

A Review: Relating Low Level Features to High Level Semantics in CBIR

Nikita Upadhyaya¹ and Manish Dixit²

^{1,2}Department of CSE/IT

Madhav Institute of Technology and Science
Gwalior, Madhya Pradesh, India

¹nikita6492@gmail.com, ²dixitmits@gmail.com

Abstract

Content based image retrieval is the technique of effective retrieval of digital pictures from an oversized store of variety of pictures. In CBIR, the focus is mostly on extracting features from the queried image and from the images stored in the database for finding the similarity between these features to retrieve images which are similar visually. CBIR becomes tougher once focus goes to reducing the semantic gap or the linguistics gap between low level features and high level semantics. This survey provides a short summary regarding low level features and high level linguistics that are thought of in CBIR for economical and correct retrieval.

Keywords: CBIR; Low level features; High level semantics; Texture; Color;

1. Introduction

With the advancement of devices for image acquisition, the digital image information is additionally increasing oftentimes. Therefore, to access any explicit image from this large pool of image information, it becomes necessary for user to own an economical image retrieval tool that is correct and processes pictures in less time. Seeing this issue, several researchers worked on this domain to introduce the techniques of economical retrieving systems. There are two approaches followed: text-based and content based. TBIR used annotations that were mapped manually to the pictures and these images were employed by the DBMS for retrieving purpose. In CBIR, content means that description concerning any explicit topic or description of one thing. So, this method follows a process of matching the outline of a specific image with the descriptions of assorted connected pictures to seek out the similar pictures consistent with the user queried image. The working of CBIR framework shown in Figure 1 is as follows:

1. Image Acquisition: Image is acquired using digital devices and image database is created.

2. Image Preprocessing: Images are preprocessed before the retrieval process to enhance the images so that more similar images could be retrieved. It involves enhancement, noise removal, segmentation etc.

3. Feature Extraction: In this feature like texture, color or shape is obtained from images and a feature vector database is created. Features can be broadly classified into low level and high level features. These FV databases are used for measuring the similarity between querying image and database images.

4. Similarity Matching: Similarity matching techniques are used for acquiring percentage of similarity between the queried image and images existing in the database.

5. Output/ Retrieved Images: These are the final outcomes after the whole process is done by matching.

6. User Interaction/Feedback: User can communicate with the system for classifying the images whether they are relevant or not. This process goes on until the user gets satisfied.

CBIR system permits a user for submitting his query in an image form to the system. Now the system computes the feature vectors of the queried image, after that matching of feature vectors of database images is done to search out the similar images as per user's queried image. Content here refers to color, shape, texture or spatial location [1] [2].

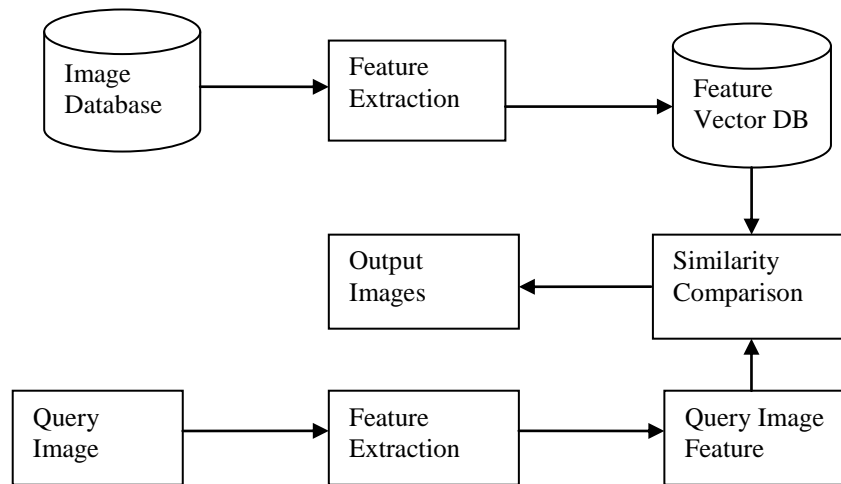


Figure 1. Block Diagram of CBIR

The main interest of researchers is how to minimize the semantic gap and currently a lot of work is going on this subject.

