Image Features Evaluation Using New Algorithm Proposed For Reducing Image Feature Number & Size Stored In Database

Shahlaa T. Abdulwahab*

Received on: 10/1/2011
Accepted on: 7/4/2011

Abstract

This study proposes technique that capable of reducing image features size and number stored in the database. The proposed technique depends on the image content of numerical values for the three basic colors (red, green and blue) and then stores it in the database and to be used for image retrieval. This technique has been developed based on recent image retrieval procedures that include Color Descriptor Matrix, YCbCr Color Space and Discrete Cosine Transform. Those procedures have been applied sequentially on the image and finally Kekre's Transform has been applied in the last stage of this technique to evaluate image features and reduce its stored size in the database. The validity and accuracy of the proposed technique have been evaluated through experiments by applying Kekre's Transform on Color Descriptor Matrix instead of using Kekre's Transform directly on the image in order to reduce its feature stored size. Another experiments have been tested and evaluated that include the application of YCbCr Color Space on the Color Descriptor Matrix and finally Kekre's Transform to be executed and explore the image features size and compare it with the previous stage. The effect of applying the Discrete Cosine Transform on the YCbCr Color Space and finally the Kekre's Transform on the image features size has been studied and compared with the previous step. It is concluded that the best reduction in image features size stored in the database can be obtained only when Kekre's Transform applied in the last step of the proposed technique with unchanged threshold based image retrieval ratios. Parametric study has been conducted to investigate the effect of applying the new algorithm on both isolated and mixed image groups. Good precision ratios of 82% and 65% have been obtained for the isolated and mixed image groups respectively.

Keywords: Image Retrieval; Kekre's Transform; YCbCr Color Space; Discrete Cosine Transform.

***Education College of Ibn Al-Haitham , University of Baghdad / Baghdad**

https://doi.org/10.30684/ etj.29.6.12
University of Technology-Iraq, Baghdad, Iraq/2412-0758

إستخراج خواص الصورة بإستخدام خوارزمية جديدة لتقليل عدد وحجم محتوياتها المخزونة في قاعدة البيانات

الخلاصة

يتناول هذا البحث طريقة مقترحة لتقليل حجم وعدد خواص الصورة المخزونة في قاعدة البيانات حيث تعتمد هذه الطريقة على محتويات الصورة من قيم رقمية للالوان الاساسية (الاحمر والاخضر والازرق) وخزنها في قاعدة البيانات ومن ثم استخدام هذه الخواص في استرجاع الصور. إن الطريقة المقترحة في هذا البحث تم تطويرها بناءاً على طرق سابقة تُستخدم في استرجاع الصورمن قاعدة البيانات مثل (Color Descriptor Matrix) و (YCbCr Color Space) و (Discrete Cosine Transform). تتضمن الطريقة المقترحة تنفيذ طرق الاسترجاع السابقة بشكل متتالي على الصورة ومن ثم تطبيق Kekre's Transform في اخر مرحلة واستخراج خواص الصورة بهدف تقليل عدد و حجم الخواص التي يتم خزنها في قاعده البيانات. لقد تم التاكد من دقة وصلاحية الطريقة المقترحة في هذا البحث من خلال إجراء عدد من التجارب والدراسات في خزن واسترجاع خواص الصور في قاعدة البيانات. تضمنت بعض التجارب تطبيق ال Kekre's Transform على Color Descriptor Matrix واستخراج عدد و حجم الخواص ومن ثم مقارنتها مع تطبيق Kekre's Transform مباشرة على الصورة. وتضمنت الدراسات ايضا تطبيق YCbCr Color Space على Color Descriptor Matrix ثم تنفيذ Kekre’s Transform عليه واستخراج خواص الصورة ومقارنة حجم الخواص مع المرحلة التي قبلها.تمت ايضا في هذه الدراسة تطبيق Discrete Cosine Transform على YCbCr Color Space ومن ثم تطبيق Kekre's Transform عليها وإستخراج خواصها ومقارنتها بالمراحل التي قبلها. لقد تمت مقارنة حجم وعدد خواص الصورة المخزونة في قاعدة البابنات في كل مرحلة عن التي قبلها. أظهرت نتائج الفحوص ان تطبيق ال Kekre's Transform في اخر مرحلة من الخوازمية المقترحة هي الافضل من ناحية تقليل حجم وعدد الخواص التي ستخزن في قاعدة البيانات مع مراعاة عدم تغير نسبة الاسترجاع وذلك لانها تعتمد على Threshold في الاسترجاع. كما تم إجراء دراسة خاصة تتعلق بتطبيق الخوارزمية المقترحة على مجموعات الصور المنفصلة والمختلطة في قاعدة البيانات. أظهرت نتائج الدراسة معدل نسب إسترجاع جيدة تبلغ 82% و 65% في حالة الصور المنفصلة والمختلطة على التوالي.

