내용기반 영상검색 시스템의 분석 및 발전 방안

Kulwinder Singh[†]· Ming Ma[†]· 박동원[††]· 안성옥[††]

요 약

내용기반 영상검색 분야에서의 활발한 연구로 지난 수년간 기술과 성능 면에서 괄목할 성장을 이룩해 내었다. 본 논문에서는 기존의 영상검색 시스템을 체계적으로 분석하여 아직까지 남아있는 취약점 및 개선 부분에 대하여 기술하였다. 특히, 의미론적 영상검색에 대하여 주안점을 두어, 시스템 향상을 위하여 심도있게 연구가 진행 되어야 할 분야의 방향 및 주제를 분류하고 분석하여 제안하였다.

Anatomy of Current Issues on Content-Based Image Retrieval

Kulwinder Singh[†]· Ming Ma[†]· DongWon Park [††]· Syungog An[††]

Abstract

In the past few years, enormous improvements have been obtained in the field of content-based image retrieval (CBIR). This paper presents a comprehensive survey on the current CBIR systems and some of their challenging technical aspects, which stand as an obstacle on its way to become successful. Furthermore, we have focused on the current state of semantic image retrieval and also we have suggested future promising directions for further research.

Keywords : Content gap, Query type, Content-based image retrieval.

† 비 회 원: 배재대학교 IT 공학부 석사과정
† 비 회 원: 배재대학교 IT 공학부 석사과정
†† 비 회 원: 배재대학교 IT 공학부 부교수
†† 종신회원: 배재대학교 IT 공학부 교수
논문접수:2003년 7월 21일, 심사완료:2003년 10월 16일

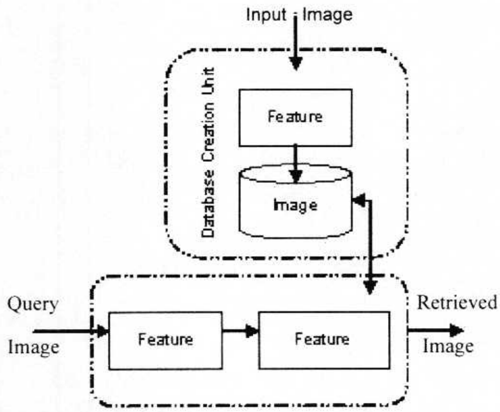
1. Introduction

"Content Based Access To Multimedia Data" is a highly active field of research from past few decades due to exponential growth of multimedia data, rapid expansion of World Wide Web, the increased memory & processor capacity allowing storage of large amounts of digital data and the need to handle queries and browse in large image databases. Image databases exist for storing art collections, satellite images, medical images, trademark images and general collections of photographs. Usually, the way of searching these collections is by still using human indexers to select keywords for their images or simply by browsing. However image databases have opened the way to content-based retrieval. Both academic and commercial development communities have paid lots of attention towards CBIR systems in recent years. The main objective of such systems is to enable the users model queries such as, "retrieve images similar to a given image" from a large image database. The CBIR systems need to extract low level or high level features of an image, index them using appropriate structures and efficiently process user model queries providing the efficient result. So we can define CBIR as "*the process of retrieving desired images from a large collection on the basis of features (such as color, texture and shape) that can be automatically extracted from the images themselves.*" Visual features of an image are much more versatile compared with text-annotation, and the amount of visual data is already enormous and still expanding very rapidly. The introduction to content based image retrieval method has been made with the hope to cope with these special characteristics of visual data. It has been widely accepted that the image retrieval techniques are the combination of both low-level visual features revealing the more detailed perceptual aspects and high-level semantic features underlying the more general conceptual aspects of visual data. Actually any of these two types of features is not capable to retrieve or manage visual data effectively or efficiently. In spite of the great efforts made by the researchers to combine these two aspects of visual data, the gap still exists between them. Intuitive and heuristic approaches do not provide us with satisfactory performance. Therefore, there is an urgent need of finding the latent correlation between low-level features and high-level concepts and merging them from a different perspective. How to find this new perspective and bridge the gap between visual features and semantic features has been a major challenge in this research field. To overcome this challenge, on the one hand, one can go along the direction of searching for more low-level features that can improve the performance of the current content-based retrieval schemes and on other hand we can go along the direction of searching more high level features or even we can find very efficient result by the integration of low and high level features.

