Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 12 (5): 416-427, 2012
ISSN 1812-5654 / DOI: 10.3923/jas.2012.416.427

Median and Laplacian Filters based Feature Analysis for Content Based Image Retrieval Using Color Histogram Refinement Method

Fazal Malik and Baharum Baharudin
Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Malaysia

Abstract: Median and Laplacian filters are used to remove noise from images but some mount of information is also lost. Edge extraction and sharpening methods are used to restore the information lost by median and Laplacian filters. Histogram is used to extract features from filtered image but it has problem that images with diverse appearance will have the same histograms because the spatial information in image does not preserve. To preserve spatial information, we quantize histograms into bins. In each bin the statistical features are calculated using the spatial information of regions. For similarity Sum-of-Absolute Differences (SAD) is used to calculate distance between query and database images. Retrieved images are displayed according to the optimized threshold value of the percentage of maximum of distance values. Experiments on the Corel database give results which show that the statistical features of histogram using spatial information are robust in retrieval of images based on Laplacian filter.

Key words: Content-based image retrieval, sum-of- absolute differences, median filter, Laplacian filter, color histogram

INTRODUCTION

The huge collection of digital images is collected due to the improvement in the digital storage media, image capturing devices like scanners, web cameras, digital cameras and rapid development in internet. Due to these reasons there is a need of an efficient and effective retrieval system to retrieve these images for visual information in different fields of life like medical, medicine, art, architecture, education, crime preventions etc. (Pass and Zabih, 1996; Rani and Saravanan, 2011). In 1970's the first approach was text-based. In this approach images are manually annotated and are retrieved by key words. But this approach has two drawbacks: first is to annotate huge number of images, this requires a lot of human labor and the second is the different subjective perceptions of human for example Lilly flower can be annotated as water lilies, flowers in pond etc. Due to these disadvantages, in 1980's another approach emerged called Content Based Image Retrieval (CBIR) (Pass and Zabih, 1996). CBIR automatically retrieves images by their visual contents such as color, shape and texture instead of keywords (Zhao *et al*., 2009).

Some CBIR systems have been developed under different categories as commercial, production, research and demonstration. For example QBIC, ADL, BDLP, Virage, AltaVista, SIMPLIcity, etc., a detail survey can be found by Veltkamp and Tanase (2002).

The CBIR system can be divided into two steps: feature extraction or indexing and similarity measurement or searching. In the first step of indexing, the low level features are extracted from the images by the algorithm and then these features are represented in a form called feature vector. These feature vectors of all images are stored in a database. In the second step similarity is measured to rank the images by calculating the distance between the query image feature vector and feature vectors of database images using Euclidian distance or other distance methods. If the distance is zero or small then that image is similar or relevant to the query image (Rani and Saravanan, 2011; Mohamed *et al*., 2009). To rank the features for efficient and accurate recognition and retrieval of images the dimension of feature vector must be reduced (Wang and Shitong, 2009).

For the retrieval of mammogram images two steps approach is used. The images are grouped in clusters having same basic information using neural network in the first step. In the second step search for the query image with cluster of images having same characteristics is performed using genetic algorithm (GA) (Jose and Mythili, 2009).

Different algorithms have been developed for the extraction of features. The color histogram features are extracted from images for content based image retrieval to retrieve and classify the images in the database according to the user query image. The histogram features

Corresponding Author: Fazal Malik, Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Malaysia

are mean, standard deviation, skewness, energy and entropy where mean reveals brightness, standard deviation indicates contrast, skewness shows intensity level distribution about mean, energy describes distribution of intensity levels in image and entropy represents distribution of pixel values in intensity levels. The performance o f this algorithm in terms of extraction of features and similarity measurement is high (Sergyan, 2008).

The texture and color features are integrated to retrieve the relevant images from the image database. Color features of image are computed by using histogram technique while texture features are computed by calculating statistical histogram features like entropy, smoothness and uniformity (Thawari and Janwe, 2011).

As image is a matrix of rows and columns, these rows and columns are used to extract statistical features like mean and standard deviation, in image for the retrieval of relevant images in CBIR. Color image is the combination of three color components Red, Green and Blue (RGB). Mean and standard deviation are computed row wise and columns wise in all three components R, G and B of the image and a six dimensions feature vector is calculated. For similarity measurement Euclidean distance is used (Kekre and Patil, 2009).

The color histogram refinement method is used to get the color and shape features of the objects in image. The histogram of the grayscale image is quantized into bins. In each bin the similar colored connected regions are determined. The number of coherent and incoherent pixels in each region of each bin is determined. A pixel is coherent if it is present in the same colored region otherwise incoherent. The number and average of coherent and incoherent clusters are calculated. Additional features are also computed but only in coherent clusters. These features consist of sizes of largest, median and smallest clusters, the major axis length, minor axis length, ellipse angle and variances of largest, median and smallest clusters. These features are not affected by the orientation of image. For distance calculation Euclidean distance is used to retrieve images (Park *et al*., 2008).

There are certain problems with the CBIR. The First problem is that how to represent and describe the visual contents of image to compare a given image with the relevant images in image database since the digital image consists of pixels. These pixels should be described in a way to get most feasible features of image such that to enable the computer to retrieve the most relevant images. The second problem is that how to properly select and use these features in the similarity function to get best performance in terms of precision and accuracy.

The algorithm should be such that to describe the contents of image in the form of certain features which will be represented to get a vector of these features. The aim and objective of the feature vector is to select those features that will give more accuracy in result.

