Region-based image retrieval with high-level semantics – A Comprehensive Investigation

Mrs Savita S. Sheelavant, Dr Andhe Dharani

***Abstract*— In this era of information highly accurate data is crucial for all requirements. The current investigation helps in achieving highly accurate image retrieval as close to human interpretation. This paper attempts to provide a comprehensive review and characterize the problem of the semantic gap that is the key problem of region-based image retrieval and the current attempts in high-level semantic-based image retrieval being made to bridge it. In this paper, the latest contributions in research on different methods of image retrieval systems are described and major categories of the state-of-the-art techniques in narrowing down the 'semantic gap' are presented. Finally, based on existing technologies and the demand from real-world applications, a few promising future research directions are suggested.**

Index Terms— **Image databases, image segmentation, ontology, relevance feedback, Semantic region, semantic template, Support Vector Machine (SVM), Binary Decision Tree (BDT), Region Based Image Retrieval (RBIR), semantic learning, query image, foreground region, Artificial Neural Networks (ANN)**

I. Introduction

Advances in data storage and image acquisition technologies have enabled the creation of large image datasets. In this scenario, it is necessary to develop appropriate information systems to efficiently manage these collections. Conventional content-based image retrieval (CBIR)[1] systems index images by their own visual contents such as color, texture and shape. The CBIR technology has been used in several applications such as fingerprint identification, biodiversity information systems, digital libraries, crime prevention, medicine, historical research, among others. The region-based image retrieval (RBIR) systems extract the images based on region of interest. During the past decade, remarkable progress has been made in both theoretical research and system development. However, there remain many challenging research problems that continue to attract researchers from multiple disciplines. Not many techniques are available to deal with the semantic gap presented in images and their textual descriptions.

Manuscript received September 04, 2013.
Mrs Savita S. Sheelavant, Asst prof Dept of MCA RV College of Engineering Bangalore
Dr Andhe Dharani, Professor, Dept of MCA RV College of Engineering Bangalore

A. *High Level features:*

Low level image features can be related with the high level semantic features for narrowing down the gap of image semantics.

Humans tend to use high-level features (concepts), such as keywords, text descriptors to interpret images and measure their similarity. While the features automatically extracted using computer vision techniques are mostly low-level features (cooler, texture, shape, spatial layout, etc)[2]. In general, there is no direct link between the high-level concepts and the low-level features. Though many sophisticated algorithms have been designed to describe color, shape, and texture features, these algorithms cannot adequately model image semantics and have many limitations when dealing with broad content image databases. Extensive experiments on RBIR systems show that low-level contents often fail to describe the high level semantic concepts in user's mind, 'Semantic gap'. Therefore, to further improve retrieval accuracy, a RBIR system should reduce the 'semantic gap' between low-level image features and human semantics [8,9]. Another advantage of semantic-based image retrieval is that it supports query by keywords or textual descriptions which is more convenient for users.

B. *Semantic gap:*

In Information Retrieval (IR), the semantic gap is the difference between what computers store and what users expect via their queries. There are several reasons for the existence of those gaps such as homonymy and synonymy in text retrieval, or the typical difference between low-level representations and keyword-based queries in image retrieval.

Techniques for reducing the 'semantic gap[18]' can be roughly classified into five categories.

(1)Using machine learning tools to associate low-level image features with high-level semantics For example, Fei-Fei et al. developed an incremental Bayesian algorithm to learn generative models of object categories and tested it on images of 101 widely diverse categories.

(2) Introducing relevance feedback (RF) into retrieval loop for continuous learning of users' intention. Considering the interaction with the details in an image (such as points, lines and regions), Nguyen and Worring proposed a framework to dynamically update the user- and context-dependent definition of saliency based on RF.

(3) Exploring domain knowledge to define ontology for image annotation. For instance, Guus et al. designed an annotation strategy to search photograph collections using the background knowledge contained in ontology.

(4) Making use of multiple information sources such as the textual information obtained from the Web and the visual content of images for Web image retrieval.

(5) Generating semantic templates (STs) to support semantic-based image retrieval. Chang et al. introduced the idea of semantic visual template (SVT) to link low-level image feature to high-level concepts for video retrieval [22][23]. Many systems exploit one or more of the above techniques to implement high-level semantic-based image retrieval.