Introduction

In recent years, the rapid increase in the number of images in various databases has led to a parallel increase in the public demand of image retrieval systems having preferably content based search capabilities. These phenomena led to the implementation of many content based image retrieval systems [1], [2]. However, without effective image retrieval it is impossible to make use of the huge image databases. There are basically three ways of retrieving previously stored multimedia data, i.e., free browsing, text-based retrieval and content-based retrieval [3].

Content-based image retrieval (CBIR) seeks techniques to re-encode contents of images and techniques to measure the similarity of two images based on the coding, in order to index and retrieve images which are semantically related to a visual information request without fully understanding the semantics carried by these images [4]. CBIR is considered as the process of retrieving desired images from huge databases based on extracted features from the image themselves (without resorting to a key word). Features are derived directly from the images and they are extracted and analyzed by means of computer processing [5].

Typical CBIR system performs two major tasks. The first one is feature extraction (FE), where a set of features, called image signature or feature vector, is generated to accurately represent the content of each image in the database. A feature vector is much smaller in size than the original image, typically of the order of hundreds of elements (rather than millions). The second task is similarity measurement (SM), where a distance between the query image and each image in the database using their signatures is computed so that the top “closest” images can be retrieved [6].

In this paper, an image retrieval algorithm has been proposed and used to reduce extensively image features numbers and size stored in the database.

The current paper is organized as follows: In Section 2, image description and approximations has been presented. Color space conversion application has been presented in Sections 3. Image retrieval algorithms have been presented in Sections 4 and 5. The proposed algorithm for Image organization and retrieval by color features is presented as in section 6. Some experiments with the prototype system are shown in section 7 and conclusions and future work aims are planned in section 8.

Image Description And Approximations

In our system we use a color code book, created on the basis of vector quantization in RGB color space with 3 basic colors (red, green and blue). The goal of color feature extraction is to obtain compact, perceptually relevant representation of the color content of image retrieval.

Color Descriptor Matrix

In order to define 2-D vector Color Descriptor Matrix and represent the spatial color information in images, a combination of image quantizationand the perceptual color model is used [4]. In order to create this structure, the whole image is divided into NxN (where N=4, 8, 16, 32) equal parts. This matrix stores the dominant colors for the image blocks. The original images were NxN quantized and were represented as NxN blocks (or sub images). The creation and retrieval time and database size are examined depending on the size of N in [7] and made the conclusion that N=16 is the most adequate for this purposes.

In order to create this index structure the whole image is divided into 256 equal parts. This matrix stores the coefficient of the dominant color from the selected color code book in the corresponding part of the image. The original images are 16X16 quantized and represented as 16X16 color blocks as shown below [7];

$$CDM = \begin{bmatrix} C_{0,0} & C_{0,1} & \cdots & \cdots & C_{0,15} \\ C_{1,0} & C_{1,1} & & & \cdots \\ \cdots & & & & \cdots \\ \cdots & & & & \cdots \\ C_{15,0} & \cdots & \cdots & \cdots & C_{15,15} \end{bmatrix} \quad \ldots\ldots(1)$$

Where CDM is the Color Descriptor Matrix and $C_{i,j}$ are the color descriptors for the dominant color in the i, j element of the image.

Color Space Conversion Application

A color space is a mathematical representation of a set of colors. The three most popular color models are [8]:

1. RGB (used in computer graphics), R'G'B' (gamma corrected RGB)
2. YIQ, YUV, YCrCb used in video systems
3. CMYK (used in color printing).

The color space conversion algorithm is used for converting image data from one color space (RGB) to another color space (YCbCr).

YCbCr Color Space

YCbCr, is a family of color spaces used as a part of the color image pipeline in video and digital photography. YCbCr is an opponent-color system. The first channel, Y, is the luma component of the image. Luma is the brightness of the image. The chrominance (chroma) of the image is held on the other two components. The blue chroma component is labeled Cb, and Cr is the red chroma component. YCbCr is not an absolute color space, but instead a way of encoding the RGB color model. Therefore, the actual color representation is dependent on the RGB color space in which it was derived from [9].

YCbCr is defined by a linear transformation of RGB color space [10]:

Luminance

$$Y = 0.299*R + 0.587*G + 0.114*B - 128 \quad \ldots.(2)$$

Blue chrominance

$$Cb = 0.169*R - 0.331*G + 0.500*B \quad \ldots.(3)$$

Red chrominance

Cr = 0.500*R – 0.419*G – 0.081*B ….(4)

The difference between YCbCr and RGB is that YCbCr represents color as brightness and two color difference signals, while RGB represents co as red, green and blue.

The Discrete Cosine Transform

Like other transforms, the Discrete Cosine Transform (DCT) attempts to decorrelate the image data. After decorrelation each transform coefficient can be encoded independently without losing compression efficiency [11].

