*International Journal of Computer Applications (0975 – 8887)*
*Volume 105 – No. 11, November 2014*

# Empirical Evaluation of Dissimilarity Measures for Content-based Image Retrieval

Marziyeh Hojati
M.S.C,
Engineering Department, Science and Art University
Daneshjoo Blvd., Yazd, Iran;

Mehdi Rezaeian
Assistant Professor, Electrical and Computer Engineering Department, Yazd University
Daneshgah Blvd., Safa-ieh, Yazd, Iran;

## ABSTARCT
In this paper, the performance of various distances in image retrieval and image classification is evaluated based on color and texture features. The evaluation has been done in two classification: k-nearest neighbors and support vector machine(SVM). Given SVM classification is a learning system and any learning system is susceptible to error, therefore in this study a method is proposed for the user interaction. In this method if an error occurs in the first implementation of SVM classification or an image is displayed incorrectly, the next executions show similar images or to inform the user that the image is not in the database.The results of the experiment will be presented and investigated based on color histogram, color moment, color correlogram, gabor features, local binary pattern and wavelet transform in a database.

## Keywords
content-based image retrieval, color feature, texture feature, support vector machine, k-nearest neighbors.

## 1. INTRODUCTION
In the last three decades, the CBIR has become a very important research area in image processing and multimedia databases. The aim of researchers in the field of CBIR is to find features and distinctions in pictures which can be used to identify the similarities in images and to properly guide the image query. In a CBIR system, the user selects his or her desired image known as the query image and the system identifies and retrieves all the similar images from the database. To determine the similarity between two images, the distance between their feature vectors is calculated. Using different criterions in determining the similarity of two images will have different results. Therefore, a measure that can provide the distance between two images in regard to their different high-level characteristics or their semantic features is considered to be a valuable criterion in image retrieval[1], [2].

This paperis organized as follows:

The second section deals with the description of features and some methods for the extraction of color and texture feature are suggested. In the third section, the similarity measures and the similarity determination distances are introduced. The fourth section describes the performance evaluation of the similarity criterions. The results of the evaluation and interpretation as well as the proposed method are given in the fifth section. The final section is a summary of the results presented.

## 2. FEATURE DESCRIPTION
In this study, two color and texture features are used. Color histogram, color moment, color correlogram are used as color features. To extract texture features, Gabor filters, wavelet transform and LBP operator are applied. Since the shape property used to retrieve the images is limited to special applications, there is not much emphasis on them. In the following, these features will be described.

### 2.1.Color Histogram
Color is one of the most common and determining visual features in image retrieval; Because of it is persistent in case of changes in size, direction, prospects and distortion of an image. This feature has been used in many common image retrieval systems.

RGB color space is widely used due to its simplicity in use and understanding, as well as hardware support. However, it has some shortcomings in technical issues related to image processing systems. For instance, the correlation between the three components means that if there is a change in brightness factor of an image, each of the three RGB components will change. Therefore, this space is not often used in image retrieval systems. HSV color space is aperceptual space, so is suitable for use in retrieval systems [3]. Constituent dimensions of this space define colors based on wavelength (H), the degree of saturation(S) and the degree of brightness (V). In this study, the HSV color spaceis used. The histogram represents the cumulative distribution of points in the three color channels, in a way that each histogram component is equal to the total number of pixels belonging to a color or a color set. Due to its low computational load, this method is widely used in image retrieval systems. Furthermore, it is persistent against image resizing, changes in the angle, prospects and image distortion [3], [4].

### 2.2.Color Moment
This method is one of the most successful methods of extracting color in retrieval systems. First moment,second and third, respectively, namely mean, variance and the degree of color bias provide the color distribution of the image efficiently and optimally. Using the third moment of the color, along with the first two moments, increases the overall efficiency of the retrieval, but this moment is itself sensitive to the viewing angle and may be a reason for reduced performance. This characteristic principally has a limited power ofseparation, so it can be used to filter out the images in the early stages ofthe query [2].

## 2.3. Color Correlogram

This feature not only represents the color distribution of pixels, but also the spatial correlation of pairs of pixels. Color Correlogram vector is a which is indexed by color pairs. This vector provides better efficiency in retrieval system as compared to the histogram and CCV. However, due to the numerous dimentions, it has high a computational load [4].

