NSF Workshop* on Visual Information Management Systems

Editor: Ramesh Jain
Electrical and Computer Engineering
University of California at San Diego
La Jolla, CA 92093-0407

1 Executive Summary

One of the most important technologies needed across many traditional areas as well as emerging new frontiers of computing, is the management of visual information. For example, most of the Grand Challenge applications, under the High Performance Computing and Communication (HPCC) initiative, require management of large volumes of non-alphanumeric information, computations, communication, and visualization of results. Considering the growing need and interest in the organization and retrieval of visual and other non-alphanumeric information, and in order to stimulate academic projects in this area, a workshop on Visual Information Management Systems (VIMS) was sponsored by the National Science Foundation. This workshop was held in Redwood, CA, on February 24-25, 1992. The goal of the workshop was to identify major research areas that should be addressed by researchers for VIMS that would be useful in scientific, industrial, medical, environmental, educational, entertainment, and other applications.

The major findings of the workshop were that VIMS require new techniques in all aspects of databases, computer vision, and knowledge representation and management; and that such techniques are best developed in the context of concrete, practical applications. VIMS will provide impetus and testbeds for many techniques being explored for the future database systems. Researchers from image processing and understanding, knowledge representation and knowledge based systems, and databases must work very closely to develop VIMS. Such systems should be developed in the context of applications that will be of immediate interest in industrial, medical, or scientific contexts. Without concrete applications and ambitious implementation projects, most of the important and difficult issues are likely to be ignored. Considering the interdisciplinary nature of the research in this area, a few major research projects in this area are essential for its growth. Increased emphasis on HPCC by many Federal agencies can help in the rapid development of VIMS technology. Similarly, by addressing some of the Grand Challenges, researchers interested in VIMS can understand critical issues and develop techniques to solve them, in a concrete and useful context.

Parallel processing is essential for implementing VIMS. As is well known, the processing of images is one of the most computation-intensive tasks. For entering images in databases and for performing required operations at query time, an enormous volume of data must be processed. Parallel computing will be essential for implementing a VIMS that can insert images in reasonable time and provide fast response to user queries. The computational requirements of video databases are likely to be one of the most demanding. It is very likely that video databases will require research in highly parallel-pipelined architectures.

In interdisciplinary research areas such as VIMS, most important and difficult problems usually fall through the cracks. The three most relevant areas for the development of VIMS are: databases, computer vision, and knowledge representation. Data compression, fault-tolerant real time access to image data through networks, and parallel processing issues should be addressed in the context of databases for VIMS. VIMS should not be considered as an application of the existing state of the art in any

*This workshop was supported by Database and Expert Systems, and Robotics and Machine Intelligence Programs of the Division of Information, Robotics, and Intelligent Systems, in the Directorate for Computer and Information Science and Engineering of National Science Foundation. The workshop participants were: Bruce Berra (Syracuse), Alfonso F. Cardenas (UCLA), Shi-Kuo Chang (U.Pittsburg), Richard Cicone (CIESIN), Edward Fox (VPI), William Grosky (Wayne State), Allen R. Hanson (U.Mass), James Hollan (Bellcore), Arding Hsu (Siemens), Hiroshi Iwamoto (LSU), S.Sitharama Iyengar (LSU), H. V. Jagadish (AT+T), Ramesh Jain (U.Michigan; now UCSD), Rangachar Kasturi (Penn State), Guy Lohman (IBM), Patrick E. Mantey (UCSC), David McKeown (CMU), Rajiv Mehrotra (U.Kentucky), Howard Moraff (NSF), George Nagy (RPI), Makishi Nakayama (LIFE), Marie-Anne Neimat (H.P.), Wayne Niblack (IBM), Dragutin Petkovic (IIBM), Arturo Pizano (Ricoh), Anca Ralescu (LIFE and U.Cincinnati), Ramin Samadani (Stanford), Hanan Samet (U.Maryland), Jorge L. C. Sanz (IBM) Linda Shapiro (U.Washington), Elliot Soloway (U.Michigan), Richard Stamm (MasPar), Michael Stonebreaker (Berkeley), Hemant D. Tagare (Yale), George Thoma (NIH), Sam Thompson (ERIM), Terry Weymouth (U.Michigan), Gregory W. Zack (Xerox), Maria Zemankova (NSF). The views expressed in this report represent the opinions of the participants, not necessarily those of the National Science Foundation. A copy of the complete report can be obtained by writing to the editor of the report.

of these fields to manage and process images. Database researchers must understand the issues specific to managing and processing images and other forms of data by granting them the same status that has been given to alphanumeric information. Computer vision researchers should identify features required for interactive image understanding, rather than their discipline's current emphasis on automatic techniques, and develop techniques to compute features in interactive environments. Most knowledge representation research has been concerned with symbolic knowledge. For VIMS and HPCC applications, techniques for representing symbolic and non-symbolic representations at the same level will be required. Reasoning approaches that can deal with such representations will be useful not only in VIMS, but in many other applications also. Finally, performance issues pose a significant challenge in all aspects of VIMS, from memory organization to information retrieval.

2 Preamble

One of the most important technologies needed across many traditional and emerging applications is the management of visual information. Every day we are bombarded with information presented in the form of images. So important are images in our world of information technology, that we generate literally millions of images every day, and this number keeps escalating with advances in imaging, visualization, video, and computing technologies. Advances in video technology and its marriage with computing are resulting in the video-computing discipline. High Performance Computing and Communications (HPCC) is emerging as a key technology for asserting international leadership in industrial, medical, scientific, defense, and environmental areas. Improved computational methods and information management tools are critical in order to enhance the national competitive edge across broad sectors of the economy.

It would be impossible to cope with this explosion of image information, unless the images were organized for rapid retrieval on demand. A similar situation occurred in the past for numeric and other structured data, and led to the creation of computerized database management systems. In these systems, large amounts of data are organized into fields and important or *key* fields are used to index the databases making search very efficient. These information management systems have changed several aspects of the modern society. These systems, however, are limited by the fact that they work well only with numeric data and short alpha-numeric strings. Since so much information is in non-alphanumeric form (such as images, video, speech), to deal with such information, researchers started exploring the design and implementation of image databases. But creation of mere image repositories is of little value unless there are methods for fast retrieval of images based on their content, ideally with an efficiency that we find in today's databases. We should be able to search image databases with image-based queries, in addition to alphanumeric queries. The fundamental problem is that images, video, and other similar data differ from numeric data and text in format, and hence they require a totally different technique of organization, indexing, and query processing. We need to consider the issues in visual information management, rather than simply extending the existing database technology to deal with images. We must treat images as one of the central sources of information rather than as an appendix to the main database.

A few researchers have addressed problems in image databases (see the bibliography). Most of these efforts in image databases, however, focussed either on only a small aspect of the problem, such as data structures or pictorial queries, or on a very narrow application, such as databases for pottery articles of a particular tribe. Other researchers have developed image processing *shells* which use several images. Clearly, visual information management systems encompass not only databases, but aspects of image processing and image understanding, very sophisticated interfaces, knowledge-based systems, compression and decompression of images. Moreover, memory management and organization issues start becoming much more serious than in the largest alphanumeric databases. In failing to address any of these topics, one may either address only theoretical issues, or may work in a microcosm that will, at best, be extremely narrow in its utility and extensibility.

It is clear that the tremendous progress in processing speed and memory technology has made it not only possible, but also attractive to design Visual Information Management Systems (VIMS) for many disparate applications. People already call the 90s *the decade of imaging.* On considering any of the Grand Challenge problems, such as weather forecasting, air pollution, the earth's biosphere, genome research, or the education network, it becomes clear that the existing database technology must be extended in several new dimensions, from managing tertiary memory to representing an object at varying degrees of detail. Many of the current issues in databases, such as interconnecting heterogeneous databases, are also important to VIMS. Moreover, VIMS have several of their own problems that must be addressed for making progress in challenging industrial and medical applications.

Considering the growing need and interest in the organization and retrieval of visual and other non-alphanumeric information, and the insufficient number of academic projects in this area, a workshop on visual information management systems was sponsored Robotics and Machine Intelligence, and Database and Expert Systems Programs of the National Science Foundation. The aim of the workshop was to bring together active researchers in databases, object- oriented systems, image and signal processing, multi-media, and other related areas to discuss important issues in managing the large amount of visual information that will play a key role in designing information systems of the future. In addition to the researchers in the above and related areas, a few researchers and practitioners interested in applying these systems were also invited. The participants were drawn from academic institutions, research centers, and industry. More details of the workshop are provided in the report. The goal of the workshop was to identify major research areas that should be addressed by researchers for VIMS useful in scientific, industrial, medical, environmental, education, entertainment, and other applications.

2.1 Format of the Report

This report is the final result of the workshop. Sections 3-8 present group reports. Each group report first identifies the major research issues in that area and then gives its recommendations for research to advance theory and practice of VIMS. The conclusion section was prepared by the editor based on group reports and discussions at the workshop. Finally, we decided to include a bibliography that contains papers related to the theme of the workshop. This bibliography was prepared by Jeff Bach, Ramesh Jain, and Deborah Swanberg of the University of Michigan.

