Bringing High Level Meanings from an Image

Sagarmay Deb

Central Queensland University 400 Kent Street Sydney 2000 NSW, Australia

Abstract   Content-based image retrieval is a difficult area of research in multimedia systems. The research has proved extremely difficult because of the inherent problems in proper automated analysis and feature extraction of the image to facilitate proper classification of various objects. To segment the image to extract meaningful objects and then classify it in high-level like table, chair, car and so on has become a challenge to the researchers in the field. The gap between low-level features like color, shape, texture, spatial relationships and high-level definitions of the images is called the semantic gap. It is very important we find a viable solution of how to extract meaningful definition of an image from low level features so that we can identify the image automatically without any human intervention. As we know billions of images are being generated from various sources all over and it's extremely time consuming and expensive to identify these images manually. When we can identify these vast number of images automatically without human intervention then we can classify them and create databases based on these classifications. These databases could then be used to enhance machine learning of artificial intelligence. Until we solve these problems in an effective way, the efficient processing and retrieval of information from images will be difficult to achieve. In this paper we explore the possibilities of how we can extract high-level meanings from an image before or after the segmentation of the image in an automatized way. We attempt to use a database of data dictionary for the purpose which we believe has the potential for solving the problem of semantic gap in content-based image retrieval.

Keywords   High-Level Definition, Low-Level Definition, Semantic Gap, Content-Based Image Retrieval

1. Introduction

Research on multimedia systems and content-based image retrieval has gained momentum during the last decade. Content-based image retrieval (CBIR) is a very difficult area in the access of multimedia databases simply because there still exist vast differences in the perception capacity between a human and a computer. There are two basic problems that still remain unresolved in the area although some progresses have been made[19]. The first one is the problem of efficient and meaningful image segmentation where we break-up a particular image into meaningful parts based on low-level features like color, texture, shape and spatial locations. Developing a segmentation algorithm which will meaningfully segment all images is yet an open problem in image analysis[8]. The second one is the vast gap existing for an image between low-level features mentioned earlier and high-level or semantic expressions contained in the image like the image of a car, a house, a table and so on[21]. To develop efficient indexing techniques for the retrieval of enormous volumes of images being generated these days, we need to achieve reasonable solutions to these above-mentio ned two problems. But only in very limited and selected

cases, some kinds of solutions have been achieved with apparently promising experimental results. In this paper we focus our attention on the second problem. The research identifies few issues causing this gap, for example, failure to capture local image details with low level features, unavailability of semantic representation of images, inadequate human involvement in the retrieval, and ambiguity in query formulation[9].

Section one gives an introduction of the area. Section two provides an analysis of the works done in the field. Section three presents experimental approach to solve the problem. Section four suggests future directions of research. We put our concluding remarks in section five.

2. Analysis of Works done

Attempts have been made to bridge the gap between low-level features and high-level concepts in the domain of web document retrieval. This work concerns a technique, Latent Semantic Indexing (LSI), which has been used for textual information retrieval for many years. In this environment, LSI is used to determine clusters of simultaneously occurring key words, sometimes, called concepts, so that a query which uses a particular keyword can then retrieve documents perhaps not containing this keyword, but containing other keywords from the same cluster. The use of this technique for content-based web document retrieval, using both keywords and image features

to represent the documents has been made. Two different approaches to image feature representation, namely, color histograms and color anglograms, are adopted and evaluated. Experimental results show that LSI, together with both textual and visual features, is able to extract the underlying semantic structure of web documents, thus helping to improve the retrieval performance significantly[20].

At the earlier stage of CBIR, contextual search has been suggested to narrow the semantic gap whereby the exact definition of an image could be obtained by first retrieving a standard annotated image and then modifying the image to suitable level to obtain the exact shape one is looking for[1]. In another attempt user feedback is used to cluster images into different thematic groups and this feedback is used for the global improvement of the system instead of just in the scope of one query. This leads to a system of continuously learning the semantics of the image base and has led to a significant increase both in recall and precision[2]. Techniques for visual and text feature descriptor extraction, for visual and text feature descriptor integration, and for data clustering in the extraction of perceptual concepts; and on proposing novel ways for discovering perceptual relationships among concepts have been mentioned[3].

A method for image retrieval was introduced using Gaussian color scale space. It uses local features to describe small pieces of images where each piece is labelled. Such a label is denoted by a 'word' of the image. LSI is applied on the words and images and text are combined by concatenating the words in an image to the textual words describing the image[4].