This paper is divided into six sections; Section 2 describes about what semantic gap is. Section 3 and 4 describe about low level features and high level semantics respectively. Section 5 gives a survey about literature. Section 6 describes about performance measures and in the last section paper is concluded.

2. Semantic Gap

Accuracy in CBIR is dependent on the subject on which CBIR is working and the user perspective. According to humans, HL features (concepts) are keywords, text, visual features *etc* by which they interpret images and find similarity between numbers of images. But automatic systems only extract LL features like color, shape, texture, spatial location *etc* from the image. So if we try to see in a broaderview, there's no direct bridge between high level features by which human understand and low level features which are extracted. Many experiments and studies have been done in CBIR but LL features obtained were unable to relate to the HL semantics of human's mind [4, 5].

In Ref [6], the semantics of images are broadly divided into two levels:

1. Local Semantic Level: Local semantic features say about the existence of independent objects or combination of objects in an image. For this purpose, queries formed can be like 'Retrieve images with mountains' (single object) or 'Retrieve images with mountains, rivers and trees' (combination of objects).

2. Thematic or Global Semantic Level: Thematic semantic features refers to the definition/description about the image or topic of the image. It sees the image with a

global perspective. Queries formed could be on a particular topic like 'Retrieve images of flood'. The thematic feature examines all image's objects and how they are related spatially to each other.

In Ref [3], the CBIR queries are categorized into three levels for better understanding:

1. Level 1: Retrieve images based on the LL features like color/shape/texture/spatial location. Typical query would 'be found images like this'. Here, 'this' would be a particular thing.

2.Level 2: Retrieve images having a particular object in it like finding images having bike in it.

3. Level 3: Retrieve images having some higher level logical reasoning or some significance about particular objects which are taken as a query. Typical query would be like retrieving images of riots which has information like an angry crowd of people.

Queries which come under level 2 and 3 are HL semantic queries and the results generated by them are closer to the human understanding having the richness of HL semantics and therefore CBIR system which comes under this category is known as semantic image retrieval. The gap between level 1 and level 2 or 3 is termed as semantic gap [6]. Thus a system of CBIR should work in such a manner to minimize the semantic gap between LL features and HL semantics [4].

3. Low Level Features

CBIR retrieves pictures on the idea of the content sent as question into the system. Here content will be text (keywords/annotations) or visual descriptors (color/shape/texture /spatial location). The most focused part within the framework of CBIR is that the feature extraction. The strength of CBIR depends on the strength of the algorithmic program used for feature extraction. The characteristics of the image which may be visually seen from the feature vectors and once extracting these feature vectors, a feature database is made that is employed for similarity activity [1]. Feature descriptors which are used for extracting feature vectors can be local or global. Global descriptors are those which take image features as a whole and local descriptors are those which consider the features of objects/regions to represent the whole image [1] [7]. LL image features include texture, shape, color and spatial location.

3.1. Color

The color feature is one in every of the essential and wide utilized in CBIR. Color feature is that a part of the image that cannot be modified with regard to orientation and size of the object, therefore it's straightforward to analyze it and extract it. Color descriptor extracts the proportion of color within the question image and is employed for similarity activity with the database pictures having a huge or less similar percentage of color to get output pictures that are similar.

The techniques used to represent the color description of an image are:

1. Color Moments [8][9]

Color moment tells about the distribution of the colors in image. CM are rotation and scaling invariant *i.e.*, color moment values taken out before rotation/scaling of the image are similar to the values taken out after rotation/scaling.

The first order CM is mean which can be explained as an average distribution of color in an image.

$$M_{ij} = \sum_{j=1}^K \frac{1}{K} S_{ij} \quad (1)$$

Where K is the number of pixels in the image and S_{ij} is the value of the j^{th} pixel of the image at i^{th} color channel.

The second order CM is standard deviation which is obtained by calculating the square root of variance.

$$\sigma_{ij} = \sqrt{\left(\frac{1}{K} \sum_{j=1}^K (S_{ij} - M_i)^2\right)} \quad (2)$$

Where M_i is the mean value, or a first color moment, for the i^{th} color channel of the image.

