

---

**COMPARATIVE STUDY OF CONTENT BASED IMAGE RETRIEVAL  
TECHNIQUES FOR BACTERIAL CELLS CLASSIFICATION**

*By*

**A. E. E. ElAlfi**

*Dept. of Computer Science  
Faculty of Specific Education  
Mansoura, Egypt*

**A. R. M. Elzainy**

*Dept. of Home Economy  
Faculty of Specific Education  
Mansoura, Egypt*

**M. E. E. ElAlmi**

*Dept. of Computer Science  
Faculty of Specific Education  
Mansoura, Egypt*

**S. A. Shaban**

*Dept. of Computer Science  
Faculty of Specific Education  
Mansoura, Egypt*

**Research Journal Specific Education**

**Faculty of Specific Education**

**Mansoura University**

**ISSUE NO. 35, JULY. 2014**

مجلة بحوث التربية النوعية – جامعة المنصورة

العدد الخامس والثلاثون – يوليو ٢٠١٤

---



## **COMPARATIVE STUDY OF CONTENT BASED IMAGE RETRIEVAL TECHNIQUES FOR BACTERIAL CELLS CLASSIFICATION**

*A. E. E. ElAlfi*<sup>\*</sup>

*M. E. E. ElAlmi*<sup>\*\*</sup>

*A. R. M. Elzainy*<sup>\*\*\*</sup>

*S. A. Shaban*<sup>\*\*\*\*</sup>

### **Abstract**

A major challenge in microbial ecology is to develop reliable and facile methods of computer assisted microscopy that can analyze digital images of complex microbial communities at single cell resolution, and compute useful quantitative characteristics of their organization and structure without cultivation. This paper performs a fair comparative study of three most common feature extraction methods for content based bacterial image classification; Color Coherence Vector (CCV), Gray Level Co-occurrence Matrix (GLCM) and Color Moments. The comparison is made on the basis of these parameters: Run time, CPU time, Recall, F-measure and Precision measures. Performance of Feature Extraction techniques based on the number of retrieved images, Run time and CPU time. Experiments results are given to analyses the effectiveness of each techniques and how that Color Moments technique is more suitable than other techniques.

**Keywords:** Bacteria - Image processing – CBIR-CCV- GLCM- Color Moments.

### **1. Introduction**

Bacteria are unicellular microscopic organisms which can only be seen through microscope. Bacteria exist in different sizes and shapes and they measure in micro-meter (which is a millionth part of a meter). Bacteria are found everywhere and in all types of environments. There are numerous types of bacteria in the world. Bacteria are mainly classified based on their shapes, biochemistry and staining methods [1].

<sup>\*</sup>Dept. of Computer Science Faculty of Specific Education Mansoura, Egypt

<sup>\*\*</sup>Dept. of Computer Science Faculty of Specific Education Mansoura, Egypt

<sup>\*\*\*</sup>Dept of Home Economy Faculty of Specific Education Mansoura, Egypt

<sup>\*\*\*\*</sup>Dept. of Computer Science Faculty of Specific Education Mansoura, Egypt

Bacteria are the microorganisms which have both the positive and negative impacts for human. They are beneficial for human because they are used in a number of industries like for producing dairy products such yogurt, cheese etc. They are also used in leather industry for making leather. Bacteria's are also economical beneficial organisms because Nitrogen fixing bacteria's, increases the fertility of the soil by processing nitrogen in the soil. Bacteria are also important for human health because they are present in our body and are producing a vitamin type in our body. Moreover, bacteria are also used in manufacturing medicines like anti-bacterial medicines. Along with the beneficial importance of bacteria, they also have some harmful effect on the human body.

Figure 1 shows some types of harmful bacteria.

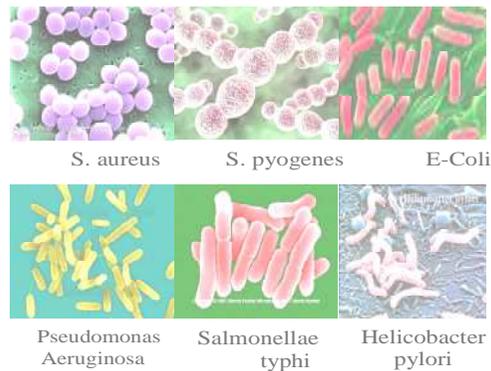


Fig. 1 Types of harmful Bacteria

The direct approach to examine the microbe's world from its own perspective is microscopy, which is one of the most important techniques in microbial ecology. The value of quantitative microscopy in microbial ecological studies can be increased even further when used in conjunction with computer-assisted image analysis [2].

