A Smart Content-Based Image Retrieval Approach Based on Texture Feature and Slantlet Transform

Original Scientific Paper

Hakeem Imad Mhaibes
Middle Technical University,
Kut Technical Institute
Baghdad, Iraq
hakeem.emade@mtu.edu.iq

Qahtan Makki Shallal
Management Technical College of Basra
Basrah, Iraq
qahtan.makii@stu.edu.iq

May Hattim Abood
Al Iraqia University,
Computer Engineering Department, College of Engineering,
Baghdad, Iraq
may.hattim@gmail.com

Abstract *– With the advancement of digital storing and capturing technologies in recent years, an image retrieval system has been widely known for Internet usage. Several image retrieval methods have been proposed to find similar images from a collection of digital images to a specified query image. Content-based image retrieval (CBIR) is a subfield of image retrieval techniques that extracts features and descriptions content such as color, texture, and shapes from a huge database of images. This paper proposes a two-tier image retrieval approach, a coarse matching phase, and a fine-matching phase. The first phase is used to extract spatial features, and the second phase extracts texture features based on the Slantlet transform. The findings of this study revealed that texture features are reliable and capable of producing excellent results and unsusceptible to low resolution and proved that the SLT-based texture feature is the perfect mate. The proposed method's experimental results have outperformed the benchmark results with precision gaps of 28.0 % for the Caltech 101 dataset. The results demonstrate that the two-tier strategy performed well with the successive phase (fine-matching) and the preceding phase (coarse matching) working hand in hand harmoniously.*

Keywords*: Image Processing, Information retrieval, CBIR, Slantlet Transform, Features extraction, Similarity measure*

1. INTRODUCTION

Today Due to the information explosion on the Internet, an image retrieved system has become a well-known method used to seek the most similar images for a given specified query image from a large database of images [1]. As a result, one of the most difficult problems in computer vision is retrieving and managing desired images from large image databases [2, 3]. Researchers have recently focused on difficult problems in CBIR in various domains such as machine learning, pattern recognition, and computer vision, among others. Generally, there are two classes of image retrieved methods. Text-based image retrieval (TBIR) and CBIR [3, 4].

In TBIR, images are denoted by headings and keywords in the database, and the manual text is used to search and retrieve images depending on users labeling ways [5]. The limitations of TBIR are (1) inaccurate results because users wrongly annotate images, (2) the overall image information cannot be described by a single keyword, and (3) it is a time-consuming process since it requires manual annotation of the images [6-8]. To overcome the limitations of TBIR, a new CBIR method of image retrieval was invented by a researcher. CBIR extracts image features and descriptions content such as color, texture, and shapes from images to find similarities from a huge database [1, 9]. The general structure of CBIR is illustrated in Fig1.

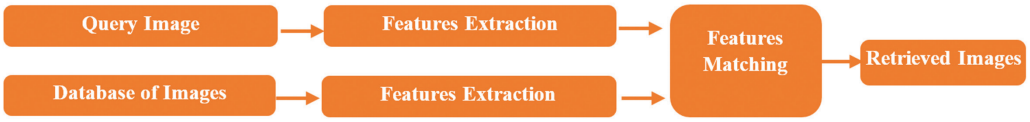

Fig. 1. The general structure of CBIR

CBIR consists of two key steps: feature extraction and matching [10]. At first, features for both the query image and database images are extracted and stored in features vectors format and then matched to find similar images using a predefined distance measurement formula. According to the least similarity index, the matched images are ranked [11]. The main objective of the present work is to develop a new method for image retrieval with higher accuracy compared to the state of art technique by introducing a new technique consisting of two phases, coarse matching, and good matching, to extract useful texture features from both spatial and frequency domains to represent image characteristics and find the closest matching image. In the first phase, a shortlist of thirty images closely resembling the query image is produced. The matching process is further refined at the fine-matching phase, where the shortlist will be used as the feeder or input. In the end, the closest matching image will prevail.

2. RELATED WORKS

Several CBIR algorithms using various feature extraction methods have been presented for image retrieval applications. Each technique competes to locate the best similar photos to the query image with the highest precision rate [1]. This section examines state-of-the-art CBIR techniques currently in use.

In 2016, S. Somnugpong et al. proposed a new CBIR that prioritizes spatial information in images by combining color correlograms and the Edge direction histogram. In the case of the same image but different colors, a color correlogram will treat information about spatial color correlation, whereas EDH will provide geometry information. Simple calculations, such as the Euclidean Distance between the query image and the image in the database, are used to evaluate performance. This work proves that containing the combination of image texture and spatial correlation of pairs of color in feature provides more robustness to a high changing image. Although using color correlation remains some limitation in the color term, EDH will fulfill the image semantic in terms of texture and geometry. Accuracy achieved in this method reached 73 % [12].

In 2018, Nazir et al. proposed a novel colour and texture-based CBIR technique. Local features such as the discrete wavelet transform and colour histogram features are extracted locally, whilst global features such as the edge histogram descriptor are extracted globally. As a result, Manhattan distance measurement was used to compare the feature vectors of the query image and the feature vectors of the images in the dataset. According to the study's findings, it has a 73.5 percent accuracy rate [13].

In 2019, S. P. Rana et al. presented the CBIR method, which combines three invariant features and is implemented across five different picture databases, achieving substantial precision values compared to previous techniques. It employs a feature vector of 247, which is relatively long compared to other algorithms and may result in increased time complexity. As a result, reducing the length of feature vectors will be a future issue because time is also a critical attribute of CBIR systems. Another consideration would be the purification of obtained results through relevant feedback (RF). The query results would be filtered through human intervention, and the feedback would be fed back to the system for better outcomes in the next iteration. Due to the large feature vector dimension, the suggested approach has a high computational cost. According to the outcomes of the study, the recommended approach had a 65 percent accuracy rate [14].