In this study we propose a CBIR algorithm to find the solution of the problems in the CBIR. This algorithm is based on the color histogram refinement method using median and Laplacian filters as preprocessing steps to reduce the noise and provides enhanced sharpened images with more detail information. Color histogram is divided into bins. The number of regions is determined in bins. The statistical moments mean and standard deviation are calculated in each bin by using the areas of regions to get feature vector which is used for image retrieval.

REVIEW OF LITERATURE

For the last decade many methods and algorithms have been developed for CBIR.

A comprehensive review about CBIR of 200 references is given by Smeulders *et al.* (2000). This study discusses the detail working status of CBIR like image type, semantic role, semantic gap and computation of feature extraction, similarity of features, image retrieval and relevant feedback for enhancement of systems.

Islam *et al.* (2010) provides a review of vast research, for the last few decades which has been performed for enhancement of image techniques for image processing, extraction of features for automatic retrieval of images and detection of fingerprint for secure authentication.

Mostly images consist of some noise and unwanted information. They should be removed from images before processing for retrieval by using filters. Different filters method can be used for removal noise. A median filter is applied on images for enhancement as a preprocessing step. Though this filter improves the image quality but it creates another problem that some amount of edge information of objects in images is lost. This edge information is recovered by applying edge extraction method. Then histogram features are extracted from the enhanced filtered image by quantizing histogram into bins and in each bin the average of pixels is computed which are combined to form a feature vector for retrieval of images. The results showed that median filter with edge gives good results (Zhao *et al*., 2009).

To improve the effectiveness of noise removal of median filter (Ge and Song, 2011) proposed Adaptive

Regularized Possibilistic Linear Models Based Median Filter (ARBMF) which combines weight of input signal with weight of out of median filter.

The normalized histograms are quantized in 48 bins in each component of the RGB color image. Thus for each image a feature vector of total 48×3 = 144 features is created. For similarity measurements, Euclidean distance is used to calculate distance between query image feature vector and database image feature vectors. The images are ranked by the similarity distance values and the results show good performance in term of effectiveness (Chakravarti and Meng, 2009).

Two prominent features, texture and color are combined. This combination of features are customized by binary partitioned tree to search the similar images for the given user query. The HSV (Hue Saturation and Value) color space is used for color feature because this color space is close to human eye perception. Color histogram is applied to Hue and Saturation components only. While the Value component which represents brightness of image, is not considered for features extraction. For texture features wavelet technique is used on grayscale image to decompose the image into regions (Mansoori and Jamzad, 2009).

In Murala *et al.* (2009) approach, a comparison is presented by using various texture and color descriptors in a novel approach in which these texture and color features are combined to retrieve similar images for the CBIR. Color Histograms (CH) are generated in all three components R, G and B of RGB image. Histograms are quantizing into 64 bins in each component to get color histogram features. For texture feature calculation, the image is divided into sub-bands by using transformation of Standard Wavelet and Gabor Wavelet. The retrieval results of the techniques, Color Histogram (CH), Wavelet Transform (WT), Gabor Wavelet Transform (GWT), only, Color Histogram (CH) with Gabor Wavelet Transform (GWT) and Color Histogram (CH) with Wavelet Transform (WT), are compared. The best result is given by the combination of Color Histogram (CH) with Gabor Wavelet Transform (GWT).

The critical problems in CBIR are: how to extract and represent low level features from the image and then how to use and select these features as query in comparison to retrieve the similar images from the image database. The solution of these problems is proposed by Choras *et al.* (2007) in which low level features, shape, texture and color are extracted and then these features are integrated to retrieve images from database. This method is started with the extraction of regions of interest (ROIs) using Gabor filter technique. Then in next step features are extracted in these regions efficiently and with high speed. The texture features are extracted by using Gabor features technique with a threshold value. For the extraction of color features in ROIs, two techniques are used, color histograms and color moments in YUV color space. The calculation of shape features is performed in the ROIs by using Zermikes moments. These features provided efficient similarity for the query image with database images.

The region of interest (ROI) is extracted from real-world car images by subtraction of background. Then image with ROI is divided into sub blocks of equal. The spectral texture features which are extracted from blocks are inputted to neural network for classification images (Nagarajan and Balasubramanie, 2008).

The CBIR algorithm will be efficient and effective if it is fast in computation of feature extraction and accurate in the result after similarity calculation. An efficient and effective novel method is proposed by Park *et al.* (2010) in which Global and local color features are extracted. The global color features are extracted by generating histograms in RGB color space while for the local features Genetic Algorithm (GA) is used in HSV color space. Efficiency and accuracy of this method is high as compared to the previous method.

The Rough Set Histograms (RSH) are constructed using rough set theory inn HSV space to get efficient and accurate segmentation of images (Guo-quan and Zhan-ming, 2011).

In Wu and Wei (2010) approach, color and texture features are integrated. The texture features are extracted by using Dual-Tree Complex Wavelet (DT-CWT) transform and Rotated Wavelet Filter (RWF) techniques. Color features are extracted by creating color histogram in two color spaces HSV and RGB. The results of this method are efficient.

Single region is better than whole image as a query example for retrieval of images and SVM is used for classification (Iskandar *et al.*, 2008).

The retrieval of images can be region-based by segmenting into regions of objects of different shapes and sizes using color features but some images would be segmented into large number of regions due high complexity. This will lead to high storage and computation cost (Zhang *et al.*, 2008).