This paper is organized as follows. Section 2 deals with the components of CBIR while section 3 classifies the query levels. We have highlighted some of the current CBIR systems in section 4 followed by modes of interactions in section 5. Section 6 reveals some of the hurdles on the way of CBIR. Section 7 presents the current state of Semantic Image retrieval and issues for further research. Conclusions and Recommendations are available in section 8.

2. Components of CBIR

Feature extraction is the most important component in a content-based image retrieval system. Most CBIR techniques fall into two categories: manual and computational.



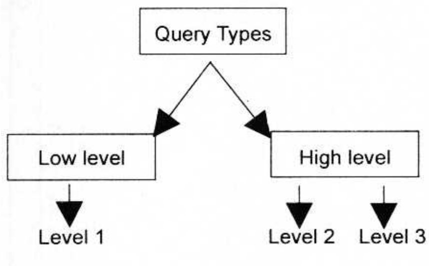
(Fig.1) Model of computational features extraction

One of the basic techniques for computational features extraction is shown in the Fig. 1.

In manual approaches, a human expert may identify and annotate the essence of an image for storage and retrieval. In a broad sense, features may include both text-based descriptions (keywords, annotations, etc.) and visual features extraction is the basis of any CBIR technique. Widely used features include color, texture, shape and spatial relationships.

3. Classification of query levels

Primarily query types are divided into two broad concepts and then subdivided into three levels ordered by increasing complexity as shown in Fig. 2:



(Fig.2) query types

Low-level concept is extracted from visual contents of an image.

- Level 1. Retrieval by *primitive* features like color, texture, shape or the spatial information of image elements.

But *High-level* concept is not extracted directly from visual features of an image. But they represent the relatively more important meanings of objects and scenes in the images that are perceived by human beings.

- Level 2. Retrieval by *derived* (*logical*) features, involving some degree of logical illation about the identity of the objects depicted in the image. It can be divided further into: (a) Retrieval of objects of a given type. (b) Retrieval of individual objects or persons.

To answer the queries at this level, human perception can be used.

- Level 3. Retrieval by *abstract* attributes, involving a significant amount of high-level reasoning about the meaning and purpose of the objects or scenes depicted. It can be divided further into: (a) Retrieval of named events or types of activity. (b) Retrieval of pictures with emotional or religious significance.

Complex reasoning and subjective judgment are needed to answer queries at this level.
The above definitions of classification of image query types are very useful in evaluating different image retrieval techniques. Most evaluations in the rest of this paper are based on this definition. Level 2 and Level 3 together are usually referred to as *semantic* image retrieval and the gap between Level 1
and Level 2 as *semantic* gap, which is the most significant gap at present.

4. Current CBIR Systems

Several systems have been developed in recent years in the research area of content-based information retrieval, notably:

- QBIC (Query By Image Content) by IBM,
- Virage by Virage, Inc.,
- Photobook by MIT Media Lab.,
- VisualSEEK by Columbia University,
- RetrievalWare by Excalibur Technologies Corp.,
- NeTra by the University of California, Alexandria Digital Library,
- IRIS by German Software Development Laboratory of IBM and the AI group of the University of Bremen,
- CORE by the University of Singapore.
- Amore (Advanced Multimedia Oriented Retrieval Engine) by C & C Research Laboratories NEC USA, Inc.
- Blobworld by Computer Science Division, University of California, Berkeley.
- Berkeley Digital Library Project by University of California, Berkeley.
- CBVQ (Content-Based Visual Query) by Image and Advanced Television Lab, Columbia University, NY.
- MARS (Multimedia Analysis and Retrieval System) by Department of Computer Science, University of Illinois at Urbana-Champaign, further developed at Department of Information and Computer Science, University of California at Irvine, CA.
- MetaSEEk by Image and Advanced Television Lab, Columbia University, NY, USA.
- MIR (Multimedia Information Retrieval System) by Center of Excellence for Document Analysis and Recognition, University at Buffalo, NY, USA.

It is also hard to compare the performance of different CBIR systems, mainly because it is hard to construct a global benchmark database and ground truth. Most of all current CBIR systems operate at *level 1*. The commonest features used in CBIR are mathematical measures of color, texture, and shape. CBIR at *level 2*: most recent researches are concentrated on *scene recognition* and *object recognition*. So it has been a major challenge to bridge the semantic gap. Reports of CBIR at level 3 are very rare.