The discrete cosine transform (DCT) is closely related to the discrete Fourier transform. It is a separable linear transformation; that is, the two-dimensional transform is equivalent to a one-dimensional DCT performed along a single dimension followed by a one-dimensional DCT in the other dimension. The definition of the two-dimensional DCT for an input image A and output image B is [12];

$$B_{pq} = \alpha_p \alpha_q \sum \sum A_{m,n} \cos \frac{\pi(2m+1)p}{2M} * \cos \frac{\pi(2n+1)q}{2N},$$

$0 \le p \le M-1,\ 0 \le q \le N-1$ ….(5)

$$\alpha_p = \begin{cases} 1/\sqrt{M} & , p = 0 \\ \sqrt{2/M} & , 1 \le p \le M-1 \end{cases},$$

$$\alpha_p = \begin{cases} 1/\sqrt{N} & , q = 0 \\ \sqrt{2/N} & , 1 \le q \le N-1 \end{cases}$$ ….(6)

Where;

p and q are variables ranging as shown in the above equations (Equ. No. 5 and 6).

M and N are the row and column size of an input image A, respectively.

The DCT tends to concentrate information, making it useful for image compression applications and also helping in minimizing feature vector size in CBIR .

Kekre's Transform

Kekre's transform matrix can be of any size NxN, which need not to be an integer power of 2. All upper diagonal and diagonal elements of Kekre's transform matrix are 1, while the lower diagonal part except the elements just below diagonal is zero. Generalized NxN Kekre's transform matrix can be given as [13];

$$K_{NxN} = \begin{bmatrix} 1 & 1 & 1 & \dots & . & . & 1 & 1 \\ -N+1 & 1 & 1 & \dots & . & . & 1 & 1 \\ 0 & -N+2 & 1 & \dots & . & . & 1 & 1 \\ . & . & . & \dots & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 0 & . & . & \dots & . & . & 1 & 1 \\ 0 & 0 & 0 & \dots & . & . & -N+(N-1) & 1 \end{bmatrix}$$ ….(7)

The formula for generating the element K_{xy} of Kekre's transform matrix is,

$$K_{xy} = \begin{cases} 1 & ; x \le y \\ -N+(x-1) & ; x = y+1 \\ 0 & ; x > y+1 \end{cases}$$.(8)

Matching Process & Threshold Determination

The matching process helps to evaluate if two elements are equal comparing both elements and calculating the distance between them. In the case of color descriptors, the matching process helps to evaluate if two images are similar [10]. Many Current Retrieval systems take a simple approach by using typically norm-based distances (e.g., Euclidian distance) on the extracted feature set as a similarity function [6]. The query image has been processed same as before to calculate image features and to be stored in the database. The calculated features will be compared with other image features already stored in database by evaluating the Euclidian distance. The direct Euclidian distance between the two descriptors of image stored in the database (P) and query image (Q) can be calculated as below;

$$ED = \sqrt{\sum_{i=1}^{n}\left(Fp_i - Fq_i\right)^2} \qquad \text{….(9)}$$

Where Fp_i and Fq_i are the feature vectors of the image stored in the database P and query image Q respectively with size n.

If the calculated distance is zero, the two images match, while if the result is very close to zero, the two images are similar to each other. These distances are calculated between the query image and database images which is sorted in ascending order. By taking maximum distance into consideration, only 25% of the maximum Euclidian distance to be used throughout this study as standard threshold.

The Proposed Algorithm

In the proposed algorithm, Color Descriptor Matrix, YCbCr color space ,Discrete Cosine Transform and Kekre Transform have been used as tools for image retrieval depends on its features through image processing until find a numerical values to be the unique image features depending on the three parts (red, green and blue color array). Then, this image features have to be stored with the document image in the database in order to use it later on to discover about if the image in database and which images like it.

The idea of the proposed algorithm depends on using of Color Descriptor Matrix, YCbCr color space, Discrete Cosine Transform and Kekre's Transform in sequential way to find an algorithm with new features aid for generation image feature. The algorithm proposal index key generation can be shown in Fig. 1.

In the current study, the Euclidian distance can be expressed as follows;

$$D = \sqrt{\sum_{i=1to16}(RRKQ-RRKI)^2} + \sqrt{\sum_{i=1to16}(RGKQ-RGKI)^2} + \sqrt{\sum_{i=1to16}(RBKQ-RBKI)^2} + \sqrt{\sum_{i=1to16}(CRKQ-CRKI)^2} + \sqrt{\sum_{i=1to16}(CGKQ-CGKI)^2} + \sqrt{\sum_{i=1to16}(CBKQ-CBKI)^2} \qquad \text{….(10)}$$

Where:

RR = Row red, RG= Row green, RB = Row blue, K = Kekre, I = Image, Q = Query image, C = Column.

Proposed Algorithm Implementation

Implement of the proposed algorithm is performed on Pentium IV PC with Visual Basic Compiler (V.6) for image processing and Microsoft Access Software for database storage in ten stages as shown in Fig. 2 and stated below.

Stag1: input image and scanning (split image in R, G and B planes).

Stage 2: Split the image into 16 X 16 blocks (size of image =128*128 pixels).