A copy of the complete report can be obtained by writing to the author of this paper.

3 Applications

From among many potential applications areas for visual information management systems, the Applications group identified four Grand Challenge application domains: Geographic/Environmental Information Systems, Engineering/Scientific Visualization Systems, Medical Information Systems, and Education Systems. Each of these domains provide numerous technical challenges to design, develop and evaluate various critical components of the sophisticated visual information management systems of tomorrow. These domains have large distinct user communities, and represent many of the characteristics which are unique to VIMS. These four application areas and their features are described in the Appendix A. Here we give a brief idea of these challenges.

- Modern visual technology can provide powerful mechanisms for educating our population in a cost-effective and exciting way. Under the "Educational Grand Challenge," we recommend building a nationwide educational network that will provide a visual repository of the best available lectures, videos, interactive classes, etc. to all Americans at a moderate cost. By providing powerful interfaces for navigating and browsing through this "encyclopedia", education can be made fun for most of the population.

- Increasing engineering productivity is a critical need for American global competitiveness. Much of the information that engineers deal with in the design and testing of new items is visual. Drawings and other documents, as well as newer types of visual information are rapidly gaining importance as computational prototyping comes into widespread use. This prototyping requires computer images as input, often in the form of solid CAD models, and generates images as output, in the form of visualization of system behavior and of performance data from system simulations. The role of VIMS becomes more significant when one considers the fact that in most cases a team of engineers must work on a design project.

- Medicine is a domain in which images are critical elements of the diagnostic process. Due to the extensive use of imaging methods in modern medicine, Picture Archiving and Communication Systems (PACS) have become very attractive. These systems use only header information for retrieval of images. VIMS would allow users to retrieve the following information:
  - Where was the tumor initially [detection, location]?
  - What happened to the tumor [temporal analysis]?
  - Assessment of the tumor [detect and delineate tumor bulk; is it benign or malignant]?
  - Assessment of the therapeutic response [did the therapy reduce the tumor]?

- Satellites have been recording several dozen images per day, and each image may contain more than 100 megabytes of data. NASA's Earth Observing System will gather an unprecedented amount of data. In response to this, universities and research centers will be required to generate and maintain appropriate visual information management systems for this geographical/environmental data. In the public sector, the scope of such information systems encompasses the global interests of military and supra-national government organizations; country-wide concerns such as federal agriculture, transportation, weather, mapping, and census departments; the planning and record-keeping operations of regional agencies; and the public safety, transportation, and revenue-raising operations of individual municipalities. In the private sector, interested parties include mineral exploration and exploitation companies, shipping companies (trucking and railroad), communications (telephone), and public utilities (gas and electric).

3.1 Technical Challenges

What is important from a VIMS design point of view is that these systems share many of the following characteristics:

- Large Volumes of Complex Objects: Each of the objects that is represented and managed in a VIMS is typically a complex entity requiring large storage and processing resources. A typical object, for example, would be an image or a sequence of images with their associated information. A typical VIMS is expected to contain a large collection of such complex objects. Storage (tertiary), retrieval and transfer rate, and automated insertion of new objects are critical issues in such a system. In particular, data objects for continuous media, such as video, are common in VIMS. There is very little experience in dealing with such data objects in databases. The problems posed by these data sets are among the most challenging.

- Read/Write Operations: READ operations are generally more frequent than WRITE operations as in any database system; users extracting information from the system are usually different from those who enter and update information. In some applications, such as those with continuous image streams, real-time processing to keep up with a large rate of data input-output becomes a critical issue. These streams present a different set of issues to be addressed by databases. With these data sets, the system has to deal with large arrays that are strictly time ordered. There is very little experience in dealing with such objects in databases.

- Feature Extraction and Automated Indexing: In a VIMS system it is essential to at least partially automate data capture operations. This requires data preprocessing and automated indexing. Automated indexing requires automated (or operator-assisted) feature extraction and (potentially semantic) interpretation. Feature extraction and interpretation are also essential for many content-addressed database operations.

- Data Representation and Organization: In many of the application domains for VIMS, data must be represented in multiple scales to cater to the needs of users requiring different levels of abstraction. If data is not explicitly represented at multiple scales, they need to be computed as and when requested. Such interpretations and data handling potentially depend on the semantics of the domain. Data requires both spatial and temporal organization. Other issues in data representation and organization include spatial/temporal resolution or granularity, and the ability to efficiently access the data specified in users' queries, especially queries having spatial and temporal predicates.

There are many research problems and issues which must be addressed before the design of fully-functional VIMS become practical. These issues were addressed by different groups. Their reports address the technical challenges that VIMS research community must address to develop systems for the applications addressed by the Applications group.

3.2 Recommendations

We recommend that the research community, with support from NSF, focus on defining the problems and developing practical solutions for two of the application domains. Specifically, we recommend that NSF should:

- Support research efforts to develop critical components of visual information management systems;
- Support research and development teams in labs, industry, and universities to validate technologies;
- Support teams to develop methodologies for scaling (and feasibility) to build large systems;
- Support building large systems;
- Support the establishment of benchmarks for performance evaluation.

4 Data Models

In determining data modeling needs for heterogeneous, distributed, multi-media databases, we assumed that the database would have the ability of storing and indexing such media as images, video, and speech signals. The basic observation that emerged at the workshop was that for effective description of content and information, current data modeling techniques would have to be expanded to include complex pointer structures and the description of functions.

The fundamental idea behind a good data model is that a useful representation should contain primitives that map naturally onto the concepts of the usage and need. A detailed analysis and understanding of particular domains is needed to identify effective primitives. This leads directly to the need for the implementation of prototypes, a need that was identified by many of the other groups as well.

4.1 Representational Needs

Two prevailing data models are used in current systems: relational and object-oriented. Commercial relational databases have become popular in the last decade. Object-oriented databases are now gaining popularity. To retain the elegance of the relational model, while exploiting powerful features offered by object-oriented modeling techniques, extensible databases are being designed both in academia and industry. In addition to these basic data modeling paradigms, extensions are currently evolving in research.

The relational model is not adequate to represent many of the richer data structures required by VIMS. As an example: most images are arrays, and most relational systems don't support an array data type. A key contribution that object-oriented database management systems provide is the ability to nest various type constructors (arrays, lists, tuples) to any level, but this is done at the price of associative access with a system-optimized, declarative language. Additional representational needs arise when we consider that domain dependent information (beyond the structure and surface semantics of the data) is needed for insertion and query processing in VIMS. This domain dependent information will be used to extract features from images that will be stored in the database and used in queries. Computer vision research has demonstrated the critical role played by the application dependent information in extraction of features. In the following, we will refer to this information a "knowledge base." It must be mentioned here that this knowledge base is not necessarily rule based.

The history of database technology is full of bringing into the database items which were previously under user control and which had to be specified by the user. This process continues, as is clear from the trends in temporal and spatial databases. Given this, the new requirements for the data modeling capabilities will be:

- Schemas as entities (say "schema"): In any data model that expresses a representation of knowledge, many of the "objects" being represented will be parts of the representation of data; in other words, we need a recursive data model. In a way, this facility is provided in object-oriented data models in their definition of objects. What is needed here is the ability to define image regions that are labeled as domain entities as attributes for other entities.

- Complex (structured) entities/attributes/relations: In representations of knowledge (that is, descriptions of the semantics or data objects) we will encounter structure in not only the entities ("a is a set of b's") but also in the attributes ("the x of a is a set of b's") and the relations ("when the x of a is 5, the y of b is a set of c's, else the y of b is a set of d's").

- Mulit-dimensional features and relations: Multi-dimensional feature spaces are a characteristic of image style indexing. Color, for example, is usually a three-valued tuple treated as a three-dimensional point in (R,G,B) space. A data model should support the representation of such dimensional spaces, especially the measures of distance in such a space and methods for clustering, projection, and defining sub-spaces and partitions. For example, in color space, you might want to represent "red" as a subspace of the whole color space and as a specialization of the general representation of "color." Are space and time special dimensional spaces or just a specific case of more general multi-dimensional spaces?

- Mixture of procedural and declarative representations: While all current data models support a declarative description of the data (and by extension of any "extra data objects"), many types of knowledge, especially in feature extraction and image processing are of the form "do this, then do this, then if this is true, do this...". These are essentially procedural. They are not well expressed as attributes, relations, or entities and, while we could express them as methods, we would like to be able to access the "innards" of the processes more easily. In addition, they may require the storage and retrieval of intermediate results in the database. Thus, it seems reasonable to attempt to capture much of this information in a procedural representation. Making the capabilities of this representation part of the data model invites opportunities for optimization, generation of user interfaces for program manipulation within the database, and (ultimately) the use of the procedural descriptions as part of the query language. For example, a user may arrive at a query, not by creating a partial description of the data objects that are desired, but by specifying part of the program that would "recognize" the data items that are desired.

- Compiled functional objects: The above item (on procedural representations) suggests that some procedural representations might be "closed" (optimized, compiled, and thus made into "black boxes") while others would be "open." Open procedural representations would be especially useful if the user were supported in a navigation over partial procedures or if there were methods for query into the procedural representations ("I don't remember what it is but I remember part of the method for extracting some feature of it"). This suggests that the data model support methods in various stages of "cooked": from raw to fully stewed.

