



Efficient Image Retrieval using Region based Image Retrieval

Ramesh K Kulkarni

Department of Electronics and Telecommunication,
Vivekanand Institute of Technology,
University of Mumbai,

Niket Amoda

Department of Electronics and Telecommunication,
Vivekanand Institute of Technology,
University of Mumbai,

ABSTRACT

Early image retrieval techniques were based on textual annotation of images. Annotating images manually is a cumbersome and expensive task for large image databases, and is often subjective, context-sensitive and incomplete. Content based image retrieval, uses the visual contents of an image such as color, shape, texture, and spatial layout to represent and index the image. The Region Based Image Retrieval (RBIR) system uses the Discrete Wavelet Transform (DWT) and a k-means clustering algorithm to segment an image into regions. Each region is represented by means of a set of features and the similarity between regions is measured using a specific metric function on such features.

Keywords

Content based image retrieval, K-Means Algorithm, Discrete Wavelet Transform, Region Based Image Retrieval.

1. Introduction

Early image retrieval techniques were based on textual annotation of images. Through text descriptions, images can be organized by topical or semantic hierarchies to facilitate easy navigation and browsing based on standard Boolean queries. It was widely recognized that a more efficient and intuitive way to represent and index visual information would be based on properties that are inherent in the images themselves.

Content based image retrieval, uses the visual contents of an image such as color, shape, texture, and spatial layout to represent and index the image. In typical content based image retrieval systems, the visual contents of the images in the database are extracted and described by multi-dimensional feature vectors. The feature vectors of the images in the database form a feature database.

To retrieve images, users provide the retrieval system with example images or sketched figures. The system then changes these examples into its internal representation of feature vectors. The similarities / distances between the feature vectors of the query example or sketch and those of the images in the database are then calculated and retrieval is performed with the aid of an indexing scheme. The indexing scheme provides an efficient way to search the image database.

Recent retrieval systems have incorporated users' relevance feedback to modify the retrieval process in order to generate perceptually and semantically more meaningful retrieval results

A visual content descriptor can be either global or local. A global descriptor uses the visual features of the whole image, whereas a local descriptor uses the visual features of regions or objects to describe the image content. To obtain the local visual descriptors, an image is often divided into parts first.

Some of the widely used techniques for extracting color, texture, shape and spatial relationship features from images are now described briefly.

Instead of exact matching, content based image retrieval systems calculate the visual similarities between a query image and the images in a database. Accordingly, the retrieval result is not a single image but a list of images ranked by their similarities with the query image. Different similarity/distance measures will affect retrieval performances of an image retrieval system significantly. Some of the commonly used similarity measures are: Euclidean Distance, Mahalanobis Distance and Bhattacharyya Distance

Another important issue in content-based image retrieval is effective indexing and fast searching of images based on visual features. Because the feature vectors of images tend to have high dimensionality and therefore are not well suited to traditional indexing structures, dimension reduction is usually used before setting up an efficient indexing scheme. One of the techniques commonly used for dimension reduction is principal component analysis (PCA). It is an optimal technique that linearly maps input data to a coordinate space such that the axes are aligned to reflect the maximum variations in the data.

After dimension reduction, the multi-dimensional data are indexed. A number of approaches have been proposed for this purpose, including R-tree (particularly, R*-tree), Linear quad-trees, K-d-B tree and Grid files

2. wavelet Transforms

Signals generated from natural sources such as digital images are often non-stationary in nature i.e. their content varies in time or space.

Frequency analysis of stationary signals can be effectively achieved by projecting the signal onto a set of infinite spatial extent basis functions using the Fourier transform:

$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{-2j\pi ft} dt \quad (1)$$

where $X(f)$ represents the global frequency of the signal. Similarly, effective frequency analysis of non-stationary signals can be achieved by projecting the signal onto a set of spatially localized basis functions using the wavelet transform.

$$\Psi(a, b) = \int_{-\infty}^{+\infty} x(t) \Psi_{ab}^*(t) dt \quad (2)$$

where $a, b \in \mathbb{R}$ and $\psi_{ab}(t)$ is the translated and scaled version of the mother wavelet $\psi(t)$ given by

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) \quad (3)$$

Different choices of a and b result in the many different possible wavelet bases at different scales and translations.