II. Methods used in RBIR with high-level semantics

Here the various approaches and algorithms or methodologies for retrieving images with high-level semantics are discussed.

A. Region-based image retrieval with high-level semantics using object ontology and Relevance Feedback.

In this approach, ontology is employed to allow the user to query an image collection using semantically meaningful concepts (semantic objects), as in [42]. Simple object ontology is used to enable the user to describe semantic objects, like "tiger," and relations between semantic objects, using a set of intermediate-level descriptors and relation identifiers. The architecture of this indexing scheme is illustrated in Figure 1.

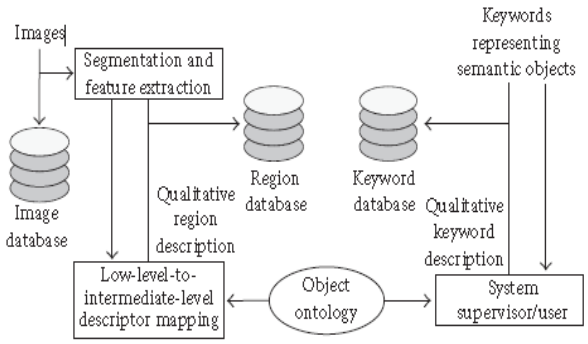

Figure1: Indexing system overview: low-level and intermediate-level descriptor values for the regions are stored in the region database; intermediate-level descriptor values for the user-defined keywords (semantic objects) are stored in the keyword database.

The simplicity of the employed object ontology serves the purpose of it being applicable to generic image collections without requiring the correspondence between image regions and relevant identifiers be defined manually. The object ontology can be expanded so as to include additional descriptors and relational identifiers corresponding either to slow-level region properties (e.g., texture) or to higher-level semantics which, in domain-specific applications, could be inferred either from the visual information itself or from associated information (e.g., text).

A query is formulated using the object ontology [5][24] to provide a qualitative definition of the sought object or objects (using the intermediate-level descriptors) and the relations between them [9]. As soon as a query is formulated, the intermediate-level descriptor values associated with each desired object/keyword are compared to those of each image region contained in the database.

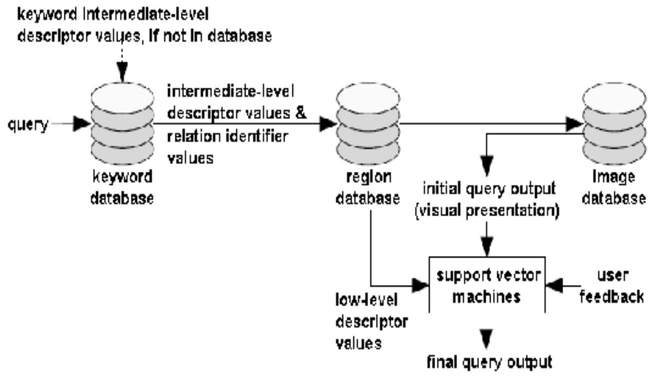

Figure 2: Query process overview

A sample query and how that query is processed and results are shown in figure 3.

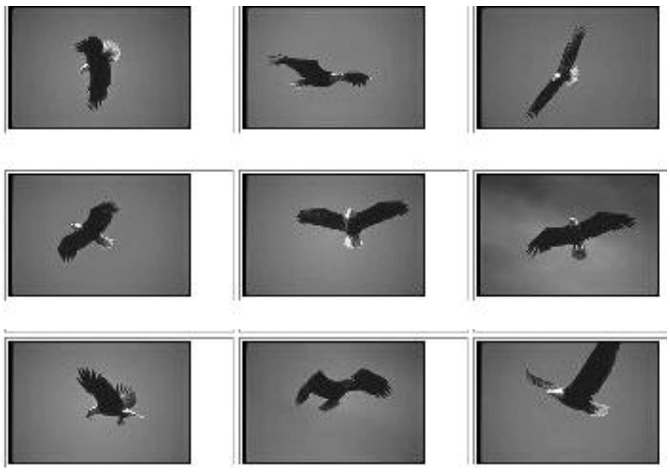

Figure 3: Results for single-object queries of bald eagle and blue sky.