Stage 3: Apply color descriptor matrix to calculate the dominant color for each block and then, redraw the image with that color.

Stage 4: Apply YCbCr Color space on each block and redraw image.

Stage 5: Apply Discrete Cosine Transform on each block and redraw image.

Stage 6: Evaluate row and column mean vectors for image in previous stage.

Stage 7: Applying Kekre's Transform on row and column mean vectors.

Stage 8: Applying Similarity Measure; Euclidian distance.

Stage 9: Selection of images less that predefined threshold.

Stage 10: Retrieval images.

Experiments

For the purpose of the experiments, test image collection consists of 75 original images of 24 bits true color has been selected. The test images were divided into 5 groups (sunset, snow trees, horses, sharks and palm fronds), each containing 15 images as shown in Fig. 3. For simplicity, all images are pre-processed to be 128x128 sizes before decomposition.

To investigate the proposed algorithm, the following experiments have been made.

Image features vector elements are calculated using Kekre's transform based on the image size. In case of 128x128 pixel image and for gray color space, both row mean and column mean vectors will consists of 128 elements. Kekre's transform will be applied on the row and column mean vectors to evaluate the image feature vectors to be stored later on in the database. The number of image features stored in the database will be 256 elements (128+128). The using of Kekre's transform only will results in large number of image features that require large memory storage in database and time of processing during matching. In case of RGB color space, the number of the row mean vector will be multiplied by 3, i.e., will be 384 elements and the same for the column mean vector. Consequently, the size of the image features stored in the database will be 768 (384 for each row and column mean vectors). Furthermore, as the size of the image became big, the image features numbers will be large.

In the current study, Color Descriptor Matrix procedure has been used to reduce the image features numbers. This has been done through dividing the image of 128x128 pixels into 16x16 blocks. Each block will be analyzed to find out the dominant color and re-draw each block with the dominant color only (i.e., the block size will be 8x8 pixels only) as illustrated in Fig. 4. This will results in appropriate reduction in the imagefeatures numbers obtained through row and column mean vectors calculations compared to Kekre's transform application only. In gray color space, the size of each row and column mean vectors will be 16 elements only, i.e, the total number of image features stored in the database will be 32 elements (16+16) compared to 256 in case of Kekre's transform. While, in case of RGB color space, the size of the row mean vector will be multiplied by 3, i.e., will be 48 elements and the same for the column mean vector. Consequently, total number of image features stored in the database will be 96 (48 for each row and column mean vectors) rather than 768 for Kekre's transform application.

Since the proposed algorithm improves image features calculations by reducing its numbers, the time of matching between the query image and the other stored images in database will be reduced and consequently time reduction in image retrieval process could be achieved especially for large database.

In the proposed algorithm, the division process of the image will not affect the original image properties and characteristics or the algorithm accuracy since it will be used only to evaluate image features which will be used later on in the last stages of the algorithm implementation. Therefore, maintaining of the original image properties stored in the database could be achieved throughout the current proposal.

YCbCr Color Space is implemented after Color descriptor Matrix application to reduce the image features vector sizes, while implementation of Discrete Cosine Transform after applying YCbCr Color Space will results in more reduction in feature vector sizes of the image. Finally, Kekre's Transform will be applied to both row and column means vectors to evaluate the image features which will be stored in the database.

It is found that the application of Color Descriptor Matrix on an image that converted from RGB color space to YCbCr color space will results in good reduction in image features sizes as shown in Fig. 5a. While evaluating the row and column means vectors for the same and employing Kekre's transform, will leads to more reduction of image features sizes. Also, it is found that the applications of Discrete Cosine Transform on an image in YCbCr Color Space will reduce image features sizes and consequently reduces the memory required to store the image features in database as illustrated in Fig. 5b. In spite of the processed image (Fig. 5b) is not clear, accurate and clear image features can be evaluated precisely from it and will be used to find the row and column mean vectors.

Row and column mean vectors for each color array (Red, Green and Blue) have been calculated through the last stage of the proposed technique (Discrete Cosine Transform) and listed in Table 1. Kekre's Transform has been used in the last stage to evaluate the image feature vectors which becomes less in numbers and size required to store in database as listed in Table 2. In all studies conducted in this paper, retrieval accuracy has been maintained without any reduction.

Comparison between the employed procedures based on image features size is illustrated in Table 3 for five different groups of images (snow, palm fronds, horse, shark and sunset). In each stage, the row and column mean vectors have been evaluated and Kekre's Transform has been applied to find out the reduced image features sizes. It is clear from the table that applying Kekre's Transform at the last stage of each process results in more and more reduction in image feature sizes. While applying Kekre's Transform directly on an image will increase the image features sizes and hence increasing memory required and time of image retrieval.

15 images of sunset of 128x128 pixels have been selected to evaluate features vector for each image and stored in the database. The accuracy of the proposed algorithm can be illustrated in Fig. 6.

