RAJIV MEHROTRA

Content-Based Image Modeling and Retrieval

A database system models and manages an abstracted real world (or mini-world) pertinent to the problem at hand in terms of alphanumeric data. The semantic associated with any piece of alphanumeric data is known to or derived by the users of the database. This conventional approach to data modeling and management is not well suited for the effective management of imagery data. In an image database management system, the desired information/semantics associated with the imaged mini-world needs to be automatically (or semi-automatically) extracted and appropriately modeled to facilitate content-based retrieval and manipulation of data. In this article, the key issues in content-based image data modeling and retrieval are discussed. A system called MUSEUM is briefly presented to illustrate some of the approaches used to resolve the main challenges of conent-based data modeling and retrieval.

INTRODUCTION

Several application areas have emerged that require effective and efficient management of image data. Examples of such application areas include digital library, medicine, defense, space exploration, law enforcement, environmental monitoring and control, museum or historic collection management, electronic publishing/advertising, education, and entertainment. In most of these areas, several large repositories of image data already exist and only a very small fraction of collected data is ever analyzed due to the lack of effective image database management techniques. The growing list of applications combined with the advances in the areas of image analysis and database management caused an ever increasing interest in image database systems over the last decade. The major hurdle was posed by the memory and the computational requirements of an image database management system. The common saying, "an image is worth a thousand words" turned out to be an understatement. In recent years, rapid improvements in computer hardware, memory management, and display devices have made it feasible to develop practical and user-oriented image database management systems.

The management of image databases involves a close interaction of database and machine vision technologies. Unfortunately, until recently, almost all reported efforts in the development of image database management technology did not consider this close interaction. As a result, some of the key issues were ignored by both the scientific communities resulting in systems with very little practical applications. Most of the proposals from the database community were extensions of the conventional database model that treated images as an appendix to the alphanumeric data. In these systems, pointers (or references) to images are allowed as attributes and very limited processing and analysis of image contents is involved. In other words, images and alphanumeric data are not treated equally since image data-based constraints cannot be employed for data retrieval/selection (Grosky & Mehrotra, 1992). However, some of these efforts assumed that the machine vision community would provide the desired content-based image processing and analysis methods (Aslandogan et al., 1995). The image analysis activities of the machine vision scientific community completely ignored the database-related issues (e.g., image representation, analysis, and recognition in a large and flexible database environment). It is now clear that conventional approaches to database management and image analysis are not well suited to the management of image and other nonalphanumeric data (Grosky & Mehrotra, 1992; Grosky, 1994). The key challenges are posed by the contents of the images to be managed. Image databases are of little use without content-based image data description and retrieval. In the following sections, the challenges posed by the imagery data from the viewpoints of data modeling (description) and data retrieval are discussed.

CONTENT-BASED IMAGE DATA MODELING

A database represents an abstracted real world (mini-world) pertinent to the problem at hand in terms of its entities and relationships. Every piece of datum in a database conveys some application domain-dependent information (or semantics). In a traditional database, the information about the modeled mini-world conveyed by a piece of alphanumeric data is known to, or derived by, the users. In an image database, raw images by themselves are of limited use unless the embedded application- or user-dependent semantics can be somehow extracted and used in image data retrieval and manipulation. In other words, information about the imaged mini-world contained in images needs to be extracted and appropriately modeled in the database (Grosky, 1994; Grosky & Mehrotra, 1992; Gupta et al., 1991). Therefore, an image database management system must be capable of representing images in terms of their contents (image properties, objects and their attributes, and relationships among objects) and the associated application- and user-dependent semantics and knowledge. From the modeling viewpoint, the content-based image databases can be broadly classified into two groups:

1. *Mini-world associated with images is known*—In an image database of this type, the images are of a known (or fixed) mini-world. In these cases, the objects, scenes, events, and visual concepts that can appear in an

image are known a priori. There is usually only one application domain-dependent interpretation of each database image. Therefore, the contents of any database image can be represented by a predetermined modeling scheme. For some application domains, model-based techniques can be employed to extract (automatically or semi-automatically) the desired content-based image representation and to process content-based image retrieval queries. An image database for a manufacturing application (containing images of parts, components, tools, machineries, and products) is an example of such a database.