A. Discrete Wavelet Transform (DWT)

For the discrete case the continuous convolution in the above equation is replaced by adiscrete summation:

$$c_{m,n} = \langle x, \Psi_{m,n} \rangle \equiv \sum_l x(l) \Psi_{m,n}(l) \quad (4)$$

where $c_{m,n}$ are the wavelet coefficients. The convolution of the scaling function with the signal is implemented at each scale through the iterative filtering of the signal with a low pass FIR filter h_n . The approximation coefficients $a_{m,n}$ at each scale can be obtained using the following recursive relation:

$$a_{m,n} = \sum_k h_{2n-k} a_{m-1,k} \quad (5)$$

where the top level $a_{0,n}$ is the sampled signal itself. In addition, by using a related highpass FIR filter g_n the wavelet coefficients can be obtained using the further recursive relation:

$$c_{m,n} = \sum_k g_{2n-k} a_{m-1,k} \quad (6)$$

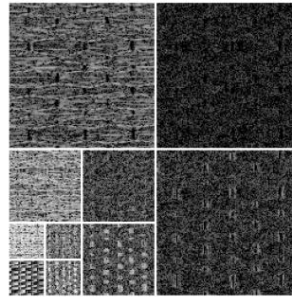
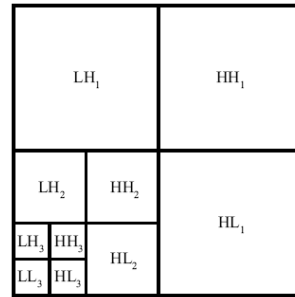
To reconstruct the original signal, the analysis filters can be selected from a biorthogonal set which have a related set of synthesis filters. These synthesis filters \tilde{h} and \tilde{g} can be used to perfectly reconstruct the signal using the reconstruction formula:

$$a_{m-1,l}(f) = \sum_n [\tilde{h}_{2n-l} a_{m,n}(f) + \tilde{g}_{2n-l} c_{m,n}(f)] \quad (7)$$

B. Extension of DWT to Two Dimensions

To extend the wavelet transform to two dimensions it is just necessary to separately filter and down sample in the horizontal and vertical directions. This produces four subbands at each scale. Denoting the horizontal frequency first and then the vertical frequency second, this produces high-high (HH), high-low (HL), low-high (LH) and low-low (LL) image subbands. By recursively applying the same scheme to the low-low subband a multiresolution decomposition can be achieved.

Figure 1(a) shows the normal layout of such a wavelet decomposition. At each scale the subbands are sensitive to frequencies at that scale and the LH, HL and HH subbands are sensitive to vertical, horizontal and diagonal frequencies respectively. Figure 1(b) shows a DWT decomposition of a texture image. This image shows the variation between wavelet subbands highlighting the scale and orientation selectivity of the transform. Also, Figure 1(d) shows a DWT decomposition of the Barbara test image (shown in Figure 1(c)). This image also shows the scale and orientation selectivity of the DWT. High energy subband regions pick out the texture content at different scales and orientations.



1 (a) Labelled Subbands 1(b) Magnitude of DWT of Texture image



1 (c) Barbara Test Image 1 (d) Magnitude of DWT of Barbara Image

Fig: 1 Two Dimensional Wavelet Transform

C. Extension of DWT to Two Dimensions

The oldest, and simplest, example of a mother wavelet is the Haar function, and is composed of a pair of rectangular pulses:

$$\Psi(x) = \begin{cases} 1 & 0 \leq x < \frac{1}{2} \\ -1 & \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

A standard two-dimensional Haar wavelet decomposition of an image is very simple to code. It involves a one-dimensional

decomposition on each row of the image, followed by a one-dimensional decomposition on each column of the result.

