CONTENT-BASED FEATURE FUSION REPRESENTATION FOR MARINE INVERTEBRATES

Mas Rina Mustaffa[1*], Noris Mohd Norowi[2], Sim May Yee[3]

[1,2,3]Department of Multimedia, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia

Email: MasRina@upm.edu.my[1](corresponding author), Noris@upm.edu.my[2], mayyeesim@outlook.com[3]

DOI: https://doi.org/10.22452/mjcs.vol33no3.1

ABSTRACT

Marine species representation and retrieval is crucial for its studies and conservation. The images of these animals are usually captured underwater with complex background, at different angle, position, and size, which makes it very hard to provide a good representation with the current methods. Most of the current methods only support content-based representation for marine life images with clear background (taken in laboratory or in environments which have been set up), containing just one animal in an image, or the animal is positioned nicely at the centre of the image. Responding to these important needs, a multi-feature method for Content-based Image Retrieval (CBIR) that employs colour, shape, and texture information of marine life images is proposed. The colour feature vectors are obtained by extracting first and second order of Colour Moments. Shape information is constructed through the implementation of Discrete Wavelet transform up to four sub-bands and the extraction of Canny edge feature. Texture features are obtained with the Zernike Moments (ZM) of order four and the extraction of few Grey Level Co-occurrence Matrix properties. We conducted two experiments to determine the best order of ZM as well as to measure the retrieval performance of the proposed descriptor. Retrieval results based on marine invertebrate and Fish4Knowledge datasets clearly shown that the proposed method has effectively obtained the best precision value at 11 standard recall levels (72.42%) and MAP value (67.7%). The proposed method is further measured based on the statistical two-tailed paired t-test and has revealed a significant improvement in retrieval effectiveness.

Keywords: ***Feature Fusion, Colour, Shape, Texture, Content-based Image Retrieval (CBIR)***

1.0 INTRODUCTION

Many digital images are being generated each day which result in higher demand for Content-based Image Retrieval (CBIR) systems that can manage and retrieve images effectively and efficiently [1]. Applications of CBIR are very broad, ranging from medicine (e.g. searching for images with similar case of breast cancer), to law (e.g. determining trademarks and logos which have been registered), biodiversity conservation [2], and many more. The needs for aquatic life management and conservation for the coming generation is becoming more apparent nowadays. Proper records are definitely crucial due to its large number in quantity as well as large variety in type. Frequently, experts that involve in sea life management and protection will need to manually identify and manage numerous species through the captured images where this can be quite troublesome and time consuming. Such tasks can actually be accomplished faster and more effectively through automatic content-based representation and retrieval.

Effective feature extraction method is the heart of any CBIR as it will determine how similar an image is to a given query. Feature extraction approaches though are generally domain-specific where an approach may well provide good representation for one domain but then perform poorly for other domains [3]. Marine life species are diverse and rich in features, where they come in various colours, shapes, sizes, and textures. Natural marine life images are usually captured under the conditions of complex background, at different angle, position, and size. This makes it challenging to provide a comprehensive representation and proper retrieval of the intended domain. To our knowledge, few researches have been done on marine life images, however, the images used for evaluation are mostly with clear background (taken in laboratory or in environments which have been set up), containing just one animal in an image, or the animal is positioned nicely at the centre of the image. Though this is ideal, it is not usually the case in real world situations. Therefore, the current methods may not able to effectively represent this complex type of marine life images. Owing to this motivation, this research aims to propose a CBIR feature extraction method that can effectively represent marine animals.

2.0 LITERATURE REVIEW

Marine species representation and retrieval are crucial for its field of studies and conservation [4]. Based on several literature reviews, we have identified few existing feature extraction approaches for marine life. Authors in [5] chose to focus on retrieving the contour shape of fish otoliths using Elliptic Fourier Descriptor. Otoliths are positioned in the ear of a fish where it can be analysed to define gender, age, inhabitants, and class, hence they can offer essential and related information for environmental studies. However, this can endanger the fishes as the person needs to have a physical or direct contact with the fishes to examine the otoliths as it is found in the interior part of the fish's ear. This may be useful for in depth studies on fishes but for the purpose of fish species representation and retrieval, it is sufficient to just use images.

In the work done by [6], colour and texture features are used to characterise fish species. Colour, statistical texture, and wavelet-based texture features are first extracted from each of the sub- image which will generate six sets of feature vector. Six fresh water fish species are used as the dataset. Colours in the HSV space are used to represent the colour information while Grey-scale Histogram and Grey Level Co-occurrence Matrices (GLCMs) are used to generate the texture feature vector. The proposed method was able to provide reasonable retrieval result. However, the images used for evaluation are of simple ones, captured in farms and only a certain part of the fish (not a full image of a fish) was photographed. Thus, the complete figure of a fish is not being considered in this study.

The authors in [7] contributed to an image segmentation technique for fish, where they recommended the combination of segmentation algorithm based on *K*-means together with mathematical morphology. *K*-means clustering algorithm is improved to support fish images with complex background. Firstly, images are converted into grey-scale and the contour of the fish body is then obtained through the mathematical morphology. It is a great effort on the authors' side to utilise natural fish images, when often, images with clear background are used. However, the application of mathematical morphology as feature extraction leads to high computation time due to the nature of the connected component analysis.

Shape feature is used in [8] for fish recognition where the first order of the Hu Moment Invariant is integrated with the geometric shape descriptor (eccentricity feature). The proposed method was able to provide a reasonable recognition rate but it is only tested on binary fish images. This method may need to be extended to include colour feature when applying on coloured fish images with complex background.

Feature fusion method for fish is introduced in [9] where the authors proposed the integration of Coloured Pattern Appearance model with Fourier descriptor. Colour segmentation based on region growing method is also utilised. The region growing method is known to involve high computation time. Additionally, Fourier descriptor could not represent the local features of an image which may result in a less effective method.

Colour and shape features have been considered for describing marine invertebrate images in [10]. Colour Moments is used to represent colour information while Discrete Wavelet transform and Canny edge detector are applied to extract shape details. The proposed method has successfully provided good retrieval effectiveness but further exploration on other features like texture could be done to support larger and more complex marine life dataset.

Mary and Dharma [11] described the coral reef images based on an Improved Local Derivative Pattern (ILDP) descriptor. The method was able to extract the diagonal directions of the pattern based on local derivative variations to represent colour and texture features. The work however did not explore on shape feature. It is probably not needed based on the sample of images shown in the paper which focused directly on the textural part of the coral reefs, rather than the whole image with background.