In 2019 Q. Zheng and colleagues developed a new end-to-end CBIR based on a deep convolutional neural network (CNN) and a differential learning algorithm. Compared to the standard method of matching the image by employing deep convolutional activation features as the feature vector. They make the method's retrieval process easier and reduce the problem of semantic gaps. In the dataset manufacturing stage, the approach creates an image matching dataset based on the gravitational field model, which includes adding the similarity score label for each image. As the number of returned images increases, the retrieval accuracy of the system decreases slightly and eventually becomes stable at a high value. The limitation of their work is that more enhancement is required in terms of the speed of the training and testing stages and more time to construct the gravitational field database. The retrieval accuracy for this strategy is 88.5 percent on average [15].

In 2020, Mohammed Q et al. proposed a new method for CBIR. Three descriptors developed in this approach to increase matching image accuracy for medical images retreival. In this approach, student or a medical specialist is assumed to be a user and the query is a medical image. The average matching accuracy was 70 % [16].

In 2020, Akshara et al. developed a novel CBIR approach that based on coarse-to-fine progressive RS image descriptor and decoded in partical compressed JPEG 2000 domain. At first, decodes the code-blocks that associated only to the coarse wavelet resolution, to discards the most dissimilar images to the query based on the similarity measurement. Following that,

the code-blocks connected with the remaining images' sub-sequent resolution are decoded, and the most irrelevant images are removed using computin similarities based on image attributes associated with both resolutions. This is accomplished by the use of the pyramid match kernel similarity measure, which gives higher weights to traits associated with finer wavelet resolution than to those associated with coarse wavelet resolution [17].

In 2021, Mounika proposed a new CBIR based on integrated machine learning and convolution neural network and incorporate with principal component analysis (PCA) for extracting salient features from the images. The proposed method focused on artificial intelligent based deep learning approach to estimate semantic features. It uses Euclidean distance as a similarity measurement [18].

In 2021, Ali Ahmed and Sara Muhamed suggested a new method for CBIR. In this research, a set of color and texture attributes is proposed to be employed in both fusion procedures. The early fusion stage uses an early combination of eighteen color characteristics and twelve texture features to create a single vector representation, while the late fusion stage uses three of the most popular distance measures. Our results on two typical picture datasets show that our suggested method has good retrieval results when compared to the traditional method of employing single features descriptors, as well as acceptable retrieval results when compared to various state-of-the-art methods. For the Corel-1K and GHIM-10K datasets, our suggested technique has an overall accuracy of 60.6 percent and 39.07 percent, respectively [19].

3. THE AIM AND OBJECTIVE OF THE STUDY

The goal of this work is to offer a novel CBIR approach that is more accurate than the current state of the art. As a result, two key objectives must be satisfied.

1. Represent picture properties that involves extracting effective texture features from both the spatial and frequency domains.
2. Create a two-tier picture retrieval approach that incorporates both spatial and frequency-based data.

4. RESEARCH METHODOLOGY

CBIR is influenced by various factors, including feature extraction strategy, utilization of appropriate features, similarity measurement approach, mathematical conversion, and reaction mechanisms. In CBIR, all of these elements are crucial. Improving some of the prompting components can result in a more effective retrieval process [20].

4.1. FEATURE EXTRACTION

Feature extraction is a fundamental step in a CBIR system. Based on low- and high-level features, images represent their features. Color, shape, and texture are low-level features, while high-level features are based on machine learning techniques [21]. One of the most important advantages of color-space conversion is that it allows choosing the most appropriate among the various color-spaces (RGB). The majority of color spaces are three-dimensional. HSV and YCbCr have a higher inherent value than RGB and are more similar to human perception. This study uses the Y and H channels from YCbCr and HSV, respectively. Features extraction perceives the luminance component in a better way than the chrominance component [22]. The formula for calculating the Y, Cb, and Cr channels from R, G, and B channels is defined by (1).

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.177 \\ -0.299 & -0.587 & 0.886 \\ 0.701 & -0.587 & -0.114 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} 16 \\ 128 \\ 128 \end{pmatrix} \quad (1)$$

For HSV color conversion, the dominant spectral component, red, blue, or yellow, is represented by the hue (H). When white is added to a pure color, the color changes; the less white, the more saturated the color becomes. This is the saturation level (S). The brightness of color is represented by the value (V). The following formulas can be used to convert a pixel's HSV values from its RGB representation:

$$H = arctan \frac{\sqrt{3}(G - B)}{(\mathrm{R} - \mathrm{G}) + (\mathrm{R} - \mathrm{B})} \quad (2)$$

$$S = 1 - \frac{min\{\mathrm{R}, \mathrm{G}, \mathrm{B}\}}{V} \quad (3)$$

$$V = \left(\frac{\mathrm{R} + \mathrm{G} + \mathrm{B}}{3}\right) \quad (4)$$

The texture is easy to recognize but more difficult to define. It is orthogonal to color because it is normally defined solely by grey levels. Smoothness, coarseness, and regularity are all texture characteristics [23]. As a result, four texture features are computed and extracted from each block of the Y and H images in this study: variance, coefficient of variation, energy, and entropy. The variance represents the deviation of each pixel in the block from the mean and is calculated using Eq. 5;

$$\sigma^2 = \frac{1}{W \times W} \sum_{i=1}^{W} \sum_{j=1}^{W} (I(i,j) - m)^2 \quad (5)$$

