I.J. Image, Graphics and Signal Processing, 2020, 6, 19-32
Published Online December 2020 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2020.06.03

A Comparative Evaluation of Feature Extraction and Similarity Measurement Methods for Content-based Image Retrieval

S.M. Mohidul Islam
Computer Science and Engineering Discipline, Khulna University, Khulna, Bangladesh
E-mail: mohid@cse.ku.ac.bd

Rameswar Debnath
Computer Science and Engineering Discipline, Khulna University, Khulna, Bangladesh
E-mail: rdebnath@cseku.ac.bd

Received: 15 March 2020; Accepted: 02 July 2020; Published: 08 December 2020

Abstract: Content-based image retrieval is the popular approach for image data searching because in this case, the searching process analyses the actual contents of the image rather than the metadata associated with the image. It is not clear from prior research which feature or which similarity measure performs better among the many available alternatives as well as what are the best combinations of them in content-based image retrieval. We performed a systematic and comprehensive evaluation of several visual feature extraction methods as well as several similarity measurement methods for this case. A feature vector is created after color and/or texture and/or shape features extraction. Then similar images are retrieved using different similarity measures. From experimental results, we found that color moment and wavelet packet entropy features are most effective whereas color autocorrelogram, wavelet moment, and invariant moment features show narrow result. As a similarity measure, cosine and correlation measures are robust in maximum cases; Standardized L_2 in few cases and on average, city block measure retrieves more similar images whereas L_1 and Mahalanobis measures are less effective in maximum cases. This is the first such system to be informed by a rigorous comparative analysis of the total six features and twelve similarity measures.

Index Terms: RST invariant, color, texture, shape, similarity, comparative evaluation

1. Introduction

The massive use of the internet with its huge amount of information has led to a high focus on the science of information retrieval as a mean to organize, store, represent and access information items. Image is one of the most used items for information retrieval. The systems such as search engines, e-business systems, online tutoring system, GIS, image archive, and so on involve retrieving image data based on pictorial content [1]. Manual browsing the database to search for similar images would be impractical since it takes a lot of time and requires human intervention, and the performances of indexing and retrieval vary with personal description of the image. The more practical way is to use Content-Based Image Retrieval (CBIR) technology. “Content-based" means that the searching process will analyze the actual contents of the image rather than the metadata such as keywords, tags, and/or descriptions associated with the image [2]. Finding similar images is indeed a challenging task since thousands of images are involved. The problems for searching a large image database are: 1) the computational time required for feature extraction is quite high 2) the high-dimensional feature vectors cannot be mapped to an optimized access method for fast and accurate indexing 3) difficult to develop Rotation, Scaling and Translation (RST) invariant system [3]. There exist several methods such as [3-7] for large dataset searching; however, performances considering above all problems still are not satisfactory.

In this paper, various RST invariant and computationally effective color, texture, and shape features are used. Color is the visual perceptual property corresponding in humans to the categories called red, green, blue, yellow, etc. [8]. Texture represents the surface structure, observes surface homogeneity, and corresponds to visual perception which is defined by six visual texture properties - contrast, roughness, regularity, directionality, coarseness and line likeness [9]. It also describes the relationship between the surfaces to the surrounding environment [10]. Shape is the contours and shapes of objects represented in the image. Shape features are less developed than their color and texture counterparts because of the inherent complexity of representing shapes [11]. For searching an image in an image database, these features values are extracted for query image and images in the database. All the features of an image form a feature vector. Then these feature vectors are evaluated using similarity measures to find the similar images.

Our research objective is to find out the most effective feature extraction method(s) and similarity measurement method(s) in case of content based image retrieval, which will help the future researchers. To do this, we undertook a rigorous comparison of two different color feature extraction methods- color moment and color autocorrelogram; three different texture feature extraction methods- wavelet moment, wavelet energy, and wavelet packet entropy, and one shape feature extraction method- Hu's seven invariant moments. Again, we focused here on twelve measures, they are: L_1, L_2, Standardized L_2, Normalized L_2, Mahalanobis, City block, Minkowski, Chebyshev, Cosine, Correlation, Spearman's rank correlation, and Relative deviation. These are chosen because of their extensive and successful applications to many datasets in CBIR. Several methods [2-7], [10-13] exist that use some of the RST invariant feature extraction methods and/or similarity measurement methods mentioned above, but a comprehensive evaluation applying all these methods does not exist. Thus it is not clear from the literature which feature or which similarity measure performs better among the many available alternatives. It is also currently poorly understood, what are the best combinations of feature extraction and similarity measurement methods in CBIR. The present study allows determinations of the features as well as similarity measures that perform best for similar image retrieval among the many available alternatives. It also allows identification of the best performing combinations of them.

For experiments, we used mostly applied Wang database [14] which contains ten categories of image, hundred images in each category. From the experimental results, we found that color moment and wavelet packet entropy features are most effective features whereas color autocorrelogram, wavelet moment, and invariant moment features show worse performance among all features used here. Again, as a similarity measure, cosine and correlation measures show the best accuracy in maximum cases; Standardized L_2 in few cases and on average, city block measure retrieves more similar images, whereas L_1 and Mahalanobis measures show the worse result in maximum cases.

The remainder of the paper is organized as follows. In section II, the materials and methods are described in details. The Results and comparative evaluations are outlined in Section III and we draw our conclusion in the last section.

2. Materials and Methods

For content based Image retrieval, we first extract the image feature using various feature extraction methods and then similar image are retrieved using various similarity measurement methods, which in turn will help to find the effective feature extraction method(s) and similarity measurement method(s). Thus the Research involves two major steps: Low-level content or visual feature Extraction from images and similar image retrieval of query image from the database.