To evaluate the performance of the proposed image retrieval algorithm, two parameters Precision and Recall have been selected and used that derived from all CBIR experiments. The Precision is the fraction of the relevant images which has been retrieved (from all retrieved) and can be calculated as follows:

$$\text{Precision} = A / (A+B) \qquad \text{….(11)}$$

While, Recall is the fraction of the relevant images which has been retrieved (from all relevant):

$$\text{Recall} = A / (A+D) \qquad \text{….(12)}$$

Where; A is the relevant retrieved images, *B* is the non-relevant retrieved images and *D* is the relevant not retrieved images.

Experiments are executed with 15 images from each of the five categories and calculated the average precision and average recall parameters for all of them. The results obtained are given in Tables 4 and 5 and shown in Figs. 7 and 8. Table 4 lists the results in case of isolated image categories stored in the database, while the results for the mixed categories are listed in Table5. Last row in the table is indicating the average precision and recall for all 75 queries from the five categories of images in the database. It is noted that very good retrieval average precisions of 82% and 65% have been obtained using the proposed algorithm for both the isolated and mixed image groups respectively.

Conclusions

Based on the above experiment results, the following conclusion can be drawn:

1. The proposed algorithm has best performance in terms of retrieval accuracy and speed because it results in great reduction of image features numbers and sizes compared to other techniques.
2. In terms of the memory space required for storing image features in database, the proposed technique presents less memory space.
3. Application of Kekre's Transform on Color Descriptor Matrix instead of applying Kekre's Transform directly on the image results in good reduction of image features sizes.
4. Adopting of YCbCr Color Space after Color Descriptor Matrix and then applying Kekre's Transform results in more reduction in image features sizes. Better results can be achieved thought applying Discrete Cosine Transform after YCbCr Color Space and adopting Kekre's Transform in the last stage.
5. The division process implemented in the current technique has no effect on image properties and characteristics or on retrieval accuracy.
6. Good retrieval precision ratios of 82% and 65% have been obtained using the proposed technique for both isolated and mixed image groups in database.

References

[1]-Goodrum, A., "Image Information Retrieval: An Overview of Current Research", Special Issue on Information Science Research, Vol. 3, No. 2, 5 pp., 2000.

[2]-Rui, Y. Huang, T.S., "Image Retrieval: Current Techniques, Promising Directions and Open Issues", Journal of Visual Communication and Image Representation, Vol. 10, pp. 1-23, 1999.

[3]-Zheng, W.M. Lu, Z.M. Burkhardt, H.," Color Image Retrieval Schemes Using Index Histograms Based on Various Spatial-Domain Vector Quantizes", International Journal of Innovative Computing, Information and Control, Vol. 2, No. 6, December, pp. 1317-1326, 2006.

[4]-Valova, I. Rachev, B.," Retrieval by Color Features in Image Databases", University of Rousse, Department of Computer Systems, and Technologies, 7017 Rousse, Bulgaria, 16 pp., 2007

[5]-Latha, Y. M. Jinaga , B.C. Reddy, V.S.K., " A Precise Content-based Color Image Retrieval: Lifting Scheme", ICGST-GVIP Journal, ISSN: 1687-398X , Vol. 8, Issue 1, pp. 25-32, 2008.

[6]-Kekre, H.B. Thepade, S.D.," Boosting Block Truncation Coding with Kekre's LUV Color Space for Image Retrieval", International Journal of Electrical and Electronics Engineering, Vol. 2, No. 3, pp. 172-180, 2008.

[7]-Valova, I. Rachev, B., "Image Databases–An Approach for Image segmentation & Color reduction Analysis & Synthesis", International Conference CompSysTech'2003, Sofia, Bulgaria, 19-20 June, 9 pp, 2003.

[8]-ITU-R, "Parameter Values for the HDTV Standards for Production and international program Exchange", BT.709-4, pp. 1-20, 2000. [9]-Monaco, M.K., "Color Space Analysis for Iris Recognition", M.Sc. Thesis in Electrical engineering, College of Engineering and Mineral resources, West Virginia University, 103 pp, 2007.

[10]-Verdaguer, S.L., "Color Based Image Classification and Description", M.Sc. Thesis, Telecommunication Engineering & Management, Universitat Politecnica De Catalunya, October , 92 p.p, 2009.

[11]-Khayam, S.A.," The Discrete Cosine Transform (DCT): Theory and Application", Department of Electrical & Computer Engineering, Michigan State University, 32 pp, 2003.

[12]-Kekre, H.B. Thepade, S.D.Maloo, A." Image Retrieval using Fractional Coefficients of Transformed Image using DCT and Walsh Transform", International Journal of Engineering Science and Technology, Vol. 2(4), pp. 362-371, 2010.

[13]- Kekre, H.B. Athawale, A. Sadavarti, D.," Algorithm to Generate Kekre's Wavelet Transform from Kekre's Transform", International Journal of Engineering Science and Technology, Vol. 2, No.5, pp. 756-767, 2010.

Table 1: Row and column mean vectors values for an image.