The third order color moment is skewness which calculates the asymmetric color distribution in the image and the shape of the color distribution.

$$SK_i = \sqrt[3]{\left(\frac{1}{K} \sum_{j=1}^K (S_{ij} - M_i)^3\right)} \quad (3)$$

2. Color Correlogram [10, 11]

CC is calculated by combining both pixel's color distribution and spatial correlation of pairs of colors in the image. A table for color correlogram of image can be created by indexing it to color pairs where k^{th} entry for (m,n) tells about the probability of finding a pixel of color n at a distance k from a pixel of color m in the image.

3. Color Histograms [12]

Color histogram elucidates the distribution of composition of color in an image. It depicts various kinds of colors present in the image and the number of pixels present in each color group. There are two types of color histograms global color histogram and local color histogram. In a global color histogram, the whole image is taken into account as a region (doesn't include spatial location of pixels) and this technique is used for matching similar images, whereas in local color histogram, an image is divided into m number of blocks or regions and for every region global color histogram is computed and final output will be the summation of all global color histogram outputs. It is also used in image matching.

4. Color Coherence Vector [13, 14]

Color coherence vector works by combining both color histogram and spatial location information. This vector classifies each pixel as either coherent or incoherent. A Coherent pixel is one when it is a part of a huge similarly colored region. An Incoherent pixel is one when it is a part of small similarly colored region. A CCV stores coherent v/s incoherent pixels for each color range.

3.2. Texture [15]

Texture can be illustrated as the occurrence of visual pattern in an image having the properties of uniformity that are not resulted because of the presence of only single color or intensity in an image. Textural properties are defined for a region in an image or sub image. The two basic properties of texture patterns are firstly its structure and secondly its specification.

1. Structure of texture patterns

A formal definition for texture patterns is difficult to administer. However, if seen from a logical sense of view, a texture may be thought-about as a set of macroscopic regions

and its structure is set by a group of repetitive patterns that have little primitives or parts organized per a special rule referred to as 'placement rule'. Texture can be defined in mathematical way as:

$$S = P(e)$$

Where P is the placement rule or relation and e is the element and S is a function which is a texture and we can get texture if P satisfies the above statement.

Texture can be divided into two classes 'Structural' and 'Statistical'. From the statistical sense of view, texture is seen from a macroscopic view, which is a functional combination of P and e. When seen from structural opinion, more detailed information about P and e is needed.

2. Specification of the texture patterns

To deduce the parameters and their specifications for texture patterns, description regarding the texture properties is needed. So for this some set of features are needed which measures all input patterns and give accurate and well measured results. The six visual texture properties were contrast, coarseness, directionality, regularity, line likeness and roughness.

1. Coarseness

Coarseness is outlined because the means of analyzing the amount and size of texture patterns. The texture pattern is classified into whether or not a pattern is coarse or fine. The coarse texture pattern is that the pattern of enormous size in a small range, whereas the fine texture pattern is of tiny size in a large range.

2. Contrast

Contrast values are the difference between the intensity values of pixels which are adjacent in an image. Contrast can be classified into two classes as whether the image is of high contrast or low contrast. A high contrast image has contrast values greater than texture primitives and a low contrast image has contrast values less than texture primitives.

3. Directionality

Directionality gives the information about the inclination of the texture patterns. It is a global property of the region of the image. It also tells about the shape and how texture primitives are placed.

4. Line Likeness

Line Likeness tells about texture primitives' shape. When the texture primitives cannot be discriminated using directionality then line likeness feature is used to determine whether texture primitive is a straight line or a sort of wave (blob).

5. Regularity

Regularity offers the knowledge concerning the arrangement of the texture primitives. Regularity is when the texture patterns are uniformly organized whereas irregularity is when the texture patterns are haphazardly organized.

3.3.Shape [16]

Shape is one of the essential basic visual features that are used to give data concerning about image content. While analyzing shapes within the image, a representation of an object or shape is done using either shape boundary or shape boundary along with interior content. For accurate retrieving of images, it is required that shape descriptors find similar shapes from the pool of images effectively.