Multi-level features global, local and pixel are extracted and are combined in a big vector for improved retrieval and classification of images by SVM. Principal component Analysis (PCA) is used for the reduction of the big dimensionality of feature vector (Mueen *et al.*, 2007).

The mostly used technique for the extraction of low level information is histogram. This technique is modified to get improved and refined histogram. This method also called color histogram refinement. This refined histogram method divides the histogram into buckets of pixels and each bucket is divided into classes of pixels of same local properties. The histograms of query image are compared with database images of pixels of same local properties in buckets. The performance is better than normal histogram using spatial information of pixels in buckets (Liu *et al*., 2007).

PROPOSED METHODOLOGY

In this study our proposed CBIR algorithm is based on analysis of statistical color histogram features using median, edge extraction and Laplacian filters. The statistical features are extracted in histograms using the spatial information for each filter separately. Before applying histogram to the image for the extraction of features, preprocessing is performed. The block diagrams of proposed algorithm step wise and as a whole are shown in Fig. 1 and 2.

Preprocessing: It can be seen in Fig.1 that the input RGB image is acquired. As the color images consist of three components therefore the computational cost of feature extraction will be high. To reduce computation cost the color images are converted into grayscale (Anjum and Javed, 2007). The grayscale image is converted to Histogram Equalized (HE) image, to make the image's intensity levels equal to get high contrast image. Then median, edge extraction and Laplacian filters are applied to get more enhanced image. The preprocessing steps using Laplacian filter is shown in Fig. 3.

Median filter: Images are also consisted of noise. Before applying processing techniques on images to extract low level features, the images are needed to be pre-processed to remove unwanted information and to get enhanced

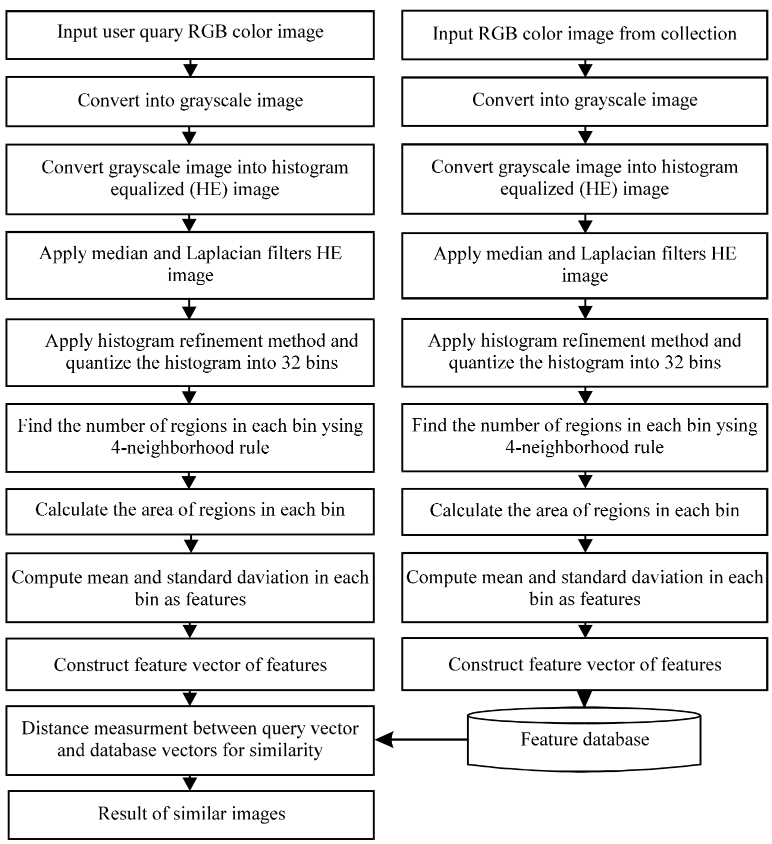

Fig. 1: Step wise block diagram of proposed method

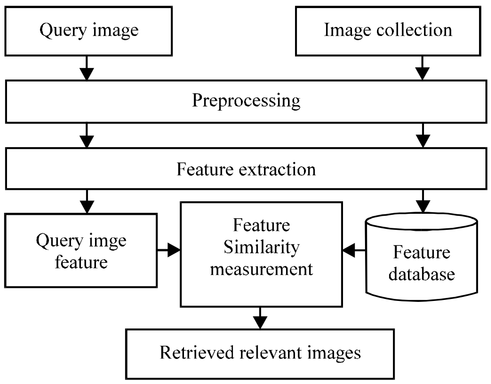

Fig. 2: Block diagram of the entire proposed process

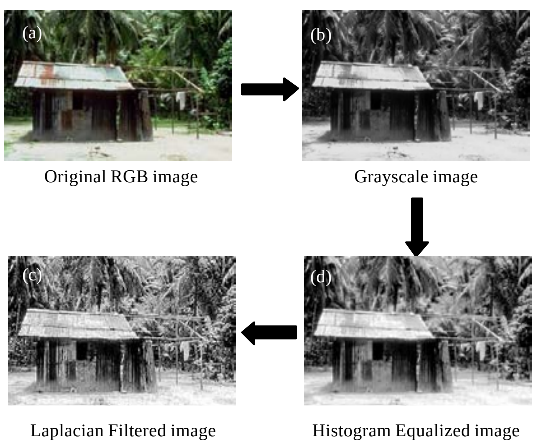

Fig. 3(a-d): Preprocessing of proposed method using Laplacian filter

images with relevant information only. To remove noise a technique is applied on image called filter. The process to remove noise from image by applying filter is called filtration. Median filter is based on neighborhood operations. It consists of a window which is encompassed over image to order (rank) pixels in the image area and then replace central pixel with determined values. Median filter replaces the value of a pixel by median of the gray levels in the neighborhood of that pixel (Gonzalez *et al*., 2004). This filtered image is used for feature extraction.

