

**ABSTRACT**

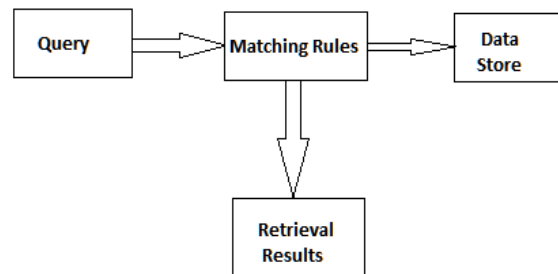
Image retrieval is the basic requirement, task now a day. Content based image retrieval is the popular image retrieval system by which the target image to be retrieved based on the useful features of the given image. CBIR has an active and fast growing research area in both image processing and data mining. In marine ecosystems the captured images having lower resolution, transformation invariant and translation capabilities. Therefore, accurate image extraction according to the user query, relevancy is a challenging task. Thus in this paper we incorporate a detailed survey on Marine images retrieval, in addition of that a new scheme is proposed for enhancing the retrieval capability of existing systems.

**KEYWORDS:** Information Retrieval, IR Models, Techniques of CBIR, Colour, Texture, Shape

**INTRODUCTION**

Information retrieval is the activity of obtaining information resources relevant to an information need for a collection of information resources. Searches can be based on metadata. Many universities and public libraries use Information Retrieval System to provide access to books, journals and other documents. Web search engines are the most visible IR applications. An information retrieval process begins when user enters a query into the system. Queries are formal statements of information needs. In IR query does not uniquely identify a single object in a collection. Instead, several objects may match the query, perhaps with different degrees of relevancy. An object is an entity that is represented by information in a database. Depending on the application the data objects may be, for ex text documents, images, videos, audios and mind maps. The document themselves is not stored or kept directly in the IR system, but are instead represented in the system by metadata. Most IR systems compute a numeric data on how well each object in the database matches the query and rank the objects according to this value. The top ranking objects are then shown to the user. The process may then be iterated if the user wishes to refine the query. Automated information retrieval systems are used to reduce what has been called "Information Overload". Many universities and public libraries use IR systems to provide access to books, journals and other documents. Web search engines are the most visible applications. In order to find the useful information

from the large data source is a complex and much frustrating task. Therefore, efficient and content based systems are required to enhance the traditional way of information retrieval [1].



**Fig 1: Basic model of Information Retrieval System.**

There are many tools that are available on the web for effective search like filtering, retrieve knowledge, retrieve relevant information etc. IR typically seeks to find documents in a given collection that are "about" a given topic or that satisfy a given *information need*. The topic or information need is expressed by a query, generated by the user. Documents that satisfy the given query in the perception of the user are said to be "relevant." Documents that are not about the given topic are said to be "non-relevant." Alternatively, an IR engine may "rank" the documents in a given collection. There are two basic measures for assessing the quality of information retrieval.

1) **Precision:** This is the percentage of the retrieved documents that are relevant to the query.

2) **Recall** -This is the percentage of the documents relevant to the query were retrieved [1].

#### A. Traditional retrieval techniques:

**Full Text Scanning:** The most straightforward way of locating the documents that contain a certain search string (term) is to search all documents for the specified string. And Compare the characters of the search string against the corresponding characters of the document. If The Query is complicated Boolean Expression is used to match the substring [2].

**Signature Files:** Word oriented index-structures based on hashing, which Maps words to a bit mask and to a pointer to the original document. It compresses a document into a 'signature'. The resulting document signatures are stored sequentially in a separate file called signature file, which is much smaller than the original file, and can be searched much faster[2].

**Inversion Files:** A list of sorts words, each associated with a set pointers to the page in which it occurs. Each document can be represented by a list of keywords which describe the contents of the document for retrieval purposes. The keywords are stored eg: alphabetically in the index file. For each keyword we maintain a list of pointers to a qualifying documents Inverted files do better than signature files for most applications. Used in nearly all commercial systems [2].