From the literature on content-based representations for marine species, it can be concluded that present methods have largely been driven by the extraction of limited features and images used for evaluation are of simple ones. The existing methods may not able to represent well marine life images captured naturally underwater. Further explorations are needed to construct a feature extraction method that is able to represent well marine life images that are captured under the conditions of complex background, at different angle, position, and size.

CBIR utilises content-based features such as colour, shape, and texture to explore and retrieve images, as an alternative of searching for images based on text-based descriptions. These visual content can be used independently or combined to represent more image information. Among the content-based features, colour could be considered as the most essential one due to the reasons that humans distinguish colour better in comparison to other visual content, it is not sensitive to image size, direction, and also insensitive to complex background. Some of the current colour-based descriptors can be found in [12- 15]. One of the most superior colour descriptors is the Colour Moments where it is recognised for its capability in analysing image's colour distribution. The method involves simple computation and produce small feature size. Colour moments can be implemented on any colour model by extracting up to three moments (mean, standard deviation, and skewness) for each colour channel.

Another equally important feature to colour is shape. Shape is vital in describing different objects which may have the same colour content. Apart from that, human is more prone to remember and describe images based on the objects contained in it rather than the colour details. That are two general categories for shape–based methods. The first type, contour-based methods, focus on representing an image based on the boundary details of objects or shapes. A good representation based on contour-based is the Canny edge detector [16]. It works by analysing the changes of intensity level within the image. Essential structural details from various objects can be extracted through this way and this intensely decrease the total data to be processed. Therefore, this technique has lower tendency to be affected by the noises originate in the image compared to other methods, hence, true edges can be discovered.

Contour-based approaches are suitable if the system's attention is solely on differentiating shapes based on its silhouette or contour. However, since this approach usually use just a small fragment of the shape (the contour), it is normally not invariant to noise and variations. This approach clearly could not be utilised successfully for images without sufficient contour details or with stronger region content. These restrictions can be resolved by using the second type of shape descriptors, known as the region-based methods. Region-based methods is said to be closer in bridging the semantic gap since it considers more pixels within a shape area or object of an image for representation. Due to the consideration of more pixels within a region for analysis in comparison to only the contour information, region-based approaches are usually more robust and have wider application, less sensitive to shape defection, and generally can offer more precise retrieval. The following literature discussed some of the well-known region-based descriptors [16-18]. Discrete Wavelet transform (DWT) is an example of a region-based method that can accurately define a shape's region at various decomposition level. The multi-resolution property that comes together with DWT allows for the coarse and fine details in a signal to be separated, including discontinuities and sharp spikes while at the same time preserving the actual structure of the data. DWT also only requires small feature size as representation with low computation time. The time required for processing is also decreased as the scale increases.

Texture feature is another crucial aspect in desribing images to differentiate the various surface properties. This is commonly done by describing the brightness's spatial distribution of a grey-scale image and categorise according to its directionality, regularity, coarseness, and few other properties. Example of objects with texture are brick wall, finger print, clouds, rocks, lizard skin, and rugs. Texture similarity is very useful when distinguishing images that have similar color content for example sky-sea, leaves-grass, and a few others. Texture methods can be categorised as structural-based, statistical-based, or spectral-based. Few texture descriptors have been proposed for CBIR [19 – 23]. Grey Level Co-occurence Matrix (GLCM) [24] is one of the earliest and most successful texture descriptors up until now. It calculates how frequent a particular pair of grey-level pixels at a certain distance d along a certain direction θ appears in the image. The representation according to this method is the various information of second order statistics of grey-level pixel pairs in an image such as energy, entropy, contrast, and homogeneity. GLCM is superior due to the reasons that it is simple to compute and provide extensive pattern representation of an image as it considers spatial pattern distribution at various distances and orientations. Zernike Moments (ZM) on the other hand is introduced by Fritz Zernike to define optical aberrations. It is recently been applied in image processing and computer vision as multi-spectral texture representation. ZM features can be obtained by mapping an image onto a set of Complex Zernike Polynomials. There are two parameters that need to be determined when using ZM, which are the number of order n and repetition m. Order of ZM refers to the number of concentric circular division and degree of repetition will determine the number of circular sectors. Therefore, different order used will result in different repetitions to be computed, hence leading to different image representation. Zernike polynomials are orthogonal to each other in nature, and thus, properties of an image can be characterised without redundancy or overlap of information among the moments [20]. Apart from that, ZM also provides multi-level information, rotationally invariant, and robust to noise and shape.

Due to the exclusive characteristics being extracted by different methods, there is still a need to construct a feature extraction technique that integrates the color, shape, and texture features based on the methods highlighted above. In our previous paper, the fusion method of colour and shape features have been constructed. It provides great result. We are extending the work done in [10] to include texture features so that more effective method can be developed for marine life content-based image representation and retrieval.

3.0 PROPOSED FEATURE FUSION METHOD

Fig. 1 below shows the complete steps of the proposed feature fusion method. The novelty of this method is based on the combination of the feature extraction methods used to accurately represent the respective colour, shape, and texture information of an image.

