

An Intelligent Multi-resolution and Co-occurring local pattern generator for Image Retrieval

Shikha Bhardwaj^{1,2,*}, Gitanjali Pandove¹ and Pawan Kumar Dahiya¹

¹Department of Electronics and Communication, DCRUST, Murthal, Sonapat, India

²UIET, Kurukshetra University, Kurukshetra, India

Abstract

Content-based image retrieval (CBIR) is a methodology used to search indistinguishable images across any vast repository. Texture, Color and Shape are among the most prominent features of any CBIR system. Two texture descriptors namely Gray level Co-occurrence matrix (GLCM) and Discrete wavelet transform (DWT) have been utilized here for the formation of a hybrid texture descriptor, denoted as (Co-DGLCM). To enhance the retrieval accuracy of the proposed system, a framework of an Extreme learning machine (ELM) with Relevance feedback (RF) has also been used. This technique provides simultaneously spatial relationship and information related to frequency in co-occurring local patterns of an image. Two benchmark texture databases namely Brodatz and MIT-Vistex have been tested and results are obtained in terms of accuracy, total average recall and total average precision which is 96.35% and 97.34% respectively on the two databases.

Keywords: Gray level co-occurrence matrix, Discrete wavelet transform, Content-based Image retrieval, Extreme learning machine, Relevance feedback, Brodatz dataset, MIT-Vistex Dataset

Received on 23 February 2019, accepted on 25 May 2019, published on 27 June 2019

Copyright © 2019 Shikha Bhardwaj *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.10-6-2019.159344

1. Introduction

This paper highlights an efficient hybrid texture descriptor namely combined DWT-GLCM (Co-DGLCM) which is an attractive combination of Gray level co-occurrence matrix (GLCM) and Discrete wavelet transform (DWT). To improve the classification accuracy of the proposed system, an Extreme learning machine (ELM) has been used. Finally, based on the user's feedback, the concept of Relevance feedback (RF) has also been used to refine the inputs of an ELM working as a classifier.

1.1. Motivation

The evolution of the cyberspace technology and various digital repository techniques led to a wide enlargement in commercial and research applications in the field of digital images. Additionally, due to the enormous increase in

multimedia content such as text, images, audio, video, etc., the storage requirement is a great issue of concern. Therefore, many multimedia compression techniques like JPEG, MP3, H.263, TIFF, etc., have been developed to solve this issue. Diverse versions of the compressed encoded streams, consisting of H.264 and H.265 [1] can be used for Bandwidth conservation. Therefore, as the multimedia content is increasing, there is a cumulative aggravation for techniques related to data storage and retrieval. To remove the bottleneck related to image retrieval, there is a need for an accurate and fast recovery method to precisely retrieve the data from these vast multimedia storehouses. The content of these databases can be images or text [2].

Based on the database content, the recovery of images can be categorized into two main types: (i) Content-based (ii) Text based. Recovery of images based upon text is the conventional method of retrieving images as it suffers from many pitfalls and the prominent among them are; it is a clumsy process to retrieve the images, though images have to be elucidated with textual descriptors and to describe

*Corresponding author. Email:shikpank@yahoo.com

multimedia content with text, it becomes too difficult. For example, complex images with perplex appearance becomes strenuous to be described with textual information. It also suffers from vast spelling errors based on homonyms etc. Therefore, an image recovery system denoted as Content-based image recovery or retrieval (CBIR) based on perceptible basic attributes of a given image is key to this problem.

This work is concentrated on specifically texture feature extraction techniques and the formation of an intelligent textual hybrid system. Classification of images is also one of the most peculiar tasks and many types of classifiers can be deployed for this purpose. Here, the classifier based on neural network has been implemented, whose results are incomparable to other customary classifiers like support vector machine (SVM), Bayesian classifier, etc. Finally, to update the training images of ELM classifier, the technique of Relevance feedback has also been implemented. The diverse amount of distance metrics can also be employed to find the resemblance between an input image and total repository images.

1.2. Main Contributions

An effective technique which integrates Gray level co-occurrence matrix (GLCM) and Discrete wavelet transform (DWT) and designated as combined DWT-GLCM (Co-DGLCM) has been implemented in this paper to extract texture features. Finally, to make the proposed system intelligent and accurate, ELM classifier has been used. Lastly, if the obtained results are less than the desired threshold then, the technique of Relevance Feedback has been applied which refines the input of the classifier based on the user's feedback and optimized results are obtained. The proposed technique has been analysed on standard benchmark texture datasets namely Brodatz and MIT Vistex and overall precision, recall and accuracy have been calculated. The framework for the presented work is described in the following pattern: Introduction based on the requirement for an optimum storage and retrieval technique has been discussed. Also, the motivation behind the concerned topic is given in section 1 and typical related work is dispensed in Section 2. Feature extraction techniques are described in Section 3. In Section 4 architectural framework with distance metric for similarity calculation are given. Experimental analysis and their outcomes are presented in Section 5. Lastly, Conclusion and future trends are described in Section 6.

2. Related work

Images are considered as an integral part of multimedia. Since many decades, recovery of images happens to be a very potent experimentation area. Texture describes the discernable patterns in an image and their spatial positioning. Texture can be extracted from the whole image or from its specific part. Anu Bala et al. proposed a system based on

Local Texton XOR patterns (LTxXORP). Here, texton patterns are used for shape analysis. Finally, XOR operation is performed on different pixels consisting of a middle pixel and its adjacent neighbors [3]. LTxXORP is utilized here for shape analysis and moreover this technique does not include any intelligence based machine learning model for reducing the semantic gap.

Correlation between intensity values of a center pixel and a specific pixel is given by Prithaj et al. [4] using Local neighborhood Intensity Patterns (LNIP). For the purpose of comparison, the used technique is compared to distinct techniques like Local Binary Pattern (LBP) [5-6], Local derivative pattern (LDP) including Local ternary pattern (LTP). Also, the texture images have been tested and analysed by using Local tri-directional weber patterns (LTriWPs) [7]. LBP produces very long histograms of images and slows down the process of recognition. The outcomes of LTP are variant with gray scale transformation of intensity values. Therefore, second order statistical texture extraction technique, GLCM is utilized in the proposed work.

Gabor Wavelets have also been used for the extraction of features related to a facial dataset. Later, Fisher's discriminant based indexing is being utilized to reduce the dimensions of the resultant feature vectors [8]. Ekta Walia et al. [9] described the results of some local texture extraction methodologies on Log Gabor filter response. Log Gabor filter is stated to be better in performance as compared to Gabor filter. Gabor wavelet has large dimension of feature vectors, therefore another wavelet domain technique, DWT is utilized here, as it has less memory requirement and small feature vector size.

Wei Zhang et al. [10] proposed a technique for texture representation denoted by Normalized difference vector (NDV). Bag of Visual Words is also being utilized in this paper. Two improved versions of Local Binary descriptor (LBP) with wavelet transform have been used for texture classification [11]. The employed techniques are specified by wavelet domain local statistical binary pattern (WLSBP) and directional wavelet domain local statistical binary pattern (dWLSBP). Again, LBP suffers from some drawbacks and an efficient retrieval system cannot be designed.