Shape descriptors (representation and description techniques) are categorized into two categories in a broad sense, *i.e.*, contour based and region based methods. The categorization is done on the basis of whether the shape descriptor extracts the shape features using the contour or boundary of the object or take the whole region of the object. Furthermore, each of these two is divided into two classes' structural and global approaches respectively. Global approach deals with the shape as a whole, whereas structural approach deals with the shape as segments/ports.

1. Contour Based Methods

The contour based method considers the information about the boundary of the shape. This method is further divided into two classes, firstly; Continuous (global) approach which does not divide the boundary of shape into sub-parts and a feature vector is formed from the integral boundary to represent the shape. The similarity measurement is done by calculating the metric distance between acquired feature vectors. Second is a discrete approach which is also known as a structural approach because it divides the boundary of shape into sub parts called primitives. These primitives are represented using strings or graphs (tree). The similarity analysis is done by using graph matching or string matching.

2. Region Based Methods

Region based methods consider the information belonging to the interior portion of shape, *i.e.* region along with the exterior portion *i.e.*, boundary. Here also this method can be divided into two categories, structural and globally on the basis of whether the shape is segmented or not respectively.

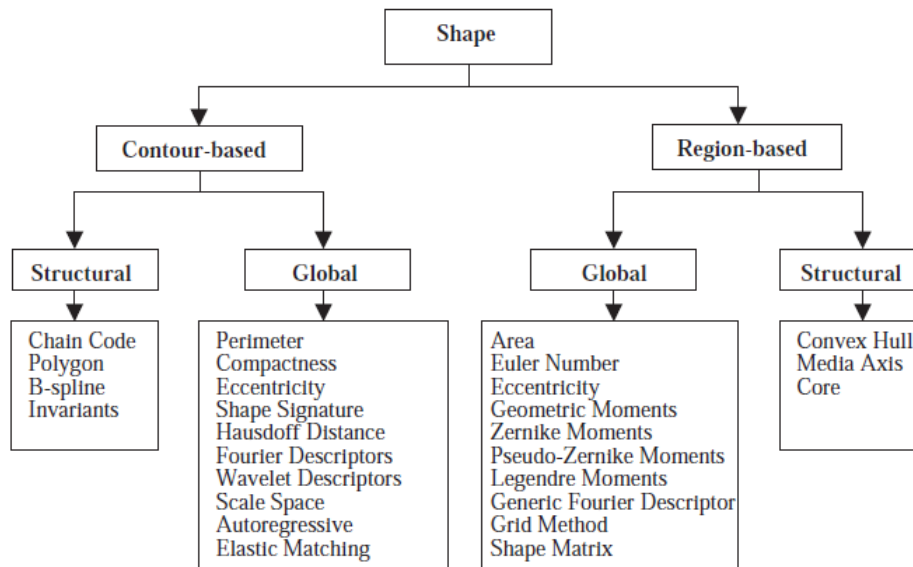
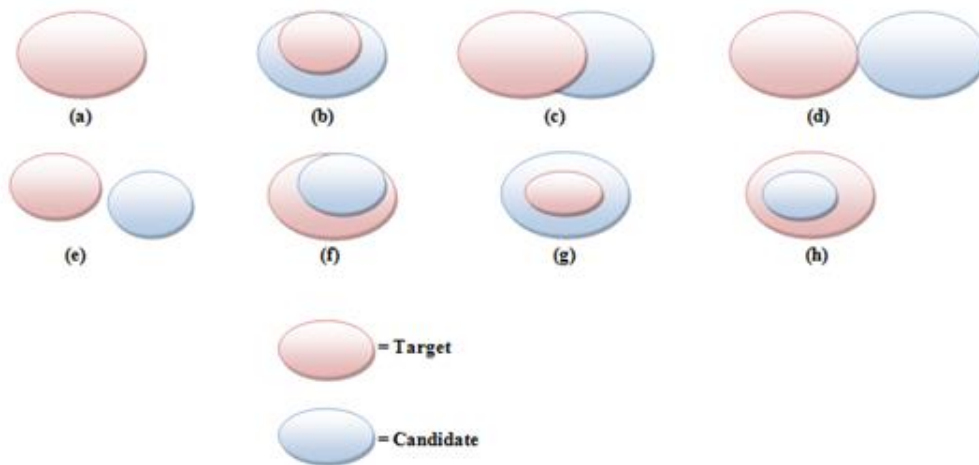