Where W represents the size of the block, $I(i,j)$ is image intensity value at the pixels i and j, and m is the mean intensity level calculated by the equation;

$$m = \frac{1}{W \times W} \sum_{i=1}^{w} \sum_{j=1}^{W} I(i,j) \quad (6)$$

The coefficient of variation is used to compare the degree of variation from one pixel to the next. It is defined as the standard deviation divided by the mean. Coefficient of variation measures the relative dispersion defined such as:

$$\eta = \frac{\sigma}{m} \quad (7)$$

Energy calculates the brightness value. A brighter pattern will have greater energy than a darker one. It is given as;

$$E = -\frac{1}{W \times W}\sum_{i=1}^{W}\sum_{j=1}^{W}\left(I(i,j)\right)^2 \quad (8)$$

Shannon entropy is used in this proposed algorithm to estimate the average amount of data contained in each block, given as;

$$\mathrm{E}_s = \frac{1}{W \times W}\sum_{i=1}^{W}\sum_{j=1}^{W}\left(I(i,j)\right)^2 log\left(I(i,j)\right)^2 \quad (9)$$

4.2. SLANTLET TRANSFORM (SLT)

The execution of the discrete wavelet transform (DWT) is based on filter bank iteration, although the processing time is greater because DWT is not well adapted to capture all of the information in a signal when the optimal time localization is unclear. Selesnick [24] proposed a new technique for this purpose, a Slantlet Transform (SLT), which gives a new parameter that is orthogonal to DWT [25].

A 2D SLT decomposition separates an image into four components, as shown in Fig. 2, High-High (HH), Low-Low (LL), High-Low (HL), and Low-High (LH), where H and L represent the high and low-frequency bands, respectively.

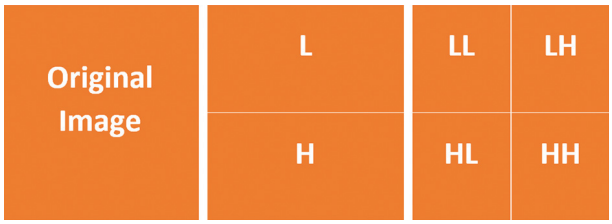

Fig. 2. 2D SLT scheme for splitting an image

The innovative image information is stored in the image's low-frequency band, denoted as LL. The HH, LH, and HL high- and medium-frequency bands, on the other hand, communicate information about the image's contour, edge, and other features. High coefficients characterize the image's important information. Meanwhile, the insignificant coefficients are brushed aside as meaningless information or background noise. In order to acquire the best outcomes in subsequent procedures, these small coefficients should be neglected.

The SLT method is commonly applied to the gray-scale channel; however, in this work, the SLT is applied to each RGB channel independently. An image is separated into three independent channels, R, G, and B, before being decomposed and partitioned into 32x32 pixel blocks for each channel - the block size was chosen empirically through a series of trials ranging from 8x8 to 64x64 pixels. Then, the SLT is used to bring out the texture features in each block [26].

4.3. QUALITY MEASUREMENT

The quantitative assessment is performed based on of three parameters, accuracy, precision, and recall, see following equations;

$$accuracy = \frac{\mathrm{TN} + \mathrm{TP}}{\mathrm{TN} + \mathrm{TP} + \mathrm{FN} + \mathrm{FP}} \quad (10)$$

$$\mathrm{precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad (11)$$

$$\mathrm{recall} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \quad (12)$$

The precision metric indicates the number of recovered images relevant to the query in the initial evaluation of an image retrieval system. The recall is the number of successfully retrieved images related to the query [27]. The suggested retrieval method's correctness is specified as the accuracy metric. Concerning that, a confusion matrix and its measuring parameters' definition are given in Table 1.

Table 1. Definition of the measuring parameters

Parameter	Definition
TP (True Positive)	Related images correctly retrieved by the query.
FN (False Negative)	Related images were wrongly rejected by the query.
FP (False Positive)	Not related images were wrongly retrieved by the query.
TN (True Negative)	Not related images were correctly rejected by the query.

4.4. DATASET

The standard datasets used to validate the proposed image retrieval technique are WANG V1.0, WANG V2.0 and Caltech 101. About 40 to 800 images per category and each picture is approximately 300 × 200 pixels in size. Description of the used datasets as shown in Table 2.

4.5. FEATURE MATCHING

There are several numbers of similarity measurements methods that can be used to compute the similarity between a query image and images in the database. In this work, Euclidean Distance is used to compute color, shapes, and texture features. The following is how the Euclidean Distance is calculated [28].

$$d(x_1, x_2) = \sqrt{\sum_{i=1}^{i=n}(x_1(i) - x_2(i))^2} \quad (13)$$

Where $x1$ is the feature vector of query image, and $x2$ is the feature vector of images in database.

5. PROPOSED METHOD

This section provides detailed discussions on the proposed CBIR method, as shown in Figure 3. It involves two main phases:

(1). Coarse Matching Phase: where a shortlist of images will be selected from the dataset that have been considered to match with the query image, and choose 30 the closest-match images

(2). Fine Matching Phase: In this stage, the query image will be matched with the shortlist's images to find the closest match images.

5.1. COARSE-MATCHING PHASE

This phase determines a shortlist from which the closest-match image is drawn. It is the same as removing all dissimilar photos from the collection and leaving just those similar resembling the query image. Algorithm 1 and Fig. 4 (a) illustrated the main steps of this phase.