A. Feature Extraction

Feature extraction involves reducing the number of resources required to describe a large set of data. We have used rotation, and scaling invariant color moment or color autocorrelogram for color feature extraction with/without RST invariant wavelet moment and/or wavelet energy and/or wavelet packet entropy for texture feature extraction with/without RST invariant Hu's invariant moment for shape feature extraction. Thus the query image can be RST invariant with respect to the images in the database. It means that the method can find similar images of query image from a database that are geometrically transformed. The algorithm for feature extraction is shown in *Algorithm 1.*

```
Algorithm 1: Algorithm for feature extraction
Step 1: Load the resized image.
Step 2: Find the color features of the image and create a color feature vector, and/or
Step 3: Find the texture feature of the same image and create a texture feature vector, and/or
Step 4: Find the shape feature of the same image and create a shape feature vector.
Step 5: Combine the results of Step 2 to Step 4 to create final feature vectors.
Step 6: Do the step 1 to 5 for all images of the database and store it.
```

In the following, we describe how to color, texture and shape features are extracted and a feature vector is created.

1) Color Feature Extraction using Color Moments: The basis of color moment lays in the assumption that the distribution of color in an image can be interpreted as a probability distribution. Probability distributions are characterized by a number of unique moments [8]. The greatest advantage of using color moment comes from the fact that there is no need to store the complete color distribution; this greatly speeds up image retrieval since there are fewer features to compare [8]. Since color moments encode both shape and color information they are a good feature to use under changing lighting conditions [4]. In addition, the color moments have the same units, which allows for comparison between them [5]. We use the first two color moments. The steps are:

Step 1: the image is separated into three color channels (R, G, and B).

Step 2: color moments are computed from each channel. The first color moment is the mean which is defined by

$$E_i = \sum_{j=1}^{N} \frac{1}{N} p_{i,j} \tag{1}$$

where N is the number of pixels in the image and $P_{i,j}$ is the value of the j-th pixel of the image at i-th color channel (3, 2, 1= i). The second color moment is the standard deviation which is defined by

$$\sigma_i = \sqrt{\left(\frac{1}{N}\sum_{j=1}^{N}\left(p_{i,j} - E_i\right)^2\right)} \tag{2}$$

where E_i is the mean value or first color moment for i-th color channel of the image (3, 2, 1= i).

Step 3: we get total 6 color moments for each image. Finally, we create the color Moments feature vector, $C_{cm} = (E_1, E_2, E_3, \sigma_1, \sigma_2, \sigma_3)$.

2) Color Feature Extraction using Color Autocorrelogram: An autocorrelogram can be stored as a table, indexed by color *i*, where *d*-th entry shows the probability of finding a pixel *i*, from the same pixel at distance *d*. Autocorrelogram shows the spatial correlation between identical colors only. Experiments show that correlogram and autocorrelogram both are computationally expensive. To find the color autocorrelogram feature, we have used the following steps [15]:

Step 1: Convert the RGB image to an indexed image using *minimum variance quantization* and *map* contains at most 64 colors i.e. quantize image into 64 colors (4×4×4) and maps each color in the original image to the closest color in the new map. No dithering is performed.

Step 2: Convert the indexed image and the corresponding colormap, *map*, to the true color image, RGB. The indexed image is an $m \times n$ array of integers. The colormap, *map*, is a three-column array of values in the range [0, 1]. The true color image output, RGB, is an $m \times n \times 3$ array.

Step 3: 4 distances are defined between neighbor pixels intensities. In our experiment, we choose distance d = 1, 3, 5, 7.

Step 4: 1×64 correlogram values (vector) are computed using neighbor pixels.

Step 5: reshapes the result vector of step 4 into a 4×4×4 matrix.

Step 6: construct final correlogram using distance values that are set in step 3 (the result is a 4×4×4 matrix).

Step 7: reform the result matrix of step 6 to a vector format. Finally, we get 1×64 feature vector, $C_{ca,}$ containing the color autocorrelogram.

3) Texture Feature Extraction using Wavelet Moments: The distribution of texture in an image can be interpreted as probability distributions which are characterized by a number of unique moments (e.g. Normal distributions are differentiated by their mean and variance). We use 2nd and 3rd moments. The steps to find the wavelet moments are as follows:

Step 1: convert the true color image RGB to grayscale image. Then the result is normalized to get value in the range [0, 1].

Step 2: perform single-level two-dimensional discrete Daubechies wavelet decomposition.

Step 3: compute the 2nd and 3rd wavelet moments. The second wavelet moment is the standard deviation which is defined by

$$\sigma = \sqrt{\left(\frac{1}{N}\sum_{j=1}^{N}\left(p_j - E\right)^2\right)} \tag{3}$$

The third wavelet is the Skewness which is a measure of the degree of asymmetry in the distribution and is defined by

$$S = \sqrt[3]{\left(\frac{1}{N}\sum_{j=1}^{N}\left(p_j - E\right)^3\right)} \tag{4}$$

In the above equations, E is the mean value or first wavelet moment.

Step 4: we get a total two wavelet moments from each image and finally create the wavelet moment feature vector, $T_{wm} = (\sigma, S)$.

4) Texture Feature Extraction using Wavelet Energy: Energy distribution over wavelet sub-bands is applied to find texture features [16]. To find the wavelet energy based texture feature, the following steps are followed:

Step 1: Perform the wavelet decomposition of the image by Daubechies wavelet, at level 2. Outputs are the decomposition vector, C and the corresponding bookkeeping matrix, S.

Step 2: Find E_a, which is the percentage of energy corresponding to the approximation, and vectors E_h, E_v, E_d, which contain the percentages of energy corresponding to the horizontal, vertical, and diagonal details respectively from (C, S).

Step 3: Finally compute the first two wavelet energy moments (Mean and Standard deviation) from Energy values of step 2, thus we get total eight wavelet Energy moments from each image and finally create the wavelet energy moments feature vector, $T_{we} = (E_1, E_2, E_3, E_4, \sigma_1, \sigma_2, \sigma_3, \sigma_4)$.

5) Texture Feature Extraction using Wavelet Entropy: Entropy is the expected value (average) of the information contained in each image. Wavelet packet entropy is a new approach with the ability to analyze transient features of non-stationary signals or images [17]. The steps are:

Step 1: separate the image channels,

Step 2: decompose each image channel into different sub-bands using Daubechies wavelet to obtain the approximation and the detail (horizontal, vertical, and diagonal) wavelet coefficient matrices.