- Mapping(s): There are issues about the existence of multiple levels at which data models could exist, and some question of the levels at which they are actually needed. Generally, storage layout and access is described by an internal model, the semantics of the application data is described by a conceptual model, and the user "interpretation" and "support" is described by a view model. In addition, there are implicit or explicit mappings between the internal model and the conceptual model and between the conceptual model and the view model. Much of the knowledge base type of information can be placed at the level of a view. In this approach, then, representations of application specific complex relations and semantics would be built upon mappings to some basic set of objects and relations. To what extent this general approach captures the requirements outlined above is not clear.

- Layered model: A novel 4-layered model, called VIMSYS, has been proposed to model data in VIMS. This model can be thought of as an architecture for organizing many of the ideas described above. It falls easily within the capabilities of current data modeling techniques to express such layered architectures. Data modeling techniques along this direction are needed. Currently, there is no experience available in implementing such a model for real applications.

4.2 Recommendations

The following set of questions and issues deserves further study. Research in these areas is required for implementation of VIMS. In the following, we have expressed several research issues in the form of questions that need answers for making progress in integrating images and video in databases.

- The coupling or division of data model, database, and user interface: To what extent can a data model be "compiled" to yield a user interface? What additional requirements does the expectation of being able to do this put on the data model? Is such a coupling desirable? Should a strict division between the data model (the database) and the user interface be observed? What, if anything, is lost by such a division?

- Heterogeneous databases and networking: This should be considered as several problems because there are connections among data models, across databases with the same data model, and across and between data and knowledge for the same database. Should we push for integration or is it better to support inter-operability? To what extent are the problems of dealing with heterogeneous databases those of mapping one data model to another? Can the methods that we might use in one database to support alternative interpretations (as from multiple doctors in a medical database), in a mapping to support alternative views, and in multiple databases to support multiple underlying representations, be unified? Would there be an advantage, such as an economy of representation, in doing this?

- Dimensional Features: Are there general operations and representations for all "dimensional" spaces? Are there unique dimensions that require special treatment (such as temporal logic, or spatial relations, such as "above," and "to the left of")? Do we need explicit representations of temporal events? Of cause and effect? Which types of time (events in the world, events in the age of data, and events in the interaction of the user with the database) do we need to represent? Why?

- Non-atomic operations: Just how rich or "open" does a representation of process have to be? Is a rule-based type of representation sufficient? Why can't we just use methods written in our favorite programming language? Do we really need to support the "migration" to/from atomic, compiling, and partial descriptions of function? How do we (within the data model) support the need to store intermediate results? What effects does the evolution of operators have on the data model? How to support versioning mechanisms? Do we want to support mechanisms for tracing and rolling back the application of operators to the data, especially when these operators might produce intermediate results?

- Alternative interpretation and uncertainty: Feature extraction, labeling, decisions made in processing, and assignments of value by humans (as in diagnosis) are all examples of interpretation. Generally, interpretation carries with it some amount of ambiguity, error, and uncertainty. Do we want to be able to model the possibility of alternative interpretations (with or without uncertainty)? Do we need special means of modeling uncertainty, or is this just another attribute of the data? Does modeling alternative interpretations make new requirements for a data model, or is it just the representation of alternative views? Can the same mechanisms being developed to handle the ambiguities that arise in distributed heterogeneous databases be used to handle alternative interpretations, views, ambiguity, and uncertainty within database systems (or other way around)?

- Integrity constraints: What new integrity constraints arise when multi-dimensional spaces are admitted? When spatial dimensions are used? When temporal dimensions are considered? When alternative interpretations and uncertainty are allowed?

5 Feature Based Indexing

In general, one can define indexing as a mechanism that reduces the search space of an operator without losing any relevant information. Database systems organize data by indexing it, using keys or other similar entities. In most cases, these entities are selected by the designer of the database, who considers frequently required searches by the system. We can further refine the definition of indexing for image and pictorial databases as a mechanism for supporting the efficient search of a collection of images and other related information, based upon intrinsic properties of the images and/or their semantic content. There are a variety of types of features to index and they include entities/objects, attributes, relationships and derivations of these. In addition, each of these types may be pictorial or non-pictorial.

5.1 Different Types of Indices in VIMS

VIMS will require indices of many different types. Depending on applications, conventional symbolic to image properties may be used to access information. Much research is needed in developing different indexing mechanisms based on different pictorial and non-pictorial attributes. We first discuss different types of indexing mechanisms and then suggest research areas related to these indexing mechanisms.

5.1.1 Pictorial-Object Index

In the pictorial case, one may think of entities/objects and attributes as those that are apparent in a picture, such as roads and rivers in a map or tumors in a medical image. Pictorial relationships can be associations between pictorial entities/objects. For example, geometric relations between objects in a picture such as left, right, distance, far, near, and so on. A problem in using these objects and relations is that images must be first segmented and the relations among the objects must be identified. The segmentation problem still remains a challenge to the computer vision community. Interactive segmentation techniques are required for VIMS, but these techniques have received, at best, scant attention from the researchers in computer vision. Finding the relations among objects has two problems: the spatial relations are usually not precisely defined and the number of such relations may become too large to be stored in a database.

5.1.2 Derived Attributes

Derived pictorial attributes are formed through a variety of means including mathematical functions to develop such measures as area, small, large, etc. Derived pictorial entities/objects are composed of computed attributes, such as contour lines, histograms, or pie charts. Such attributes have been common in many databases as part of object properties. These attributes, however, have rarely been used as indices, and it appears that in many applications, they may be used so. In those cases, a user may want to use symbolic (linguistic) values of these attributes as indices.

5.1.3 Low-level Indices

In many applications, low-level image properties such as color and texture can be very efficient as indices to an image database. Many recent research results, for example, have shown the effectiveness of color histograms and fractal dimensions in characterizing images. Quantitative measures of texture have been developed. In dynamic vision, characteristics of difference pictures may be used as motion indicators. In the last few years, there has been a strong interest in qualitative vision. Researchers are trying to characterize and recover qualitative descriptions of attributes in images. These attributes may be quite useful in VIMS. More low level attributes may be found useful in indexing images in different applications.

5.1.4 Non-pictorial Indices

Non-pictorial entities/objects and attributes are those that we normally think of in terms of current database systems. Non-pictorial relationships can be associations among entities/objects that exist independently of their position in the picture, such as city X receives money from county Y. Derived non-pictorial attributes are formed through the application of mathematical or other means to form such aggregates as average age, largest value and so on.

5.1.5 Spatio-temporal

In VIMS dealing with either a sequence of images or spatial aspects, such as in geographic applications, spatial location or temporal events may be used for indexing. Some simple applications of spatial indexing are common in many early image database systems. Geographic information systems extensively use spatial indexing. Increasing applications of VIMS will require the development of indexing techniques based on spatial location and relations among objects.

Temporal events and segmentation of frame sequences based on the story presented in the sequence will play important roles in indexing video. In addition to segmenting a sequence only along the time axis, the contents of individual images in a video database will be used for indexing video databases to retrieve either images or subsequences. Very little research has been done in this area.

5.2 Recommendations

- The ability to perform pictorial operations on entities/objects, attributes, and relationships depends upon either the extension of or new database technology. For instance, we must develop extensions to current systems or new systems that give us the capability of defining and installing content and semantic based operators. This implies that we must also develop the associated non-alphanumeric access mechanisms.

- Efficient indexing methods for high dimensions need to be developed, balancing efficient use of index memory with efficient retrieval (i.e. $O(logn)$) methods. Except in the field of information retrieval, current "multi-dimensional" indexing methods are oriented toward relatively few dimensions, say 10 to 20.

- We must have the ability to handle imprecision in a variety of contexts, such as content, semantic, or fuzzy query representation.

- With these new data types, additional indexing capabilities are required. For instance, temporal indices may be required for a series of images. Also, due to the high cost of indexing, partial indexing may be required. For instance, one may choose to index only certain values of an attribute. This also leads to higher dimensional indices in which only certain values from a set of attributes are combined to form an index.

- In order to match different objects, methods to measure similarity between them should be developed. In VIMS, the matching of objects using similarity based on several features will be one of the most common and fundamental operations. Techniques to judge similarity among different patterns will definitely depend on the application. What is needed is the development of a general theory of similarity that will be useful across several applications. These measures should also express distance between patterns.

- The features used for indexing in a VIMS should be easily expressible in intuitive terms. The quantitative measures for parameterizing attributes should be similar to those used by humans to help them express their queries. Thus, research in expressing pictorial attributes used by humans in quantitative terms should be performed.

6 Memory Management

Memory management concerns issues that impact how data is stored, represented, and moved between various levels of the memory hierarchy. Key issues that affect memory management in visual information management systems are listed below. Two points that are common to many of these issues and should be emphasized at the outset are:

- Memory management in VIMS deals with large units of data (e.g. still images of size 2k x 2k x 24 bits, video sequences of thousands of frames, say 512x512x8 each), potentially huge volumes of data (large libraries of many images, collections of video sequences, graphics, audio as well as text and tabular data), and high bandwidth requirements, with the result that the memory management requirements in VIMS applications are among the most demanding of any systems today; and

- These demands are rich and diverse, depending strongly on the application and the particular queries required. As mentioned for other topics considered in this report, memory management techniques must be dependent on the queries required in any particular system.

