FISH IMAGE ANALYSIS: FUSION OF MOMENT-BASED AND DIRECTIONAL FEATURES IN COLOUR SPACE

Mas Rina Mustaffa[1*] , Noorul Shuhadah Osman[2], Shyamala Doraisamy[3], and Hizmawati Madzin[4]

[1, 2, 3, 4] Multimedia Department, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia

Emails: MasRina@upm.edu.my[1*], GS42122@student.upm.edu.my[2], Shyamala@upm.edu.my[3], Hizmawati@upm.edu.my[4]

ABSTRACT

This study introduces an innovative approach to content-based image retrieval (CBIR) specifically designed for fish species identification. The proposed method integrates shape, colour, and texture features using Zernike Moments Invariant (ZMI) and Local Directional Pattern (LDP), applied to the momentgram and the hue channel of the HSV colour space. This fusion ensures invariance to transformations such as rotation, scaling, and translation, enabling robust performance on natural images with varying orientations and quality. The method was evaluated using the Fish4Knowledge dataset, consisting of 27,370 images, with 30% randomly selected as query images. Experimental results demonstrate that the proposed method achieved a mean average precision (MAP) of 84.17%, significantly outperforming comparable state-of-the-art approaches. Statistical analysis using two-tailed paired t-tests confirms its superiority. By combining global shape descriptors, local texture features, and colour properties, this method delivers a comprehensive representation of fish images. The inclusion of moment-based descriptors enhances its robustness against low-resolution images and noise. This research underscores the importance of combining diverse features within CBIR systems and offers a significant improvement in retrieval accuracy, contributing to domain-specific applications such as sustainable fisheries management and aquaculture research.

Keywords: Colour feature; Content-based image retrieval; Fish identification; Local directional pattern; Zernike moments invariant.

1.0 INTRODUCTION

Fish play a significant role in human life, serving as a vital source of protein and essential nutrients. Beyond their consumption as food, advancements in processing technologies have enabled fish to be transformed into various products, such as dietary supplements and convenience foods. The aquatic food industry contributes immensely to global socio-economic development, underscoring the critical importance of effective fisheries management [1]. Addressing challenges such as declining fish stocks, setting fishing quotas, identifying breeding grounds, and combating fraud in processed seafood requires systematic and accurate monitoring methods [2]. Since the nutritional content of fish varies significantly across species, precise identification is essential to support sustainable aquaculture programmes and ensure food security [3].

Traditional methods for identifying fish species rely on manual examination, requiring significant labour, expertise, and time. These approaches are impractical when managing large datasets of fish images, particularly in marine environments [4]. Content-Based Image Retrieval (CBIR) systems have emerged as transformative tools, enabling the automated retrieval of relevant images from extensive databases [5]. CBIR processes involve extracting features such as shape, colour, and texture, which are then stored in a database [6]. A query image is analysed to extract its features, which are subsequently compared to database entries using similarity measures. Images are ranked and retrieved based on their closeness to the query image, significantly improving the efficiency of fish species identification [7].

Proper feature selection is crucial in CBIR systems. Over-extraction of features can lead to redundancy, affecting accuracy and computational efficiency, while insufficient features may fail to represent the query image adequately [8]. Fish species are primarily distinguishable by their shape, colour, and texture. Marine fish, however, present unique challenges due to varying orientations, sizes, and angles in images, necessitating feature extraction methods that are invariant to transformations such as rotation, scaling, and translation [9].

Recent advancements in CBIR research have demonstrated the value of fusing global and local features with colour and texture properties. Methods combining Zernike Moments Invariant (ZMI) for global shape features and Local Directional Pattern (LDP) for texture analysis have shown significant potential [10], [11], [12]. Hue-based colour features have further enhanced species differentiation, particularly for natural images where shape and texture are insufficient. Studies have also highlighted the efficacy of hybrid approaches that incorporate multiple features, improving retrieval accuracy for aquatic datasets [13], [14]. For instance, Mustaffa et al. [15] demonstrated the advantage of multi-feature fusion, while Dewan and Thepade [16] addressed low-resolution and noisy conditions with advanced descriptors.

Despite these advancements, challenges persist in uncontrolled marine environments, particularly due to issues like varying lighting conditions, background clutter, and complex fish orientations that negatively impact retrieval accuracy. While existing hybrid CBIR frameworks, which integrate moment-based descriptors with texture and colour properties, have shown improvements, they still struggle with robustness against severe noise, geometric distortions, and real-world underwater variability. This paper addresses these specific challenges by proposing an enhanced hybrid framework that refines feature extraction through the combination of ZMI, LDP, and hue channel features, aiming to improve retrieval accuracy and resilience in complex and noisy marine environments.

2.0 RELATED WORKS

Content-Based Image Retrieval (CBIR) systems have become an essential component across various fields, especially in fisheries management and aquaculture, for efficiently classifying and retrieving image datasets. Over the years, significant progress has been made in CBIR, particularly through advancements in feature extraction techniques, hybrid frameworks, and deep learning methodologies.

Hybrid feature extraction approaches have garnered considerable attention due to their ability to address the limitations of single-feature techniques. Rani et al. [17] demonstrated the effectiveness of combining fractional Hartley transform with hybrid features for image retrieval, achieving improved retrieval accuracy. This work complements the findings of Khan et al. [18], who proposed an effective hybrid framework for CBIR. Their approach integrates feature fusion methods that combine local and global descriptors, enabling robust performance in diverse and large-scale datasets. The framework excels in handling variations in image attributes, such as shape and texture, demonstrating its applicability across complex retrieval scenarios. Additionally, Madhu and Kumar [19] proposed a hybrid feature extraction technique for content-based medical image retrieval, integrating Discrete Wavelet Transform (DWT), Principal Component Analysis (PCA), and Gray-Level Co-Occurrence Matrix (GLCM). Their method, which combines segmentation and clustering techniques, has proven effective in enhancing retrieval accuracy across varied datasets, offering insights transferable to other domains, including aquatic datasets. In a broader context, Chu et al. [20] showcased the value of hybrid frameworks beyond image retrieval, highlighting how integrating multiple technologies such as blogs and Facebook enhanced collaborative learning and data sharing in educational environments. This cross-domain evidence reinforces the versatility and effectiveness of hybrid systems in improving performance across complex and data-rich applications.

Deep learning methods have further transformed CBIR by bridging the semantic gap between low-level features and high-level image understanding. Mittal et al. [7] conducted a comprehensive survey of deep learning techniques for underwater image classification, highlighting their effectiveness in addressing challenges such as light absorption, scattering, and noise inherent in underwater environments. Their review emphasised the advantages of convolutional neural networks (CNNs) in learning robust feature representations, particularly for distorted and noisy data. Saleh et al. [9] expanded upon this by surveying the application of computer vision and deep learning techniques for fish classification in underwater habitats. Their study demonstrated how CNNs could effectively overcome the complexities of underwater environments, achieving higher precision and robustness compared to traditional methods. Similarly, Liu and Yang [21] investigated the exploitation of deep texture features for CBIR. Their work introduced the Deep Texture Feature Histogram (DTFH), combining classical texture features with deep learning approaches, enabling the capture of intricate texture details critical for robust and accurate image retrieval.