Step 3: Again, the first approximation matrix is decomposed for finding second level approximation and detail (horizontal, vertical, and diagonal) wavelet coefficient matrices. Thus we get a total 24 coefficient matrices.

Step 4: Each coefficient matrix is used to calculate the quantitative values of log energy wavelet entropy to select the texture features of images. Generally speaking, it performs log energy entropy operation over the coefficients of wavelet transform and is defined by

$$W_i = \sum_j \sum_k \log \left(p[j,k] \right)^2 \tag{5}$$

with the convention $log(0) = 0$, $p[j, k]$ is the $[j, k]$th entry of the wavelet coefficient matrix and $i=1,\ldots\ldots,24$. Wavelet packet entropy is RST invariant. All feature are combined to create a vector, $T_{wp} = (W_1, W_2, W_3, \ldots, W_{24})$.

6) Shape Feature Extraction using Hu's Invariant Moments: An image moment is a certain particular weighted average (moment) of the image pixels' intensities or a function of such moments, usually chosen to have some attractive property or interpretation. To find the Image invariant moment based shape feature, we calculate the Hu's seven invariant moments [18] for each image. The steps of this feature extraction are as follows:

Step 1: convert the true color image RGB to grayscale image. Then the result is normalized to get value in the range [0, 1].

Step 2: Geometric raw moments are computed to get the image centroid.

Step 3: translation invariant central moments are computed using the centroid. Thus we get image orientation.

Step 4: Invariants with respect to both translation and scale can be constructed from central moments by dividing through a properly scaled *zero*-th central moment.

Step 5: Invariants with respect to translation, scale, and rotation are constructed. These are *Hu's moment invariants* $(I_1, I_2, \ldots, I_7)$. The first one is analogous to the moment of inertia around the image's centroid, where the pixels' intensities are analogous to physical density. The last one is skew invariant, which enables it to distinguish mirror images of otherwise identical images.

Step 6: In step 5 we get total seven features for each image and finally shape feature vector, $S_{im} = (I_1, I_2, \ldots, I_7)$ is created.

All the feature vectors of an image described above form a final feature vector and the final feature vectors of the training images are stored in the database.

7) Creating Final Feature Vector: We get the color moment based color feature vector, C_{cm}, color autocorrelogram based color feature vector, C_{ca}, wavelet moment based texture feature vector, T_{wm}, wavelet energy based texture feature vector, T_{we}, wavelet packet entropy based texture feature vector, T_{wp}, and moment invariants based shape feature vector, S_{im}. The final feature vector, u, is created by combining one or more of these feature vectors. The feature vector u is a column vector.

B. Similar Image Retrieval

Algorithm 2 is used for finding similar images from the database.

Algorithm 2: Algorithm for similar image retrieval

Step 1: Load stored feature values of all images of the dataset.

Step 2: Load resized query image

Step 3: Extract features from the query image and combines them as described in Step 2 to Step 5 of *Algorithm 1*. Thus the column feature vector, v, for query image is created and we get, $v = \left(C_{cm}^{/}, C_{ca}^{'}, T_{wm}^{/}, T_{we}^{'}, T_{wp}^{'}, S_{im}^{/}\right)$.

Step 4: Find the distance between the feature vector of the query image, v, and the feature vector of each image of the database, u, using similarity measures and do it for all training images.

Step 5: Select the images that are more similar i.e. in short distances from the query image and the corresponding images are retrieved as similar images of the query image.

The measures that are used to find the similarity [19-21] i.e. to find the distance of each training image from query image are shown in the following:

Minkowski distance	:	$d = \sqrt[\lambda]{\sum_{i=1}^{k} \lvert u_i - v_i \rvert^{\lambda}}$
Cityblock distance	:	$d = \sum_{i=1}^{k} \lvert u_i - v_i \rvert$
Relative Cityblock or L_1 distance	:	$d = \sum_{i=1}^{k} \frac{\lvert u_i - v_i \rvert}{1 + u_i + v_i}$
Euclidean or L_2 distance	:	$d = \sqrt{\sum_{i=1}^{k} (u_i - v_i)^2}$
Standardized L_2 distance	:	$d = \sqrt{\sum_{i=1}^{k} (u_i - v_i)^2 V^{-1}}$
Normalized L_2 distance	:	$d = \sqrt{\sum_{i=1}^{k} (u_i - v_i)^2 S^{-1}}$
Mahalanobis distance	:	$d = \sqrt{\sum_{i=1}^{k} (u_i - v_i)^2 C^{-1}}$
Chebyshev distance	:	$d = \max_{i}^{k} \lvert u_i - v_i \rvert$
Cosine Similarity	:	$d = \frac{\sum_{i=1}^{k} u_i v_i}{\sqrt{\sum_{i=1}^{k} u_i^2 \sum_{i=1}^{k} v_i^2}}$
Correlation (Pearson's)	:	$d = \frac{\sum_{i=1}^{k} \left(u_i - \bar{u}\right)\left(v_i - \bar{v}\right)}{\sqrt{\sum_{i=1}^{k} \left(u_i - \bar{u}\right)^2} \sqrt{\sum_{i=1}^{k} \left(v_i - \bar{v}\right)^2}}$
Spearman's rank correlation	:	$d = \frac{\sum_{i=1}^{k} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i=1}^{k} (r_{ui} - \bar{r}_u)^2} \sqrt{\sum_{i=1}^{k} (r_{vi} - \bar{r}_v)^2}}$
Relative deviation	:	$d = \frac{\sqrt{\sum_{i=1}^{k} (u_i - v_i)^2}}{{}^{1}\!/_{2}\left(\sqrt{\sum_{i=1}^{k} u_i^2} + \sqrt{\sum_{i=1}^{k} v_i^2}\right)}$

In the above equations, *d* is the corresponding distance between two image feature vectors *u* and v, *k* is the total number of image features, *λ* is the order of the Minkowski distance, *V* is the variance, a *n*×*n* diagonal matrix whose *j*th diagonal element is S(*j*)^2, where *S* is the vector of standard deviations, *S* is the standard deviation of the query image in Normalized Euclidean Distance, *C* is the covariance matrix of each training image and *C* is positive definite, $\bar{u} = \frac{1}{n}\sum_{i} u_i$ and $\bar{v} = \frac{1}{n}\sum_{i} v_i$, r_{ui} and r_{vi} are the feature-wise rank vectors of *u* and v respectively and $\bar{r}_u = \frac{1}{k}\sum_{i=1}^{k} r_{ui} = \frac{(n+1)}{2}$ and $\bar{r}_v = \frac{1}{k}\sum_{i=1}^{k} r_{vi} = \frac{(n+1)}{2}$.