Algorithm 1: Coarse-Matching Phase

1. Let qi be an image.
2. Convert colour image RGB of q into YCbCr using Eq. (1), and into HSV using Eq. (2), Eq. (3), and Eq. (4).
3. Select Y and H from YCbCr and HSV, respectively.
4. Split the Y and H into n blocks of 128x128 pixels, respectively. Where n depends on image size.
5. For each block of Y, calculate Variance f_1, Coefficient of variance f_2, Entropy f_3 and Energy f_4.
6. For each block of H, calculate Variance f_5, Coefficient of variance f_6, Entropy f_7 and Energy f_8.
7. Concatenate all the features to form a block-based feature vector $Fi=\{f\_1, f\_2, \ldots, f\_8\}$.
8. Repeat Step 5 to Step 7 until all blocks are exhausted.
9. Upon completion, concatenate all block-based feature vectors to form an image-based feature vector of the query image $F_{total}=\{F_1, F_2, \ldots, F_m\}$.
10. Repeat 1 to 9 for query image to build its feature vector.
11. Use Eq. (13) to calculate the similarity between the feature vector of query image and database images to produce distances $d_{coarse}=\{d_1, d_2, \ldots, d_n\}$.
12. Arrange the distances in ascending order (i.e. the smallest distance at the top and the biggest one at the bottom) and pick the top 30 images to form a shortlist.

Table 2. The Datasets Used

Datasets	No. of Images	No. of Categories	No. of images per category	Image size	No. of blocks	No. of features
WANG V1.0	1000	10	100	256x384 or 384*256	6	48
WANG V2.0	10000	100	1000	128*256 or 256*128	2	16
Caltech 101	9146	101	40-800	300*200 or 200*300	4	32

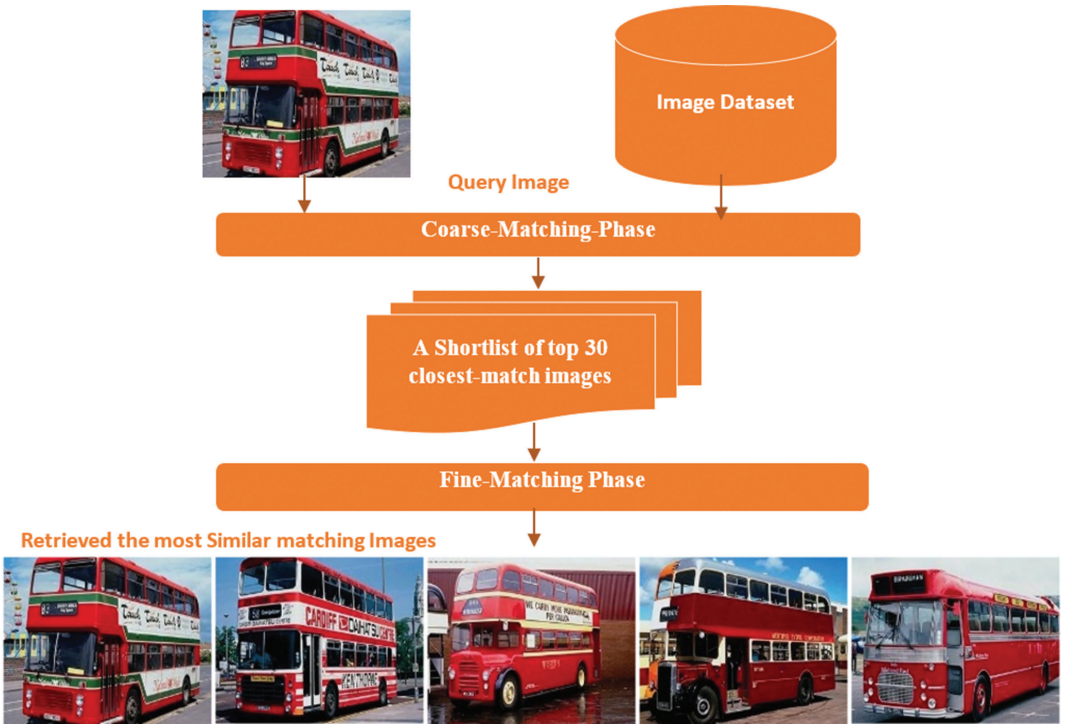

Fig. 3. Proposed diagram of Coarse-to-Fine matching image retrieval based on Slantlet transform.

5 .2. FINE-MATCHING PHASE

In this phase, a shortlist of images is used instead of the entire image collection. This speed up the CBIR. Algorithm 2 and Figure 4 (b) illustrated this phase.

This study uses the Slantlet transform. Afterwards, three features, mean, standard deviation and maximum value, are calculated. Algorithm 2 illustrates steps of the fine-matching;

Algorithim 2: Fine-matching phase

1. Let q_i be an image (RGB).
2. Partition each R, G and B channel into n blocks of 8x8 pixels.
3. For each block of R, decompose the block using SLT and calculate mean f_1, standard deviation f_2 and maximum f_3.
4. For each block of G, decompose the block using SLT and calculate mean f_4, standard deviation f_5 and maximum f_6.
5. For each block of B, decompose the block using SLT and calculate mean f_7, standard deviation f_8 and maximum f_9.
6. Concatenate all the features to form a block-based feature vector $F_i=\{F_1, F_2,\ldots, F_n\}$.
7. Repeat Step 3 to Step 9 until stop condition.
8. Upon completion, concatenate all block-based feature vectors to form an image-based feature vector of the query image $F_{total}=\{F_1, F_2,\ldots, F_m\}$.
9. Repeat 1 to 8 for query image to build its feature vector.
10. Use Eq. (13) to calculate the similarity between the feature vector of query image and shortlist images obtained from first phase to produce 30 distances $d_{fine}=\{d_1, d_2,\ldots, d_{30}\}$.
11. Choose the image with the least distance as the closest match.

