A Comprehensive Review on the Relevance Feedback in Visual Information Retrieval

Manisha Bhimrao Waghmode
Department of Computer Science and Engineering
N.B.Navale Sinhgad College of Engineering, Solapur, India.
Solapur.
manishaw7@gamil.com

Abhijit V.Mophare
Department of Computer Science and Engineering
N.B.Navale Sinhgad College of Engineering, Solapur, India.
Solapur.
mophare_abhi@yahoo.com

Abstract— Visual information retrieval in images and video has been developing rapidly in our daily life and is an important research field in content-based information indexing and retrieval, automatic annotation and structuring of images. Visual information system can make the use of relevance feedback so that the user progressively refines the search result by marking images in the result as relevant , not relevant or neutral to the search query and then repeating the search with the new information. With a comprehensive review as the main portion, this paper also suggested some novel solutions and perspectives throughout the discussion. Introduce the concept of Negative bootstrap, opens up interesting avenues for future research.

Keywords- Bootstrapping, CBIR (Content Based Image Retrieval), Relevance feedback VIR (Visual Information Retrieval).

I. INTRODUCTION

There has been a renewed spurt of research activity in Visual Information Retrieval. Basically two kinds of information are associated with a visual object (image or video): information about the object, called its metadata, and information contained within the object, called visual features. Metadata is alphanumeric and generally expressible as a schema of a relational or object-oriented database. Visual features are derived through computational processes typically image processing, computer vision, and computational geometric routines executed on the visual object. The simplest visual features that can be computed are based on pixel values of raw data, and several early image database systems [1] used pixels as the basis of their data models.

In many specific applications, the process of visual feature extraction is limited by the availability of fast, implementable techniques in image processing and computer vision [2]. If each image/region is represented by a point in a feature space, relevance feedback with only positive (i.e., relevant) examples can be cast as a density estimation or novelty detection problem; while with both positive and negative training examples it becomes a classification problem, or an on-line learning problem in a batch mode, but with the following characteristics associated with this specific application scenario related to small sample issue, asymmetry in training sample and real time requirement[8],[12].

Since last decades a large number of content-based image retrieval (CBIR) technologies (see the recent survey in [1]) have been developed to help users retrieve the desirable database photos using the query by example framework. In these systems, at first a user needs to provide example images as queries. Then, the database images are ranked based on the visual similarities between the query images and the database images. Due to the so-called semantic gap between the low-level visual features (e.g., color, texture, shape) and the high-level semantic concepts, initial retrieval results are frequently unsatisfactory to the users. To bridge the semantic gap, relevance feedback methods have been proposed to learn the user's search intention, and these methods proven effective in improving the retrieval performance of CBIR systems [1]. This paper provides a comprehensive survey of the recent technical achievements in the research area of content based image retrieval. The rest of the paper is organized as follows. Related work is reviewed in Section II. We detail the basic terminology of the visual information retrieval in Section III. relevance feedback of image retrieval in Section IV. Challenging issues can be depicted in Section V. We conclude the paper in Section VI.

II. RELATED WORK

Initially developed in document retrieval (Salton 1989), relevance feedback was transformed and introduced into content-based multimedia retrieval, mainly content-based image retrieval CBIR)[3]-[6].Interestingly, it appears to have attracted more attention in the new field than the previous. A variety of solutions has been proposed within a short period and it remains an active research topic. The reasons can be that more ambiguities arise when interpreting images than words, which makes user interaction more of a necessity; and in addition, judging a document takes time while an image reveals its content almost instantly to a human observer, which makes the feedback process faster and more sensible for the end user[7].

The 1990s saw the launch of a number of experimental CBIR systems, one of the earliest of which, and certainly the best-known, was QBIC (Query By Image Content) [79]. Other systems, including Blobworld, Excalibur, MARS, Photobook and VisualSeek followed; comprehensively surveyed in [53,80], a comparative evaluation was also undertaken [81]. The Benchathlon network (http://www. benchathlon.net/) was established with the aim of developing benchmarking facilities in support of the experimental CBIR environment.