## 2.4.Gabor Filter

Another important characteristic of an image is texture which, from a visual perspective, is a uniform pattern created from the presence of more than one color or grayscale value. Using the Gabor filter is one of the most common filter-based methods to extract texture. Gabor filter can be used to model the reaction of the human visual system. This filter works in both space and frequency domain. In the spatial domain, the cores of the Gabor filter is obtained by a Gaussian function multiplied by a directional sinusoidal function. As a result, this filter produces a favorable response in the parts of the image which locally have a certain directional and spatial frequency [2]. To apply this filter, it is better to use the HSV color space.

## 2.5. Wavelet Transform

The underlying reason of the application of wavelet analysis in systems are things such as media concisely, orthogonality, symmetry, zero moments, etc. However, not all these characteristics are completely satisfied in classical wavelets. Therefore, multi-precision wavelets and hybrid wavelet resulting from the expansion of the classical wavelet are introduced. The idea behind multi-precision wavelets is a more general mode of classical wavelets. That is, instead of using aparticular scale function, several scales functions called multi-precision scaling functions are used. Wavelet transform includes HatMexican, Morlet, CoifletHaar and Daubechies, among which Haar filter is mostly used in retrieval systems due to its ease [5].

## 2.6. LBP Operator

Local binary pattern first was first proposed by Ojala et al [6]. The experiments carried out showed that LBP is a strong function in showing the texture. Therefore, it is widely used in the analysis of local texture. The way in which LBP works is that a 3×3 square mask is considered such that the pixel being calculated is the central pixel of this pattern. The desired pixel is compared to its eight neighbors. If each of the eight neighbors of the central pixel was larger than the central pixel, it will be replaced by one and if it is smaller, it will be replaced by zero. Finally,the value of the centeral pixel will be replaced by the weighted binary summation of neighboring pixels. Thus, by moving the 3×3 mask in all the image, the desired texture is achieved.

# 3. SIMILARITY MEASURES

It is assumed that for all or some of the features used in the image retrieval system, a total of N close images to the query image are obtained. To obtain the output images,a score should be assigned to each image based on its similarity to the query image and finally the best K images are selected displayed to the user as the output. In order to do this, similarity criterion and weighting factors are used. The similarity criterion is a non-negative ascending function that defines the degree of similarity of feature vectors, and the weighting factor determines the degree of importance of these features in determining the final images. The way in which the similarity criterion is chosen is important in increasing the accuracy of the output results.

Various distances between the vector $x_s$ and $x_t$ are defined as follows:

**Minkowski distance [7]:** is the generalized metric distance:

$$D_{st} = \left(\sum_{i=1}^{n} |x_s - x_t|\rho\right)^{1/\rho} \qquad (1)$$

**Euclidean distance [8]:** Defined as the distance between two points give by the Pythagorean Theorem. Special case of the Minkowski metric where p=2.

$$D_{st} = \sqrt{(x_s - x_t)(x_s - x_t)'} \qquad (2)$$

**Standardized Euclidean distance [8]:** Defined as the Euclidean distance calculated on standardized data, in this case standardized by the standard deviations.

$$D_{st} = \sqrt{(x_s - x_t)V^{-1}(x_s - x_t)'} \qquad (3)$$

Where V is the n-by-n diagonal matrix whose $j^{th}$ diagonal element is $S(j)^2$, where S is the vector of standard deviations.

**Mahalanobis distance [8]:** Defined as the Euclidean distance normalized based on a ovariance matrix to make the distance metric scale-invariant.

$$D_{st} = \sqrt{(x_s - x_t)C^{-1}(x_s - x_t)'} \qquad (4)$$

Where C is the covariance matrix

**City block distance [8]:** Also known as Manhattan distance, it represents distance between points in a grid by examining the absolute differences between coordinates of a pair of objects. Special case of the Minkowski metric where p=1.

$$D_{st} = \sum_{j=1}^{n} |x_{sj} - x_{tj}| \qquad (5)$$

**Chebychev distance [8]**: Measures distance assuming only the most significant dimension is relevant. Special case of the Minkowski metric where p = ∞

$$D_{st} = \max_j\{|x_{sj} - x_{tj}|\} \qquad (6)$$

**Cosine distance [9]:**Measures the dissimilarity between two vectors by finding the cosine of the angle between them.

$$D_{st} = 1 - \frac{x_s x_t{}'}{\sqrt{(x_s x_s')(x_t x_t')}} \qquad (7)$$

**Correlation distance [9]:** Measures the dissimilarity of the sample correlation between points as sequences of values.

