

## **Performance Analysis of Image Retrieval Using Web Mining**

**K. Chandrakanth<sup>1</sup>, R. Vijayanandh<sup>2</sup>**

<sup>1,2</sup>*Department of Computer Applications J.J. College of Engineering and Technology, Tamil Nadu, India*

**ABSTRACT :** *The current image retrieval systems are successful in retrieving images, using keyword based approaches. However, they are incapable to retrieve the images which are context sensitive and annotated inappropriately. Content-Based Image Retrieval (CBIR) aims at developing techniques that support effective searching and browsing of large image repositories, based on automatically derived image features. The current CBIR systems suffer from the semantic gap. Though a user feedback is suggested as a remedy to this problem, it often leads to distraction in the search. To overcome these disadvantages, novel interactive keyword based image retrieval and integrating text with image content are proposed to enhance the retrieval accuracy. Also GOOGLE search engine is used as a back end to search and retrieved images with their link. The robustness of the result obtained by the proposed method is shown by various performance analyses like different web browsers, different internet service providers and etc.*

**Keywords:-** *Context sensitive, Annotation, Content-Based Image Retrieval, Google search engine*

### **I. INTRODUCTION**

The interest towards image retrieval is increased due to the rapid growth of the World Wide Web. The need to find a desired image from a collection is shared by many groups, including journalists, engineers, historians, designers, teachers, artists, and advertising agencies. The image needs and usages vary considerably among the users in these groups. The users may require access to the images, based on primitive features, such as color, texture or shape, or associated text. The technology to access these images has also accelerated phenomenally. The current approaches are broad and inter-disciplinary, mainly focused on three aspects of image research which are text-based retrieval, content-based retrieval and interactive based image retrieval. Early techniques are based on the textual annotation of images.

Many techniques have been developed for text-based information retrieval. and they proved to be highly successful for indexing and querying web sites. Their success may also shed some light on the area of image retrieval, because the relatively mature theories and techniques of text-based information retrieval may be applicable to the image domain. Text-based image retrieval uses traditional database techniques to manage images. Through text descriptions, images can be organized by topical or semantic hierarchies to facilitate easy navigation and browsing based on standard Boolean queries. Although text-based methods are fast and reliable when images are well annotated, they are incapable of searching in unannotated image collections. The generalization of the information retrieval from the text domain to the image database is, however, non-trivial. The greatest obstacle arises from the intrinsic difference between the text and image in representing and expressing information.

It is known that the search engines can extract text documents effectively and however, they are very poor in extracting multimedia documents such as image, video and audio. Say for example, in image retrieval applications, the query can be either in the form of image or text keywords. In the text based approach, the images are manually annotated by text evidence and indexed effectively to perform image retrieval. However, these types of systems have two major drawbacks in annotating the keywords [1, 2]. The first drawback is that the major level of human intervention is necessary for manual annotation. The second one is that the inaccuracy annotation due to the subjectivity of human perception. However, this is the only available information from which the properties of the images are extracted. To overcome these drawbacks in text-based retrieval system, content based image retrieval (CBIR) has been introduced [3, 4]. There are various attempts has been proposed to use image contents as the basis to retrieve images and index [3, 5] words. In the text based approach, the images are manually annotated by text evidence and indexed effectively to perform image retrieval. However, these types of systems have two major drawbacks in annotating the keywords. The first drawback is that the

major level of human intervention is necessary for manual annotation. The second one is that the inaccuracy annotation due to the subjectivity of human perception. However, this is the only available information from which the properties of the images are extracted. To overcome these drawbacks in text-based retrieval system, content based image retrieval (CBIR) has been introduced. There are various attempts has been proposed to use image contents as the basis to retrieve images and index. The main problem in current content based approaches focused only on color, texture, shape features such as geometric and statistical to build image contents and performance of the images are not focused.

As these features are low-level, they are effective only in matching images almost identical in contents, and also retrieve the performance of the images with the Timing. In our proposed approach, maximum use of HTML document content structures, including: image TAGs and its related attributes are used. This paper proposes a faster image retrieval mechanism, which is tested on a large number of documents. For this purpose, the documents are fetched, using a web crawler. The content of the documents is segregated into text and images. From the text, keywords are extracted and these keywords are considered to be the relevant. In the following sections of the paper, the related works are presented. In section 3, the proposed method is elaborated along with a suitable example. The experimental result is presented in section 4. The conclusion is given in the last section of the paper.