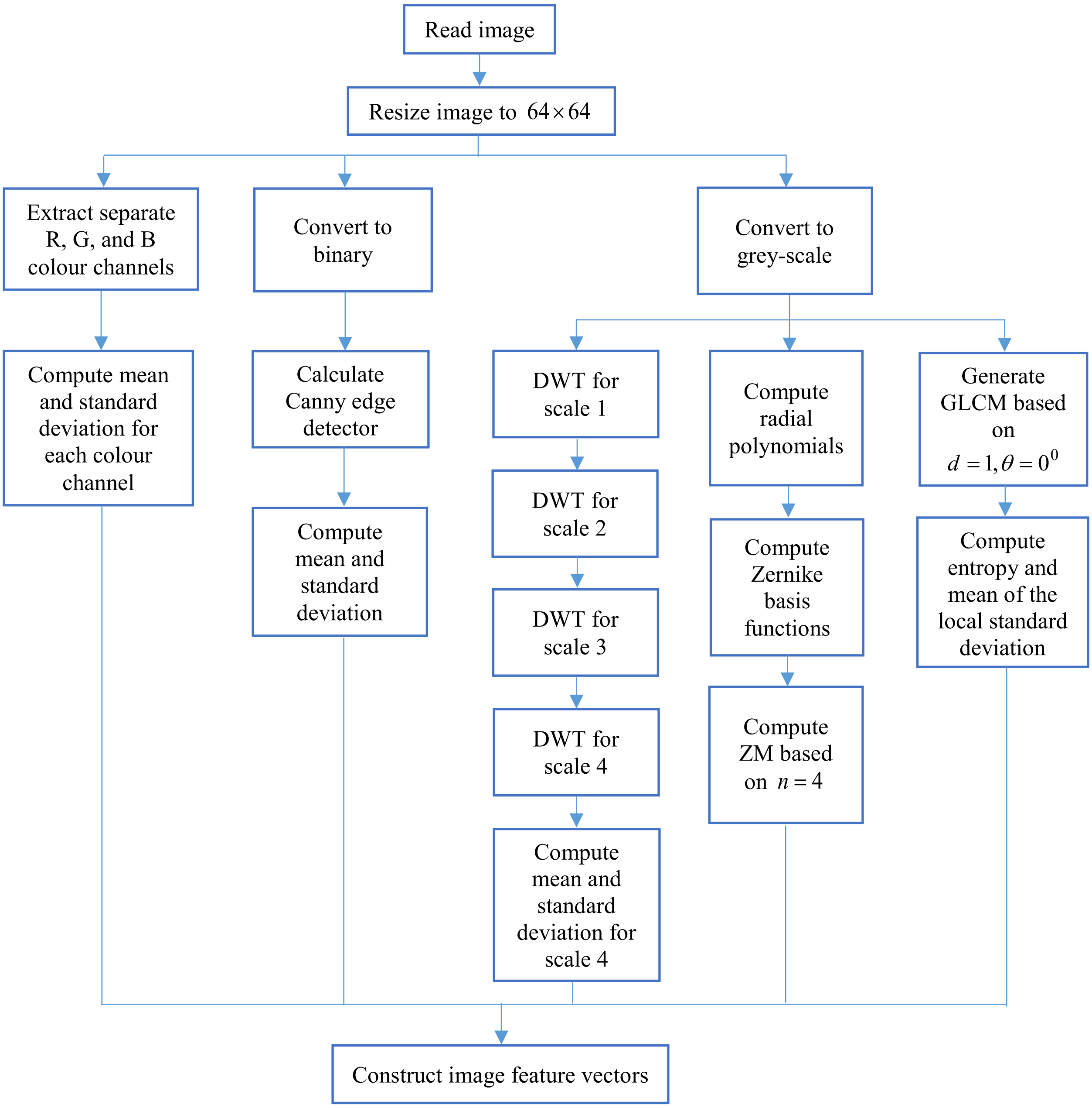

Fig. 1: Steps involved for the proposed content-based feature extraction method

Each of the components in Fig. 1 will be explained in the following sub-sections.

3.1 Image Pre-processing

First of all, for ease of computation and management, we have resized all images to a standard size of 64×64. This is because the images in the dataset comes in various sizes and some may be too large to be processed efficiently.

3.2 Colour Representation Based on Colour Moments

In order to extract colour information, separate *R*, *G*, and *B* channels are obtained. First order and second order colour moments are then computed on each of the channel where the mean μ and variance σ are calculated based on the Equations (1) and (2) below.

$$\mu_i = \frac{1}{N}\sum_{j=1}^{n} f_{ij}, \tag{1}$$

$$\sigma_i = \left(\frac{1}{N}\sum_{j=1}^{n} \left(f_{ij} - \mu_i\right)^2\right)^{1/2}, \tag{2}$$

where f_{ij} is the *i*-th colour component of pixel *j* and *N* is the number of pixels in an image. This will result in six feature vectors for colour representation.

3.3 Shape Representation Based on Discrete Wavelet Transform and Canny Edge Descriptor

Next, shape features are extracted based on two descriptors, namely Discrete Wavelet Transform (DWT) and Canny edge descriptor. DWT of a function can be generated from a mother wavelet ψ through dilation and translation processes. ψ is constructed from the scaling function ϕ, satisfying the two-scale difference equation as expressed in Equation (3) below.

$$\phi(t) = \sqrt{2}\sum_{k=-\infty}^{\infty} h_{(k)}\phi(2t-k), \tag{3}$$

where $h(k)$ is the scaling coefficient. The mother wavelet $\psi(t)$ is defined in Equation (4).

$$\psi(t) = \sqrt{2}\sum_{k=-\infty}^{\infty} g_{(k)}\phi(2t-k), \tag{4}$$

where the wavelet coefficients $g_{(k)} = (-1)^k h(1-k)$.

Therefore, wavelet decomposes a signal into wavelet function $\psi_{m,n}(t)$, where *m* is the scale or dilation index and *n* is the time or space index. There are various wavelet filters which can be utilised as the mother wavelet for the DWT computation. Some of the well-known mother wavelets are Haar, Daubechies, Coiflet, and Symmlet. For this work, we have chosen Coiflet 1 due to the reason that it uses six scaling and wavelet function coefficients. This will increase pixel averaging and differencing which produce a smoother wavelet and provide better image representation. Before computing the DWT, the image will be converted to normalised grey-scale values. The DWT is then implemented up to four sub-bands based on the chosen mother wavelet. According to [25], some decomposition levels perform better than others. Different levels of sub-band provide different details of information related to the image. Coarse resolution or higher frequency sub-bands generally describe the coarse information of an image. They are less invariant to noise and accumulation errors. Finer resolution or lower frequency sub-bands on the other hand normally define image's details. Due to the noise introduced in the original image and errors accumulated in the process of computation, the information are becoming less important. Plus it also has lost important information of the original image. Authors of [25] stated that the recognition rates obtained by them tend to decrease as the level of sub-bands increases. Due to these reasons, the intermediate sub-bands are usually more favourable as they usually contain the balance criteria between these two ends. Therefore, only sub-band four of the wavelet decomposition levels is utilised in this work for further processing. Once the DWT features have been obtained, the mean and standard deviation are computed which produced the final eight feature vectors for DWT.

Apart from DWT, the proposed method also consider the canny edge descriptor to represent shape. Canny edge detector is performed on a binary image where the mean and standard deviation are then captured. Canny is utilised instead of other edge detection methods because we found that it can detect edges better especially in noisy state.

3.4 Texture Representation Based on Zernike Moments and Grey Level Co-occurrence Matrix

As for texture representation, we have computed the statistical-based Grey Level Co-occurence Matrix (GLCM) and spectral-based Zernike Moments (ZM) on each of the normalised grey-scale image. Let *C* be the GLCM, so $C_{a,d}(i,j)$ will be the co-occurrence of pixels with grey values *i* and *j* at a given distance *d* and in a given direction θ. For this work, we have considered the common pair of these parameters, which are $d = 1$ and $\theta = 0^0$. Other (d, θ) pair combinations could also be used but more *d* and θ will lead to more GLCM being generated, hence leading to larger feature vectors and higher computation time. Out of the chosen (d, θ) pair, one GLCM is obtained. Haralick [24] introduced few statistical features like contrast, uniformity, mean, variance, inertia moments, etc, to represent texture. For this work, the entropy and mean of the local standard deviation are then calculated on the generated GLCM where this will result in nine feature vectors. Standard deviation is calculated using the standard formula while Entropy for the GLCM is computed using the following Equation (5).

$$Entropy = \sum_i \sum_j P(i,j) \log P(i,j), \tag{5}$$

where $P(i,j)$ is referring to the GLCM matrix value at location (i,j).