The application of CBIR to aquatic image datasets has been explored extensively. Hasegawa and Nakano [4] proposed a robust fish recognition system utilising foundation models to facilitate automatic fish resource management. Their approach integrates advanced feature extraction techniques with deep learning models to enhance the accuracy of fish species identification, even in challenging underwater environments. This aligns with the findings of Veiga and Rodrigues [22], who utilised Vision Transformers for fine-grained fish classification across datasets of varying sizes. Their work demonstrated the effectiveness of Vision Transformers in capturing intricate patterns and details necessary for accurate classification, highlighting the importance of advanced deep learning architectures for improved retrieval performance and biodiversity monitoring.

Feature fusion has emerged as a key strategy for improving CBIR systems. Wang et al. [23] introduced a two-stage CBIR framework that employs sparse representation and feature fusion techniques. Their approach effectively combines diverse feature types, such as texture and shape descriptors, to enhance retrieval performance across a range of datasets. Ju and Xue [24] adopted an alternative strategy by leveraging an improved AlexNet model for fish species recognition. Their method demonstrated significant improvements in accuracy by refining feature extraction processes, making it particularly effective for aquatic datasets.

In addition to these advancements, CBIR systems have been applied to various domains requiring robust image classification and retrieval capabilities. Silva et al. [25] introduced a scalable open-source framework for machine learning-based image collection, annotation, and classification, focusing on automatic fish species identification. Their framework demonstrated the effectiveness of integrating machine learning techniques for image-based classification tasks, providing a flexible and adaptable solution for applications in fisheries management and related fields. Similarly, Tadepalli et al. [26] proposed a CBIR system that leverages Gaussian–Hermite moments in conjunction with firefly and grey wolf optimisation algorithms. Their approach optimised feature selection and similarity measures, significantly improving retrieval performance across diverse datasets, highlighting the adaptability of CBIR systems to a variety of practical challenges.

Beyond aquatic applications, recent studies have focused on underwater image enhancement techniques to support CBIR systems. Zhang et al. [27] proposed an underwater image enhancement method based on weighted wavelet visual perception fusion, which effectively improved visual quality by addressing colour distortion and noise. This approach enhanced image retrieval accuracy in challenging underwater environments. Similarly, Panetta et al. [28] introduced a comprehensive benchmark dataset for underwater object tracking and employed generative adversarial networks (GANs) for image enhancement. Their method demonstrated significant improvements in correcting underwater distortions, providing a robust foundation for object tracking and image retrieval tasks. These advancements underscore the importance of innovative image enhancement techniques in overcoming the challenges of underwater datasets, thereby supporting more effective CBIR systems.

Recent research in image retrieval has demonstrated significant advancements through the use of mathematical moments, optimisation techniques, and multimodal frameworks. Hassan et al. [29] developed an efficient biomedical image retrieval system leveraging radial associated Laguerre moments. Their method achieved high retrieval accuracy and efficiency, addressing the unique challenges posed by feature extraction in medical imaging. Similarly, Fadaei and Rashno [30] proposed a CBIR framework that combines wavelet and Zernike features, optimised using the particle swarm optimisation algorithm. This approach effectively reduced computational complexity while maintaining precision in retrieval tasks. Building on these advancements, Bibi et al. [31] introduced a multimodal framework for CBIR using a query-by-visual-search mechanism. This framework integrated diverse features to enhance adaptability and retrieval performance across varied datasets. Together, these works highlight the growing importance of combining mathematical and computational techniques to improve the effectiveness and efficiency of image retrieval systems in diverse domains.

The field of CBIR continues to evolve with innovative approaches that integrate advanced learning techniques and user feedback to improve performance. Zhao et al. [32] introduced a scalable sub-graph regularisation framework that incorporates long-term relevance feedback to enhance retrieval accuracy and efficiency, particularly in dynamic and evolving datasets. Alarcão et al. [33] further advanced CBIR methodologies with ExpertosLF, a dynamic late fusion system that utilises online learning and real-time relevance feedback to optimise retrieval processes and adapt to user preferences. Meanwhile, Yang et al. [34] applied contrastive learning to CBIR for the automatic recognition of in situ marine plankton images, demonstrating its effectiveness in managing complex and heterogeneous datasets. These studies underscore the growing role of learning-based strategies and feedback mechanisms in developing adaptive and high-performance CBIR systems.

The literature review highlights notable advancements in CBIR systems, particularly in the context of aquatic image datasets. Research has focused on hybrid feature extraction techniques, such as the integration of Zernike Moments Invariant (ZMI) and Local Directional Pattern (LDP), to enhance retrieval accuracy by combining texture, shape, and colour properties. Feature fusion methods and optimisation strategies have further improved system performance, enabling the effective retrieval of relevant images across diverse datasets. Additionally, deep learning models have been utilised to bridge the gap between low-level features and high-level image representation, addressing challenges such as noise and variability in underwater environments. However, current methods often fail to fully address the complexities introduced by transformations, including rotation, scaling, and translation, which are crucial for robust image retrieval. Furthermore, the challenge of balancing computational efficiency with effective feature representation persists. These limitations emphasise the need for the proposed framework, which integrates ZMI, LDP, and hue channel features to provide robust, efficient, and accurate CBIR, particularly for fisheries management and aquaculture applications.

3.0 METHODOLOGY

In the offline process, encoding the colour features begins with converting the original image into the HSV colour space. The hue component of the HSV image is processed using the Local Directional Pattern (LDP), and the resulting features are stored in the colour database. For the extraction of shape and texture features, the image is transformed into grayscale. LDP is directly applied to the grayscale image to extract texture features, which are subsequently saved in the texture database. For shape features, the Zernike Moment Invariant is applied to the grayscale image to produce a momentgram. This momentgram undergoes processing with the 2D Dual Tree Complex Wavelet Transform (2D-DTCWT), after which LDP is calculated on the output from the 2D-DTCWT. These shape features are then stored in the shape database. This procedure is carried out for all images in the dataset.

In the online process, the query image undergoes the same feature extraction steps. Once the features are extracted, they are compared with those in the respective databases using a similarity measurement method. The system retrieves images based on similarity values, where a lower similarity value signifies a greater resemblance to the query image. Fig. 1 depicts the framework of the proposed system.