2. *Unknown and variable mini-world*—The images in such a database do not belong to any fixed mini-world and there is no a priori knowledge about the objects, scenes, and events that can appear in images. A database of a family's picture collection, a database of an explorer's image collection, or a database of images of a museum's collection are examples of such image databases. In such a database, most images are very different from other database images in terms of their contents. Therefore, in general, a predefined set of objects and relationships cannot be used to describe all the database images. Also, model-based image processing and analysis approaches are not directly applicable. Instead, capabilities to dynamically describe (or associate a mini-world) with each of the database images and to manipulate these descriptions is essentially required. In such a database, image processing and analysis methods are needed to interactively or automatically develop models for objects, events, and scenes found in the user-defined mini-world of an image. Such models can then be used in a model-based approach to partially and fully represent other images in terms of previously modeled objects, events, and scenes and to facilitate content-based image information manipulation and retrieval.

In general, the interpretation of an image or a visual concept (e.g., beautiful or serene) may vary from user to user. Therefore multiple user- or application-dependent mini-worlds (interpretations) can be associated with each of the database images. Each user can be characterized by the collection of his/her descriptions (or models) of objects, events, scenes, and concepts. Such user profiles can be effectively used to develop user-oriented descriptions of database images and to process a user's queries in accordance with that user's profile. For example, a user's definition of the concept "colorful" should be utilized to respond to his/her queries like, "retrieve all colorful pictures and retrieve database images at least as colorful as a given query image." If the definition of "colorful" does not exist in that user's profile, then the definition provided by another user, whose profile is found to be most similar to the current user, can be selected to create the corresponding response.

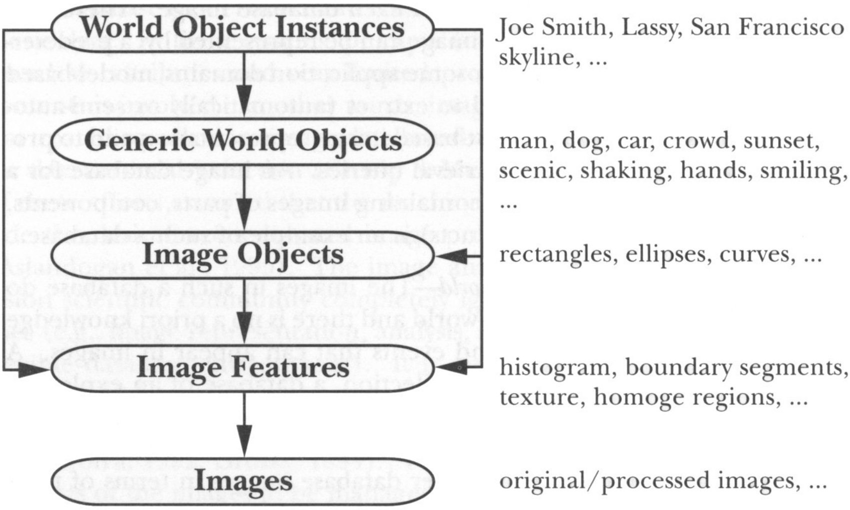

Figure 1: Levels of Abstraction in MUSEUM

It should be clear from the above discussion that image contents pose several challenges with respect to the representation or the modeling of images. We are currently developing an image database system which utilizes a data model called MUlti-SEmantic Unstructured data Model (MUSEUM). MUSEUM resolves the above-mentioned modeling issues and it has the following key features:

- a generalized object-oriented model capable of representing both the structured data (i.e., images with a known mini-world) as well as the unstructured data (i.e., images where the mini-world is dynamically associated);
- ability to dynamically associate, compose, and modify data (image) description;
- flexibility of switching from one view of an image, a group of images, or the entire database, to the other and to simultaneously access and manipulate multiple views of an image, group of images, or the entire database; and
- effective management of user profiles derived from their definitions of abstract concepts and descriptions of mini-worlds.

In MUSEUM, database images and visual concepts are described using a multilevel abstraction hierarchy. The main levels of abstraction are shown in figure 1. At the lowest levels are database images or example images. At the next level of description, an image is characterized in terms of its properties such as background/foreground colors, dominant colors, histograms, and texture properties. Description of images in terms of objects—such as image regions, boundary segments, and contours—and relationships among them forms the next level of abstraction. At the next level of abstraction, images are described in terms of generic objects, relationships, and concepts such as man, dog, car, crowd, horizon, sunset, cloudy, colorful, and smile. At the highest level of abstraction, images are described in terms of specific instances of the generic world objects. For example, a man may be described as Joe Smith, a dog may be described as Lassie, an image may be described as the San Francisco skyline. The image descriptions at any of these abstraction levels can be multilevel and can be derived from—or mapped to—the descriptions at the lower levels of abstraction. In MUSEUM, this multilevel description of images is composed of two parts—i.e., mandatory part and optional part. The mandatory description components are found in all database images, whereas the optional description components are image and/or user-dependent. The mandatory and optional description components are dependent on the nature of the image database. For example, descriptions of images of a fixed mini-world image database may not have any optional description components.