3. Region Based Image Retrieval

The Region Based Image Retrieval (RBIR) system uses the Discrete Wavelet Transform (DWT) and a k-means clustering algorithm to segment an image into regions. Each region is represented by means of a set of features and the similarity between regions is measured using a specific metric function on such features. The implementation of RBIR can be divided into two parts: Image Pre-processing and Image Retrieval

A. Image Pre-processing

The detailed algorithm for the pre-processing stage is given below.

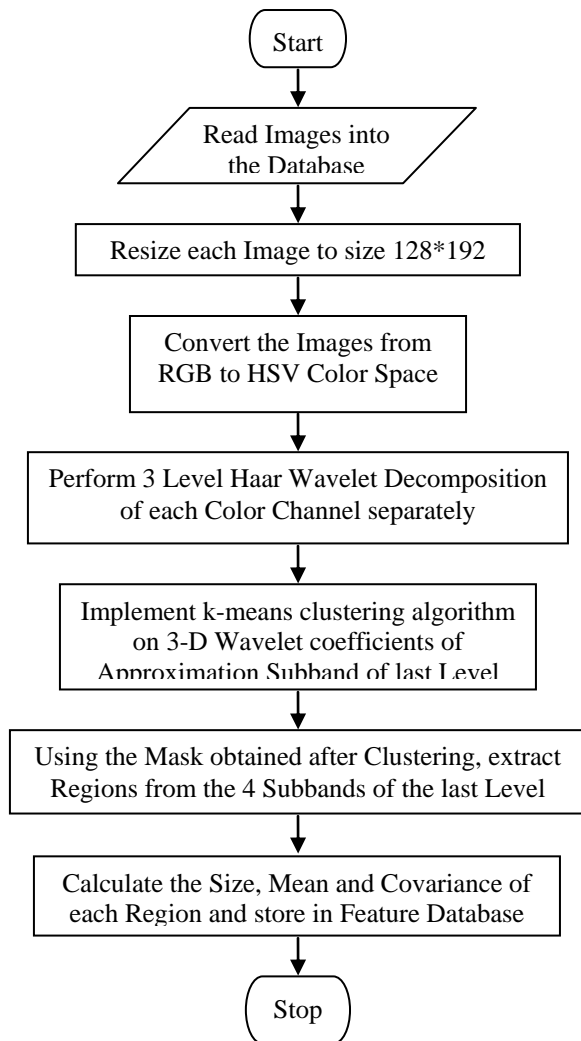


Fig: 2 Flow

The algorithm requires that all the images in the database and the query image be of the same size. A size of 128*192 was chosen for all of the images. If the image had a different size, it was first resized to 128*192 and then the pre-processing operations were carried out on the image.

In the RBIR application, each image is divided into the corresponding color channels (i.e. H, S and V) and the DWT was applied separately to each color channel. The j -th Wavelet

coefficient of subband B ($B \in \{LL, LH, HL, HH\}$, where L stands for “low” and H for “high”) and DWT level l is a 3-D vector i.e.

$$\omega_j^{l:B} = (\omega_{0j}^{l:B}, \omega_{1j}^{l:B}, \omega_{2j}^{l:B}) \quad (9)$$

Where each component refers to a color channel c ($c \in \{0, 1, 2\}$). The energy of $\omega_j^{l:B}$ on the c and d channels is then defined as:

$$e_{cdj}^{l:B} = \omega_{cj}^{l:B} \cdot \omega_{dj}^{l:B} \quad (10)$$

When $c=d$, $e_{cdj}^{l:B}$ is called the channel energy of channel c , whereas when $c \neq d$, $e_{cdj}^{l:B}$ is termed the cross-correlation energy between channels c and d . The energy vector

$$e_j^{l:B} = (e_{00j}^{l:B}, e_{01j}^{l:B}, e_{02j}^{l:B}, e_{11j}^{l:B}, e_{12j}^{l:B}, e_{22j}^{l:B}) \quad (11)$$

captures both color and texture information through channel and cross-correlation energies, respectively. This is known to be one of the most robust methods for the representation of texture features.