Image processing and computer modeling are important tools in most medical imaging domains, and have more recently started to attract the attention of the biological community and to take a growing role in biological imaging applications.

Todate, many of the biological and microbiological data analysis entail a substantial amount of human intervention. Manual procedures are based on subjective human interpretation, are prone to large variability between the human experts, are time consuming and are of great cost [3].

Digital image processing and pattern recognition techniques are used in conjunction with microscopy for quantitative studies of microbial ecology. These techniques provide an important quantitative tool to analyze the structures and spatial features of complex microbial communities [4].

One of the most important and yet most tedious tasks performed during microscopical analysis of microbial communities is the classification of observed cells into known morphological categories and recognition of new categories as well if new distinct characteristics are captured [5].

Content Based Image Retrieval (CBIR) is a technology that in principle helps to organize digital image archives by their visual content. By this definition, anything ranging from an image similarity function to a robust image annotation engine falls under the purview of CBIR [6].

In CBIR systems, images automatically indexed by summarizing their visual features. A feature is a characteristic that can capture a certain visual property of an image either globally for the entire image or locally for regions or objects. Color, texture and shape are commonly used features in CBIR systems [7].

Feature extraction is the basis of content-based image retrieval. It is the process of extracting features from the image such as color, shape and texture. It computes a numerical or alphabetical representation of some attribute of digital images [8].

The main goal of CBIR is efficiency during image indexing and retrieval, thereby reducing the need for human intervention in the indexing process [6]. The computer must be able to retrieve images from a database without any human assumption on specific domain [9].

The process of CBIR consists of three stages [10]:

- (1) Image acquisition
- (2) Feature Extraction
- (3) Similarity Matching

Figure 2 shows Architecture of CBIR system [11] .

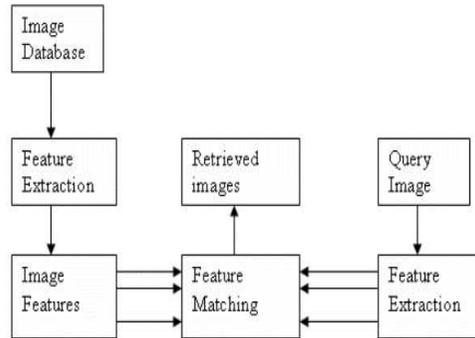


Fig.2 Architecture of CBIR system

For the given image database, features are extracted first from individual images. The features can be visual features like color, texture, shape, region or spatial features or some compressed domain features. The extracted features are described by feature vectors. These feature vectors are then stored to form image feature database. For a given query image, we similarly extract its features and form a feature vector. This feature vector is matched with the already stored vectors in image feature database. Sometimes dimensionality reduction techniques are employed to reduce the computations. The distance between the feature vector of the query image and those of the images in the database is then calculated. The distance of a query image with itself is zero if it is in database. The distances are then stored in increasing order and retrieval is performed with the help of indexing scheme .

## 2.Survey of Content Based Image Retrieval

In CBIR systems, a feature is a characteristic that can capture a certain visual property of an image either globally for the entire image or locally for regions or objects . The low level features commonly used in CBIR are color, texture, shape and edge.

### 2.1 Color Features

Color features are extracted using color moments, color histogram, color coherence vector, invariant color histogram, and dominant color. color moments and color coherence vector are explained in the following section.

### 2.1.1 Color Moments

There are four color moments used for color feature extraction. These moments are: the mean, the standard deviation, the skewness and the kurtosis [12].

The first color moment ( $E_i$ ) can be calculated by using the following formula [13]:

$$E_i = \sum_{j=1}^N \frac{1}{N} P_{ij} \quad (1)$$

Where:

$N$  = number of pixels in the image

$P_{ij}$  = value of the  $j$ th pixel of the image at the  $i$ th color channel.

The second color moment ( $\sigma_i$ ) can be calculated by using the following formula;

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

The third color moment is the skewness ( $S_i$ ). It can be calculated by using the following formula;

$$S_i = \sqrt[3]{\left(\frac{1}{N} \sum_{j=1}^N (p_{ij} - E_i)^3\right)} \quad (3)$$

The fourth color moment is the kurtosis ( $K_i$ ). It can be calculated by using the following formula;

$$K_i = \sqrt[4]{\left(\frac{1}{N} \sum_{j=1}^N (p_{ij} - E_i)^4\right)} \quad (4)$$

### 2.1.2 Color Coherence Vector

Color's coherence is defined as the degree to which pixels of that color are members of large similarly-colored regions. The significant regions are importance in characterizing images. Colored pixels are either coherent or

incoherent. Coherent pixels are part of some sizable contiguous region, while incoherent pixels are not. A color coherence vector (CCV) represents this classification for each color in the image [14].