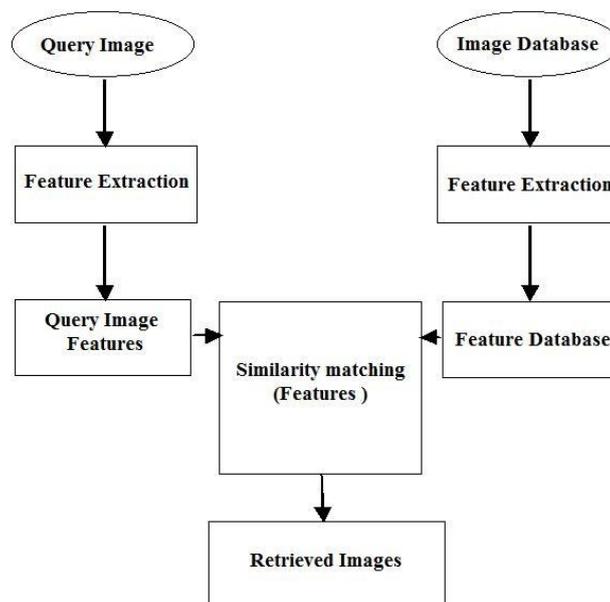
## **II. LITERATURE REVIEW**

Most of the existing information retrieval systems are text-based and use the textual information for retrieving images. However images frequently have little or no accompanying textual information for complementing indexing and semantic capturing approaches. The text-based ontology and classification schemes have been developed for image description. Indexing based on text is considered as having ability to represent general and specific instantiations of an object. The retrieval of images from the Web documents has received a lot of attention recently and there are various approaches extended especially for text-based information retrieval techniques for retrieving images from WWW [2, 6, 7, 8, 9, 10, 11].

Based on the above discussion, it is noticed that the information available in HTML documents should be used effectively. In addition, the visual features of images may not be suitable for generic domain say for images in WWW and suitable approach is required for retrieving images from WWW without low-level features. The research on web image retrieval focused with developing techniques to make use of information obtained from HTML TAGs especially TAGs related images. In this paper, the text present in the <img src> TAG is analysed for retrieving relevant images from WWW. Each attribute in the TAG is carefully considered and categorized into four levels. A suitable weight is assigned to each level such that importance of each level is considered and accordingly the weight is assigned.

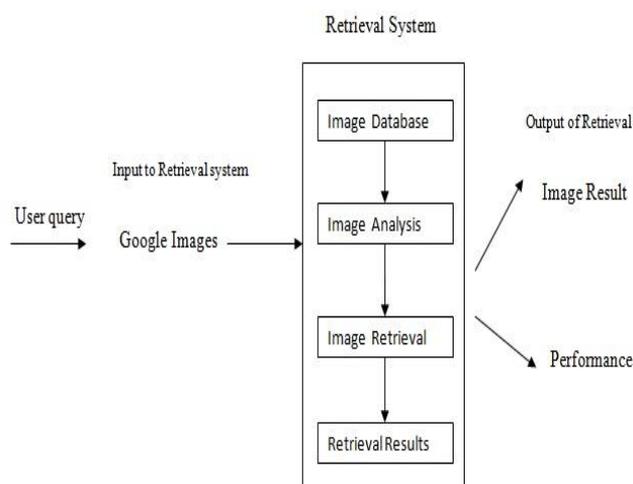
## **III. PROPOSED WORK**

In this work, the importance of the text or attributes present inside the image TAGs present in HTML document is considered and analysed. In this Work, We analyze a image based on their performance with the different browsers and the different service providers. The proposed image retrieval system is based on decomposition of database images using online for keyword based image. Fig. 1 shows the step-by-step process of the proposed method.



**Fig. 1 Step-by-step process of the proposed method**

The Fig. 2 shows how the keyword based image retrieval give their performance.



**Fig. 2 How the keyword based image retrieval give their performance**

The highly efficient image matching takes place using Google Search Engine. The use of progressive retrieval strategy is to provide balance between computational complexity and retrieval accuracy. During the image retrieval process decomposition, feature extraction and Performance are placed and give a clear and accurate result of an image.