K-Means Clustering: The k-means algorithm partitions the observations in the given data into k mutually exclusive clusters, and returns a vector of indices indicating to which of the k clusters it has assigned each observation.

Each cluster or partition is defined by its member objects and by its centroid, or center. The centroid for each cluster is the point to which the sum of distances from all objects in that cluster is minimized.

$$\sqrt{(p_1 - q_1)x^2 + (p_2 - q_2)x^2 + \dots + (p_n - q_n)x^2} = \sum_{i=1}^n (p_i - q_i) \quad (12)$$

Also, when using the Euclidean distance, each centroid is computed as the mean of the points in the cluster. For example, if the data set has three dimensions and the cluster has two points: $X = (x_1, x_2, x_3)$ and $Y = (y_1, y_2, y_3)$, then the centroid Z becomes $Z = (z_1, z_2, z_3)$, where $z_1 = (x_1 + y_1)/2$ and $z_2 = (x_2 + y_2)/2$ and $z_3 = (x_3 + y_3)/2$.

The k-means algorithm is a two-phase iterative algorithm that minimizes the sum of point-to-centroid distances, summed over all k clusters:

- 1) The first phase uses what the literature often describes as “batch” updates, where each iteration consists of reassigning points to their nearest cluster centroid, all at once, followed by recalculation of cluster centroids. This phase provides a fast but potentially only approximate solution as a starting point for the second phase.
- 2) The second phase uses what the literature often describes as “on-line” updates, where points are individually reassigned if doing so will reduce the sum of distances, and cluster centroids are recomputed after each reassignment. Each iteration during this second phase consists of one pass through all the points.

B. Image Retrieval

The detailed algorithm for the Image Retrieval phase is given below.

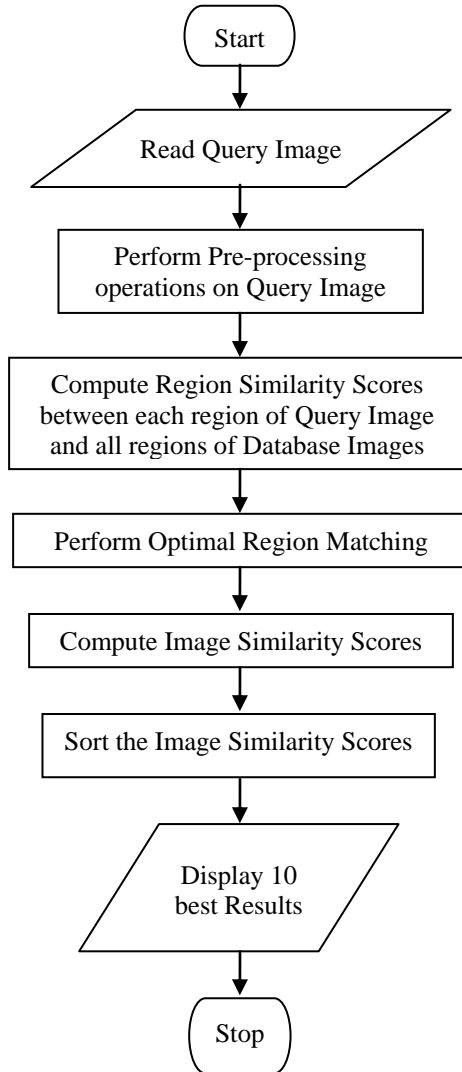


Fig: 3Flow Diagram

After reading the query image, the same pre-processing operations of image resizing, RGB to HSV conversion, DWT decomposition, k-means clustering and feature extraction must be performed on the query image. At the end of the pre-processing operation, the sizes, mean vectors and covariance matrices of the regions of the query image would be obtained.