A pixel is coherent if the size of its connected component exceeds a fixed value  $\tau$ ; otherwise, the pixel is incoherent. The color coherence pair's vector for the image consists of [15] :

$$\langle (\alpha_1 + \beta_1), \dots, (\alpha_n + \beta_n) \rangle$$

Where;  $\alpha_n, \beta_n$  are the number of coherent pixels of the  $n$ th discrete color and the number of incoherent pixels respectively.

## 2.2 Texture Features

Texture is another important property of images. Various texture representations have been investigated in pattern recognition and computer vision [16]. Texture features are extracted using Gray Level Co-occurrence matrix (GLCM), Gabor Transform and Tamura Features [17]. Gray Level Co-occurrence matrix (GLCM) is explained in the following section.

### 2.2.1 Gray Level Co-occurrence matrix

The GLCM is created from a gray-scale image. It finds how often a pixel with a gray-level value  $i$  occurs either horizontally, vertically, or diagonally to adjacent pixels with the value  $j$  [18]. It is given by the relative frequency of the occurrences of two gray-level pixels  $i$  &  $j$ , separated by  $d$  pixels in the  $\theta$  orientation and  $\theta$  is the direction. The  $d$  can take values 1, 2, 3, etc., and  $\theta$  can take values  $0^\circ$  (horizontal),  $90^\circ$  (vertical),  $45^\circ$  and  $135^\circ$  (diagonal) [19].

The GLCM is used for texture feature extraction. These features are: the energy, contrast, correlation and the homogeneity.

The first GLCM texture feature can be calculated by using the following formula [20] ;

$$Energy = \sum_{i,j} p(i, j)^2 \quad (5)$$

Where:

$i, j$  are a single pixel.

$p(i, j)$  is the probability.

The second GLCM texture feature can be calculated using the following formula;

$$Contrast = \sum_{i,j} |i - j|^2 p(i, j) \quad (6)$$

The third GLCM texture feature is the correlation. It can be calculated using the following formula;

$$Correlation = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j) p(i, j)}{\sigma_i \sigma_j} \quad (7)$$

Where  $\mu_i$  represents the horizontal mean,  $\mu_j$  represents the vertical mean in the matrix,  $\sigma_i$  and  $\sigma_j$  represent dispersion around the mean of combinations of target and neighbor pixel.

The fourth GLCM texture feature is the homogeneity. It can be calculated using the following formula;

$$Homogeneity = \sum_{i,j} \frac{p(i, j)}{1 + |i - j|} \quad (8)$$

### ***Pattern discrimination***

There are many methods that can be used for pattern classification such as Weighted Euclidean distance measure [21], Correlation coefficient method, Logarithmic magnitude distance measure, Minimum mean distance rule, Artificial Neural Networks (ANN) [22], decision tree [23] and K-means [24]. In this paper the Weight Euclidean Distance (WED) is used.

$$d(v, v^k) = \sqrt{\sum_{i=1}^n p_i (v_i - v_i^k)^2} \quad (9)$$

Where:

$v_i$  to balance the variations in the dynamic range.

$p_i$  the weight added to the component.

$k$  is the matched image index.

$$P_i = \frac{N}{\sum_{k=1}^N (v_i^k - \bar{v}_i)^2} \quad (10)$$

N = the number of images in databases.

$$\bar{v}_i = \frac{\sum_{k=1}^N v_i^k}{N} \quad (11)$$

### **3. Performance Evaluation**

Evaluation of retrieval performance is a crucial problem in content-based image retrieval (CBIR). Many different methods for measuring the performance of a system have been created and used by researchers. The most common evaluation measures used in CBIR are precision and recall which are defined

as [11]:

$$\text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}}$$

$$\text{Recall} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images in database}}$$

A single measure that trades off precision versus recall is the F-measure which is the weighted harmonic mean of precision and recall [25]:

$$\text{F-Measure} = \frac{2 \cdot (\text{precision} \times \text{recall})}{(\text{Precision} + \text{recall})}$$

### **4. A Comparative Study**

A comparison among feature extraction methods (Gray Level Co-occurrence Matrix, Color Moments and Color Coherent Vector) for content based bacterial image classification is presented.