**Clustering:** The basic idea in clustering is that similar documents are grouped together to form clusters. The cluster hypothesis is that it closely associated documents that tend to be relevant to the user requests. Each document is processed once and is either assigned to one \_or more\_ if overlap is allowed of the existing clusters or it creates a new cluster. The basic reason is that clustering can reveal the intrinsic Structure of a collection, e.g., by topic, subtopic, etc. [2]

#### B. Information retrieval process:

The goal of information retrieval (IR) is to provide users with those documents that will satisfy their information need. An IR model governs how a document and a query are represented and how the relevance of a document for a user query is defined [2].

**Retrieval Models:** For effectively retrieving relevant documents by IR strategies, the documents are typically transformed into a suitable representation. Each retrieval strategy incorporates a specific model for its document representation purposes. A retrieval model specifies the details of:

- (a) Document representation
- (b) Query representation
- (c) Retrieval function

#### C. Classes of Retrieval Models

**Boolean model:** A document is represented as a set of keywords. Queries are Boolean expressions of keywords, connected by AND, OR, and NOT, including the use of brackets to indicate scope. And Output is the weather Document is relevant or not. No partial matches or ranking It is one of the popular retrieval models because: it is easy to understand for simple queries and provides clean formalism. Boolean models can be extended to include ranking. In these reasonably efficient implementations are possible for normal queries [2].

**Statistical Model:** This Retrieval model is based on similarity between query and document Output documents are ranked according to similarity to query. Similarity based on occurrence frequencies of keywords in query and document. And hence Automatic relevance feedback can be supported:

- (a) Relevant documents  
"added" to query.
- (b) Irrelevant documents  
"subtracted" from  
query.

A document is typically represented by a bag of words (unordered words with frequencies).Bag which is set that allows multiple occurrences of the same element. The vector space and probabilistic models are the two major examples of the statistical retrieval approach. Both models use statistical information in the form of term frequencies to determine the relevance of documents with respect to a query [2].

**Vector-space model:** Query and documents are represented as vectors of index terms and Similarity is calculated using COSINE similarity between two vectors Normalization may also be applied [2].

**Probabilistic model:** The probabilistic retrieval model is based on the Probability Ranking Principle, which states that an information retrieval system is supposed to rank the documents based on their probability of relevance to the query, given all the evidence available. The principle takes into account that there is uncertainty in the representation of the information need and the documents [2].

**Language modeling:** In the simplest form of automatic text retrieval, users enter a string of keywords that are used to search the inverted indexes of the document keywords. This approach retrieves documents based solely on the presence or absence of exact single word strings as specified by the logical representation of the query. Clearly this approach will miss many relevant documents because it does not capture the complete or deeper meaning of the user's query [2].

**Content Based Image Retrieval (CBIR):** The importance and need of having Content Based Image Retrieval (CBIR) systems is to take care of growing digital image databases over the vast internet. Content based image retrieval is the popular image retrieval system by which the target image to be retrieved based on the useful features of the given image. Content based image retrieval is the application of computer vision to the image retrieval problem. In this approach instead of being manually annotated with textual keywords, images would be indexed using their own visual contents. The visual contents may be color, texture and shape. This approach is said to be a general framework of image retrieval. The CBIR focuses on the image 'features' to enable the query and have been the recent focus of studies of image databases. The features further can be classified as low-level and high-level features. Users can query example images based on these features such as texture, color, shape, region and others. By similarity comparison the target image from the region repository is retrieved. The color aspect can be achieved by the techniques like averaging and histograms. The texture aspect can be achieved by using transforming or vector quantization. The shape aspect can be achieved by using gradient operators or morphological operators [3].

#### **D. Conclusion of overall**

Information retrieval is the activity of obtaining information resources relevant to an information need for a collection of information resources. The goal of information retrieval is to provide users with those documents that will satisfy their information need. An IR model governs how a document and a query are represented and how the relevance of a document for a user is defined.

Image retrieval is a fast growing and challenging research area with regard to both still and moving images. Many content based image retrieval system prototypes have been proposed, but few are used as commercial systems. CBIR aims at searching image databases for specific image that are similar to a given query image. It also focuses on developing new techniques that support effective searching and browsing of large digital image libraries based on automatically derived imagery features.