Region Similarity : The similarity between two regions $R_{q,i}$ (represented by the feature vector $[\mu_{R_{q,i}}, CR_{q,i}3, size(R_{q,i})]$) of a query image I_q and $R_{s,j}$ (represented by the feature vector $[\mu_{R_{s,j}}, CR_{s,j}3, size(R_{s,j})]$) of a database image I_s is computed as

$$r_{sim}(R_{q,i}, R_{s,j}) = d(R_{q,i}, R_{s,j}) \quad (13)$$

where $d()$ is a distance function. The distance $d(R_{q,i}, R_{s,j})$ between the regions $R_{q,i}$ and $R_{s,j}$ is a weighted sum, taken over the four frequency subbands, of the distances between color-texture descriptors, plus an additional term that takes into account the difference between the relative sizes of the two regions.

In the present work, all the frequency coefficients are equally weighed i.e. $\gamma_B = 1$ for $B \in \{LL, LH, HL, HH\}$. The second term takes into account the difference in size between the regions by multiplying it with a coefficient that favors matches between large regions. The distance $d_B(R_{q,i}, R_{s,j})$ between two regions on the frequency subband B is computed by using the Bhattacharyya Metric:

$$d_B(R_{q,i}, R_{s,j})^2 = \frac{1}{2} \ln \left(\frac{\frac{|c_{R_{q,i}}^{3;B} + c_{R_{s,j}}^{3;B}|}{2}}{|c_{R_{q,i}}^{3;B}|^{\frac{1}{2}} * |c_{R_{s,j}}^{3;B}|^{\frac{1}{2}}}} \right) + \frac{1}{8} \left[(\mu_{R_{q,i}}^B - \mu_{R_{s,j}}^B)^T * \left(\frac{c_{R_{q,i}}^{3;B} + c_{R_{s,j}}^{3;B}}{2} \right)^{-1} * (\mu_{R_{q,i}}^B - \mu_{R_{s,j}}^B) \right] \quad (14)$$

4. Implementation of RBIR

After the optimal region assignment has been performed, the next step is to compute the Image Similarity Score. This score is obtained by simply adding the region similarity scores of the matched regions. The final step is to sort the image similarity scores so obtained and then display the images having the least distance from the query image.

Adding Images to the Database: The steps involved in adding images to the database are:

- 1) Run the 'Feature Database Generation Population' program. A Matlab GUI appears as shown in Figure 4
- 2) Select the images from the folder Image_Database and click on 'Add' to generate a feature database.
- 3) The images selected will appear in the listbox.
- 4) Click on the 'Done' after selecting the images to be added to the database.
- 5) Feature Database of the image will be generated.



Fig: 4Feature Database population

Image Retrieval: The steps involved in searching for images are:

- 1) Start the 'Region Based Image Retrieval' program. A GUI appears as shown in Figure 5 (a).
- 2) Select a query image from the folder Image_Query.
- 3) The query image will be displayed Figure 5 (b) along with the 16 matches of the most similar images available in the database.

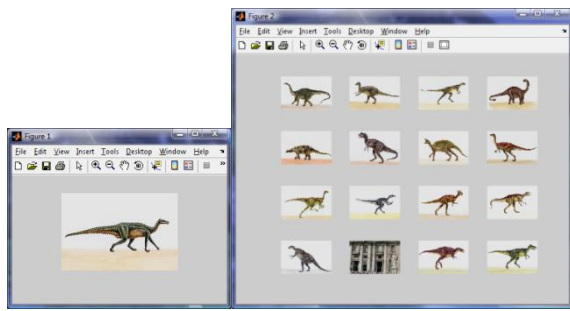


Fig: 5 (a) Fig: 5 (b)

Fig: 5 Region Based Image Retrieval

6. Results

To test the RBIR application, a database consisting of 180 general images, was used. The 180 images could roughly be categorized into 9 groups, each group consisting of 20 similar images. In addition 9 query images, each query corresponding to one of the groups were taken.