Each of the feature extraction technique has their own strong and weak points.

The extracted features by using previous methods are described by feature vectors.

These feature vectors are then stored to form image feature database. Thus, three image feature databases have been formed , one to each feature extraction method . For a given query image, the previous methods have been applied and similarly extract its features and form a feature vector. This feature vector is matched with the already stored vectors in image feature database. Recall , f-measure and precision measures have been calculated to each method for a query image to know the best method of them. Run time and CPU time have been calculated to each method also.

### 5. Experimental Results

An image database is used for bacteria classification [26]. It includes 150 images . It is consisted of 3 classes of bacteria namely , Bacilli, Cocci and Spiral . Matlab program is developed and used for image retrieval.

Table 1 shows Run time and CPU time for feature extraction methods

F. Extraction Methods	Time			
	Learn Time (sec)		Test Time (sec)	
	CPU	Run	CPU	Run
GLCM	53.9	53.5	250.9	2.1
CCV	5.9e+003	5913	20.4	20.6
Color Moments	28.4	20.0	290.2	1.0

When tracing CPU time and Run time results, one can find that color moments results are better than Gray Level Co-occurrence Matrix and Color Coherent Vector .

Figure 3 shows bacilli Bacterial cell as a query image.



Fig. 3 bacilli Bacterial Cell

Figure 4 shows the resulted images due to using GLCM feature extraction method for content based bacterial image retrieval.



Fig.4 Retrieved Images by using GLCM Feature Extraction Method.

Figure 5 shows the resulted images due to using CCV feature extraction method for content based bacterial image retrieval.

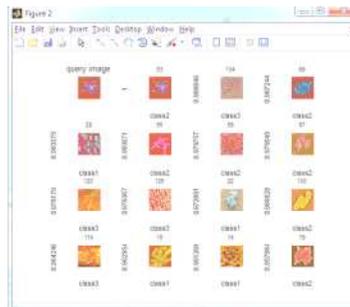


Fig.5 Retrieved Images by using Color Coherent Vector Feature Extraction Method.

Figure 6 shows the resulted images due to using Color Moments feature extraction method for content based bacterial image retrieval.



Fig.6 Retrieved Images by using Color Moments Feature Extraction Method.

Figure 7 shows recall curve for Color Coherent Vector feature extraction method.

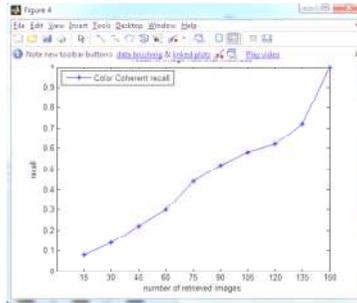


Fig.7 Recall Curve for Color Coherent Vector Feature Extraction Method

Figure 8 shows recall curve for GLCM feature extraction method.

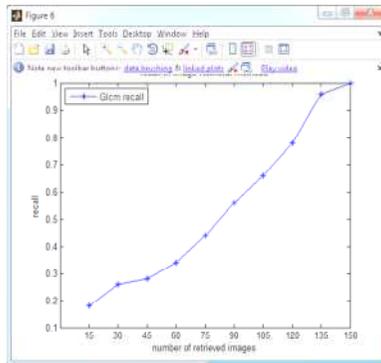


Fig. 8 Recall Curve for GLCM Feature Extraction Method

Figure 9 shows recall curve for Color Moments feature extraction method.

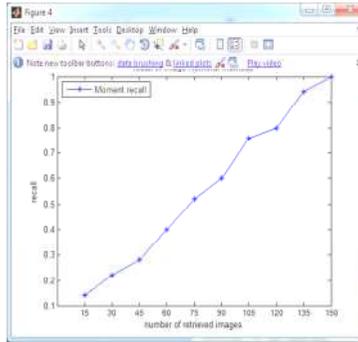


Fig.9 Recall Curve for Color Moments Feature Extraction Method

From the previous figures one conclude that Color Moments results are better than GLCM and Color Coherent Vector methods.

Figure 10 shows Precision curve for GLCM feature extraction method.

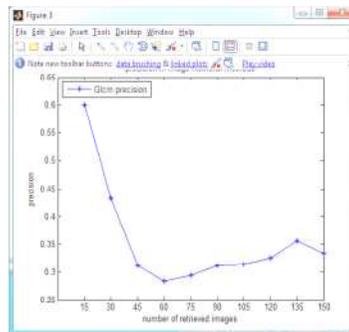


Fig.10 Precision Curve for GLCM Feature Extraction Method

Figure 11 shows Precision curve for Color Coherent Vector feature extraction method.