CONTENT-BASED IMAGE RETRIEVAL

The central task of any database management system is to retrieve records/objects that satisfy a set of specified constraints. In image databases, an important class of data retrieval is content-based retrieval of images. In content-based retrieval, images whose contents satisfy the specified constraints are retrieved or selected. Content-based image retrieval queries can be classified into two broad classes:

1. *Queries involving no image processing/analysis*—in these queries, no processing or analysis of database images is required and no query images are given. Examples are: (1) retrieve all images containing at least one automobile in front of a house, (2) retrieve pictures containing a smiling man. The symbolic descriptions (automatically extracted and/or user specified) associated with database images are used to select the desired images. These queries can be processed using traditional approaches.
2. *Queries involving image processing/analysis*—these queries involve one or more images that are processed to extract the associated desired symbolic information. The extracted description is compared against the description of database images to select images that satisfy the

specified constraints. Examples of such queries are: (1) retrieve all images containing one or more objects similar to the object in a given query image, (2) retrieve all images that are similar to a given query image in terms of image color and texture features.

To efficiently process content-based image retrieval queries, various levels of descriptions of database images need to be organized in efficient secondary storage-based index structures. These indexes are searched to find descriptors and hence images that satisfy the specified constraints. Complex queries can be efficiently processed using the incremental refinement process. An incremental refinement process starts by selecting images that satisfy a subset of the specified constraints. This initial response is refined in several stages until all the remaining constraints are satisfied. For example, consider a facial image database. To retrieve facial images similar to a query facial image, first the nose of the query face nose must be used to select facial images with a similar nose. This initial response can be refined by selecting other features in the query face one by one—e.g., hair region, mouth, eyes, and so on. The user can review the response at each stage of refinement and select the query feature for the next stage of refinement or elect to terminate the refinement process. Query processing by incremental refinement can be used with a multiresolution image representation scheme. In this case, the initial response to a query can be generated using a coarse representation, and this response can be refined using finer (less lossy) representations.

To illustrate the key steps involved in the design of a content-based image retrieval system, we consider the problem of shape similarity-based image retrieval (Flickner et al., 1995; Gary & Mehrotra, 1995; Grosky & Mehrotra, 1990; Jagadish, 1991; Petland et al., 1994). In shape similarity-based queries, constraints are therefore specified in terms of similarity of shapes. An example is, "retrieve images that contain at least one shape similar to the given query shape." In this case, the key issues to be resolved are:

- *Shape representation*—how can the shapes present in an image be represented? How can the selected representation be extracted automatically or semiautomatically from images?
- *Shape similarity definition*—what criteria or measures should be used to automatically determine the similarity or dissimilarity of two shapes? The similarity measure should be consistent with the human interpretation of shape similarity.
- *Access or index structures*—how should shapes and related representations be organized to enable efficient searches for shapes that satisfy the specified shape similarity-based constraints?

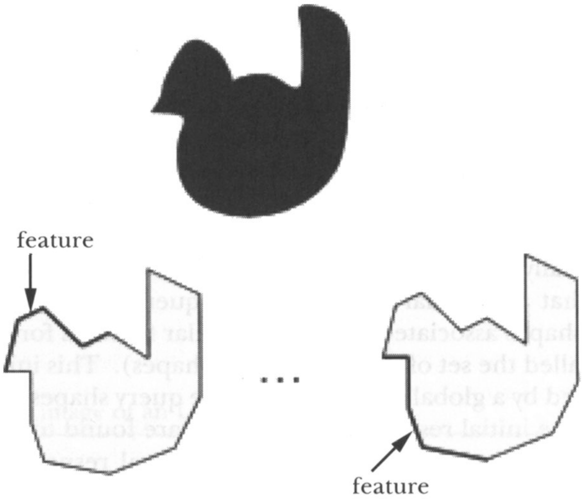

Figure 2. A shape and its structural feature.

