Content-based Image Retrieval for Fish based on Extended Zernike Moments-Local Directional Pattern-Huecolor Space

Noorul Shuhadah Osman, Mas Rina Mustaffa, Shyamala C. Doraisamy, Hizmawati Madzin

Abstract: ***Scholars have been fascinated in the areas of the description and representation of fish species images so the Content-based Image Retrieval is adopted. Proposals have been made to use various techniques like the fusion of Zernike Moments (ZM) and Local Directional Pattern (LDP) to obtain good image representation and description results for feature extraction. To elaborate, ZM is characteristically rotation invariant and it is very robust in the extraction of the global shape feature and the LDP serves as the texture and local shape feature extractors. Nevertheless, extant studies on ZM-LDP fusion are only adopted for gray-level. The role of color is substantial for the fish. The proposal is that the ZM-LDP method is improved so that it can bring out the color features for the fish-domain effectively. By computing the LDP on the Hue plane of the HSV color space of the image, the color information is obtained. Improved ZM-LDP fusion to be able to obtain color information (Extended Zernike Moments-Local Directional Pattern-Hue Color Space) is experimented on Fish4Knowledge (natural image) dataset consists of 27370 images and able to achieve Mean Average Precision of 77.62%. Based on the experimental results, it is shown that the proposed method has successfully achieved higher accuracy compared to other comparable methods. A statistical comparison based on the Two-tailed paired t-test was carried out and has proven that the retrieval performance of the proposed method is improved.***

Keywords: Zernike moment, Local directional pattern, Fish, Content-based image retrieval, Color feature

I. INTRODUCTION

It is evidenced by a body of literature that a specific compact feature extraction method is needed to develop an effective and efficient Content-based Image Retrieval(CBIR) system for different types of domain or dataset images.

Lately, with the increasing application demands, underwater exploration has received more attention and has become one of the main topics of study, such as marine species [1-2]. Fish is one type of marine life that has been given attention by researchers from various fields[3-4]. It is because fish have important roles in human life. Fish are not only the source of food and livelihoods, but each fish species has different benefits as a source in the vitamins supplied and its application and function in medications. This makes people interested in studying more about the fish. Feature extraction is the most important phase in CBIR system, and the extraction of wrong feature might not accurately represent the image[5]. However, very little research has been conducted regarding feature extraction technique specifically for fish domain. Most studies that have been conducted are to use commercial fish images, where the images are of good quality (clear background and good illumination and captured in lab or under controlled environments, i.e.cameras, fixed background, lighting and objects in the water, and known number and types of fish). In addition, the datasets used for most of the previous studies are different from one another. Besides, most of the previous works are focusing on different aspects or phases of CBIR system (segmentation, feature extraction, semantic and fusion method). Each category of fish has different color, shape and texture features. Fish lives in natural environment, where color plays an important role to differentiate the fish. Due to this condition, it is important to have a specific feature extraction technique for the CBIR system for fish domain, which able to extract the color, texture and shape features of the fish [3, 6].

II. RELATED WORKS

In the early 80s, the marine fishery resources had experienced depreciation. This situation has raised awareness among the public to improve fisheries resources. Fish is one of the many marine lives that require conservation and protection for future generation. Fish have a wide range of benefits in everyday human life. It is not only a source of foods and livelihoods, but also because of its unique shape, color and texture, that makes it one of the favorite animals. The number of fish categories found is high and increasing, explaining why a proper documentation is vital. Often, professionals that involve in fish studies and conservation have been tasked with the complex and often difficult assignment of manually identifying aquatic species

Revised Manuscript Received on May 22, 2019.

Noorul Shuhadah Osman, Multimedia Department, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor Darul Ehsan. Malaysia

Mas Rina Mustaffa, Multimedia Department, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor Darul Ehsan. Malaysia

Shyamala C. Doraisamy, Multimedia Department, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor Darul Ehsan. Malaysia

Hizmawati Madzin, Multimedia Department, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor Darul Ehsan. Malaysia

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

and fish species when this can be done much faster and more effectively with the assistance of an automatic content-based representation and retrieval [7]. Thus, many researchers in the CBIR field have taken the initiatives to design representation and retrieval approaches specifically for fish domain. To always be informed about the fish population and to conduct a study on fish behavior is a very challenging thing for sea life scientist.

There are few fish CBIR that have been established in past studies. A system to recognize fish is developed by adopting the morphological and lexicon ontology services [8]. Besides the ability of the system to automatically import existing fish knowledge base into the system, this system also able to construct the new one for different type of dataset properly. The processing of the images includes discarding the noise, setting the direction, turning to grayscale, and sharpening the edge. The dataset used is a total of twelve species of fish images, which shows that even though two images have different colors, but the matching rate is high. However, the limitation of this study is that it only accounts for fish images with clean background and good illumination. The image is a photograph of only certain parts of the fish. The color feature is altogether ignored.

Multi-class Support Vector Machine is proposed by[9]. It is a classification method specifically for fish. The extraction of color features, statistical texture features and wavelet-based texture features (from color and texture sub images) will give six groups of feature vectors. The Hue-Saturation-Value (HSV) color features were extracted for the color feature, while Grayscale Histogram and Grey Level Co-occurrence Matrices method are adopted for the extraction of texture features. The dataset used are the very popular six freshwater fish images in China. When the color properties are considered, better retrieval accuracy is achieved. However, the dataset images used is a photographed of only a certain part of the fish. The shape feature is totally ignored in this study.

In [10], the authors suggested a fusion of mathematical morphology and *K*-means clustering segmentation algorithm to segment the fish image. To get the fish contour and body boundary, the mathematical morphology method is adopted, with the random natural fish images being the dataset. The proposed method is anticipated to address the difficulty to segment the fish image in complex background. That said, the method is tested with different values of *k* for different images, but it did not mention what value is the best for fish in a large database. This study did not bring into attention the color and texture feature.