#### **IV. EXPERIMENTAL RESULTS**

In this paper, the performance of the proposed approach is evaluated using Google Image dataset and it is available at <http://www.sGoogle.com/images/>. It is created from English Google Images downloaded from static Google.com/Images. Each page contains Category, Portal, Template, User, Image or Wikipedia in the URL are removed from the snapshot, as well as redirect pages. The dataset consist of many images with their

links and the raw file is about 524 MB. The present work is compared with the work by Xue et al, 2007. Though the work has been done 2007, we have not found to the best our knowledge any similar work in recent time by considering <IMG SRC> TAG for retrieving images from WWW. The query set is constructed by randomly selecting 150 keywords for measuring retrieval accuracy. Both the proposed and other work are executed on these query set retrieved top-100 results for comparison. All the experiments are conducted in computer system with Intel(R) Xeon(R) CPU @ 2.40 GHz with 12.0GB RAM configuration and have measured the recall for measuring the efficiency of the proposed approach.

In fact, 30% of the documents are again cross checked for finally updating the ground truth information. Based on the ground truth, we have evaluated the performance using three well-known performance metrics such as Service Providers, Browsers and Providers and Browsers. Service Providers is defined as the where the ground Performance value for relevant documents is known. In this we measure a performance by identifying the top-70 results for every query and taken the average for comparing the corresponding top-70 results. we have shown the average Performance for various Service Providers. Average value of the proposed and title extraction approaches are 96% and 35% respectively for the top-10 results.

To calculate the performance of the Various Browsers by verifying with the various Browsers such as Internet Explorer, Chrome, Firefox. For each query, the performance is calculated and the average is calculated for the Timing. For instance, the average performance value for the top-10 and top-70 pages is 21.7% and 96.3% respectively.

To calculate the Overall performance of different Browsers and Different Service providers. It is the combination of both the Browsers and Service providers and it is taken as a weighted average of both Browsers and service providers value. While the best value is 0 to 0.5 seconds and also present in milli seconds to retrieve the images. In our proposed approach, the value is s at top-10 and top-70 is 36% and 76% and 10% and 4% for title extraction method. It is observed that the performance enhancement of the proposed approach in terms of Performance score is around 400 % for top 10 nearest neighbor. This trend can also be noticed for other nearest neighbors. The overall performance is nearly 80-90% accurate manner to retrieve the images.

Table 1 represents the performance in seconds of the proposed method in retrieving the images using different browsers with different providers.

Fig. 3 and Fig.4 shows the performance in seconds of the proposed method in retrieving the images using different browsers with different providers and sample result of the proposed method respectively.

**TABLE I Performance-On Retrieving The Images**

Providers	Browsers	Performance (in sec)
<b>Vodafone</b>	IE	0.001
	Firefox	0.015
	Chrome	0.026
<b>Aircel</b>	IE	0.007
	Firefox	0.015
	Chrome	0.01
<b>Airtel</b>	IE	0.02
	Firefox	0.025
	Chrome	0.012
<b>Bsnl</b>	IE	0.035
	Firefox	0.012
	Chrome	0.006
<b>Docomo</b>	IE	0.01
	Firefox	0.025
	Chrome	0.5
<b>Reliance</b>	IE	0.015
	Firefox	0.2
	Chrome	0.007

