual features (color, shape, texture, and spatial orientation) and these features have been used for image retrieval [1–9]. Jing *et al*. [10] suggest that a single (global) signature computed on the whole (structure) image does not sufficiently represent the important features of each individual shape in an image, and there is a gap between the visual features and semantic concepts. To overcome this problem, a region-based method was introduced [9,10,14,15], which focuses users' perception about the contents of the image. The methods proposed in [3,11,12,13] segment the structure images into various homogeneous regions according to the shape, and compare the query and target images region-wise. Amanatiadis *et al*. [8] have studied about shape descriptors for shape-oriented image retrieval, and report that the region-based descriptors yield better results. Zhang *et al*. [14] have proposed a method, which identifies salient region and its surrounding regions, and they are combined together. The combined region is defined as an extended salient region (ESR). It extracts visual content from the ESR, using the Bag of Words model, based on Gabor, SIFT, and HSVH features and proposes a graph-based model for the visual content nodes, which represent the input image.

Song *et al*. [15] have proposed a method, based on diagonal texture structure descriptor (DTST), which converts an RGB color image to HSV color space and

segregates the background and foreground; and extracts multi-region features from the background and foreground. A histogram is constructed for the multi-region features and a one-dimensional vector is formed by concatenating the histograms. The terminology of global feature concept arises due to multi-region feature extraction; as suggested in [10], the features of the multi-regions do not overcome the drawbacks of the global features. In recent years, deep learning algorithms [16-18] have turned the attention of the research community of the computer vision, including image retrieval [19–21]. In [16], a deep convolutional learning neural network is introduced, which detects image descriptors. It suggests three kinds of retraining techniques: (i) unsupervised retraining, if no information is available, except dataset; (ii) retraining with relevant information, if the labels of the training datasets are available; (iii) relevance feedback-based retraining, if feedback is available from users. A recent survey on the convolution neural network (CNN) reports that the CNN demands high memory and computational time, and it requires skill and experience to select suitable hyperparameters, such as the learning rate, kernel sizes of convolutional filters, the number of layers etc. [22]. Since the CNN demands high memory and computational time, it is difficult to deploy a volume of data on the mobile platform (restriction on memory resources) and on the PC environment with network. Moreover, it could also consume more time to retrieve images through the Internet.

The distance measures play a significant role in matching and retrieving the images, especially, in content-based image retrieval (CBIR) system. The distance measures could be categorized into pairwise approach and distribution-based approach. The pairwise approach, first, extracts features, then matches them in a one-to-one manner. In this case, the users could not fix the number of images to be retrieved. To overcome this problem, in recent works, the extracted features are assumed to be a Gaussian random field or Markov Random field or Sigma field or Gibbs field, etc. and apply test statistics (distribution-based approach) to compare the similarity of the images. Moreover, in the context of CBIR, many statistical tests-of-hypothesis-based methods take the images directly without extracting features, i.e. avoid a separate feature extraction procedure. Since the distribution-based approach by itself computes statistical features like variance, standard deviation, covariance, mean vector, probability, etc., the computational complexity of the feature extraction is considerably reduced. Also, it facilitates the users to match and retrieve a specific number of images, by fixing the significance level of the test statistic accordingly the user's own convenience. This is difficult in the case of pairwise approach, because there is no such an option.

Johnson and Sinanovic [23] have analyzed the distance measures -- Kullback-Leibler divergence (KLD), Chernoff distance (CD), and Bhattacharyya distance (BD) -- with a viewpoint of statistical properties, and suggested that the CD measure performs better than the KLD and BD measures. Hence, an empirical study was conducted for image retrieval, based on CD measure, and the obtained results have been reported in [7]. The method proposed in this paper has been extended from [7] by incorporating a noise removal method, constructing a proper image database and its feature vector database. The databases were categorized using the *fuzzy c-means* algorithm. The performance of the proposed method was measured by precision and recall, and the ANMRR methods. The obtained results evidence the reports in [23].

1.1. Motivation

The segmented homogeneous regions of a structure image or the whole texture image assumed to be a Gaussian process [4,9,24]. In this paper, an attempt was made to test whether a texture image has been distributed to a Gaussian process or not.

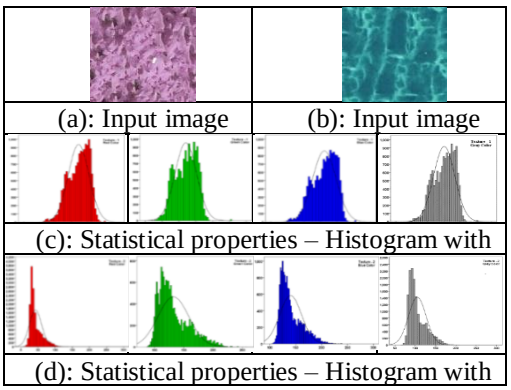

Fig. 1 Texture images and their distributional properties

The outcome of the test exposes some kind of texture images have not been distributed to Gaussian process because they do not follow the symmetric property of the Gaussian distribution, i.e. the histogram with normal curve presented in Figs. 1(c) and 1(d) are skewed to left and right tails respectively. The obtained results evidence this and the results have been presented in Figs. 1(a) and 1(b). At this juncture, in the case of fully automated image retrieval system, it is believed that the CD could lead to better retrieval results than the KLD and BD measures even if the whole texture image or regions of a structure image distributed to a Gaussian process or a non-Gaussian. Because the CD adapts itself according to the nature of the image data. Moreover, despite the ESR [14]

and DTSD [15] methods claim region-based feature extraction, actually they extract features globally. Because the DTSD method selects multi-region features from background and foreground, and the ESR method extracts features from the extended salient region (covers multiple of regions). The DCNN also extracts features globally. As a result, these motivated us to develop the proposed method.