B. RBIR system with high-level semantics derived using Decision Tree (DT) learning.

Every image in the database is segmented into different regions, represented by their color and texture features. The DT induction process is based on the concept of top-down induction of DTs. For image feature discretization, a set of (Semantic Template) STs is generated for the concepts defined in our database. A ST is the representative feature of a concept and is calculated from the low-level features of a collection of sample regions. DT-ST converts low-level color/texture features into color/texture labels, thus avoiding the difficult image feature discretization problem. For tree simplification, DT-ST[14] employs a hybrid of pre-pruning and post-pruning techniques in order to resolve the noise and tree fragmentation problems. As a result, the tree grows in a well-controlled manner and the classification performance is improved. Based on the decision rules derived by DT-ST, each region in an image is associated with a high-level concept.

This system supports both query by specified region and query by keyword. For query by region, it's assumed that every image contains a dominant region that represents the semantic concept of the image. When the user submits a query region, the system obtains the query concept using DT-ST and then returns those images that contain region(s) of the same concept as that of the query. In the case of query by keyword, the system will return images with region(s) matching the query concept. Experimental results demonstrate that this system significantly improves the accuracy of image retrieval, compared with a system without using high-level semantics as shown in figure 4.

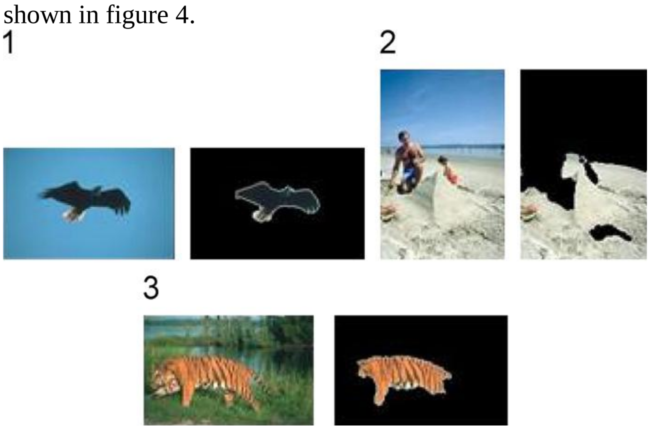

Figure 4: Examples of query image and the dominant region

C. Region Based Image Retrieval by Extracting the Dominant Region and Semantic Learning

The Semantic Region Based Image Retrieval (SRBIR)[6] system automatically segments the dominant foreground region, consisting of the semantic concept of the images. The system segments an image into different regions and finds the dominant foreground region in it, which is the semantic concept of that image. Then it extracts the low-level features of that dominant foreground region. The Support Vector Machine-Binary Decision Tree (SVM-BDT) is used for semantic learning and it finds the semantic category of an image. The low level features of the dominant region of each category image are used to find the semantic template of that category. The SVM-BDT is constructed with the help of these semantic templates. The high level concept of the query image is obtained using this SVM-BDT. Similarity matching is done between the query image and the set of images belonging to the semantic category of the query image and the top images with least distances are retrieved.

Algorithm for extracting the dominant foreground region of an image:

1. An RGB image is read and the indexed image is obtained from it. The indexed image is used to get back the color from the corresponding gray scale image
2. The gray scale image is obtained from the color image
3. Noise is removed by applying median filtering
4. The edges of the image are found by using 'canny edge detection'.
5. Smoothing of the image is done to reduce the number of connected components
6. Find the connected components of the image
7. The component number for the background image is 0. Among all the connected components excluding the background component, the biggest connected component in the image is found
8. For the pixels that are in the maximum connected component, the original pixel value from the indexed image is copied and for all the remaining pixels the value is set to zero. This biggest connected component is treated as the dominant region
9. Make the dominant region obtained as a solid region
10. Now the solid region is converted back into a color image, using the color mapping.

Samples of extracted dominant region from images are listed in figure 5.