Row Mean Vectors			*Column Mean Vector*		
DCTR	**DCTG**	**DCTB**	**DCTR**	**DCTG**	**DCTB**
20.560	19.937	20.187	20.687	21.000	28.937
14.625	7.9375	19.125	16.437	10.062	16.500
0.0625	0.1250	0.1875	0.0625	0.0625	0.1875
4.8125	2.7500	6.6875	5.4375	3.5000	5.8125
0.0625	0.1250	0.1875	0.0625	0.1250	0.1875
3.0625	1.5625	3.7500	3.4375	1.9375	3.1875
0.0625	0.1250	0.1875	0.0625	0.1250	0.2500
2.2500	1.3125	3.1875	2.4375	1.6875	2.8125
0.0625	0.1250	0.2500	0.0625	0.1250	0.1875
1.1250	1.3125	1.3750	0.0000	0.1250	0.3125
0.0625	0.2500	0.5000	0.3125	0.0000	0.4375
0.9375	1.3750	1.7500	0.0625	0.1250	0.5625
0.6250	0.1250	0.2500	0.0625	0.1250	0.1875
0.8750	0.8750	1.1250	0.0000	0.1875	0.3125
0.1250	0.2500	0.5625	0.3125	0.0625	0.5000
0.7500	1.2500	1.6250	0.0625	0.1875	0.5625

Table 2: Kekre Row and column mean vectors values for an image.

Kekre Row Mean Vectors			*Kekre Column Mean Vector*		
KDCTR	**KDCTG**	**KDCTB**	**KDCTR**	**KDCTG**	**KDCTB**
3.9	2.46	3.81	3.09	2.46	3.81
-18.75	-18.72	-17.64	-18.89	-19.85	-26.94
-12.82	-6.72	-16.58	-14.64	-8.91	-14.5
0.84	0.61	1.18	0.71	0.46	0.79
-3.32	-1.69	-4.51	-3.99	-2.54	-4.13
0.54	0.44	0.77	0.38	0.2	0.44
-1.71	0.64	-1.9	-2.15	-1.16	-1.81
0.35	0.35	0.55	0.17	0.09	0.21
-1.02	-0.39	-1.33	-1.32	-0.89	-1.39
0.21	0.28	0.32	0.02	-0.01	0.09
-0.32	-0.32	-0.24	0.05	-0.01	0.02
0.15	0.15	0.14	-0.09	0.04	-0.03
-0.18	-0.27	-0.32	0.01	0	-0.08
0.09	0.12	0.14	0.01	0	0.04
-0.11	-0.07	-0.07	0.02	-0.02	0.01
0.03	0.05	0.03	-0.04	0	-0.03

Table 3: Image features size evaluated for 5 groups of images using different techneques after applying Kekre's transform at the end of each technique.

Image/Group	***Kekre's Transform only (bytes)***	***Color Descriptor Matrix→Kekre's Transfom (bytes)***	***Color Descriptor Matrix→YCbCr Color Space→Kekre's Transfom (bytes)***	***Color Descriptor Matrix→YCbCr Color Space→Discrete Cosine→Transform Kekre's Transfom (bytes)***
Snow Trees	6247	819	818	753
Palm fronds	6371	800	800	767
Horses	6391	840	802	762
Sharks	6297	831	792	770
Sunset	6339	816	803	758

Table 4: Retrieval results for the proposed algorithm over row and column mean vectors at 25% of maximum Euclidian distance (Isolated image groups).

Group	Total retrieved images	Relevant retrieved images	Precision	Recall
Sharks	40	24	0.69	0.11
Horses	17	15	0.93	0.07
Snow trees	21	16	0.83	0.07
Sunset	27	24	0.92	0.11
Palm Fronds	32	19	0.74	0.08
Average Precision			0.824	0.087

Table 5: Retrieval results for the proposed algorithm over row and column mean vectors at 25% of maximum Euclidian distance (Mixed image groups).

Group	Total retrieved images	Relevant retrieved images	Precision	Recall
Sharks	109	59	0.59	0.26
Horses	17	15	0.93	0.07
Snow trees	67	25	0.64	0.11
Sunset	153	54	0.50	0.24
Palm Fronds	142	49	0.58	0.22
Average Precision			0.65	0.180

```
Algorithm:
Inputs:  Database document images.
Output: image features.
Begin
Image scanning; split image in R, G and B planes
{Color descriptor matrix}
    Split (image values)      {divide the image 16*16 Blocks}
For j1 = 0 To n Step 8  {where n = image width and image height}
         For i1 = 0 To n Step 8
             For i = i1 To i1 + 7
                For j = j1 To j1 + 7
Compute Histogram for each block (red, green and blue color array) and chose the color
have largest value of each color array.
                Next j
             Next i
         Next i1
Next j1
Re-draw the image again with dominant colors
Apply;
{YCbCr Color Space},
{Discrete cosine Transform}
Calculate row and column mean vectors for each color array (red, green and blue).

Calculate the Kekre's Matrix

Apply;
Kekre Transform on row and column means vectors to evaluate features vectors.

{Apply similarity measure Euclidian distance for each red, green, blue row mean and
column mean}

{selection of image less than predefine threshold}

end
```

Figure. 1: The proposed algorithm.