Edge extraction: However median filtering removes the noise from images but some black specks are left around the border. These black points on border are due to the default padding of zeros (0's). Some amount of information in image like edge information is lost (Gonzalez *et al*., 2004). To restore the edge information of median filtered image, a technique called canny edge detection is used to determine the edge information in image before applying median filter (Zhao *et al*., 2009).

0	1	0
1	-4	1
0	1	0

Fig. 4: Laplacian filter 3×3 mask

This is the most powerful edge detector. This technique detects two edge points, strong and weak using two threshold values T1 and T2 such that T1<T2. If the pixel values greater than T2 then the edge values are strong and if pixel values are in between T1 and T2 then these are called weak edge pixels. At last the canny technique connects weak edges to the strong edges by using 8-connection. Edge detection technique is used to determine edges before applying median filter to image. The edge information of the median filtered image is restored by the already extracted edge information (Gonzalez *et al*., 2004). Thus the features are extracted from the median filtered image with edge extraction method using spatial information and are used for image retrieval. The results are analyzed on the basis of this filter technique.

Laplacian filter: Laplacian filter is a linear filter. In this filter a window or mask with some values works with values of image pixels in the neighborhood. The values in filter window are called filter coefficients. The result of this filter is the sum of products of the filter coefficients and the corresponding image pixel values. This filter gives an image with strong edges.

Let f(x,y) is an original image and $\nabla^2 f(x,y)$ is Laplacian image such that:

$$\nabla^2 f(x,y) = \frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2} \tag{1}$$

$$\frac{\partial^2 f}{\partial x^2} = f(x+1,y) + f(x-1,y) - 2f(x,y) \tag{2}$$

$$\frac{\partial^2 f}{\partial y^2} = f(x,y+1) + f(x,y-1) - 2f(x,y) \tag{3}$$

$$\nabla^2 f = [f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1)] - 4f(x,y) \tag{4}$$

To get the filtered image, at all points (x, y) in (4) can be convolved with 3×3 mask in Fig. 4.

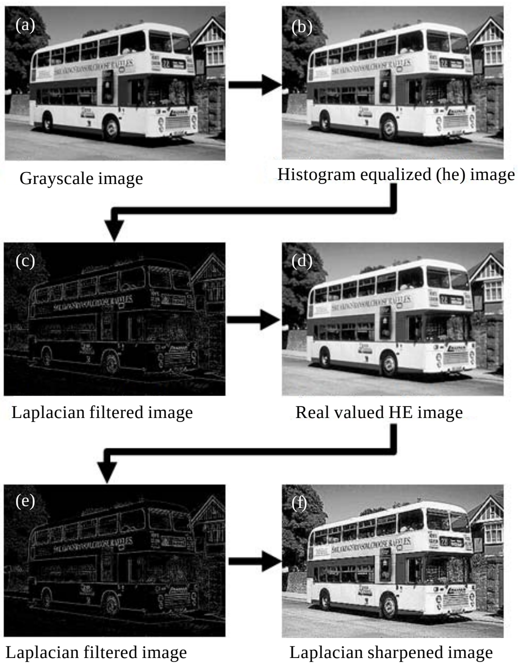

Fig. 5(a-f): Sharpening process using Laplacian filter

Before using the Laplacian filter, the image is converted into grayscale as shown in Fig. 5a, to get a single component of image and then it is again converted into histogram equalized image f to get enhanced image in Fig. 5b. The image f is filtered with Laplacian filter to get a filtered image. g_1 with edges of objects in image as show in Fig. 5c. But all pixel values in g_1 are positive and these values must be negative because of the negative value -4 at the center of the mask as shown in Fig. 4. For this purpose the histogram equalized image f is converted into real valued image f_2 as in Fig. 5d. This image f_2 is again filtered with Laplacian filter to get image g_2 with edges information as shown in Fig. 5e. But during the filter process of image g_2 some amount of information is lost. To restore this information and get an enhanced and sharpened image g, this Laplacian filtered image g_2 is subtracted from the real valued image as calculated:

$$g = f_2 - g_2 \tag{5}$$

where, g is the sharpened and enhanced image with detailed information as shown in Fig. 5f. This process is also called sharpening of image (Gonzalez *et al*., 2004). The features are extracted from the image g for retrieval and analysis.

Color histogram: Color is the most prominent and important feature of image because it is the dominant part of human visual perception. It is used to retrieve images in CBIR. For this purpose various color methods have been used. In these methods color histogram is popular one and mostly used method. Color histogram has the frequency of occurrence of each color in an image. Color histogram is divided into bins of color and each pixel having a specific color belongs to a color bin of that color. It has the characteristics that it represents the global information of the image (Jin, 2009; Swain and Ballard, 1991). These global features representation of the image is very useful in the queries in which the matching of the images is based on the whole appearance. Color histograms are very fast in computation of features (Park *et al*., 2008) . It has no effect to the small changes in the scenes. It is useful and widely used for the images which require invariance in translation and rotation (Park *et al*., 2010). That query which requires retrieving the images with the same scenes but with different circumstances of illuminations then color histogram is not a suitable technique for such queries. Spatial information in the color histogram does not maintain due to which the same histograms will be extracted for the images with diverse appearances. In other words an image with many very small red spots has a histogram similar with the histogram of the image which has a single large red area (Jin, 2009).

