REPORT OF THE **DELOS-NSF WORKING GROUP ON DIGITAL IMAGERY FOR SIGNIFICANT CULTURAL AND HISTORICAL MATERIALS**[*]

Co-Chairs

Ching-chih Chen, Simmons College, Boston, USA, Co-chair
Alberto Del Bimbo, University of Florence, Italy, Co-Chair

EU Members

Giuseppe Amato, CNR-IEI, Pisa, Italy
Nozha Boujemaa, INRIA - Roquencourt, France
Patrick Bouthemy, INSA INRIA - Rennes, France
Joseph Kittler, University of Surrey, UK
Ioannis Pitas, Aristotle University of Thessaloniki, Greece
Arnold Smeulders, University of Amsterdam, The Netherlands

US Members

Kirk Alexander, Princeton University, USA
Kevin Kiernan, University of Kentucky, USA
Chung-Sheng Li, IBM T. J. Watson Research Center, USA.
Howard Wactlar, Carnegie Mellon University, USA
James Z. Wang, Pennsylvania State University, USA

1. INTRODUCTION

Recent revolutionary breakthroughs in computing and communications with the epoch-making arrival of the Internet have begun to demolish artificial disciplinary boundaries and to open vast new fields of interdisciplinary research. One major area was outlined in the recent report to the US President by the President's Information Technology Advisory Committee (PITAC), entitled *Digital Libraries: Universal Access to Human Knowledge* [US PITAC, 2001]. In its cover letter PITAC defines digital libraries as "the networked collections of digital text, documents, images, sounds, scientific data, and software that are the core of today's Internet and tomorrow's universally accessible digital repositories of all human knowledge." One of the chief impediments to broadly useful access to digital libraries, however, is the sharp cleavage in the academic research community between science and humanities. The division is particularly detrimental to research and practical development in digital libraries, because computer scientists cannot adequately provide for universal use of the world's cultural heritage without a deep understanding of the relevant materials. The DELOS/NSF Working Group for Digital Imagery of Significant Cultural and Historical Materials was organized to bring together specialists who study these priceless materials with technologists who have the expertise to help mine them in new ways to make them universally available to the world's population.
(http://www.delosnsf-imagewg.unifi.it/activities.htm).

Our world is rich with relatively inaccessible and increasingly vulnerable repositories of unique paintings, sculptures, and other works of art and fragile hand-written records in a plethora of styles

---

[*] The Report is prepared under the joint editorship of Ching-chih Chen and Kevin Kiernan.

and scripts on clay and stone and wood and canvas and cave walls, on parchment and paper and papyrus, not just in libraries and museums, but also in churches, temples, and mosques, and in the living museums of the longest inhabited cities and villages throughout the world. But time, natural disasters, thieves, vandals, and terrorists, are ever busy destroying them. For example, for the past thirty years the museums of Europe have routinely checked bags for bombs, and evacuated their premises in the middle of the day for bomb scares.   A unique Native American pictograph known as the *Blue Buffalo* was recently destroyed by an unknown vandal, while Michelangelo's *Pieta* was damaged 30 years ago by a crazed, hammer-wielding, zealot who thought he was Jesus Christ. Not long ago the Taliban blew up two 1500-hundred-year-old monumental Buddhas in the Bamiyan valley in Afghanistan.   Indeed, as we were drafting this report, two priceless paintings by the Dutch master Vincent van Gogh were stolen from the Van Gogh Museum in Amsterdam [CNN, 7 December 2002]. On September 11, 2001 – widely known as "911," along with the horrific loss of life in the World Trade Center, the world lost original Rodins and many other irreplaceable works of art in the destruction of the "museum in the sky" on the 105th floor of the North Tower. As our museums have now become the "soft targets" of terrorists, it suddenly becomes a matter of our health, security, and economic well-being to archive, preserve, and even restore our most significant cultural and historical materials in digital libraries.

