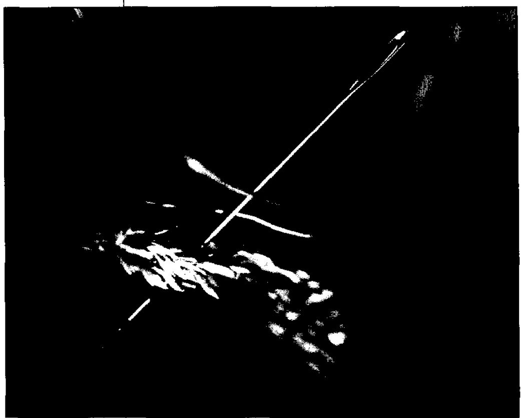

Content-Based Image Retrieval Systems

Venkat N. Gudivada
Ohio University

Vijay V. Raghavan
University of Southwestern Louisiana

Images are being generated at an ever-increasing rate by sources such as defense and civilian satellites, military reconnaissance and surveillance flights, fingerprinting and mug-shot-capturing devices, scientific experiments, biomedical imaging, and home entertainment systems. For example, NASA's Earth Observing System will generate about 1 terabyte of image data per day when fully operational. A content-based image retrieval (CBIR) system is required to effectively and efficiently use information from these image repositories. Such a system helps users (even those unfamiliar with the database) retrieve relevant images based on their contents. Application areas in which CBIR is a principal activity are numerous and diverse:

- art galleries and museum management,
- architectural and engineering design,
- interior design,
- remote sensing and management of earth resources,
- geographic information systems,
- scientific database management,
- weather forecasting,
- retailing,
- fabric and fashion design,
- trademark and copyright database management,
- law enforcement and criminal investigation, and
- picture archiving and communication systems.

With the recent interest in multimedia systems, CBIR has attracted the attention of researchers across several disciplines.

PREVIOUS APPROACHES

Previous approaches to content-based retrieval have taken two directions.[1-4] In the first, image contents are modeled as a set of attributes extracted manually and managed within the framework of conventional database-management systems. Queries are specified using these attributes. Attribute-based representation of images entails a high level of image abstraction. Generally, the higher the level of abstraction, the lesser is the scope for posing ad hoc queries to the image database. Attribute-based retrieval is advocated and advanced primarily by database researchers.

The second approach depends on an integrated feature-extraction/object-recognition subsystem to overcome the limitations of attribute-based retrieval. This subsystem automates the feature-extraction and object-recognition task that occurs when the image is inserted into the database. However, automated approaches to object recognition are computationally expensive, difficult, and tend to be domain specific. This approach is advanced primarily by image-interpretation researchers.

RECENT RESEARCH

Recent CBIR research recognizes the need for synergy between these two approaches. Toward this goal, efforts draw upon ideas from areas such as knowledge-based systems, cognitive science, user modeling, computer graphics, image processing, pattern recognition, database-management systems, and information retrieval. This confluence of ideas has culminated in the introduction of novel image representations and data models, efficient and robust query-processing algorithms, intelligent query interfaces, and domain-independent system architectures. These advances have brought CBIR systems from their infancy to a state of reasonable maturity.

Primitive versus logical

Current approaches to CBIR differ in terms of which image features are extracted, the level of abstraction manifested in the features, and the degree of desired domain independence. There are two major categories of features: primitive and logical. Primitive, or low-level, image features such as object centroids and boundaries can be extracted automatically or semiautomatically. Logical features are abstract representations of images at various levels of detail. Some logical features may be synthesized from primitive features whereas others can only be obtained through considerable human involvement. Logical features denote the deeper domain semantics manifested in the images.

Current approaches differ in terms of image features extracted, their level of abstraction, and the degree of domain independence.

Trade-offs

In developing CBIR systems, there is an inherent trade-off between the degree of automation desired for feature extraction and the level of domain independence realized in the system. CBIR systems can be developed with emphasis on automatic and dynamic feature extraction. Although some features may be determined a priori, these systems can dynamically compute the required primitive features and synthesize the logical ones, both under the guidance of a domain expert. This approach is ambitious and aims at sophisticated CBIR; systems employing this approach are most suitable for applications involving relatively small image collections and when retrieval is performed exclusively by domain experts. We call this the *dynamic feature-extraction* approach.