3. Experiment Simulation and Result Analysis

For experiments, we have used ten categories of image, hundred images in each category, i.e. total 1000 images of the Wang database [14]. All images are scaled into the same size (384×256). As an example of retrieved result, Fig. 1 shows the retrieved images for using *color moment, wavelet energy, and invariant moment* features together for *correlation* similarity measure.

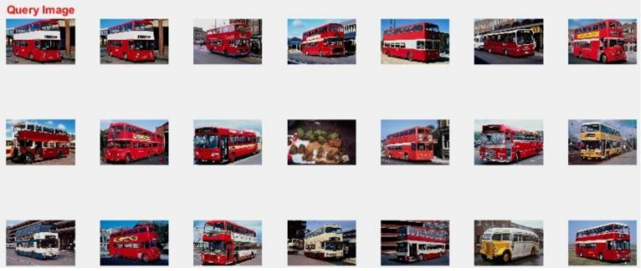
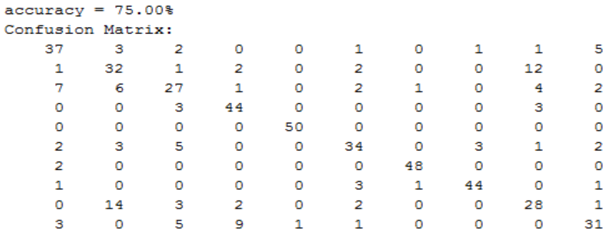


Fig.1 Retrieved result using correlation for a sample Bus image

Accuracy is a vital parameter for evaluation as it is a direct measurement of quality and user satisfaction of the image retrieval process [22]. To evaluate, we have used the holdout method i.e. we divide each image category into two sets so that the training set contains 50 images and test set contains another 50 of each category. That means, total 500 images of the database are used for training and the rest 500 images are used for testing. Accuracy is calculated from the confusion matrix. Within the diagonal of the confusion matrix, the numbers represent the number of samples that are correctly retrieved. Numbers not found within the diagonals are errors [6]. For example, Fig. 2 shows the confusion matrix and corresponding accuracy for *color moment* feature with *Chebyshev* distance (from our implementation using Matlab).



```
accuracy = 75.00%
Confusion Matrix:
    37     3     2     0     0     1     0     1     1     5
     1    32     1     2     0     2     0     0    12     0
     7     6    27     1     0     2     1     0     4     2
     0     0     3    44     0     0     0     0     3     0
     0     0     0     0    50     0     0     0     0     0
     2     3     5     0     0    34     0     3     1     2
     2     0     0     0     0     0    48     0     0     0
     1     0     0     0     0     3     1    44     0     1
     0    14     3     2     0     2     0     0    28     1
     3     0     5     9     1     1     0     0     0    31
```

Fig.2 Sample confusion matrix with accuracy value

A. *Performance Analysis using Various Features and Similarity Measures*

To evaluate the performance, accuracy of each feature extraction method for each similarity measure is calculated. Table 1 shows the short form for various feature extraction and similarity measurement methods, used for simplicity in the later parts of this paper.

Table.1 Used Short form for (a) Feature Extraction and (b) Similarity Measurement methods

Feature Extraction Method	**Short form**
Color moments	CM
Color autocorrelogram	CA
Wavelet moments	WM
Wavelet energy	WE
Wavelet packet entropy	WP
Hu's invariant moment	IM

(a)

Similarity measure	**Short form**	**Similarity measure**	**Short form**
L_1	L_1	Chebyshev	Cheby
L_2	L_2	Cosine	Cos
Standardized L_2	S.L_2	Correlation	Corr
Mahalanobis	Mahal	Spearman's correlation	Spear
City block	City	Normalized L_2	N.L_2
Minkowski	Mink	Relative deviation	R.D.

(b)

1) Retrieval using Color Feature: Table 2 shows the accuracy using various color features for various similarity measures. Color moment shows better result than color autocorrelogram for all types of similarity measures. For color moment, most of the measures show better accuracy but Spearman's rank correlation is not a good choice here whereas Mahalanobis is the best choice. For color autocorrelogram, no measure shows comparative accuracy but L_1 shows little better result than other measures whereas Standardized L_2 and Mahalanobis show the worst results.

Table.2 Accuracy (in %) for various color features

	CM	CA
L_1	84	30
L_2	80	25
$S.L_2$	81	13
Mahal	88	13
City	85	28
Mink	80	25
Cheby	75	15
Cos	84	20
Corr	85	24
Spear	61	15
$N.L_2$	80	23
R.D.	79	15

2) Retrieval using Texture Feature: Table 3 shows the accuracy using various texture features for various similarity measures. Wavelet packet entropy shows the best result than Wavelet Moment and wavelet energy methods for all measures except L_1. Again, it shows better accuracy than the best color feature (color moment), though color moment is better than other texture features. L_1 distance does not provide better result for wavelet entropy and also for wavelet moment, though it shows better result than other measures when wavelet energy is used as a feature. Standardized L_2 distance is a better choice for wavelet moment feature, and correlation measure is for wavelet entropy. In case of wavelet energy, Mahalanobis distance can retrieve only the query image itself as a similar image, so it is not a good choice in this case at all.