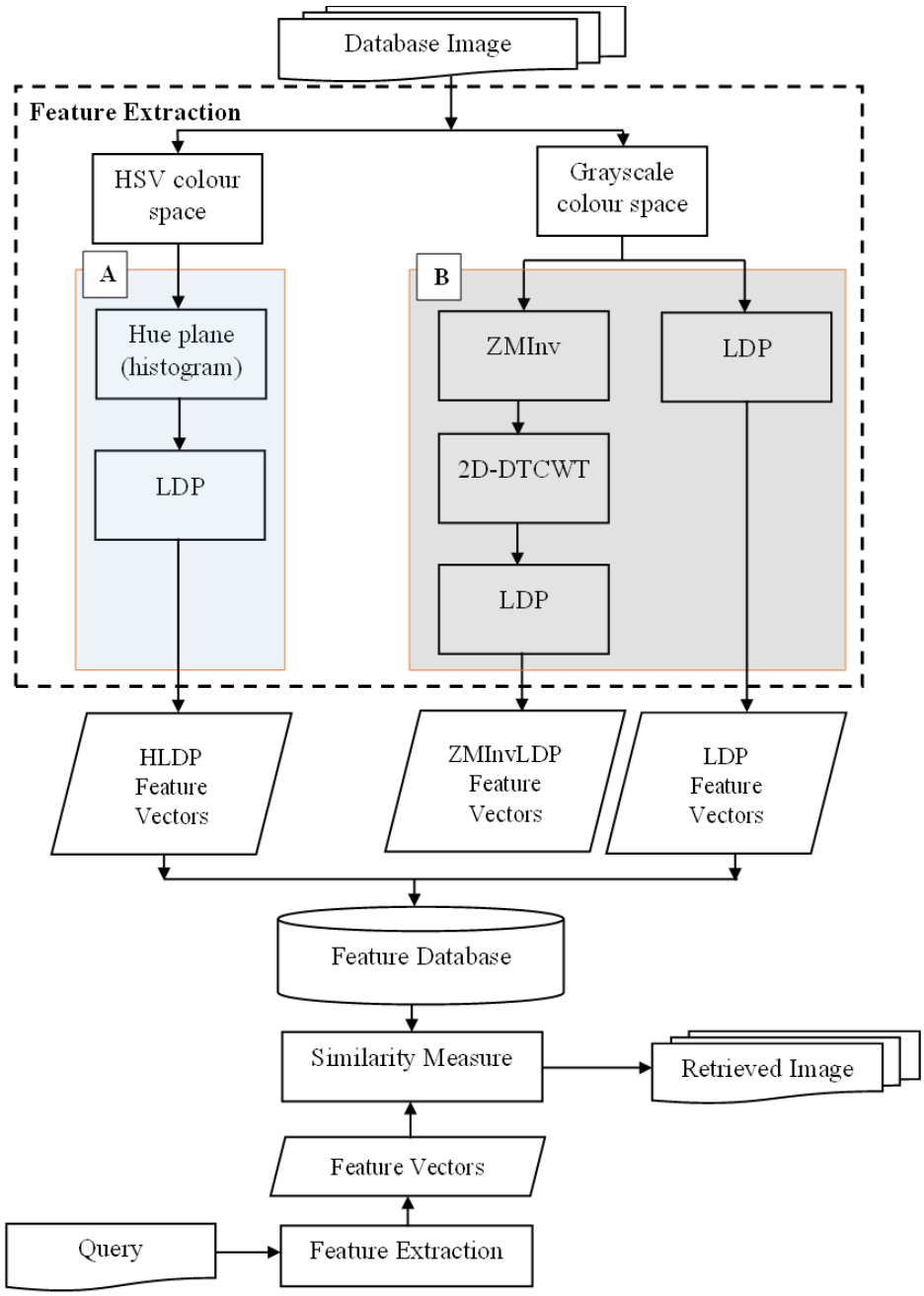


Fig. 1: Framework of the proposed CBIR system

3.1 Query Image/ Database Image

The dataset utilised in this research was obtained from the Fish4Knowledge platform (natural image) [35]. A total of 27,370 images were used as the dataset, with 30% of these images randomly selected to serve as query images. The dataset includes images from multiple fish species, with some species having more images than others, reflecting the natural distribution and availability of data. The dataset details, including the number of images per species, are available at the Fish4Knowledge ground truth repository [36]. To provide better insight into the nature and complexity of the dataset, Fig. 2 displays sample images from the Fish4Knowledge dataset. These images highlight the variability in fish orientations, lighting conditions, and background clutter inherent in the dataset, which pose challenges for CBIR methods.


01. *Dascyllus reticulatus*


02. *Plectroglyphidodon dickii*


03. *Chromis chrysura*


04. *Amphiprion clarkii*


05. *Chaetodon lunulatus*


06. *Chaetodon trifascialis*


07. *Myripristis kuntee*


08. *Acanthurus nigrofuscus*


09. *Hemigymnus fasciatus*


10. *Neoniphon sammara*


11. *Abudefduf vaigiensis*


12. *Canthigaster valentini*


13. *Pomacentrus moluccensis*


14. *Zebrasoma scopas*


15. *Hemigymnus melapterus*


16. *Lutjanus fulvus*


17. *Scolopsis bilineata*


18. *Scaridae*


19. *Pempheris vanicolensis*


20. *Zanclus cornutus*


21. *Neoglyphidodon nigroris*


22. *Balistapus undulatus*


23. *Siganus fuscescens*

Fig. 2: Sample images from the Fish4Knowledge dataset according to species, illustrating variability in fish orientation, lighting conditions, and background complexity

3.2 Grayscale Colour Space

As the extraction of shape and texture features does not require colour information, the images are converted into grayscale colour space. This transformation reduces computational time and eliminates redundant features that could otherwise impact the system's accuracy.

3.3 HSV Colour Space

Colour features are extracted by converting the dataset images into the HSV colour space. The hue channel is then isolated from the remaining channels for further processing.

3.4 Local Directional Pattern (LDP)

LDP is employed in three distinct feature extraction processes. For colour feature extraction, it is calculated on the Hue channel of the image. For shape features, LDP is applied to the momentgram derived from the Zernike Moment Invariant (ZMInv). In the case of texture features, LDP is directly applied to the grayscale image.

LDP is a pixel-based descriptor that operates using edge responses. It computes relative edge responses in eight directions for each pixel in an image using the Kirsch mask. These calculations generate a code representing the relative strength of edge magnitudes. The magnitude emphasises key features such as corners, boundaries, and edges. The LDP formula is presented in Equation 1.

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

(1)

The number of neighbouring pixels, N is fixed at eight, with m_i and m_k representing the eight directional edge responses based on the Kirsch mask and the k-th most significant response, respectively. The calculation $z = m_i - m_k$ is used. LDP identifies the k most significant directions to produce the LDP binary code. In this study, k is set to three, meaning the three most significant directional bit responses of the 8-bit LDP code are assigned the value one, while the remaining bits are set to zero. The value k = 3 is widely adopted in the literature, as altering k impacts system performance.

Within a set of 8 bits, there are 56 possible combinations where exactly three bits have the value one. These combinations define the histogram vector or feature set of the LDP. All eight directional responses are computed using Kirsch masks, represented as $\{M_i\}$. Based on k = 3, the top three responses are set to one, and the remaining responses are assigned a value of zero, resulting in an 8-bit binary pattern. The LDP decimal code is then determined by assigning weights to the 8-bit binary pattern. Fig. 3 illustrates the process of generating the LDP code for a 3 × 3 pixel neighbourhood.

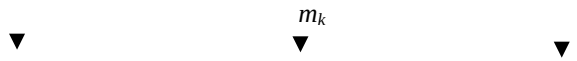
85	44	67
32	51	9
78	39	42

$\{m_i\}$

100	380	228
372		244
4	84	468

0	1	0
1		0
0	0	1

	14 8	



LDP Binary Code: 10010100, Corresponding Decimal Code: 148

Fig. 3: Computation of LDP code for a 3 × 3 neighbourhood

3.5 Zernike Moment Invariant (ZMInv)

ZMInv is a technique employed to calculate the image moment. It involves computing Zernike Moments (ZMs) using Geometric Moments (GMs). Initially, the image is represented as a series of partial intensity slices by applying intensity slice representation [37] and the algorithm outlined in [38].