CBIR systems can also be developed that achieve a reasonable degree of domain independence at the cost of not having a completely automated system for feature extraction. We refer to this approach as *a priori feature extraction*. A set of primitive features is extracted, and all logical features are derived only when the image is inserted into the database. Primitive features are typically derived semiautomatically, while the logical features are extracted manually or semiautomatically. Queries are processed using both primitive and logical features.

Query classes

Regardless of which approach is used, generic query classes[5] facilitate CBIR through retrieving by

- color,
- texture,
- sketch,
- shape,
- volume,
- spatial constraints,
- browsing,
- objective attributes,
- subjective attributes,
- motion,
- text, and
- domain concepts.

An image retrieval system featuring all these query classes will have reasonable generality for dealing with diverse applications.

Color and texture queries let users select images containing objects specified accordingly. The notion of an image object is domain dependent and represents a semantic entity of interest in an application. Retrieval by sketch lets users outline an image and then retrieves a like image from the database. This class can be thought of as retrieving images by matching the dominant edges. The shape class of queries has a counterpart in 3D images referred to as *Retrieval by Volume*. The spatial constraints category deals with a class of queries based on spatial and topological relationships among the objects in an image. These relationships may span a broad spectrum ranging from directional relationships to adjacency, overlap, and containment involving a pair of objects or multiple objects.

Retrieval by browsing is performed when users are vague about their retrieval needs or are unfamiliar with the structure and types of information available in the image database. The objective attributes query uses attributes like the date of image acquisition or the number of bedrooms in a residential floor-plan image and is similar to Structured Query Language retrieval in conventional databases. Retrieval is based on an *exact match* of attribute values. In contrast, a subjective attributes query is characterized by the presence of attributes that may be interpreted differently by each user. For example, in a mug-shot database, the attribute *eyebrow shape* assumes one of three values: arched, normal, or straight. One user may assign the normal value for the eyebrow shape, while another may interpret the value as arched.

Retrieval by motion facilitates retrieving spatiotemporal image sequences depicting a domain phenomenon that varies in time or geographic space. Some applications require retrieving images based on associated text. Such a need is modeled by retrieval by text. Note that processing this query involves natural language processing and information retrieval techniques.[6]

The above query classes can be used as fundamental operators in formulating a class of complex queries referred to as *Retrieval by Domain Concepts*. An example of this is "Retrieve images of snow-covered mountains."

Not all the above generic query classes are necessary, however, for a given image retrieval application. For example, a real estate marketing application may require only retrieval by browsing, objective attributes, shape, and spatial constraints. In such a case, the a priori feature-extraction approach helps generate an application-specific system from a generic one by retaining only the necessary query classes. Though this approach doesn't entail the level of sophistication of the dynamic feature-extraction approach, it appears to be the most promising and realistic. Since users are not expected to be familiar with the image-interpretation task, the a priori feature-extraction approach has broader appeal and suits naive and casual users.

It should be emphasized that true CBIR can only be achieved by synergistically employing various generic query classes in a way transparent to the user. In other words, Retrieval by Domain Concepts queries need to be expressed as a composition of the fundamental operators in a domain-independent way without involving the user. Furthermore, images don't occur in isolation; CBIR issues must be addressed in a broader information retrieval context. For example, diagnostic medical images are retrieved not only in terms of image contents but also in terms of other information associated with the images (like text describing a physician's diagnosis, treatment plan, and the final outcome). Hence, from the physician's viewpoint, the text associated with diagnostic medical images is as central as the content of the image itself. Therefore, information retrieval techniques[6] have an important complementary role in CBIR.

Images don't occur in isolation; CBIR issues must be addressed in a broader information retrieval context.

IMPORTANT RESEARCH AREAS

Although recent progress in CBIR is impressive, exploring the following areas will help realize its full potential.

Data modeling

An expressive data model is at the heart of the CBIR system. This model determines supported query classes and the range of applications that can thus be modeled. A data model for CBIR should feature a rich set of modeling constructs to capture the information necessary for processing the types of query classes discussed earlier. Existing data models are not comprehensive and thus are inadequate for domain-independent CBIR.

Logical features and algorithms

Domain-independent and storage-efficient logical features are required to realize the data model. Furthermore, efficient and robust algorithms based on logical features must be devised for processing various generic queries interactively.

Query-specification scheme

Domain experts will probably prefer a sophisticated graphical interface for specifying queries. This interface should support several generic query classes, with query specification for each class based on natural and efficient schemes. These schemes should be integrated under a graphical environment to provide a unified view of querying the image database.