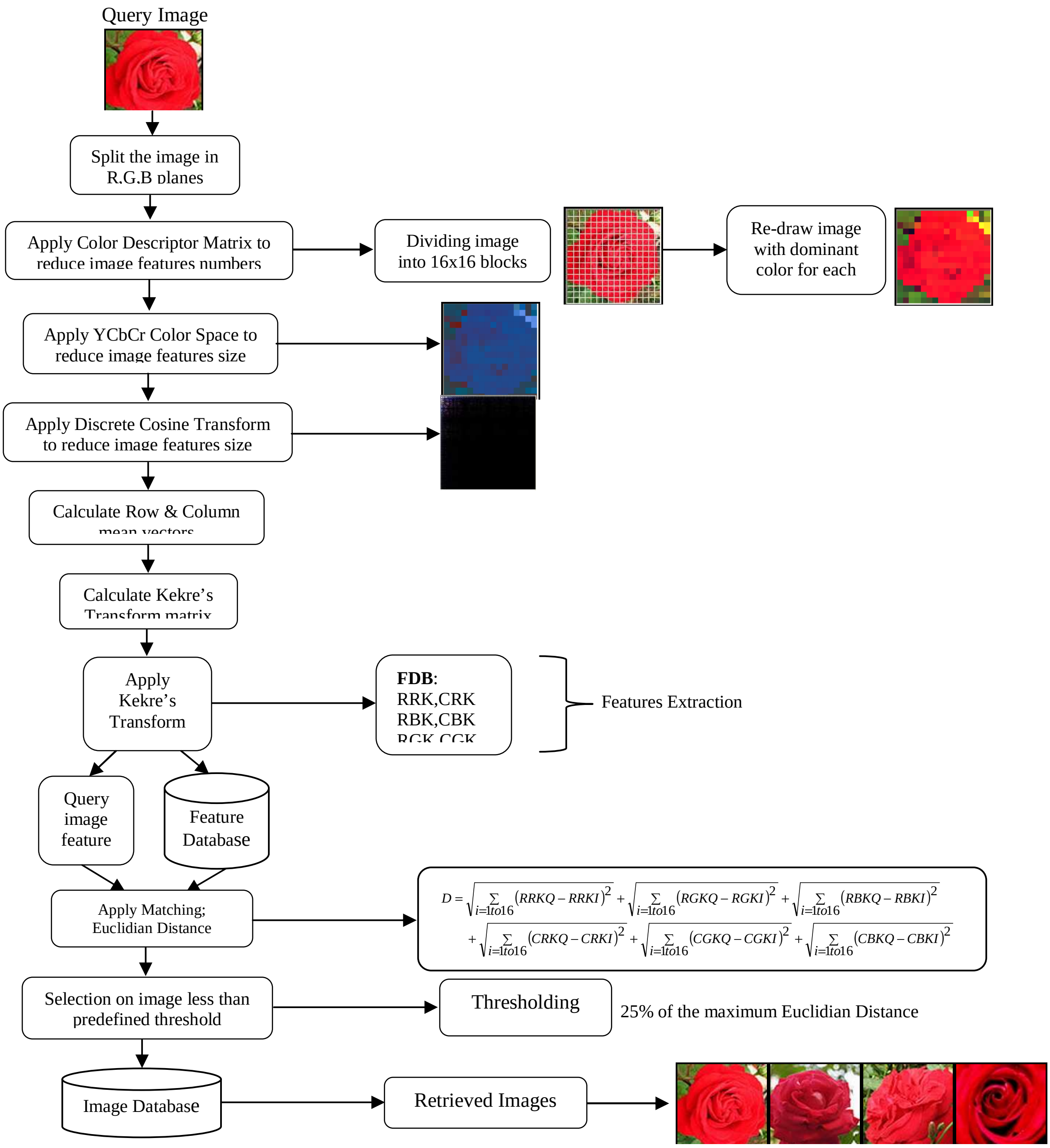


Figure. 2: Flow chart of the proposed algorithm.

(c) Palm fronds

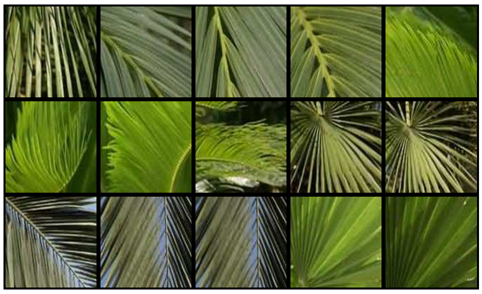
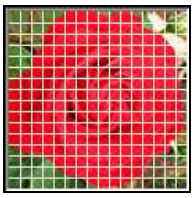
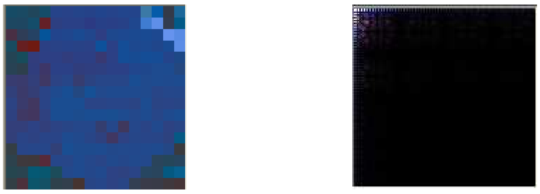


Figure. 3: Experiment Groups (5 categories present in the database).