3.6 Feature Vectors/ Feature Database

The term 'feature vectors' refers to the attributes extracted from an image after undergoing the feature extraction process, while 'feature database' pertains to the repository of extracted features for the dataset images. In framework A, the Hue component of the HSV colour space produces feature vectors of size 64 × 64, and applying LDP to the Hue channel results in 56 feature vectors. In framework B, applying LDP directly to the grayscale image yields 56 feature vectors, whereas the application of ZMInv on the grayscale image generates 8×8 feature vectors (known as the momentgram). When LDP is applied to the momentgram, an additional 56 feature vectors are created. Both frameworks A and B ultimately extract a total of 56 feature vectors each for colour, texture, and global and local shape, resulting in a combined total of 168 features.

To handle the potential challenges of feature selection and ensure that only the most informative features are used for retrieval, this study adopts a careful feature fusion approach. While all extracted features (colour, texture, and shape) are initially considered, the framework employs dimensionality reduction strategies to avoid feature redundancy and overfitting. Specifically, feature normalisation is applied to scale values uniformly, ensuring balanced contribution across different feature types. Moreover, experiments were conducted to analyse the impact of feature combinations on retrieval accuracy, allowing the identification of optimal feature sets that enhance performance while maintaining computational efficiency. This approach ensures that the most discriminative features are retained, improving overall retrieval accuracy without introducing unnecessary complexity.

3.7 Similarity Measure

In this study, the Euclidean Distance (refer to Equation 2) is utilised to calculate the similarity value between query images and database images for each group of feature vectors. The study includes three distinct groups of features: ZM feature vectors, LDP feature vectors, and Hue LDP feature vectors.

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

The variable d_p represents the distance for the *p*-th group of feature vectors, while G^{Di} and G^{Q} denote the feature vectors of the *j*-th database image and the query image, respectively. The parameter *N* indicates the total number of feature vectors. Additionally, G^{Dij} and G^{Qj} specifically correspond to the feature vectors of the *i*-th database image and the query image.

The distances for each group of features are calculated separately. The overall distance, *D*, is obtained by summing the distances of each feature vector group as expressed in Equation 3:

$$D = d_1 + d_2 + d_3 \quad (3)$$

Here, d_1, d_2 and d_3 represent the distances calculated for the ZM, LDP, and Hue LDP feature vectors, respectively.

4.0 EXPERIMENTAL SETUP

This section describes the experimental setup designed to evaluate the proposed method. It includes the objectives of the experiments, the parameters for assessing retrieval performance, and the statistical and benchmark methods

utilised. Each aspect has been carefully planned to provide a thorough evaluation of the ZM-LDP approach, focusing on its effectiveness in representing colour, shape, and texture features. The experiments also aim to demonstrate its resilience to transformations such as rotation, scaling, and translation, as well as to identify improvements that enhance its fusion capabilities.

4.1 Objective of the Experiments

The aim of the experiments is to refine the ZM-LDP approach, enabling it to effectively represent colour, shape, and texture features while remaining invariant to transformations such as rotation, scaling, and translation.

Prior to this, a preliminary experiment is conducted to determine the most appropriate image size and the optimal order value for ZMInv. The size of an image affects the amount of information it contains. Both excessive and insufficient detail can impact the accuracy of the image retrieval system. Reducing an image to a very small size results in the loss of critical information, whereas using an excessively large image introduces redundant data, leading to increased computation time. Therefore, careful evaluation of image size is essential.

A similar consideration applies to the order and repetition values used for ZMInv. Feature extraction outcomes vary depending on the chosen order and repetition. The number of repetitions is determined by the order. For instance, an order of one includes $m = 0$ and $m = 1$ as repetitions, where $m = 0$ corresponds to a horizontal shift, and $m = 1$ corresponds to a vertical shift or tilt of the image.

4.2 Parameters to Measure the Retrieval Performance

Three distinct measurement methods are employed to evaluate the performance of the proposed approach and benchmark methods.

4.2.1 Average 11-Point Precision-Recall

The Average 11-Point Precision-Recall method is used to compare the effectiveness of multiple retrieval systems. Precision represents the proportion of retrieved items that are relevant, while recall denotes the proportion of relevant items successfully retrieved [39]. The formulas for calculating precision and recall are provided in Equations 4 and 5.

$$Precision = \frac{Retrieved\ Relevant\ Document}{Retrieved\ Document} \tag{4}$$

$$Recall = \frac{Retrieved\ Relevant\ Document}{Relevant\ Document} \tag{5}$$

Precision and recall metrics do not account for the ranking of retrieved items. To evaluate ranked lists, a precision-recall curve is employed. This curve is generated using the interpolated precision at 11 specific recall levels: 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0. The formulas used to compute precision and recall at a given rank are provided in Equations 6 and 7.

$$Precision\ at\ Rank\ (p') = \frac{Retrieved\ Relevant\ Document\ at\ Rank}{Retrieved\ Documents\ at\ Rank} \tag{6}$$

$$Recall\ at\ Rank\ (Rank') = \frac{Retrived\ Document\ at\ Rank}{Relevant\ Documents} \tag{7}$$

Equation 8 is utilised to determine the interpolated precision (p_x) at a specific recall level. For any recall at rank (r'), where $r' \geq r, r'$ is defined as the maximum precision (p') observed at that rank.

$$p_x = Max\ r' \geq r\ Precision\ at\ Rank\ (p')\ of\ Recall\ at\ Rank\ (r') \tag{8}$$

Equation 9 is applied to compute the Average Precision for a given recall level.

$$Average\ Precision = \frac{\sum_{i=1}^{N} p_x}{N}\ x = \{0.0, 0.1, 0.2, 0.3, \ldots, 1.0\} \quad (9)$$

The total amount of data is denoted as *N*, while the interpolated precision at a specific recall level *x* is represented as p_x.

4.2.2 Mean Average Precision

Mean Average Precision (MAP) is widely recognised as one of the most effective single-value metrics for comparing the accuracy of multiple retrieval systems [40]. Average Precision (AP) refers to the mean of un-interpolated precision values across all ranks of retrieved images for a given query. Meanwhile, MAP represents the mean of the AP values across all queries. The calculations for AP and MAP are defined by Equations 10 and 11, respectively [41].

$$AP = \frac{\sum_{i=1}^{M} p}{M} \quad (10)$$

$$MAP = \frac{\sum_{i=1}^{N} AP}{N} \quad (11)$$

The total number of queries is represented by *N*, while *M* denotes the total number of images retrieved.

4.2.3 Two-tailed Paired *t*-test

The statistical method employed in this research is the two-tailed paired *t*-test. This test is commonly used to determine whether a sample lies within or outside a specified range of values, referred to as the critical region [42]. The critical region comprises the two tails under the curve of the distribution graph. The hypothesis is accepted if the sample falls within the critical region; otherwise, the null hypothesis is retained.