$$D_{st} = 1 - \frac{(x_s - \bar{x}_s)(x_t - \bar{x}_t)'}{\sqrt{(x_s - \bar{x}_s)(x_s - \bar{x}_s)'}\sqrt{(x_t - \bar{x}_t)(x_t - \bar{x}_t)'}} \qquad (8)$$

Where $\bar{x}_s = \frac{1}{n}\sum_{j=1}^{n} x_{sj}$ and $\bar{x}_t = \frac{1}{n}\sum_{j=1}^{n} x_{tj}$

**Spearman distance [10]:** Measures the dissimilarity of the sample's Spearman rank correlation between observations as sequences of values.

$$D_{st} = 1 - \frac{(r_s - \bar{r}_s)(r_t - \bar{r}_t)'}{\sqrt{(r_s - \bar{r}_s)(r_s - \bar{r}_s)'}\sqrt{(r_t - \bar{r}_t)(r_t - \bar{r}_t)'}} \qquad (9)$$

Where $r_{sj}$ is the rank of $x_{sj}$ taken over $x_{1j}, x_{2j}, \dots x_{mj}$, $r_s$ and $r_t$ are the coordinate-wise rank vectors of $x_s$ and $x_t$, i.e., $r_s = (r_{s1}, r_{s2}, \dots r_{sn})$ and

$$\bar{r}_s = \frac{1}{n}\sum_{j=1}^{n} r_{sj} = \frac{(n+1)}{2}, \bar{r}_t = \frac{1}{n}\sum_{j=1}^{n} r_{tj} = \frac{(n+1)}{2}$$

## 4. PERFORMANCE EVALUATION OF SIMILARITY CRITERIA

To assess the performance of the defined distances, a database containing 2000 images were used and it is assumed that the semantic features are clear. This database consists of 20 semantic classes each containing 100 images. All images in each classhave the same semantic features, though they may have different low-level features. The format and the size of the database images are JPG and 384×256 respectively. The semantic groups used include: people, buses, decoration, elephants, flowers, horses, mountains, sunsets, food, fruit, road, seat, monument, stork, backgrounds, planets, airplanes, ocean, dinosaurs and butterflies. Figure 1 shows an image of each semantic groupas an example.

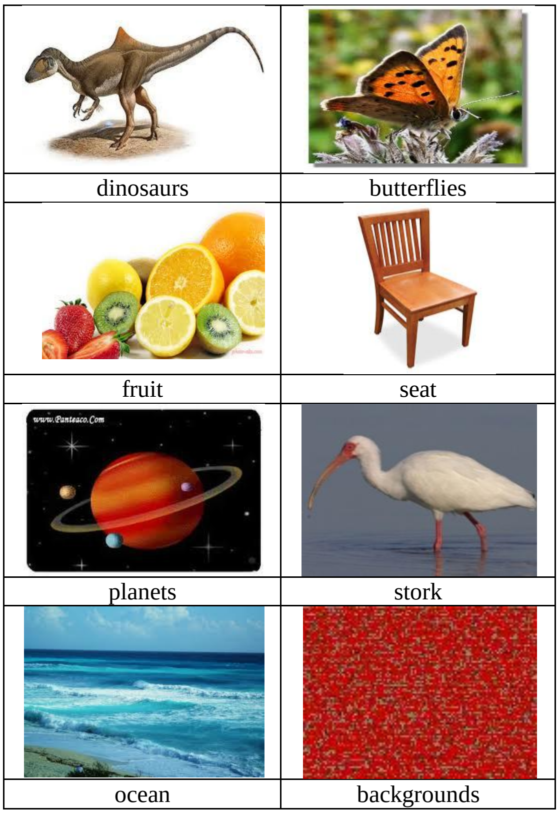




**Fig1: Examples of images in each semantic group**

### 4.1. K- Nearest Neighbors Classification Criterion

In most systems, nearest-neighbor search is used for searching for an image and finding similar images, which is, based on the definition by Chiu [5], this searching methods is after a point in the feature space which is at the minimum distance from the feature vector of the query image. Various methods are proposed for obtaining this collection of pictures, most of which use methods of indexing images and segmenting the feature space. Thus, the k nearest images to the desired image are extracted, where k is a predetermined value or set by the user. For doing the classification using each of the distance, each of the 2000 image in the database is presented in turn as the unlabeled image to the classifier. Then, the actual label of the image is compared to the label set by the classification algorithm to calculate classification rate of each distance. Figure 2 is an example of cbir code using k-nearest neighbors classification method.