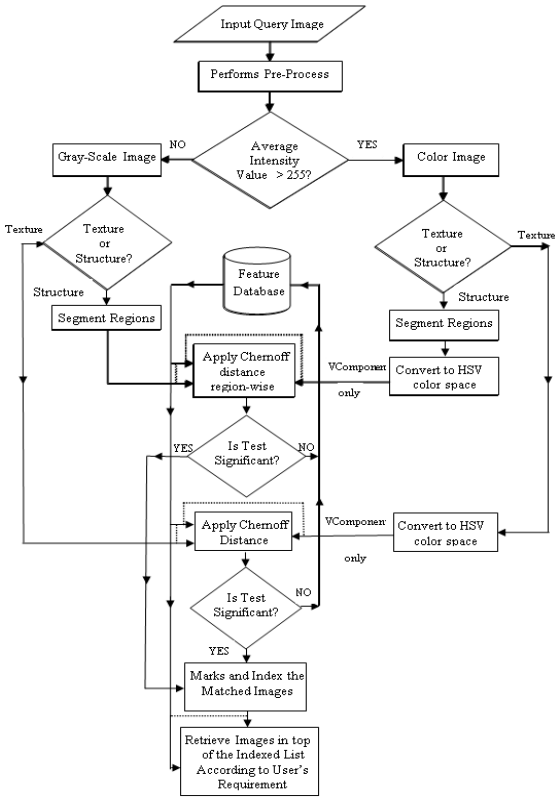

Fig. 2. Overview of the propose work

1.2. Outline of the proposed method

The proposed method, first, pre-processes the input query image, that is, removes noise, and examines whether it is a color or grayscale. If the query image is a color, then it is converted to HSV color space; otherwise, it is treated as a grayscale and subjected to experiment as it is. Further, the query image is examined whether it is texture or structure. If it is a structure, then it is segmented into various homogeneous regions according to the structure of the image; and the regions are counted. If the number of regions in the query and target images match, then the proposed method proceeds to the Chernoff test; otherwise, the matching process is dropped. One can easily identify whether the query and target images match or not, at the beginning of the retrieval process, by comparing the regions of the query and target images; by which the computation of the Chernoff test statistic can be avoided and the time complexity of the retrieval process could be reduced considerably. If the query image is a texture, the whole image is subjected to the experiment. If the CD value between the two images is less than or equal to the critical value of the Chi-square table, then it is inferred that the two images are same or similar; and they have been marked and indexed. The overall procedure of the proposed method is diagrammatically represented in Fig. 2.

The rest of the paper is organized as follows. Section 2 discusses the pre-processing method, and the proposed adaptive CD measure. Section 3 illustrates the image and its feature vector database construction while Section 4 deals with the performance measure. The experimental results are illustrated in Section 5. A discussion and concluding remarks are presented in Section 6 and Section 7 respectively.

2. Proposed image retrieval method

Many researchers, usually, assume the images as a Gaussian random field [4,9,24,25]. But, practically, this assumption is not applicable to all kinds of images. The Fig. 1 evidences this. In an automatic image retrieval system, it is difficult to identify the nature of the image data. In such cases, the KLD and BD measures are not appropriate because these measures strictly adhere to the Gaussian properties [25–28]. But, in such a context, the CD measure could be applied because it adapts itself, despite the images did not distribute to Gaussian process [25,27,28]. This is also one of the reasons why the proposed method is a fully automated.

2.1 Pre-processing

2.1.1 Image denoising

The presence of noise could diminish the accuracy of the retrieval rate. Thus, it is indispensable to remove the noise before extracting the features. In order to remove or eliminate the effects of the noise, the proposed method adopts a pre-processing technique used in [29], which improves the efficacy of the proposed retrieval method, even if the images were rigorously affected by different kinds of noises. The denoising technique is expressed in Eq. (1).

$$\|\mathbf{f} - \mathbf{q}\|^2 \le \mathbf{r}^2, \qquad (1)$$

where, $\|\,.\,\|$ is the Euclidean norm; q is the mean of the intensity value, which can be expressed by $q = \sum_{i=1}^{p+n} \beta_i \theta(f_i)$, β_i is a Lagrange multiplier; $\theta(f_i)$ is a vector of the pixel intensity values; **f** represents the intensity value of the pixel to be classified whether it is affected by noise or not. To classify the pixel intensity value **f**, the distance between the pixel and the mean intensity value is calculated. If this distance

IJASRM
International Journal of Advanced Scientific Research and Management, Volume 4 Issue 4, April 2019
www.ijasrm.com
ISSN 2455-6378

is less than or equal to radius, r, i.e. it satisfies the condition in Eq. (1), then **f** is accepted as a normal pixel; otherwise, it is identified as noise affected.

$$\left.\begin{aligned} r^2 &= \|f_u - q\|^2 = (f_u - q).(f_u - q) \\ &= (f_u).(f_u) + (q).(q) - 2(f_u).(q) \\ &= K(f_u, f_u) + K(q, q) - 2K(f_u, q) \end{aligned}\right\} \quad (2)$$

$$r^2 = K(f_u, f_u) + \sum_{i=1}^{p+n}\sum_{k=1}^{p+n}\beta_i\beta_k K(f_i, f_k) - 2\sum_{i=1}^{p+n}\beta_i(f_i, f_u) \quad (3)$$

where, $K(f_u, f_u)$ is the Mercer kernel function, which is denoted by $K(f_u, f_u) = (f_u).(f_u)$. The Eq. (3) can be derived by applying the conditions of the Lagrange-multiplier theorem and the Karush-Kuhn-Tucker complementarity [30].

The above denoising technique was applied to the images in the newly constructed image database, and the features were extracted. The extracted features were formed as a feature vector database (FV^{db}).

2.1.2 Image segmentation

Fig. 3. row 1: actual image; row 2: segmented shapes.