A content-based soft annotation (CBSA) procedure for providing images with semantic labels had been proposed. The annotation procedure starts with labelling a small set of training images, each with one single semantic label (e.g., forest, animal, or sky). An ensemble of binary classifiers is then trained for predicting label membership for images. The trained ensemble is applied to each individual image to give the image multiple soft labels, and each label is associated with a label membership factor[5].

In order to have a more meaningful way to fill the semantic gap, simple classifications of types of image and of types of user are proposed which provides the basis for the specification of a broadly encompassing evaluation study, which will employ the image/user type classification and the expert domain knowledge of selected user groups in the construction of segmented test collections of real queries, images and relevance judgements for a better-informed view on the nature of semantic information need[6].

In medical CBIR research, an ontology of 14 gaps has been defined that addresses the image content and features, as well as system performance and usability. In addition to these gaps, seven system characteristics that impact CBIR applicability and performance have been identified. The framework can be used a posteriori to compare medical CBIR systems and approaches for specific biomedical image domains and goals and a priori during the design phase of a medical CBIR application, so that systematic analysis of

gaps provides detailed insight in system comparison and helps to direct future research[7].

This paper explores the ways in which both top-down, ontologically driven approaches and bottom-up, automatic-annotation approaches to provide retrieval facilities to users, are used. Then it discusses many of the current techniques to combine these top-down and bottom-up approaches to further the goal of improved retrieval effectiveness[10].

This research makes a methodological contribution to the development of faceted vocabularies and suggests a tool for the development of more effective image retrieval systems. The research project applied an innovative experimental methodology to collect terms used by subjects in the description of images drawn from three domains. The resulting natural language vocabulary was then analysed to identify a set of concepts that were shared across subjects. These concepts were subsequently organized as a faceted vocabulary that can be used to describe the shapes and relationships between shapes that constitute the internal spatial composition -- or internal context -- of images[11].

This paper attempts to provide a comprehensive survey of the recent technical achievements in high-level semantic-based image retrieval. Major recent publications are included in this survey covering different aspects of the research in this area, including low-level image feature extraction, similarity measurement, and deriving high-level semantic features. Five major categories of the state-of-the-art techniques are identified in narrowing down the 'semantic gap': (1) using object ontology to define high-level concepts; (2) using machine learning methods to associate low-level features with query concepts; (3) using relevance feedback to learn users' intention; (4) generating semantic template to support high-level image retrieval; (5) fusing the evidences from HTML text and the visual content of images for WWW image retrieval[12].

This research demonstrates how the semantic gap can be narrowed through the use of ontology which represent task-specific attributes, objects, and relations, and relate these to the processing modules available for their detection and recognition[13].

The use of relevance feedback (RFb) and the inclusion of expert knowledge have been investigated to reduce the semantic gap in CBIR of mammograms. Tests were conducted with radiologists, in which their judgment of the relevance of the retrieved images was used with techniques of query-point movement to incorporate RFb[14].

Guided by the rich textual information of Web images, this framework tries to learn a new distance measure in the visual space, which can be used to retrieve more semantically relevant images for any unseen query image. A ranking-based distance metric learning method is proposed for image retrieval problem, by optimizing the leave-one-out retrieval performance on the training data[15].

Based on ontology model which represents spatial knowledge in order to provide semantic understanding of image content, a retrieval system is developed. It is based on

two modules: ontological model merging and semantic strategic image retrieval. The first module allows developing ontological models which represent spatial knowledge of the satellite image, and managing uncertain information. The second module allows retrieving satellite images basing on their ontological model[16].

In CAIM project image context-awareness is used as a means of supporting image-based information retrieval in distributed and mobile environments as image can be associated with multiple contexts and knowledge of these contexts can enhance the quality of the information retrieval process. According to the authors, context-awareness can be used for identifying image semantics and relationships, and may thus contribute to closing the semantic gap between user information requests and the shortcomings of current content-based image retrieval techniques[17].

The relevance feedback technique has been used to reduce the semantic gap between high-level semantics and low-level visual features. As most of the existing interactive content-based image retrieval (ICBIR) systems require a substantial amount of human evaluation labor, which leads to the evaluation fatigue problem that heavily restricts the application of ICBIR, a solution based on discriminative learning is presented. It extends an existing ICBIR system, PicSOM, towards practical applications. The enhanced ICBIR system allows users to input partial relevance which includes not only relevance extent but also relevance reason. A multi-phase retrieval with partial relevance can adapt to the user's searching intention in a from-coarse-to-fine manner[18].

Another relevance feedback and latent semantic index based cultural relic image retrieval system is presented. First, the optimum weights that can be used for iterative retrieval is computed, then a semantic image link network is constructed to store the semantic correlation information between images, which is obtained from memorized relevance feedbacks. Following image relevance feedback, Latent Semantic Indexing is applied to image retrieval, which helps saving in storage and estimating the hidden semantic relationship among images. To illustrate the potential of such an approach a prototype image retrieval system has been developed and Preliminary experimental results on a database containing about 2000 images demonstrate the effectiveness of the proposed model[22].