For ZM, first of all the radial polynomials will need to be calculated followed by the computation of the Zernike basis functions. The ZM features can then be constructed by projecting the image onto the basis functions. The discrete form of ZM for an image of size $N \times N$ is shown in Equation (6).

$$Z_{nm} = \frac{(p+1)}{\lambda_N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) R_{nm}(x,y) e^{-jm\theta_{xy}}, \tag{6}$$

where *n* is a non negative integer representing the order of the radial polynomial, *p* is a positive or negative integer satisfying that $n \geq 0, 0 \leq m \leq n$, and $n - |m|$ are even, $j = \sqrt{-1}$, $|r| \leq 1$, $0 \leq r_{xy} \leq 1$, and normalisation factor λ_N must be the number of pixels located in the unit circle π in the continuous domain. The transform distance r_{xy} and the phase θ_{xy}, at the pixel (x, y) are calculated based on the respective Equations (7) and (8).

$$r_{xy} = \frac{\sqrt{(2x - N + 1)^2 + (2y - N + 1)^2}}{N} \tag{7}$$

$$\theta_{xy} = \tan^{-1}\left(\frac{N - 1 - 2x}{2 - N + 1}\right) \tag{8}$$

It is crucial to determine the most suitable order *n* and repetition *m* when applying ZM. Order of ZM refers to the number of concentric circular division and degree of repetition will determine the number of circular sectors. Therefore, different order used will result in different repetitions to be computed, hence leading to different image representation. Due to this reason, we have conducted an experiment to determine the most suitable order value to be adopted. Out of conducting an empirical study, it has been found that $n = 4$ which generate nine order moments ZM (0, 0), ZM (1, 1), ZM (2, 0), ZM (2, 2), ZM (3, 1), ZM (3, 3), ZM (4, 0), ZM (4, 2), and ZM (4, 4) are the most suitable value for this work. These values are utilised to obtain the magnitude and angle of the ZM which are then utilised as the feature vector.

3.5 Feature Fusion Vectors

All of these generated feature vectors will then be combined to represent an image. To calculate the distance between an image with all images in the database, Manhattan distance function is used. According to this distance metric, the distance (or dissimilarity), d_{ij} between two features, x_i and x_j at position k can be calculated according to the Equation (9) below.

$$d_{ij} = \sum_{k=1}^{n} \left| x_{ik} - x_{jk} \right| \tag{9}$$

4.0 EXPERIMENTAL SETUP

This research is conducted based on the quantitative experimental approach. According to Leedy and Ormrod [26], experimental study is done to investigate the possible influences that one factor or condition may have on another factor or condition. This approach is suitable for this study since the main focus of this research is to examine the possible influences that the proposed method have in increasing the retrieval effectiveness compared to the existing comparable methods.

There are two experiments being conducted for this work. Experiment one is to determine the most suitable value for the order of Zernike Moment. Different order will lead to different number of repetitions being executed for the ZM. It is very important to determine the number of order so that images will be represented well according to the optimum number of repitition. Second experiment is to determine the retrieval effectiveness of the proposed feature fusion method on the marine life image dataset compared to the work done in [10] and few other benchmark methods.

4.1 Hardware and Software Specifications

A prototype CBIR system has been implemented to evaluate the retrieval effectiveness of the proposed method versus the comparable methods. All of the experiments mentioned in this work have been conducted on a desktop running on Microsoft Windows XP Professional operating system, with Pentium® Dual-Core E5200 2.50GHz, and 4 GB Random Access Memory. MATLAB has been used as the programming language for development where it enables the complex and computationally intensive tasks to be performed faster.

4.2 Dataset Collection

There are two image datasets utilised for this work. For the first experiment in determining the optimum order for ZM method, a dataset consists of 50 marine invertebrates is used. It has five image categories which include sea anemone

(images 1 - 10), sea urchin (images 11 - 20), jellyfish (images 21 - 30), sea cucumber (images 31 - 40), and starfish (images 41 - 50). There are 10 images to a category. All images are used as query. This dataset is similar to the one used for our earlier works mentioned in [10]. The ground truth of this dataset is based on the 10 similar images within a class. Samples of the mentioned dataset can be found in Fig. 2.

Fig. 2: Samples of the marine invertebrate dataset

In order to evaluate the retrieval effectiveness of the proposed feature fusion method, a larger image collection is used. For this purpose, natural fish images from Fish4Knowledge has been utilised where it contains 23 fish image categories with 27370 fish images in total. Each category contains images of a fish species which have been verified by marine biologist experts [27]. This will provide the ground truth. The number of images for each category varies from one to another and they vary significantly in terms of colour, shape and texture. Since the dataset is extremely large, we were not able to run and process all of the images due to time limitation. Therefore, a subset of this dataset is utilised for our experiment. Approximately 5% of images from each of the 23 category are randomly chosen to act as the dataset as well as the queries, which is equivalent to 1446 images. The dataset can be downloaded from http://groups.inf.ed.ac.uk/f4k/GROUNDTRUTH/RECOG/. Fig. 3 shows few samples of the Fish4Knolwedge fish image dataset.

Fig. 3: Samples from Fish4Knowledge image dataset [26]

4.3 Evaluation Measurement

The proposed method needs to be evaluated in order to measure its effectiveness. In this work, the retrieval effectiveness of the proposed method will be compared with the benchmark methods using various retrieval effectiveness measurements explained below. These measurements are known for evaluating the retrieval performance of methods in the CBIR field.