A texture feature extraction technique specified by Local quantized edge binary pattern (LQEBP) is implemented in HSV color space. These quantized patterns are used for the extraction of local information [12]. The concept of extraction of spectral textural features has been reported by Joydeb kumar et al. [13]. Here power law transform (PLT) has been employed for the same.

An image retrieval system based on Rotated wavelet transform (RWT), Discrete wavelet transform (DWT), Dual-tree complex wavelet transform (DT-CWT) and finally Dual tree rotated complex wavelet filter (DT-RCWF) and their different combinations have been proposed by Das et al. [14]. These structures have been developed by using DWT by changing some of its attributes but it is only the combination of DWT and GLCM in the proposed technique which produces an effective hybrid texture descriptor.

In order to fetch textural data for browsing, a system based on Gabor wavelet have been described by Manjunath et al.

[15]. For the withdrawal of texture attributes, Hartley domain and another directional transform based on [16] Hartley domain (DirHT) is also specified. Hartley transforms based on Wrapping (WDirHT) and similarly, Hartley transform based on Overlapping (ODirHT) are the two implemented techniques here.

The application of machine learning in image retrieval has increased significantly in recent years. Image classification has evolved into a famous research area. Varied machine learning based classifiers have been utilized for image retrieval. But, Extreme learning machine (ELM), which is based on neural networks, can be considered as a type of Deep learning network. Though, it utilizes a single hidden layer and is a feed-forward neural network but has unexcelled results as compared to other machine learning based classifiers [17]. Kaya et al. [18] proposed a texture based retrieval system based on LBP and GLCM independently. For classification accuracy of butterfly images, ELM classifiers has also been deployed. But, this system lacks the encapsulation of Relevance feedback which has also been added to an ELM model in the proposed work to retrain the images of ELM. Support vector machine (SVM) have also been used for classification purpose. For the extraction of features Convolutional neural networks (CNN) have been used [19].

Refinement of the user's query is considered as an important tool in aggravating the accuracy of the CBIR system. Therefore, the technique of Relevance feedback has played a prominent role in improving the query of a particular system based on the feedback obtained by the user.

Relevance feedback can also be used in combination with many meta-heuristics. Sequential forward selector (SFS) meta-heuristic is utilized in combination with relevance feedback using a single iteration. But, only Relevance feedback without any machine learning model do not yield desired results and both these concepts have been utilized in the implemented technique. Many distance metrics have also been tested and analysed here [20].

In literature, the texture descriptors do not provide information concerning frequency relating to co-occurring local patterns of an image. But, the proposed texture descriptor is an intelligent and normalized fusion descriptor which contributes to producing a highly effective and accurate system with diversified resolution and prominent interconnection among pixels of an image.

3. Feature Extraction Techniques

In order to extract texture features, two divergent techniques namely Gray level Co-occurrence Matrix (GLCM) and Discrete Wavelet Transform (DWT) have been utilized here. To boost the outcome capability of the CBIR system, an efficient method which is an integration of these two techniques (Co-DGLCM) is developed and augmented results are obtained. To improve the accuracy of the proposed system, Extreme learning machine (ELM) as a classifier has also been employed. Lastly, if the attained results are less than the desired value then, Relevance feedback is utilized to rectify the training of ELM and enhanced results are obtained.

3.1. Discrete Wavelet Transform

For analyzing a signal in time-frequency zone, initially Fourier based transforms like Discrete Fourier transform (DFT) [21] and Discrete Cosine Transform (DCT) [22] are used. But, these traditional methods are deficient, as they do not convey the local information of an image. Also, the images regenerated by these techniques are of deprived standard, especially at the edges because of the presence of high-frequency bristly components [23]. In the wavelet domain, both Gabor wavelet and Discrete Wavelet Transform are highly prominent. But, due to a large dimension of feature vectors, Gabor wavelet takes more time in image analysis as compared to DWT. So, Discrete Wavelet transform is utilized here as one of the applied technique which has many advantages like: (1) Less memory requirements due to small feature vector dimension (2) Multi-resolution analysis (3) Precise feature extraction.

A one dimensional DWT putrefies an input $I(x) = L^2(R)$ relatively with reference to wavelet function $\psi(x)$ with [24] scaling function $\phi(x)$ and is denoted as

$$(x) = \sum_{k \in Z} S_{j_0, k} \phi_{j_0, k}(x) + \sum_{j \geq j_0} \sum_{k \in Z} W_{j, k} \psi_{j, k}(x) \quad (1)$$

where $\phi_{j_0, k}(x) = 2^{\frac{j_0}{2}} \phi(2^{j_0} x - k) \quad (2)$

And $\psi_{j, k}(x) = 2^{\frac{j}{2}} \psi(2^j x - k) \quad (3)$

Here, $S_{j_0, k}$ and $W_{j, k}$ denotes the scaling and wavelet coefficients respectively.

A two-dimensional wavelet transform can be systematically resolved by employing to each column of an image, a 1-D filter bank and subsequently implementing the same 1D filter bank to consequential coefficients of each row. Thus, by continually filtering the resultant picture coefficients in a tier of row and column fashion, until the desired wavelet transform is obtained. In this method, solitary sub-image which is in low pass region (LL1) and a trio of sub-images obtained in high pass regions (HL1, LH1 and HH1) are acquired after an initial stage of disintegration [25]. Among the acquired three high pass sub-images, one high-low frequency vertical band (90°) (HL1), one low-high frequency horizontal band (0°) (LH1) and single high-high frequency diagonal element band (45°) (HH1) are created. To achieve a higher level of image decomposition, this procedure repeatedly takes place on each and every low pass sub-image.

These sub-images obtained in the low pass region represent similar low or little resolution of native input and sub-images obtained in high pass regions represent changes of brightness level in a vertical, horizontal and diagonal orientation. DThe triad partition of wavelet dissolution of an input $Z = I(x, y)$ depicting the order of $m \times n$ pixels is shown in Fig. 1

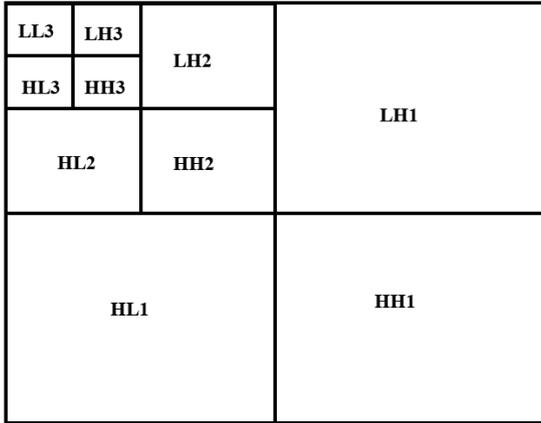


Figure 1. Three level 2D DWT image dissolution

3.2. Gray Level Co-occurrence Matrix

For the withdrawal of texture features, the prominent second order statistical methods are Gray level Co-occurrence matrix (GLCM), Gray level Difference matrix (GLDM) and Gray level run length matrix (GLRLM) [26]. Among these, GLDM is too trouble free in calculation and is based on the difference between gray level pixel pairs. But, its main drawback is that with the change in gray level variance, it also becomes variant. On the other hand, GLRLM [27] is based on counting the number of runs which in turn depicts gray levels in an image. This technique suffers from meagerness to represent the pattern of an image and also its computational involvement is high. Therefore, the best second-order statistical texture feature extraction technique is GLCM which has many dominant attributes like: (1) Rotation invariance (2) Diverse applications (3) Simple implementation and fast performance (4) Numerous resultant (Haralick) parameters.