The idea of bootstrapping is to start from a small set of training samples, and successively acquire more training samples at each bootstrapping iteration to derive better classifiers. To achieve this, we need a way for the system to evaluate the quality of newly annotated samples. This can be achieved by using the co-training technique [4] in which two "orthogonal" classifiers independently confirm the quality of newly annotated samples, and learn from each other's results.

Relevance (or similarity) judgments are often subjective, and it is extremely difficult if not impossible to devise a single metric that can be consistently used to determine relevance between two images. Perceived relevance by humans varies not only between individuals, but it also varies for a single individual according to one's perspective of the specific task at hand. Such approaches are common in testing environments [2].

Over the years it has been observed that it is too ambitious to expect a single similarity measure to produce robust perceptually meaningful ranking of images. As alternative, attempts have been made to augment the effort with learning-based techniques. We summarize possible augmentations to traditional image similarity based retrieval in table I [7].

III. BASIC TERMINOLOGY OF VISUAL INFORMATION RETRIEVAL

A. *Analysis of the user's need*

In comparison with studies of users' needs for still image content, search requests and behavior in the context of film and video material has received comparatively little attention. Studies involving archival film collections have been reported, but a fully comprehensive study of user interaction with moving images is still awaited. The World Wide Web contains a great quantity of image and other visual information such as videos, movies, and comic strips, that may belong to structured collections (e.g., museum collections) or be independent (e.g., images found in Web pages in the form of individuals' photographs, logos, and so on). Tools for effective retrieval of this information can prove very useful for many applications. In the current section, here try to show why such tools are indispensable for users, what applications people may need them for, and what services users may ask them to accomplish.

B. *Image Retrieval Systems*

Since the early 1990s, content-based image retrieval has become a very active research area. Many image retrieval systems, both commercial and research, have been built. Most image retrieval systems support one or more of the following options [6]: random browsing, search by example, search by sketch, search by text (including key word or speech) and navigation with customized image categories.

Systematic studies involving actual users in practical applications still need to be done to explore the trade-offs among the different options mentioned. Figure 1 shows the basic structure of web image retrieval system.

TABLE I: COMPARISON OF LEARNING TECHNIQUES IN THEIR APPLICATION TO IMAGE RETRIEVAL

Augmentation	Purpose	Techniques	User Involvement	Drawbacks
Clustering	Meaningful result visualization, faster retrieval, efficient storage	Side-information, kernel mapping, k-means, hierarchical, metric learning	Minimal	Same low-level features, Poor user adaptability
Classification	Pre-processing, fast/accurate retrieval, automatic organization	SVM, MIL, statistical models, Bayesian classifiers, k-NN, trees	Provide training data (not interactive)	Training introduces bias, many classes unseen
Relevance Feedback	Capture user and query specific semantics, refine rank accordingly	Feature reweighting, region weighting, active learning, boosting	Significant (interactive)	Same low-level features, increased user involvement

Here, we will select a few representative systems and highlight their distinct characteristics. Traditional image retrieval mainly bases on the text, using keywords, or free text to descript each image and using text-matching to search. However, the current computer vision technology is not mature, fail to extract automatically of the keywords and semantic information, manual extraction is time-consuming on the one hand, and on the other hand is subjective; at the same time, some visual information of image such as texture and shape are difficult to describe the text accurately[10]. Therefore, content based image retrieval(CBIR) is proposed. This technology extracts visual image features automatically by machine such as color, texture, shape, object location and mutual relations match the images of the database and sample images in the feature space, then search out the similar image of the sample.

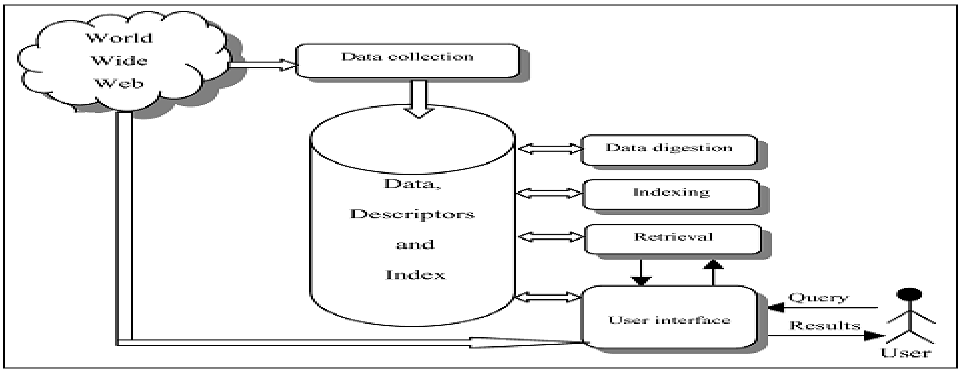

Figure 1: General Structure of a Web Image Search Engine and Its Main Tasks.