The shape in the structure images were segmented into various homogeneous regions, based on the intuitionistic *fuzzy c-means* technique [31], and features were extracted. The actual image and its segmented shapes have been presented in Fig. 3.

2.2 Basis of adaptive distance measure

Let F^{c1} and F^{c2} be two classes of image datasets. Let F be an image with intensity values f_{ci} ($c \in$ [red, green, blue]; i = 0, 1, 2, …, 255), which are non-negative random variables with absolutely continuous distribution function. Let $p(F|F^{c1})$ be the conditional probability of the image, F, belongs to the image dataset, F^{c1}; and the conditional probability of the image, F, belongs to the image dataset, F^{c2}, is $p(F|F^{c2})$. The probability of classification error to assign the image, F to either F^{c1} or F^{c2} can be bounded by the Chernoff bound as in Eq. (4).

$$P(\varepsilon) \le p_{c1}^{\lambda} p_{c2}^{1-\lambda} \int p^{\lambda}(F|F^{c1}) p^{1-\lambda}(F|F^{c2}) dy \quad (4)$$

where, ε is the error bound; p_{c1} and p_{c2} represent the *a priori* probability of the image dataset, F^{c1} and image dataset, F^{c2}; and $P_{c1}, P_{c2} \in [0,1]$. The derivation of the adaptive distance measure is discussed in the next section.

2.2.1 Proposed adaptive distance measure

Let $f_{(c)}^{(q)}$ and $f_{(c)}^{(t)}$ be intensity values of c-th color (c = 1, 2, 3: red, green, and blue) of a pixel at location (k, l) of the query image F^q, and the target image F^t distributed independent and identical to Gaussian process with mean vector $M^{(.)}$ and covariance matrix $\Sigma^{(.)}$, that is, $F^q \sim N(M^q, \Sigma^q)$ and $F^t \sim N(M^t, \Sigma^t)$ with *a priori* probability P_i^q and P_i^t (i: 0, 1, 2, …, 255), respectively. The P_i^q and P_i^t represent *a priori* probability of the i-th intensity value of the query and target images; $P_i^q, P_i^t \in [0,1]$. The integral part of the Eq. (4) can be written as in Eq. (5), from which a closed-form of the expression for the Chernoff upper bound, ε_u, can be derived as,

$$\int \left(\frac{f^q}{F^q}\right)^{\lambda} \left(\frac{f^t}{F^t}\right)^{1-\lambda} dx = e^{-C(\lambda)} \quad (5)$$

$$C(\lambda) = CD(F^q, F^t) = \frac{P_i^q P_i^t}{2}(M^q - M^t)^T \Sigma_{pooled}^{-1}(M^q - M^t) + \frac{1}{2}\left(\log\left|\Sigma_{pooled}^{-1}\right| - \log\left(\left|\Sigma^q\right|^{P_i^q} - \log\left|\Sigma^t\right|^{P_i^t} \right) \right). \quad (6)$$

In Eq. (5), the term C(λ) is called the Chernoff distance, which can be written as in Eq. (6). In this case, the optimum P_i^q can be derived by solving the expression in Eq. (7) using the mean vector $M^{(.)}$ and the covariance matrix $\Sigma^{(.)}$ of the query and target images. The $P_i^t = 1 - P_i^q$ and $0 \le P_i^q, P_i^t \le 1$.

$$\Sigma_{pooled} = P_i^q \Sigma^q + P_i^t \Sigma^t \; ; \; M^{(.)} = \left[m_r \;\; m_g \;\; m_b \right]^T \text{ and}$$

$$\Sigma = \begin{bmatrix} \sigma_{rr} & \rho\sigma_{rg} & \rho\sigma_{rb} \\ \rho\sigma_{gr} & \sigma_{gg} & \rho\sigma_{gb} \\ \rho\sigma_{br} & \rho\sigma_{bg} & \sigma_{bb} \end{bmatrix} = \begin{bmatrix} \sigma_r^2 & \rho\sigma_{rg} & \rho\sigma_{rb} \\ \rho\sigma_{gr} & \sigma_g^2 & \rho\sigma_{gb} \\ \rho\sigma_{br} & \rho\sigma_{bg} & \sigma_b^2 \end{bmatrix}$$

$$\frac{dCD(F^q, F^t)}{d(P_i^q, P_i^t)} = \frac{1 - 2P_i^q}{2}(M^q - M^t)^T \Sigma_{pooled}^{-1}(M^q - M^t) = 0 \quad (7)$$

N_i^q and N_i^t represent the number of pixels of the i-th intensity value of the query and target images respectively.

2.2.2 Significance and matching

If the query image, F^q and the target image, F^t are same or similar, then the terminology, *null hypothesis*, is used and denoted by H_o; if the images differ, which is termed as *alternative hypothesis* and denoted by H_a. The target image is retrieved, if the test formulated by a set of $FV_c \subset FV^{db}$, called acceptance region, that is, accept H_o if $F^q \in FV_c$ at the level of significance α. Otherwise, reject H_o, that is, accept H_a, means the query and target images differ. The level of significance means, the probability of the query and target images are to be same or similar at a particular (probability) point. To test the similarity of the F^q and F^t, a hypothesis is framed as follows.

Test of Hypothesis

$$H_o: \ CD(Fq, Ft) \le \chi^2_\alpha \ \text{(similar)}$$
$$H_a: \ CD(Fq, Ft) > \chi^2_\alpha \ \text{(dissimilar)}$$

The CD statistic is asymptotically distributed to Chi-square (χ^2) distribution, so that the computed CD value is compared to the critical value of the Chi-square table with degrees of freedom $(n_q + n_t - 1)$ at the level of significance ***α***; n_q and n_t are the number of pixels of the query and target images respectively. If the $CD \le \chi^2_\alpha$ at the level of significance ***α,*** then it is inferred that the query and target images are same or similar; otherwise, the two images differ. Based on the significant CD value, the images are marked and indexed in an ascending order; the indexed images are retrieved.