Table.3 Accuracy (in %) for various texture features

	WM	WE	WP
L_1	6	64	4
L_2	48	51	81
$S.L_2$	65	23	86
Mahal	9	1	70
City	61	54	85
Mink	48	51	81
Cheby	30	45	80
Cos	51	54	86
Corr	54	59	88
Spear	48	15	80
$N.L_2$	48	51	81
R.D.	40	51	80

3) Retrieval using Shape Feature: Table 4 shows the accuracy using Hu's moment based shape feature for various similarity measures. Hu's Invariant moments shows the better result for some measures than some color and texture features. But in most cases, shape feature does not provide a better result than color moment and any of texture features. In this case, Spearman's rank correlation is not a good choice whereas L_2, Minkowski, Normalized L_2, or Relative deviation is a better choice.

Table.4 Accuracy (in %) for shape feature

	IM
L_1	55
L_2	63
$S.L_2$	58
Mahal	50
City	59
Mink	63
Cheby	59
Cos	43
Corr	44
Spear	25
$N.L_2$	63
R.D.	63

4) Retrieval using Color and Texture Features together: Table 5 shows the accuracy using various color features with various texture features for various similarity measures. Color moment with Wavelet packet entropy shows the best accuracy than all other methods for the standardized L_2 measure and also show better result than using only color moment or only wavelet entropy feature. Though color moment with other texture feature improves the retrieving result for some measures than using only the corresponding texture feature but color moment alone can provide better result than these ones, in most of the cases. Using multiple texture features with color moment provides better result than using only wavelet moment or wavelet energy but provides worse result than using only wavelet entropy. Using three texture feature together do not provide better accuracy than using two texture features together.

Color autocorrelogram with wavelet entropy shows better result for cosine and correlation measure. Though color autocorrelogram with all texture features improves the retrieving result for many measures than using only the color autocorrelogram but all texture features separately themselves provide better result than this. That is, color autocorrelogram is not a good choice at all. Here using all three texture features together do not provide better result than using two texture features together or even using only wavelet entropy feature. Again, using both wavelet energy and wavelet entropy with color autocorrelogram does not show better result than using only wavelet entropy with color autocorrelogram. In other cases, using multiple texture features shows better result than using one of the corresponding features.

When wavelet energy is used as a feature in any combination, Mahalanobis is not a good choice. But when wavelet moment or wavelet entropy is used in any combination, L_1 is not a good measure at all. Overall cosine or correlation show better result and on average, city block shows some good result.

Table.5 Accuracy (in %) for various color features with various texture features

	CM+ WM	CM+ WE	CM+ WP	CM+ WM+ WE	CM+ WM+ WP	CM+ WE+ WP	CM+ WM+ WE+ WP
L_1	5	85	4	5	3	4	4
L_2	81	80	81	83	81	81	81
S.L_2	66	23	90	23	78	23	23
Mahal	6	1	74	1	6	1	1
City	84	85	85	85	85	85	85
Mink	81	80	81	81	81	81	81
Cheby	75	75	80	75	80	80	80
Cos	89	86	86	89	86	86	86
Corr	89	86	88	89	86	86	86
Spear	46	84	84	46	55	85	58
N.L_2	81	80	81	81	81	81	81
R.D.	80	79	80	81	80	80	80

(a)

	CA+ WM	CA+ WE	CA+ WP	CA+ WM+ WE	CA+ WM+ WP	CA+ WE+ WP	CA+ WM+ WE+ WP
L_1	5	48	5	6	5	4	4
L_2	44	20	81	46	81	81	81
S.L_2	58	23	50	21	73	23	23
Mahal	8	1	39	1	8	1	1
City	58	38	85	60	85	85	85
Mink	44	20	81	45	81	81	81
Cheby	26	19	79	30	80	80	80
Cos	50	24	86	48	86	86	86
Corr	51	25	86	49	86	86	86
Spear	43	18	26	43	46	33	49
N.L_2	43	20	81	46	81	81	81
R.D.	36	20	80	48	80	80	80

(b)

5) Retrieval using Color and Shape Features together: Table 6 shows the accuracy using various color features with shape feature for various similarity measures. Color moment with invariant moment shows better result than using only color moment or Invariant moment. Though color moment with Invariant moment gives good result for some measures but overall it is not a better combination than color moment with wavelet entropy.

Table.6 Accuracy (in %) for various color features with shape feature

	CM + IM	CA + IM
L_1	88	34
L_2	81	25
S.L_2	83	19
Mahal	74	15
City	86	34
Mink	80	25

Cheby	76	15
Cos	84	20
Corr	85	21
Spear	75	16
N.L$_2$	80	25
R.D.	79	19

Color autocorrelogram with Invariant moment reduces the result accuracy i.e. it is worse than using only color autocorrelogram or Invariant moment. That is, color autocorrelogram itself only or with other features is not a good feature at all.

When we use shape feature with the color feature, L$_1$ always provides the best result and Mahalanobis always provides worse result. Chebychev measure is also a bad choice. City block is a good choice here.

6) Retrieval using Texture and Shape Features together: Table 7 shows the accuracy using various texture features with shape feature for various similarity measures.

Wavelet entropy with invariant moment shows better result than other combinations here and also gets better result than using only Hu’s invariant moment but gives worse result than using only wavelet entropy feature. Again, wavelet moment with invariant moment does not provide better result than using any of them singly. But if we use wavelet energy with invariant moment then we get better accuracy than using only wavelet energy but gives worse result than using only invariant moment feature. Again, using two texture features do not show better result than using only wavelet entropy with the invariant moment. For few measures, using both wavelet moment and wavelet energy features shows better result than using only one of them. Using all three texture features together do not provide better result than using two texture features together or even using only wavelet entropy feature. Invariant moment with Wavelet Moment and wavelet energy shows worse results than other combinations here.

Mahalanobis and L$_1$ (except for wavelet energy and invariant moment feature combination) are worse measures and cosine and correlation are better measures here. City block also shows comparative accuracy.