In this study, a significance level of 5% is used to evaluate hypotheses. The two frameworks are considered significantly different if the *p*-value is smaller than the significance level. Larger *t*-values correspond to smaller *p*-values, indicating a greater difference between the two systems. This implies a discernible difference in retrieval performance between the proposed system and the benchmark method. Equations 12, 13, 14, and 15 are utilised to calculate the *t* and *p* values.

The average value of the sample:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (12)$$

Here, *N* represents the size of the sample, and x_i denotes the individual data points within the sample.

The standard deviation of the sample:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2} \quad (13)$$

In this context, *N* denotes the sample size, x_i represents the individual data points within the sample, and μ is the average value of the sample.

The *t*-value:

$$t = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \quad (14)$$

Here, $\underline{\mu_1}$ and $\underline{\mu_2}$ represent the means, σ_1^2 and σ_2^2 denote the variances, and N_1 and N_2 correspond to the sample sizes for the first and second systems being compared.

The degree of freedom (v) is calculated using the formula:

$$v = \frac{\left(\frac{\sigma_1^2}{N_1}+\frac{\sigma_2^2}{N_2}\right)^2}{\frac{1}{N_1-1}\left(\frac{\sigma_1^2}{N_1}\right)^2+\frac{1}{N_2-1}\left(\frac{\sigma_2^2}{N_2}\right)^2} \quad (15)$$

Here, σ_1^2 and σ_2^2 are the variances, and N_1 and N_2 are the sample sizes for the first and second systems being compared. To obtain the *p*-value from a statistical *t*-distribution table, both the calculated *t*-value and the degree of freedom are required.

4.3 Benchmark Methods

The benchmark methods employed in this study include the Fusion Method [15], ZM+LDP [42], and EZLH (colour-only) [43]. These approaches were chosen as each incorporates distinct criteria, enabling a clear evaluation of the features that enhance the proposed method.

5.0 RESULTS AND DISCUSSION

In this section, the outcomes of the proposed image retrieval framework are presented and analysed. The evaluation focuses on assessing the framework's retrieval performance using established metrics, such as Mean Average Precision (MAP) and Precision-Recall curves. Experiments were conducted to determine the effectiveness of the feature fusion approach and the robustness of the system in handling various challenges, including image transformations and low-resolution inputs. The findings are compared against benchmark methods to highlight the advantages and limitations of the proposed approach.

5.1 Optimal Image Size

An initial experiment was carried out to determine the optimal image size for use in the system. The framework of EZLH, being the first to be tested, was utilised for this purpose. The dataset images ranged in size from 20 × 20 pixels to 200 × 200 pixels. For the experiment, image sizes of 32 × 32 pixels, 64 × 64 pixels, 128 × 128 pixels and 256 × 256 pixels were selected. These sizes were chosen as they are commonly used in computer vision, balancing the need to avoid both excessive redundancy and loss of critical information from the original images.

System performance was evaluated using MAP, considering the top 10 retrieved results. The Fish4knowledge dataset, comprising natural images, was employed. Given that the smallest number of images in any category was 16, the top 16 images from each category were selected for both the dataset and query images, resulting in a total of 368 images. The results indicated that 64 × 64 pixels was the most suitable image size for this system.

5.2 Optimum Order of ZMInv

As each order of ZMInv contributes differently, an experiment was conducted to determine the optimal order value for use with the dataset images. The system's performance was assessed using MAP, and the same dataset used in the previous experiment to determine the suitable image size was employed. It was found that among the tested values, an order of four resulted in the highest retrieval accuracy, providing sufficient detail to accurately represent the images while maintaining manageable computational costs. This finding is significant as it underscores the importance of optimising feature extraction parameters to ensure a balance between performance and efficiency. Moreover, the ability to extract accurate shape features from fish images, characterised by distinct, yet sometimes subtle, contours, was crucial to achieving higher retrieval accuracy.

Fig. 4 presents the Average 11-Point Precision-Recall graph, while Table 1 shows the MAP values for each category and the overall performance of the proposed and benchmark methods. The results demonstrate that the proposed feature fusion method outperforms the benchmark approaches. The key to the superior performance of the proposed method lies in its fusion of multiple feature types: colour, texture, and both local and global shape information. As fish images are inherently rich in colour, relying only on texture and shape features would have been insufficient to represent the images fully. The role of colour is particularly important in natural images, where

variations in colour can distinguish between species and environmental contexts that are not always well-represented by shape and texture alone.

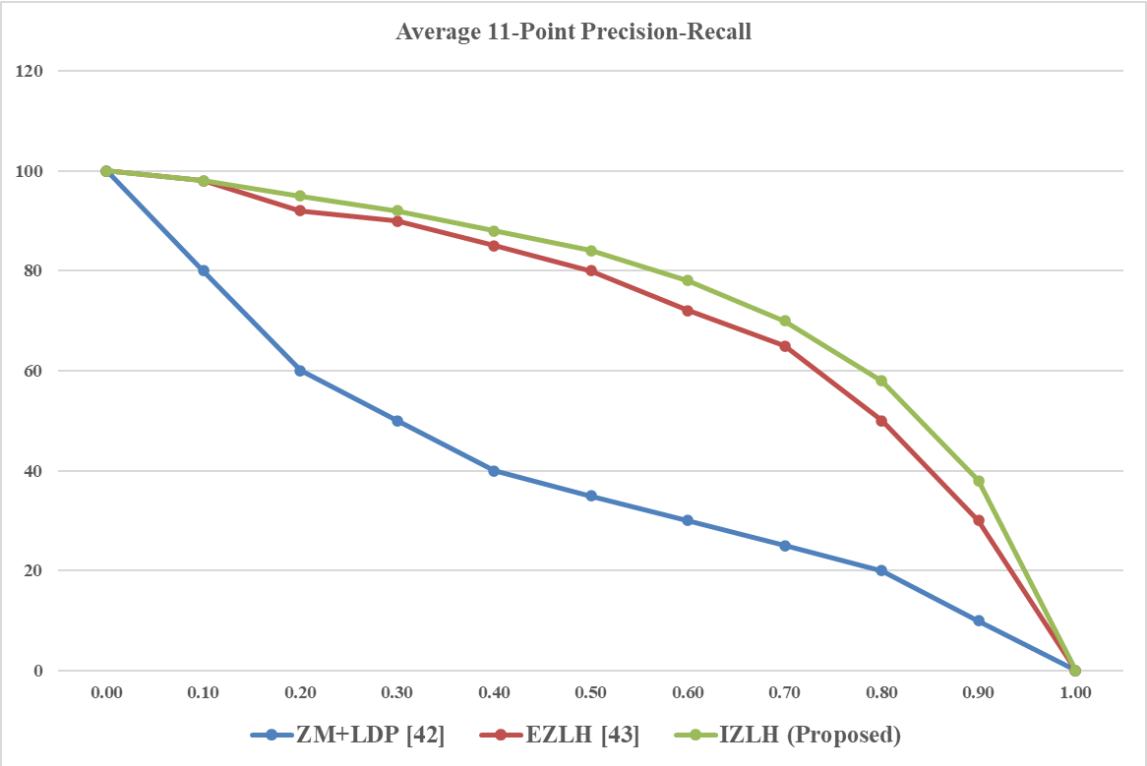


Fig. 4: Graphs of Average 11-Point Precision-Recall for the proposed and benchmark methods

The incorporation of colour features in EZLH and IZLH proved to be highly beneficial, as demonstrated by the results in Table 1. While ZM+LDP ignores colour entirely, which limits its ability to differentiate objects in highly detailed images, EZLH and IZLH utilise colour to provide a more discriminative feature set. This discriminative power stems from the fact that colour is one of the most noticeable and distinguishing attributes in natural fish images, where shape and texture may not be as clearly defined due to the fluid, dynamic nature of the subjects.