2.2.3 Advantages of the adaptive distance measure

The proposed method adapts itself accordingly the nature of the image because the *a priori* probabilities, P_i^q and P_i^t, are computed, using the expression in Eq. (7), based on the query and target images; and the *a priori* probability of P_i^q is not fixed arbitrarily to $1/2$ as in BD [25,26,32]. Thus, in the context of fully automated CBIR, the proposed method leads to better results than the KLD and BD measures for any kind of images.

3. Image and Feature Databases Settings

In order to implement the proposed method, an image database is constructed with 2,21,751 images. To maintain the heterogeneity of the image datasets, the images collected from benchmark datasets, and subjected to experiment. Of them, 3,573 images from Corel 10K image databases [33]; 4,288 images from VisTex image database; 5,656 images from CalTech image database [34]; 4,296 images from Holidays image database [35]; 2,508 images with size 128×128 were photographed by a digital camera; 2,517 images with size 128×128 downloaded from websites.

To validate the proposed method is robust for rotation, scaling and noise, the images were rotated through 90^o, 180^o, and 270^o, and scaled in different sizes. Moreover, to emphasize the effici`ency of the proposed method, 4,589 noised images have been considered for the experiment. In addition to that 1,524 multi-spectral satellite images have been incorporated in the image database. There are totally 2,21,751 [(((3573+4288) ×16) + (5656 + 4296 + 2508 + 2517)) × 3 (rotated through 90^o, 180^o, and 270^o) +44931 (scaled) + 4,589 (noisy) + 1,524 (satellite) = 2,21,751] images in the image database.

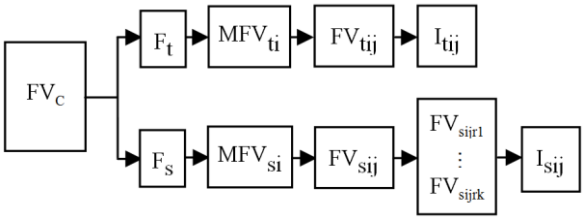

Fig. 4. Architecture of the feature vector and its images.

Based on these image collections, an image database and a feature vector database (FV^{db}) were constructed together. A link has been established between the FV and their corresponding images in the image database as depicted in Fig. 4. The FV_c denotes the classes of the FV^{db}; F_t and F_s represent the features of the texture and structure images; MFV_{ti} and MFV_{si} denote the median value of the i-th class of the FV^{db}; FV_{tij} represents the j-th FV of the i-th class of the texture images; FV_{sijr} represents the FV of the r-th region of the j-th FV (median of all the FVs of the regions) of the i-th class of the structure images; I_{tij} and I_{sij} represent the image, which corresponds to the j-th feature vector of the i-th class of the texture and structure features, respectively.

3.1 Feature database construction

The coefficient of variation (CV) was computed for each image in the image database using the expression in Eq. (8), and the CV was compared to a threshold ***t***. If the CV > ***t,*** then the image was assumed to be a structure (due to various features with different shapes, normally, the variation among the pixels of the structure image is larger); otherwise, it was assumed to be a texture. The threshold ***t*** was fixed at 25 percent, i.e. ***t*** = 0.25. The threshold ***t*** was fixed after conducting a rigorous empirical analysis. Based on this test, the structure and texture images

were labeled and maintained separately in the image database. This helps to search the required image only from either the structure part or the texture part instead of searching the entire database; by which the searching time could be reduced considerably.

$$CV = \frac{\sigma}{\bar{f}} \times 100 \quad (8)$$

where, σ and $\bar{f}$ are the standard deviation and the mean value of the pixel intensity values.

The structure images were segmented into various homogeneous regions. The regions are denoted by π_r and labeled with a numeric, i.e. r=1,2, …, m. The texture image considered as it is (not segmented). If the query image was a color, then it was modeled to HSV color space; otherwise, the image was treated as a grayscale and subjected to experiment as it is. The proposed retrieval method was employed on H, S, and V components individually for extracting features. On each region of the image, sample size N^t, mean vector M^t, covariance matrix Σ^t, and *a priori* probability P_i^t were computed; they were formed as a feature vector (FV^t), which belongs to FV^{db}, i.e., $FV^t \in FV^{db}$. The extracted features were clustered into various classes, which are homogeneous, using intuitionistic *fuzzy c-means* algorithm. A median value was computed for each class. After the FV^ts were indexed, based on the median, a link has been established between the images in the image database and their corresponding FV^t of each class. The extracted FV^q compared with the index (median value) of the FV^ts in the FV^{db}, and the same or similar images were identified based on the expression in Eq. (6). The identified images and their FV^q have been categorized into the corresponding class of the FV^{db}. If the FV^q does not match with any class of the FV^{db}, a new class was created and the FV^q has been categorized into the new FV^t class.

Algorithm

Input: Input query image, I^q, of size M×N.
Output: Retrieved Images.
Step 1: Uploads I^q.
Step 2: Performs pre-process with overlapping windows of size 3×3.
Step 3: Computes average intensity value for the whole image, I^q.
Step 4: *If* the average intensity > 255, I^q is color *then* continue next Step
else I^q is grayscale *goto* Step 6.
Step 5: Converts I^q to HSV color space.
Step 6: Computes CV, *If* CV > 0.25, I^q is structure *then goto* Step 10
else I^q is texture, continue next Step.
Step 7: Computes P^q, M^q, and Σ^q and forms FV^q
Step 8: Computes CD between FV^q and FV^t.
Step 9: *If* $CD \le \chi^2_\alpha$ *goto* Step 14
Step10: Segregates shapes into various homogeneous regions, and computes P^q, M^q, and Σ^q; forms FV^q region-wise.
Step 11: Compute CD region-wise.
Step 12: *If* $CD \le \chi^2_\alpha$ *then goto* Step 11 until all regions are tested.
Step 13: If all regions pass the CD test, then marks the image.
Step 14: Consider another FV^t from FV^{db} and *goto* Step 8.
Step 15: Retrieves the marked images.