Table.7 Accuracy (in %) for various texture features with shape feature

	WM+ IM	WE+ IM	WP+ IM	WM+ WE+ IM	WM+ WP+ IM	WE+ WP+ IM	WM+ WE+ WP+ IM
L$_1$	6	75	4	5	4	4	4
L$_2$	43	59	81	46	81	81	81
S.L$_2$	64	23	86	21	80	23	23
Mahal	8	1	65	1	6	1	1
City	58	66	85	61	85	85	85
Mink	45	59	81	45	81	81	81
Cheby	29	55	80	29	80	80	80
Cos	49	68	86	46	86	86	86
Corr	51	68	86	44	86	86	86
Spear	44	61	71	45	53	73	54
N.L$_2$	43	59	81	45	81	81	81
R.D.	36	59	80	43	80	80	80

7) Retrieval using Color, Texture, and Shape Features together: Table 8 shows the accuracy using various color features with various texture features and shape feature for various similarity measures. Color moment and invariant moment with wavelet moment show worse accuracy and that with wavelet entropy shows better accuracy than all other methods that use color moment feature. Again, using shape feature with color and texture feature do not show better result than using only color and texture feature. In very few cases, using shape feature with the color moment and wavelet moment shows better result than using only the later. Again, using two texture features do not show better result than using only wavelet entropy with the color moment and invariant moment. For very few measures, using both wavelet moment and wavelet energy shows better result than using only one of them.

Color autocorrelogram and invariant moment with wavelet energy show worse accuracy and that with wavelet entropy shows better results than all other methods that use the color autocorrelogram feature. Again, using shape feature with color and texture feature do not show better result than using only color and texture feature together. For few measures, using shape feature with color autocorrelogram and wavelet moment shows better result than using the later together. Using all three texture features together does not show better accuracy.

Table.8 Accuracy (in %) for various color features with various texture and shape features

	CM+ WM+ IM	CM+ WE+ IM	CM+ WP+ IM	CM+ WM+ WE+ IM	CM+ WM+ WP+ IM	CM+ WE+ WP+ IM	CM+ WM+ WE+ WP+ IM
L_1	5	89	4	5	4	6	4
L_2	83	80	81	81	81	81	81
S.L_2	68	23	90	23	80	23	23
Mahal	6	1	65	1	8	1	1
City	84	85	85	85	85	85	85
Mink	81	80	81	81	81	81	81
Cheby	75	75	80	75	80	80	80
Cos	89	86	86	89	86	86	86
Corr	88	86	86	89	86	86	86
Spear	18	83	81	45	55	80	58
N.L_2	81	80	81	81	81	81	81
R.D.	80	79	80	81	80	80	80

(a)

	CA+ WM+ IM	CA+ WE+ IM	CA+ WP+ IM	CA+ WM+ WE+ IM	CA+ WM+ WP+ IM	CA+ WE+ WP+ IM	CA+ WM+ WE+ WP+ IM
L_1	5	58	5	5	5	5	4
L_2	44	26	81	46	81	81	81
S.L_2	59	23	50	23	74	23	23
Mahal	6	1	31	1	6	1	1
City	59	48	85	61	85	85	85
Mink	44	26	81	46	81	81	81
Cheby	29	20	80	31	80	80	80
Cos	50	29	86	49	86	86	86
Corr	53	28	86	50	86	86	86
Spear	44	16	25	44	48	30	50
N.L_2	43	26	81	48	81	81	81
R.D.	39	26	80	48	80	80	80

(b)

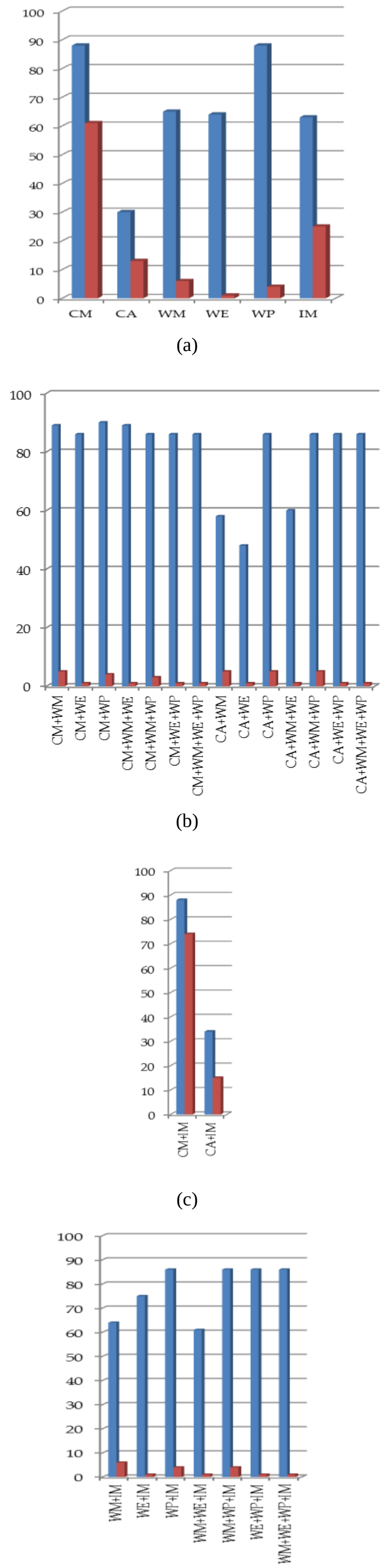
As a similarity measure, Mahalanobis and L_1 (except for wavelet energy and invariant moment feature combination with any of color feature) are worse measures and cosine and correlation are better measures here. City block also shows comparative accuracy. Standardized L_2 measure with the color moment, wavelet entropy, and invariant moment show the best accuracy but we see from table 5 that color moment with wavelet entropy for this measure shows the same best accuracy without using invariant moment together.