In this paper the nature of the relevance feedback problem in a continuous representation space in the context of CBIR is analysed. Emphasis is put on exploring the uniqueness of the problem and comparing the assumptions, implementations, and merits of various solutions in the literature. An attempt is made to compile a list of critical issues to consider when designing a relevance feedback algorithm. With a comprehensive review as the main portion, this paper also offers some solutions and perspectives throughout the discussion[23].

3. Experimental Approach

We show in the diagram below how we could bring out the semantic meaning of an image using a data dictionary.
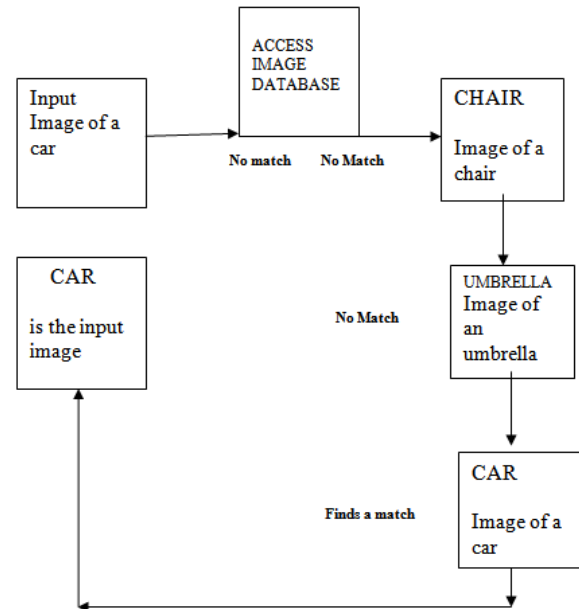
The input image is only the image without any high-level definition. This image accesses the database but could not find similarity of image with that of a chair. It then goes to second image and again could not find a similarity with the image of an umbrella. Then it goes to third image but this time it finds similarity with the image of the database and pick up the high-level definition from that image which is defined as CAR and imprints this definition in the input image to define it as a car. That way we can bring out the high-level or semantic definition of an image.

Of course the time taken by the system to access a large database to find out a match could be very long. Hence we have to work on it further as to how to access the database fast enough to provide the output in a reasonable period of time.

4. Suggested Future Directions of Research

In Section 2, we presented various attempts to extract semantic meanings from an image to fill-in the semantic gap between low-level features and high-level semantic meanings. These include Latent Semantic Indexing (LSI), contextual search, user feedback, data clustering in the extraction of perceptual concepts, content-based soft annotation (CBSA), image classifications, ontology, top-down, ontologically driven approaches and bottom-up, automatic-annotation approaches, using machine learning methods to associate low-level features with query concepts, using relevance feedback to learn users' intention, generating semantic template to support high-level image retrieval, fusing the evidences from HTML text and the visual content of images for WWW image retrieval, use of ontology which represent task-specific attributes, objects, and relations, and

relate these to the processing modules available for their detection and recognition, use of context-awareness for identifying image semantics and relationships to contribute to closing the semantic gap between user information requests and the shortcomings of current content-based image retrieval techniques , enhanced ICBIR system which allows users to input partial relevance which includes not only relevance extent but also relevance reason for a multi-phase retrieval where partial relevance can adapt to the user's searching intention in a from-coarse-to-fine manner.

Although these are good, constructive progresses in solving the problem of semantic gap in CBIR, they cannot define the semantic meanings of an image specifically. They can contribute to some broad classification of the image in certain groups.

To solve this problem we have to develop devices to define the semantic meanings of an image very specifically from low-level features and that should be done automatically without users' interaction.

It is undeniable that complexities of the problem make it difficult to obtain a solution. We suggest a solution where we create a data dictionary with learning capabilities which contain images along with their corresponding relevant keyword/s. When a query image comes it would search the data dictionary and find a match and pick up the corresponding keyword/s to display the semantic meanings of the image on the screen for the users.

5. Conclusions

In this paper we studied the problem of semantic gap between low-level features and high-level semantic definitions in content-based image retrieval. We referred to various relevant techniques developed and presented them so that a full-fledged picture of the problem of semantic gap can be obtained. We suggested to create data dictionary of images with semantic definition attached to each individual image. Then when an input image can find a similarity match with an image of the database we can extract the semantic meaning of the image from the database and attach this to the input image. That way we can extract semantic meaning of an image from input image. Also we suggest some future direction of research in the area which we believe could contribute to the solutions. There has to be reasonable response time for the input image to retrieve relevant image from the database. Also efficient image retrieval system should be in place so that accurate and relevant image from the database is selected based on input image. Since this is a very important area of research with major implications in all spheres of life beginning with medical images, the necessity of this study cannot be overestimated.