**Query Image**

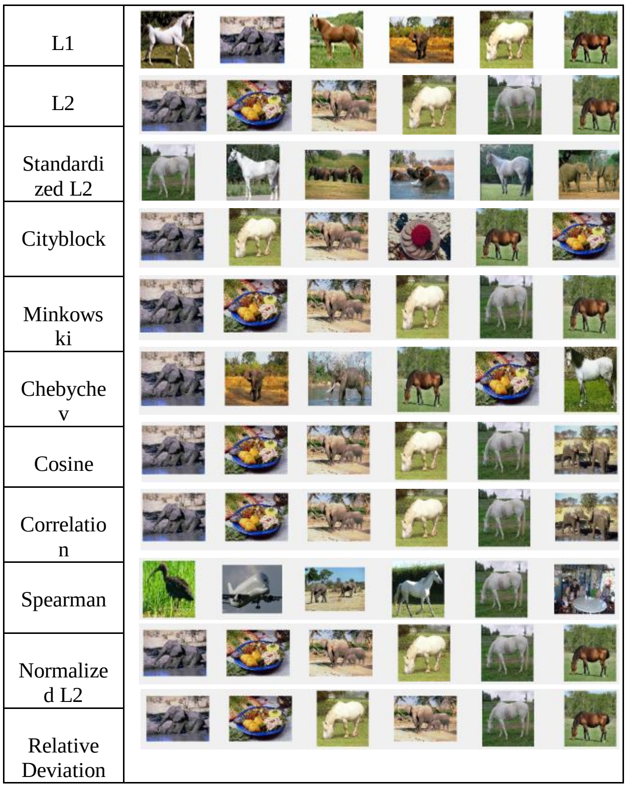



**Fig2: Example of image retrieval using k-NN classification**

## 4.2. Retrieval using Support Vector Machine

The machine learning algorithm predicts the category of the query image which is nothing but the semantic concept of the image. Therefore, instead of finding a similarity between the query image and all images in the database, only the similarity between the query image and the images in the same category as that of the query image is found. Also, when the entire database is searched, the resultant retrieved images are from various categories. However, when the machine learning techniques are used, the retrieved images include only images belonging to the same category for the category of the query image (semantic concept) is predicted.

## 4.3. Multi-class Support Vector Machine

Since the SVM is essentially a binary divider. Multi-class pattern recognition can be achieved by combining two-class support vector machines. There are typically two viewpoints for this purpose. One of them is the strategy of "one vs.all" to classify each pair of classes and the remaining classes, and the other is the strategy of "one vs. one" to classify each pair. For multi-class problems, the general solution is to reducethe multi-class problemto several binary problems. Each problem is solved using a binary divider. Then, output of the binary SVM's are combined, and thus the multi-class problem is solved [11]. In our case, thereare 20 classes of images, so there are 20 different classes and a multi-class problem is converted to binary classes. SVM always acts in pairs. Figure 3 is an example of image retrieval using SVM.