#### **RELATED WORK**

Challenges of marine images are low resolution, translation and transformation invariant. **Ahsan Raza Sheikh** [3] implemented a method that is Gradient Vector Flow (GVF); it has been implemented in a lot of image processing applications. Inspired by its fast image restoration algorithms author applied GVF to

marine images. And evaluated different automated segmentation techniques and found GVF showing better retrieval results compared to other automated segmentation techniques.

Image mining is the rising concept which can be used to extract potential information from the general collection of images. Target or close images can be retrieved in a little fast if it is clustered in a right manner. **A. Kannan**[4] combined the concepts of CBIR and Image mining and introduces a new clustering technique in order to increase the speed of the image retrieval system.

CBIR systems perform retrieval based on the similarity defined in terms of extracting features with more objectiveness. But, the features of the query image alone will not be a sufficient constraint for retrieving images. Hence **Dr. V. Mohan** [5] proposed a new technique Color Image Classification and Retrieval using an image technique for improving user interaction with image retrieval systems by fully exploiting the similarity information.

**Ms. K. Arthi** [7] proposed an efficient image retrieval algorithm based on Color Co-occurrence Matrix (CCM). The CCM for each pixel of an image is found using the Hue Saturation Value (HSV) of the pixel and then compared with CCM of the images in the database and the images are retrieved.

**S. Kousalya**[9] done the experiments with the color similarity mining technique of extracting color from specific images and retrieving the similar pixels using the Euclidean distance measure, pixels are grouped on the basis of the nearest neighbor algorithm. This paper evaluates the given images and measures the pixels in image query processing.

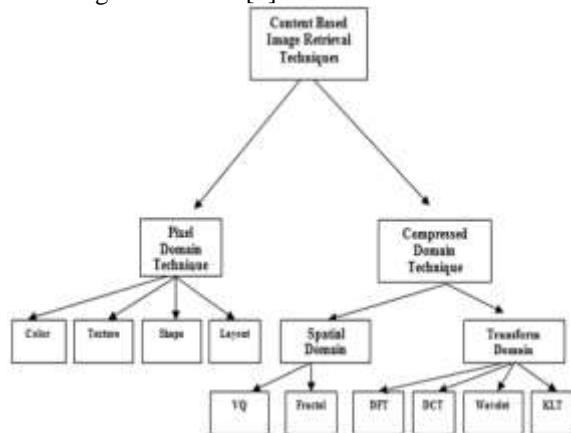
The perception of the Human System of Image is based on the Human Neurons which hold the  $10^{12}$  of information; the Human brain continuously learns with the sensory organs like the eye, which transmits the image to the brain, which interprets the image.

**Rajshree S. Dubey** [10] examines the State-of-art technology. Image mining techniques which are based on the Color Histogram the texture of his image. The query image is taken, then the Color Histogram and Texture is taken and based on this the resultant image is output.

#### **Content Based Image Retrieval Systems**

From a practical point of view, content-based image Retrieval techniques can be divided into two main domains: pixel and compressed domain techniques. Figure exhibits the hierarchy of the techniques. In the pixel domain, the values of individual pixels in the image matrix are used directly for making visual indexes. On the other hand, in the compressed

domain, transformed data, which is the result of mapping the original image matrix into another domain, are employed for feature extraction and retrieval. Different approaches for visual content analysis, representation, and their application indexing and retrieval is brief. Image retrieval techniques, promising directions, and open issues are surveyed. Image coding algorithms will provide the capability of producing image features in compressed domain. Pixel domain and compressed domain techniques are explained with more details in the following subsections [4].



**Fig 2: Classification of Content Based Image Retrieval Techniques**

In content based image retrieval system, the three main features are calculated:

**Retrieval based on Color:** Several methods for retrieving images on the basis of color similarity are being used. Each image is added to the database is analyzed and a color histogram is computed which shows the proportion of pixels of each color within the image. Then this color histogram for each image is stored in the database. During the search time, the user can either specify the desired proportion of each color (75% olive green and 25% red, for example), or submit a reference image from which a color histogram is calculated. The matching process then retrieves those images whose color histograms match those of the query most closely [4].