B. Comparison with Existing works

Most of the results from this study are not exactly comparable with the analyses provided in the original studies due to differences in the setup of experimental design. To the best of our knowledge, no previous work exists that shows complete analysis applying all methods used in this study. However, the reported results confirm that color moments and wavelet packet entropy as applied here perform equally as well, or even better, compare to existing works on the same dataset.

C. Discussion

If we consider the better accuracy using as less as possible feature and better accuracy for maximum measures for that feature then we get four better feature combinations: (i) using only color moment (except for Spearman's correlation measure) (ii) using only wavelet entropy (except for L_1 measure) (iii) color moment with wavelet packet entropy (best for standardized L_2 measure), and (iv) color moment with Hu's invariant moment (except for Mahalanobis measure). Again, if we consider the worst case results, we see it happens when color autocorrelogram or wavelet moment (in some cases wavelet energy) is used as a feature. That means as a feature, color moment or wavelet entropy shows better results. Adding invariant moment based shape feature provides very less contribution on similar image retrieving compare to color and texture feature. Fig. 3 shows the highest and lowest accuracies for using various features.

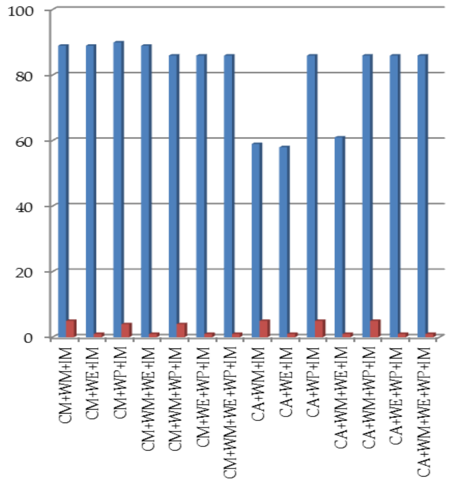

100
90
80
70
60
50
40
30
20
10
0
CM
CA
WM
WE
WP
IM
100
80
60
40
20
0
CM+WM
CM+WE
CM+WP
CM+WM+WE
CM+WM+WP
CM+WE+WP
CM+WM+WE+WP
CA+WM
CA+WE
CA+WP
CA+WM+WE
CA+WM+WP
CA+WE+WP
CA+WM+WE+WP
CM+IM
CA+IM
WM+IM
WE+IM
WP+IM
WM+WE+IM
WM+WP+IM
WE+WP+IM
WM+WE+WP+IM

(a)

(b)

(c)

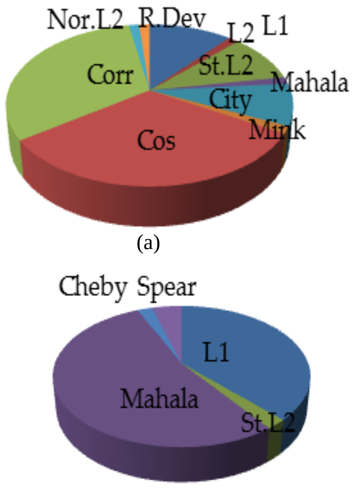
(d)



(e)

Fig.3. Feature extraction methods with their highest accuracy (blue colored) and lowest accuracy (red colored) for (a) Color, texture, shape feature separately (b) Color and texture features together (c) Color and shape feature together (d) Texture and shape features together (e) Color, texture, and shape features together

In the case of similarity measures, cosine or correlation shows the best accuracy in maximum cases. It shows some narrow results in the only case of using wavelet energy as a feature. L_1 distance is a good choice in the cases when wavelet energy is used but not a good choice in other cases. Mahalanobis is a worse measure in maximum cases of similar image retrieving. In case of using wavelet energy feature, Mahalanobis distance retrieves no other similar images except itself because in this case, the covariance matrix is not positive definite. But in this case, city block and relative city block (L_1) show higher accuracy because it calculates the absolute value difference. The measures such as L_2 as well as Spearman’s rank correlation (which is a square of Euclidean distance between two rank vectors), Standardized L_2, Normalized L_2, Minkowski, Relative deviation, Mahalanobis; all of these methods squared the distance value before summation, this places great emphasis on those features for which the dissimilarity is large. Chebyshev, City block or L_1 is a winner in this case. But Chebyshev distance does not treat each feature as equally important, which affects the retrieval result. Cosine and correlation similarity measures find similarity between two vectors with the origin as the reference, indicates how closely the two vectors point in the same direction, and because of these nature of RST invariant characteristics, cosine and correlation measures retrieve more similar images, hence shows higher accuracy. But Standardized L_2 shows the best accuracy above all when color moments and wavelet packet entropy is used as a feature. Fig. 4 shows the performance chart for various similarity measures.



(a)

(b)

Fig.4. Accuracy for various similarity measures (a) Best accuracy (b) Worst accuracy

Fig. 5 shows the average accuracy (in percentage), for various similarity measures. We found that though city block rarely shows the best accuracy but on average, city block shows better accuracy.

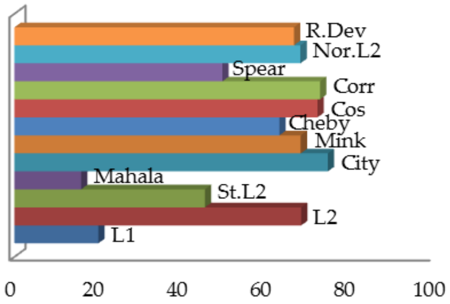


Fig. 5. Average accuracy shown by various similarity measures

Though, there have some methods [2-7], [10-13] that use some of our used feature extraction and/or similarity measurement methods, but a comprehensive evaluation applying all these methods does not exist. It is also currently poorly understood, what are the best combinations of feature extraction and similarity measurement methods in CBIR. This is the first such system to be informed by a rigorous comparative evaluation of the features and similarity measures, used here.