In contrast, casual and naive users may prefer to specify their queries by using natural language. For example, it is far easier to express a query such as "Show me images of snow-covered mountains" in natural language than it is to sketch an image of a mountain and sprinkle it with snow texture. Mapping such a natural language query onto generic query classes remains to be investigated.

The interface should also display query results in a context that makes them easier to interpret. Information-filtering and user-modeling techniques have a major role in refining user queries and determining their general context.

Similarity-based retrieval

One of the main differences between an image retrieval system and a database system is the former's ability to rank-order database images by the degree of similarity with the query image (that is, similarity-based retrieval[5]). Database systems typically process queries based on exact match. A theoretical framework for similarity-based retrieval should be developed.

Access structures for logical features

When an image database is on the order of tens of thousands of images, suitable access structures and methods become critical for efficient query processing. Such access schemes should help retrieve only those images from secondary or tertiary storage that have the most potential for being relevant to the user's query. In essence, access schemes should act as filters in eliminating nonrelevant images while ensuring that the relevant ones are not dismissed. Query-processing algorithms are applied only to images that pass through the filter. Existing access schemes are based on spatial access methods and assume that all image features are numeric and represent images as points in multidimensional space. However, not all image features are numeric. Novel access structures and methods must be devised.

Eliciting relevance feedback and improving retrieval effectiveness

Since subjectivity and imprecision are usually associated with specifying and interpreting subjective attributes, the query processor should be designed to deal interactively with these problems at the query specification or processing time.[7] The query interface may be designed to function as a knowledge-based system to guide users through the query-specification process and to facilitate user-relevance feedback and incremental query reformulation. User involvement in providing the relevance feedback should be at a conceptual level. That is, users should not be forced to explain feedback in terms of low-level image features. For example, the user may provide relevance feedback by simply labeling a set of retrieved images relevant, nonrelevant, or somewhat relevant. Although there appears to be no research in this direction, Jung and Gudivada[7] is a starting point.

Distributed aspects

With the existence of the information superhighway, image repositories are likely to exist and evolve in a decentralized fashion. Methods must be investigated to provide network-transparent access to such distributed image collections. This aspect does not seem to have been investigated.

Tools for feature extraction and semantics capture

Easy-to-use automated tools for extracting primitive features and semiautomated tools for capturing logical features are indispensable for developing large, realistic image retrieval applications. Comprehensive tools are yet to emerge.

Performance and usability issues

Three aspects of interest in this context are effectiveness, efficiency, and usability. *Effectiveness* is a measure of the relevance of retrieved images to a query as perceived by a user. System responsiveness and interactivity for various user queries are measured by the *efficiency* factor. *Usability methods* are useful for measuring and evaluating human performance and preference in using the CBIR system. Suitable metrics should be developed to characterize these factors.

IN THIS ISSUE

This theme issue provides a sampling of current CBIR research. Its five articles collectively address various CBIR issues. "Query by Image and Video Content: The QBIC System" by Flickner et al. provides a good introduction to the subject. QBIC is a comprehensive, operational CBIR system based on the a priori feature extraction approach. The features are extracted semiautomatically. The mature QBIC system allows queries on large image and video databases based on example images, user-constructed sketches and drawings, color and texture patterns, and camera and object motion. The ideas presented here will be useful to practicing engineers interested in developing CBIR systems.

Brink, Marcus, and Subrahmanian's "Heterogeneous Multimedia Reasoning" addresses CBIR issues in a broader context and proposes certain principles around which multimedia database systems can be built in a domain-independent fashion. The authors propose a framework based on the notion of media abstraction, which plays a pivotal role in defining access structures, query language, and query-processing algorithms. Their MACS system is an implementation of this framework and provides a stand-alone basis for querying multimedia databases. The authors also illustrate how MACS can be expanded to access a variety of databases, data structures, and general software packages using Hermes, a framework for general problem solving. This system is also based on the a priori feature-extraction approach through manually extracted features.

In the third article, "Chabot: Retrieval from a Relational Database of Images," Ogle and Stonebraker report on a project that integrates a relational database system with content analysis techniques. The system design is based on the a priori feature-extraction approach, with features extracted automatically. This article demonstrates how text-based search criteria can combine with content-based criteria to achieve good retrieval results. In the current version, content analysis is primarily based on color histograms, and query-optimization techniques are used to minimize the number of histograms considered when a query is processed. The Chabot system is designed to eventually manage an image database comprising over 500,000 digitized multiresolution images.

The data model should feature a rich set of modeling constructs to capture the information necessary for processing query classes.