Feature extraction is one of the significant phases in CBIR system. It is a process to obtain feature vectors of the image. The extraction of wrong features would be the precursor to imprecise retrieved images[11]. To better represent the fish image, a lot of studies have been done and the best approaches are explored so that the retrieval will be more precise. There are high-level (semantic) and low-level features (shapes, color and texture)[12]. There is always a different between the way human and computer interprets images (semantic gap). It happens when the user seeks to retrieve the images with the same low-level feature but with different meanings. Another question raised is how to narrow down the semantic gap. The visual information that can be obtained from the image content itself is categorized as low-level features, which are extracted by the system automatically. The obvious criteria to differentiate between fish species are through their color, shape and texture information.

Being widely used feature in CBIR, color is one of the most expressive visual features. The image size and orientation does not affect the performance of the retrieval system and relatively robust to background complications[13]. In image retrieval, color histogram is the most popular approach for color feature extraction.

Color Histogram (CH) has been utilized as the color feature extractor for the work in [14]. The algorithm works well and it is not difficult to implement. However, CH has a very large dimension and it does not take into account the color's spatial correlation.

The fusion of Haar Wavelet Transform (HWT) with Color Moments (CM) to extract texture and color features of the image is then proposed by [15]. Even though the dataset used is natural images with complex background, the fusion achieved higher retrieval accuracy compared to the other comparable methods. However, CM cannot depict all color information of the image and there is no spatial information recorded.

Local Binary Pattern (LBP) is proposed by [16]. It is computationally efficient and at the same time can perform well in monotonic illumination variation. To improve LBP descriptor to be invariant to illumination, combination of LBP with six different color descriptors- RGB-LBP, nRGB-LBP, Transformed Color LBP, Opponent-LBP, nOpponent-LBP and Hue-LBP are proposed [17]. From the list, Hue-LBP gives the best performance and it shows a strong property of invariant to illumination. However, LBP only discover a small neighborhood area which can only provide very limited local information of an image. It is also insensitive to random noise and non-monotonic change in illumination.

To improve LBP, Local Directional Pattern (LDP) is proposed [18]. Both LBP and LDP are originally developed as texture and local contour-based shape feature extractors. But the way how LDP calculates the image feature makes it not susceptible to noise and non-monotonic illumination variations. However, the drawback is, by computing LDP directly on a full image, the information extracted might not be enough to best represent the local features of the image. The improvement of LDP method is then proposed by [19]. Instead of full image, LDP is applied on image patches. They also proposed the adoption of a shape-encoded image, which employed DT-CWT as the edge detection technique. Specifically, in this work, LDP descriptor is improved with additional weight patches (pLDPw) and wavelet-based shape-encoded images with LDP-weight patches (pWLDPw). The retrieval accuracy increased as opposed to the traditional LDP. However, it is not able to extract color feature.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Zernike Moments (ZM) is rotation invariant and very powerful in extracting global shape feature. In research done by [19], the performance of ZM only, the fusion of ZM with proposed pLDPw and pWLDPw are investigated. The methods are tested on several types of databases, and compared with a lot of other methods available. For the databases considered, pWLDPw method proved itself to be the best compared to other tested methods. However, the limitation is all dataset used are simple grayscale images with clear edges, where the color feature is not taken into account.

The extension of LDP operator proposed by [19], by allowing the extraction of color information is thought to be timely. Based on the literature review above, it can be concluded that most of the datasets used for evaluation are not of natural fish images, and previous works mostly did not consider the image's color feature. Fish lives in natural habitat and is always on the move, where this makes it almost impossible to properly capture a clear image of a fish. Therefore, there is a need for a method that integrates both global and local shape features, and allow for the representation of color information so that the accuracy of the system can be increased.

III. EZLH AS COLOR DESCRIPTOR

The proposed improved ZM-LDP fusion is named as EZLH. EZLH is motivated from the research work by [19]. The ZM and LDP were fused by [19] to extract local and global shape features as well as texture feature of the images and higher retrieval result was obtained. The proposed extension to include color representation that will be done to the ZM-LDP fusion is believed to generate higher retrieval result. For color feature vector extraction process, the database images are first converted to HSV color space. Then, the Hue histogram of the images is generated. LDP is then computed on the generated histogram. The extracted feature vectors are then saved as HLDP feature vectors. For the local and global shape features and texture feature vector extraction processes, all database images are first converted to grayscale images. The application of ZM and LDP on the grayscale images is executed simultaneously and the extracted vectors are saved as ZM feature vectors and LDP feature vectors. All saved feature vectors are saved in feature database. The processes are performed offline.

For online process, for each selected query image, the image will be imposed into the same feature extraction process. All of the query feature vectors extracted during feature extraction process are then stored. The distance between the color feature vector of the query image and all database images stored in database are computed by the system using the Euclidean Distance which is widely used in computing similarity for feature extraction [20-21].Smaller distance value is representative of the most similar database image to the query image. The most similar images will be displayed and sorted from top left to bottom right of the retrieval window. Fig. 1shows the framework for EZLH and the explanation for each of the components in the figure are detailed out in Sections 3.1 to 3.7.