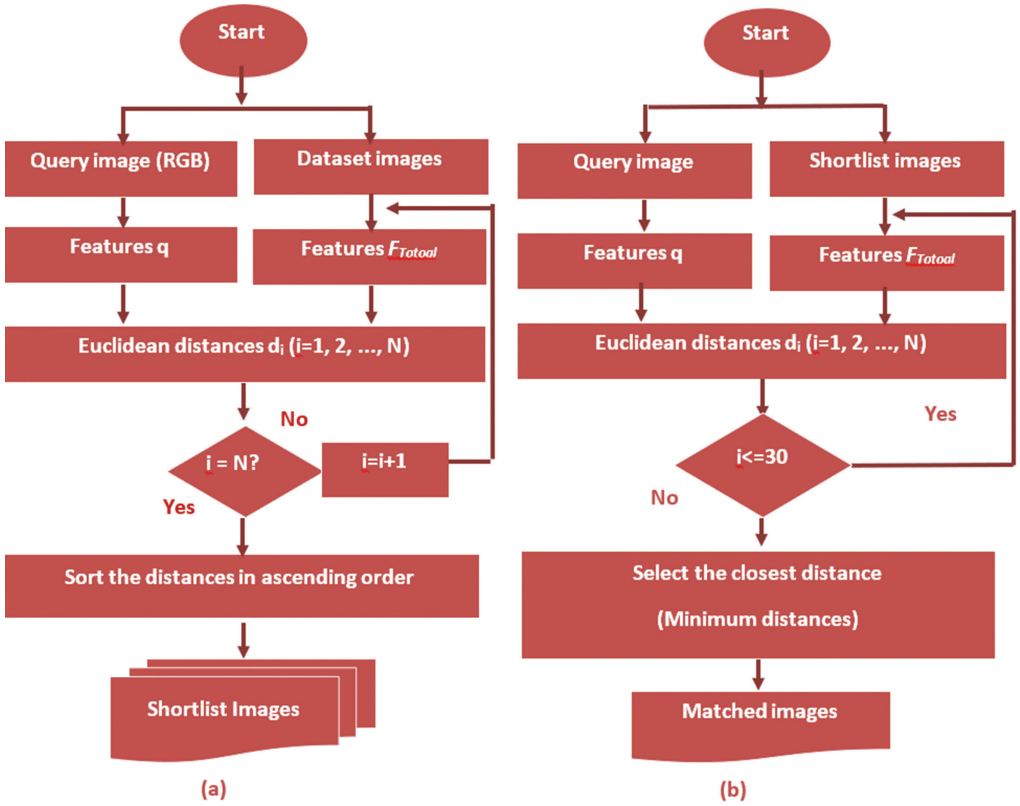

Fig.4. (a) The Coarse-Matching Phase, (b) Fine-Matching Phase

6. EXPERMINTAL RESULTS AND DISCUSSION

MATLAB (R2020a) was employed as a software platform programming on Windows 10 in this study. The tests are performed on an Acer Intel(R) Core (TM) i3-2310M CPU (2.10 GHz) with 4 GB RAM and a 1 TB hard drive. Several studies were carried out in this regard, utilizing three well-known datasets (WANG V1.0, WANG V2.0 and Caltech 101). Furthermore, the proposed approach focuses on precision and accuracy to assess the value of the recommended technique's high accuracy and consistency under standard image retrieval processes. The findings of the performance evaluation are presented in the form of tables.

6. 1. RESULT OF COARSE-MATCHING PHASE

This study arbitrarily chose one query image from each class found in the datasets. The results are arranged and presented class by class. Due to limited spaces, only ten classes from each dataset are carried out.

For WANG V1.0, classes are (Africa, Beach, Monuments, Buses, Dinosaur, Elephant, Roses, Horse, Mountain and Food). For WANG V2.0, classes are carried out (Shuttle, Building, Sky, Nightfall, Cowboys,

Flag, Mountains, desert, boat and Forest). Moreover, for Caltech 101, 101 queries have been made, and the ten classes are (Accordion, Cellphone, Chair, Chandelier, Cougar_face, Crab, Crocodile_head, Cup, Dalmatian, Dollar_bill and Airplanes).

Table 3, 4, and 5 provides the accuracy and precision measures for the first phase using WANG V1.0, WANG V2.0 and Caltech 101 datasets respectively.

Table. 3. Accuracy and precision using Coarse-Matching For WANG V1.0.

Classes	Accuracy	Precision
Africa	0.8794	0.3828
Beach	0.8739	0.3504
Monuments	0.8829	0.4029
Buses	0.8905	0.4469
Dinosaur	0.9897	0.9532
Elephant's	0.8748	0.3561
Roses	0.9098	0.5540
Horse	0.9275	0.6482
Mountains	0.8801	0.3870
Food	0.8642	0.2927
Average	0.8973	0.4774

Based on Table 3, it can be seen that the suggested approach achieved high accuracy and an acceptable level of precision, with averages of around 0.90 and 0.48. In comparison to the other classes, the Dinosaur class achieved the greatest precision (0.95). This is not surprising since the photographs in this category have a smooth background that contrasts with the item or foreground. In contrast, the Food class scored the lowest accuracy, below 0.30, because to the structure of the query picture, which comprises several little things of varied shapes and hues that form an uneven-textured surface. Ironically, one notable aspect of the data is that the accuracy is always more than 0.86 despite the precision being much lower (except for the Dinosaur class). This is because the coarse matching approach created a large number of false positives (FP), the irrelevant photos incorrectly obtained by the query.

Overall, the suggested technique obtained greater accuracy and precision, which averaged around 0.95 and 0.45 for the one hundred classes in the total dataset, respectively. In connection to this, the Flag class achieved the greatest levels of accuracy (0.98 and precision (0.84)) in comparison to the other classes. Contrarily, the Cowboys class obtained the lowest accuracy, less than 0.30. This is because of the structure of the query picture, which comprises several items of varied forms and hues, resulting in a patchy texture surface. Consistent with the results of the Wang V1.0, the accuracy has never been greater than the precision. This demonstrated that the coarse matching method created a substantial number of false positives (FP).