C. Content-Based Image Retrieval

The core technology of CBIR is using visual features of the image to search. In essence, it is an approximate matching technique, combining computer vision, image processing, image understanding and databases, and other fields of technological achievements. A typical CBIR system structure is shown in Figure 2. Retrieval process and the algorithm process as follows:

- Analyze the content of the image, extract visual features of each image and then store in the feature database, the features may be text based and visual based. Textual (text based) features are keywords, tags, annotations etc. Visual (visual based) image features are color, shape, texture etc. The visual features are further classified as general features and domain specific features. General features are color, texture, shape and domain specific features are application dependent for e.g. human faces and finger prints.
- During the image retrieval ,only providing fuzzy description of the image is enough, such as a sample image or a sketch, select a feature extraction method to extract features of the case diagram;
- Select the similarity comparison; match the case diagram of the characteristics and features of the library.
- Return the result to the user by similarity from large to small sequencing.
- When the results back to the user, he can select feature composition through interactive feedback and adjust the weights of the various features, and finally get satisfactory results.

Content-based image retrieval is the modern image retrieval system. The Content based image retrieval systems are used to extract image features, index those using appropriate structures and efficiently process user queries providing the required answers. The query processing includes segments and features extraction and search in the feature space for similar images. In Content based image retrieval system various techniques are bought together effectively for the same purpose as image processing, information retrieval and database communities. It is also called query-by-image content and content-based visual information retrieval.

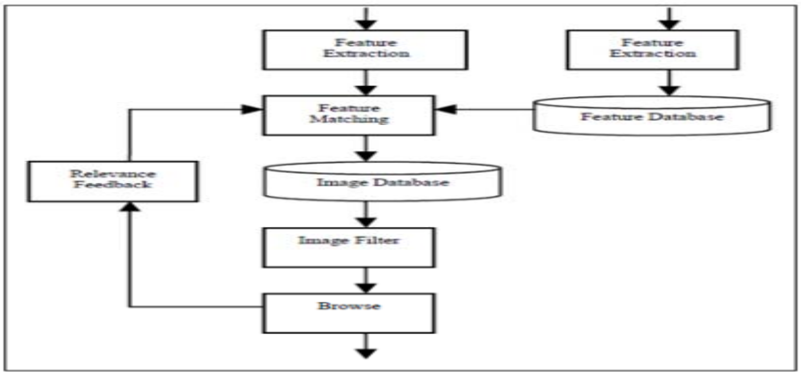

Figure 2: Basic Structure of Content-Based Image Retrieval.

D. Feature Extraction Techniques

The characteristic of image retrieval based on low-level feature is that its features can be obtained directly from the image, while image retrieval based on high-level semantic features are not only to solve the mathematical model, description, search algorithm and other issues, but also possess semantic features to consider ambiguity, uncertainty, and dependence on the natural language description [11].

1. *Retrieval based on color feature*: Color feature is the most intuitive and obvious characteristics of the image, generally using the histogram to describe. It has a fast speed, low storage space requirements, and is not sensitive to the scale and rotation of the image, so it receives wide attention. Now image retrieval based on color feature has become an important means of search. It is mainly divided into two categories: global color feature search and retrieval of local color feature.
2. *Retrieval based on texture feature*: Basically use statistical characteristics of texture or structural characteristics to describe texture feature, and the nature based on the airspace can be converted to the frequency domain, so the commonly used texture description methods can be divided into statistical method, frequency domain method and structure method, and in image retrieval they can be used together.
3. *Retrieval based on shape feature*: Shape-based image retrieval uses shape feature of the target image to search, it is a very important aspect of content-based image retrieval. One important part of Shape-based image retrieval is the research on shape feature extraction.