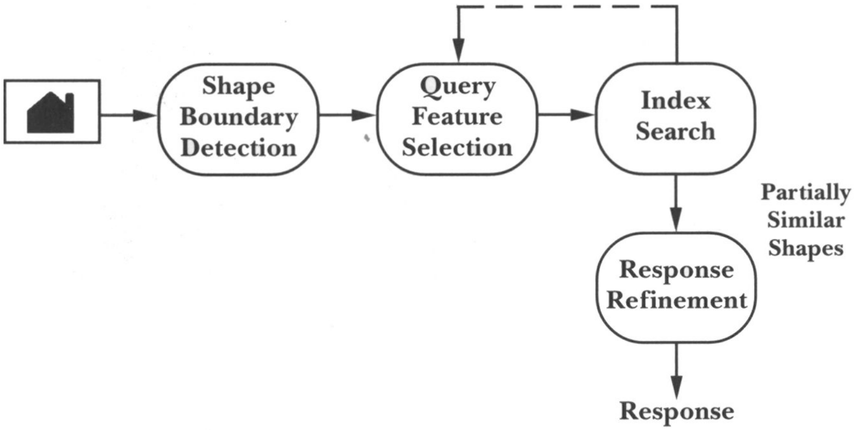

Figure 3: Two-dimensional shape similarity-based query processing

In MUSEUM, a two-dimensional (i.e., almost flat) object image is represented by an ordered set of boundary points (e.g., maximal curvature points or vertices of the polygonal approximation). Each shape is further represented by a set of structural features, which is a fixed size set of adjacent points (or line segments) of its representation. A shape and its structural features are shown in figure 2.

Each structural feature is represented as a point in a multidimensional space. Similarity between two structural features is measured by the Euclidian distance between the corresponding points in the multidimensional space. Any multidimensional point access method (Nievergelt et al., 1984, p. 84; Robinson, 1981; Seeger & Kriegel, 1990) can be used to organize the structural features of all the shapes in the database. Associated with each structural feature is a list containing information about where, and in which shape, that structural feature appears. The key steps involved in processing a shape-similarity-based query are shown in figure 3.

The boundary-based query shape representation is first developed, and then a structural feature, called the query feature, is selected (automatically or by the user). The index is searched to find structural features that are similar to the selected query features, and the list of database shapes associated with these similar features form the initial response (called the set of partially similar shapes). This initial response is then refined by a global comparison of the query shapes with each of the shapes in the initial response. Shapes that are found to satisfy the specified global similarity constraints form the final response to the query. Further details of this technique can be found in Gary and Mehrotra (1993) and Mehrotra and Gary (1995).

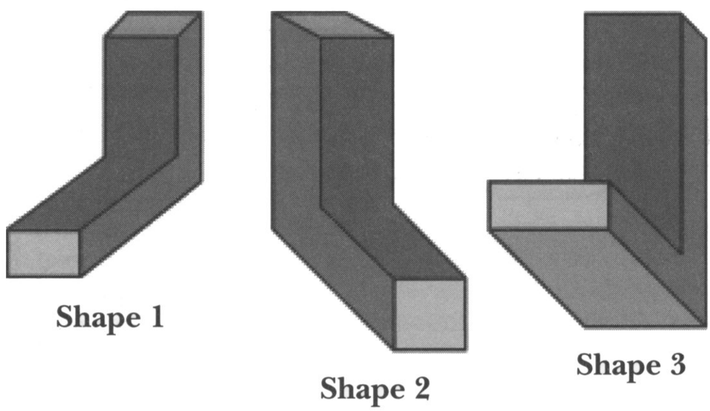

Figure 4. Qualitatively Similar and Different Three-Dimensional Shapes

MUSEUM also supports shape-similarity-based retrieval of images of three-dimensional objects (Mehrotra & Gary, 1996). Similarity of images of three-dimensional shapes is determined by their qualitative appearances in their respective image. Qualitative appearance of a three-dimensional object in an image is defined by the visible surfaces (or faces) and their qualitative characteristics. For example, in figure 4, shapes 1 and 2 are considered qualitatively similar as they have qualitatively similar visible surfaces. Shape 3 is considered to be qualitatively different from the other two.

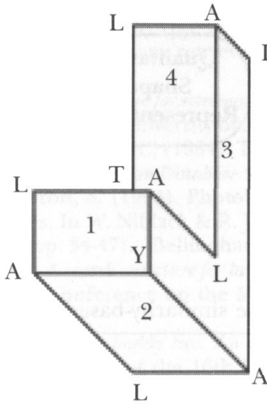

Figure 5. Labeled image of an L-shaped object

In MUSEUM, an image of a three-dimensional shape is represented by a character string composed of substrings, each representing the qualitative appearance of a region in the image in terms of its vertex types. Figure 5 shows an image of an L-shaped object with labeled vertices. The qualitative appearance of this shape is represented by the character string

ALALALTL:ALTAY:AYAL:AYALAL:ALATL,

where ":" (colon) is the region string separator. In this string, the first substring (i.e., the character string before the first ":" character) represents the silhouette region, and the following four substrings respectively represent regions labeled 1, 2, 3, and 4.

Two shape images with the same character string representation are considered to be qualitatively similar. If only some of the leading substrings match, the corresponding shapes are considered to be partially similar. The degree of similarity is determined by the number of matching leading substrings. In this case, a suitable extension of any efficient string-matching index structure can be used to organize the database of shapes and their representations.