Table. 4. Accuracy and precision using Coarse-Matching for WANG V2.0.

Classes	Accuracy	Precision
Shuttle	0.9317	0.3764
Building	0.9322	0.3819
Sky	0.9384	0.4425
Nightfall	0.9328	0.3874
Cowboys	0.9231	0.2915
Flag	0.9815	0.8396
Mountains	0.9277	0.3378
Desert	0.9536	0.5863
Boat	0.9335	0.3942
Forest	0.9443	0.4982
Average	0.9510	0.4480

Table. 5. Accuracy and precision using Coarse-Matching For Caltech 101.

Classes	Accuracy	Precision
Accordion	0.9753	0.5265
Cellphone	0.9803	0.6497
Chair	0.9714	0.5132
Chandelier	0.9581	0.5906
Cougar_Face	0.9661	0.4800
Crab	0.9676	0.5326
Crocodile	0.9757	0.4868
Cup	0.9770	0.5751
Dalmatian	0.9663	0.4681
Dollar_Bill	0.9776	0.5458
Average	0.9670	0.5368

Based on Table 5, the coarse matching achieved better accuracy and precision results, averaging around 0.967 and 0.537 respectively, for the entire dataset. In the same context, based on the results of Precision eleven queries, which represent 10% of total queries, three classes, the Cougar face, Crocodile head and Dalmatian, have attained lower precision, which is below 0.50, compared to the rest. Because of the features shows that these images share a common trait (i.e. complex background), which naturally leads to uneven-texture surfaces. Conversely, the Cellphone class, which has clear contrast that separates the cellphone from its background, has outper-

formed the rest in terms of accuracy and precision. Again, the accuracy is always higher than the precision in all cases. This undoubtedly reaffirmed the findings that the coarse matching technique generated a lot of FP.

6.2. RESULT OF FINE-MATCHING PHASE

This section provides results of the fine-matching phase to evaluate the performance. The findings are then compared to the coarse matching results to see how effectively it reduces FP concerns and improves precision. Again, the experimental findings are provided in the following order: Wang V1.0, then Wang V2.0, and finally Caltech 101 datasets. At the conclusion of the fine matching, the query should return just the image with the nearest proximity among all of the images in the shortlist generated by the coarse matching.

The aforementioned 10 classes of the WANG V1.0 dataset are compared in terms of average precision in Table 6.

Table. 6. Performance study of recovered images acquired by coarse matching and fine matching techniques for WANG V1.0.

Classes	Accuracy	Precision
Africa	0.3828	0.672
Beach	0.3504	0.6113
Monuments	0.4029	0.7607
Buses	0.4469	0.873
Dinosaur	0.9532	1
Elephant's	0.3561	0.7643
Roses	0.554	0.9667
Horse	0.6482	0.9333
Mountains	0.387	0.6817
Food	0.2927	0.6243
Average	0.4774	0.7887

Table 6 demonstrates unmistakably that the precision produced by fine matching is far superior than the precision obtained by coarse matching, with an average of 0.79, while retaining a high level of accuracy. The dinosaur received flawless scores for both accuracy and precision. The accuracy of the remaining classes ranged between 26.0 and 42.0 percent. Fine texture details uncovered by the Slantlet transform are principally responsible for the substantial gain in accuracy.

Similar to the preceding Wang V1.0, a total of one hundred inquiries were conducted for Wang V2.0, consistent with earlier efforts. Performance metrics are provided in Table 7.

The results clearly show that the precision attained by fine matching is significantly better than that of the coarse matching with an average of around 0.79.

Similarly, for Caltech 101. The results of precision, which measure the performance of the fine matching relative to the coarse matching, are given in Table 8.

Table. 7. Performance study of recovered images acquired by coarse matching and fine matching techniques for WANG V2.0.

Classes	Accuracy	Precision
Shuttle	0.3764	0.588
Building	0.3819	0.773
Sky	0.4425	0.8083
Nightfall	0.3874	0.895
Cowboys	0.2915	0.6383
Flag	0.8396	1
Mountains	0.3378	0.7927
Desert	0.5863	0.9543
Boat	0.3942	0.6473
Forest	0.4982	0.7807
Average	0.4536	0.7878

Table. 8. Performance study of recovered images acquired by coarse and fine matching phases for Caktech

Classes	Accuracy	Precision
Accordion	0.5265	1
Airplanes	0.6009	1
Beaver	0.4898	0.8817
Binocular	0.7324	0.8903
Bonsai	0.417	0.915
Brontosaurus	0.5349	0.8667
Buddha	0.4961	0.892
Camera	0.583	0.9383
Cannon	0.5047	0.8233
Cellphone	0.6497	0.9493
Chair	0.5132	0.856
Chandelier	0.5906	0.8957
Cougar_Body	0.5998	0.8135
Cougar_Face	0.48	0.922
Crab	0.5326	0.921
Crayfish	0.5961	0.761
Crocodile	0.4868	0.8113
Crocodile_Head	0.5722	0.7943
Cup	0.5751	0.9577
Dalmatian	0.4681	0.8877
Dollar_Bill	0.5458	0.9057
Dolphin	0.531	0.8587
Average	0.5494	0.8882

As anticipated, Caltech 101 frequently beat Wang V1.0 and Wang V2.0, as seen in Tables 8, where precision was around 0.89 and accuracy was practically flawless. Accordion and Airplanes were awarded perfect scores. The Crayfish has the lowest accuracy and precision of all the models, both of which are consistent with the earlier results. The findings shown that the precision gained by Fine-matching is 33,00 percent greater than

that of coarse matching, with an average of around 0.88, while achieving almost flawless accuracy. However, the fine-matching has effectively addressed the shortcomings of the coarse-phase. It has been noticed that the decrease of FP has not only influenced precision favorably, but also accuracy positively. It is mostly attributable to the Slantlet transform displaying minute texture details.