Additionally, by integrating LDP with colour features in HLDP, the system capitalised on LDP's robustness to noise and its ability to handle non-monotonic illumination variations. These attributes help preserve important features under different environmental conditions, such as variations in lighting or partial occlusion, which are common challenges when working with real-world images. LDP's ability to compute edge responses for each pixel in the momentgram (which encodes rotational, scaling, and translational invariance) enabled the system to maintain high accuracy even when the images underwent basic transformations.

The combination of these features which are colour, texture, and invariant shape provides a comprehensive representation of the images, significantly improving retrieval accuracy. This comprehensive approach allows the system to better handle the variability in the dataset, where fish images may differ substantially in terms of orientation, lighting, and scale.

Table 1: MAP values for each category and the overall system performance of the proposed and benchmark methods

Species	Fusion Method [15]	ZM+LDP [42]	EZLH [43]	IZLH (Proposed)
1	75.41	37.03	76.49	76.79
2	61.28	40.06	69.09	85.12
3	60.48	51.18	65.98	84.45
4	68.56	23.57	84.61	85.64
5	73.19	47.86	75.29	76.55
6	30.53	41.84	68.64	90.35
7	73.63	34.79	82.7	84.58
8	39.48	36.79	73.72	74.83

9	33.01	58.48	83.93	98.8
10	65.13	65.73	77.96	97.63
11	31.75	33.89	57.24	91.88
12	31.65	23.01	73.13	79.14

13	39.77	56.2	71.73	72.29
14	38.64	43.2	63.86	72.93
15	37.98	27.95	89.89	90.06
16	62.74	38.9	72.18	77.8
17	28.57	42.18	86.38	87.61
18	62.55	27.8	82.97	91.65
19	15.97	30.54	79.82	80.15
20	28.70	33.19	70.16	84.06
21	3.57	34.2	80.26	88.33
22	40.36	42.13	79.43	85.51
23	30.71	42.3	73.87	79.75
MAP	**44.94**	**39.69**	**75.62**	**84.17**

The addition of invariance to basic transformations such as rotation, scaling, and translation was shown to have a profound impact on retrieval accuracy. When applied to ZM+LDP, the accuracy improved by 44.48%, highlighting that the ability to handle transformations is critical in natural image retrieval tasks. Fish images are often captured under varying angles and scales, which makes invariance a crucial feature for accurate retrieval.

Interestingly, when colour features (EZLH) were added to ZM+LDP, the accuracy improved by 35.93%. This suggests that, while colour plays a vital role in distinguishing natural fish images, invariance to transformations has a more significant effect on performance. This insight reinforces the importance of addressing both the content (colour) and the context (invariance) in image retrieval systems. The results indicate that, for this specific dataset, handling basic transformations was a more critical factor for improving retrieval performance than the addition of colour information.

When examining the Fusion Method [15], it becomes evident that incorporating multiple feature types such as shape, texture, and colour into a single framework enhances the discriminative ability of the system. The method proposed by Mustaffa et al. also considered fusion of features, but with some key differences in implementation. The fusion method demonstrated strong performance by combining multiple features, but the addition of invariant features (such as those in IZLH) further improves the system's robustness to transformations.

The key distinction between our method and the Fusion Method [15] lies in the emphasis on rotation, scaling, and translation invariance, which, as shown in the results, has a more pronounced effect on accuracy. While the Fusion Method [15] offers a solid foundation for combining features, the integration of invariance to transformations, as demonstrated in IZLH, significantly boosts retrieval performance in dynamic and real-world scenarios, where images may vary in terms of orientation, scale, and resolution.

One important observation from the experiments is that the number of feature vectors (56) used in each framework did not significantly affect the overall system performance. This suggests that the quality of the features and the manner in which they are combined are far more important than the sheer quantity of the features extracted. By effectively combining colour, texture, and shape features, the system is able to represent the images in a way that maximises the discriminative power, regardless of the number of features considered.

The results presented here clearly show that the proposed feature fusion method, which integrates colour, texture, and invariant shape features, offers superior performance compared to the benchmark methods, including ZM+LDP and the Fusion Method [15]. The addition of invariance to transformations plays a crucial role in improving retrieval accuracy, particularly for dynamic images like those of fish, where changes in orientation, scale, and resolution are common. Colour features further enhance the system's performance, providing the necessary discriminative power to differentiate between images with subtle variations. The findings indicate that a holistic approach, incorporating both feature diversity and transformation invariance, is key to achieving high retrieval accuracy in real-world image datasets.

An analysis of sample images from the retrieval results offers further insight into the system's performance. Correctly retrieved images (Fig. 5, top row) demonstrate the framework's ability to handle challenges like variations in fish orientation, scale, and background clutter. In contrast, incorrectly retrieved images (Fig. 5, bottom row) often result from factors such as high inter-species visual similarity, partial occlusions, or inconsistent lighting conditions, which can affect feature extraction and similarity measurements. These examples provide valuable qualitative context to the retrieval performance, complementing the quantitative results discussed.

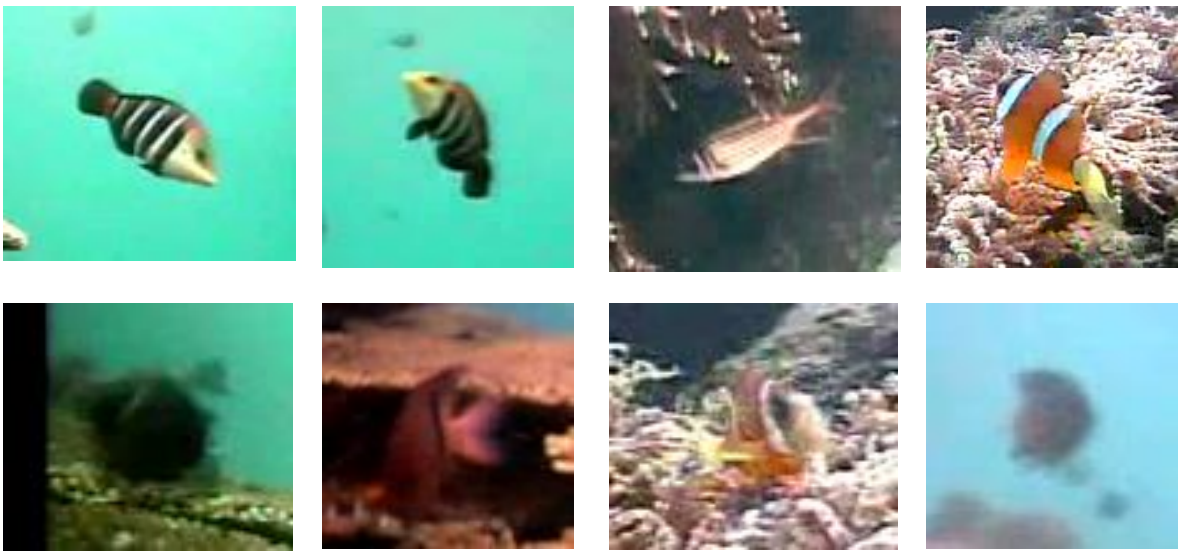


Fig. 5: Examples of correctly (top row) and incorrectly (bottom row) retrieved fish images. Correct retrievals show the system's robustness in handling varying orientations and background clutter, while incorrect retrievals highlight challenges such as inter-species visual similarity, occlusion, and lighting issues.