4. Performance measure

To measure the performance of the proposed method, the Average Normalized Modified Retrieval Rank (ANMRR) measure was employed. Since it is a single measure of the performance, which considers both the number and order of the ground truth items that appear in the top retrievals [32]. The ANMRR and its derivatives have been expressed in Eqs. (9), (10), (11), and (12). Also, the performance of the proposed method was measured in terms of precision (P) and recall (R) [36], which are expressed in Eqs. (13), and compared with the state-of-the-art methods. The ANMRR value ranges zero to one; the lower values indicate the better retrieval rate.

$$ANMRR = \frac{1}{N_s}\sum_{q=1}^{N_q} NMRR(q) \quad (9)$$

$$NMRR(q) = \frac{AVR(q) - 0.5\left(1+N_q\right)}{1.25K_q - 0.5\left(1+N_q\right)} \quad (10)$$

The normalized modified retrieval ranking (NMRR(q)) score takes values between zero (whole ground truth found) and one (nothing found) irrespective of size of the ground-truth for query image, q, NG(q). The average rank (AVR) for a single query is computed as

$$AVR(q) = \frac{1}{N_s}\sum_{k=1}^{N_q} Rank(k) \quad (11)$$

$$Rank(q) = \begin{cases} Rank(k), & \text{if } Rank(k) \le K(q) \\ 1.25K(q), & \text{if } Rank(k) > k(q) \end{cases} \quad (12)$$

The *Rank*(*k*) of the *k*-th item is defined as a position at which it is retrieved. An item with higher rank was given a constant penalty, if a number $K_a \ge N_a$ was chosen. The K_q is generally chosen to be $2N_q$.

$$P = \frac{tp}{tp+fp}, \text{ and } R = \frac{tp}{tp+fn} \quad (13)$$

where, tp denotes the number of relevant images retrieved; fp denote the number of irrelevant images retrieved; fn represents the number of relevant images not retrieved.

IJASRM
International Journal of Advanced Scientific Research and Management, Volume 4 Issue 4, April 2019
www.ijasrm.com
ISSN 2455-6378

5. Experiments and results

In order to validate the proposed method, the image database constructed in Section 3 was considered for the experiment. The input query was pre-processed, based on the expression in Eq. (2), and tested to identify whether it is a texture or structure. The input query was segmented into various homogeneous regions as discussed in Section 2.1.2, if it is a structure. The proposed method, first, identifies the number of shapes π_q, and compares with that of the π_t of the $FV^t \in FV^{db}$. If it satisfies the criterion in Eq. (14), then the system proceeds the retrieval process; otherwise, it was dropped.

$$\left(\left(1-\frac{\pi_q}{\pi_t}\right)\times 100\right) \leq t_\alpha, \qquad (14)$$

where, t_α is a criterion to match the shapes in the images, and it was fixed to 20%, i.e. $t_\alpha = 0.2$. Either the query image or the target image, which has a smaller number of shapes placed in the numerator and the larger number placed in the denominator. The N^q, M^q, Σ^q, and P^q were computed on each region of the query image. If the query image is a texture, then the features were extracted from the whole image. The similarity of the query and target images was identified by employing the adaptive CD measure, which computes the distance between the FVs, N^q, M^q, Σ^q, P^q, of the query image and the FVs, N^t, M^t, Σ^t, P^t, of the target image.

To validate the proposed method, a number of structure images were subjected to experiment; owing to space constraint, for a sample, only the face images have been presented in Fig.5. The image in Fig. 5(a) was inputted as a query, for which, the proposed method retrieved the images in row 1 of the Fig. 5(b) at the level of significance, 0.05, i.e., $\alpha = 0.05$. At $\alpha = 0.10$, the images in rows 1 and 3 were retrieved. Similarly, all the images in Fig. 5(b) were retrieved while $\alpha = 0.18$.

Moreover, the CalTech and Holidays images were subjected to experiment; the images in column 1 of the Fig. 6 given as an input query to the proposed system, for which, the images in column 2 were retrieved at $\alpha = 0.05$; images in columns 3 and 4 were retrieved while fixing $\alpha = 0.12$; the images in columns 2–6 were retrieved at $\alpha = 0.20$.

Fig. 5(a): input query image of size 128 × 144. (b): retrieved target images.

Fig. 6. CalTech and Holidays images.

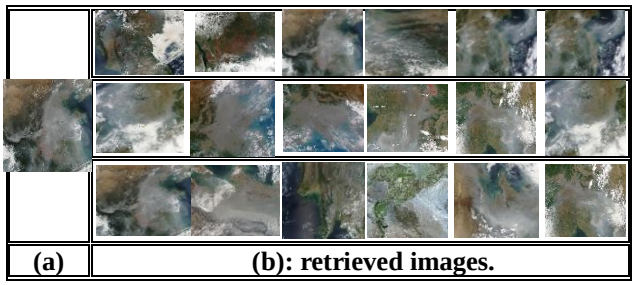

Fig. 7. Satellite images - Water-bank images.

An attempt was made to test the applicability of the proposed method for multi-spectral satellite images. For a sample, a few of them have been presented in Fig. 7. The proposed method retrieved the images in column 3 of the Fig. 7(b) at $\alpha = 0.10$; the images in columns 3 and 4 were retrieved at $\alpha = 0.15$; images in columns 3–6 were retrieved while fixing $\alpha = 0.20$; all the images in Fig. 7(b) were retrieved while $\alpha = 0.25$. Though, the proposed method was not developed exclusively for satellite imagery retrieval, it yields good results. The obtained results have been presented in Table 1. The line graphs were drawn for the precision and recall values obtained at various significant levels, and they have been presented in Figs. 8, and 11. A Bar chart also drawn for the precision values, and it has been shown in Fig. 9.