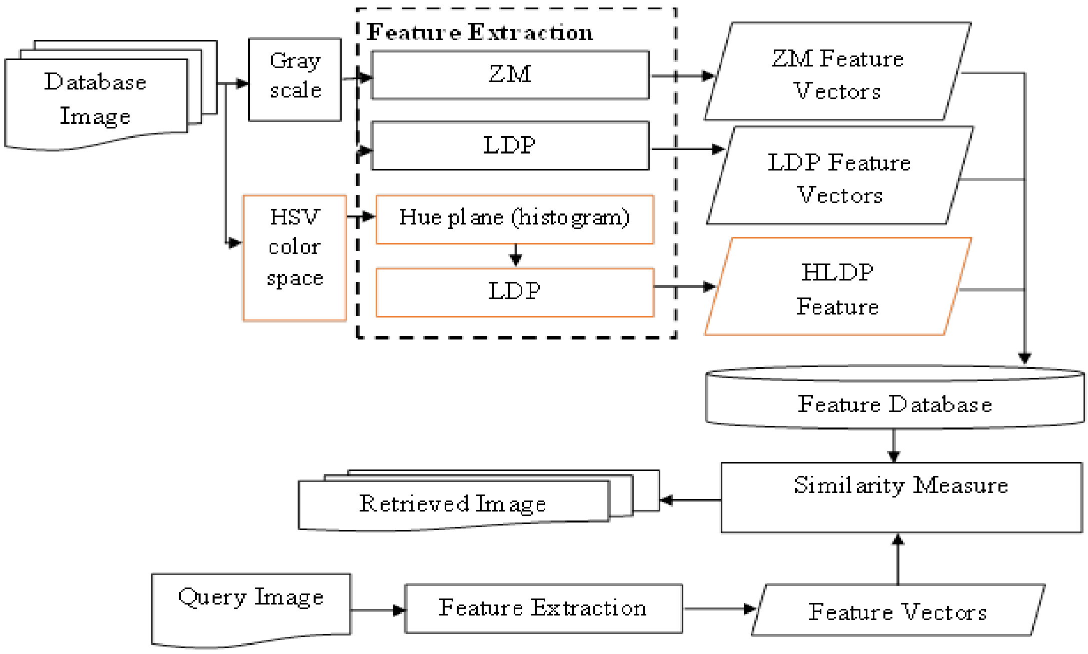


Fig. 1 CBIR System Framework for the EZLH

Query Image/ Database Image

All of the 27370 Fish4knowledge natural images are utilized as the dataset for this work. 30% of the total images are then utilized as query.

Since different image size will bear different information and will affect the accuracy of the image retrieval system directly, the image size has to be studied first. An experiment to find the suitable image size is first carried out.

EZLH framework is used for this experiment.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The system's performance is calculated based on the top 10 retrieved and is measured by Mean Average Precision (MAP). Since the images range from 20 × 20 pixels to 200 × 200 pixels, the size 32 × 32 pixels, 64 × 64 pixels, 128 × 128 pixels and 256 × 256 pixels are considered good for testing. These sizes are commonly used in computer vision and it is not too small or too large from the original image, hence the information loss is reduced. From the experiment conducted, it is found that the most appropriate image size is 64 × 64 pixels. The details of the result can be found in Section 5.

Grayscale Color Space

Since the information regarding color is not required in obtaining the shape and texture feature, therefore the grayscale color space appears to be the best choice. It gives only single intensity value for each pixel, which makes the computation is less complicated and subsequently reducing the computation time.

HSV Color Space

Having converted the Red-Green-Blue (RGB) image to HSV color space, the histogram of its Hue plane is extracted. Due to its stability, HSV color space is chosen because of its popularity in many fields [22]. The color is described better and it does not deviate much from how people experience color. Hue is used for color discrimination. Saturation serves to measure how much white color is contained in the pure color, while value gauges how dark the color is. Therefore, Hue plane is selected. Before the histogram of the Hue plane of the image is extracted, the RGB image is first resized to 64 × 64 pixels and after that it is converted to HSV color space. Then, the color information of the image is encoded by applying LDP on the extracted Hue plane of the image. Hue plane is extracted from the resized image based on the Eq. 1[17] below:

$$Hue = arctan\left(\frac{O_1}{O_2}\right) = \arctan\left(\frac{\sqrt{3}\,(R-G)}{R+G-2B}\right) \quad (1)$$

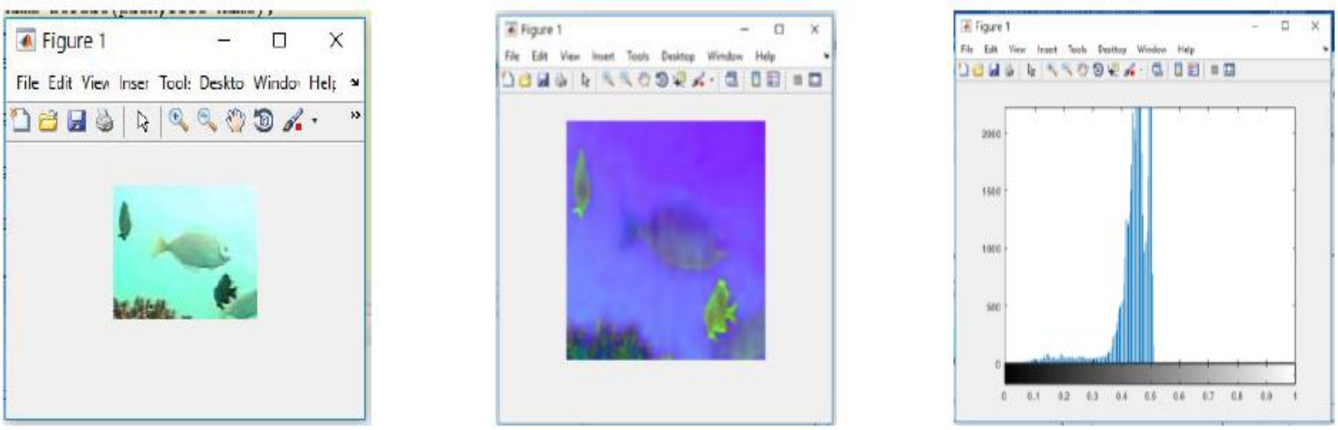
A histogram with 64 × 64 feature vectors is then saved for the next process. Fig. 2 leads us to the sample of an RGB image, the image in HSV color space, and the generated histogram for the Hue plane.