It is a traditional and analytical technique used to extract texture features for the retrieval and classification of different images. [28] The spatial correlation enclosed by the pixel duplet in any given image is also computed by this geometrical method. A co-occurrence matrix denoted by $P_{dis}(m,n)$ specifying grey levels, contains information about two pixels: Gray level content m is denoted by the first pixel and content n is denoted precisely by the second pixel which further is separated by a distance denoted by dis . These specifications are chosen according to a specific angle. Matrices produced by this technique gives the gray level spatial frequencies which further gives association among pixels which are adjacent and have distinct distances amidst them [29].

Therefore, the description of GLCM is as follows:

$$P_{dis,\theta}(m,n) = P_r(I(p_1) = m \wedge I(p_2) = n \wedge \|p_1 - p_2\| = d_{is}) \quad (4)$$

In the above-given equation (4) p_1 and p_2 denote positions

in gray scale image I , the probability is denoted by P and θ denotes the range of different angle directions given by 0° , 45° , 90° and 135° . Therefore, a GLCM image is represented by d as its vector used for movement, δ as its radius and θ as orientation. A generalized GLCM matrix can be represented by Fig. 2.

Gray Tone	0	1	2	3
0	# (0,0)	# (0,1)	# (0,2)	# (0,3)
1	# (1,0)	# (1,1)	# (1,2)	# (1,3)
2	# (2,0)	# (2,1)	# (2,2)	# (2,3)
3	# (3,0)	# (3,1)	# (3,2)	# (3,3)

Figure 2. A generalized GLCM

Therefore, for a given specific test image with gray tone values, a GLCM matrix is formed which is the spatial co-occurrence dependence matrix.

Prominent four types of GLCM feature parameters which are subjected to be used for the extraction [30] of textual content of an image and are denoted by:

$$Contrast = \sum_m \sum_n [P(m,n) * (m - n)^2] \quad (5)$$

The disparity between the topmost and the bottom most conterminous pixel sets is given by Contrast intensity.

$$Correlation = \frac{\sum_m \sum_n (m - u_x) * (n - u_y) * P(m,n)}{\sigma_i \sigma_j} \quad (6)$$

The correspondence between a reference pixel and its adjoining pixels in an image is diagnosed by making use of Correlation. It considers the mean and standard deviation of a matrix by encapsulating both the row and column of that particular matrix.

$$Homogeneity = \sum_m \sum_n \frac{P(m,n)}{1 + |m - n|} \quad (7)$$

In the spatial domain, the proximity among gray levels in an image is defined by the term homogeneity.

$$Energy = \sqrt{\sum_m \sum_n P(m,n)^2} \quad (8)$$

Energy of a texture denotes the cyclic consistency of gray level allocation in an image.

3.3. Extreme Learning Machine

Extreme learning machine (ELM) is a single-layer feed-forward neural network (SLFN), designed to be used for classification and regression applications. It was initially

introduced by Huang et al. [17]. ELM has been successfully utilized in many research problems like pattern recognition, classification, fault identification, over-fitting, etc. In this system, the input weights are selected without any conscious decision and by using a specified analytical technique, its output weights are decided. This algorithm indeed is based on a single hidden layer and number of nodes in the hidden layer is an important parameter to be decided. ELM has many advantages as compared to many state-of-the-art traditional algorithms like least human involvement, [31] swift learning speed, complimentary universal capability, convenience to use, varied kernel functions, etc. [32]. The basic diagram of ELM is shown in Fig. 3.

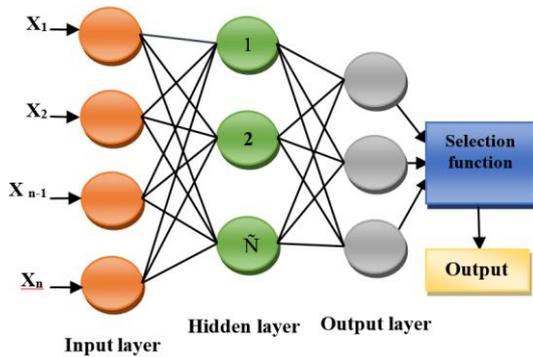


Figure 3. A basic form of Extreme Learning Machine

For a given, N different samples (X_i, t_i) , where $X_i = [x_1, x_2, x_3, \dots, x_n]^T \in \mathbb{R}^n \times \mathbb{R}^m$ (where $i=1, 2, \dots, N$). An ELM is a SLFN with \tilde{N} hidden nodes, then its activation function $f(x)$ can be represented as:

$$\sum_{i=1}^{\tilde{N}} \beta_i f(u_i x_j + v_i) = o_j \quad (9)$$

In the above equation, the weight vectors which connects the input nodes to the i^{th} hidden nodes are given by $u_i = [u_{i1}, u_{i2}, u_{i3}, \dots, u_{in}]^T$ and $\beta_i = [\beta_{i1}, \beta_{i2}, \beta_{i3}, \dots, \beta_{im}]^T$ represents the connecting weight vectors between the i^{th} hidden nodes and the output nodes. The threshold value for the hidden nodes is given by v_i whereas u_i and x_j is the inner product.

The above equation can be concisely written as:

$$H = \beta O \quad (10)$$

Where

$$H = \begin{bmatrix} f(u_1, x_1, v_1) & \dots & f(u_{\tilde{N}}, x_1, v_{\tilde{N}}) \\ \vdots & \vdots & \vdots \\ f(u_1, x_N, v_1) & \dots & f(u_{\tilde{N}}, x_N, v_{\tilde{N}}) \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix} \text{ and } O = \begin{bmatrix} O_1^T \\ \vdots \\ O_1^T \end{bmatrix}$$

Here, the hidden layer output matrix of ELM is represented by H . Also, the hidden matrix H becomes invertible, if the number of given samples \tilde{N} becomes equal to hidden node parameters H . However, the learning attributes of ELM, u_i, v_i and the hidden nodes can be allocated randomly, in the absence of any input data, therefore the output weights β of a linear system can be calculated by implementing least square technique and is given by:

$$\hat{\beta} = H^\dagger T \quad (11)$$

Where H^\dagger signifies the Moore-Penrose conception in one of its version. So, from equation (11), we can see that the output weights are calculated by using a simple mathematical equation, thereby avoiding any lengthy procedures. Thus, the algorithm of ELM can be summarized in three steps, which are as follows:

- (1) Assigning of hidden node parameters u_i and v_i , where $i=1, 2, \dots, \tilde{N}$.
- (2) Calculation of hidden layer matrix H , by using activation function.
- (3) Output weights calculation by using $\beta = H^\dagger T$.