A shape-similarity-based query is processed in two stages as shown in figure 6. In the first stage, the query shape representation is used to search the index structures to find the string with the most number of matching leading substrings. The database shapes associated with this string form the initial response. Then if requested, this initial response is refined by retaining shapes having character string representations that completely match the query shape character string.

Note that there is always a loss of information in a content-based representation of images. There is a trade-off between the memory and computational requirements and loss of information. The memory and

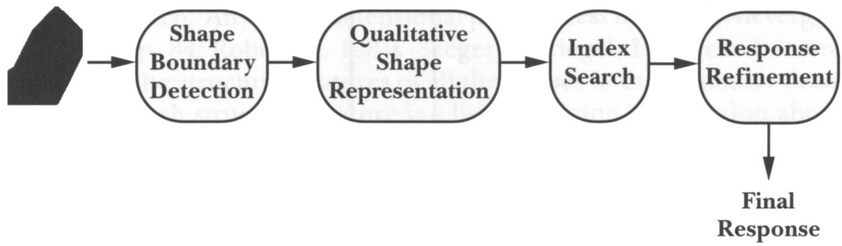

Figure 6. 3-dimensional shape similarity-based retrieval

computational requirements are higher for finer representation schemes. The quality of image representation directly determines the query response quality (i.e., number of images in a query response that are not consistent with the user's interpretation). Multiresolution image representation combined with query processing by incremental refinement provides a scheme in which the trade-off between the response quality and query processing time can be controlled by the users.

CONCLUSION

Recent advances in several digital computation technologies have made it possible to store and manage large repositories of imagery data. In recent years, it has become clear that such image databases are useful only if the database management schemes permit content-based retrieval of images. Image content and associated interpretations pose several challenges. In this discussion, key database design issues pertinent to content-based modeling (representation) and retrieval of images have been reviewed.

REFERENCES

Aslandogan, Y. A.; Thier, C.; Yu, C. T.; Liu, C.; & Nair, K. R. (1995). Design, implementation, and evaluation of SCORE. In *Proceedings 1995 IEEE Data Engineering Conference* (pp. 280-287). Taiwan.

Flickner, M. et al. (1995). Query by image and video content: The QBIC system. *IEEE Computer, 28*(9), 23-32.

Gary, J. E., & Mehrotra, R. (1993). Similar shape retrieval using a structural feature index. *Information Systems, 18*(7), 525-537.

Grosky, W. I. (1994). Multimedia information systems—A tutorial. *IEEE Multimedia, 1*(1), 12-24.

Grosky, W. I., & Mehrotra, R. (1990). Index-based object recognition in pictorial data management. *Computer Vision, Graphics, and Image Processing, 52*(3), 416-436.

Grosky, W. I., & Mehrotra, R. (1992) Image database management. In M. C. Yovits (Ed.), *Advances in Computers* (vol. 35, pp. 237-253). New York: Academic Press.

Gupta, A.; Waymouth, T.; & Jain, R. (1991). *Semantic queries with pictures: The VIMSYS model* (Proceedings of the 17th International Conference on Very Large Databases) (pp. 69-79). Barcelona, Spain.

Jagadish, H. V. (1991). *A retrieval technique for similar shapes* (Proceedings ACM SIGMOD Conference on the Management of Data) (pp. 208-217). Denver, Colorado.

Mehrotra, R., & Gary, J. E. (1995). Similar-shape retrieval in shape data management. *IEEE Computer, 28*(9), 57-62.

Mehrotra, R., & Gary, J. E. (1996). *A technique for retrieval of similar images of three-dimensional objects.* Unpublished technical report, University of Missouri—St. Louis.

Nievergelt, J.; Hinterberger, H.; & Sevcik, K. C. (1984). The grid file: An adaptable symmetric multikey file structure. *ACM Trans. on Database Systems, 9*(1), 38-71.

Petland, A.; Picard, R. W.; & Sclaroff, S. (1994). Photobook: Tools for content-based manipulation of image databases. In W. Niblack & R. Jain (Eds.), *Storage and retrieval for images and video databases II* (pp. 34-47). Bellingham, WA: SPIE.

Robinson, J. T. (1981). *K-D-B-tree: A search structure for large multidimensional dynamic indices.* (Proceedings ACM SIGMOD Conference on the Management of Data) (pp. 10-18). Ann Arbor, MI.

Seeger, B., & Kriegel, H. P. (1990). *The buddy tree: An efficient and robust access method for spatial database systems.* (Proceedings of the 16th International Conference on Very Large Databases) (pp. 590-601). Brisbane, Australia.