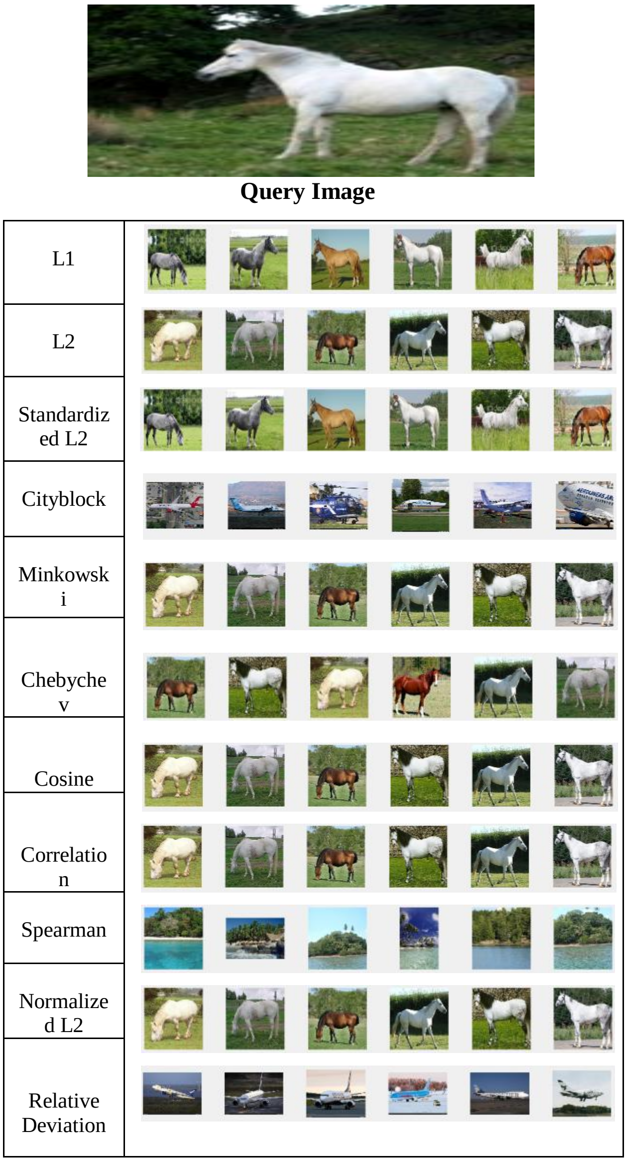



**Fig3: Example of image retrieval using SVM classification**

# 5. EXPERIMENT RESULTS AND INTERPRETATION

In this paper, 11 mentioned distances for semantic image retrieval and classification based on color and texture features using support vector machine are compared separately, and results are presented here.

The bar charts shown in Figures 4 and 5 related to the above-mentioned distances based on color moment, color correlogram, color histogram, Gabor feature, LBP operator and wavelet transform features inSVM and KNN methods.

The experiment is performed in a way that each time two query image from an image semantic group is tested against each single distance. The number of returned images is set to 20 images. In the all charts, the vertical axis shows the average number of returned images using specified distance in the horizantal axis. The blue colorcurve represents the image retrieval using svm classification and the red color curve also represents the image retrieval by knn classification.

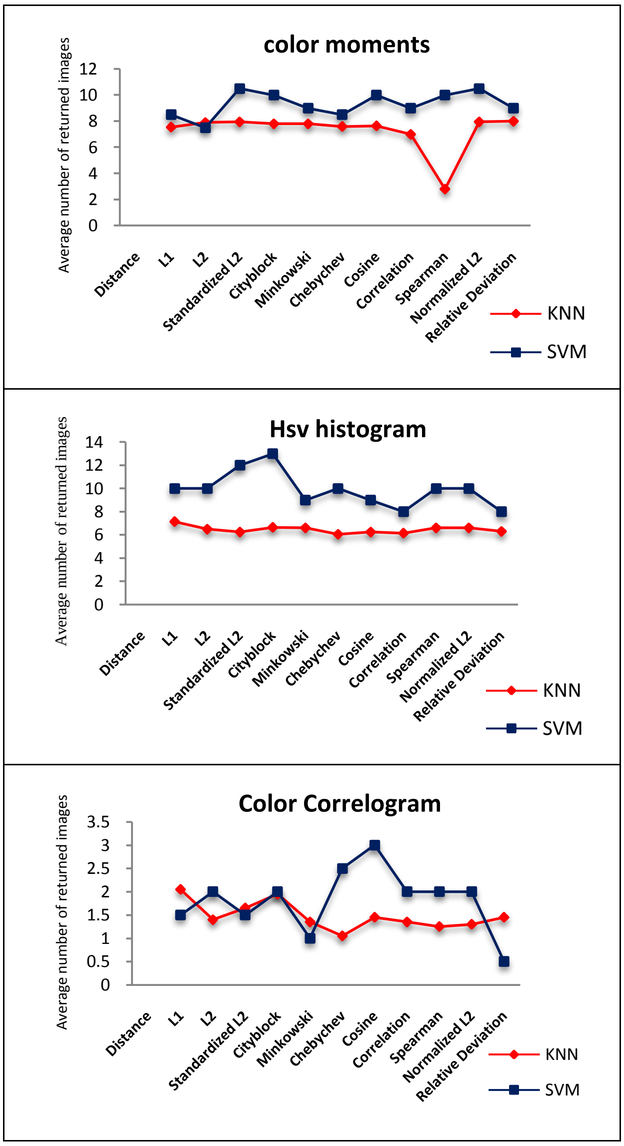
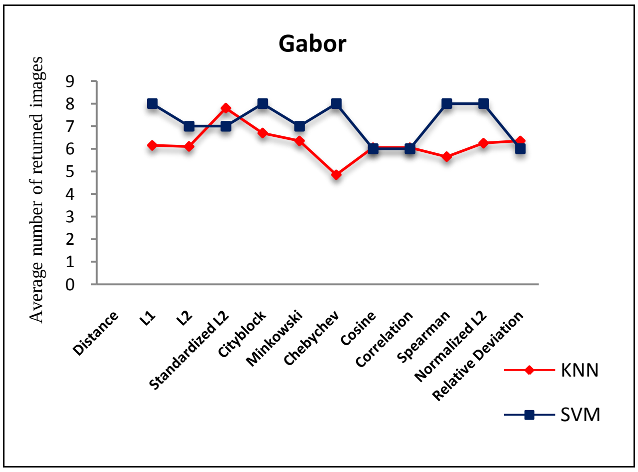