IV. RELEVANCE FEEDBACK IN IMAGE RETRIEVAL

The basic idea of image retrieval based on relevance feedback is that in the retrieval process, allowing users to assess and mark the results of image retrieval, find out which results are relevant to the query image, which are not relevant, then feedback the relevant information that the users mark to the system as training samples for learning, and instruct next image retrieval ,so make the results more in line with the needs of users. A broader application of relevance feedback method modifies the query vector on the one hand, on the other hand, using feedback information to change the weight of each feature vector in the formula, highlighting the more important vector of the query. The relevance feedback method based on support vector machine (SVM) is that learning the tagged cases and negative samples in the feedback process each time, establish SVM classifier as a model, and use the model to search [9].

A. Negative Bootstrapping

Negative Bootstrap, an iterative negative ensemble learning strategy, to select relevant negatives from many user-tagged images, without the need of new annotation. Bootstrapping aims to use a small set of labeled samples to kick-start the learning process with a large unlabeled corpus. To achieve bootstrapping, we need a way for the system to evaluate the quality of newly annotated samples. This can be achieved by using the co-training technique [1] in which two “view-independent” methods independently confirm the quality of newly annotated samples, and learn from each other’s results. To accomplish this, we exploit the evidences from both the HTML text and visual content features of an image by developing two “orthogonal” classifiers – one based on text, and the other on visual content features. The classifiers are developed using probabilistic Support Vector Machine (pSVM) [11]. There are two notable differences between negative bootstrap and active learning. First, in contrast to active learning which requires human interaction to label examples selected in each round, negative bootstrap selects relevant examples without human interaction. Second, different from active learning wherein the new input data is supposed to be unlabeled and comprised of positive and negative examples; our setting assumes that the new input data contains negatives only. Hence, in active learning, examples the system is most uncertain about, namely closest to the decision boundary [8], are considered informative. Negative bootstrap, by contrast, selects negative examples which are most misclassified, i.e. falling on the positive side and distant from the boundary. Inclusion of such negatives in training pushes towards a tight boundary in the area of the target class, yielding classifiers with better discrimination ability. Negative bootstrap opens up interesting avenues for future research.

B. Text-based Classifier

The text descriptions of a web page often give useful hints on what an embedded image is about. However, while textual contents may contain information that captures the semantics of the embedded image, it also contains other descriptions that are not relevant to the image. These “noises” lead to poor retrieval performance. To improve retrieval performance, we need to extract textual contents that are relevant to an image while avoiding irrelevant information. There are several places where relevant text may be found, namely, (i) image file name; (ii) page title; (iii) alternate text (ALT-tag); and (iv) surrounding text. Most current approaches employ the first three features [4] as they are easy to extract, and tend to provide the most accurate description of the embedded image. However, our empirical studies show that they often do not give sufficient information on an image. The image file name is often abbreviated and may not be recognized as meaningful words. The page title may be too general for the embedded image as there may be more than one image or topic in a web page. Moreover, a large number of images do not even have alternate text. Thus, while the first three features provide reasonably accurate description of an image, they are often inadequate, leading to poor recall.

C. Visual-based Classifier

As discussed in earlier section almost all existing systems use a combination of color histogram, texture and statistical shape features to model the visual contents of images [3]. These visual features have been found to be too low-level to adequately model image contents. As a result, they are effective only in matching highly identical images, and will fail if there is diversity among relevant images, or when the query is looking for object segments in the images. These problems point to the need to develop an adaptive feature representation scheme, where the representation could adapt to the characteristics of the images in the category. Here, we explore the development of adaptive features for texture, to be used in conjunction with color histogram. We do not use shape feature as it is often unreliable and is easily affected by noise.

Comparing with conventional wavelet-based texture features, the main advantages of adaptive texture features are that they are efficient and provide an accurate reflection of local texture properties. This is because through matching pursuit, we are able to obtain the most appropriate representation for an image with fewest significant coefficients. As in text, we explore two visual representations of images to compare our adaptive matching pursuit features with traditional visual content features.