Figure 2. Classification of Shape Representation and Description Techniques

3.4. Spatial Location

In addition to color, texture and shape, spatial location information is also used in region classification. The information which we get using spatial location is as how objects of the scenery are related spatially to each other or spatially placed in the image. Spatial Context modeling [18] is based on the relationship between objects as to how they are placed according to spatial coordinates. Eight spatial relationships are:

- Equal
- Cover
- Overlap
- Touch
- Disjoint
- Covered-by
- Contains
- Inside



(a) Equal (b) Cover (c) Overlap (d) Touch (e) Disjoint (f) Covered-by (g) contains (h) Inside

Figure 3. Eight Spatial Relationships

4. High Level Semantics

Semantics in general terms describe about the denotation of something. In programming, semantics tells about the meaning and format of syntaxes written in the program code, in the same manner when discussed in the field of image processing, semantics tells about the interpretation of images from the user level perspective. In CBIR, LL features are extracted which do not justify the user's perspective and his critical thinking properly. This loophole is called semantic gap as discussed earlier in the starting of paper. So to bridge this gap, researchers proposed methods by which retrieval based on HL semantics was possible.

Now the question is how the relation is established between LL features and HL semantics. These five methods to solve this problem by reducing the semantic gap:

1. Using object ontology to define high-level concepts [19,20]
2. Using machine learning tools to associate low level features with the query concepts [21, 22]
3. Introducing relevance feedback (RF) into the retrieval loop for continuous learning of users' intention [20, 23]
4. Generating semantic template (ST) to support high-level image retrieval [24, 25]
5. Making use of both the visual content of images and the textual information obtained from the Web for WWW (the Web) image retrieval [25-27]

4.1. Object Ontology

In a broader view, ontology is a way by which formal naming, definition and properties of entities can be given along with interrelationships between them. Here entities can be objects of any type for which conceptualization is to be done. If we talk about image processing, here object ontology uses image semantics for defining itself. It defines different levels for assigning LL image features to it. Each level of object ontology defines region attributes which are well understood from users' level. It is accounted as the intermediary level descriptor for image [28]. Suppose an image is acquired of scenery. The sky in, it takes the upper region in it and is of blue color. Then a descriptor can be assigned to the sky's color, whether it is of blue high, blue medium or blue low.

Object Ontology deals with mainly four parameters:

a. Color: In this Lab model is used. L stands for luminance, for a red-green and b for blue-yellow. Luminance can be categorized as very low, low, medium, high and very high. The color components can be partitioned as low, medium, high or none when neither of the two colors is present.

b. Position: Position can be viewed as the region's placement according to spatial coordinates in the image. Positions can be determined with respect to horizontal and vertical axis.

c. Size: Size determines the size of the region, whether it is small, medium or large.

d. Shape: It tells about the shape of the region, whether it is little oblong, medium oblong or very oblong. In this way images can be divided by assigning descriptors to image features and relating it with HL semantics according to human knowledge.

4.2. Machine Learning

This method is acceptable, but is very difficult to implement. Machine learning is partitioned into two categories, *i.e.*, supervised learning and unsupervised learning. In supervised learning, numbers of images are taken and a classifier is used to determine semantic category label. Bayesian classifier is the most common and effective technique used for classification of images into various classes. Most general classes according to which images are classified are indoor or outdoor. Another technique is a neural network in which inputs in the form of images are given and it establishes relationships between LL features and HL semantics. One more method is Decision tree which images are given as training dataset and a tree is constructed and after that, this tree is used for classifying database images (test image dataset) whether they are relevant or not. In unsupervised learning, images are organized into groups and a group label or name is given to each of it. The main concept in this organization is to minimize inter group similarity and maximize intra group similarity so that in retrieval results images from similar groups are retrieved.