6.1 Key Issues

Memory management issues will become increasingly important with new applications such as those posed in the Grand Challenges. In addition to the amount of memory required to store the massive volume of data, issues related to query processing and performing other operations will require development of techniques to organize memory at every level. Key issues include: spatial reasoning, storage hierarchies and management of secondary and tertiary memory, managing video data and other sequences of images, data representation, buffer pool and other memory management, and compression.

6.1.1 Spatial Reasoning

Managing visual information requires spatial reasoning, and answering queries on layout, proximity, containment, overlap, and intersection of image objects and regions. Such queries are common, for example, in GIS, medical, and technical drawing applications. The queries can be both location based ("What is at or near location x?") as well as feature-based ("Does y occur in region x?"). In general, it is important to store spatial relationships implicitly in a spatial data structure rather than explicitly (e.g. "river 1 crosses area 2"). Various structures such as quadtrees, R trees, R+ trees, grid files, and z-indexing are therefore necessary, allowing images to be organized to answer questions related to attributes at particular absolute and relative locations.

Secondly, the spatial coordinates maintained within the set of attributes for an object need specific support and index structures. They should not be treated just as any other attribute, but represented in a form that suits their operational purpose, which often involves 2-D and 3-D distance-related functions, retrieval based on position and proximity, and so on.

6.1.2 Storage Hierarchies and Related Staging, Caching, and Clustering Methods

Visual information management systems require extremely large data volumes, up to the petabyte range. At such volumes, storage hierarchies are necessary, consisting of primary, secondary, and tertiary stores. These stores will include optical disk and, for the very large applications, tape. A variety of new memory management techniques are necessary. These include: (1) Staging schemes: These move data up and down the hierarchy and stage it for anticipated access. They need to be optimized over primary, secondary, and tertiary storage. Depending upon the application, they are not necessarily symmetric over the uppath (moving from tertiary storage up) and downpath (moving from primary storage down). In many VIMS applications, update operations are rare and this may be used to gain efficiency not possible in conventional schemes. (2) Memory caching schemes: Methods such as LRU may not be suitable, and need to be replaced with summary form methods (e.g. pyramidal and multi-resolution) with different caching schemes. The high dimensionality of data may require a special caching mechanism that depends not only syntactical context, but also on the semantic context of the application. (3) Clustering methods: These involve placing data likely to be accessed together in logically close places in storage, and are necessary to improve efficiency, particularly for slower storage devices such as tape.

6.1.3 Managing Video Data

Real time video data, because of its volume and time-dependence, requires many of its own memory management techniques. These include:

- Caching and storage layout schemes to enable real-time delivery of the data.
- Buffering techniques, ensuring that data is passed through the system with minimal copies and guaranteed delays;
- For many cases, "bufferless" delivery. Typically, video data should not be buffered as other data, but directly "dropped" on the requesting application, for instance, directly to a video window, bypassing the main DBMS and the associated buffer layers used by other data types. This leads to the development of specially designed video servers with specific functions for video data delivery.
- New indexing methods, allowing, for example, queries such as "Begin showing the sequence where event X happens".
- Storage methods for the video data that allow various forms of playback. Will the video be played backwards? Is it necessary to retrieve: the sequence of a region of interest? a moving region of interest? a time subsampled sequence? All require different storage and memory management methods, and interact strongly with compression methods used.

6.1.4 Other Sequences of Images

In addition to video sequences, there are other types of image sequences. Two examples are longer time interval repeated images as in remote sensing, and sequences made up of different versions and overlays of an image, such as in the red-lining of engineering drawings. For sequences of repeat coverage images from remote sensing, key issues for memory management are support (e.g. clustering methods on storage devices) for change detection and multi-temporal multi-spectral classification. For office documents and engineering drawings, issues are support for change control, and red lining with comments (e.g. "I tried to insert flange A but it would not fit at this point"), where a bitmap overlay is associated with a drawing.

6.1.5 Data Representation

Suitable data representation is key to effective memory management when managing visual information, and is perhaps the single most important issue affecting performance. High performance data representations and structures are essential. Examples of structures for image data are pyramids and quadtrees. Efficient task-dependent representations for secondary and tertiary memory need to be developed. Efficient structures are also needed for image objects and sets of objects, and may include trees and hierarchies, with particular objects/regions represented as (possibly multi-resolution) rectangular covers, polygons, or spline approximations. Specific representations support compression, retrieval of nonoccluded objects based on shape, retrieval of occluded objects based on shape, searches for contained and overlapping objects, progressive or reduced (thumbnail) display, and video display.

The utility of these representations is highly dependent on the nature of the queries, and representations that are efficient for certain types of queries may be inefficient for other types. Thus, the search for efficient representations is not absolute, but rather dependent on the nature of queries expected, and on the space-time tradeoff that is to be performed.

A particular case of data representation concerns images with multiple attributes per pixel, such as three color (e.g. RGB) components for digitized color images, or n components in multispectral remote sensing and GIS applications. The choice must be made between storing the data in a format where each pixel is a vector-valued n-tuple, or as n separate "planes". This choice is application specific. Storing the data as vectors allows fast operations in cases where all components are needed (e.g. some multispectral classification procedures), whereas storing as separate planes allows easy insertion and deletion of new planes (principal components, ratios of existing planes, etc.) Each method also has associated compression schemes that may be optimized for common operations. For example, compression methods can be optimized to support fast display of standard RGB color images.

6.1.6 Buffer Pool, Page Pinning, and Other Memory Strategies

Applications such as CAD and those in medicine and anatomy which involve data with a rich, hierarchical spatial structure may map naturally to object-oriented implementations. Navigating within the structures, displaying graphic results, and answering queries based on the structures and their hierarchies requires buffer management techniques different than methods applied to sets-of-records structures. These strategies must consider issues such as which components to pin in memory and how to handle concurrent queries and updates – i.e. there may be a need for an object-based paging system, or to organize data in contexts to allow caching based on the context and the task. These issues are particularly important in real-time and interactive applications.

Other memory management strategies performed by a DBMS need to be changed to manage visual data. A DBMS currently can take control of certain functions from the operating system (such as pinning pages in memory) to increase its efficiency. With visual data, additional functions may fall into this category, such as holding pages containing the root nodes of an R tree or the top layers of a pyramidal representation of an image.

Related issues arise when large repositories of visual information reside on network-based servers, with multiple users requesting data. In such cases, managing and caching data must be done within the limits of the network bandwidth and client and server memory capacities. A key issue is caching. Are large volumes of image or video data cached on the server or on the client?

Other issues in buffer management include how to take advantage of the large real memories, say 1000 megabytes, soon to be available in desktop workstations; how to store data from a single image to allow minimal paging for proximity access; storage of z-axis information for 3D rendering; and storage of large indices that span page boundaries. Conventional approaches store images in a matrix format. For many operations, this format is clearly unsatisfactory.

6.1.7 Compression

Compression methods affect essentially all aspects of memory management for visual information systems. Techniques are application dependent, and both lossless and lossy methods are needed, as are both "raw" compression methods and those that allow indexing (e.g. into subareas, into image content). Fractal based and model based compression techniques offer promising, very high compression rates, typically lossy. Important issues include whether one can quickly browse through compressed images without decompressing them and still get meaningful information, and the allowing of progressive and/or reduced display?

6.2 Recommendations

We must know the tasks and queries a user will perform in a VIMS system in order to design efficient memory management techniques. Thus a first step is to study application domains and find the important operations required to satisfy common queries across applications. Knowing the nature of the queries will enable us to identify the techniques that will be necessary to handle the data requirements. Knowing the queries will also greatly facilitate the choice of appropriate data structures and indexing methods. Some key issues to be addressed in memory management are:

- Design of data structures that are suitable for spatial reasoning,
- Development of techniques for storage hierarchies,
- Management of video and other sequences of data,
- Management of large repositories of image data on network, and
- Developing compression techniques that are efficient and also allow browsing and recovery of some information without decompression.

7 Query Processing

Query processing for VIMS is complicated by the numerous demands placed upon the system. In some cases, the VIMS will be an embedded system, where query processing really means providing low-level support for searches, following links, and communicating (fuzzy) sets of results. In standalone systems, query processing will be more visible, and visual, to support in an integrated fashion needs which today are being met by a variety of separate systems: hypertext, hypermedia, multimedia, information retrieval, GIS, knowledge based systems, or DBMS.

Depending on user needs and perspectives, VIMS may need to support a variety of perspectives, such as:

- for a video editor or producer/director, access to specific shots or scenes, based on a rough description or sketch, or referring to the content of an audio track or text in an image, or characterized by descriptive (meta) data, or having particular content, time, theme, or source.
- for a GIS user, complex special purpose queries relating to: planning, what-if questions, locating a particular shape that has desired features, etc.
- for an image archive user, supporting rapid browsing and scanning with the aid of user-defined filters (regarding content, features, origin, etc.)