The precision and recall measures are one of the widely used to evaluate retrieval effectiveness due to its simple calculations and results obtained are easily interpreted. Apart from that, the results obtained from these measurements are usually visualised through the graph representations, which makes it easier to analyse. A precision rate, $P(Q)$ can be defined as the number of retrieved images similar to the query image, $Y(Q)$ among the total number of retrieved images, N. The equation is as shown in Equation (10) below.

$$P(Q) = \frac{Y(Q)}{N} \tag{10}$$

A recall rate, $R(Q)$ is defined as the number of retrieved images, which are similar to the query image, $Y(Q)$ among the total number of images similar to the query in the database, A. The equation is as shown in Equation (11) below.

$$R(Q) = \frac{Y(Q)}{A} \tag{11}$$

The precision and recall measures only evaluate the quality of an unordered set of retrieved documents. To evaluate ranked lists, the precision at rank *rnk*, $P_{rnk}(Q)$ and the recall at rank *rnk*, $R_{rnk}(Q)$, can be calculated based on the following Equations (12) and (13) below.

$$P_{rnk}(Q) = \frac{Y_{rnk}(Q)}{N_{rnk}} \tag{12}$$

$$R_{rnk}(Q) = \frac{Y_{rnk}(Q)}{A}, \tag{13}$$

where Y_{rnk} *(Q)* is the number of retrieved images similar to the query image at rank *rnk*, N_{rnk} is the total number of retrieved images at rank *rnk*, and *A* is the total number of images similar to the query in the database. To facilitate the computation of the average performance over a set of queries where each may retrieve different number of relevant images, the precision rate is usually interpolated against the 11 standard recall levels $(0.0, 0.1, 0.2, ..., 1.0)$ for standardisation. This representation is known as the 11 standard precision-recall and can be obtained based on the following Equation (14) below.

$$P(r_j) = \max_{rnk_j \le rnk \le rnk_j + 1} P(rnk), \tag{14}$$

where *j* is the 11 standard recall levels $(0.0, 0.1, 0.2, ..., 1.0)$.

Average precision, *AvgPrecise (Q)* on the other hand is calculated by taking into consideration the average of precisions for each image that are relevant to the query image *Q*, as shown in Equation (15) below.

$$AvgPrecise(Q) = \left(\sum_{i=1}^{A} \frac{I_i}{rank\ (I_i)} \right) / A, \tag{15}$$

where *A* is the number of images similar to the query in the database and *rank* (I_i) is the rank of I_i among *A* relevant images.

AvgPrecise (Q) are always in the range of 0 to 1. A higher value represents a better performance in terms of retrieval rate. However, for the 11 standard precision-recall, in practice, a method with a high overall precision will most probably have a low overall recall. This is because while trying to retrieve all relevant items to a query, some irrelevant items are also retrieved. Due to this, the precision and recall often trade off against each other.

The average of each retrieval effectiveness measurement is then calculated. This includes finding the average of the image queries within an image class or among images in the image dataset. The Mean Average Precision (MAP) is defined in Equation (16) below.

$$MAP_r = \frac{1}{N_q} \sum_{k=1}^{Nq} AvgPrecise(I_{k_r}) \tag{16}$$

where N_q represents the number of query images, I_k represents the *k* query image, and I_{k_r} represents the recall level *r* for the *k* query image. In order to find the average of the query images within the image class, N_q represents the total number of query images in the specified class. For a global averaging, N_q represents the total number of query images in the image dataset.

Two-tailed paired *t*-test is a statistical test used to determine whether the sample falls inside or outside a certain range (critical area) of values. The critical area is a two sided area below the curve graph. If the tested sample falls in between of the critical area, then the hypothesis is accepted, otherwise it is labelled as null hypothesis. For this research, the significant level used for the hypothesis to be rejected is 5%. A method is considered to be significant if the *p*-value is less than the significant value. Larger *t*-values translate into smaller *p*-values as well as indicating larger difference between the two approaches. In other words, there should be a difference in retrieval performance between both, proposed and benchmark method. Equations (17) – (20) are used to find *t* and *p* value.

Mean of the sample:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i, \tag{17}$$

where N is the sample size and x_i is the data of the sample.

Standard deviation of the sample:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}, \tag{18}$$

where N is the sample size, x_i is the data of the sample and μ is the mean of the sample.

t-value:

$$t = \frac{\overline{\mu_1} - \overline{\mu_2}}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \tag{19}$$

Where $\overline{\mu_1}$ and $\overline{\mu_2}$ is the mean, σ_1^2 and σ_1^2 is the standard deviation, and N_1 and N_2 is the sample size of the first and second methods to be compared.

Degree of freedom:

$$v = \frac{\left(\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}\right)^2}{\frac{1}{N_1 - 1}\left(\frac{\sigma_1^2}{N_1}\right)^2 + \frac{1}{N_2 - 1}\left(\frac{\sigma_2^2}{N_2}\right)^2}, \tag{20}$$

where σ_1^2 and σ_1^2 is the standard deviation, and N_1 and N_2 is the sample size of the first and second methods to be compared. The *t*-value and degree of freedom value can be used to find *p*-value from a statistical table (*t* distribution).

5.0 RESULTS AND DISCUSSION

Two retrieval experiments employing the Query by-Example (QBE) paradigm are conducted for this work. The first experiment is to explore on the most suitable value of order n for the ZM. The second experiment is conducted to evaluate the retrieval effectiveness of the proposed content-based feature fusion method in comparison to the benchmark methods.

5.1 Experiment on Determining the Order of Zernike Moment

Few possible values for the ZM's order are experimented. We tested on the order $n = 0$ up to the order of $n = 10$. We conducted the evaluation on 50 marine invertebrate images which have been used in our previous experiment on marine life [10]. All 50 images are used as query. Fig. 5 shows the outcome of the experiment based on the precision at 11 standard recall levels measurement while Table 1 provides the experimental result based on the Mean Average Precision (MAP) measurement . For Fig. 5, the x-axis of the graph represents the 11 standard recall levels while the y-axis represents the average precision values at the 11 standard recall levels. From this figure, it can be observed that the most suitable value of n is 4. This will generate nine order moments ZM (0, 0), ZM (1, 1), ZM (2, 0), ZM (2, 2), ZM (3, 1), ZM (3, 3), ZM (4, 0), ZM (4, 2), and ZM (4, 4) as the ZM's feature vector. From the figure it can also be observed that the order of 4 is able to consistently retrieve the correct images at a much higher rank compared to the other orders.