**Fig4:Comparison the performance of distances for color features in SVM and KNN classification**



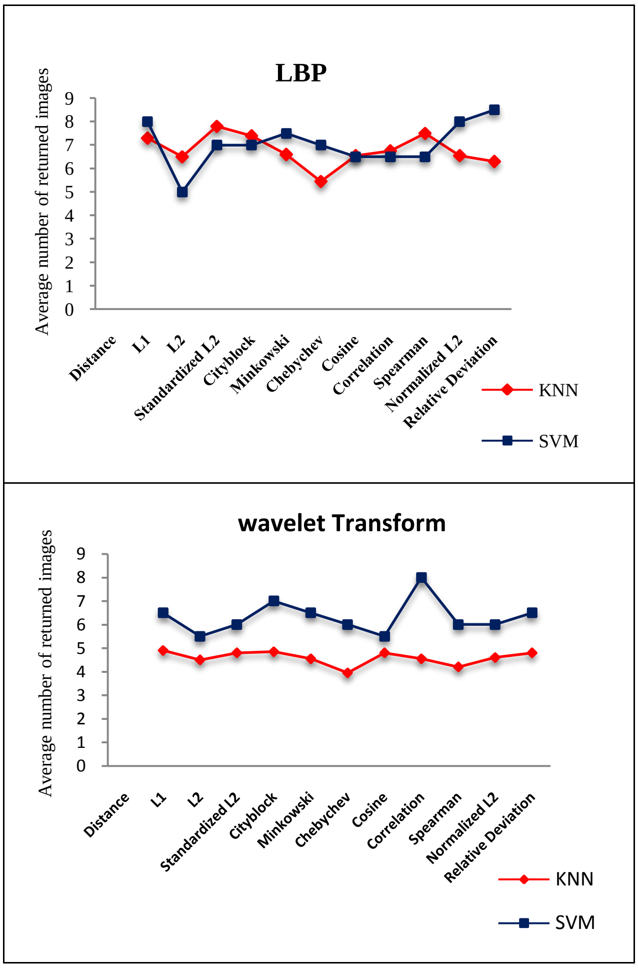




**Fig5:Comparison the performance of distances for texture features in SVM and KNN classification**

The results of the K-nearest neighbors is as follows:

The best results for image retrieval based on color feature are respectively associated with color moments, color histogram and color correlogram. Better results based on color moments are obtained with Standardized$L_2$ and Normalized$L_2$ distances. Regarding colorhistogram, $L_1$ and as for color correlogram, $L_1$ and Cityblock yield better results.

The best results regarding image retrieval based on texture are associated with LBP operator, Gabor feature and the wavelet transform, respectively. Better results based on LBP operator are obtained with Standardized$L_2$ and spearman distances. In Gabor feature, Standardized$L_2$ and in wavelet transform, $L_1$ and Cityblock give more favorable results.

The results of the support vector machine is as follows: The best results for image retrieval based on color feature are respectively associated with color histogram, color moments and color correlogram. Better results based on color histogram are obtained with $L_1$‘ $L_2$ and cityblock distances. Regarding color moment, Normalized$L_2$ and Standardized$L_2$ yield and as for color correlogram, cosine better results.