**Retrieval based on Texture:** The ability to match on texture similarity can often be useful in distinguishing between areas of images with similar color. A variety of techniques have been used for measuring texture similarity in which the best established, rely on comparing values of what is known as second order statistics calculated from query and stored images. From these it is possible to calculate measures of image texture such as the degree of contrast, coarseness, directionality and regularity or periodicity and randomness [4].

**Retrieval based on Shape:** Unlike texture, shape is a fairly well-defined concept and there is considerable evidence those natural objects primarily recognized by their shape. A number of features characteristic of object shape is computed for every object identified within each stored image. Queries are then answered by computing the same set of features for query image, and retrieving those stored images whose features most closely match those of the query. Two main types of shape feature are commonly used global features such as aspect ratio, circularity and moment invariants and local features such as consecutive boundary segments. Shape matching of three-dimensional objects is a more challenging task, particularly where only a single 2-D view of the object in a question available [4].

- i. Using these extracted features color images is properly defined.
- ii. Only color and texture of an image cannot describe the hidden shapes in an image, therefore shape feature is required to embed with the existing CBIR system.
- iii. Image segmentation helps to recognize an image with low computational complexity. In addition of that the BLOB (Binary Large Object) analysis can improve the accuracy of the system [3].

Ahsan Raza Sheikh [3] worked on Marine based images which having the following challenges:

1. Low resolution.
2. Translation.
3. Transformation invariant.
4. Segmentation: not better accuracy, but improve efficiency.

**Manual Segmentation:** To test the performance of the system we applied manual segmentation and compare our automated segmentation results with manually segmented results. Initially in pre-processing, the images are manually segmented to only have region of interest and rest black background so our techniques can only work in the region. Once the images are segmented, features are extracted using the feature extracting followed by distance matrix on them [3].

**Automated segmentation:** performing automated segmentation on marine images had a lot of issues, as there were a lot of information on the marine images and the quality of the images was issued. To overcome a lot of these issues each technique went through a lot of processing to have maximum of region of interest and to have minimum of background. There are two ways in which the automated segmentation was performed: [3]

- i. Saliency Detection.

- ii. Region growing using texture.
- iii. Region growing using color.

When segmentation is applied to images, the one most prominent issue is the distortion or missing region in an image while segmentation. To overcome this, morphology is applied to the images. Morphology removes small, distortion areas and add missing region in the image [3].

- (a) Search only on marine life images, not on other images.
- (b) Using segmentation, color and texture information lost.

### PROPOSED SOLUTION

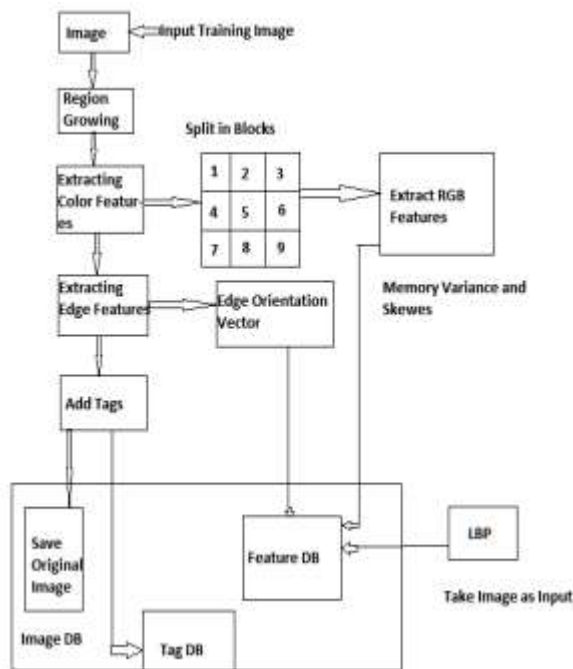


Fig 3: Proposed Training Model

**Region Growing:** Region Growing is a simple region-based image segmentation method. It is also classified as a pixel-based image segmentation method since it involves the selection of initial seed points. This approach to segmentation examines neighboring pixels of initial seed points and determines whether the pixel neighbors should be added to the region.

#### Goals & Objectives:

- (1) Improving search relevancy in terms of precision and recall.
- (2) Use pseudo relevance instead of user feedback relevancy.
- (3) Use normalized feature vector for overcoming the storage overhead.