4.3. Relevance Feedback

Relevance feedback makes the whole retrieval process interactive by involving the users' opinion in the process. The user enters the query into the system and retrieval results are shown to the user. The user decides whether the output is relevant or not. A machine learning algorithm is used to get the users' feedback and enhance retrieval performance. This process goes on until users' satisfaction is met.

4.4. Semantic Templates

Semantic templates [25] are a set of mappings between HL semantic concepts and LL visual features. It's a function of concept representative determined using sample images.

Semantic template is of significance because it enhances the performance level of retrieving but also recognizes the content which user entered as a query from his perspective.

4.5. WWW Image Retrieval

WWW images have more characteristics than images which are stored in small databases. Web image has properties like its URL, which is a hierarchical structure that can tell about the category of images. Web image retrieval systems like Google or AltaVista uses a text based approach, though the retrieval result contains many similar images, but it is not sure that the resulting images contain user content or not and also user has to go through a long list for searching a particular image which is time consuming. A lot of research is going on in this domain for combining text based and visual based approaches for retrieving images from web [26].

5. Literature Survey

1. In Ref [8], image retrieval based on color is done by using color moment invariants. Rather than being fixed in a given color space, color representations are computed from each image. This technique allows the representation of features to be accurate and compact. The proposed technique image descriptors which are small, and by using a two-stage clustering technique, it makes itself adapted to the context of the image.

2. In Ref [9], the proposed method for CBIR uses both color and texture features. This method encodes a nominal amount of spatial information in the color index for enhancing color indexing technique's discriminating power. By exploiting a color feature, the image is divided into three equal size non overlapping regions. The first three moments of the color distribution are obtained from each region's three color channels and after that indexing is done. Gabor texture descriptors are used for texture features. Weights are assigned to every feature, respectively, and similarity analysis is done. Experiments showed that this method had a high retrieval accuracy than any other conventional methods which uses color moments and texture features.

3. In Ref [10], color feature which has been proposed is called a color correlogram which is new in the domain. This feature includes spatial correlation of colors. It can also describe about the global distribution of colors as to how they are spatially correlated locally. This feature allows the indexing of images and comparing it with each other. The spatial correlation information allows discrimination of images.

4. In Ref [11], the proposed method combines supervised learning techniques with color correlogram which is used for image indexing and retrieval. The learning technique used here is relevance feedback which allows the entry of feedback information from the user back into the system for enhancing the retrieval accuracy by learning the query and metric.

5. In Ref [12], an efficient image similarity matching techniques has been presented. In this, the image is represented as a graph. The steps involved in this method are: firstly image colors are quantized and then histograms are calculated for each color. Now a weighted undirected graph is constructed for each color of the query image and database images. Lastly, minimum cost matching is calculated for each graph and sum is calculated of all colors and considering it as a distance between two images. This technique is robust and invariant to rotation and translation.

6. In Ref [13], a multi CCV mixing location information is proposed as indexing vector. Using single CCV did not give an efficient and robust retrieval result. In this method, an image is partitioned into core region and minor region. Now a hash table is used to store pixel according to color set. Multi CCV mixing location for each region is calculated and the results of both the regions are combined which tells the distance between two images and similarity is measured.

7. In Ref [23], a new method using color coherence region vectors to segment an image is presented. By using CCRV's region characteristics can be described. Also, color histogram is combined with spatial information to give excellent performance not only for region based image retrieval but also for image retrieval.

8. In Ref [17], a method is proposed for 3D object recognition and shapes from images by combining B-splines and neural network. B-splines are used for curve representation and back propagation neural network is used for text/markings recognition. From image curves, object shape can be determined using stereo imaging and type of object can be determined by recognizing or reading text/markings on the objects based on features that are rotation, translation or scaling invariant.

9. In Ref [20], an image retrieval methodology is proposed for searching of images, heterogeneous in nature, in a huge collection. A fully unsupervised segmentation algorithm is used to partition the image into regions. Low level features are extracted from these resulting regions and these features are mapped to intermediate level descriptors. This mapping reduces the semantic gap by ignoring irrelevant images during the retrieving process using these intermediate level descriptors.