Concerning the preceding study, numerous writers developed a system based on special domain feature extraction and combined it with frequency domain approaches to improve image retrieval and achieve a high precision average. However, the precision of our suggested approach, which uses Slantlet transfer to extract features from colour channels, was the highest. Table 9 clearly shows that the proposed technique outperformed the others..

The drawbacks and limitations of the previous methods are mostly resolved. These achievements were mainly attributed to the following contribution; (1) It incorporated two types of texture features, spatial- and SLT-based. The former was extracted from HSV and YCbCr colour spaces, while the latter was drawn from the SLT-decomposed image. Here, two original ideas were brought forth. The way spatial-based texture features are drawn via HSV and YCbCr. Moreover, another idea is the way the SLT-based texture feature is computed via RGB instead of greyscale. Consequently, only the first two rows of the SLT matrix are used for each channel to extract texture features. (2) A novel Coarse-to-Fine approach is introduced instead of the threshold-based approach, which was previously adopted by the majority. Like a coarse-grained sugar filter, which separates granulated sugar from caster sugar. In this study, the query image is first coarsely matched against the entire dataset to form a shortlist. Afterwards, the fine matching is applied using the shortlist as the input, from which the most likely candidate that resembles the query image is chosen.

7. CONCLUSIONS AND FUTURE WORK

Despite encouraging milestones, many challenging issues remain unresolved, such as FP and low precision. Hence, against the backdrop, this study has proposed a two-tier approach that worked for hand in hand in achieving both high precision and high accuracy, in which the coarse matching became the feeder to the fine matching. Texture characteristics, according to the Caltech 101 dataset, are reliable and capable of producing excellent results while also being resistant to low resolution. It also demonstrated that spatial-based texturing alone cannot achieve the desired precision and that the SLT-based texture feature is the ideal partner. Despite the promising results of the proposed strategy, there are still many areas for development.

For the issue of an FP, the best precision reached to date is [0.885 – 0.888], indicating that the problem is still there. Furthermore, it is not easy to achieve complete precision due to inadequacies in the ground truth. Some photographs were incorrectly sorted manually, resulting in different images being placed into the same class while nearly identical images were grouped into multiple classes. Interclass similarity and intra-class dissimilarity are terms used to describe this situation. As a result, a more accurate ground truth has been long required.

Table. 9. Precision comparsion between the proposed method and state-of-the-art

Dataset	Proposed CBIR	S. Somnugong et al. 2016 [12]	Nazir et al. 2018 [13]	S. P. Rana et al. 2019 [14]	Q. Zheng et al. 2019 [15]
WANG V1.0	0.788	0.723	0.693	0.753	0.735
WANG V2.0	0.787	0.756	0.703	0.642	0.761
Caltech 101	0.888	0.725	0.735	0.645	0.885

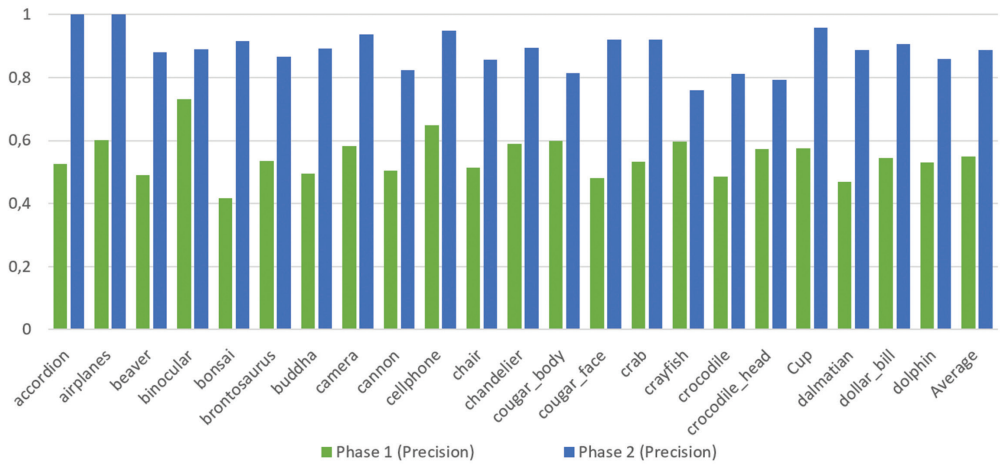

Fig. 5. Comparsion of performance evaluation using proposed two phases in terms of Precision.

8. REFERENCES:

[1] I. M. Hameed, S. H. Abdulhussain, B. M. Mahmmod, "Content-Based Image Retrieval: A Review of Recent Trends", Cogent Engineering, Vol. 8, No. 1, 2021, pp. 1-37.

[2] M. K. Alsmadi, "Content-Based Image Retrieval Using Color, Shape and Texture Descriptors and Features", Arabian Journal for Science and Engineering, Vol. 45, No. 4, 2020, pp. 3317-3330.

[3] S. Hossain, R. Islam, "A New Approach of Content Based Image Retrieval Using Color and Texture Features", Current Journal of Applied Science and Technology, Vol. 51, No. 3, 2017, pp. 1-16.

[4] M. Danish, R. Rawat, R. Sharma, "A Survey: Content Based Image Retrieval Based on Color, Texture, Shape & Neuro Fuzzy", Image, Vol. 3, No. 5, 2013, pp. 839-844.