(a) Input image (b) 16x16 blocks image (c) Image with dominant color

Figure.4: Example of Color Descriptor Matrix application (RGB color space).



(a) (b)

Figure 5:Results of applying: (a):Color Descriptor Matrices on YCbCr color space image,
(b): Descrete Cosine Transform application on an image feature.

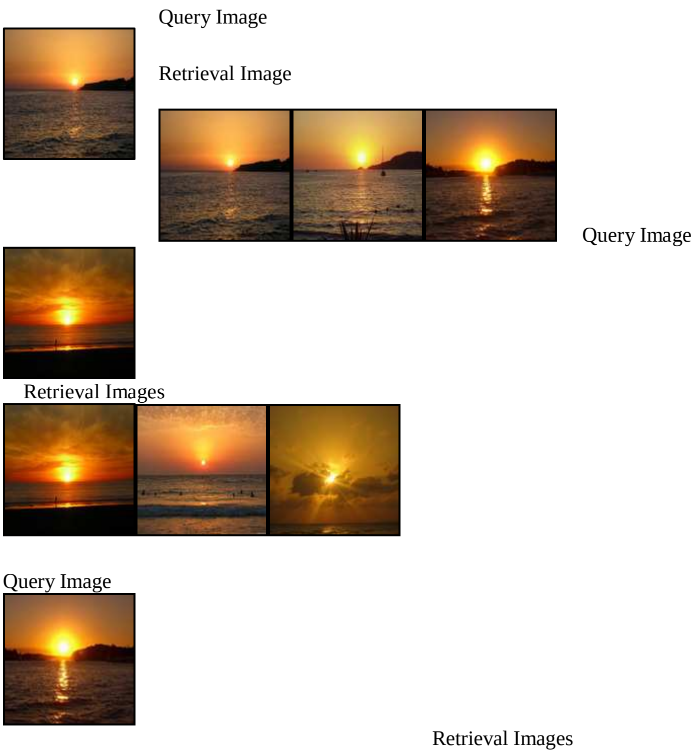


Figure 6: Sample images retrieved for sunset as query image.

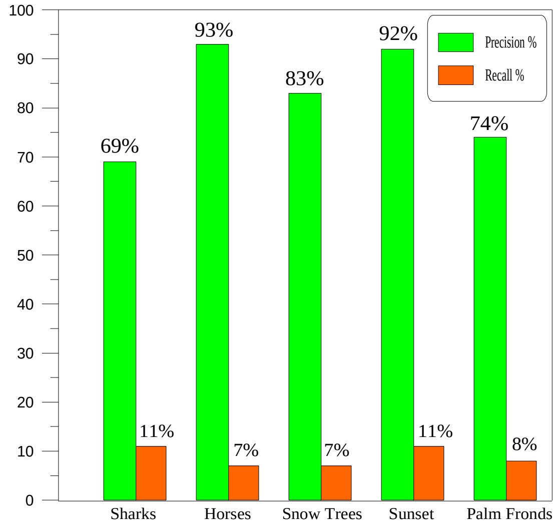


Figure. 7: Performance evaluation of retrieval results for the proposed algorithm over row and column mean vectors at 25% of maximum Euclidian distance (Isolated image groups).

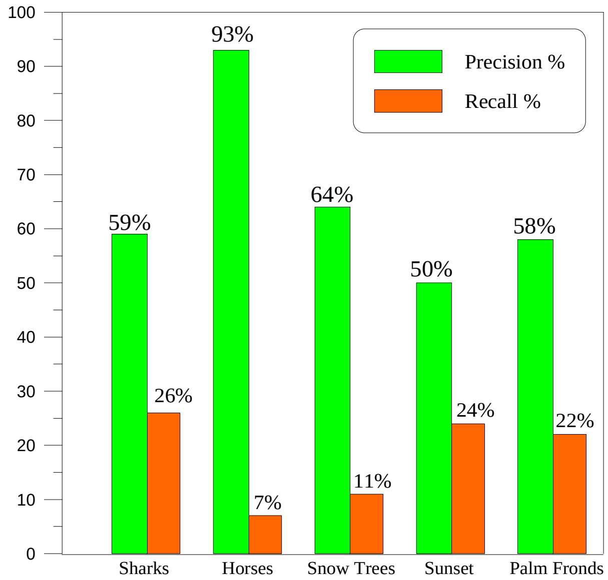


Figure. 8: Performance evaluation of retrieval results for the proposed algorithm over row and column mean vectors at 25% of maximum Euclidian distance (Mixed image groups).