10. In Ref [21], a bootstrapping approach for annotating large collection of images. The method starts from a small set of training labeled images and iteratively annotates large set of unlabeled images. To do this, two statistically independent classifiers are used for co-training and co-annotation of unlabeled images. Along with this, a decision model is used to elucidate the concepts they learned from regions extracted using different segmentation method.

11. In Ref [25], an image retrieval system is integrated with semantic template. In the process of retrieving, relevance feedback is used by which semantic templates are generated automatically and a network is created using WordNet. By the help of semantic templates relevant images are returned to the user according to the keyword used to query and even those relevant images are also returned which are not annotated by keyword.

12. In Ref [26], a method is proposed to organize WWW images which come in search result. On the basis of the web context, three representations have been given. Firstly, representation based on visual feature, secondly representation based on textual feature and last representation induced from the image link graph. Using spectral techniques, search results were clustered into different semantic categories. Several images are selected as image representatives according to ImageRanks for each category which helps the user for better understanding of topics of image.

6. Performance Measures

A number of methods have been developed to measure the performance of retrieval systems. Most commonly used performance measures are recall and precision. These are generally represented as precision v/s recall graph. The standard definitions for these two measures are [29]:

Table 1. Contingency Table for Retrieval Results

	Relevant	Non-Relevant	
Retrieved	a	b	a+b
Not Retrieved	c	d	c+d
	a+c	b+d	a+b+c+d

Precision (P): Precision (P) is defined as the ratio of the number of relevant images obtained to the number of total obtained images. Precision tells about the fraction of obtained images that are relevant.

$$P = \frac{a}{a + b} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}}$$

Recall (R): Recall (R) is defined as the number of relevant images obtained over the total number of relevant images available in the database. Recall tells about the fraction of relevant images that are retrieved.

$$R = \frac{a}{a + c} = \frac{\text{Number of relevant images retrieved}}{\text{Number of relevant images in database}}$$

7. Conclusion

This paper gives a comprehensive summary about content based image retrieval, semantic gap, LL features and HL semantics. A literature review of work done in CBIR is also given. CBIR is still an immature technique which needs more enhancements for better retrieval results. Researchers are still working on the main issue of CBIR which is the reduction of semantic gap. Though a lot of work has been done in this domain, but still a generic approach is not yet developed for image retrieval which uses HL semantic parameters. As there is no proper technique available which reduces semantic gap fully, future research directions are suggested.

References

- [1] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics", *Pattern Recognition*, vol. 40, no. 1, (2007), pp. 262-282.
- [2] N. S. Vassilieva, "Content-based image retrieval methods", *Programming and Computer Software*, vol. 35, no. 3, (2009), pp. 158-180.
- [3] J. P. Eakins and M. E. Graham, "Content-based image retrieval, a report to the JISC Technology Applications programme", (1999).
- [4] X. S. Zhou and T. S. Huang, "CBIR: from low-level features to high-level semantics", In *Electronic Imaging*, International Society for Optics and Photonics, (2000), pp. 426-431.
- [5] A. Mojsilovic, and B. Rogowitz, "Capturing image semantics with low-level descriptors", In *Image Processing, 2001. Proceedings. 2001 International Conference on*, IEEE, vol. 1, (2001), pp. 18-21.
- [6] Y. Song, W. Wang, and A. Zhang. "Automatic annotation and retrieval of images." *World Wide Web* 6, no. 2, (2003), pp. 209-231.
- [7] N. S. Vassilieva, "Content-based image retrieval methods", *Programming and Computer Software*, vol. 35, no. 3, (2009), pp. 158-180.
- [8] X. Duanmu, "Image retrieval using color moment invariant", In *Information Technology: New Generations (ITNG)*, 2010 Seventh International Conference, (2010), pp. 200-203.
- [9] Z.-C. Huang, P. PK Chan, W. WY Ng, and D.I S. Yeung, "Content-based image retrieval using color moment and Gabor texture feature." In *Machine Learning and Cybernetics (ICMLC)*, 2010 International Conference, vol. 2, (2010), pp. 719-724.
- [10] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms", *Computer Vision and Pattern Recognition, Proceedings, 1997 IEEE Computer Society Conference*, (1997), pp. 762-768.
- [11] J. Huang, S. Ravi Kumar, and M. Mitra, "Combining supervised learning with color correlograms for content-based image retrieval", *Proceedings of the fifth ACM international conference on Multimedia*, (1997), pp. 325-334.
- [12] V. H. Vu, Q. N. Huu, and H. N. T. Thu, "Content based image retrieval with bin of color histogram." In *Audio, Language and Image Processing (ICALIP)*, (2012), pp. 20-25.
- [13] L. Wen, and G. Tan, "Image Retrieval Using Spatial Multi-color Coherence Vectors Mixing Location Information", In *Computing, Communication, Control, and Management*, 2008. CCCM'08. ISECS International Colloquium on, vol. 1, (2008), pp. 299-302.
- [14] X. Hongli, X. De and G. Yong, "Region-based image retrieval using color coherence region vectors", In *Signal Processing, 2004, Proceedings. ICSP'04. 2004 7th International Conference on*, vol. 1, (2004), pp. 761-764.