Figure 6 shows an example of the results obtained with the RBIR application. From a semantic point of view, the results obtained are particularly good i.e. all the images in this particular example are of horses.

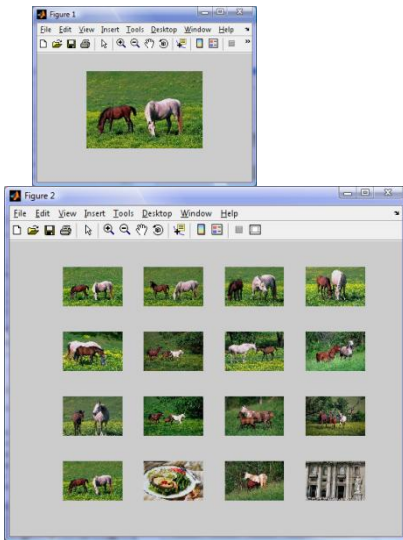


Fig: 6 RBIR Results

a. Partial Match Queries

A partial match query is a query that specifies only part of the image. In Figure 7, the query image is obtained by cropping a database image. As the results show, the RBIR application gave the complete image as the very first match.

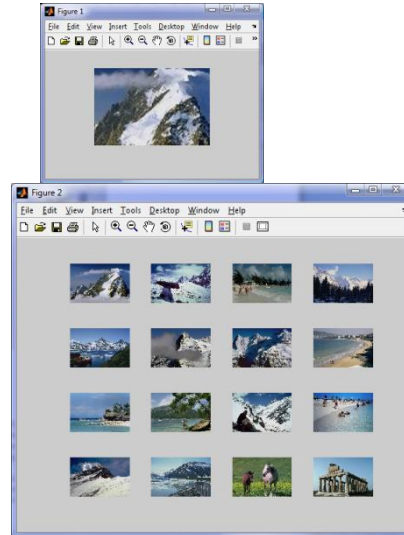


Fig: 7 Partial Match Querying

b. Scanned Queries

If the query image is scanned, it may suffer artifacts such as color shift, poor resolution, dithering effects, and misregistration. To consider the effect of scanned images on the retrieval effectiveness, the query image was first printed and then subsequently scanned. The scanned image appeared fuzzier, darker and slightly misregistered compared to the original. Figures 8 (a) and (b) show the results obtained with the original query image and the scanned query image respectively. It can be observed that there is a slight degradation in the quality of the results obtained.

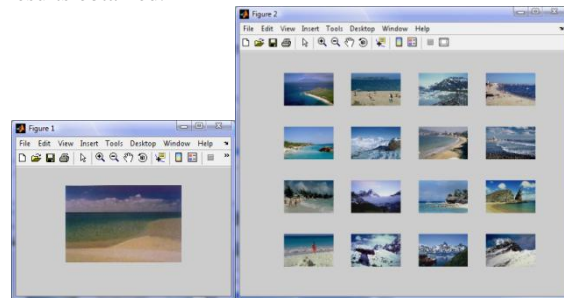


Fig: 8 (a) Result with Original Query

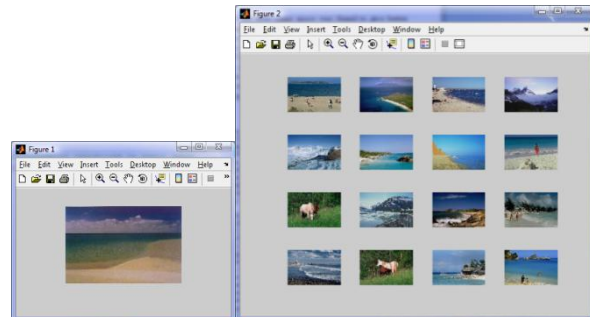


Fig: 8 (b) Result with Scanned Query

Fig: 8 Scanned Queries



c. Difficult Queries

The effectiveness of the RBIR application is confirmed when considering “difficult” queries, i.e. queries having a low number of similar images in the database. Figure 9 shows the results for a query having only two similar images in the database. The RBIR system is able to retrieve both of these images.