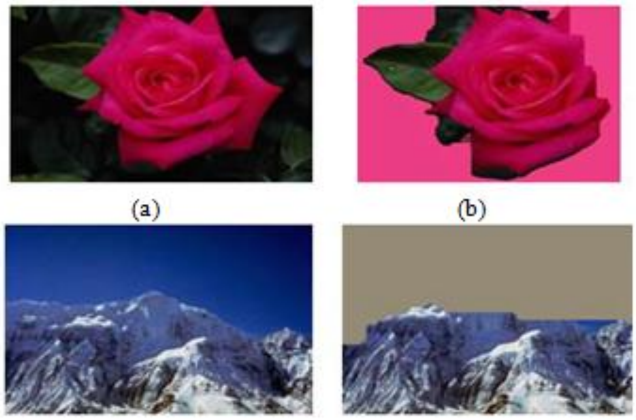

Figure 5: Image and dominant foreground region

III. FUTURE WORK:

1. The region based image retrieval with high level semantic features can be extracted from the satellite images with sensor networks.
2. Combination of eye tracker with high level semantic features:

Eye tracker:

Eye tracking is a technique whereby an individual's eye movements are measured so that the researcher knows both where a person is looking at any given time and the sequence in which their eyes are shifting from one location to another. Tracking people's eye movements can help HCI(Human Computer Interaction) researchers understand visual and display-based information processing and the factors that may impact upon the usability of system interfaces. In this way, eye-movement recordings can provide an objective source of interface-evaluation data that can inform the design of improved interfaces. Eye movements can also be captured and used as control signals to enable people to interact with interfaces directly without the need for mouse or keyboard input, which can be a major advantage for certain populations of users such as disabled individuals. We begin this chapter with an overview of eye-tacking technology, and progress toward a detailed discussion of the use of eye tracking in HCI and usability research. A key element of this discussion is to provide a practical guide to inform researchers of the various eye-movement measures that can be taken, and the way in which these metrics can address questions about system usability. We conclude by considering the future prospects for eye-tracking research in HCI and usability testing. Eye-tracking systems available today measure point-of-regard by the "corneal-reflection/pupil-center" method.

The different measurements used in eye-tracking research are fixations (described previously) and "saccades", which are quick eye movements occurring between fixations [16][17]. There are also a multitude of derived metrics that stem from these basic measures, including "gaze" and "scan path" measurements Pupil size and blink rate are also studied. All these measurements and their descriptions are listed in table 1.

By making use of these eye tracker metrics we can narrow down the 'semantic gap' between low-level image features and human semantics.

IV. CONCLUSION:

The semantic region based image retrieval looks for high-level features which are close to the human interpretation of images. Here our investigation discusses some of the methodologies used for region-based image retrieval and how high-level features are used to bridge the gap of human perception. The future work suggested with eye tracker to extract high-level features helps in retrieving the more accurate information.

Table 1: List of Eye tracking metrics

Eye-Movement Metric	**What it Measures**
Fixations	Fixations can be interpreted quite differently depending on the context. In an encoding task (e.g., browsing a web page), higher *fixation frequency* on a particular area can be indicative of greater interest in the target, such as a photograph in a news report, or it can be a sign that the target is complex in some way and more difficult to encode.
Saccades	No encoding takes place during saccades, so they cannot tell us anything about the complexity or salience of an object in the interface. However, regressive saccades (i.e., backtracking eye-movements) *can* act as a measure of processing difficulty during encoding.
Scan paths	A scan path describes a complete saccade-fixate-saccade sequence. In a search task, an optimal scan path is viewed as being a straight line to a desired target, with relatively short fixation duration at the target.
Blink rate and pupil size	Blink rate and pupil size can be used as an index of cognitive workload. A lower blink rate is assumed to indicate a higher workload, and a higher blink rate may indicate fatigue.

V. HELPFUL REFERENCES

[1] Suhasini, P.S., K.S.R. Krishna and I.V.M. Krishna, 2008, Graph based segmentation in content based image retrieval. J. Comput. Sci., 4: 699-705.

[2] Liu, S., H. Yi, L.T. Chia and D. Rajan, 2005. Adaptive hierarchical multi-class SVM classifier for texture based image classification. IEEE International Conference on Multimedia and Expo, July 6-8, Technology University, Singapore, pp: 4-4

[3] Liu, Y., D. Zhang, G. Lu and W.Y. Ma, 2007. A survey of content-based image retrieval with high-level semantics. Patt. Recog., 40: 262-282.