LDP

LDP is applied on the extracted Hue plane of the image. For every image pixel, guided by Kirsch Mask[23], the calculation of LDP on relative edge response is in eight directions. This process will generate the code from relative magnitude strength which will further highlight the prominent corners, boundaries and edges within this magnitude. The LDP calculation is based on Eq. 2 below:

$$LDP_k(x_c,y_c) = \sum_{i=0}^{N-1} \int(m_i - m_k)2^i \, , \int(z) = \begin{cases} 1, z \geq 0 \\ 0, z < 0 \end{cases} \quad (2)$$

N, number of neighboring pixels is set to eight, m_i and m_k are eight directional edge responses based on eight Kirsch mask and the kth most significant response. $z=m_i–m_k$. To yield LDP binary code, LDP will compute the k prominent directions. The calculation of LDP is depending to the value of k. The use of different k values will affect the performance of the system [24]. For this work, k is set to three because it is widely used in literature. Because of that, top three directional bit response of 8-bit LDP code are set to one, and the other five is set to zero. In a set of eight bits, the possibility of having exactly three bits with value one is given by 56 combinations or bins (C_8^3) thus the number defines the histogram vector or feature set of LDP. Based on



RGB image HSV image Histogram of Hue plane

Fig. 2 Sample of RGB Image, HSV Image, and the Generated Histogram of the Hue Channel

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication

85	44	67
32	51	9
78	39	42

$\xrightarrow{\{m_i\}}$

100	380	228
372		244
4	84	468

$\xrightarrow{m_k}$

0	1	0
1		0
0	0	1

$\rightarrow$

	148	

LDP Binary Code: 10010100, LDP Decimal Code: 148

Fig. 3 Obtaining LDP Code for 3 × 3 Neighborhood

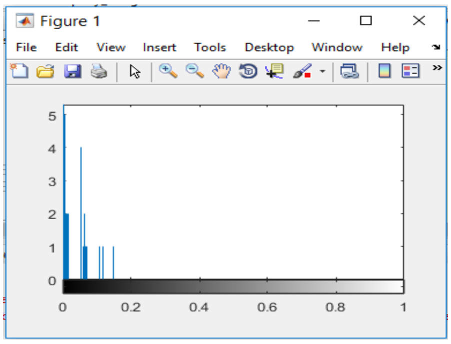


Fig. 4 Sample of LDP Histogram of Image

k=3, top three responses are assigned the value one and the rest set to zero which leads to the generation of 8-bit binary pattern. Having assigned the weights to 8-bit binary pattern, LDP decimal code is obtained. 56 feature vectors extracted are then saved as HLDP feature vectors Fig. 3 shows the process of generating the LDP code for a 3 × 3 pixels neighborhood. Fig. 4on the other hand gives us the sample of LDP histogram of the image.

ZM

The original calculation method of ZM used in [19] is adopted in this work. In order to construct ZM features, the radial polynomials and Zernike basis functions will need to be computed. ZM of $N{\times}N$ image can be obtained based on the Eq. 3 below:

$$Z_{nm} = \frac{(p+1)}{\lambda_N}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)R_{nm}(x,y)e^{-jm\,\theta_{xy}}\ , \qquad (3)$$

where n is the order of the radial polynomial, m is the number of repetition, p satisfying $n \geq 0, 0 \leq m \leq n$, and $n - |m|$ are even, $j = \sqrt{-1}, |r| \leq 1$, $0 \leq r_{xy} \leq 1$, and normalization factor λ_N is the number of pixels located in the unit circle π in the continuous domain. For this work, the same order n from [19] is adopted, which is 10.

Eqs. 4 and 5 below is used to calculate the transform distance r_{xy} and phase θ_{xy} at pixel (x, y):

$$r_{xy} = \frac{\sqrt{(2x-n+1)^2+(2y-N+1)^2}}{N} \qquad (4)$$

$$\theta_{xy} = tan^{-1}\left(\frac{N-1-2x}{2-N+1}\right) \qquad (5)$$

Feature Vectors / Color Feature

When ZM (to extract global shape feature) and LDP (to extract local shape feature) are applied on the grayscale image, it gives 36 feature vectors and 56 feature vectors respectively. For the texture feature, the direct application of LDP on grayscale image gives 56 feature vectors. The extraction of Hue plane on HSV color space provides 256 feature vectors. 56 color feature vectors are extracted when LDP is calculated on the Hue histogram of the image. The feature vectors are saved as the ZM feature vectors, LDP feature vectors and HLDP feature vectors for the online process and in feature database for the offline process, before similarity distance calculation comes into the picture. Therefore, 36 feature vectors from ZM, 56 feature vectors from LDP and 56 feature vectors from HLDP will yield a total of 148 feature vectors for the EZLH proposed.

For the first benchmark method, ZM-LDP framework by [19], the application of ZM will give 36 feature vectors and LDP 56 feature vectors. Therefore, there are 92 feature vectors in full. For the second benchmark method, [17], the application of LBP on Hue plane offers 256 feature vectors in total. Therefore, the generated feature vectors of the proposed work are still within an acceptable and comparable size with much higher retrieval accuracy, if compared to the benchmark methods.

Similarity Measure

The similarity measure used for this work is Euclidean Distance. There are three groups of feature vectors. The distance for each group is calculated using Eq. 6 below:

$$d_p(G^{Di}, G^{Q}) = \sqrt{\sum_{j=1}^{N}(G^{Dij} - G^{Qj})^2} \qquad (6)$$

$d_p$is the distance of pth group of feature vector. $G^{Di}$and G^{Q} is the feature vector of the jth database image and query image respectively. N is the maximum number of the feature vectors and G^{Dij} and G^{Qj} are the feature vector of the ith image of database image and feature vectors of the query image, respectively.

The distance for each group is performed separately. To obtain the final distance, D, the distance of each group is then summed. The formula to calculate the final distance is given in Eq. 7.

$$D = d_1 + d_2 + d_3 \quad (7)$$

IV. EXPERIMENTAL SETUP

The experiments of this research work are conducted based on the following setting.

Objective of the Experiments

The objective of experiment one is to find the most suitable image size to be used.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Different image size carries different information. Resizing image to a very small size will lead to loss of many important information while using too large image size will cause the information redundancy hence will increase computation time. Too many or too less information of image will affect the accuracy of image retrieval system. Therefore, the size of images needs to be examined in detail. The objective of experiment two is to compare the retrieval effectiveness between EZLH with the benchmark methods by [19] and [17].

Hardware and Software Specifications

All of the experiments are conducted on Windows 10 Home Single Language notebook with the Intel® Core™ i7-7500U CPU @ 2.70GHz 2.90GHz processor, 8.00 GB installed memory (RAM) and 64-bit Operating System, x64-based processor system type.