5.1 Comparative study

In order to evaluate and validate the performance of the proposed method, the ANMRR, precision (P) and recall (R) scores were computed, and compared with the state-of-the-art methods: DCNN [16], ESR [14], DTSD [15], KLD [37], and BD [32]. The ANMRR scores obtained were represented with a Bar chart, and has been presented in Fig. 10. The output results reveal that the proposed method yields better results

than the state-of-the-art methods. The average precision and recall, and ANMRR score obtained have been presented in Table 1.

5.2 Computational time complexity

The proposed and the state-of-the-art methods were implemented through Java SE 7 compiler with the system specification: Intel Core i5-4440 processor-based PC with 4GB DDR3 RAM. The time consumed for feature extraction, matching, and retrieving was measured in seconds, and the obtained results have been presented in Table 2. The output results reveal that the proposed method consumes lesser time than the state-of-the-art methods.

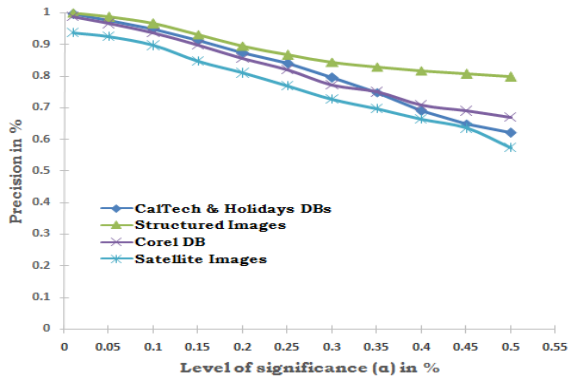

Fig. 8. Precision versus significance level (α) of the test statistic.

The DCNN uses a number of layers, such as five types of convolutional layers, sub-sampling layers with non-linear neural network activation, and a fully connected layers with appropriate activation functions. Because of a lot of computational procedures involved, it demands high computational time. The ESR method detects salient region and its surrounding regions, and combines them, then derives the extended salient region. Also, converts the rectangular coordinates to polar coordinates, and uses Bag-of-words model and a graph-based model. Finally, it performs matching and retrieving processes. As a result, it requires high computational time than the proposed method, but at the same time it is superior than the DCNN and DTSD methods. The DTSD method segments the images into foreground and background using Otsu algorithm, and the features of the multi-regions extracted; H and S components are quantized into histogram bins, which obtains a detailed description of the color differences. Finally, the DTSD is extracted, based on V component to represent the edge information as a feature. This leads to a lengthy computational procedure.

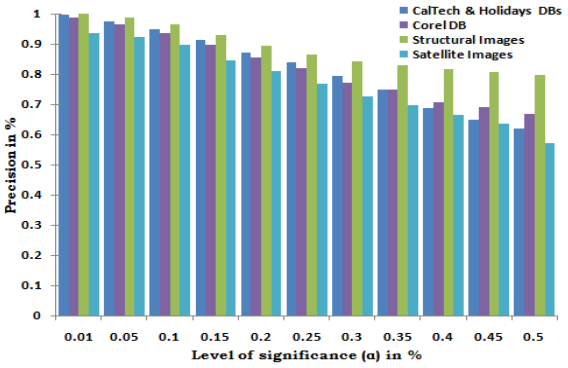

Fig. 9. Precision versus significance level (α) of the test statistic.

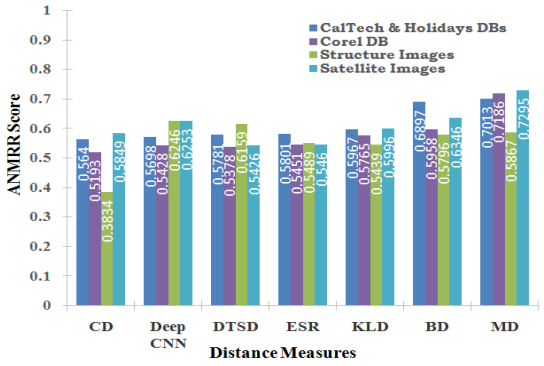

Fig. 10. Comparison of various methods.

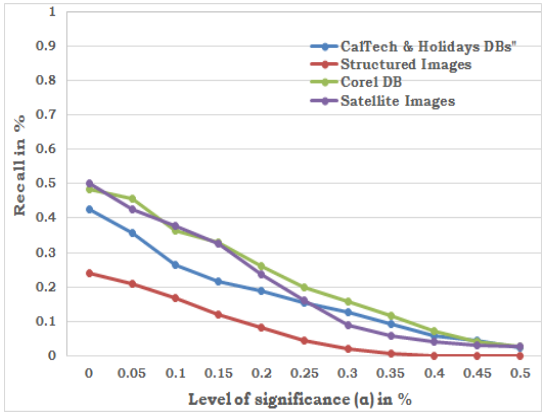

Fig. 11. Recall versus significance level (α) of the test statistic.

ISSN 2455-6378

Table 1. Performance of the proposed method with other existing methods - average precision (AP), average recall (AR), and ANMRR Score.