Research on digital imagery for significant cultural and historical materials is imperative because of its potential impact on many related sciences and engineering, its relevance to education at all levels, and its role in promoting cultural understanding. Research on digital cultural materials will undoubtedly inspire and add new rigor to many related research fields, including computer vision, artificial intelligence, information technology, data mining, and image processing. Cultural materials require specialized knowledge to prepare for ubiquitous use by everyone from scholars to the general public. Cultural materials are valuable not only to humanities research but also to technological and scientific research. Digital imagery of significant cultural and historical materials is of great value in their own right for research on computational intelligence. Applying modern computing techniques to analyze them will gain insights for general-purpose image archival, distribution, and intelligent automatic information extraction.

Education of all levels is a crucial part of our conception of collaboration and subsequent recommendations. Treasures of human culture infuse ordinary people with inspiration, imagination, and pride. Historical materials record the history of our human societies. Ancient artifacts reveal social structure, the way people normally lived, fashion and entertainment, as well as the technological level of the times. The same modern technologies used by medicine, intelligence, forensics, and space programs will bring the cultural heritages of all time with greater access.

While other working groups deal with text and multilingual approaches, our focus on digital imagery does not neglect the beauty and cultural significance of the scripts of other cultures. Formerly so foreign and inaccessible to other cultures, Chinese characters, Arabic script, and cunieform tablets, for example, can now be easily read and understood by people without foreign language capabilities, with the help of intelligent graphical user interfaces. These interfaces can even instruct users how to compose these special hand-written artifacts or can automatically translate the images into the different languages of any and all readers. With the ever-increasing importance of communication among people all over the world, it is crucial to understand and

respect cultural diversities and learn from each other. Prejudices often come from misunder-
standing, or unwillingness to understand. Cultural materials are non-violent, unbiased, cultural
ambassadors. Modern digital technologies have made it a reality to exhibit large collections of
works from multiple cultures. Since an enormous amount of historical and cultural materials have
been created, both storage and distribution raise many challenges.   Further advancing digital
technologies for archiving and distributing these materials is of great importance.

In this preliminary report we lay out an urgent interdisciplinary research agenda to pursue
collaborative projects that develop, apply, and adapt leading-edge technologies to manage and
analyze large and varied digital collections of cultural materials.

2. OVERARCHING GOAL

*Recognizing that significant cultural and historical materials are not merely data, we advocate an
organized, continuing collaboration between subject specialists and technologists to establish
sustainable and enduring digital archives of the world's cultural heritage and to provide universal
and ubiquitous online access for advanced research as well as for all levels of formal and
informal public education.*
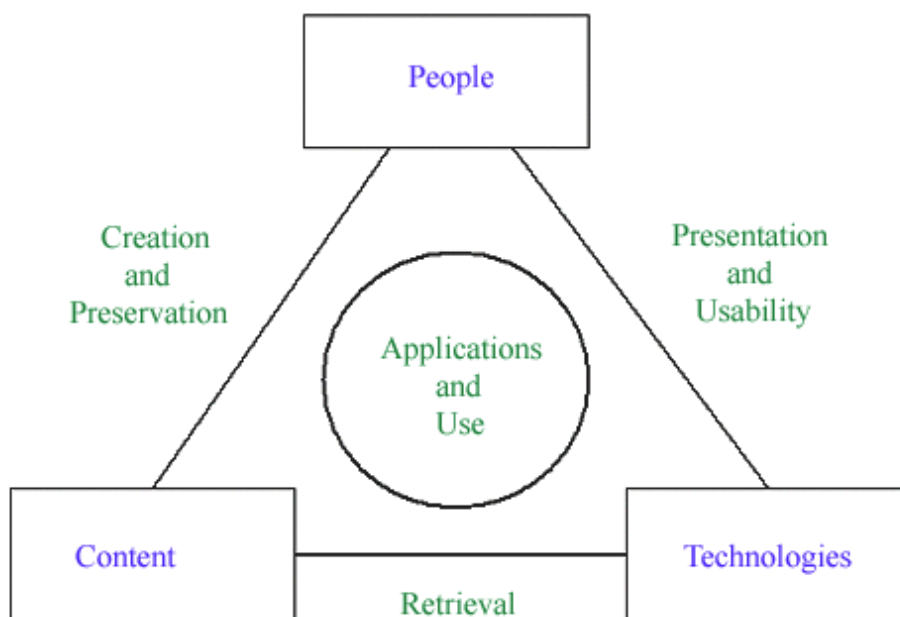
3. CONCEPTUAL FRAMEWORK