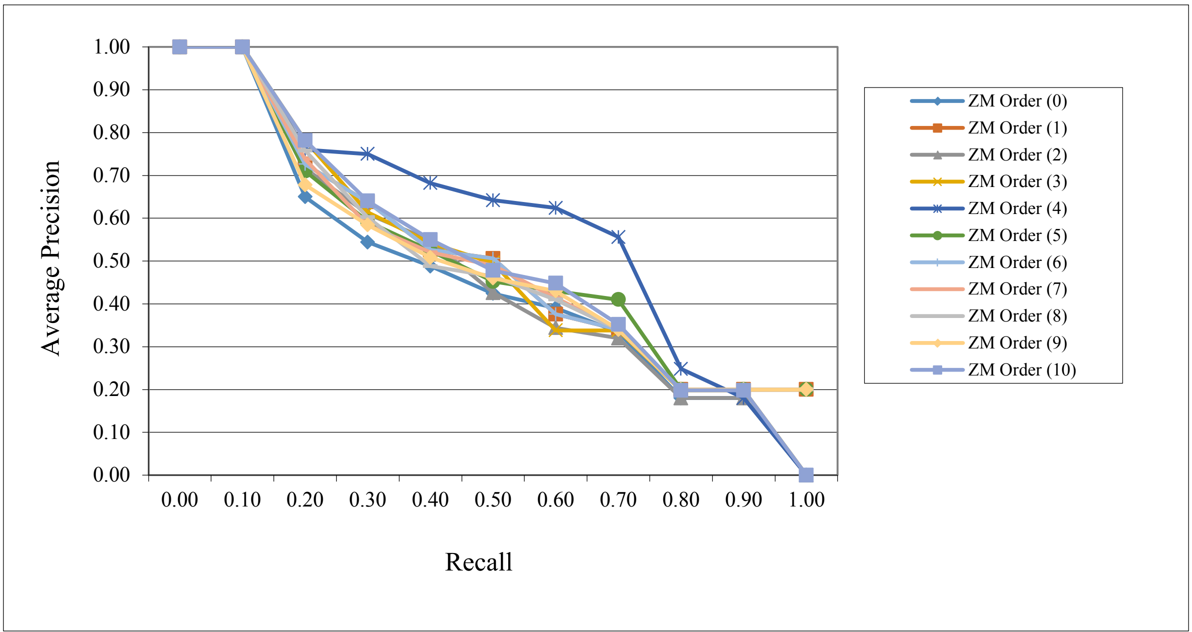

Fig. 5: 11 standard precision-recall graph representation in determining the order of ZM based on marine invertebrate images

From Table 1, it can be observed that the MAP values starting to increase from the order n = 0 up to n = 3 and achieve its optimum value when n = 4. The MAP values however start to decrease slowly (or generally becoming more consistent) starting from n = 5 and above. Since we do not see that the MAP values is improving starting from n = 5 and more orders will only result in higher computation time and feature size, hence the experiment is stopped at the value of n = 10. Lower order moments are said to be more effective for image representation and it is more robust to noise in comparison to higher order moments and it defines the global information of the image. Higher order moments on the other hand describe the fine structures of an image and making it more vulnerable to image noise since errors may be accumulated in the process of computation . Therefore, the intermediate moments can be said to be more favourable as it represents the best part of image's information between the two ends. Hence, it can be concluded that the order of n = 4 is the best value for the ZM in this work which will be utilised as part of the proposed feature fusion method.

Table 1: MAP of Zernike Moments with different order n

Order of Zernike Moments, *n*	**MAP**
0	0.41188
1	0.42732
2	0.42912
3	0.44703
4	0.47748
5	0.46900
6	0.46897
7	0.46590
8	0.46314
9	0.45940
10	0.45223

5.2 Retrieval Effectiveness Experiment on Fish4Knowledge Image Dataset

For this particular experiment, larger dataset is used in comparison to the previous experiment. Approximately 5% (1446) of the 27370 Fish4Knolwedge images comprise of 23 fish species are randomly selected for evaluation and all of these images also act as the query. The proposed feature fusion method is compared to four other methods which are colour-only (Colour Moment), shape-only (DWT and Canny edge detector), texture-only (GLCM and ZM of order 4), and fusion of colour and shape features. For each query and for each benchmark method, the precision at 11 standard recall levels and its MAP are recorded. Fig. 6 displays the result based on the first retrieval perfomance measurement while average MAP value obtained for each of the 23 fish species are tabulated in Table 2.

From Fig. 6, it can be observed that the proposed feature fusion method confirmed its validity in providing effective retrieval in comparison to the other four benchmark methods. The proposed feature fusion method not only able to obtain higher overall precision rate but it has also achieved higher precision rate at all of the 11 standard recall levels. Although the Fish4Knowledge images generally have complex background with multiple objects within the image that come in different scales, position, and direction, the proposed method is still superior and able to consistently retrieve the relevant images at higher rank. This is due to the different features being extracted by the respective proposed methods which allow for the image to be accurately represented. Colour Moments represents the spatial distribution of colours, DWT and Canny edge descriptors allow for the respective region-based and contour-based details to be represented, GLCM defines the spatial pattern distribution at various distances and orientations, and ZM extracted the texture information at multi-level in the spectral domain. All of these chracteristics that actually translated the proposed colour-shape-texture method to become effective. Colour-only and texture-only descriptors came in second where they performed almost equally, followed by the texture-only method. Shape-only method reported to perform the worst. The Fish4Knowledge dataset contain images of fishes with complex background, hence colour and texture plays an important role in this situation. Shape feature on the other hand may be less useful due to the almost-similar shape of fishes found in most of the categories.