Thus, ELM transforms a complex problem into a simpler and linear function. Fast speed, more accuracy and many more advantages contribute to making this technique more sophisticated and precise as compared to many other customized methods.

3.4. Relevance Feedback

There are many ongoing advancements in the field of Content-based image retrieval. But still, the issue of the semantic gap is prevailing over this system. Semantic gap is the difference between high-level human perception and low-level machine understanding. To overcome this gap, many types of intelligent techniques have been used with CBIR systems in order to retrieve accurate and precise results. Relevance Feedback (RF) is one of the prominent techniques which has been discussed in the literature [33].

It is a strategy which helps to refine a particular image based on the feedback obtained by the user. To search the system with massive database images, text or a combination of images annotated with text can be used as a query. Then, a set of relevant images are obtained based on a specific query image. These retrieved images are analysed by the user and finally, the query image is refined by using Relevance feedback which selects the best-matched images, based on some common features. This process works iteratively until the desired results are obtained or the user gets satisfied. This intelligent technique can also be used in combination with many other concepts like Support vector machines (SVM), Neural networks, Deep learning, machine learning, etc. [34]

The query input given by the user can be broadly classified into three types: The first category consists of a system in which a query image is composed of only keyboard text

letters. This technique has some limitations like polysemy, synonymy, homonymy etc. So, finding the desired images based on the user’s intention is a major issue of concern. A query can also be given in the form of an image, which is the second medium of inputting a query image. This technique has removed many ambiguities, which were present in the traditional method of the query by text. Also, this method has gained vast popularity in recent times due to its numerous applications in image processing. Relevance feedback can be considered as the third category of providing a query image, indeed through the iterative refinement of a user’s query image. The three basic ways of a query refinement are as follows:

- (1) Extension of Query: In this technique, the neighboring images of an actual query image are also included in it, based on the feedback obtained by the user. Thus, in a way, an expansion of an original query image is done.
- (2) Query Re-Weighting: This method enhances the weights of some prominent attributes of an image and simultaneously reduces the weights of some unimportant attributes. In this way, a query becomes more refined.
- (3) Movement of Query: A query is moved close to the required images by the adjustment in the attributes of a utilized distance function.

In this paper, to make the hybrid textural system more effective, Relevance Feedback is utilized. It works on the relevant images obtained after classification by ELM. If the obtained result is less than a defined threshold, then the role of relevance feedback comes into play. It works on the refinement process until the satisfactory results are obtained.

4. Architectural Framework

In the proposed system, a unique technique denoted by Co-DGLCM has been described which is a combination of DWT and GLCM. Both these techniques are jointly inexplicable. The superiority of both these techniques can be perceived with the help of their prominent characteristics like: DWT belongs to the family of wavelets which dispense localized frequency data unambiguously. DWT has an awesome temporal resolution which precisely apprehends both location and frequency statistics from an image. It also has swift computation due to consumption of a series of diverse filters. Similarly, GLCM has highly factual results and takes very little computational time. Therefore, the proposed approach is doubtlessly relevant as it is based on a combination of hybrid texture descriptor and a machine learning model with human’s feedback in the form of Relevance feedback. This approach can be disintegrated into five major phases as shown in Fig. 4.

Phase 1: Pre-processing

In the first phase, a pre-processing step is employed. Brodatz database consists of 13 categories and each image bears the size of 512x512. Here, every single image is further split,

producing 16 resultant images. Finally, this database contains a total of 1456 (7x16x13) images resized to 128x128, 13 categories with 7 angle directions (0°, 30°, 60°, 90°, 120°, 150° and 200°).

Vistex dataset consists of 40 initial images and each image bear the size of 512x512. Finally, a database of 640 images is created by partitioning the original 512x512 image into 128x128 sixteen non-overlapping images (40x16).

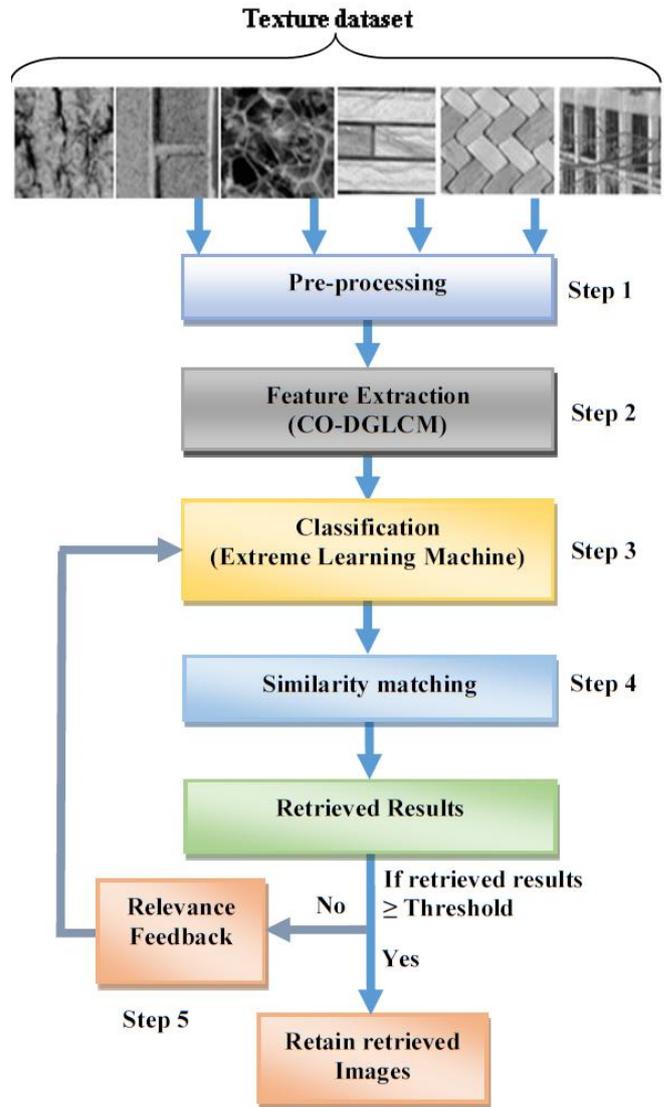


Figure 4. The architecture of the proposed system

Phase 2: Feature Extraction and Normalization

This second phase of the proposed system is based on the extortion of image features with the help of Discrete wavelet transform (DWT) and wavelet coefficients are obtained and the resultant matrix is formed. Then Gray level Co-occurrence matrix (GLCM) feature vector with four angle directions viz. 0°, 45°, 90° and 135° is obtained. Finally, both feature vectors are normalized by using minimum-maximum normalization which is denoted by:

$$\text{Obtained normalised features} = \frac{y - \text{minimum}}{\text{maximum} - \text{minimum}} \quad (12)$$

Here, y = value of a particular feature
 minimum = Bottom value of every single feature vector
 maximum = Highest value of every single feature

Normalization brings the feature dimensions into a common range. Both obtained feature vectors are fused and concatenated to form the final and resultant feature vector after normalization. These feature extraction and normalization steps are applied to all database images also. Z-score normalization and Decimal scaling normalization are also among the prominent normalization methods. But, the main disadvantage of Z-score Normalization is that it always achieves a normal distribution. But, if this condition is not met, then the vague results are produced.