There are two software used in this research work, MATLAB R2015b and Microsoft Excel. The Matlab platform is used to solve most engineering and scientific problems. Its matrix-based feature is the best way to perform mathematics calculation. In image processing, it is often used to calculate the image feature and distance measure. In this research, this software is used as the development tool. The proposed EZLH, [19] and [17], and all benchmarks methods are developed and computed by using this software.

Microsoft Excel is a spreadsheet-based developed by Microsoft. Some of the features are for calculation and as graphing tool.

The purpose of using this software is to calculate *p*-value and *t-value* of Two-tailed paired *t*-test, to generate the Average 11 Standard Precision-Recall graphs and make comparison between each of the proposed framework and benchmark methods.

Data Collection

In order to see the impact of the proposed color feature, experiments are performed on Fish4Knowledge (natural image) dataset. The dataset is available at this website: http://groups.inf.ed.ac.uk/f4k/GROUNDTRUTH/RECOG/.

The dataset consists of 27370 images which are classified into 23 groups according to their respective species, which provide the ground truth. The number of images in each category is different from each other. Each image may be different in terms of fish position, scale, orientation, color, shape and texture. Fig. 5 shows the sample of the dataset.

Table 1 illustrates the fish species name and the number of images detected. The images were obtained with the help from fish detection and tracking software [25]. The images are then analyzed and categorized manually according to the species using the instructions issued by marine biologists [26].

To avoid bias, the query image is selected randomly by generating the file name using ‘randomly select name’ coding in Microsoft Excel. 30% out of 27370 RGB images in total from Fish4Knowledge (natural image) dataset image are used as query.

Parameters to Measure the Retrieval Performance

All data are gathered and analyzed to evaluate the research objectives. The effectiveness of the proposed and benchmark methods are calculated based on the commonly used retrieval effectiveness measurements in CBIR field which include Average 11 Standard Precision-Recall and Mean Average Precision (MAP). Two-tailed paired *t*-test is also utilized to determine whether the observed differences in performance between the investigated methods are significant or not.

Average 11 Standard Precision-Recall

Average 11 Standard Precision-Recall serves to compare the performance of more than one different retrieval system. This measurement is widely used in determining the accuracy of a retrieval system [11, 19, 27-29].Precision is interpreted as the fraction of number of retrieved items deemed relevant, while recall is the fraction of the relevant items retrieved. Eqs. 8 and 9establish the formula that computes the precision and recall value.

$$precision = \frac{retrieved\ \ relevant\ \ document}{retrieved\ \ documents} \quad (8)$$

$$recall = \frac{retrieved\ \ relevant\ \ document}{relevant\ \ documents} \quad (9)$$

To evaluate the ranked list, the precision-recall curve is used. The interpolated precision at 11 recalllevel (0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0) is needed to generate the curve. The formula of this calculation is established in Eqs. 10 and 11.

$$precision\ at\ rank\ (p') = \frac{retrieved\ \ relevant\ \ document\ \ at\ rank}{retrieved\ \ documents\ \ at\ rank} \quad (10)$$

$$recall\ at\ rank\ (r') = \frac{retrieved\ \ relevant\ \ document\ \ at\ rank}{relevant\ \ documents} \quad (11)$$

The interpolated precision (p_x) at certain recall, r is defined as the highest precision at rank *(p’*) found for any recall at rank (r'), $r' \geq r$. The value of interpolated precision at certain recall is established in Eq. 12.

$$p_x = \max_{r' \geq r} precision\ at\ rank\ (p')\ of\ recall\ at\ rank\ (r') \quad (12)$$

To compute the Average Precision at certain recall, Eq. 13 is adopted.

$$Average\ Precision = \frac{\sum_{i=1}^{N} P_x}{N} \quad (13)$$

N is the total number of data, P_x is the interpolated precision at recall x.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

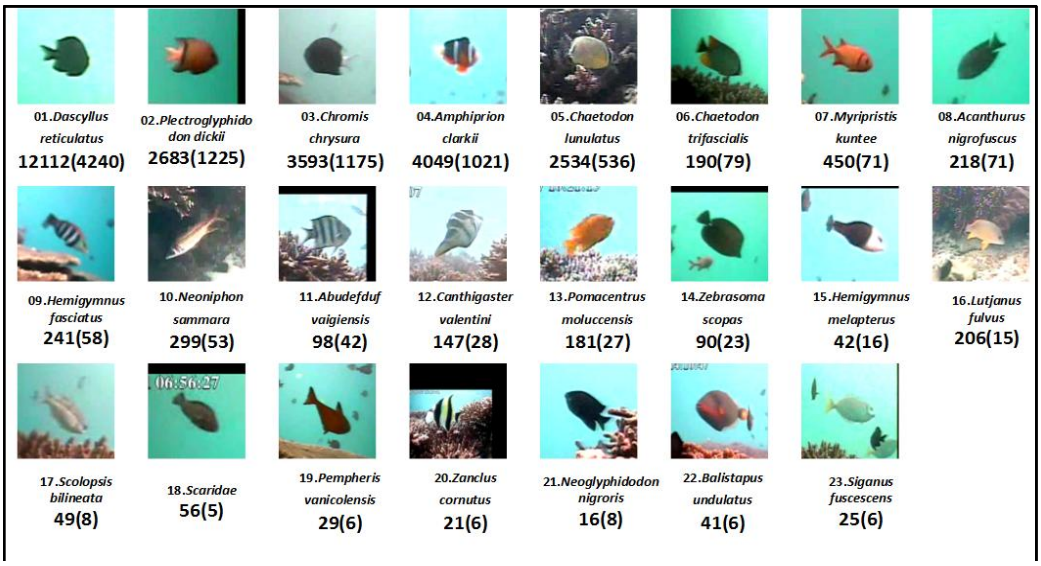


Fig. 5 Sample images from Fish4Knowledge (natural image) dataset

Table. 1 Fish ID, Species Name and Number of Detection

Species ID	Number of Detection
01.Dascyllus reticulatus	12112
02.Plectroglyphidodon dickii	2683
03.Chromis chrysura	3593
04.Amphiprion clarkii	4049
05.Chaetodon lunulatus	2534
06.Chaetodon trifascialis	190
07.Myripristis kuntee	450
08.Acanthurus nigrofuscus	218
09.Hemigymnus fasciatus	241
10.Neoniphon sammara	299
11.Abudefduf vaigiensis	98
12.Canthigaster valentini	147
13.Pomacentrus moluccensis	181
14.Zebrasoma scopas	90
15.Hemigymnus melapterus	42
16.Lutjanus fulvus	206
17.Scolopsis bilineata	49
18.Scaridae	56
19.Pempheris vanicolensis	29
20.Zanclus cornutus	21
21.Neoglyphidodon nigroris	16
22.Balistapus undulatus	41
23. Siganus fuscescens	25