Many users will have vaguely defined information needs, and often will need to recognize what they are looking for rather than describe it. The query process will be interactive and iterative, with the system accepting feedback to narrow or reformulate searches or change its link-following behavior, and to refine any queries that are given. "Like-this" comments, tentative constraints, indications of preference, and estimates of utility — all will be considered. In many situations, ambiguity and imprecision will be inherent aspects of the system, making it especially difficult to deal with huge collections, and the expensive processing required to handle large and complex objects.

For many applications, matching of known objects or queries against entries in the collection must take place in rapid fashion. This is especially difficult since in many situations transforms must be applied before matching, and since ranking of results is more expensive than fetching a fixed set. Ranking presumes the system will learn and keep statistics or other representations of preferences, but this is especially difficult in large collections, when there are a multitude of users, and where queries are complex or only rough approximations to user needs. Application specific problems arise, such as how to deal with space and time in geometric and geographic database systems, or how to manage semantic problems when text descriptions are included.

Key problems of managing ambiguity and applying methods for combination of evidence must be addressed in general, and also must be related to specific types of objects, relationships, and features.

To understand the requirements of query processing for VIMS, it is essential to consider the kinds of results that are desired. In some cases, such as for analysis or for initial exploration, only statistics are required. This is particularly useful for counting objects of a certain type or that satisfy certain criteria, or develop classification schemes. Statistics may also be supplied by the VIMS in response to a query, if it becomes clear that the results will be too voluminous, or will take too long to materialize.

In other circumstances, other forms of summaries should be provided. When a good deal of text is available, an abstract may be appropriate, but otherwise, extracts from (several) relevant text fields or sections may suffice. Other types of summaries include thumbnail, low resolution, or monochrome versions of images, or very low frame rate video or animations. In addition to these more cost effective (for storage and transmission) forms of results, there are a variety of objects.

Objects may be simple or composite. In the latter case, there may be a need for synchronization, as of video and audio streams for a digital motion picture. Composites also may be very large, though some basic objects are also voluminous, such as high resolution maps in raster form. Regarding type, VIMS objects may be images, video sequences, maps, etc. They may be selected, derived, or synthesized from the collection.

While in some cases sets of objects are returned, for many purposes it is not clear exactly which objects (best) satisfy the query. Answers may be tentative, and have associated probability, set membership value, or an amount of supporting evidence. Regardless of the formal underlying model (e.g., fuzzy set theory, Bayesian probabilities, Dempster-Shaffer), it will usually be possible to provide the top n answers, or the top n% of the answers, to give users a good sampling, especially if ranking in descending order of likelihood of relevance is the goal.

The results of query processing go further, often into some type of task-related data manipulation. While in traditional DBMS this usually means sorting, computing statistics, or reporting (to screen or hard copy), for VIMS there is a wider range of processing. Purposes include visualization, stimulating and enhancing creativity, facilitating data analysis, and supporting synthesis of new forms. Thus, the system must support flexible control of processing, and may include some type of (scripting) language(s). This may tie-in with an integrated or related expert or knowledge-based system, or information retrieval system. For efficient operations, special indexing or lazy evaluation methods (especially in cases of large databases), may be required.

7.1 Basic Issues

Query processing for VIMS involves three key aspects: languages, operations, and performance. The first of these, the choice of a query language, often receives the most attention in the database arena and is often a first focus for standards work.

7.1.1 Query Languages

Basic dynamic views should be supported by a query language for a VIMS, and in addition it must be possible to have multiple dynamic views, views with uncertainty, and version control/reference.

Some query languages will be visual. Partial or approximate drawings used in queries allow handling of one type of uncertainty. Graphic tools will be needed to allow capture and sketches of images, and there must be mappings into the query language or related internal representation. Other mappings must be available, say between image features (e.g., face height) and derived features (e.g., round face), so that queries, in whatever formalism is selected, can refer to attributes at all levels of detail and analysis.

There are many query language formalisms. On the one hand, we have linguistic forms, including linguistic variables, and natural language. On the other hand, we have logics and algebras. Standards have given us a merger of these, such as SQL and its recent forms, like SQL 3, with extensive object-oriented capabilities. However, there is a largely unresolved issue of how the parts of a query are combined. The principles of combination support handling of uncertainty, based on fuzzy sets, probability distributions, or collections of evidence. The query language should also support refined expression, whether by the user by way of weights, or by the system by way of learning or training techniques.

7.1.2 Operations

Query processing in VIMS involves several types of important operations. Matching is particularly important, often done under different constraints, and usually rests upon some definitions of similarity of objects. Other predicates and functions are needed, especially those that are feature based, those dealing with spatial phenomena, and those relating to image processing. The latter usually are expensive in terms of time and/or space.

The operations in a VIMS also must have their own properties. In general they must be dynamic, easily modifiable, and extensible. This can usually be achieved if operations can be defined in a powerful language, relative to features and object classes, which themselves will change as the schema undergoes revisions.

7.1.3 Performance

Fast VIMS query processing depends upon having a high performance system. There must be support of large objects, and the operating system and networking software therefore must avoid copying of those objects any more than is necessary

The actual query processing must make use of an optimizer that considers the formal query along with the database and system characteristics, to figure out the most efficient search approach. In many cases this is done under certain constraints, such as synchronization required between two streets, or real-time performance. Indeed, in the case of systems with sensors and VIMS, one might think of having an enormous amount of input (for insert and update) that is handled as part of a straightforward query process. Similarly, a great deal of visualization results involve query processing and then presentations, sometimes handled by the application program. In these two cases, then, we need very high bandwidth I/O. This must take place in the context of user or application sessions, over which refinement of queries, by users or the computer, and handling of like-this or by-example queries, can be handled relatively quickly. Making all this work is quite difficult, suggesting that a variety of research investigations should be carried out.

7.2 Recommendations

7.2.1 Short Term Research and Development

In the next several years, VIMS systems should be built with advanced query processing capabilities. It is crucial that task analysis be undertaken to identify representative queries and query classes. This should lead to the collection of a large amount of data, that should ultimately facilitate testing, tuning, and judgment regarding expressive power.

Another short term line of research is to integrate the query processing capabilities found in a GIS, information retrieval system, knowledge-based system, or DBMS, into one facility, with an elegant, terse, but understandable query language. Tuning of performance is an important attendant area for further studies.

7.2.2 Long Term Research and Development

In the long term, we see three important challenges that should be met, and which would require significant advances in query processing.

- Research is needed to support a wide range of data manipulation activities. This problem must be solved if we are to devise systems for integrated Global Change – where land, satellite and aerial sources must be reconciled and fit into a unified, coherent representation. Also, the problem of large scale Data Management Systems, where any format and any document preparation system can be interfaced, will require considerable research in handling a range of Data Manipulation operations.

- There is the challenge to support large image and video bases, with fast processing of content-based, query by example, and like-this queries. The size of the objects, the ways to perform matching, the approaches to combining information — all present important challenges demanding technical solutions and empirical testing.

- There is a need to develop high level data manipulation languages. These should have iconic, symbolic, and scripting capabilities. They must also support synchronization, such as of media types in streams, as well as registration of new operations and coordination of the operations.

7.2.3 High Risk Research and Development

Important problems, which are unlikely to be solved within the next 5 years, also exist in abundance. Three are of particular importance:

- There is the need for real-time image understanding. This will enable vast new opportunities regarding driverless vehicles and flexible manufacturing.

- There is the need to handle enormous (10^{16} byte) databases, and to guarantee desired response time to complex queries, and real-time response in critical activities.

- There is the challenge of integrating machine learning with query optimization, so that performance can be automatically tuned.

8 Interfaces

During the 80's, the way in which one "wrote" changed due to the ready availability of desktop information technology. The ability to draw on previous documents and cut/paste to a draft, the ability to make rapid changes, the ability to produce professional looking documents on the desktop, etc. transformed writing. Better or worse is not the issue; that can be argued ad nauseam. The point is simply that the practices that people engaged in changed in some significant way.

The 90's are bringing about yet another major transformation in workplace practices: whereas in the 80's, the computer was primarily used for specific applications, for instance, a word processor, a database, or a spreadsheet, in the 90's the computer is being woven into the very fabric of the workprocess. No longer will someone "go to the computer to get an answer;" rather, the workplace artifacts themselves will be computational engines, connected to other workplace artifacts. In a word, information technologies are becoming the new infrastructure for organizations; such an infrastructure facilitates communication, creation of artifacts, decision-making, etc.

Multimedia databases are a major factor in the evolution of such an organizational infrastructure. VIMS are more than just collections of video clips or images. Rather, VIMS promise to permit a better integration of the computer with its surroundings; then, people can apply a computational engine to transform/operate on that captured view. Keyboards allowed the capturing of text; scanners, microphones, camcorders, cameras, pens, etc. are now legitimate input devices to "the computer." While we will have to endure multimedia office memos, just as we endured all sorts of experimentation with fonts during the "early days", the interesting issue will be how VIMS transform our workplace practices beyond the obvious.

VIMS introduce two fundamental changes to the workplace's technology-based infrastructure:

- In Quality: We are no longer restricted to ASCII text strings; photo-realistic images, video clips, sounds, etc. are becoming first class citizens, i.e., that there is no (great) penalty for using non-textual data items. Given how much "image- makers" (cameras, movies, camcorders, etc.) have impacted our lives, the routine availability of such media-rich material at the desktop will surely bring its share of surprises.