[4] M. R. Naphade and T. S. Huang, "Extracting semantics from audio-visual content: the final frontier inmultimedia retrieval," *IEEE Transactions on Neural Networks*, vol. 13, no. 4, pp. 793–810, 2002.

[5] VasileiosMezaris, Ioannis Kompatsiaris, Michael G. Strintzis, "Region-Based Image Retrieval Using an Object Ontology and Relevance Feedback", EURASIP Journal on Applied Signal Processing 2004:6, 886–901

[6] I. Felci Rajam and S. Valli, "SRBIR: Semantic Region Based Image Retrieval by Extracting the Dominant Region and Semantic Learning", Journal of Computer Science 7 (3): 400-408, 2011

[7] M.R. Naphade and T.S. Huang, "Extracting semantics from audio-visual content: the final frontier in multimedia retrieval," IEEE Trans. on Neural Networks, vol. 13, no. 4, pp. 793–810, July 2002.

[8] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz, "Efficient and effective querying by image content," *Journal of Intelligent Information Systems*, vol. 3, no. 3/4, pp. 231–262, 1994.

[9] J.R. Smith and S.-F. Chang, "Visualseek: A fully automated content-based image query system," in ACM Multimedia, 1996, pp. 87–98.

[10] I. Kompatsiaris, E. Triantafillou, and M. G. Strintzis, "Region-Based Color Image Indexing and Retrieval," in *Proc. IEEE International Conference on Image Processing*, Thessaloniki, Greece, October 2001.

[11] H. Lu, S. Lu & G. Yang, "Robust Eye Tracking in Video Sequence" Journal of Circuits, Systems, and Computers Vol. 21, No. 1 (2012).

[12] J. Lim, D. Ross, R. Lin and M.-H. Yang, Incremental learning for visual tracking, Advances in Neural Information and Processing Systems, 17, eds. L. Saul, Y. Weiss and L. Bottou (MIT Press, 2005), pp. 793 800.

[13] H. Tan and Y.-J. Zhang, Detecting eye blink states by tracking iris and eyelids, Pattern Recogn. Lett. (2006) 667 675.

[14] Ying Liu, Dengsheng Zhang∗, Guojun Lu, "Region-based image retrieval with high-level semantic susing decision tree learning" Pattern Recognition 41 (2008) 2554 – 2570

[15] Y. Liu, D.S. Zhang, G. Lu, W.-Y. Ma, A survey of content-based image retrieval with high-level semantics, Pattern Recognition, 40 (1) (2007)262–282.

[16] Poole, A., Ball, L. J., & Phillips, P. (2004). In search of salience: A response time and eye movement analysis of bookmark recognition. In S. Fincher, P. Markopolous, D. Moore, & R. Ruddle (Eds.), *People and Computers XVIII-Design for Life: Proceedings of HCI 2004*. London: Springer-Verlag Ltd.

[17] Alex Poole and Linden J. Ball, "Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects"

[18] http://en.wikipedia.org/wiki/Semantic gap, 2010.

[19] CamTu Nguyen, Takeshi Tokuyama, "Bridging Semantic Gaps in Information Retrieval: ContextBased Approaches"

[20] G. Carneiro, A. B. Chan, P. J. Moreno, and N. Vasconcelos. Supervised learning of semantic classes for image annotation and retrieval. IEEE Trans. Pattern Anal. Mach. Intell., 29(3):394{410, 2007.

[21] Hsin-Chang Yang , Chung-Hong Lee, "Image semantics discovery from web pages for semantic-based image retrieval using self-organizing maps" , Expert Systems with Applications 34 (2008) 266–279.

[22] Anuja Khodaskar and Dr. S.A. Ladke, "Content Based Image Retrieval with Semantic Features using Object Ontology" International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 4, June - 2012

[23] Alaa. M Riad, Hamdy. K Elminir and Sameh. Abd-Elghany. Article: A Literature Review of Image Retrieval based On Semantic Concept. *International Journal of Computer Applications* 40(11):12-19, December 2012.

[24] Hichem Bannour and Céline Hudelot, "Towards Ontologies for Image Interpretation and Annotation" CBMI'2011,IEEE pp 211-216.