The best results regarding image retrieval based on texture are associated with Gabor feature, LBP operator and the wavelet transform, respectively. Better results based on, Gabor feature are obtained with chebychev and cityblock distances. In LBP operator, Relative deviation and $L_1$ and in wavelet transform, correlation give more favourable results.

The objective of content-based image retrieval research is find discovery methods, features and distinctions that we as humans use to identify similar images as well as directing the image query properly. Research has shown that in a generic system for image retrieval, using a single feature is not enough [7]. As it can also be observed here,the best results are those of the graphin Figure 6 where all features are used.

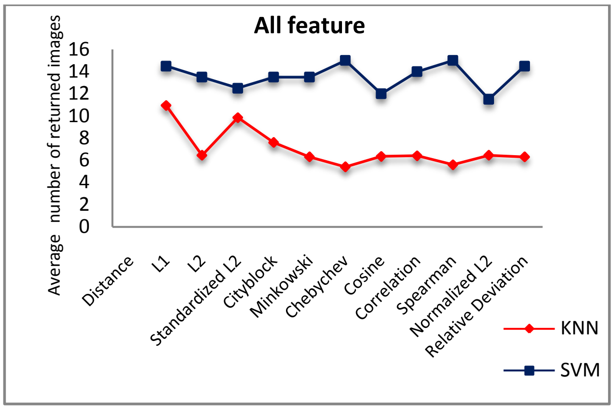



**Fig6: Comparison the performance of distances for all features in SVM and KNN classification**

## 6.THE PROPOSED METHOD

As shown in Figure 3, all the images are retrieved from the same class. Since any learning system is prone to error, there is a chance of viewing the wrong answer during the run. Notably, SVM doesn't reach complete compliance during the training time, so the correctanswer is not expected during the testing phase. That is why the detection accuracy and effectiveness confirmation are defined. Up until now, no learning method has ever been defined with a 100% efficiency.

Therefore, in this research, a method has been proposed to interact with the user so that if an error occurs or a wrong image is displayed during the first time that SVM classification runs, insubsequent runs, similar images are shown to the user or ensure him that the desired image is not in the database.

Since the best results are of Figure (6), the proposed method was tested in this mode (SVM with all features). Therefore, 100 experiments with SVM code including all the color and texture features were performed. The results areas follows: In 38% of cases, the user had between 0 to 2 interactions with SVM and accessed the similar images to the query. Similarly, in 29% of the cases, the user had 3 to 6, in 21% of the cases had 7 to 10, in 10% of the cases had 11 to 15, and in 2% of the cases had 16 to 20 interactions with the SVM and accessed the desired results. The diagramin figure 7 shows the results of these tests.

In this chart, the vertical axis shows the percent of accuracy rate and the horizantal axis reprecents the number of user interaction with svm classification.

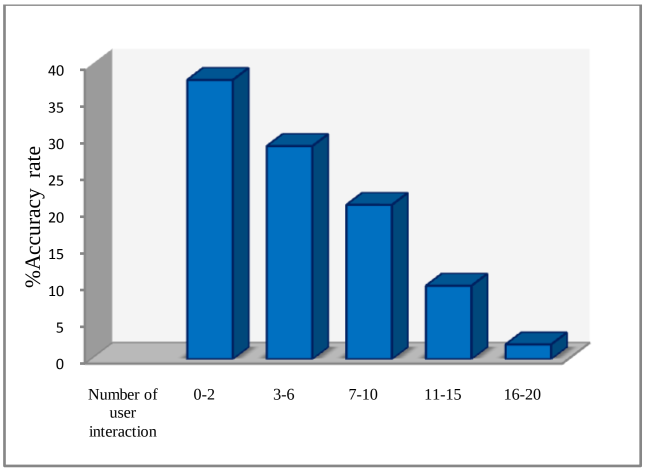



**Fig7:The exprimental results of proposed method**

With regard to the fact that the entire database is searched to determine the similarity, the time required to search and the properties of the machine on which is the search is done are shown in Table1.

**Table1: Properties of machine and processing time for a query in the system**

| Machineproperties | |
|---|---|
| Intel Cor5 | Processor |
| 1GB | Memory usage |
| Matlab | Used software |
| 10 Second | Processing time of a query with K-NN classification |
| 20 Second | Processing time of aquery with SVM classification (train and test) |

## 7. CONCLUDING

The choice of an appropriate distanceas a measure to determine the similarity between two images based on human perception can play an important role in reducing the semantic gap and bringing the low-level features and the semantic features closer to each other.