4. Conclusion

In this work, we conducted a thorough evaluation to identify the most accurate (and RST invariant) feature extraction methods for various similarity measures. The results of this work also highlight the most accurate combination of features and similarity measures. Experimental results show that the best accuracy is found for color moment and wavelet packet entropy feature for the standardized L_2 measure. But in maximum cases, cosine and correlation show the best result and on average, city block shows better results on different orientation of query and database image. The present results are relevant to the extent that the dataset employed is representative of the characteristics of similar image retrieval dataset in common use. Of course, analysts when using the present benchmark comparison results to inform their analyses should consider the degree of similarity of their datasets to the datasets in the study.

Acknowledgment

This research is funded by Information and Communication Technology (ICT) division, Ministry of Post, Telecommunication and Information Technology, Peoples Republic of Bangladesh government through grant ICT fellowship.

References

[1] Yu, C., Brandenburg, T.: ‘Multimedia Database Applications: Issues and Concerns for Classroom Teaching’, IJMA, 2011, 3, (1), pp. 1-9.

[2] Sclaroff, S., Taycher, L., Cascia, M.L.: ‘Imagerover: A content-based image browser for the World Wide Web’. Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries, San Juan, Puerto Rico, June 1997, pp. 2–9.

[3] Prathiba, T., Darathi, G.S.: ‘An efficient content based image retrieval using local pattern’, IJAREEIE, 2013, 2, (10), pp. 5040-5046.

[4] Duanmu, X.: ‘Image retrieval using color moment invariant. Proc. Int. Con. Information Technology’, Las Vegas, USA, April 2010, pp. 200-203.

[5] Chitkara, V.; Nascimento, M.A.; Mastaller, C.: ‘Content-based image retrieval using binary signatures’, Department of Computing Science, University of Alberta, Canada, September 2000.

[6] Singla, A., Garg, M.: ‘CBIR approach based on combined HSV, auto correlogram, color moments and Gabor wavelet’, IJECS, 2014, 3, (10), pp. 9007-9012.

[7] Han, J., Ma, K.: ‘Fuzzy Color Histogram and Its Use in Color Image Retrieval’, IEEE Trans. Image Process., 2002, 11, (8), pp. 944-952.

[8] Stricker, M.A., Orengo, M.: ‘Similarity of color images’, Proc. SPIE Con. the International Society for Optical Engineering, San Jose, February 1995, pp. 381-392.

[9] Tamura, H., Mori, S., Yamawaki, T.: ‘Texture Features Corresponding to Visual Perception’, IEEE Trans. Syst., Man, Cybern., 1978, 8, (6), pp. 460-473.

[10] Sakhare, S.V., Nasre, V.G.: ‘Design of feature extraction in Content Based Image Retrieval (CBIR) using color and texture’, IJCSIS, 2011, 1, (2), pp. 57-61.

[11] Castelli, V., Bergman, L.D.: ‘Image Databases: Search and Retrieval of Digital Imagery’, (John Wiley & Sons, 1st Edn., 2002).

[12] *Haralick, R.M., Shanmugam, K., Dinstein, I.:* ‘Textural Features for Image Classification’, *IEEE Trans. Syst., Man, Cybern., 1973, 3, (6), pp. 610–621.*

[13] Latecki, L.J., Lak¨amper, R., Wolter, D.: ‘Shape similarity and visual parts’, Proc. Int. Con. Discrete Geometry for Computer Imagery, Naples, Italy, November 2003, pp. 34–51.

[14] ‘Wang Groups: Modeling Objects, Concepts, Aesthetics, and Emotions in Big Visual Data’, http://wang.ist.psu.edu/docs/home.shtml, Accessed 10 July 2018.

[15] Tungkasthan, A., Intarasema, S., Premchaiswadi, W.: ‘Spatial Color Indexing using ACC Algorithm’, Proc. Int. Con. ICT and Knowledge Engineering, Bangkok, Thailand, December 2009, pp. 113-117.

[16] Dua, S., Acharya, U.R., Chowriappa, P., et al.: ‘Wavelet-based energy features for glaucomatous image classification’, IEEE Trans. Inf. Technol. Biomed., 2012, 16, (1), pp. 80-87.

[17] Langley, P.: ‘Wavelet entropy as a measure of Ventricular beat suppression from the Electrocardiogram in Atrial Fibrillation’, Entropy, 2015, 17, (9), pp. 6397–6411.

[18] Hu, M.K.: ‘Visual Pattern Recognition by Moment Invariants’, IRE Trans. Info. Theory, 1962, 8, (2), pp. 179–187.

[19] *Mahalanobis P.C.: ‘*On the generalized distance in statistics’, *Proc. National Institute of Sciences of India, Calcutta, India, April 1936, pp. 49–55,*

[20] *Cantrell, C.D.: ‘Modern Mathematical Methods for Physicists and Engineers’, (Cambridge University Press, 1st Edn. 2006).*

[21] ‘Distance metrics’, http://numerics.mathdotnet.com, Accessed 19 July 2018.

[22] Chandha, A., Mallik, S., Johar, R.: ‘Comparative study and optimization of feature extraction techniques for Content Based Image Retrieval’, IJCA, 2012, 52, (20), pp. 35-42.

Authors’ Profiles

S.M. Mohidul Islam is an Associate Professor at the Computer Science and Engineering Discipline, Khulna University, Bangladesh. He received his B.Sc. Engg. and M.Sc. Engg. degree from Khulna University, Bangladesh. His research interests include Machine learning, Data Mining, Pattern Recognition, and Digital Image Processing.

Rameswar Debnath is a Professor at the Computer Science and Engineering Discipline, Khulna University, Bangladesh. He received his Ph.D. from the University of Electro-Communications (UEC), Tokyo, Japan in 2005. His research interests are statistical machine learning (in particular supervised learning; support vector machines and kernel methods), and its applications to pattern recognition, image processing, bioinformatics and natural language analysis.

How to cite this paper: S.M. Mohidul Islam, Rameswar Debnath, " A Comparative Evaluation of Feature Extraction and Similarity Measurement Methods for Content-based Image Retrieval", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.12, No.6, pp. 19-32, 2020.DOI: 10.5815/ijigsp.2020.06.03