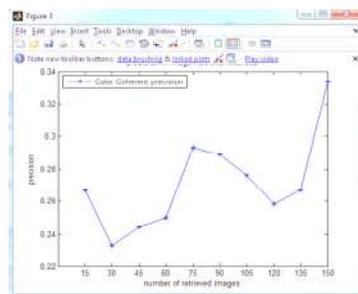


Fig.11 Precision Curve for Color Coherent Vector Feature Extraction Method

Figure 12 shows Precision curve for Color Moments feature extraction method.

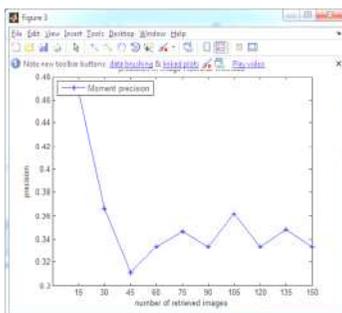


Fig.12 Precision Curve for Color Moments Feature Extraction Method

When tracing Precision curve results, one can find that Color Moments results are better than GLCM and Color Coherent Vector .

Figure 13 shows f-measure curve for GLCM feature extraction method.

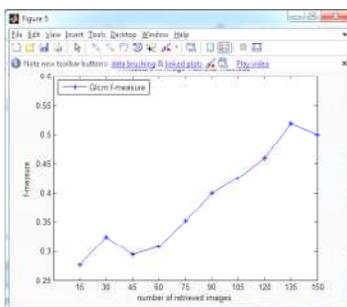


Fig. 13 F-measure Curve for GLCM Feature Extraction Method

Figure 14 shows f-measure curve for Color Coherent Vector feature extraction method.

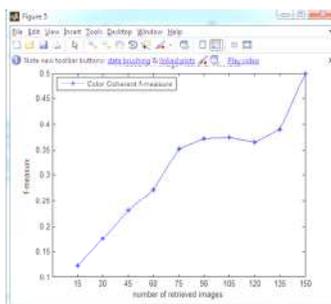


Fig.14 F-measure Curve for Color Coherent Vector Feature Extraction Method

Figure 15 shows f-measure curve for Color Moments feature extraction method.

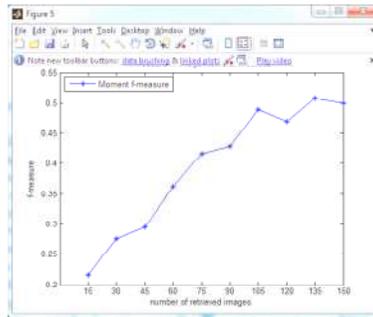


Fig.15 F-measure Curve for Color Moments Feature Extraction Method

When tracing f-measure curve results, one can find that Color Moments results are better than GLCM and Color Coherent Vector .

## 6. Conclusion

Feature extraction techniques play an important role in content based bacterial cells classification.

In this paper a comparative study of different feature extraction techniques is presented. The selected GLCM, CCV and Color Moments techniques are used for performance evaluation. Based on the experimental result it was concluded that Color Moments technique consumes least CPU time and least Run time , also better than other feature extraction methods in recall, precision and f-measure performances. This can lead to better classification for bacterial cells and can be used for multiple purposes such diagnosis of bacterial diseases, help researchers in the identification of the type of bacteria and their relation to the corruption of food. Also it can help students in the Department of Microbiology to understand the subject of bacteria classification. The identification of bacterial cell growth and the time of cell division have a great benefit in sensitivity tests against a particular drug when designing effective medication to control bacterial disease.

## References

1. P.S.Hiremath and Parashuram Bannigidad (2009) : Automatic Classification of Bacterial Cells in Digital Microscopic Images, International Journal of Engineering and Technology, Vol(2), No( 4), p: 9.
2. P.S. Hiremath and Parashuram Bannigidad (2010): Automatic Classification of Bacilli Bacterial Cells in Digital Microscopic Images using Active Contour Model , International Journal of Advances in Science and Technology, Vol(1), No(5 ), p: 5216.