In "Automatic Indexing and Content-Based Retrieval of Captioned Images," Srihari examines the role of text as collateral information for image feature extraction and content-based retrieval. She describes the Piction system, which uses the photograph caption to identify human faces in the photograph. Piction illustrates how information from both text and images can be used to compute similarity between a query and database images. The system is based on the a priori feature-extraction approach with automatically extracted features.

Finally, Mehrotra and Gary's "Similar-Shape Retrieval in Shape Data Management" examines the key issues involved in rigid or articulated shape descriptions, similarity measures, and indexing schemes. It describes a technique for matching similar shapes and an associated prototype system to demonstrate how key issues are resolved in designing a typical shape retrieval system.

WE HOPE THAT THIS ISSUE succeeds in exposing readers to current developments in CBIR and helps stimulate further research in this scientifically exciting and commercially important field. ∎

Acknowledgments

The initial impetus for our CBIR research came from the US Army Research Office (ARO) through Grant No. DAAL03-89-G-0118. We acknowledge the support of ARO and thank David Hislop for his encouragement. We also thank the article reviewers for their professionalism and insightful, detailed reviews, and for responding under stringent time constraints. Ted Lewis, Ed Parrish, and Joan Francioni are acknowledged for their timely help in guiding us through the steps involved in bringing out this theme issue.

References

1. W. Grosky and R. Mehrotra, Guest Editors, Special Issue on Image Database Management, *Computer*, Vol. 22, No. 12, Dec. 1989.
2. R. Jain, "NSF Workshop on Visual Information Management Systems," *SIGmod Record*, Vol. 22, No. 3, Dec. 1993, pp. 57-75.
3. A.D. Narasimhalu, Guest Editor, Special Issue on Content-Based Retrieval, *ACM Multimedia Systems*, Vol. 3, No. 1, Feb. 1995.
4. A. Ralescu and R. Jain, Guest Editors, Special Issue on Advances in Visual Information Management Systems, *J. Intelligent Information Systems*, Vol. 3, No. 3, July 1994.
5. V. Gudivada, V. Raghavan, and K. Vanapipat, "A Unified Approach to Data Modeling and Retrieval for a Class of Image Database Applications," in *Multimedia Database Systems: Issues and Research Directions*, S. Jajodia and V. Subrahmanian, eds., Springer-Verlag, New York, 1995.
6. G. Salton, *Automatic Text Processing*, Addison-Wesley, Reading, Mass., 1989.
7. G. Jung and V. Gudivada, "Adaptive Query Reformulation in Attribute-Based Image Retrieval," in *Third Golden West Int'l Conf. Intelligent Systems*, Kluwer Academic Publishers, Boston, Mass., 1994, pp. 763-774.

Venkat N. Gudivada *is an assistant professor of computer science in the School of Electrical Engineering and Computer Science at Ohio University. His research interests are in content-based retrieval of multimedia data with an emphasis on data modeling, efficient processing of spatial similarity queries, distributed information retrieval, and relevance feedback for query reformulation and effective retrieval.*

Gudivada completed a B.Tech degree in civil engineering at Jawaharlal Nehru Technological University, an MS degree in civil engineering at Texas Tech University, and MS and PhD degrees in computer science at the University of Southwestern Louisiana. Vijay Raghavan and Gudivada are coauthoring a book, Image and Video Database Management: Data Models, Query Interfaces, and Search Algorithms, *to be published by Cambridge University Press in 1996. Gudivada is a member of the IEEE Computer Society and ACM.*

Vijay V. Raghavan *is a Distinguished Professor of Computer Science at the University of Southwestern Louisiana. He previously taught at the University of Regina and St. Mary's University, both in Canada. Raghavan specializes in designing effective retrieval strategies to improve retrieval system performance by utilizing prior system-user interactions. The application domains include both text and image databases. In the context of traditional business databases, he has developed approaches for multiple query processing, index selection, and database mining.*

Raghavan completed a B.Tech degree in mechanical engineering at the Indian Institute of Technology, Madras, an MBA degree at McMaster University, and a PhD degree in computer science at the University of Alberta. Raghavan cochaired the NSF Workshop on Future Directions in Information Retrieval in 1991. He has served on numerous conference committees and is currently an ACM National Lecturer. He is a member of the ACM and the IEEE.

Readers can contact Venkat N. Gudivada at the School of Electrical Engineering & Computer Science, Ohio University, Athens, OH 45701.