To preserve the spatial information in histogram an algorithm is proposed by Liu *et al*. (2007) in which color histograms refinement method is used such that pixels of the same color are classified into coherent and incoherent clusters. This method is also called color coherence vector (CCV).

Quantization: The value of each pixel indicates a specific color that can be represented in various color spaces of three components like; Red, Green and Blue (RGB) and Hue, Saturation and Value (HSV). Each component R, G and B, in RGB color space consists of 0 to 255 pixel values or intensity levels. The intensity levels for color histogram will be large using these three color components due to which the computation speed for feature extraction will be increased. To reduce the computation speed, the RGB color image is converted into grayscale image of only 0 to 255 levels.

In our proposed algorithm we do not use the 256 levels of the grayscale image. We will further reduce it to certain levels to increase the speed of computation. The process to divide the image in levels or bins of frequencies of same color is called quantization. In this paper we quantize the filtered grayscale image of 256 levels into 32 bins so that to reduce the computations as shown in Fig. 1 The histogram is then quantized into L = 32 bins such that:

$$H= \{h(b_1), h(b_2) \ldots h(b_L)\} \tag{6}$$

where, $h(b_i)$ is the frequency of pixel values in bin b_i and H is the histogram of L bins.

Feature extraction: For feature extraction the color histogram refinement technique is used. Color histogram is quantized into 32 bins to increase the speed of extraction of feature. Each bin is divided into connected regions of pixels using 4-neighborhood rule. The number of regions in each bin is determined. Then the area of each region is calculated. Two color moments are used to calculate features, the first-order moment called mean and second order moment called standard deviation. The mean represents the brightness of image and standard deviation represents the contrast. The dark image has low mean and bright image has high mean. The low contrast image has low and high contrast has high standard deviations. The means and standard deviations are calculated in each bin using the areas of regions.

The statistical color features mean and standard deviation are calculated in histogram bins of H. Let μ_j is the mean and $\acute{o}_j$ is the standard deviation in a particular bin j, where j = 1, 2, 3..., L and then these two features can be calculated by using the statistical measurements (Jia and Wang, 2003; Wang and Jia, 2000) as:

$$\mu_j = \frac{1}{N}\sum_{i=1}^{N} A_{ji} \tag{7}$$

$$\sigma_j = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(A_{ji} - \mu_j)^2} \tag{8}$$

where, A_{ji} is the area of the ith region in jth bin and N is the total number of regions in each bin j.

After the calculation of color features, the feature vector FV is constructed as:

$$FV= \{\mu_1,\mu_2,\mu_3 \ldots\ldots \mu_L,\sigma_1,\sigma_2,\sigma_3 \ldots\ldots \sigma_L\} \tag{9}$$

For all images in database the feature vectors FVs are computed and stored in database to be retrieved. The feature vector of the user query is constructed in the same way and compared with feature vectors of database for similarity and retrieval of relevant images as show in Fig. 1.

Similarity measurements: Once the database of the images with feature vectors is created, then the user can give an image as a query to retrieve the relevant images from the database. The feature vector of the query image is computed by using the same algorithm that is used for feature vectors of images in database.

For the measurement of the similarity between the query image and the database images, the difference between query image feature vector and database image feature vectors is calculated. For this purpose the Sum-of-Absolute Differences (SAD) (Kodituwakku and Selvarajah, 2010) is used to calculate the difference between the query image feature vector and database feature vectors for the similarity.

Let query feature vector is represented by Q_f and database feature vector by D_f then the distance is calculated as:

$$\Delta d = \sum_{i=1}^{n} |(Q_f(i) - D_f(i))| \tag{10}$$

where, n is the number of features in feature vector, i = 1, 2... n. Both query and database images are same for $\Delta d = 0$ and small values of Δd represent most similar images to the query image. The distance Δd of the query image is calculated with all images in database. The distance values are arranged in ascending order. The smallest distance values will be on top which correspond to the most similar images and values at bottom correspond to irrelevant images. The top most images are displayed to the users which are the required images according to the threshold value of the maximum of distance.

IMAGE RETRIEVAL

Retrieval of the images is performed in three phases for the analysis of histogram features based on three filter methods, median, median with edge extraction and Laplacian filters. Before feature extraction some preprocessing steps are performed on image to enhance up to certain level such that the required information in image becomes more apparent for extraction of features. When the input RGB image is acquired then in the first step it is converted into grayscale for the extraction of features. Grayscale requires fewer computations as compared to color image. The grayscale image is converted to histogram equalized image. Histogram equalization is a technique to convert image's intensity levels into equal levels such that to enhance contrast and image has more details. For further enhancement of image median and Laplacian filters are used. Image retrieval for median, median with edge extraction and Laplacian will be performed in three phases using histogram method.