The relevance feedback method based on support vector machine (SVM) is that learning the tagged cases and negative samples in the feedback process each time, establish SVM classifier as a model, and use the model to search. In each feedback process, if the user’s sample image has similar properties with the feature space, that is, as a support vector of a sample, then the sample because of distance has no effect on the SVM classifier, so although feedback from users tag images limited, but for the establishment of SVM classifier is sufficient to effectively control the generalization ability of machine learning. SVM method is the core function of the main difficulties in the selection, the selection of kernel function directly affects the classifier's generalization ability [9].

V CHALLENGING ISSUE FOR FURTHER DIRECTIONS

Development in the field of raises visual information retrieval many research challenges such as:

- To further improve the color characteristics of the retrieval results, from the color of physical, visual and psychological aspects of a comprehensive in-depth.
- Finding new connections, and mining patterns. Text mining techniques might be combined with visual-based descriptions.
- Furthermore shape measurement method still can’t distinguish shape well; it cannot express the similarity between shapes effectively. The research of image retrieval based on shape feature is still a challenging research topic.
- New image annotation techniques need to be developed because there are no such techniques available which properly deal with semantic gap.
- The collaborative bootstrapping approach, initially developed for text processing, can be employed effectively to tackle the challenging problems of multimedia information retrieval on the Web.
- The consistency and scalability of the co-training approach.

VI. CONCLUSIONS

In this paper, past and current technical achievements in content based image retrieval system are reviewed. This review presented the most important aspects of image retrieval from the World Wide Web and its basic terminology. It is concluded that although significant amount of work has been done in this area but still there is no generic approach for high-level semantic-based image retrieval. To design a full-fledged image retrieval system with high-level semantics requires the integration of primitive feature extraction and high-level semantics extraction parameters. Open research issues are identified and future research directions suggested.

ACKNOWLEDGEMENT

I thank all the anonymous reviewers and editors for their valuable comments and suggestions to improve the quality of this manuscript.

REFERENCES

[1] A. Blum & T. Mitchell. Combined labeled data and unlabelled data with co-training. Proceeding of Annual Conference on Computational Learning Theory. 1998.

[2] Arnold W. M. Smeulders and Xirong Li, CeesG. M. Snoek, "Bootstrapping Visual Categorization With Relevant Negatives" IEEE Transactions On Multimedia, VOL. 15, NO. 4, JUNE 2013 933)

[3] Chen J-Y, Bouman CA and Dalton J (2000). Hierarchical browsing and search of large image databases. IEEETrans. on Image Processing 9(3): 442-445.

[4] Chen Y, Zhou XS and Huang TS (2001). One-class SVM for learning in image retrieval. Int’l Conf on Image Processing, Greece.

[5] Guo G, Zhang HJ and Li SZ (2001). Boosting for content-based audio classification and retrieval: an evaluation.Microsoft Research Technical Report: MSR-TR-2001-15.

[6] Hechao Yang, Xuemei Zhou "Research of Content Based Image Retrieval Technology" Workshops(ISECS ’2010)

[7] Huamin Feng1,2, Rui Shi1 and Tat-Seng Chual "A Bootstrapping Framework for Annotating and Retrieving WWW Images" (MM’04, OCTOBER 10–16, 2004, ACM 1-58113-000-0/00/0000)

[8] Ishikawa Y, Subramanya R and Faloutsos C (1998). MindReader: query databases through multiple examples. Int’l Conf. on Very Large Data Bases (VLDB), NY.

[9] J.C. Platt , A.J. Smola, P. Bartlett, B. Scholkopf & D. Schuurmans (Eds). ‘Advances in Large Margin Classifiers’MIT Press, 1999

[10] K. Yanai and K. Barnard, “Probabilistic web image gathering,” in Proc. ACM MIR, 2005, pp. 57–64.

[11] Santini S and Jain R (2000). Integrated browsing and querying for image database. IEEE Trans. Multimedia7(3).

[12] Xiang Sean Zhou*, Thomas S. Huang "Relevance Feedback in Image Retrieval: A Comprehensive Review ACM 2004.

[13] Wikipedia, Journals and white papers.