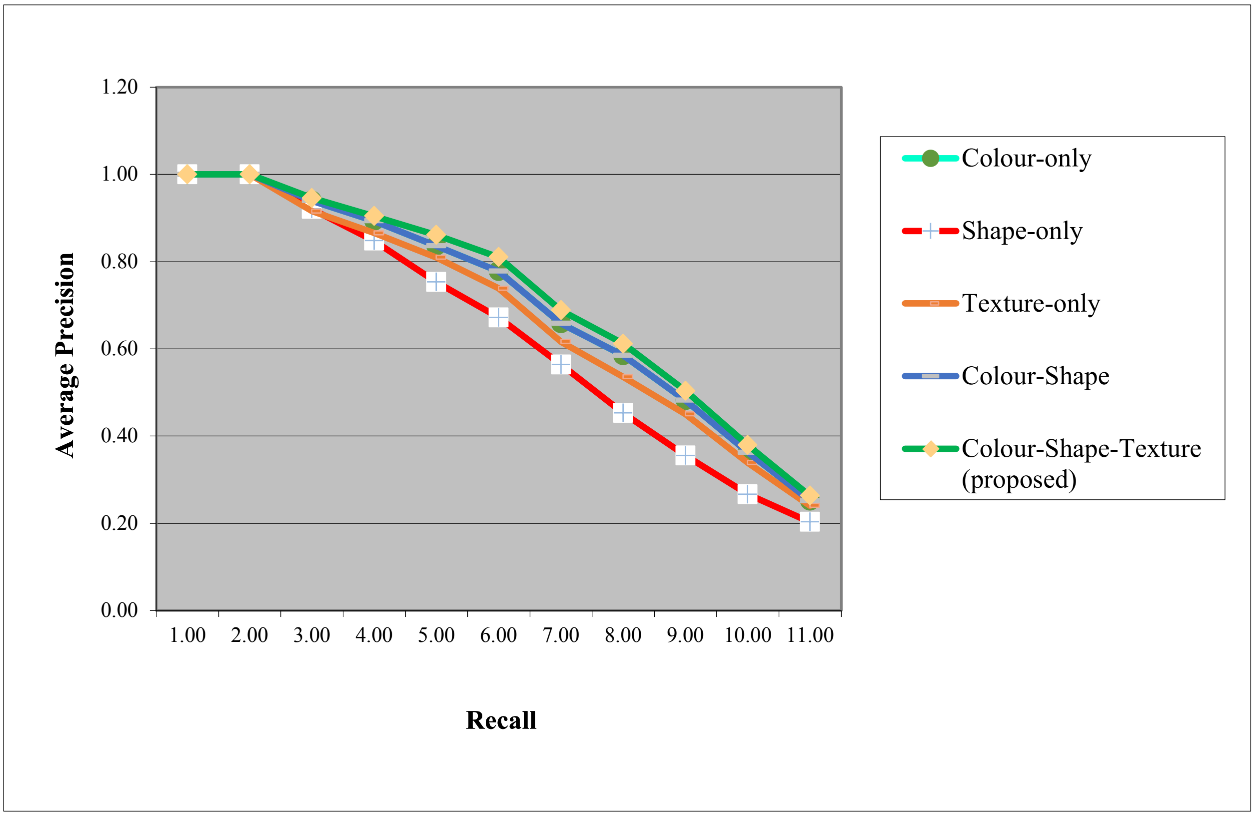

Fig. 6: 11 standard precision-recall graph representation for the proposed and benchmark methods based on Fish4Knowledge image dataset

Table 2 tabulates the retrieval results for 23 fish species of Fish4Knowledge image dataset based on MAP retrieval effectiveness measurement. For each different fish species, the retrieval value of the method achieving better result than the rest are put in bold. From Table 2, we can see that the proposed method achieved the highest average MAP value compared to the benchmark methods (67.70%). Out of the 23 fish species, the proposed method has also successfully obtained the highest MAP value for more than 50% of the categories and mostly marked a precision rate of 70% or above.

Based on the two different retrieval effectiveness measurements, it can be observed that the proposed method showed to be the most effective in comparison to the benchmark methods. Therefore, it can be concluded that the proposed method has proven its stability as well as a consistent performance.

Table 2: MAP value for each species and overall system of the proposed and benchmark methods

Species	Colour-only	Shape-only	Texture-only	Colour-Shape [10]	Colour-Shape-Texture (Proposed)
1	0.93709	0.90827	0.86833	0.93990	**0.94536**
2	0.79562	0.61098	0.58407	0.81368	**0.87506**
3	0.86528	0.74779	0.72848	0.91160	**0.91607**
4	0.90888	0.87511	0.87801	0.89350	**0.92458**
5	**0.93901**	0.90782	0.91799	0.90280	0.93764
6	0.51228	0.39569	0.54473	0.59953	**0.68213**
7	0.73335	0.51162	0.63130	0.78350	**0.84128**
8	0.66136	0.62644	**0.79654**	0.70318	0.73466
9	0.30515	0.21635	0.34125	0.51207	**0.52335**
10	0.75637	0.76633	**0.94309**	0.38704	0.59475
11	0.47773	0.25737	0.30173	0.87600	**0.89875**
12	**1.00000**	**1.00000**	0.92000	0.97578	**1.00000**
13	0.67588	0.46907	0.44681	0.68769	**0.73481**
14	0.42858	0.35988	0.52185	0.54931	**0.55763**
15	0.46000	0.42790	0.44550	**0.62906**	0.59748
16	0.60759	0.57367	**0.70088**	0.48262	0.56426
17	0.44400	0.41433	0.48598	0.49333	**0.50000**
18	0.46865	0.26752	0.35549	0.69304	**0.70431**
19	0.22250	0.17161	0.13833	**0.30837**	0.29170
20	0.27394	0.23233	**0.36871**	0.31011	0.36171
21	**0.50357**	0.48430	0.30994	0.38933	0.35290
22	0.57218	0.43996	0.56095	**0.78104**	0.75504
23	0.28419	0.22040	0.20000	**0.30632**	0.27749
MAP	**0.60144**	**0.51673**	**0.56478**	**0.64908**	**0.67700**

To investigate the significance of the retrieval effectiveness improvement contributed by the proposed method, a statistical comparison based on the two-tailed paired *t*-test is done. Let $Y_e, e \in \{\{R\},\{RF\},\{NRF\},\{ENRF\}\}$ be the mean average precisions (retrieval accuracies) for each method. The null hypothesis $H_0 : Y_e > 0.05$ and the alternative hypothesis is $H_1 : Y_e < 0.05$. The larger *t*-values is the more likely that the difference is significant. The statistical test result is as shown in Table 3 below. From the results, we can see that the proposed feature fusion method of colour, shape, and texture features shows a statistically significant improvement in retrieval performance when compared to few existing works.

Table 3: Paired *t*-test at a significance level of 0.05. *p*-values less than 0.05 indicate significantly different results

Method	***t*-value**	***p*-value**	**Null Hypothesis (H_0)**	**Remark**
Colour-only vs. Proposed method	2.7657	8.2701E-03	Reject	**Proposed method > Colour-only**
Shape-only vs. Proposed method	3.8568	3.7100E-04	Reject	**Proposed method > Shape-only**
Texture-only vs. Proposed method	2.7947	7.6671E-03	Reject	**Proposed method > Texture-only**
[10] vs. Proposed method	2.4743	1.7278E-02	Reject	**Proposed method > [10]**

6.0 CONCLUSIONS AND FUTURE WORKS

CBIR specifically for marine animals needs further investigations. This is because marine species are diverse and rich in features, where they come in various colours, shapes, sizes, and textures. Natural marine life images are usually captured under the conditions of complex background, at different angle, position, and size, and these require an effective and all-rounder approach to be used for representation. Inspired by these challenges, this research proposed a content-based feature extraction method specifically for marine life images. For this purpose, few features such as colour, shape, and texture are considered and an integrated approach has been constructed. The improved technique is compared to that of colour-only, shape-only, texture-only, and colour-shape descriptors [10] using Fish4knolwedge image dataset. Three measurements which include precision-recall curve, MAP, and two-tailed paired *t*-test statistical method are used. Outcomes of the conducted experiments pointed out that the proposed method achieved significantly improved results than the benchmark methods for all performance measurements.