Image Database	Proposed Method			DCNN			DTSD Method			ESR Method			KLD Method			BD Method		
	AP	AR	ANMRR	AP	AR	ANMRR	AP	AR	ANMRR	AP	AR	ANMRR	AP	AR	ANMRR	AP	AR	ANMRR
CalTech & Holidays	0.8223	0.1779	0.5486	0.7925	0.2269	0.5698	0.7285	0.3018	0.5781	0.7051	0.3195	0.5801	0.7293	0.3052	0.5967	0.7037	0.3256	0.6897
Corel	0.8227	0.2078	0.5193	0.8116	0.2185	0.5428	0.8125	0.2197	0.5378	0.7924	0.2259	0.5489	0.8098	0.2207	0.5765	0.7736	0.3014	0.5958
Structure Images	0.8850	0.0813	0.3834	0.7654	0.2251	0.6246	0.7557	0.2914	0.6159	0.8053	0.2206	0.5451	0.7413	0.2895	0.5839	0.8018	0.2218	0.5796
Satellite Images	0.7805	0.2284	0.5749	0.7764	0.2206	0.6253	0.7598	0.2895	0.5426	0.7954	0.2289	0.5460	0.7359	0.3052	0.5996	0.7021	0.3317	0.6346

Hence, it demands more time than the proposed method, whereas it requires lesser time than the DCNN method. Though, the KLD and BD methods require low computational time, they yield low retrieval rate. The obtained results show that the proposed method consumes lesser time than the existing methods.

Table 2. Comparison of time (second) taken by different methods for feature extraction and, matching and retrieving.

Time taken	Proposed method	DCNN	DTSD	ESR	KLD	BD
Feature	0.683	0.978	0.846	0.802	0.735	0.698
Matching and retrieval	0.045	0.086	0.071	0.061	0.069	0.054
Total	0.728	1.064	0.917	0.863	0.804	0.752

6. Discussion

The proposed method was compared with the state-of-the-art methods in terms of precision, recall, ANMRR score, and computational time complexity. The proposed method yields better results for structure images, and marginally high results for CalTech and Holidays, and Corel image datasets compared to the state-of-the-art methods. Moreover, it yields moderate results for satellite images with less computational time. Because some of the CalTech and Holidays, Corel images, and satellite images have attributed to more texture or semi-structure properties than the structure. So that the proposed method faces a bit of difficulty in segregating the regions. This is the reasoning, the proposed method yields moderate results for CalTech and Holidays, Corel images, and satellite images. The reason behind the proposed method outperforms the state-of-the-art methods, in the case of structure images, it extracts features region-wise and matches region-wise, but the state-of-the-art methods extract features globally and retrieve. This may be the reason the state-of-the-art methods yield moderate results for CalTech and Holidays, Corel images, and satellite images. Generally, the features extracted globally do not give better results for structure images in the context of the automated CBIR. The proposed method retrieves exactly the same and very similar images while fixing α at or less than 0.15, at the same time it misses to retrieve the relevant (not similar) images; but it returns same and all the relevant images while fixing α at or greater than 0.2 (i.e. increasing critical region).

As discussed earlier, in an automatic image retrieval system, it is difficult to identify the nature of the image data. The proposed method overcomes this problem, because the Chernoff distance adapts itself according to the nature of the images, but the KLD and BD measures do not adapt themselves. The KLD and BD, also, have a constraint, such that the images should have been distributed to the Gaussian process, whereas there is no such a constraint in the case of the Chernoff distance. Hence, in the case of an automatic image retrieval system, the CD measure leads to better results than the KLD and BD measures. Moreover, the BD is a special case of the Chernoff distance [26]. Therefore, the proposed method is more appropriate to an automatic image retrieval system when compared to the state-of-the-art methods.

7. Conclusion and future extension

The proposed method converted the query image from RGB to HSV color space, if it is a color; otherwise, the image was treated as a grayscale. The structure query image was segmented into various homogeneous regions, and features extracted from each region, if it is a structure; otherwise, the image was treated as a texture, and the whole image subjected to the experiment as it is. The extracted features were formed as a FV and classified into different classes, based on the median value of the FVs of each class. Each class contains similar FVs, whereas the median (index of the class) value of each class differs from each other. The features extracted from each region of the structure image were formed as a FV and labeled with a numeric; this labeling method considerably reduces the searching and matching time. The obtained results were compared to the state-of-the-art methods, which reveal that the proposed method yields better results.

The proposed method can be extended to pattern recognition and matching, especially, face recognition, iris recognition, and video retrieval, because it adapts itself accordingly the nature of the images. This method is very suitable for video retrieval and classification, because most of the video frames contain structure objects. An Eigen space can be formed by deriving eigenvectors from each object (region) of the frame. Based on the eigenvectors, the video frames could be matched and retrieved.

Acknowledgement

The first author thanks the University Grants Commission, India, which extended financial support for this research work under the MRP scheme, File No. 42-145/2013(SR).

References

[1] Huang, J., Kunar, S.R., Mitra, M., Zhu, W.J., Zabih, R., Image indexing using colour correlogram. In: Proc. IEEE Comp. Soc. Conf. Comp. Vis. and Patt. Recog. 1997, 762-768, (1997).

[2] Stricker, M., Orengo, M., Similarity of colour images. In: Storage and Retrieval for Image and Video Databases, Proc. (SPIE 2420), pp. 381-392, (1995).

[3] Fuh, C-S., Cho, S-W., Essig, K., Hierarchical colour image region segmentation for content-based image retrieval system. IEEE Transactions on Image Processing vol, 9(1), 156-162, (2000).

[4] Seetharaman, K., Jaikarthic, M., Statistical Distributional Approach for Scale and Rotation Invariant Colour Image Retrieval Using Multivariate Parametric tests and Orthogonality Condition, Journal of Visual Communication and Image Representation, vol. 25(5), 727-739, (2014).

[5] Hsieh, J-W., Eric, W., Grimson, L., Spatial template extraction for image retrieval by region matching. IEEE Transactions on Image Processing, 12(11), 1404-1415, (2003).

[6] Belongie, S., Carson, C., Greenspan, H., Malik, J., Recognition of images in large databases using colour and texture, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(8), 1026-1038, (2002).