Image retrieval phases

- **Phase 1:** In this phase histogram equalized image is filtered with median filter to further enhance the image. For features extraction color histogram refinement method is applied on filtered image. The filtered image is quantized into histograms of 32 bins using (6). In each quantized bin the number of connected regions is determined. Then mean and standard deviation are calculated in each bin using the area of regions by using (7) and (8). By using L = 32 then total 64 features are extracted and a feature vector FV is generated for the image using (9). All the images are stored in database with feature vectors. The feature vector of the user query is generated in the same way. This query image feature vector is compared with database and distance is computed between them using (10). Thus we get the retrieval result for median filter using histogram technique
- **Phase 2:** In this phase median filter is used with edge extraction method. As during median filtration some amount edge information is lost. To restore edge information, canny edge detection method is used after median filtration. The features for database image and query image are extracted in the same way as in phase-1 using color histogram refinement method. The distance is computed between query image and database image using (10) to retrieve relevant images
- **Phase 3:** The Laplacian filter is applied on histogram equalized image in this phase. The Laplacian filter gives a sharpened and enhanced image using real valued image. The features are extracted in same way as in Phase-1 and Phase-2 using color histogram refinement method. The distance is computed between images using (10)

Threshold value for query: The distance between the query image and database image is calculated by using (9). The distance values for all images in the database are calculated and are stored in a list in ascending order in main memory of computer. Top most values represent the relevant images. Various percentage values of maximum of calculated distance values are selected as threshold values.

Table 1: Average precision and recall for 10 categories using median filter

Categories	Total retrieved	Relevant retrieved	Precision (%)	Recall (%)
People	36	11	31	11
Beaches	25	9	36	9
Buildings	36	8	22	8
Buses	36	11	31	11
Dinosaurs	25	16	64	16
Elephants	36	7	19	7
Roses	25	8	32	8
Horses	36	18	50	18
Mountains	25	8	32	8
Foods	36	10	28	10
Average	32	11	34	11

Table 2: Average precision and recall for 10 categories using median filter with edge extraction

Categories	Total retrieved	Relevant retrieved	Precision (%)	Recall (%)
People	36	10	28	10
Beaches	25	7	28	7
Buildings	25	7	28	7
Buses	25	9	36	9
Dinosaurs	21	14	67	14
Elephants	25	6	24	6
Roses	23	6	26	6
Horses	36	18	50	18
Mountains	21	8	38	8
Foods	36	10	28	10
Average	27	10	35	10

Table 3: Average precision and recall for 10 categories using Laplacian filter

Categories	Total retrieved	Relevant retrieved	Precision (%)	Recall (%)
People	36	16	44	16
Beaches	25	8	32	8
Buildings	25	9	36	9
Buses	36	13	36	13
Dinosaurs	23	19	83	19
Elephants	36	9	25	9
Roses	25	8	32	8
Horses	25	10	40	10
Mountains	25	6	24	6
Foods	36	16	44	16
Average	29	11	40	11

Query image for the median filter is using threshold value 0.2% of maximum of distance to display the result images to the users. This threshold value displays an average 32 images as a query result as show in Table 1.

Query image for the median filter with edge extraction is using threshold value 0.2% of maximum of distance to display the result images to the users. This threshold value displays an average 27 images as a query result as show in Table 2.

In Laplacian filter, query image is using a threshold value 0.4% of maximum of distance. This threshold value displays an average 29 images as a query result as show in Table 3.

RESULTS

The proposed CBIR algorithm is tested with a benchmark Corel image database which is provided by

Jia and Wang (2003) and Wang and Jia (2000). This database is free for the researchers to test their algorithms for CBIR systems. There are 1000 images in the database. The database consists of 10 categories of images which include people, beaches, buildings, buses, dinosaurs, elephants, roses, horses, mountains and foods. Each category has 100 images. The images are in PEG format with dimensions of 256×384 and 384×256 pixels. When the user put the query image, then the relevant images are displayed according to the threshold value of distance.

The proposed algorithm is started with preprocessing of images. The images are converted into grayscale images. Histogram equalization technique is applied on grayscale images. To reduce computation the histogram equalized images are quantized into 32 bins. The statistical features mean and standard deviation are computed in each bin of images and generated feature vectors. These feature vectors are stored in database with images as a first step of proposed algorithm. The feature vector of query image is also constructed and compared with database images by computing distance between them. The result is displayed to the user according to the threshold value of the maximum of distance values.

Evaluation of measurement: The efficiency and effectiveness of the retrieval systems are determined by measuring their computational complexity (Xie and Wang, 2011). There are two measurements to evaluate the performance of CBIR systems: precision and recall (Thawari and Janwe, 2011). These measurements can be defined as:

Precision: Precision is the measurement of the retrieved relevant images to the query total retrieved images:

$$\text{Precision} = \frac{A}{B} \tag{11}$$

where, A is "the relevant retrieved images" and B is "the total retrieved images".

Recall: Recall is the measurement of the retrieved relevant images to the total database images:

$$\text{Recall} = \frac{A}{C} \tag{12}$$

where, A is "the relevant retrieved images" and C is "the total number of relevant images in database".

For the analysis of features a set of query images were used of all 10 categories of images in the three phases of retrieval. The average precision and recall are calculated using (11) and (12) for all three retrieval phases. In first phase median filter is applied in the proposed method. For the image results the threshold values 0.2% of the maximum distance is used. An average 32 images are displayed to the user having 11 relevant with precision 34% and recall 11%. The recall is for total 100 images in each category. The results are show in Table 1.

In second phase median filter with edge extraction is applied with threshold value 0.2% and average total retrieved images are 27 with 35% precision and 10% recall as show in Table 2.

In third phase Laplacian filter is applied with threshold value 0.4% and average total retrieved images are 29 with 40% precision and 11% recall as show in Table 3.

Table 4 shows the average precision and recall of the proposed algorithm using three filter methods. It can be seen that the proposed algorithm using Laplacian filter gives good results as compared to median and edge filter methods.