5. Modes of interactions in retrieval system

Still, conventional image databases are text-annotated. However, there are two major problems with this method. First, creating keywords for a large number of images is time consuming. Moreover, the keywords are inherently subjective and not unique. Due to these disadvantages, automatic indexing and retrieval based on visual and semantic features came into existence. Undoutedly, human beings are much better than computers while extracting and making use of semantic information from images. Unfortunately, this goal is still beyond the reach of state-of-the-art in computer vision.

6. Obstacles on the way of CBIR

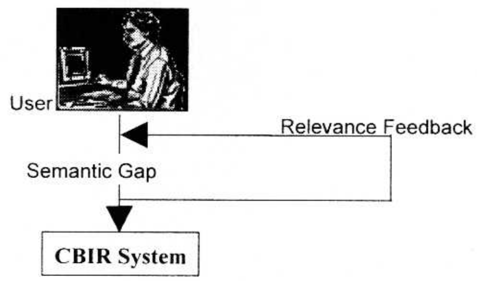
CBIR suffers from several disadvantages: -

- *Bad results due to the semantic gap and the subjectivity of human perception.*

The first point stands for the difference between the high-level CBIR concepts usually presented to users and the low-level features actually employed.
The latter addresses the fact that different persons (recipients) or the same person in different situations may judge visual content differently.

☐ *Semantic gap*

One of the major problems in content-based image retrieval is the so-called *semantic gap* shown in Fig. 3 -the mismatch between the capabilities of current CBIR systems and the needs of users.



(Fig.3) The semantic gap between the user and the retrieval system

The vast majority of current CBIR techniques can retrieve images only by similarity of appearance, using features such as color, texture or shape. User surveys indicate clearly that the majority of image users look for images with specified semantic content - types of object, scenes or individual people. Actually CBIR systems have achieved great success when dealing with pure scenery image data, but when images change from simple scenery to more semantically complex domain, such as images with human beings, the performance reduces significantly.

The reason is:
(i). These systems do not differentiate the semantic foreground objects (e.g. human beings) from background scenery.
(ii). The matching schemes are based on certain primitive statistical information extracted from the combination of the foreground and the background scene.

- *Bad querying performance*

Using (computational often very complex) distance functions for the comparison of feature vectors leads to bad, sometimes unacceptable response times.

- *Complex interfaces*

CBIR is very different from traditional text retrieval. CBIR interfaces tend to be complex and difficult to use. Additionally, average users are overtaxed by the requirement to select features and weights for a specific querying process.

- *Feature matching techniques*

Designing efficient matching algorithms make CBIR a very challenging problem.
Geometric hashing is an example of a powerful feature-based matching technique. Its main drawback is that it
requires large memory to store shape indices, thus is not well suited for very large databases.
In general, the comparison is performed either globally using techniques such as histogram matching and color layout indexing, or locally based on decomposed regions (objects) of the images. A major drawback of the global histogram search lies in its sensitivity to intensity variations, color distortions, and cropping. The color layout indexing method is proposed to alleviate this drawback. But it is in general sensitive to shifting, cropping, scaling, and rotation. Recognizing these deficiencies and the fact that human beings are capable of this technically complex performance, researchers try to relate human perception to CBIR systems.

- *Web Searching*

For images on the web, even though some good work has taken place, technical breakthroughs are needed to make the image search engines comparable to their text-based counterpart.
One major technical barrier lies in linking the low-level visual feature indexes used in most systems today to more desire semantic-level meanings. Based on preliminary on-line experiments, we have observed that subject browsing and text-based matching are still more popular operations than feature-based search options. That is partly the reason that commercial image retrieval systems on the web typically use customized subject categories to organize their image collection. Usually, different image retrieval systems focus on different sections of users and content. As a result, the indexing features and the subject taxonomies are also different, causing the concern of interoperability. Several research systems on image metaservers have investigated frameworks for integrated access to distributed image libraries.