[7] Seetharaman, K., Chembian, W.T., Statistical Distribution-Based Color Image Retrival, International Conference on Graphics and Signal Processing (ICGSP2017), Nanyang Technological University, Singapore, from 24.06.2017 to 28.06.2017, ACM Digital Library, pp. 6–9, (2017).

[8] Amanatiadis, A., Kaburlasos, V.G., Gasteratos,A., Papadakis, S.E., Evaluation of shape descriptors for shape-based image retrieval, IET Image Processing, 5(5), 493–499, (2011).

[9] Seetharaman, K., Image retrieval based on micro-level spatial structure features and content analysis using Full Range Gaussian Markov Random Field model, Engineering Applications of Artificial Intelligence, 40, 103-116, (2015).

[10] Jing, F., Li, M., Zhang, H.J., Zhang, B., An efficient and effective region-based image retrieval framework. IEEE Transactions on Image Processing, 13(5), 699-709, (2004).

[11] Jing, F., Zhang, B., Lin, F. Z., Ma, Y. W., Zhang, H. J., A novel region-based image retrieval method using relevance feedback. In: Proc. 3rd ACM Int. Workshop on Multimedia Information Retrieval (2001).

[12] Hsieh, I-S., Fan, H. C., Multiple classifiers for colour flag and trademark image retrieval," IEEE Transactions on Image Processing, 10(6), 938-950, (2001).

[13] Salah, M.B., Mitiche, A., and Ayed, I.B., Multiregion Image Segmentation by Parametric Kernel Graph Cuts, IEEE Transactions On Image Processing, 20(2), 545-557, (2011).

[14] Zhang, J., Feng, S., Li, D., Gao, Y., Chen, Z., Yuan, Y., Image retrieval using the extended salient region, Information Sciences, 399, 154–182, (2017).

[15] Song, W., Zhang, Y., Liu, F., Chai, Z., Ding, F., Qian, X., Park, S.C., Taking advantage of multi-regions-based diagonal texture structure descriptor for image retrieval. Expert Systems With Applications, 96, 347–357, (2018).

[16] Tzelepi, M., Tefas, A., Deep convolutional learning for Content Based Image Retrieval, Neurocomputing, 275, 2467-2478, (2018).

[17] Deng, L., A tutorial survey of architectures, algorithms, and applications for deep learning, APSIPA Trans. Signal Inf. Process. 3, e2, pp. 1-29, (2014).

[18] Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, V., Alsaadi, F.E., A survey of deep neural network architectures and their applications, Neurocomputing, 234, 11–26, (2017) .

[19] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, M.S. Lew, Deep learning for visual understanding: a review, Neurocomputing, 187, 27–48, (2016).

[20] G. Hinton, L. Deng, D. Yu, G.E. Dahl, AR. Mohamed, N. Jaitly, A. Senior, V. Van Houcke, P. Nguyen, T.N. Sainath, Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups, IEEE Signal Process. Mag. 29(6), 82–97.

[21] J. Ngiam, A. Khosla , M. Kim, J. Nam, H. Lee , A.Y. Ng, Multimodal deep learning, in: Proceedings of the 28th International Conference on Machine Learning (ICML), pp. 689–696, (2011).

[22] Gua, J., Wang, Z., Kuenb, J., Ma, L., Shahroudy, A., Shuai, B., Liub, T., Wang, X., Wang, G., Cai, J., Chenc, T., Recent advances in convolutional neural networks, Pattern Recognition, 77, 354–377, (2018).

[23] Johnson, D.H., and Sinanovic, S., Symmetrizing the Kullback-Leibler Distance, A technical report, Rice University, Houston, pp.1-10, (2001).

ISSN 2455-6378

[24] Kato, Z., Pong, T-C., A Markov random field image segmentation model for colour texture images, Image and Vision Computing, 24, 1103–1114, (2006).
[25] Moustafa, A. K. T., Ferrie, F. P., A Note on Metric Properties for Some Divergence Measures: The Gaussian Case, Asian Conference on Machine Learning, Workshop and Conference Proceedings 25:1-15, (2012).
[26] Fukunaga, K., Introduction to Statistical Pattern Recognition, Academic Press. Inc., New York, 2nd Ed. pp. 97-102, (1990).
[27] Cover, T.M., and Thomas, J.A., Elements of Information Theory, Wiley, New York, (1991).
[28] Rueda, L., Herrera, M., Linear Dimensionality Reduction by Maximizing the Chernoff Distance in the Transformed Space, Pattern Recognition, 41(10), 3138-3152, (2008).
[29] Liu, B., Xiao, Y., Yu, P. S., Hao, Z., and Cao, L., An Efficient Approach for Outlier Detection with Imperfect Data Labels, IEEE Transactions on Knowledge and Data Engineering, 26(7), 1602-1616, (2014).
[30] V. Vapnik, Statistical Learning Theory. London, U.K.: Springer-Verlag, 1998.
[31] Huang, C-W., Lin, K-P., Wu, M-C., Hung, K-C., Liu, G-S., Jen, G-H, Intuitionistic fuzzy c-means clustering algorithm with neighborhood attraction in segmenting medical image, Soft Computing, 19(2), 459–470, (2015).
[32] Yang, Y., and Newsam, S., Geographic Image Retrieval Using Local Invariant Features, IEEE Transaction on Geoscience and remote Sensing, 51(2), 878 – 832, (2013).
[33] www.corel.com.
[34] http://vision.caltech.edu.
[35] http://lear.inrialpes.fr/~jegou/data.php
[36] Powers, D.M.W., Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation, Journal of Machine Learning Technologies, 2(1), 37-63, (2011).
[37] Sumana, I.J., Lu, G., and Zhang, D., Comparison of Curvelet and Wavelet Texture Features for Content Based Image Retrieval, IEEE International Conference on Multimedia and Expo, pp. 290-295, (2012). DOI 10.1109/ICME.2012.90.