### CONCLUSION

Information retrieval is the activity of obtaining information resources relevant to an information need for a collection of information resources. Searches can be based on metadata. An information retrieval process begins when the user enters a query into the system. Queries are formal statements of information needs. In IR query does not uniquely identify a single object in a collection. Content based image retrieval is the popular image retrieval system by which the target image to be retrieved based on the useful features of the given image. Content based image retrieval is the application of computer vision to the image retrieval problem.

### FUTURE WORK

After conducting the detailed literature review of existing methodologies of refined image retrieval technique for marine life images are proposed which is in the near future, the proposed method will implements using JAVA technologies and their comparative performance are reported in terms of time and space complexity with their precision and recall of search.

### References

- [1] Available at [informationretrieval.org.en.wikipedia.org/wiki/information-retrieval](http://informationretrieval.org.en.wikipedia.org/wiki/information-retrieval).
- [2] I Shrivastava, "Analysis on Information Retrieval System: A Survey", International Journal of Research in Engineering Technology and Management, Vol. 02(04).
- [3] A. R. Sheikh, S. Mansor, M. H. Lye, and M. F. A. Fauzi: "Content-Based Image Retrieval System for Marine Life Images using Gradient Vector Flow", International Conference on Signal and Image Processing Applications (ICSIPA), IEEE 2013.
- [4] A. Kannan, Dr. V. Mohan, and Dr. N. Anbazhagan, "Image Clustering and Retrieval using Image Mining Techniques", International Conference on Computational Intelligence and Computing Research (ICICR), IEEE 2010.
- [5] Dr. V. Mohan, and A. Kannan, "Color Image Classification and Retrieval using Image Mining Techniques", International Journal of Engineering Science and Technology, Vol. 2(5), pp.1014-1020, 2010.
- [6] J. Chao, A. Al-Nuaimi, G. Schroth and E. Steinbach, "Performance Comparison of Various Feature Detector-Descriptor Combinations for Content-Based Image Retrieval with JPEG-Encoded Query Images", MMSP, IEEE 2013.

- [7] Ms. K. Arthi, and Mr. J. Vijayaraghavan, "Content Based Image Retrieval Algorithm Using Color Models", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2(3), March 2013.
- [8] Mrs. P. Jayaprabha, and Dr. Rm. Somasundaram, "Content Based Image Retrieval Methods Using Self Supporting Retrieval Map Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2(1), January 2012.
- [9] S. Kousalya, and Dr. A S Thanamani, "Image Color Extraction and Retrieval using Classification Techniques", International Journal of Computer Science and Mobile Computing, Vol. 2( 8), August 2013.
- [10] R S. Dubey, N Bhargava, and R Choubey, "Image Mining using Content Based Image Retrieval System", International Journal on Computer Science and Engineering, Vol. 02(07), pp. 2353-2356, 2010.
- [11] P Maheshwari, and Dr. N Srivastava : "Retrieval of Remote Sensing Images using Color & Texture Attribute", International Journal of Computer Science and Information Security, Vol. 4(1 & 2), 2009.
- [12] Dr. C. J Venkateswaran, S. Murugan, and Dr. N. Radhakrishnan, "An Useful Information Extraction Using Image Mining Techniques form Remotely Sensed Image", International Journal on Computer Science and Engineering, Vol. 02(08), 2010.
- [13] C. L Devasena, T. Sumathi, and Dr. M. Hemalata, "An Experimental Survey on Image Mining Tools, Techniques and Applications", International Journal of Computer Science and Engineering, Vol. 3(3) March 2011.
- [14] J-H Su, W-J Huang, P S. Yu, and V S. Tseng, "Efficient Relevance Feedback for Content-Based Image Retrieval by Mining User Navigation Patterns", IEEE Transactions on Knowledge and Data Engineering, Vol. 23(3), March 2011.
- [15] R R Londhe, and V P Pawar, "Analysis of Facial Expression using LBP and Artificial Neural Network", International journal of Computer Applications, Vol. 44(21), April 2012.
- [16] P Rai, and S Singh, "A Survey of Clustering Techniques", International Journal of Computer Applications, Vol. 7(12), October 2010.