In Decimal scaling, the normalization is achieved by moving the decimal point of input values. But, this technique is generally based on calculating the least and the largest value of given data which is difficult in some cases. Moreover, if the assumption about these values is done, then also the results could be impermissible. Therefore, on the basis of these conclusions, Min-Max Normalization is preferred because the correspondence among all the data values is conserved, without any bias introduction [35].

Phase 3: Classification

The third phase of this proposed system is based on classification. Here, the hybrid features of GLCM and DWT are applied in the form of an input to the ELM. Here, the ELM classifier is initially trained with the complete specific database. The output of this step is the texture images with meticulous categorization. With the correct categorization, the performance and accuracy of the proposed system is enhanced. The working of the ELM network depends on the type of activation function used and hidden layer total neurons. Various working parameters are given in Table 1. Here, Radial basis function (RBF) is used as the activation function.

Table 1. Working parameters of the proposed technique

Number of Input Features	16+672
Number of Output (classes)	13 (Brodatz); 8 (Vistex)
ELM Kernel Type	Radial Basis Function (RBF)
Kernel Parameters	100
Algorithm	Feed-forward single hidden layer neural network

Phase 4: Similarity matching

After classification by the ELM network, the whole training dataset is successfully classified into varied categories. Now, based on the query image, a similarity matching is done between a given query image and the respective category to which it belongs. These categories are formed as a result of ELM classification. After the results of similarity calculation are acquired, the resultant images are arranged in increasing

order based on the utilized distance metric. A result of zero with regard to distance metric exhibits accurate resemblance between two images. Many types of distance metrics are utilized for the purpose of calculating similarity. Some of the prominent distance metrics which are used in similarity calculation are given under:

$$\text{Distance}_{\text{Euclidean}} = \sqrt{\sum_{i=1}^n (|I_i - D_i|)^2} \quad (13)$$

$$\text{Distance}_{\text{Manhattan}} = \sum_{i=1}^n |I_i - D_i| \quad (14)$$

$$\text{Distance}_{\text{Cosine}} = 1 - \cos\theta = \frac{X \cdot Y}{\|X\| \|Y\|} \quad (15)$$

$$\text{Distance}_{\text{Minkowski}} = \left[\sum_{i=1}^n (|I_i - D_i|)^{1/P} \right] \quad (16)$$

$$\text{Distance}_{\text{Spearman}} = 1 - \frac{6 \sum_{i=1}^n ((\text{rank}(X_i) - \text{rank}(Y_i))^2)}{n(n^2 - 1)} \quad (17)$$

Here, I_i denotes the input query image and D_i depicts all database images related to Euclidean, Manhattan and Minkowski distances whereas X_i and Y_i denote i_{th} values of sequences X and Y related to Spearman and Cosine distances.

Phase 5: Relevance Feedback

In the last phase of the proposed system, relevance feedback is used. The result obtained after ELM classification is compared to a threshold value, which is set at 95%. If the obtained results are greater than or equal to 95% then, the same results are retained. But, if it is less than the threshold then, relevance feedback is applied in the form of two iterations. This process rejects the non-relevant images based on user's feedback and finally, the ELM data is updated and retrained. Now, based on these refined images, final top N images are retrieved. This process enhances the accuracy of the proposed system to a significant value.

4.1. Evaluation metrics

In CBIR systems, after the recovery of indistinguishable images from a huge repository of images, the capability of a particular system can be concluded with respect to many evaluation parameters [36-37]. Precision and Recall are the most well-known evaluation metrics. These are defined using the given equations:

$$\text{Precision}(P_i) = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} \quad (18)$$

$$\text{Recall}(R_i) = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images in the database}} \quad (19)$$

In the proposed work, total database images are regarded as query input images and afterwards, Precision and Recall are calculated. Total (overall) average precision and total (overall) average recall based on the entire image set are

given by the following equations:

$$\text{Total average Precision} = \frac{1}{|DB|} \sum_{k=1}^{|DB|} \text{Precision}(P_i) \quad (20)$$

$$\text{Total average Recall} = \frac{1}{|DB|} \sum_{k=1}^{|DB|} \text{Recall}(R_i) \quad (21)$$

In these above equations, total images in the database are represented by *DB*.

5. Experimental Results and Analysis

Two texture datasets, one Brodatz [38-39] and another MIT-Vistex [40] are chosen for experimentation and analysis. In each experiment overall average precision, overall average recall and accuracy are calculated. Various texture extraction techniques like Local Binary Pattern (LBP), GLCM, DWT, Gabor wavelet and Curvelet transform have been analyzed, tested and compared with the proposed method. The performance of the proposed technique outperforms these state-of-the-art texture feature extraction techniques. Some sample images from both the datasets are shown in Fig.5.



Figure 5. Some sample images from Brodatz and MIT-Vistex datasets

In the initial assessment, Brodatz texture dataset is evaluated. In this experiment, each image is further divided into 16 resultant images, finally forming 1456 (7x16x13) images of the size 128x128. Here, 7 denotes angle orientations and 13 denotes different categories. In the present experiment total dataset images are used as query input images and final resultant images are recovered in the retrieval pattern of 25, 35...up to 65. Graphs depicting average precision (total) and average recall (total) are shown in Fig. 6(a) and 6(b).

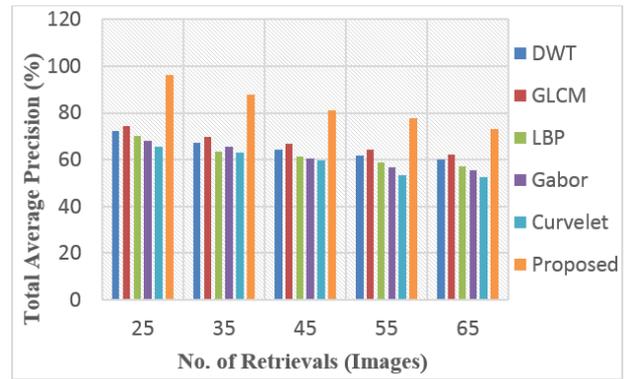


Figure 6(a). Total Average Precision (%) Brodatz

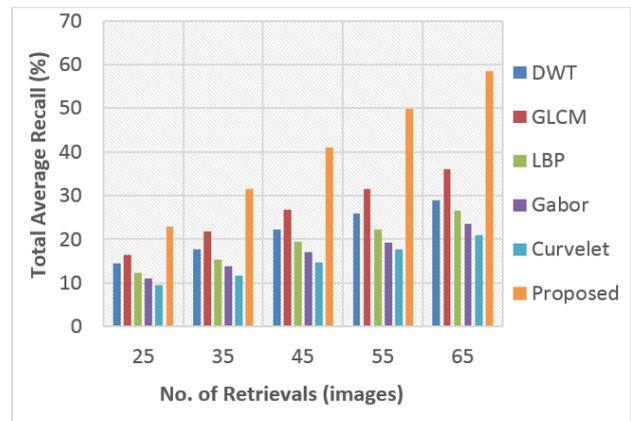


Figure 6(b). Total Average Recall (%) Brodatz

It is clearly undoubted from the comparative analysis of the above techniques that the proposed technique, outperforms the other compared techniques and total average precision of 96.35% is obtained on the retrieval of 25 images. This proposed texture descriptor is also benefitted by the encapsulation of ELM classifier and relevance feedback. Therefore, the results of precision and recall are outstanding as compared to some commonly used texture extraction techniques. Retrieval result for a particular Query image on the Brodatz database is shown in Fig. 7.