Mean Average Precision (MAP)

MAP is referred as a single value measurement that compares more than one different retrieval system. It is the average of the average precision of all queries. Average precision (AP) is interpreted as the average of un-interpolated precision values at all ranks of retrieved documents for each query[17, 30-31]. Eqs. 14 and 15are used to compute the AP and MAP values.

$$AP = \frac{\sum_{i=1}^{M} P}{M} \quad (14)$$

$$MAP = \frac{\sum_{i=1}^{N} AP}{N} \quad (15)$$

N is the total number of query used while M is the total retrieved image.

Two-tailed paired t-test

By definition, the Two-tailed paired *t*-test is a statistical

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

test which determines whether the sample falls inside or outside the critical area of values; the critical area being a two sided area below the curve graph. In case the tested sample falls in between this area, the hypothesis is accepted, otherwise it would be applied as null hypothesis [32-33]. Thesignificant level used for the hypothesis which is to be rejected is 5%. Two frameworks are regarded as significant if the *p*-value is less than the significant value. Larger *t*-values translate into smaller *p*-values other than indicating that there is a larger difference between the two systems. In other words, there should be a difference in the retrieval performance between both systems. Eqs. 16, 17, 18 and 19 are used to obtain *t* and *p* values.

Mean of the sample

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (16)$$

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

Standard deviation of the sample:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \quad (17)$$

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

t-value:

$$t = \frac{\overline{\mu_1} - \overline{\mu_2}}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \quad (18)$$

Where $\overline{\mu_1}$ and $\overline{\mu_2}$ is the mean, σ_1^2 and σ_2^2 is the standard deviation, and N_1 and N_2 is the sample size of the first and second system for comparison.

Degree of freedom

$$v = \frac{\left(\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}\right)^2}{\frac{1}{N_1 - 1}\left(\frac{\sigma_1^2}{N_1}\right)^2 + \frac{1}{N_2 - 1}\left(\frac{\sigma_2^2}{N_2}\right)^2} \quad (19)$$

Whereσ_1^2 and σ_2^2 and N_1 and N_2 is standard deviation and sample size of the first and second systems to be compared. The *t*-value and degree of freedom value would help find the*p*-value from a statistical table (*t* distribution).

V. RESULTS AND DISCUSSION

The results derived from each experiment are discussed. Before processing, the image size is determined beforehand. Table 2 shows the result of implementing the EZLH framework on Fish4Knowledge dataset using different image sizes.

Evidently, the change in image sizes carries little impact on the MAP. Since adjusting image size to 64 × 64 pixels achieved higher MAP compared to other sizes, all dataset have their sizes adjusted to 64 × 64 pixels. This same setting is used for other experiments done in this work.

The results of Average 11 Standard Precision-Recall graph of the proposed and benchmark methods for all image categories applied are displayed in Fig. 6. The Precision-Recall and MAP value for each category, for the proposed and benchmark methods are next illustrated in Table 3. Notably, compared to the benchmark methods, the proposed method is able to achieve higher retrieval result.

From Table 3, in sum, EZLH is found to be suitable for the mentioned dataset, because compared to the other benchmark methods, the average precision accuracy achieved by EZLH is more than 50% for all categories. This experiment gives the evidence that the extraction of shape and texture features only is inadequate for natural images that have low quality and resolution. It necessitates a method that is able to extract color feature and not easily distracted by noise. The consideration of color feature had made the result more precise. HSV color space describes and interprets color similar to human, and Hue plane helps discriminate color. LDP counts each and every pixel of the images, and LDP is based on edge response, lending to easy handling of noise. It is evidenced that the LDP application on the Hue plane of the images is able to extract the color feature that can represent the image effectively. Despite the quality of the images being very low and the shape and texture of the images appear blurry, the HLDP is still able to extract the features that can depict the images and thus increase the retrieval system's accuracy. Even though HLBP regards the color feature and the method is quite similar to HLDP, the LBP features is easily distracted by random noise and non-monotonic changes in illumination [19]. Compared to HLBP which only considers the color feature, the shape and texture features are also deliberated by ZM and LDP which is a strong point for the EZLH method.

However, the result of HLBP is higher than EZLH for categories 1, 2, 3 and 21. LBP is categorized as contour-based descriptor. Therefore, it can get good result for contour-based shapes, but for region-based shapes, LDP is found to be effective. For most of the images of mentioned categories, the background is clear and simple, and the color of the fish is contradicted to the background color. The images of these categories are categorized as the image that has good shape contour. Therefore, the application of LBP on Hue plane of the mentioned categories increases the accuracy result. It shows that the information extracted is enough to represent the images.

Fish images are rich in color, where images are taken by placing the camera in deep sea (under water). This justifies why the images are not clear and the resolution is low. For a lot of the images, the shape and texture of the fish are blurry. The obvious criterion that distinguishes between images is color, so the consideration of shape and texture only would not be sufficient, hence having color is vital. However, considering the color feature for the proposed method will generate extra feature vector. There are 148 feature vectors for the proposed EZLH, 92 feature vectors for ZM-LDP [19], and 256 feature vectors for HLBP [17]. In spite of the fact that the number of feature vector of the EZLH is higher than one of the benchmark methods (ZM-LDP), the accuracy achieved is much higher and highly encouraging compared to all benchmark methods.