[5] M. Sharma, A. Batra, "Analysis of Distance Measures in Content Based Image Retrieval", Global Journal of Computer Science and Technology, Vol.14, No. 2, 2014, pp. 11-16.

[6] M. F. Sadique, B. K. Biswas, S. R. Haque. "Unsupervised Content-Based Image Retrieval Technique Using Global and Local Features". Proceedings of the 1st International Conference on Advances in Science, Engineering and Robotics Technology, Dhaka, Bangladesh, 3-5 May 2019, pp. 1-6.

[7] O. Beijbom et al. "Towards Automated Annotation of Benthic Survey Images: Variability of Human Experts and Operational Modes of Automation", PloS one, Vol. 10, No. 7, 2015, pp. e0130312.

[8] Z. Hu et al. "A Novel Method Based on a Mask R-Cnn Model for Processing Dpcr Images", Analytical Methods, Vol. 11, No. 27, 2019, pp. 3410-3418.

[9] X. Li, J. Yang, J. Ma, "Recent Developments of Content-Based Image Retrieval (CBIR)", Neurocomputing, Vol. 452, 2021, pp. 675-689.

[10] A. Latif et al. "Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review", Mathematical Problems in Engineering, Vol. 2019, 2019, pp. 1-21.

[11] X. Jiang, J. Jiang, A. Fan, Z. Wang, J. Ma, "Multiscale Locality and Rank Preservation for Robust Feature Matching of Remote Sensing Images", IEEE Transactions on Geoscience and Remote Sensing, Vol. 57, No. 9, 2019, pp. 6462-6472.

[12] S. Somnugpong, K. Khiewwan, "Content-Based Image Retrieval Using a Combination of Color Correlograms and Edge Direction Histogram", Proceedings of the 13th International Joint Conference on Computer Science and Software Engineering, Khon Kaen, Thailand, 13-15 July 2016, pp.1-5.

[13] A. Nazir, R. Ashraf, T. Hamdani, N. Ali, "Content Based Image Retrieval System by Using Hsv Color Histogram, Discrete Wavelet Transform and Edge Histogram Descriptor", Proceedings of the International Conference on Computing, Mathematics and Engineering Technologies, Sukkur, Pakistan, 3-4 March 2018, pp. 1-6

[14] S. P. Rana, M. Dey, P. Siarry, "Boosting Content Based Image Retrieval Performance through Integration of Parametric & Nonparametric Approaches", Journal of Visual Communication and Image Representation, Vol. 58, 2019, pp. 205-219.

[15] Q. Zheng, X. Tian, M. Yang, H. Wang, "Differential Learning: A Powerful Tool for Interactive Content-Based Image Retrieval", Engineering Letters, Vol. 27, No. 1, 2019, pp. 202-215.

[16] M. Q. Shatnawi, M. Alrousan, S. Amareen, "A new approach for content-based image retrieval for medical applications using low-level image descriptors", International Journal of Electrical and Computer Engineering, Vol. 10, No. 4, 2020, pp. 4363-4371.

[17] A. P. Byju, B. Demir, L. Bruzzone, "A Progressive Content Based Image Retrieval in JPEG 2000 Compressed Remote Sensing Archives", IEEE Transactions on Geoscience and Remote Sensing, Vol. 8, No. 85, 2020, pp. 5739-5751.

[18] M. Jammula, "Content Based Image Retrieval System Using Integrated ML and DL-CNN", Annals of R.S.C.B, Vol. 25, No. 4, 2021, pp. 9656-9666.

[19] A. Ahmed, S. Mohamed, "Implementation of early and late fusion methods for content-based image retrieval", International Journal of Advanced and Applied Sciences, Vol. 8, No. 7, 2021, pp. 97-105.

[20] A. Alzu'bi, A. Amira, N. Ramzan, "Semantic Content-Based Image Retrieval: A Comprehensive Study", Journal of Visual Communication and Image Representation, Vol. 32, 2015, pp. 20-54.

[21] S. Zakariya, I. A. Khan, "Analysis of Combined Approaches of Cbir Systems by Clustering at Varying Precision Levels", International Journal of Electrical & Computer Engineering, Vol. 11, No. 6, 2021, pp. 5009-5015.

[22] J. Chaki, N. Dey, "Image Color Feature Extraction Techniques: Fundamentals and Applications", 1st Edition, Springer Nature, 2021.

[23] C. Singh, K. P. Kaur, "A Fast and Efficient Image Retrieval System Based on Color and Texture Features", Journal of Visual Communication and Image Representation, Vol. 41, 2016, pp. 225-238.

[24] I. W. Selesnick, "The Slantlet Transform", IEEE Transactions on Signal Processing, Vol. 47, No. 5, 1999, pp. 1304-1313.

[25] V. SS, B. Vasista, V. N Sansthanik, "Slantlet Transform: An Efficient Approach for Compression and De-Noising of Power Quality Events", International Journal of Computer Applications, Vol. 91, No. 4, 2014, pp. 27-31.

[26] M. S. H. Al-Tamimi, "Combining Convolutional Neural Networks and Slantlet Transform for an Effective Image Retrieval Scheme", International Journal of Electrical & Computer Engineering, Vol. 9, No. 5, 2019, pp. 4382-4395.

[27] S. F. Da Silva, M. X. Ribeiro, J. d. E. B. Neto, C. Traina-Jr, A. J. Traina, "Improving the Ranking Quality of Medical Image Retrieval Using a Genetic Feature Selection Method", Decision Support Systems, Vol. 51, No. 4, 2011, pp. 810-820.

[28] H. Wu, Y. Cao, H. Wei, Z. Tian, "Face Recognition Based on Haar Like and Euclidean Distance", Journal of Physics: Conference Series, 2021, pp. 1-6.