3. Sigal Trattner and others (2004): Automatic Identification of Bacterial Types Using Statistical Imaging Methods, IEEE transactions on medical imaging, vol(23), No(7), p: 807.
4. P.S. Hiremath and Parashuram Bannigidad (2010): Automatic Identification and Classification of Bacilli Bacterial Cell Growth Phases, IJCA Special Issue on "Recent Trends in Image Processing and Pattern Recognition", p:48.
5. J. Liu, F.B. Dazzo, O. Glagoleva, B. Yu and A.K. Jain (2001) : CMEIAS: A Computer-Aided System for the Image analysis of Bacterial Morphotypes in Microbial Communities, Springer-Verlag New York Inc, p:174.
6. Nidhi Singhai and Shishir K. Shandilya(2010) : A Survey On: Content Based Image Retrieval Systems, International Journal of Computer Applications (0975 – 8887), Vol(4) , No(2), p:22.
7. S.R. Kodituwakku1 and others (2011): Comparison of Color Features for Image Retrieval, Indian Journal of Computer Science and Engineering , Vol(1), No(3) , p: 207.
8. Malti Puri (2013) : A Survey on Content Based Image Retrieval, International Journal of Computer Science & Engineering Technology (IJCSET) , Vol( 4), No(7), p: 1004.
9. Swarnalata .V and others (2013): Survey on Content Based Image Retrieval System, International Journal of Scientific & Engineering Research, Vol(4), Issue(12), ,p:89.
10. Dr. h.b. kekre and others (2011): A Survey of CBIR Techniques and Semantics, International Journal of Computer Engineering Science and Technology, Vol(3), No(5) , p: 4510.
11. Yogita Mistry and Dr.D.T. Ingole (2013) : Survey on Content Based Image Retrieval Systems, International Journal of Innovative Research in Computer and Communication Engineering, Vol(1), Issue(8), pp: 1827- 1835.
12. I.Felci Rajam and S. Valli (2013): A Survey on Content Based Image Retrieval, Life Science Journal , vol(10),No(2), p:2476.
13. H.H.Pavan Kumar Bhuravarjula and V.N.S Vijaya Kumar (2012) :A novel Content Based Image Retrieval using Variance Color Moments, International Journal of Computer and Electronics Research , vol(1), Issue (3), p:94.
14. Reshma Chaudhari and A. M. Patil (2012): Content Based Image Retrieval Using Color and Shape Features, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol(1), Issue( 5), pp:387-388.
15. Greg Pass, Ramin Zabih and Justin Miller (1996):Comparing Images Using Color Coherence Vectors, Proceedings of the fourth ACM international conference on Multimedia, vol(96), pp: 66-67.
16. Avinash N Bhute and B. B. Meshram (2013) : Content Based Image Indexing and Retrieval, International Journal of Graphics & Image Processing, Vol(3),issue (4) , p:237.

17. I.Felci Rajam and S. Valli(2013): A Survey on Content Based Image Retrieval, Life Science Journal , vol( 10),No(2), p:2476.
18. Swapnalini Pattanaik and Prof.D.G.Bhalke(2012) :Beginners to Content Based Image Retrieval, International Journal of Scientific Research Engineering &Technology (IJSRET), Vol ( 1), Issue(2), pp: 040-044.
19. Rahman M.M., Bhattacharya M.P., and Desai B.C(2007): A framework for medical image retrieval using machine learning and statistical similarity matching techniques with feedback, IEEE Trans.Inform.Technol.Biomed.,Vol(11), No(1), pp:58-69.
20. Matlab , R2012a (7.14.0.739).
21. R. Mukundan and K. R. Ramakrishnan :Moment Functions In Image Analysis Theory and Applications, World Scientific Publishing Co. Pte. Ltd. 1998 ISBN 981-02-3524-0, pp:81-85.
22. M.seetha and others (2008) : Artificial Neural Networks and Other Methods of Image Classification, Journal of Theoretical and Applied Information Technology ,pp:1039-1051.
23. Bhaskar N. Patel and others(2012): Efficient Classification of data using decision tree, Bonfring International Journal of Data Mining , Vol(2), No(1), pp:1-7.
24. Balasubramanian Subbiah and Seldev Christopher (2012): Image Classification through integrated K- Means Algorithm, International Journal of Computer Science Issues, Vol(9), Issue (2), No (2), pp:518-522.
25. M. Keyvanpour, R. Tavoli and S. Mozaffari (2014): Document Image Retrieval Based on Keyword Spotting Using Relevance Feedback, International Journal of Engineering, Vol(27), No. (1), pp: 7-14
26. <http://www.fotosearch.com/photos-images/bacteria.html>