Figure 6-9 are the screen shots of the results of user queries. Each figure has of a query image and the relevant retrieved images from the database by using the proposed algorithm. The top single image is the query image and below 9 are the relevant images. The results show that proposed algorithm has good retrieval accuracy.

Table 4: Comparison of average precision and recall of proposed algorithm based on three filter

Techniques	Precision (%)	Recall (%)
Median filter	34	11
Median with edge detection	35	10
Laplacian filter	40	11

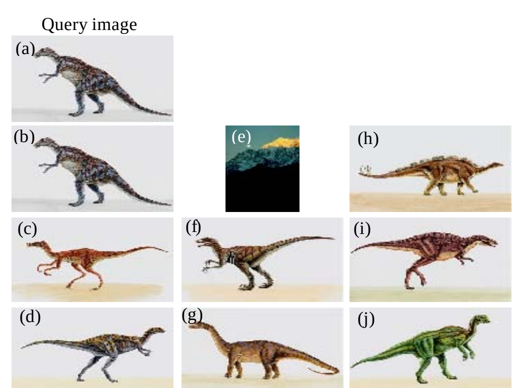

Fig. 6(a-j): Query results for dinosaurs using Laplacian filter

Fig. 7 (a-j): Query results for people using Laplacian filter

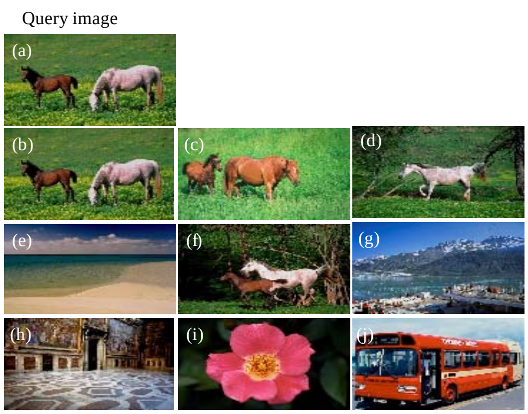

Fig. 8(a-j): Query results for horses using Laplacian filter

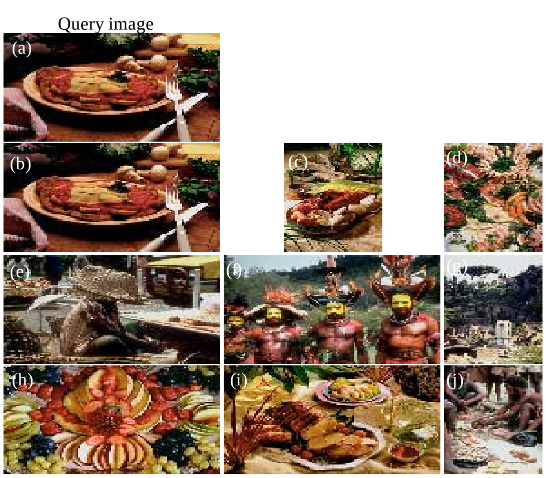

Fig. 9(a-j): Query results for foods using Laplacian filter

CONCLUSION

In this Study a CBIR algorithm is proposed which is based on the features of the different filter methods using color histogram refinement method for feature extraction. Median, median with edge extraction and Laplacian filter methods were applied on grayscale images for noise removal before applying histogram method. During median filtration edge information is lost which is restored by edge detection method while in Laplacian some amount information is also lost which is restored by subtracting the Laplacian image from grayscale image to get more enhanced sharpened image. The statistical features mean and standard deviation of quantized histograms are calculated using the spatial information of connected regions. These statistical features are used for the retrieval of relevant images. These features do not depend upon the orientation of image. In this algorithm spatial information in images are preserved. For the similarity measurement, Sum of Absolute Differences (SAD) is used to determine the difference between query image and database images. The images are ranked according to the distance measurements in ascending order. To display these ranked images to the users, we use a different threshold query approach such that an optimized threshold value is calculated by calculating percentage of the maximum value of distance measurements. Then relevant images are displayed to the users according to the optimized threshold value of the maximum of distance values.

We have used three phases for the feature analysis of filter methods in image retrieval. In first phase median filter, in the second phase median with edge extraction and in third phase Laplacian filter, are used by proposed method. The results show that the algorithm provides good result based on Laplacian filter.

REFERENCES

Anjum, M.A. and M.Y. Javed, 2007. Multiresolution and varying expressions analysis of face images for recognition. Inform. Technol. J., 6: 57-65.

Chakravarti, R. and X. Meng, 2009. A study of color histogram based image retrieval. Proceedings of the 6th International Conference on Information Technology: New Generations, April 27-29, 2009, Las Vegas, NV., pp: 1323-1328.

Choras', R.S., T. Andrysiak and N. Choras, 2007. Integrated color, texture and shape information for content-based image retrieval. Pattern Anal. Appl., 10: 333-343.

Ge, H. and W. Song, 2011. A novel adaptive regularized possibilistic linear models based median filter ARBMF for image noise suppression. Inform. Technol. J., 10: 2260-2267.

Gonzalez, R.C., R.E. Woods and S.L. Eddins, 2004. Digital Image Processing Using MATLAB. Pearson Education, Prentice Hall, New Jersley, USA.

Guo-quan, Z. and L. Zhan-ming, 2011. Research on color image segmentation based on RS for intelligent vehicle navigation. Inform. Technol. J., 10: 1432-1436.

Iskandar, D.A., J. Thom and S. Tahaghoghi, 2008. Content-based image retrieval using image regions as query examples. Proceeding of the Nineteenth Australasian Database Conference (ADC 2008), January 22-25, 2008, Wollongong, NSW, Australian Computer Society Australia, CRPIT, 75. Fekete, ACS, pp: 38-46.