- [15] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception", *Systems, Man and Cybernetics, IEEE Transactions*, vol. 8, no. 6 (1978), pp. 460-473.
- [16] D. Zhang, and G. Lu, "Review of shape representation and description techniques", *Pattern recognition*, vol. 37, no. 1, (2004), pp. 1-19.
- [17] J.-Y. Wang and F. S. Cohen, "3D object recognition and shape estimation from image contours using B-splines, unwarping techniques and neural network", *Neural Networks, International Joint Conference*, (1991), pp. 2318-2324.
- [18] M. J Egenhofer and R. D. Franzosa, "Point-set topological spatial relations", *International Journal of Geographical Information System*, vol. 5, no. 2, (1991), pp. 161-174.
- [19] P.Stanchev, D. Green and B. Dimitrov, "High level color similarity retrieval", (2003).
- [20] V. Mezaris, I. Kompatsiaris and M. G. Strintzis, "An ontology approach to object-based image retrieval." In *Image Processing, Proceedings. 2003 International Conference*, vol. 2, (2003), pp. II-511.
- [21] H.M. Feng and T.-S. Chua, "A bootstrapping approach to annotating large image collection", *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, (2003), pp. 55-62.
- [22] R. Shi, H. Feng, T.-S. Chua and C.-H. Lee, "An adaptive image content representation and segmentation approach to automatic image annotation", *Image and Video Retrieval*, (2004), pp. 545-554..
- [23] L. Zhang,, Fuzong Lin, and Bo Zhang. "Support vector machine learning for image retrieval." In *Image Processing, 2001. Proceedings. 2001 International Conference on*, vol. 2, pp. 721-724. IEEE, (2001).
- [24] S. John R. "Decoding image semantics using composite region templates." In *Content-Based Access of Image and Video Libraries, 1998. Proceedings. IEEE Workshop on*, pp. 9-13. IEEE, (1998).
- [25] Z. Yueting, Xiaoming Liu, and Yunhe Pan. "Apply semantic template to support content-based image retrieval." In *Electronic Imaging*, pp. 442-449. International Society for Optics and Photonics, (1999).
- [26] C. Deng, Xiaofei He, Zhiwei Li, Wei-Ying Ma, and Ji-Rong Wen. "Hierarchical clustering of WWW image search results using visual, textual and link information." In *Proceedings of the 12th annual ACM international conference on Multimedia*, pp. 952-959. ACM, (2004).
- [27] F.Huamin, Rui Shi, and Tat-Seng Chua. "A bootstrapping framework for annotating and retrieving WWW images." In *Proceedings of the 12th annual ACM international conference on Multimedia*, pp. 960-967. ACM, (2004).
- [28] K.Anuja, and S. A. Ladke. "Content Based Image Retrieval with Semantic Features using Object Ontology." In *International Journal of Engineering Research and Technology*, vol. 1, no. 4 (June-2012). ESRSA Publications, (2012).
- [29] G. Nidhi, and Priti Sehgal, "A refined hybrid image retrieval system using text and color" *Int. Journal of Comput. Sci* 9, (2012).