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication

Two-tailed paired *t*-test is performed to confirm that there is a significant difference between two frameworks. Let the MAP value for each method be $F_{m}, m \in \{\{ZM + LDP\}, \{HLBP\}, \{EZLH\}\}$. Hypothesis is $H_0 : F_m > 0.05$ (Similarity) and alternative hypothesis is $H_1 : F_m < 0.05$ (Non-similarity). The larger the *t* value, the more likely that the difference is significant. Table 4 highlights the Two-tailed paired *t*-test result for both methods. Result shows that EZLH has achieved a statistically significant improvement in retrieval performance as opposed to the ZM-LDP [19] and HLBP [17].

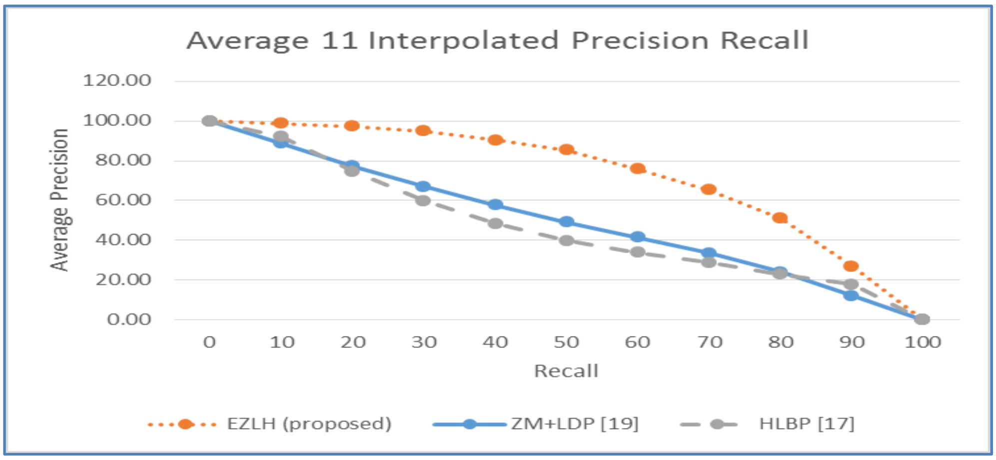


Fig. 6 Average 11 Standard Precision-Recall Graph of the Proposed and Benchmark Methods

Table. 2 Performance comparison using different image size in terms of MAP

Image size	MAP
32 × 32 pixels	48.29
64 × 64 pixels	49.04
128 × 128 pixels	48.9
256 × 256 pixels	48.97

Table. 3 MAP Value for Each Category of the Proposed and Benchmark Methods

Species	EZLH (proposed)	ZM-LDP [19]	HLBP [17]
1	76.49	37.03	85.5
2	69.09	40.06	87.3
3	65.98	51.18	70.4
4	84.61	23.57	76.1
5	75.29	47.86	74.4
6	68.64	41.84	15.1
7	82.7	34.79	43.7
8	73.72	36.79	16.2
9	83.93	58.48	16.2
10	77.96	65.73	36
11	57.24	33.89	34.6
12	73.13	23.01	24.9
13	71.73	56.2	23.5
14	63.86	43.2	21.6
15	89.89	27.95	26.1
16	72.18	38.9	36.15
17	86.38	42.18	46.4
18	82.97	27.8	25.5
19	79.82	30.54	36.2
20	70.16	33.19	48.8
21	80.26	34.2	83.21

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

22	79.43	42.13	31.8
23	73.87	42.3	55.9
MAP	75.62	39.69	44.15

Table. 4 Paired *t*-test at a significant level of 0.05. *p*-values < 0.05 indicate significantly different results

Method	*t*-value	*p*-value	Null Hypothesis (H_0)	Remark
ZM-LDP vs. EZLH	8.7047	4.0174E-11	Reject	EZLH > ZM-LDP
HLBP vs. EZLH	5.2007	4.9565E-06	Reject	EZLH > HLBP

VI. CONCLUSION

A CBIR system particularly for fish domain has been developed (EZLH). Experiment in Section 3 has shown that the extraction of shape and texture feature only is not sufficient for blur natural image with low resolution and quality. This experiment also gave evidence that when LDP is applied on Hue plane of HSV colour space of low quality image, the colour information extracted is good enough in representing the image colour, hence contributing to higher retrieval accuracy as compared to other benchmark methods. EZLH has proven itself to be one of the suitable descriptors for the dataset considered. The capability of the method in extracting colour feature makes it effective for fish species image representation and retrieval.

Based on experimental analysis, it is shown that the EZLH outperformed the two benchmark methods in the sense that it attained higher retrieval result. However, few things can be considered for future works to obtain better results.

Even though the fish images are from the same category, but the original size and colour of the images is different. The proposed CBIR system extracts the fish object together with the background. Suggestion for future work is as a pre-processing step, the background may be eliminated first so that the extraction only focuses on the fish. Then, the employment of classifier can be considered to give a boost to the system's performance. The proposed CBIR system only made use of the Euclidean distance as similarity measure, further research may settle for different similarity measures such as Manhattan Distance and Jaccard Similarity.

ACKNOWLEDGMENT

This research is financially supported by the Malaysian Ministry of Education under the Fundamental Research Grant Scheme08-01-16-1855FR.