For future works, a classifier may be considered to further improve the method. Feature selection could also be conducted to reduce the feature vector size. The proposed method also involved quite a number of approaches where for each of the approach, the most optimum value for certain parameters may need to be prior determined. This can be achieved by conducting experiments for each of the parameter involved to further improve the method.

ACKNOWLEDGEMENT

This work is supported in part by Ministry of Higher Education (MoHE) under the Fundamental Research Grant Scheme (FRGS) project code 08-01-16-1855FR.

REFERENCES

[1] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*, 2nd ed., New York: Addison-Wesley Longman Limited, 2011.

[2] A. Alzu'bi, A. Amira, and N. Ramzan, "Semantic content-based image retrieval: A comprehensive study," *Journal of Visual Communication and Image Representation*, vol. 32, pp. 20-54, 2015.

[3] C. Spampinato, S. Palazzo, B. Boom, J. Van Ossenbruggen, I. Kavasidis, R. Di Salvo, ... and R. B. Fisher, "Understanding fish behaviour during typhoon events in real-life underwater environments," *Multimedia Tools and Applications*, vol. 70, no. 1, pp. 199-236, 2014.

[4] N. S. Osman and M. R. Mustaffa, "A review on content-based image retrieval representation and description for fish," in *Proceedings of the 4th International Conference on Advanced Computer Science Applications and Technologies (ACSAT)*, 2015, pp. 39-44.

[5] R. Reig-Bolaño, P. Marti-Puig, E. Gallego, G. Masferrer, A. Lombarte, L. Ferrer-Arnau, and V. Parisi-Baradad, "Feature selection for analysing and retrieving fish otoliths using Elliptic Fourier Descriptors of shapes," in *Proceedings of the 7th International Conference on Next Generation Web Services Practices (NWeSP)*, 2011, pp. 290-295.

[6] J. Hu, D. Li, Q. Duan, Y. Han, G. Chen, and X. Si, "Fish species classification by colour, texture and multi-class support vector machine using computer vision," *Computers and Electronics in Agriculture*, vol. 88, pp. 133-140, 2012.

[7] H. Yao, Q. Duan, D. Li, and J. Wang, "An improved *k*-means clustering algorithm for fish image segmentation," *Mathematical & Computer Modelling*, vol. 58, no. 3, pp. 790-798, 2013.

[8] M. Ghosh, J. Mukherjee, and R. Parekh, "Fish shape recognition using multiple shape descriptors," *International Journal of Computer Applications*, vol. 73, no. 16, pp. 14-19, 2013.

[9] A. R. Sheikh, S. Mansor, and M. F. Ahmad Fauzi, "An automated content-based image retrieval system for marine life images," *International Journal of Advances in Computer Science and its Applications*, vol. 4, no. 1, pp. 202-206, 2014.

[10] M. R. Mustaffa, N. M. Norowi, and S. M. Yee, "Content-based image retrieval system for marine invertebrates," in *Proceedings of the 3rd International Conference on Information Retrieval and Knowledge Management (CAMP)*, 2016, pp. 71-76.

[11] N. A. B. Mary and D. Dharma, "Coral reef image classification employing improved ldp for feature extraction," *Journal of Visual Communication and Image Representation*, vol. 49, pp. 225-242, 2017.

[12] M. R. Mustaffa, F. Ahmad, R. W. Rahmat, and R. Mahmod, "Content-based image retrieval based on colour-spatial features," *Malaysian Journal of Computer Science*, vol. 21, no. 1, pp. 1-12, 2008.

[13] D. Srivastava, R. Wadhvani, and M. Gyanchandani, "A review: Colour feature extraction methods for content-based image retrieval," *International Journal of Computational Engineering & Management*, vol. 18, no. 3, pp. 9-13, 2015.

[14] K. Juneja, A. Verma, S. Goel, and S. Goel, "A survey on recent image indexing and retrieval techniques for low-level feature extraction in CBIR systems," in *Proceedings of the IEEE International Conference on Computational Intelligence & Communication Technology (CICT)*, 2015, pp. 67-72.

[15] M. R. Mustaffa, F. Ahmad, R. W. Rahmat, and R. Mahmod, "Dominant colour descriptor with spatial information for content-based image retrieval," in *Proceedings of the International Symposium on Information Technology 2008 (ITSIM '08)*, 2008, pp. 1612-1620.

[16] S. Agarwal, A. K. Verma, and N. Dixit, "Content-based image retrieval using colour edge detection and discrete wavelet transform," in *Proceedings of the International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, 2014, pp 368-372.

[17] E. Gupta and R. S. Kushwah, "Combination of global and local features using DWT with SVM for CBIR," in *Proceedings of the 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, 2015, pp. 1-6.

[18] S. Gandhani and S. Nandini, "Content-based image retrieval: Survey and comparison of CBIR system based on combined features," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 8, no. 10, pp. 155- 162, 2015.

[19] S. Selvarajah and S. R. Kodituwakku, "Analysis and comparison of texture features for content-based image retrieval," *International Journal of Latest Trends in Computing*, vol. 2, no. 1, pp. 108-113, 2011.

[20] A. Tahmasbi, F. Saki, and S. B. Shokouhi, "Classification of benign and malignant masses based on Zernike moments," *Computers in biology and medicine*, vol. 41, no. 8, pp. 726-735, 2011.

[21] X. Y. Wang, B. B. Zhang, and H. Y. Yang, "Content-based image retrieval by integrating colour and texture features," *Multimedia tools and applications*, vol. 68, no. 3, pp. 545-569, 2014.

[22] S. Sharma and P. Khanna, "Computer-aided diagnosis of malignant mammograms using Zernike moments and SVM," *Journal of digital imaging*, vol. 28, no. 1, pp. 77-90, 2015.

[23] J. K. Patil and R. Kumar, "Analysis of content-based image retrieval for plant leaf diseases using colour, shape and texture features," *Engineering in Agriculture, Environment and Food*, vol. 10, no. 2, pp. 69-78, 2017.

[24] R. M. Haralick and K. Shanmugam, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, vol. SMC-3, no. 6, pp. 610-621, 1973.

[25] O. R. Terrades and E. Valveny, "A new use of the Ridgelets transform for describing linear singularities in images," *Pattern Recognition Letters*, vol. 27, no. 6, pp. 587-596, 2006.

[26] P. D. Leedy and J. E. Ormrod. *Practical Research: Planning and Design*, 10th ed., Harlow: Pearson Education Limited, 2014.

[27] B. J. Boom, P. X. Huang, J. He, and R. B. Fisher, "Supporting ground-truth annotation of image datasets using clustering," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR)*, 2012, pp. 1542-1545.