- In Quantity: Very large databases are nothing new to the database community. They are, however, new to the workplace. The routine use of very-large, media-rich databases will be a natural consequence of organizations relying more and more on the information backbone to support their practices.

8.1 A Framework for Interface Research in VIMS

We can identify two key activities engaged in by professionals that can profitably be mediated by computing: (1) construction of artifacts – from reports to videos to CAD drawings, and (2) sharing of those artifacts and their intermediate products. Concurrent with the development of these workbenches, we must also deal with the problem of how this technology will come to be accepted by professionals at various levels of expertise. While a major component of that acceptance will be a function of organizational, cultural, etc. issues, nonetheless technology itself can be put to work towards facilitating acceptance.

8.1.1 New User Interface Metaphors

New technology is first used to emulate old technology. For example, the first computer-based spreadsheets were a simple one-for-one mapping from their paper cousins; however, current computer-based spreadsheets support functionality that is essentially unthinkable in the paper domain. Thus, the real challenge will be to move beyond the current metaphors, the current work practices, and take advantage of online, terabyte VIMS. Toward that end, below we identify work practices that are at the core of productive activity. Given their centrality, effort at developing support interface metaphors, techniques, and software is well-placed.

- Technological Support for the usability of technology. The key role that technology itself can play here is to permit adaptation of the interface; as the user grows more expert, the interface should change substantially to better support this higher level of expertise.

- Technological Support for the Construction of Artifacts. Activity is marked by the externalizations, the artifacts, that humans create. Effective artifacts are ones that are designed. Given how integral design activity/artifact production is to a profession, computer-based support can have a significant impact on the practices of professionals.

- Technological Support for the Sharing of Artifacts. Communication underlies effective work practices: professionals interact via various forms of artifacts, e.g., with questions, reports, drawings, etc.

- Enabling Technology: The key enabling technology for the above activities is the exploration of alternative metaphors for interacting with VIMS. The move from command line interfaces to point-and-click interfaces was a quantum leap, unquestionably. However, that latter style of interface too doesn't provide enough "bandwidth" for expression either. In particular, there are two major limitations in both of these interface styles: (1) the requirement that a user must directly refer to objects, and (2) the limited modalities – text, pointing – in which a user must express those references.
  - Supporting Indirect Reference: The user still refers directly to a specific item – either by name in the command line case, or by pointing in the graphical interface case. To resolve these types of references, humans make all sorts of inferences and negotiations. The challenge, then, is to develop metaphors that are comfortable for humans that enable them to refer indirectly to objects and that are operationalizable by the computer. What are metaphors for specifying objects that are both human-friendly and computable?
  - Multi-modal Input/Output: We need to develop interface metaphors that encourage multi-modal input and output. Command line input and point-and- click do not permit humans to use the range of input mechanisms they have evolved, including voice and gestures. The virtual reality systems being developed are exploring alternative input and output metaphors, for example, the eyephone. Multi-modal interfaces – and the metaphors that drive them – should be objects of serious scientific study.

8.2 Recommendations

Here then are two key suggestions for what research community, including funding agencies, might well do to foster more effective "interface" research and development for VIMS:

- The four areas are intimately interrelated and thus some mechanism needs to be identified where sharing of ideas can take place. Small workshops need to be held. Multi-investigator proposals need to be especially encouraged.

- The methodology for research and development in this area needs to move out from the laboratory and into real-life contexts. The field is in an exploratory phase right now; we need to encourage in situ style studies. Moreover, researchers should be encouraged to work for three years on a project; in situ studies take considerable resources and time to carry out effectively. One needs to iterate on the design of the study and on the artifacts in the study (the software environment, the evaluation metrics, etc.).

VIMS will impact our work practices; however, the user interface mediates that impact. If the user does not accept the functionality provided, if the user can't do what he/she wants to do, then in fact the interface limits the usefulness of the VIMS. Given the pervasiveness of VIMS and the pervasiveness of the computer in the everyday lives of professionals, the investment in a coherent interface research activity is well worth the price.

9 Conclusion

Visual information management will be essential in many applications. VIMS will be at the core of the Grand Challenge applications and in many other engineering, scientific, medical, entertainment, and business applications. Though there has been significant progress in database, computer vision, parallel processing, knowledge representation and many other relevant areas, many basic issues in designing VIMS have started receiving attention only recently. The hardware and software technology has progressed to the point that VIMS systems can be implemented for some simple applications, leading to a better understanding of the issues involved in designing VIMS systems for many other applications. It was the general feeling of the participants that, considering the importance of VIMS to many important areas, significant resources should be channelled towards research in this area.

The major findings of the workshop were that VIMS require new techniques in all aspects of databases, computer vision, and knowledge representation and management; and that such techniques are best developed in the context of concrete, practical applications. VIMS will provide impetus and testbeds for many techniques being explored under the aegis of Next Generation Database Systems. Researchers from image processing and understanding, knowledge representation and knowledge based systems, and databases must work very closely to develop VIMS. It is also believed that such systems should be developed in the context of applications that will be of immediate interest in industrial, medical, or scientific contexts. Without concrete applications and ambitious implementation projects, most of the important and difficult issues are likely to be ignored. Considering the interdisciplinary nature of the research in this area, it is strongly recommended that NSF fund a few major research efforts in this area.

Many issues in VIMS research are the core issues in High Performance Computing (HPCC). The recent surge in interest in developing computing systems to respond to the Grand Challenges is very relevant to VIMS research. Most of the Grand Challenge applications require the management of a large volume of non-alphanumeric information, computations, communication, and visualization of results. Techniques developed for VIMS will form the backbone for most of the applications of HPCC. Increased emphasis on HPCC by many Federal agencies can help in the rapid development of VIMS technology. Similarly, by addressing some of the Grand Challenges, researchers interested in VIMS can understand critical issues and develop techniques to solve them, in a concrete and useful context.

As is very well known, the processing of images is one of the most computation-intensive tasks. For entering images in databases and for performing required operations at query time, an enormous volume of data must be processed. Parallel computing will be essential for implementing a VIMS that can insert images in reasonable time and provide fast response to user queries. The computational requirements of video databases are likely to be one of the most demanding. It is very likely that video databases will inspire research in highly parallel-pipelined architectures.

The three most relevant areas for the development of VIMS are: databases, computer vision, and knowledge representation. VIMS should not be considered as an application of the existing state of the art in any of these fields to manage and process images. That is guaranteed to fail. Database researchers must understand the issues specific to making images and other forms of data by granting them the same status that has been given to alphanumeric information. Computer vision researchers will have to identify features required for interactive image understanding, rather than their discipline's current emphasis on automatic techniques, and develop techniques to compute these features in interactive environments. Most knowledge representation research has been concerned with symbolic knowledge. For VIMS and HPCC applications, techniques for representing symbolic and non-symbolic representations at the same level will be required. Reasoning approaches that can deal with such representations will be useful not only in VIMS, but in many other applications also.

More specifically, the recommendations of the workshop in the specific areas are summarized here briefly.

- Many generic database issues which are also relevant to VIMS are being addressed by the database researchers. Research in specific data modeling issues related to VIMS should be encouraged. Some of these issues include: layered data models, dimensional features, alternative interpretations and uncertainty representation, synchronization of information along different media, and combination of information from disparate sources.

- Feature based indexing for retrieval of information will be a significant feature in VIMS that is different from conventional databases. Some specific problems that need to be solved to enable feature based indexing are: use of pictorial objects as indices, selection of appropriate low-level attributes as indices, spatio-temporal features for indexing video databases, high-dimensional features for indexing, distance measures to determine similarity among entities, techniques to interpret fuzzy queries, and determination of basic pictorial operations on features and images to generate new indices.

- Considering the key role of VIMS in the Grand Challenges, major projects to address all aspects of VIMS design and implementation should be started in at least two different areas. Education and engineering design offer two such areas. These projects should be multi-investigator projects involving multiple universities, research institutes, and industrial groups.

- Multidisciplinary explorations in VIMS should be encouraged. Special attention should be given to the needs of practicing scientists to select and retrieve particular types of data, especially visual forms, like those used in Geographic Information Systems(GIS), along with attendant descriptive or metadata. Since many VIMS involve multidatabases, issues of query optimization and giving users an integrated schema to consider when forming queries are of key importance. Some areas that need special attention are: query formulation; applying measures of similarity; dealing with uncertainty and ambiguity; fuzzy queries; extracting suitable features; managing complex objects; undertaking detailed analysis of the information provided; and integration of data, information, and knowledge in a single system.

- Systems level projects should be supported and funded. Projects that address particular application problems and develop complete large scale prototypes should be encouraged. Projects that address the following two key areas in memory management should specifically be encouraged:

  1. storage hierarchies involving DASD, optical disk, and tape, and the staging and caching strategies necessary; and
  2. video data management as well as on-going work using large, still image collections.

- With the emergence of scientific computation as a major thrust behind new computing technology, the diversity among interfaces for different applications necessitates funding of ambitious application oriented projects that will address hardware artifacts, software, and domain knowledge based techniques for development of new metaphors and environments for the next generation interfaces.