REFERENCES

[1] A. Smeulders, "The Sleeping Giant Content-based image retrieval",http://framework.v2.nl/archive/archive/node/text/.xslt/nodenr-136440

[2] L.Brocker, M. Bogen, A. Cremers, "Bridging the semantic gap in content-based image retrieval systems", http://cat.inist.fr/?aModele=afficheN&cpsidt=14051641

[3] A.B.Benitez, S. Chang, "Perceptual Knowledge Construction From Annotated Image Collections", http://citeseer.ist.psu.edu/536148.html

[4] J.C. Gemert, "Retrieving Images as Text", http://citeseer.ist.psu.edu/vangemert03retrieving.html

[5] E. Chang, K. Goh, G. Sychay, G.Wu, "CBSA: Content-based Soft Annotation for Multimodal Image Retrieval Using Bayes Point Machines", http://citeseer.ist.psu.edu/chang03cbsa.html

[6] P. Enser, C. Sandom, "Towards a comprehensive survey of the semantic gap in visual image retrieval", http://cat.inist.fr/?aModele=afficheN&cpsidt=15567509

[7] R. Long, T.M. Deserno, S. Antani, "Ontology of Gaps in Content-Based Image Retrieval, Journal of Digital Imaging", Springer New York, http://www.springerlink.com/content/x554247727h481v1/

[8] D. Mehta, E.S.V.N.L.S. Diwakar, C.V. Jawahar, "A Rule-based Approach to Image Retrieval", www.iiit.net/techreports/2003_8.pdf

[9] M. Islam, "Reducing semantic gap in content based image retrieval using region based relevance feedback techniques", http://www.gscit.monash.edu.au/gscitweb/seminar.php?id=41

[10] J.S. Hare, P.A.S. Sinclair, P.H. Lewis, K. Martinez, P.G.B.. Enser, C.J. Sandom,"Bridging the Semantic Gap in Multimedia Information Retrieval:Top-down and Bottom-up approaches", http://eprints.ecs.soton.ac.uk/12737/

[11] C. Beebe, "Bridging the Semantic Gap: Exploring Descriptive Vocabulary for Image Structure", http://dlist.sir.arizona.edu/2065/

[12] Y.Liu, D. Zhang, G.Lu, W.Y. Ma, "A survey of content-based image retrieval with high-level semantics", http://www.citeulike.org/user/arjen/article/1077366

[13] C.Town, Research, http://www.cl.cam.ac.uk/~cpt23/

[14] Paulo Mazzoncini de Azevedo-Marques, Natália Abdala Rosa, Agma Juci Machado Traina, Caetano Traina Jr, Sérgio Koodi Kinoshita and Rangaraj Mandayam Rangayyan,"Reducing the semantic gap in content-based image retrieval in mammography with relevance feedback and inclusion of expert knowledge", http://www.springerlink.com/content/f9164t3k77u75232/

[15] C. Wang, L. Zhang, H. Zhang, " Learning to Reduce the Semantic Gap in Web Image Retrieval and Annotation", http://portal.acm.org/citation.cfm?id=1390396

[16] W. Messaoudi, I.R. Farah, K.s ettabaa, B. Solaiman, "Semantic Strategic Satellite Image Retrieval", http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4529997

[17] "Context-Aware Image Management", http://caim.uib.no/

[18] Z.Yang, "Dissertation for the degree of Doctor of *Science in Technology*", http://lib.tkk.fi/Diss/2008/isbn97895 12296262/, Last accessed on April 10, 2010

[19] "Bridging the Semantic Gap in Image Retrieval", http://citeseer.ist.psu.edu/497446.html"

[20] R. Zhau, W.I. Grosky, "Narrowing the semantic gap - improved text-based web document retrieval using visual features", http://citeseer.ist.psu.edu/zhao02narrowing.html

[21] R. Zhau, W.I. Grosky, "Bridging the semantic gap in image retrieval", http://citeseer.ist.psu.edu/497446.html

[22] "Relevance Feedback and Latent Semantic Index Based Cultural Relic Image Retrieval", http://www.springerlink.co m/content/1345758j68562261/

[23] X.S. Zhou, S.H. Thomas, "Relevance feedback in image retrieval: A comprehensive review", http://www.springerlink .com/content/gfpvl19h67wpgx3g/