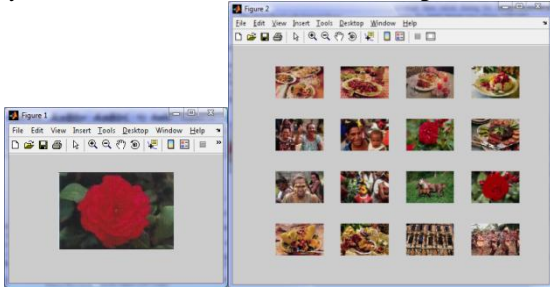


Fig: 9 Difficult Queries

d. Search Time

For an image retrieval application, the time taken for retrieval is an extremely important parameter. On the other hand, the time taken for pre-processing is not as important since the pre-processing operations have to be carried out only once.

The first entry in table below shows the average time taken to perform the pre-processing operations on the database of 180 images of size 128*192. The second entry shows the average time taken during the image retrieval phase. Here again, the query image was of size 128*192.

The experimental setup consisted of a PC with a 3.0 GHz Pentium 4 processor and 512 MB of DDR 400 RAM running MATLAB 7 on Windows XP Professional.

Pre-processing Stage	25.2 seconds
Image Retrieval Stage	1.7 seconds

6. Conclusions

Although the HSV color space was found to give better results compared to the RGB color space in, in our experiments the RGB and HSV color spaces were found to give almost equivalent results. Eventually, it was decided to use the HSV color space because it gave better results than the RGB color space in case of “difficult queries”. The figure below shows that when using the RGB color space, only one of the two matches was retrieved. On the other hand, in the HSV color space, both the matches were retrieved.

7. References

- [1] D. Lowe, “Object recognition from local scale-invariant features,” in ICCV, 1999, pp. 1150–1157.
- [2] Y.J. Zhang “A survey on evaluation methods for image segmentation”, Pattern Recognition 29 (8) (1996) 1335 - 1340
- [3] A. Jain, “Data clustering: 50 years beyond k-means,” Pattern Recognition Letters, vol. 31, no. 8, pp. 651 – 666, June 2010.
- [4] W. Zhao, H. Ma, Q. He, "Parallel K-Means Clustering Based on MapReduce," in: Cloud Computing, vol. 5931, pp. 674-679, 2009.
- [5] W. D. Arthur, S. Vassilvitskii, “K-means++: the Advantages of careful seeding,” in Proc. 2007 Symposium on Discrete Algorithms, pp.1027-1035.
- [6] Rafael C. Gonzalez, Richard E. Woods, " Digital Image Processing" , Second Edition, Prentice Hall Upper Saddle River, New Jersey 07458, TA1632.G66 2001, 698-740
- [7] Fast Multiresolution Image Querying, International Conference on Computer Graphics and Interactive Techniques, 1995: Charles E. Jacobs, Adam Finkelstein, David H. Salesin
- [8] Content-based Image Retrieval, A report to the JISC Technology Applications Programme, 1999: John Eakins, Margaret Graham
- [9] Fundamentals of Content-based Image Retrieval, Multimedia Information Retrieval and Management - Technological Fundamentals and Applications, Springer, 2002: Dr. Fuhui Long, Dr. Hongjiang Zhang, Prof. David Dagan Feng
- [10] Image Retrieval – Current techniques, Promising directions and Open issues, Journal of Visual Communication and Image Representation, 1999: Yong Rui, Thomas S. Huang, Shih-Fu Chang
- [11] Wavelet Based Texture Analysis and Segmentation for Image Retrieval and Fusion, Thesis, University of Bristol, 2002: Paul R. Hill
- [12] WINDSURF: A Region Based Image Retrieval System, Proceedings of the 10th International Workshop on Database & Expert Systems Applications, 2000: Ilaria Bartolini, Paolo Ciaccia, Marco Patella
- [13] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part based models,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, 2010.