- VIMS require partial interpretation of images, video, or other data when entered in databases. This opens a new problem area for both databases and computer vision. Research is required to develop techniques for including computer vision or similar techniques in a database environment for entering data and processing image-based queries. More emphasis is required to develop interactive image and video interpretation techniques that employ domain knowledge in order to reduce the burden on the person entering the data.

- The computational requirements for processing images and video will push existing computers to their limits. Parallel processing, pipeline architectures, neural nets and other similar approaches for feature detection, and hybrid architectures should be studied to facilitate fast insertion and query processing in VIMS.

10 Bibliography

The report to NSF includes related research material for ready reference to find the state of art in the areas related to the theme of the workshop. No effort was made to compile an exhaustive bibliography. This bibliography was prepared for the InfoScope project at the University of Michigan by Jeff Bach, Deborah Swanberg, and Ramesh Jain. Interested readers can obtain the complete report by writing to Ramesh Jain.

11 Appendix A: Application Domains

The four application domains identified in the text are now briefly described.

11.1 The Education "Grand Challenge"

It has been apparent for almost a decade that the United States' "once unchallenged preeminence in commerce, industry, science, and technical innovation is being overtaken by competitors throughout the world" because "the educational foundations of our society are presently being eroded by a rising tide of mediocrity..."[1]. There are many possible causes of this situation: insufficient resources, not enough skilled teachers, unstimulating curricula, a population that would rather watch television than do homework or read, etc. These problems, and more, affect the continuing education of adults, who find themselves increasingly undereducated as the pace of change accelerates. The necessities of earning a living, two-career families, child and parent care, etc. make it difficult if not impossible for adults to seek formal education. America's burgeoning population of older and retired people may have the time and incentive to take classes, but not the mobility and/or monetary resources.

While many of these problems can – and should – be addressed locally and non-technically, modern visual technology can provide powerful mechanisms for educating our population in a cost-effective and exciting way.

[1] The National Commission on Excellence in Education, "A Nation at Risk: the Imperative for Educational Reform", 1983, reprinted in *Comm. ACM*, 26,7 (July 1983), pp. 467-478.

We propose that NSF fund an "Educational Grand Challenge" to build a nation-wide educational network that will provide to all Americans a visual repository of the best available lectures, videos, interactive classes, etc. at a moderate cost. This repository would serve as an on-line encyclopedia and educational clearing-house to which educators would contribute educational video material and interactive learning programs, and from which individuals and institutions could interactively find and view such materials whenever and wherever it was convenient for them. Subscription and per-view fees would pay royalties to the contributors, as well as to the maintenance and improvement of the technology making this network possible. In this way, the cost of preparing high-quality educational material could be amortized by the potentially huge nation-wide audience, and the cost to the viewer could be minimized as well. The market economics of per-view fees would reward and retain in the repository the most effective materials.

Viewers would link into the network through dial-up terminals, much as they do today through interactive services such as Prodigy. However, current systems do not have the capacity to deliver digital video to large numbers of dial-up systems, a major technological prerequisite to building this network (see below). The repository would provide visual indices, menus, etc. to help potential viewers browse for information of interest, and allow the viewer to preview free "teasers" of that material before deciding to view (and pay) for it. Viewers could also provide feedback to the developer of the material, ask questions that would be relayed electronically to a lecturer, and otherwise "talk back" to the system, not just passively listen (as in existing educational TV programming). The network could also provide on-line access to bibliographic search systems and actual text of written material in either character form (if provided by publishers in soft copy) or as a sequence of images of pages. The latter is probably more cost-effective for entering existing literature than trying to recognize automatically and proof text from scanned images.

The benefits of this system to contributors, users, and the country are many. To educators, it would provide a much wider, and hence more lucrative, market for their services, rewarding the best with the largest subscriber fees. Duplication of efforts to develop curricula and other educational materials could be avoided, allowing faculty more time to experiment with new approaches and new topics, and to interact with students. The result would be greater variety of materials, greater innovation, more efficient use of key individuals, and concentration of teacher time on interacting with students (electronically!) rather than repetitively presenting the same material. But the biggest benefit by far would be to America's population, by providing a wealth of information exceeding our finest libraries today. No longer would hearing the leading lecturers in the country be limited to those with the means and mobility to travel to the lecturers' podiums. Viewers would benefit from the convenience of viewing material whenever and wherever they wanted. For children, who are increasingly sophisticated video watchers and expect rapid- fire action to maintain their interest, the network would provide a more educational alternative to the programming of commercial TV networks, particularly in the after-school hours when a parent may not be present. For working people, the availability of classes in the evening without having to leave the home would lower the threshold for initiating study, and would obviate logistic problems such as child-care. For care-givers, disabled, and older people, delivery to the home is more a necessity than a convenience. The country as a whole would benefit from a more educated population, and at significantly less cost than is currently invested in acquiring, maintaining, storing, and improving existing educational materials (e.g., in libraries nation-wide) and facilities. Recent budgetary cutbacks for local governments have resulted in severely limited hours and acquisition budgets for local libraries, which cannot hope to provide the variety and quality of material that an "on-line, video Library of Congress" can provide.

There are many technological and economic barriers to the establishment of such an educational network. There would be a huge initial cost involved in creating and populating the repository with material, but thereafter viewer fees should be able to sustain the addition of new material and gradually pay off the initial investment. Existing commercial networks such as Prodigy, which already has informational services, a limited encyclopedia, and an established customer base, may well be expanded to include the more sophisticated capabilities described here, once the technological barriers are overcome.

The technical advances required by this system are probably more challenging. Two major prerequisites at the user's end are the ability to receive video over dial-up telephone lines and lowering the cost of home terminals to below that of tuition and fees for a single class at a local university. Devices that allow a home computer to use a standard television as its display would lower the entry cost and improve the quality of the systems needed by viewers. Given the enormous market for such systems, there is already considerable development of systems approximating what we have described. Hewlett-Packard has already pre-announced a joint venture with a small company that will deliver "interactive television" but has not disclosed the entry price or hardware required to subscribe.

The "video repository", on the other hand, will require the most technical investment and innovation. Video information, whether in analog or digital format, will, without doubt, drive every aspect of this system. Cost-effective storage media for storing, cataloging, and retrieving the vast amount of video (and other format) data is crucial to keeping subscriber fees affordable. Finding and sending material of interest to large numbers of simultaneous users presents all sorts of interesting challenges to system architects. Research into how such a repository might be organized to afford the best response time and ensure timely delivery of time-dependent media such as voice and video has only begun. The user interfaces for browsing and searching for information of interest are likely to be very different from those used for today's text-oriented systems, and hence also require research and prototyping for evaluation.

Much of this research would benefit main-stream computing as well, by providing high-density low-cost storage media, improved user interfaces, and insights into organization techniques for databases far larger than any current commercial system. The techniques used by this system to store and deliver continuous-delivery media would likely become the standard for all future systems, as these media make computing more accessible, "human-like", and understandable, in less time than current I/O interfaces that require reading and typing skills. Hence this system could revolutionize computing as well as education.

11.2 Engineering and Scientific Image and Visualization Systems

Increasing engineering productivity is a critical need for American global competitiveness. Much of the information that engineers deal with today is visual, in drawings and other documents, and new types of visual information are rapidly gaining importance as computational prototyping comes into widespread use. Computational prototyping requires computer images as input, often in the form of solid CAD models, and generates images as output, in the form of visualization of system behavior and of performance data from system simulations. Managing both input and output images and integrating with existing CAD engineering imaging systems would stress today's technology, especially in trying to handle massive industrial-sized applications.

A scenario for a next-generation image information system is as follows. An engineer, in addressing a task of designing a new briefcase-sized copier, needs to browse through on-line information, including pictures, of all competitive products on the market, perhaps obtaining such market data from a publicly-available subscription service. There is a need to access and browse detailed drawings and specifications of the engineer's own company's current and past products to find components which could be re-used or adapted for this new design. As the design is developed, the engineer may want to incorporate a new concept for a compact stapling unit he learned about in a technical report from the research branch. The engineer will need to retrieve the sketches and concept description for the stapling unit electronically along with the equations or computer programs required to simulate its behavior. He/she will need to work directly with the standard solid modeling CAD system and move images into and out of its editor easily. After the product in configured, the engineer will simulate its operation and collect animation sequences of machine behavior along with still images of the simulated copier's output. In finalizing the design, the engineer may need visually to explore a complex data space of reliability vs. paper capacity vs. print quality vs. toner material in order to make final tradeoff decisions. Doing this may require interaction with the images, as in fly-by visualization. During the whole process the engineer will need to share designs, both statically and interactively, among the design team and with management. Hardcopy will need to be produced of any of the images or other data which are accessed.

The main requirements for realizing this scenario include: 1) network access to multiple heterogeneous image databases, both public and private, 2) image browsing capability, 3) color still and moving images, 3D and higher-dimensional data spaces, 4) shared image access, 5) coupling with other non-image data types, 6) dynamically computed images, and 7) massive amounts of data. Additional requirements implied by the huge data sizes and the need for interaction include 8) computational assistance in the searching and browsing tasks, for example, in invoking logic programs to filter results of queries and vision functions for content- based searching, 9) high speed communication, compression and decompression services, 10) large-format image display and interaction systems, and 11) image printing services.