- *Interaction*

Interaction, as a fundamental difference between computer vision pattern recognition and image retrieval, was already introduced in the 1992. The early research was proposed by the NEC laboratories in Japan and the MIT Media Lab. During the evolution of content-based image retrieval, it has been proved that early literature which focus on "full automated systems" and tries to find a "single best feature" does not lead to success.
There are two distinct issues existing in current image retrieval techniques, which result in the emergence of the interaction. First is the semantic gap we have mentioned in previous Section. Second is the subjectivity of human perception. Different people, or the same person under different circumstances, may perceive the same visual content differently. This is called human perception subjectivity. The subjectivity exists in various levels.
Therefore, interaction and feedback have moved into the focus of attention. For example, the QBIC system uses interactive region segmentation and the MARS proposes relevance *feedback* architecture in image retrieval, where human and computer can interact with each other to improve the retrieval performance.

- *Database Organization*

As the size of database grows larger and larger, databases can no longer be ignored as an essential component of content based retrieval system. The connection between content-based image retrieval and database research is likely to increase in the future. Already, the most promising efforts are interdisciplinary, but, so far, problems like the definition of suitable query languages, efficient search in high dimensional feature space, search in the presence of changing similarity measures are largely unsolved.
In addition, when interactive performance is essential, storage and indexing must be organized in advance. Such large data sets will have an effect on the choice of features as the expressive power, computational cost, and hierarchical accessibility determine their effectiveness. For very large data sets, a view on content integrated with computation and indexing cannot be ignored. When speaking about "indexing" in computer vision, the emphasis is still on what to index, whereas the emphasis from the database side is on how to index. The difference has become smaller recently, but we believe most work is still to be done. Furthermore, in dealing with large feature vector sizes, the expansion of query definitions and query expansions in a useful manner for a variety of user aims is still mostly unanswered.

- *Evaluation*

Any technique is pushed forward by its domain□s evaluation criterion. SNR is used in data compression, and precision and recall are used in text-based information retrieval. Good metrics will lead the technique in the correct direction while bad ones may mislead the research effort. Currently, some image retrieval systems measure performance based on the "cost/time" to find the right images. Others evaluate performance using precision and recall, terms borrowed form text-based retrieval.

Although these criteria measure the system□s performance to some extent, they are far from satisfactory. One major reason causing the difficulty of defining a good evaluation criterion is the perception subjectivity of image content. That is, the subjectivity of image perception prevents us from defining objective evaluation criteria. But still, we need to find a way of evaluating the system performance to guide the research effort in the correct direction.

7. Current state of semantic Image Retrieval

- Model-based techniques can draw on high-level reasoning, but are domain-specific and incapable of learning
- Statistical techniques less domain-specific but lack deep knowledge or ability to learn
- Adaptive techniques can continue to learn, but depend crucially on quality of user input

7.1 Issues further researches

(1) Are current techniques for shape, color and texture retrieval going to get much better?
(2) Are we wasting our time looking for better ways of segmenting images?
(3) Is automated semantic image retrieval an achievable goal?
(4) Is content-based navigation a better idea than content-based retrieval?
(5) Should we be putting all our effort into video rather than still image retrieval techniques?
(6) Can we train up search intermediaries sufficiently skilled to overcome CBIR's current imitations?
(7) Do we really know anything about user requirements?

8. Conclusions and recommendations

Presently CBIR is an unvaned technology with potency. It is pretty much a research topic at present. Although over the years a number of CBIR algorithms has been proposed, none has stood out as being particularly robust, despite the fact that each claims to perform best on some benchmark. Unfortunately there is no universally accepted benchmark for CBIR and the lack of a metric is probably one of the main causes for the poor quality of today's algorithms -- without a performance metric is it impossible to diagnose the shortcomings of a particular algorithm. Hopefully some achievements will appear in the next five to ten years. Some recommendations are made at the end of the paper to favor the development of CBIR technology.

All professionals involved in image data management need to be aware of standards development in the area, and be prepared to influence their future development if appropriate. Professionals who are working in field of image processing should keep abreast of emerging standards, like MPEG-4,MPEG-7,MPEG-21 and JPEG-2000 standard for multimedia content description, and contribute to their future development where necessary.

Reference

[1] J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz,R. Humphrey, R. Jain, and C. Shu. The virage image search engine: An open framework for image management. Symposium on Electronic Imaging: Science and Technology □ Storage & Retrieval for Image and Video Databases IV, 2670:76□ 87, 1996.