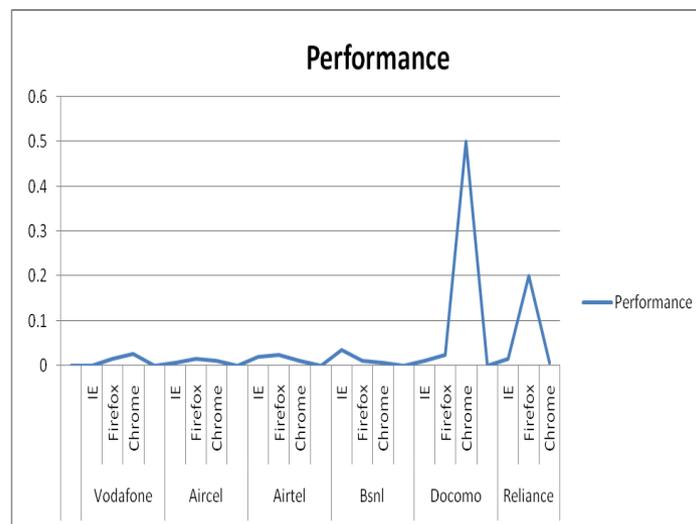


Fig. 3 Performance on retrieving the images

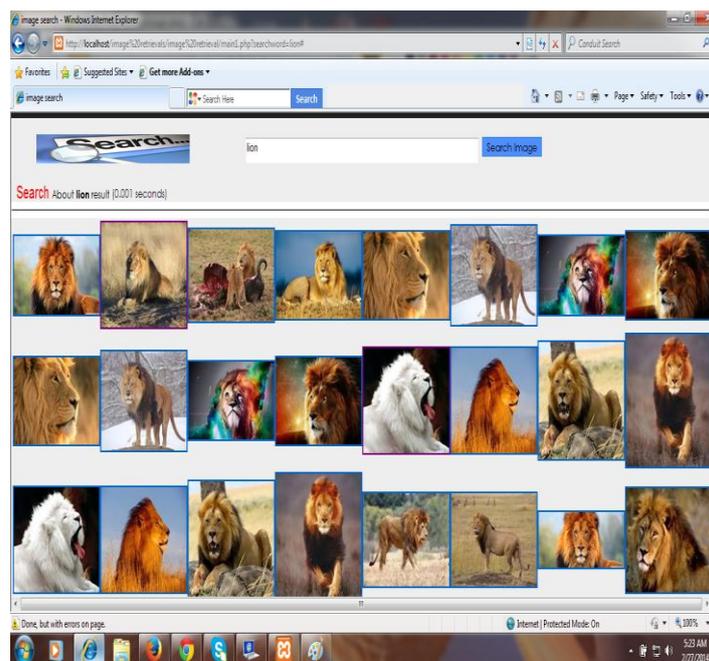


Fig. 4 Sample result of the proposed method

## V. CONCLUSION

The proposed method is to improve the Content Based Image Retrieval via the Keyword structure. Usually, retrieval system returns a large number of images for a query. It is found to be tedious to display the most relevant images in top results. This kind of retrieval is done by most popular retrieval system using various performances. Here images were retrieved using the keyword and also show the performance of the images via the different Service providers and Browsers. Based on the importance of each Providers and Browsers, Results for performance is nearly 80-90% will clearly provide the images with in nearly 0.5 seconds and their improvements are quite promising: the average time consumed by a query is reduced by 66% and close to the In future, use the retrieved patterns to group the web pages and image for efficiently capturing the relevant web pages and images.

## REFERENCES

- [1] S Edward Chang., Kingshy Goh., Gerard Sychay., Gang Wu., CBSA: Content-based soft annotation for multimodal image retrieval using Bayes Point Machines, *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Conceptual and Dynamical Aspects of Multimedia Content Description*, 13, 2003, p. 26.
- [2] Feng, H., Shi, R., chua, T.S., “ A Bootstrapping Framework For annotating and retrieving WWW images,” In *Proceedings of the ACM International Conference on Multimedia*, 2004.
- [3] Liu, Y., Zhang, D.S., Feng, D.D., A survey of content-based image retrieval with high-level semantics, *Pattern Recognition*, 40 (1), 2007, p. 262-282.
- [4] Smith, J.R., Chang, S.F., “Visual Seek: A Fully automated content-based query system”, *ACM Multimedia*, 1996, p. 87.
- [5] Hauptman, A., Baron, R.V., Chen, M. Y., Christel, M., Duygulu, P., Huang, C., Jin, R., Lin, W. H., Ng, T., Moraveji, N., Papernick, N., Snoek, C. G. M., Tzanetakis, G., Yang, J., Wactlar, H. D., 2003. *Informedia at TRECVID: analyzing and searching broadcast news* <http://www-nlpir.nist.gov/projects/tv.pubs.org>.
- [6] Hu, W., Wu, O., Chen, Z., Fu, Z., Maybank, S., Recognition of Pornographic WebPages by Classifying Texts and Images, *IEEE Transactions on Pattern Analysis And Machine Intelligence* 29(6), 2007, p.1019.
- [7] Barnard. K., Forsyth, D. A., Learning the semantics of words and pictures, *IEEE International Conference on Computer Vision* 2, 2001, p. 408-415.
- [8] Yang, H. C., Lee, C. H., Image Semantics Discovery from Web Pages for Semantic-based Image Retrieval using Self-organizing maps, *Expert Systems with Applications* 34, 2008, p. 266 - 279.
- [9] Vadivel, A., Sural, S., Majumdar, A. K., Robust Histogram Generation from the HSV Color Space based on Visual Perception, *International Journal on Signals and Imaging Systems Engineering* 1(3/4), 2008, p. 245.
- [10] Vadivel, A., Sural, S., Majumdar, A. K., Image Retrieval from Web using Multiple Features, *Online Information Review, Emerald* 33(6), 2009, p. 1169.
- [11] Yanai, K., “Generic image classification using visual knowledge on the web,” In *Proc. On ACM Multimedia*, Berkeley, USA, 2003, p. 167