Most of the technologies required for such a system exist in small systems today. The challenge is to scale them all up to industrial size and make them work together. The problems, roughly in from most easy to most difficult:

- Storage capacity: The biggest new requirement is for tertiary storage for archives. Otherwise, current trends in magnetic disks and memory are favorable.

- Image printing: Effective printers exist, and faster printers with larger format and more color capabilities will soon be available. Handling dual resolution images for display and for printing is still a problem area.

- Speed and compression: Emerging NREN, digital video and high-performance workstations are going the right direction. Large format (hence many bits) desirable for engineering will continue to stress systems but its lack will not kill engineering applications, which have already tended toward smaller sizes for this reason. There is a research opportunity for progressive-refinement systems (in which a low resolution image is sent first, then detail is increased with later bits) to facilitate browsing.

- Image types: With still and moving image compression standards at hand, hardware will make the physical handling of video images as easy as text is now. Integrating multiple data types will present a challenge in, for example, allowing queries with predicates combining images and numeric data invoked on separate databases.

- Data organization: Relational vs object-oriented? The jury is still out. Relational technology will probably be required for query speed, but the need for content-based searching makes object-oriented approaches attractive. There is much room for research on scalability and performance here.

- Shared access: Methods of sharing images and especially working with them still need to be developed. Both data issues, such as consistency and version control, and human issues such as cursor control and use of voice need to be addressed.

- Computer vision: Many exciting possibilities will open up when intelligent analysis of the images is practical, both for allowing queries based on image content and for operating on multiple images, e.g. grouping similar images. In the nearer term automatic indexing can have significant payback, such as OCR and deriving key words for each image on entry into an image database.

- System integration and interoperability: Building such systems is today almost an art. Basic research is needed on scalability, performance prediction, reliability modeling, growth strategy, and similar issues.

What would it take to develop an effective engineering and scientific image and visualization system? Many of the pieces are available in some form today. An attractive strategy is to integrate the best of these into a current state-of-the-art system, and then exercise that system on realistic data sizes. It is likely that performance will be better than expected in some areas and worse in others. Some of the refinement then would be possible on the individual components and some on the overall integrated system, as in allocation of system resources. Such an evergreen testbed system would provide a platform for testing and comparing alternative components and overall system strategies. An important enabler for this approach is the adoption of modular architecture and standard interfaces that enable software and hardware components to be removed and replaced. A modular and public-interface approach would also encourage participation by many organizations, helping to promote university-industry and university-university cooperation.

11.3 Image Databases in Medicine

Bibliographic databases in medicine, e.g., MEDLINE, have been routinely used for the last thirty years. However, medicine is a domain in which images are critical elements of the diagnostic process. Radiographs (x-ray films) are an example of a traditional source of image-based information. In recent years other imaging modalities such as CT, MRI, PET, and ultrasound images have grown in importance and today are used routinely. Other examples include 35mm slides of dermatologic conditions, histologic (tissue) slides, and images in cytology and pathology. Thus, the field of medicine is an ideal choice for the design and evaluation of tools and techniques for VIMS.

The role that an intelligent system incorporating imaging might play in, for instance, clinical oncology, might be to address the following:

- Where was the disease initially [detection, location]?

- What happened to the disease [temporal analysis]?

- Assessment of the tumor [detect and delineate tumor bulk; is it benign or malignant?]

- Assessment of the therapeutic response [did the therapy reduce the tumor?]

The user, a diagnostician, a biostatician or other investigator, could benefit from a VIMS from several points of view. One, he/she could access exemplary images from a "library" store for comparison with current images. In clinical work this type of comparison might also be used for images of the same "scene" taken at different times, to say detect the size change of a tumor. Two, images may be transferred among different sites linked by a high speed network for consultation more readily in an electronic form than as films. Three, a sequence of 2-dimensional imagery may be used to generate 3-D views for a heightened sense of perspective (it has been shown that 3-D images often reveal features easily missed in 2-D slices). Four, images may be linked to other information sources available electronically, such as bibliographic or factual databases.

The objectives of the user of a medical VIMS vary widely depending on the setting. Clinical requirements often postulate very rapid responses, i.e., on the order of a second or two. This might call for a highly compressed image, which in turn might require a lossy compression technique. On the other hand, clinical usage also comes with a strong concern for legal considerations, which could be interpreted as precluding the use of lossy techniques.

Among the desirable features of a VIMS in medicine are: Rapid access to the image store for interactive usage; image access linked to searches of bibliographic, full text or factual databases; single point-of-contact multimedia workstations to handle outputs of different databases; the ability to send images seamlessly over local networks, high speed links, and increasingly available wide area networks; interactive use of image databases where the images themselves are manipulable fields of the DBMS; the extraction of quantitative information (size, shape, and proximity of objects) from images containing multiple objects; and tools to enable computer-aided diagnosis.

Important research issues in the design of a medical VIMS are data compression, access speed, database management, and computer-aided diagnosis. These are briefly discussed below.

Data Compression: Techniques that provide a high compression ratio (CR) and are also lossless would be most desirable. However, existing lossless techniques such as DPCM, Lempel-Ziv and others do not exceed a CR of 3 or so, in contrast to lossy techniques that could deliver a CR of 30 or more. What is needed are studies by expert panels presented at consensus meetings to investigate and confirm that images (of different classes) contain all relevant information at different levels of loss. Another approach would be to mount serious research into more efficient lossless techniques.

Access Speed: Research into efficient caching algorithms, high speed communication protocols and hierarchical storage methods would be needed to provide higher access speeds.

Image DBMS: Research into multimedia database structures, pictorial query languages, spatial indexing techniques, and other tools to allow access not only to images but into images is needed.

Computer-aided Diagnosis: Tools to perform image analysis (shape extraction, texture measurement), feature extraction and link-up with content knowledge to enable the automated detection of: heart size and size changes, clustered microcalcifications in mammograms, lung nodules, etc.

11.4 Geographic/Environmental Information Systems

The largest homogeneous collection of digital image data is undoubtedly that produced since the early seventies by satellites in geo-stationary or polar orbits. Each satellite can record several dozen frames per day, and each frame may contain upwards of 100 megabytes of data. Civilian remote sensing satellites sample the surface of the earth at down to 10 meters, and record albedo in both the visible and the infrared portions of the electromagnetic spectrum. The United States Global Change Research Program has outlined a "research strategy designed to reduce key scientific uncertainties and to develop more reliable scientific predictions upon which sound policies and responses to global change can be based." In support of this effort, NASA's Earth Observing System will gather an unprecedented amount of data. The comprehensive data analysis efforts of more than just the major research universities and research centers will be required to generate and maintain appropriate visual information management systems for this geographical/ environmental data.

In the public sector, the scope of such information systems encompasses the global interests of military and supra-national government organizations; country-wide concerns such as federal agriculture, transportation, weather, mapping, and census departments; the planning and record-keeping operations of regional agencies; and the public-safety, transportation, and revenue-raising operations (based on cadastral records) of individual municipalities.

In the private sector, interested parties include mineral exploration and exploitation companies, shipping companies (trucking and railroad), communications (telephone), and public utilities (gas and electric).

The first important applications of computers to cartography were in computing the projections required to represent an (almost) spherical earth on a flat map, and elevation contours from stereophotography. More recent issues include the conversion of existing maps to digital form; generalization (the automatic preparation of a small-scale map from the large scale map), scale- independent representation; boolean operations to produce new thematic maps by overlays; the construction of gridded and triangulated digital elevation models that are more suitable for computational purposes than digital contour maps; extraction of terrain features such as peaks, ridges, valleys, and watershed boundaries; automated label placement; and self-documenting map data structures for information interchange. Some of this research has drawn on advances in computational geometry: the biennial Spatial Data Handling Symposium often features the interface between the two disciplines. An increasing portion of maps at all scales are produced with computer assistance and are made available in digital form. Digital maps are typically distributed in the form of overlays with separate files for such things as vegetation, transportation networks, topography, water cover, and political boundaries.

Research on satellite images has addressed the correction of geometric and radiometric distortion to permit map registration, tiling, and temporal comparisons; land cover segmentation and classification; and object detection (forest fires, center pivot irrigation, missile launch sites). It is now recognized that for most applications the remote-sensing data must be geo-referenced against spatially accurate maps. There has been research in combining satellite and cartographic data: it is now possible to produce realistic displays of land cover, as observed by satellite, on a 3-D elevation model obtained from cartography.

Automated cartography also borders on computer aided design (CAD): it is difficult to decide whether GIS or CAD software is better suited for utility-company external plant schematics (transmission and telephone lines, conduits, pipelines, transformers, manholes). On the database end, there has been some success in extending relational database concepts to geographic entities, and the National Center for Geographic Information and Analysis (NCGIA) has held several conferences on the subject.

Many of the research issues identified in the main body of this report are clearly relevant to this application domain; thus, without restating these issues again here, it is obvious that the design of a comprehensive and operational geographical/environmental visual information management system would be a natural choice to address the research problems in VIMS.