[2] R. Barber, M. Flickner, J. Hafner, W. Niblack, andD. Petkovic. Efficient and effective querying by image content. Journal of Intelligent Information Systems, 3:231□262, 1994.

[3] S. Servetto, Y. Rui, K. Ramchandran, and T. S. Huang. A region-based representation of images in mars. Special issue on multimedia Signal Processing, Journal on VLSI Signal Processing Systems, October 1998.

[4] J. Smith and S.-F. Chang. Visualseek: A fully automated content-based image query system. Proceeding ACM International Conference of Multimedia, pages 87-98, November 1996.

[5] J. R. Smith. Integrated Spatial and Feature Image Systems: Retrieval, Analysis and Compression. PhD thesis, Columbia Univ., 1997.

[6] Y. Rui, T. S. Huang, and S. Mehrotra, Content-based image retrieval with relevance feedback in MARS, in Proc. IEEE Int. Conf. on Image Proc., 1997.

[7] S. Mehrotra, Y. Rui, O.-B. Michael, and T. S. Huang, Supporting content-based queries over images in MARS, in Proc. of IEEE Int. Conf. on Multimedia Computing and Systems, 1997.

[8] Y. Rui, T. S. Huang, S. Mehrotra, and M. Ortega, A relevance feedback architecture in content-based multimedia information retrieval systems, in Proc. of IEEE Workshop on Content-Based Access of Image and Video Libraries, in Conjunction with IEEE CVPR□97, 1997.

[9] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, Relevance feedback: A power tool in interactive content-based image retrieval, IEEE Trans. on Circuits Systems Video Techno. (Special Issue on Interactive Multi-Media Systems for the Internet), Sept. 1998.

[10] C. Frankel, M. J. Swain, and V. Athitsos. Webseer: An Image Search Engine for the World Wide Web, Technical Report TR-96-14, Computer Science Department, University of Chicago, 1996.

[11] V. Athitsos, M. J. Swain, and C. Frankel, Distinguishing, photographs and graphics on the world wide web, in Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries, 1997.

[12] S. Sclaroff, L. Taycher, and M. La Cascia, Imagerover: A content-based image browser for the world wide web, in Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries, 1997.

[13] R. Jain, A. Pentland, and D. Petkovic, in NSF-ARPA Workshop on Visual Information Management Systems, Cambridge, MA, June 1995.

[14] Arnold W.M. Smeulder et al, Content-Based Image Retrieval at the End of the Early Years, IEEE Transactions on Pattern Analysis and Machine intelligence, Vol.22, No.12, December 2000.

[15] J. R. Smith and S.-F. Chang, Visualseek: A fully automated content-based image query system, in Proc.ACM Multimedia 96, 1996.
[16] Y.Rui, T.S.Huang and S.F. Chang: Image Retrieval:
Current Techniques, Promising Directions, and Open Issues in Journal of Visual Communication and Image Representation 10,39-62,1999

Kulwinder Singh

1998 Ravishanker Shukla 대학교 전기공학과 (공학사)
2002~현재 배재대학교 정보통신공학과 석사과정

관심분야: 영상처리, Biometric Authentication
E-Mail: singh@pcu.ac.kr

박 동 원

1983 고려대학교 전기공학과 (공학사)
1985 Florida Institute of Technology 공학석사 (전산학 전공)
1993 Texas A&M University 공학박사 (전산학 전공)
1994 ~ 현재 배재대학교 IT공학부 부교수

관심분야: 실시간통신, 영상처리
E-Mail: dwpark@pcu.ac.kr

Ma Ming

1999 내몽고 대학교 전산학과 (공학사)
2002~현재 배재대학교 정보통신공학과 석사과정

관심분야: 영상처리, 멀티미디어통신
E-Mail: singh@pcu.ac.kr

안 성 옥

1983 고려대학교 수학교육과 졸업(학사)
1985 고려대학교 대학원 졸업 (전산학전공 석사)
1989 고려대학교 대학원 졸업 (전산학전공 박사)

1991~현재 배재대학교 정보통신공학부 교수
1993~1994 펜실베이니아 주립대학교 컴퓨터학과 Post-doc
1994-1995 뉴욕주립대학교 컴퓨터학과 Post-doc

관심분야: 데이터베이스, 컴퓨터 교육, 영상처리
E-Mail: sungohk@pcu.ac.kr