To assess the statistical significance of the differences between the frameworks, a Two-tailed paired *t*-test was conducted. Let the MAP value for each method be denoted as $F_m, where\ m \in \{\{ZM + LDP\}, \{EZLH\}, \{IZLH\}\}$. The test compares the mean values of two groups, with the null hypothesis H_0 stating the difference is not significant (i.e., $F_m > 0.05$ for similarity), and the alternative hypothesis H_1 suggesting a significant difference (i.e., $F_m < 0.05$ for non-similarity). A larger *t*-value indicates a more significant difference between the methods. Table 2 presents the results of the Two-tailed *t*-test, showing that IZLH demonstrates a statistically significant improvement in retrieval performance compared to Fusion Method [15], ZM+LDP [42] and EZLH [43].

Table 2: Results of the paired *t*-test at a significance level of 0.05. *p*-values less than 0.05 suggest statistically significant differences

Method	***t*-value**	***p*-value**	**Null Hypothesis (H_0)**	**Remark**
ZM+LDP vs. IZLH	11.1796	1.9357E-14	Reject	IZLH > ZM+LDP
Fusion Method vs. IZLH	6.8639	1.8105E-08	Reject	IZLH > Fusion Method
EZLH vs. IZLH	4.1574	1.4635E-04	Reject	IZLH > EZLH

6.0 CONCLUSION

This paper introduces a novel and robust descriptor for fish species image retrieval, which effectively integrates colour, texture, and both local and global shape features while maintaining invariance to basic transformations such as rotation, scaling, and translation. The primary achievement of this work lies in the development of a feature fusion method that enhances the representation of fish images, particularly in datasets where colour is a critical distinguishing characteristic. The proposed method outperforms traditional benchmark approaches, such as ZM+LDP, in terms of retrieval accuracy, achieving a MAP value of 84.17, highlighting the importance of combining multiple feature types to improve system performance. By leveraging the discriminative power of colour, texture, and invariant shape features, the system addresses the challenges posed by dynamic and varied fish images, which are often captured under different orientations and lighting conditions.

One key contribution of this paper is the effective incorporation of invariance to basic transformations, which has proven to be crucial in improving retrieval accuracy by 44.48%. The system's ability to handle variations in rotation, scaling, and translation significantly enhances its robustness. Moreover, the combination of colour features (via EZLH and IZLH) further boosts performance, as these features play a pivotal role in differentiating species in natural fish images.

In conclusion, this paper provides a significant advancement in the field of image retrieval, offering a comprehensive and efficient solution for fish species identification. The proposed method achieves higher retrieval

accuracy than existing methods, demonstrating the effectiveness of feature fusion and invariance to transformations in handling complex, real-world image datasets.

Future work will focus on integrating machine learning classifiers, such as support vector machines (SVMs), to enhance the system's ability to distinguish subtle differences between images by leveraging both low-level and high-level features. Additionally, the incorporation of deep learning techniques, particularly convolutional neural networks (CNNs), could further automate feature extraction and improve the system's robustness to a broader range of image variations.

Expanding the dataset to include more diverse fish species and environmental conditions will also be important for improving the system's generalisability. Furthermore, exploring motion-based features could enhance the retrieval accuracy, particularly for dynamic fish images captured in motion, providing a more comprehensive solution for image retrieval in real-world scenarios.

REFERENCES

[1] S. Modeel, P. Dolkar, S. Siwach, P. Yadav, and R. K. Negi, "Role of science and technology for sustainable aquaculture development and aquatic ecosystem management," in *Role of Science and Technology for Sustainable Future: Volume 1: Sustainable Development: A Primary Goal*, Singapore: Springer Nature Singapore, 2024, pp. 277–301.

[2] S. Taheri, J. C. D. Andrade, and C. A. Conte-Junior, "Emerging perspectives on analytical techniques and machine learning for food metabolomics in the era of industry 4.0: A systematic review," *Critical Reviews in Food Science and Nutrition*, pp. 1–27, 2024.

[3] J. Gladju, B. S. Kamalam, and A. Kanagaraj, "Applications of data mining and machine learning framework in aquaculture and fisheries: A review," *Smart Agricultural Technology*, vol. 2, p. 100061, 2022.

[4] T. Hasegawa and D. Nakano, "Robust fish recognition using foundation models toward automatic fish resource management," *Journal of Marine Science and Engineering*, vol. 12, no. 3, p. 488, 2024.

[5] S. I. Moghadasi, S. D. Ravana, and S. N. Raman, "Low-cost evaluation techniques for information retrieval systems: A review," *Journal of Informetrics*, vol. 7, no. 2, pp. 301–312, 2013.

[6] X. Li, J. Yang, and J. Ma, "Recent developments of content-based image retrieval (CBIR)," *Neurocomputing*, vol. 452, pp. 675–689, 2021.

[7] S. Mittal, S. Srivastava, and J. P. Jayanth, "A survey of deep learning techniques for underwater image classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 10, pp. 6968–6982, 2022.

[8] D. Srivastava, S. S. Singh, B. Rajitha, M. Verma, M. Kaur, and H. N. Lee, "Content-based image retrieval: A survey on local and global features selection, extraction, representation and evaluation parameters," *IEEE Access*, 2023.

[9] A. Saleh, M. Sheaves, and M. Rahimi Azghadi, "Computer vision and deep learning for fish classification in underwater habitats: A survey," *Fish and Fisheries*, vol. 23, no. 4, pp. 977–999, 2022.

[10] S. Hu, Z. Cheng, G. Fan, M. Gan, and C. P. Chen, "Texture-aware and colour-consistent learning for underwater image enhancement," *Journal of Visual Communication and Image Representation*, vol. 98, p. 104051, 2024.

[11] V. G. Mahesh, C. Chen, V. Rajangam, A. N. J. Raj, and P. T. Krishnan, "Shape and texture aware facial expression recognition using spatial pyramid Zernike moments and Law's textures feature set," *IEEE Access*, vol. 9, pp. 52509–52522, 2021.

[12] G. Sucharitha, N. Arora, and S. C. Sharma, "Medical image retrieval using a novel local relative directional edge pattern and Zernike moments," *Multimedia Tools and Applications*, vol. 82, no. 20, pp. 31737–31757, 2023.

[13] W. Li, Z. Du, X. Xu, Z. Bai, J. Han, M. Cui, and D. Li, "A review of aquaculture: From single modality analysis to multimodality fusion," *Computers and Electronics in Agriculture*, vol. 226, p. 109367, 2024.