1. In this regard, eleven different distances for color and texture features in a database containing 2000 images for image retrieval and classification were compared with each other in both support vector machine and K nearest neighbors methods.

2. The results of the K nearest neighbours is as follows:

- The best results for image retrieval based on color feature are respectively associated with color moments, color histogram and color correlogram. Bette results based on color moments are obtained with Standardized$L_2$ and Normalized$L_2$ distances. Regarding color histogram, $L_1$ and as for color correlogram, $L_1$ and Cityblock yield better results.

- The best results regarding image retrieval based on texture are associated with LBP operator, Gabor feature and the wavelet transform, respectively. Better results based on LBP operator are obtained with Standardized$L_2$ and spearman distances. In Gabor feature, Standardized$L_2$ and in wavelet transform, $L_1$ and Cityblock give more favorable results.

3. The results of the support vector machine is as follows:

- The best results for image retrieval based on color feature are respectively associated with color histogram, color moments and color correlogram. Better results based on color histogram are obtained with $L_1$‘$L_2$ and Cityblock distances. Regarding color moment, Normalized$L_2$ and Standardized$L_2$ yield and as for color correlogram, Cosine better results.

- The best results regarding image retrieval based on texture are associated with Gabor feature, LBP operator and the wavelet transform, respectively. Better results based on, Gabor feature are obtained with Chebychev and Cityblock distances. In LBP operator, Relative Deviation and $L_1$ and in wavelet transform, Correlation give more favorable results.

4. Finally, the results show that a graph in which all features are used obtains a better result than any of the features mentioned. The results of this experiment also show that retrieval by support vector machines yield better results than the k-nearest neighbors.

Although existing methods of content-based image retrieval demonstrated desirable efficacy, this is an open research area which requires considerable efforts inissues such as increasing the speed of retrieval in databases with very large images as well as the accuracy of the retrieval.

## 8. REFERENCES

[1] Chadha, A., Mallik, S., Johar, R. 2012. Article: Comparative Study and Optimization of Feature-Extraction Techniques for Content based Image Retrieval. International Journal of Computer Applications, Vol.52, No.20, 35-42.

[2] Long, F., Zhang, H., Dagan, D., Feng,. 2003. Fundamentals of content-based image retrieval, in Multimedia Information etrieval and Management – Technological Fundamentals and Applications, Springer-Verlag.

[3] Einarsson, S. H. , Grétarsdóttir , R. Ý. , Jónsson , B. Þ., , Amsaleg, L. 2005. The EFF 2 Image retrieval System Prototype”, in ASTED Intl. Conf. on Databases and Applications (DBA), Innsbruck, Austria.

[4] Li, X. , S.C. , M.L. , Chen, Shyu , Furht, B. 2002. Image Retrieval by Color, Texture, and Spatial Information, in 8th nternational Conference on Distributed multimedia Systems (DMS'2002), San Francisco Bay, California, USA.

[5] Smith, J. R. , and Chang, S. F. 1996. Visual SEEk: A fully automated content-based image query system, in ACM Multimedia Conference. Boston, MA, USA.

[6] Ojala, T., Pietikäinen, M., Mäenpää, T.,2002. Multiresolution gray-sclae and rotation invariant texture classification with local binary patterns, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 24, No.7, 971-987.

[7] Gevers. Th., Smeulders , A.W.M. ,2003. Image Search Engines, An Overview, The International Society for Optical Engineering (SPIE), Vol.8.

[8] K. Yang, J. Trewn,2004. Multivariate Statistical Methods in Quality Management. McGraw-Hill Professional, 1st Edition.

[9] P. N. Tan, M. Steinbach, V. Kumar,2005. Introduction to Data Mining, eBook-EnG Addison-Wesley, 500.

[10] C. Spearman,1904. The proof and measurement of association between two things The American Journal of Psychology. ,Vol.15, 72–101.

[11] I. Felci Rajam, S. Valli, 2013. A Survey on Content Based Image Retrieval, Life Science Journal, Vol.10, No. 2.

[12] R. C. Gonzalez, E. R. Woods,2002. Digital Image Processing Book, Second Edition, Prentice Hall, Upper Saddle River, NJ.