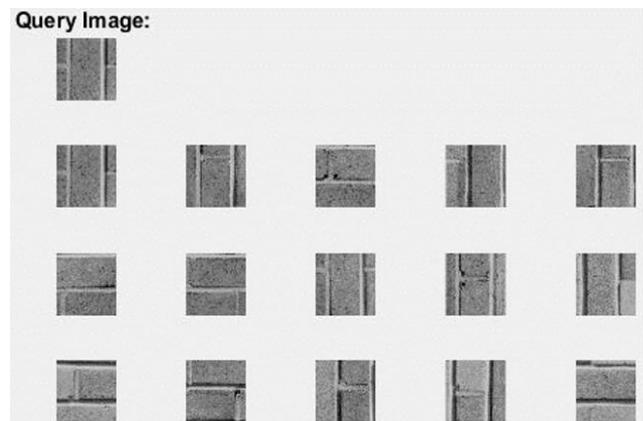


Figure 7. Retrieval result of a query image from Brodatz dataset

From Fig. 7, it can be seen that the retrieved results for a specific query image are accurately classified and belong to the respective category of query image.

The second experimental analysis is carried on another MIT-Vistex [41] dataset. This dataset consists of 40 initial images and a database of 640 images is created by partitioning the original 512x512 image into 128x128 sixteen non-overlapping images (40x16).

Total average precision and total average recall on Vistex dataset are shown in Fig. 8(a) and 8(b) and total average precision for this database is 97.34% on the retrieval of 16 images. Again in this 2nd experiment, total dataset images are regarded as input query images and recovered outcome images are procured in the sequence of 16, 32.....up to 80.

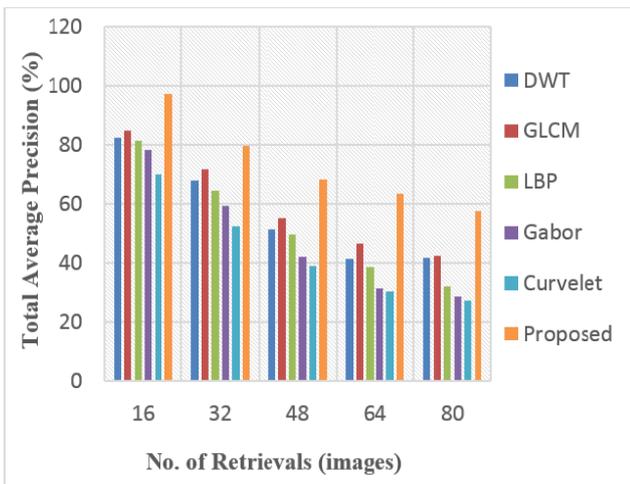


Figure 8(a). Total Average Precision (%) Vistex

Again, it can be seen from the precision values obtained for the proposed technique that, the implementation has significant results as compared to many basic texture feature extraction techniques.

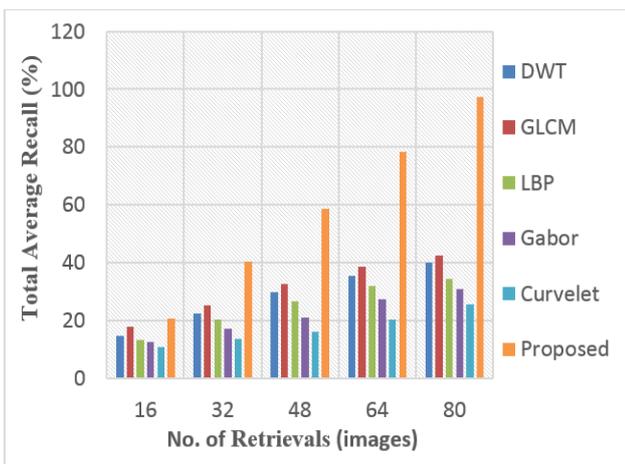


Figure 8(b). Total Average Recall (%) Vistex

A specific query image and its retrieval results for Vistex dataset are shown in Fig. 9.



Figure 9. Retrieval result of a query image from Vistex dataset

Again, it can be concluded from the retrieved results that images belonging to the same native category of the query image are retrieved and the accuracy of the system is remarkably enhanced.

5.1. Proposed technique using divergent distance metrics and state-of-the-art comparison

There are different similarity measures [42] which can be used for the purpose of calculation. Here, five distance metrics namely Euclidean, Manhattan, Cosine, Spearman and Minkowski are tested on the proposed method and the results reveal that the Euclidean distance metric outperforms the other four used distance metrics [43]. The performance of all distance metrics is shown in Table 2.

Table 2. Performance of the proposed system with various distance metrics

Distance Metric	Brodatz Precision (%)	Vistex Precision (%)
Euclidean	96.35	97.34
Manhattan	74.31	85.32
Minkowski	76.05	87.7
Cosine	78.27	88.43
Spearman	61.03	72.53

Manhattan distance is a distinctive case of Minkowski distance. It produces innumerable false negatives and does not yield accurate results. The outcomes of the Cosine distance metric are variant with changes in the applied inputs. Spearman distance is complicated and cumbersome. Sometimes, it also shows vague results. Euclidean distance metric is based on weighted and normalized attributes and has speedy computational performance. Therefore, the Euclidean

distance metric is used here.

The comparison of the proposed system with the state-of-the-art techniques in terms of average precision on Brodatz dataset is shown in Table 3.

Table 3. Comparison of the proposed system with the prominent state-of-the-art techniques

Method/Ref. No.	Brodatz	
	Total Average Precision (%)	
WLSBP [11]	89.85	
PLT [13]	92.5	
CoALTP [36]	77.94	
LNIP [4]	80.00	
LQEPB [12]	76.00	
Proposed	96.35	

The average precision of the proposed system during the various phases of its functioning is shown in Table 4.

Table 4. Average Precision (%) of the proposed system During three phases of implementation

Average Precision (%)			
Name of the Technique			
Dataset	Co-DGLCM	Co-DGLCM+ELM	Co-DGLCM+ELM+RF
Brodatz	88.93	93.37	96.35
MIT-Vistex	90.56	94.53	97.34

Similarly, the relative comparison of the proposed technique with the state-of-the-art techniques on MIT-Vistex dataset is given in Table 5.

Table 5. Comparison of the proposed system with the prominent state-of-the-art techniques

Method/Ref. No.	Vistex	
	Average Precision (%)	
LTriDP [5]	89.9	
LNIP [4]	90.00	
CoALTP [36]	86.53	
LQEPB [12]	95.00	
Proposed	97.34	

From Table 3 and 5, it is clearly visible that the average precision obtained on Brodatz and Vistex dataset is remarkably higher than many state-of-the-art prominent texture analysis techniques.