REFERENCES

1. Qin, H., Li, X., Liang, J., Peng, Y., & Zhang, C. (2016). DeepFish: Accurate underwater live fish recognition with a deep architecture. *Neurocomputing*, *187*, 49-58.
2. Sun, X., Yang, J., Wang, C., Dong, J., & Wang, Z. (2018). Low-contrast underwater living fish recognition using PCANet. In *Proc. SPIE 10615, Ninth International Conference on Graphic and Image Processing (ICGIP 2017)*, 10 April 2018.
3. Ghosh, M., Mukherjee, J., & Parekh, R. (2013). Fish shape recognition using multiple shape descriptors. *International Journal of Computer Applications*, *73(16)*, 14-19.
4. Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W. P., … & Müller, H. (2016, September). LifeCLEF 2016: multimedia life species identification challenges. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pp. 286-310. Springer, Cham.
5. Singh, C., Walia, E., & Kaur, K. P. (2018). Color texture description with novel local binary patterns for effective image retrieval. *Pattern Recognition*, 76, 50-68.
6. Wang, W., Yan, X., Zhao, L., & Shi, J. (2017, December). A new affine invariant descriptor for shape recognition. In *Computer and Communications (ICCC), 2017 3rd IEEE International Conference on*, pp. 1693-1697. IEEE.
7. Osman, N. S. & Mustaffa, M. R. (2015). A review on content-based image retrieval representation and description for fish. In *2015 4th International Conference on Advanced*
8. *Computer Science Applications and Technologies (ACSAT 2015)*, 8-10 December 2015, Kuala Lumpur, Malaysia, pp. 39-44. IEEE.
9. Zheng, X., Zhong, J., & Zhang, Y. (2011). A soa-based fish recognition system prototype, In *Intelligent Human-Machine Systems and Cybernetics (IHMSC), International Conference on*, pp. 66-69. Zhejiang.
10. Hu, J., Li, D., Duan, Q., Han, Y., Chen, G., & Si, X. (2012). Fish species classification by color, texture and multi-class support vector machine using computer vision. *Computers and Electronics in Agriculture*,*88*, 133-140.
11. Yao, H., Duan, Q., Li, D., & Wang, J. (2013). An improved k-means clustering algorithm for fish image segmentation. *Mathematical & Computer Modelling*, *58(3)*, 790-798.
12. Goyal, A., & Walia, E. (2012). An analysis of shape based image retrieval using variants of zernike moments as features. *International Journal of Imaging and Robotics*, *7(S12)*, 44-69.
13. Zarchi, M. S., Monadjemi, A., & Jamshidi, K. (2014). A semantic model for general purpose content-based image retrieval systems. *Computers & Electrical Engineering*, *40(7)*, 2062-2071.
14. Mustaffa, M. R., Norowi, N. M., & Yee, S. M. (2016, August). Content-based image retrieval system for marine invertebrates. In *Information Retrieval and Knowledge Management (CAMP), 2016 Third International Conference on*, pp. 71-76. IEEE.
15. Chakravarti, R., & Meng, X. (2009, April). A study of color histogram based image retrieval. In *Information Technology: New Generations (ITNG'09), Sixth International Conference on*, 2009. pp. 1323-1328. IEEE.
16. Sarker, M. I. H., & Iqbal, M. S. (2013). Content-based image retrieval using haar wavelet transform and color moment. *Smart CR, 3(3)*, 155-165.
17. He, D. C., & Wang, L. (1990). Texture unit, texture spectrum, and texture analysis. *IEEE transactions on Geoscience and Remote Sensing*, *28(4)*, 509-512.
18. Zhu, C., Bichot, C. E., & Chen, L. (2010, August). Multi-scale color local binary patterns for visual object classes recognition. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pp. 3065-3068. IEEE.
19. Jabid, T., Kabir, M. H., & Chae, O. (2010). Facial expression recognition using local directional pattern (LDP). In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pp. 1605-1608. IEEE.

20. Goyal, A. & Walia, E. (2014). Variants of dense descriptors and zernike moments as features for accurate shape-based image retrieval. *Signal, Image and Video Processing, 8(7)*, 1273-1289.
21. Saurav, S., Belsare, P., & Sarkar, S. (2015). Holistic Correlation of Color Models, Color Features and Distance Metrics on Content-Based Image Retrieval. *International Research Journal of Engineering and Technology (IRJET),* 39-43.
22. Soni, D., & Mathai, K. J. (2015, April). An efficient content based image retrieval system based on color space approach using color histogram and color correlogram. In *Communication Systems and Network Technologies (CSNT), 2015 Fifth International Conference on*, pp. 488-492. IEEE.
23. Srikanth, V., Maczurek, A., Phan, T., Steele, M., Westcott, B., Juskiw, D., & Münch, G. (2011). Advanced glycation endproducts and their receptor rage in alzheimer's disease. *Neurobiology of aging, 32(5)*, 763-777.
24. Kirsch, R. (1971). Computer determination of the constituent structure. *Computers and biomedical research, 4*, 315-328.
25. Shabat, A. M., & Tapamo, J. R. (2016). Directional local binary pattern for texture analysis. In *International Conference Image Analysis and Recognition*, pp. 226-233. Springer International Publishing.
26. Boom, B. J., Huang, P. X., Beyan, C., Spampinato, C., Palazzo, S., He, J., & Chen-Burger, Y. H. (2012). Long-term underwater camera surveillance for monitoring and analysis of fish populations. *VAIB12*.
27. Boom, B. J., Huang, P. X., He, J., & Fisher, R. B. (2012, November). Supporting ground-truth annotation of image datasets using clustering. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pp. 1542-1545. IEEE.
28. Singh, C., & Sharma, P. (2013). Performance analysis of various local and global shape descriptors for image retrieval. *Multimedia systems, 19(4)*, 339-357.
29. Jain, K., & Bhadauria, S. S. (2016). Performance Evaluation of Content-Based Image Retrieval on Feature Optimization and Selection Using Swarm Intelligence. *Performance Evaluation, 7(3)*, 245-249
30. Saini, A., & Singh, S. (2016). Evaluation of CBIR System by Using Low Level Features of an Image. *Evaluation, 5(6)*, 1951-1954.
31. Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 1725-1732.
32. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1-9.
33. Furuta, R., Inoue, N., & Yamasaki, T. (2018, February). Efficient and Interactive Spatial-Semantic Image Retrieval. In *International Conference on Multimedia Modeling*, pp. 190-202. Springer, Cham.
34. Dhara, A. K., Mukhopadhyay, S., Dutta, A., Garg, M., & Khandelwal, N. (2017). Content-based image retrieval system for pulmonary nodules: Assisting radiologists in self-learning and diagnosis of lung cancer. *Journal of digital imaging, 30(1)*, 63-77.