Islam, M.R., M.S. Sayeed and A. Samraj, 2010. Technology review: Image enhancement, feature extraction and template protection of a fingerprint authentication system. J. Applied Sci., 10: 1397-1404.

Jia, L. and J.Z. Wang, 2003. Automatic linguistic indexing of pictures by a statistical modeling approach. IEEE Transac. Pattern Analy. Mach. Intellig., 25: 1075-1088.

Jin, C., 2009. A statistical image retrieval method using color invariant. Proceeding of the 6th International Conference on Computer Graphics, August 11-14, 2009, Imaging and Visualization, Tianjin, pp: 355-360.

Jose, T.J. and P. Mythili, 2009. Neural network and genetic algorithm based hybrid model for content based mammogram image retrieval. J. Applied Sci., 9: 3531-3538.

Kekre, H.B. and K. Patil, 2009. Standard deviation of mean and variance of rows and columns of images for CBIR. Int. J. Comp. Inf. Eng., 3 : 1-4.

Kodituwakku, S.R. and S. Selvarajah, 2010. Comparison of color features for image retrieval. Indian J. Comp. Sci. Engine., 1: 207-211.

Liu, Y., D. Zhang, G. Lu and W.Y. Ma, 2007. A survey of content-based image retrieval with high-level semantics. Pattern Recogn., 40: 262-282.

Mansoori, Z. and M. Jamzad, 2009. Content based image retrieval using the knowledge of texture, color and binary tree structure Proceeding of the Canadian Conference on Electrical and Computer Engineering, May 3-6, 2009, St. John's, NL, 2009. CCECE '09, pp: 999-1003.

Mohamed, A., F. Khellfi, Y. Weng; J. Jiang and S. Ipson, 2009. An efficient image retrieval through DCT histogram quantization. Proceeding of the International Conference on CyberWorlds 2009. September 7-11, 2009, Bradford, pp: 237-240.

Mueen, A., M.S. Baba and R. Zainuddin, 2007. Multilevel feature extraction and X-ray image classification. J. Applied Sci., 7: 1224-1229.

Murala, S., A.B. Gonde and R.P. Maheshwari, 2009. Color and texture features for image indexing and retrieval. Proceeding of the IEEE International Advance Computing Conference, 2009. March 6-7, 2009, Patiala, pp: 1411-1416.

Nagarajan, B. and P. Balasubramanie, 2008. Neural classifier for object classification with cluttered background using spectral texture based features. J. Artif. Intell., 1: 61-69.

Park, J., Y. An, G. Kang, W. Rasheed, S. Park and G. Kwon, 2008. Defining a new feature set for content-based image analysis using histogram refinement. Int. J. Imag. Syst. Technol., 18: 86-93.

Park, S.-S., Y.-G. Shin and D.-S. Jangcor1, 2010. A novel efficient technique for extracting valid feature information. Expert Syst. Appl., 37: 2654-2660.

Pass, G. and R. Zabih, 1996. Histogram refinement for content-based image retrieval. Proceeding of the 3rd IEEE Workshop on Applications of Computer Vision, 1996 WACV '96, December 2-4, 1996, Sarasota, FL , USA, pp: 96-102.

Rani, P.M.N. and A. Saravanan, 2011. Effective semantics intensive image retrieval using fast indexing technique. Res. J. Inform. Technol., 3: 114-122.

Sergyan, S., 2008. Color histogram features based image classification in content-based image retrieval systems. Proceeding of the 6th International Symposium on Applied Machine Intelligence and Informatics, 2008. SAMI 2008, January 21-22, 2008, Herlany, pp: 221-224.

Smeulders, A.W.M., M. Worring, S. Santini, A. Gupta and R. Jain, 2000. Content-based image retrieval at the end of the early years. IEEE Transac. Pattern Analy. Mach. Intellig., 22: 1349-1380.

Swain, M.J. and D.H. Ballard, 1991. Color indexing. Int. J. Comput. Vision, 7: 11-32.

Thawari, P.B. and N.J. Janwe, 2011. Cbir based on color and texture. Int. J. Inform. Technol. Knowledge Manag., 4: 129-132.

Veltkamp, R.C. and M. Tanase, 2002. Content based image retrieval systems: A survey. Tech. Rep. No. UU-CS-2000-34. Department of Computer Science, Utrecht University, Utrecht, the Netherlands.

Wang, J.Z. and L. Jia, 2000. SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. Adv. Visual Inf. Syst., 1: 171-193.

Wang, X. and W. Shitong, 2009. Feature ranking by weighting and ISE criterion of nonparametric density estimation. J. Applied Sci., 9: 1014-1024.

Wu, J. and Z. Wei, 2010. Color and texture feature for content based image retrieval. Int. J. Digital Content Technol. Applic., 4: 43-49.

Xie, Z. and S. Wang, 2011. A new motion segmentation method for dynamic scenes. Inf. Technol. J., 10: 1942-1949.

Zhang, M., Z. Shang and J. Shen, 2008. Research on the image segmentation necessity for regions-based image processing. Inform. Technol. J., 7: 269-276.

Zhao, H., P. Kim and J. Park , 2009. Feature analysis based on edge extraction and median filtering for CBIR. Proceeding of the 11th International Conference on Computer Modelling and Simulation, 2009. UKSIM '09, March 25-27, 2009, Cambridge, pp: 245-249.