The proposed system also outperforms customary texture descriptors in terms of accuracy. The accuracy of the proposed system is given in Table 6. The confusion matrices based on classification by ELM are given in Table 7 and 8 for Brodatz and MIT-Vistex respectively.

Table 6. Accuracy of the proposed system

Dataset	Accuracy (%)
Brodatz	98.07
MIT-Vistex	99.21

From Table 6, it can be clearly visualized that the accuracy of the proposed technique by using a machine learning model, specifically an Extreme learning machine (ELM) is significant and moreover this ELM model is retrained by using a feedback from the user in terms of Relevance feedback is form of “Cherry on a cake”.

By obtaining a prominent and an effective value of accuracy, it can be concluded that the machine learning model has done an accurate classification. This performance of classification can be studied by a confusion matrix. A table which describe the performance of a classification model on a set of test data for which the true values are known. Therefore, the obtained confusion matrices for both the datasets are given in Table 7 and Table 8.

Table 7. Confusion matrix based on Brodatz dataset

Category	Confusion Matrix (Brodatz dataset)												
1	108	1	1	0	0	0	1	1	0	0	0	0	0
2	0	109	0	1	0	2	0	0	0	0	0	0	0
3	0	0	111	0	0	0	0	0	1	0	0	0	0
4	1	0	0	109	0	0	0	1	0	1	0	0	0
5	0	0	0	0	110	1	1	0	0	0	0	0	0
6	0	0	0	0	0	112	0	0	0	0	0	0	0
7	0	0	1	0	0	0	110	0	1	0	0	0	0
8	0	0	0	0	0	0	0	111	0	1	0	0	0
9	0	0	1	0	0	0	0	1	110	0	0	0	0
10	1	0	0	0	0	0	2	0	1	108	0	0	0
11	0	1	0	0	0	0	0	0	0	0	111	0	0
12	0	0	0	0	0	0	0	0	1	0	0	111	0
13	0	1	0	2	1	0	0	0	0	0	0	0	108

From the confusion matrix given in Table 7, the diagonal elements of the matrix represents the total number of retrieved images per category. Since, there are 13 categories in Brodatz dataset with each category containing 112 images in it. Therefore, these diagonal elements represents the total images out of 112 and it can be seen that from 1st category 108 images are retrieved, 109 from second category and so on. Thus, out of 112 images, a significant number of images are retrieved thus enhancing the retrieval accuracy of the proposed system.

Table 8. Confusion matrix based on MIT-Vistex dataset

Category	Confusion Matrix (Vistex dataset)							
1	80	0	0	0	0	0	0	0
2	0	80	0	0	0	0	0	0
3	0	0	80	0	0	0	0	0
4	0	0	0	78	0	0	0	2
5	0	0	0	0	80	0	0	0
6	0	1	0	1	0	76	2	0
7	0	0	0	0	0	0	80	0
8	0	1	1	0	1	0	0	77

Again, from the confusion matrix of Vistex dataset, it can be concluded that out of 80 images particularly of 8 categories, an accuracy of 99.21% is achieved.

Hence, from the results obtained on both the renowned texture datasets namely Brodatz and MIT-Vistex, it is evident that the implemented technique is significantly superior to many existing texture feature descriptors.

6. Conclusion and Future Trends

This paper depicts an innovative technique for texture feature extraction which is based on an efficient combination of Gray level co-occurrence matrix (GLCM) and Discrete wavelet transform (DWT). These techniques are utilized for the formation of a hybrid texture descriptor, namely combined DWT-GLCM (Co-DGLCM). The process of normalization has been used here to create a hybrid feature vector (HFV) from two obtained independent feature vectors. This HFV is applied as input to an Extreme learning machine (ELM) which works as a classifier. Then, based on a decisive condition of a threshold, Relevance feedback is applied. This Relevance feedback is used in the form of some iterations based on the user's feedback. This Extreme learning based-Relevance feedback framework helps in the evolution of an intelligent and modified system for learning and classification. GLCM has precise inter-pixel and inter-pattern relationship and DWT has awesome temporal resolution which precisely apprehends both location and frequency statistics from an image. To validate these facts, simulation results are provided which are finally compared with many latest texture extraction techniques. Our future work will be focused on developing a hybrid CBIR system by extracting color, texture and shape features of an image. Moreover, the feature extraction will be done by incorporating deep learning techniques like auto-encoders, deep belief networks, etc. There are many real-life multimedia applications like medical diagnosis, crime detection for the extraction of fingerprints, face recognition, pattern recognition etc, where this texture based hybrid descriptor can be utilized. Last but not the least, the concept of Internet of things (IoT) can be used for the online transfer of desired images.

References

- [1] Kalampogia, A. and Koutsakis, P. (2017) H.264 and H.265 Video Bandwidth Prediction. *IEEE Transaction on Multimedia* DOI 10.1109/TMM.2017.2713642 : pp. 1-13.
- [2] Mukane, S.M., Gengaje, S.R. and Bormane, D.S. (2014) A novel scale and rotation invariant texture image retrieval method using fuzzy logic classifier. *Computer and Electrical Engineering* **40**(8): pp. 154–162.
- [3] Bala, A. and Kaur, T. (2016) Local texton XOR patterns: A new feature descriptor for content-based image retrieval. *Engineering Science and Technology an International Journal* **19**(1): pp. 101–112.
- [4] Banerjee, P., Bhunia, A.K., Bhattacharya, A., Roy, P.P. and Murala, S. (2018) Local Neighborhood Intensity Pattern – A new texture feature descriptor for image retrieval. *Expert system and Applications* DOI:10.1016/j.eswa.2018.06.044.
- [5] Verma, M. and Raman, B. (2016) Local tri-directional patterns: A new texture feature descriptor for image retrieval. *Digital Signal Processing: A Review Journal* **51**: pp. 62–72.
- [6] Ojala, T., Pietikäinen, M. and Mäenpää, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(7): pp. 971–987.
- [7] Satya, V., Gangavarapu, K., Krishna, G., Pillutla, M. and Descriptor, A.W.L. (2016) Local Tri-directional Weber Patterns: A New Descriptor for Texture and Face Image Retrieval. *International Journal of Computer Science and Information Technology* **7**(3): pp. 1571–1577.
- [8] Elgarrai, Z., El, O., Mustapha, M. and Hakim, K. (2016) Robust facial expression recognition system based on hidden Markov models. *International Journal of Multimedia Information Retrieval* **5**(4): pp. 229–236.
- [9] Walia, E. and Verma, V. (2016) Boosting local texture descriptors with Log-Gabor filters response for improved image retrieval. *International Journal of Multimedia Information Retrieval* **5**(3): pp. 173–184.
- [10] Zhang, W. (2018) A Feature Descriptor based on Local Normalized Difference for Real-World Texture Classification. *IEEE Transactions on Multimedia* **20**(4): pp. 880– 888.
- [11] Nadu, T. (2018) Two Improved Extension of Local Binary Pattern Descriptors using Wavelet Transform for Texture Classification. *IET Image Processing Repository* **12**(11): pp. 2–13.
- [12] Sucharitha, G. and Senapati, R.K. (2018) Local quantized edge binary patterns for colour texture image retrieval. *Journal of Theoretical and Applied Information Technology* **96**(2): pp. 291–303.
- [13] Sana, J.K. (2018) PLT-based spectral features for texture image retrieval. *IET Image Processing* **12**(11): pp. 2065–2074.
- [14] Das, R., Dash, J.K. and Mukhopadhyay, S. (2013) Rotation invariant textural feature extraction for image retrieval using eigen value analysis of intensity gradients and multi-resolution analysis. *Pattern Recognition* **46**: pp. 3256–3267.
- [15] On, T., Analysis, P. and Intelligence, M. (1996) Texture Features for Browsing and Retrieval of Image Data. *IEEE Transactions on Pattern analysis and machine Intelligence* **18**(8): pp. 837–842.
- [16] Rajavel, P. (2010) Directional Hartley transform and content based image retrieval. *Signal Processing* **90**(4): pp. 1267–1278.
- [17] Mansourvar, M., Shamshirband, S., Raj, R.G., Gunalan, R. and Mazinani, I. (2015) An Automated System for Skeletal