[14] J. H. Dewan and S. D. Thepade, “Image retrieval using low-level and local features contents: A comprehensive review,” *Applied Computational Intelligence and Soft Computing*, vol. 2020, no. 1, p. 8851931, 2020.

[15] M. R. Mustaffa, N. M. Norowi, and S. M. Yee, "Content-Based Feature Fusion Representation for Marine Invertebrates," *Malaysian Journal of Computer Science*, vol. 33, no. 3, pp. 170–187, 2020.

[16] J. H. Dewan and S. D. Thepade, “Feature fusion approach for image retrieval with ordered colour means based description of keypoints extracted using local detectors,” *Journal of Engineering Science and Technology*, vol. 16, no. 1, pp. 482–509, 2021.

[17] K. V. Rani, M. E. Prince, P. S. Therese, P. J. Shermila, and E. A. Devi, "Content-based medical image retrieval using fractional Hartley transform with hybrid features," *Multimedia Tools and Applications*, vol. 83, no. 9, pp. 27217–27242, 2024.

[18] U. A. Khan, A. Javed, and R. Ashraf, "An effective hybrid framework for content-based image retrieval (CBIR)," *Multimedia Tools and Applications*, vol. 80, no. 17, pp. 26911–26937, 2021.

[19] Madhu and R. Kumar, "A hybrid feature extraction technique for content-based medical image retrieval using segmentation and clustering techniques," *Multimedia Tools and Applications*, vol. 81, no. 6, pp. 8871–8904, 2022.

[20] S. K. Chu, S. D. Ravana, S. S. Mok, and R. C. Chan, "Behavior, perceptions and learning experience of undergraduates using social technologies during internship," *Educational Technology Research and Development*, vol. 67, pp. 881–906, 2019.

[21] G. H. Liu and J. Y. Yang, "Exploiting deep textures for image retrieval," *International Journal of Machine Learning and Cybernetics*, vol. 14, no. 2, pp. 483–494, 2023.

[22] R. J. Veiga and J. M. Rodrigues, "Fine-grained fish classification from small to large datasets with Vision Transformers," *IEEE Access*, 2024.

[23] W. Wang, P. Jiao, H. Liu, X. Ma, and Z. Shang, "Two-stage content-based image retrieval using sparse representation and feature fusion," *Multimedia Tools and Applications*, vol. 81, no. 12, pp. 16621–16644, 2022.

[24] Z. Ju and Y. Xue, "Fish species recognition using an improved AlexNet model," *Optik*, vol. 223, p. 165499, 2020.

[25] C. N. Silva, J. Dainys, S. Simmons, V. Vienožinskis, and A. Audzijonyte, "A scalable open-source framework for machine learning-based image collection, annotation and classification: A case study for automatic fish species identification," *Sustainability*, vol. 14, no. 21, p. 14324, 2022.

[26] Y. Tadepalli, M. Kollati, S. Kuraparthi, P. Kora, A. K. Budati, and L. Kala Pampana, "Content-based image retrieval using Gaussian–Hermite moments and firefly and grey wolf optimization," *CAAI Transactions on Intelligence Technology*, vol. 6, no. 2, pp. 135–146, 2021.

[27] W. Zhang, L. Zhou, P. Zhuang, G. Li, X. Pan, W. Zhao, and C. Li, "Underwater image enhancement via weighted wavelet visual perception fusion," *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.

[28] K. Panetta, L. Kezebou, V. Oludare, and S. Agaian, "Comprehensive underwater object tracking benchmark dataset and underwater image enhancement with GAN," *IEEE Journal of Oceanic Engineering*, vol. 47, no. 1, pp. 59–75, 2021.

[29] G. Hassan, K. M. Hosny, R. M. Farouk, and A. M. Alzohairy, "An efficient retrieval system for biomedical images based on radial associated Laguerre moments," *IEEE Access*, vol. 8, pp. 175669–175687, 2020.

[30] S. Fadaei and A. Rashno, "Content-based image retrieval speedup based on optimized combination of wavelet and Zernike features using particle swarm optimization algorithm," *International Journal of Engineering*, vol. 33, no. 5, pp. 1000–1009, 2020.

[31] R. Bibi, Z. Mehmood, R. M. Yousaf, T. Saba, M. Sardaraz, and A. Rehman, "Query-by-visual-search: Multimodal framework for content-based image retrieval," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, pp. 5629–5648, 2020.

[32] M. Zhao, J. Liu, Z. Zhang, and J. Fan, "A scalable sub-graph regularization for efficient content-based image retrieval with long-term relevance feedback enhancement," *Knowledge-Based Systems*, vol. 212, p. 106505, 2021.

[33] S. M. Alarcão, V. Mendonça, C. Maruta, and M. J. Fonseca, "ExpertosLF: Dynamic late fusion of CBIR systems using online learning with relevance feedback," *Multimedia Tools and Applications*, vol. 82, no. 8, pp. 11619–11661, 2023.

[34] Z. Yang, J. Li, T. Chen, Y. Pu, and Z. Feng, "Contrastive learning-based image retrieval for automatic recognition of in situ marine plankton images," *ICES Journal of Marine Science*, vol. 79, no. 10, pp. 2643–2655, 2022.

[35] R. B. Fisher, Y. H. Chen-Burger, D. Giordano, L. Hardman, and F. P. Lin, Eds., *Fish4Knowledge: Collecting and Analyzing Massive Coral Reef Fish Video Data*, vol. 104, p. 319. Berlin/Heidelberg, Germany: Springer, 2016.

[36] B. J. Boom, P. X. Huang, J. He, and R. B. Fisher, "Supporting ground-truth annotation of image datasets using clustering," in Proc. 21st Int. Conf. Pattern Recognit. (ICPR), Nov. 2012, pp. 1542–1545.

[37] G. A. Papakostas, D. E. Koulouriotis, E. G. Karakasis, and V. D. Tourassis, "Moment-based local binary patterns: A novel descriptor for invariant pattern recognition applications," *Neurocomputing*, vol. 99, pp. 358–371, 2013.

[38] H. Shu, H. Zhang, B. Chen, P. Haigron, and L. Luo, "Fast computation of Tchebichef moments for binary and grayscale images," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3171–3180, 2010.

[39] I. Park and S. Kim, "Performance indicator survey for object detection," in *2020 20th International Conference on Control, Automation and Systems (ICCAS)*, Oct. 2020, pp. 284–288.

[40] A. Moffat, "Batch evaluation metrics in information retrieval: Measures, scales, and meaning," *IEEE Access*, vol. 10, pp. 105564–105577, 2022.

[41] P. M. Cahusac and S. E. Mansour, "Estimating sample sizes for evidential t tests," *Research in Mathematics*, vol. 9, no. 1, p. 2089373, 2022.

[42] A. Goyal and E. Walia, "Variants of dense descriptors and Zernike moments as features for accurate shape-based image retrieval," *Signal, Image and Video Processing*, vol. 8, pp. 1273–1289, 2014.

[43] N. S. Osman, M. R. Mustaffa, S. C. Doraisamy, and H. Madzin, "Content-based image retrieval for fish based on extended Zernike moments-local directional pattern-hue colour space," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, pp. 173–183, 2019.