- Maturity Assessment by Extreme Learning Machines. *PLOS One* doi:10.1371/journal.pone.0138493.
- [18] Kayaa, Y., Kaycib, L., Tekinc R. and Ertuğrul, O.F. (2014) Evaluation of texture features for automatic detecting butterfly species using extreme learning machine. *Journal of Experimental & Theoretical Artificial Intelligence* **26**(2): pp. 37-41.
- [19] Fu, R., Li, B., Gao, Y. and Wang, P. (2016) Content-Based Image Retrieval Based on CNN and SVM. In *Proceedings of International Conference on Computer and Communications*, Chengdu, 14-17, October (IEEE), 1-5.
- [20] Mosbah, M., and Boucheham B. (2017) Distance selection based on relevance feedback in the context of CBIR using the SFS meta-heuristic with one round. *Egyptian Informatics Journal* **18**(1): pp.1-9.
- [21] Rogelio, R., Benjamin, V., Roumen Z., Michael S., and Jose María B. (2017) Discrete Wavelet Transform and Its Application for Noise Removal in Localized Corrosion Measurements. *Hindawi International Journal of Corrosion*, **2017**(7925404): pp.1-7.
- [22] Himanshu, M. (2014) Comparison of DCT and Wavelet based Image Compression Techniques. *International Journal of Engineering Development and Research* **2**(1): pp.664-669.
- [23] Parsons, S., Arjan, B., and Martin, O. (2000) Advantages and disadvantages of techniques for transforming and analyzing Chiropteran echolocation calls. *Journal of mammology* **81**(4): pp. 927-938.
- [24] Murala, S. and Wu, Q.M.J. (2014) Expert content-based image retrieval system using robust local patterns. *Journal of Visual Communication and Image Representation* **25**(6): pp. 1324-1334.
- [25] Pandey, D. (2017) An Efficient Low Level Features of CBIR using Wavelet GLDM and SMOSVM Method. *International Journal of Signal Processing, Image Processing and Pattern Recognition* **10**(3): pp. 13-22.
- [26] Loris, N., Shery, B., Stefano, G., Emanuele, M., and Tonya, B. (2013) Different Approaches for Extracting Information from the Co-Occurrence Matrix. *PLOS one* **8**(12): pp.1-9.
- [27] Robert, S. (2016) A comparison of gray-level run length matrix and gray-level co-occurrence matrix towards cereal grain classification. *International Journal of Computer Engineering & Technology* **7**(6): pp. 9-17.
- [28] Annrose, J., and Seldev, C. (2018) An efficient image retrieval system with structured query based feature selection and filtering initial level relevant images using range query. *Optik* **157**: pp.1053-1064.
- [29] Carlos, W.D., Renata, M.C.R., and Candeias, A.L.B. (2010) Texture classification based on co-occurrence matrix and self-organizing map. In *Proceedings of International Conference on Systems, Man and Cybernetics*, Istanbul, Turkey, 10-13, October (IEEE), 2487-2491.
- [30] Mehdi, E., Aroussi, E. and Houssif, N. El. (2018) Content-Based Image Retrieval Approach Using Color and Texture Applied to Two Databases (Coil-100 and Wang). *Springer International Publishing*, https://doi.org/10.1007/978-3-319-76357-6_5 pp. 49-59.
- [31] Ding, S., Xu, X. and Nie, R. (2013) Extreme learning machine and its applications. *Neural computing and Applications* **25**(3-4): pp. 1-11.
- [32] Liu, S., Wang, H., Wu, J. and Feng L. (2015) Incorporate Extreme Learning Machine to content-based image retrieval with relevance feedback. In *Proceedings of 11th World Congress on Intelligent Control and Automation* Shenyang, China, 29 June-4 July (IEEE), 1010-1013.
- [33] Shikha, B., Gitanjali, P. and Pawan, D. (2018) A journey from basic image features to Lofty human intelligence in Content based Image retrieval system: Motivation and Applications. In *Proceedings of 5th International Conference on Computing for Global Sustainable development*, Delhi, 14-16 March 1-6.
- [34] Kien-Ping, C. and Chun Che, F. (2004) Relevance feedback and intelligent technologies in content- based image retrieval system for medical applications. *Australian journal of Intelligent Information Processing* **8**(3): pp. 113-122
- [35] <https://www3.nd.edu/~rjohns15/cse40647.sp14/www/content/lecture%20s/07%20-%20Data%20Transformation.Pdf>.
- [36] Naghashi, V. (2018) Co-occurrence of adjacent sparse local ternary patterns: A feature descriptor for texture and face image retrieval. *Optik-International Journal of Light and Electron Optics* **157**: pp. 877-889.
- [37] Khanh, L. B., Thi Thu, H. N. and Thanh, T.D. (2018) Deep Level Markov Chain Model for Semantic Document Retrieval. *EAI Endorsed Transactions on Scalable Information Systems* **5**(19): pp. 1-6.
- [38] Safia, D. and He. New Brodatz-Based Image Databases for Grayscale Color and Multiband Texture Analysis, ISRN Machine Vision [Online]. Available, 2013 http://multibandtexture.rec.herche.usherbrooke.ca/original_brodatz.Html.
- [39] <http://sipi.usc.edu/database/database.php?volume=textures>
- [40] <https://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html>
- [41] Dubey, S.R., Singh, S.K. and Singh, R.K. (2015) Local Wavelet Pattern: A New Feature Descriptor for Image Retrieval in Medical CT Databases. *IEEE Transactions on Image Processing* **24**(12): pp. 5892-5903.
- [42] Said, L., Khurshid, K. and Aman, A. (2018) Using Co-occurrence and Granulometry Features for Content Based Image Retrieval. *EAI Endorsed Transactions on Scalable Information Systems* **4**(1): pp. 1-5.
- [43] Alzu'bi, A., Amira, A. and Ramzan, N. (2015) Semantic content-based image retrieval: A comprehensive study. *Journal of Visual Communication and Image Representation* **32**: pp. 20-54.