Figure 1. Conceptual Model of Our Research Agenda

Our conceptual model (Figure 1) attempts to illustrate the relationships among people, content,
and technologies in our proposed research agenda. Our interdisciplinary research will develop
technologies to enhance the way *people* create and access the *content* of their cultural heritage.
People encompass all users, from curators and library and information scientists, to scholars,

teachers, and students in all areas of the humanities, to citizens of all cultures. Content is the vast array of significant cultural and historical materials throughout the world. Technologies are the enabling research and development in all related technical areas such as information retrieval, image processing, artificial intelligence, and data mining.

We recommend focused, interdisciplinary, research programs along the three edges and the center of the triangle, areas that traditional research programs currently neglect. The research area between people and content is the area of digital imagery *creation and preservation*. The area between content and technologies is the efficient and effective *retrieval* of the content using technologies. Research into *presentation and usability* will enhance the ability to access the content. Effective *applications and use* of the research results, under lifecycle management, will integrate research of the three related areas.

4. THE FOUR INTERRELATED RESEARCH INITIATIVES

4.1. CREATION AND PRESERVATION

Digitization of cultural artifacts should provide a lasting electronic record for scholarly and universal access, preservation, and study. At the present time, however, digitization projects are proceeding without established methods of recording precise conditions of digitization. Experts in the subject field must begin to work closely with technologists in developing digital imaging technologies for historical archiving. We need tools that automatically protect the integrity, fidelity, and security of digital images, and record any subsequent processing of them. In addition to the automatic recording of such technical metadata, these tools should provide the means for subject specialists to encode descriptive metadata to facilitate subsequent search and retrieval.

Digital imaging modalities encompass visual appearance, texture, surface shape, and sub-surface hidden structure. Multi-modal acquisition enables new insights into structure and meaning. We do not confine ourselves to any specific imaging modalities. Techniques such as photography, video, X-ray, 3-D scans, infrared, UV, and laser scans have been used successfully for different art recording purposes. Capturing cultural artifacts using different imaging modalities creates the need for efficient, automated multimodal geometric registration techniques. Novel technologies or integration of existing technologies should be developed to better facilitate the study and recording of collections of historical artifacts. New methods of multi-modal rendering and presentation are required to support different audiences and applications. Techniques are required for measurements of degradation and support of restoration. Considering a digital repository of cultural artifacts not only as an educational & art history research tool but also as a powerful tool for restoration, implies that, apart from visual data (images, x-rays, etc) and simple text/metadata information, a wealth of other research/restoration data should be stored in the repository [Pappas et al, 1999]. Such data can include physical details (e.g. dimensions), restoration details, creation data, current physical conditions, storage information, historical data, associated bibliography, spectroscopy/colorimetry measurements (along with information for the measurement location on the artifact), etc. Obviously, efficient storage, recall and presentation of this information to the user is a challenging task that requires significant research.

An important consideration is to record the provenance and any subsequent changes of the items in the collection, such as distinguishing between the original source at some point in time and other existing renderings in the archives. Historical artifacts and works of art evolve over time and users must be able to distinguish the source, the time, and the process. The needs of users also evolve over time. The recording process must both incorporate previous use and anticipate future needs. For the best, long-lasting, results, we must record artifacts with the highest resolution economically possible. Naturally, this approach creates the need for special recording procedures such as image recording in partially overlapping parts (so as to minimize geometric distortions and maximize resolution) followed by image mosaicing to synthesize the whole image. Furthermore, one should bear in mind that certain situations require that the work of art should be digitized in-place (consider for example murals and architectural monuments), a fact that creates additional problems that need to be tackled.

Although recording should indeed be performed at maximum resolution to facilitate research and restoration, storage and transmission of recorded data might impose size/speed restrictions. Thus, existing lossy/lossless compression schemes should be incorporated and new compression techniques that take into account the special requirements of digital art artifacts repositories should be pursued if necessary. Some good candidates for this purpose are multiresolution schemes that can adapt to the requirements of resolution and transmission speed of a specific application.

Creation of a lasting historical record requires a repeatable imaging process. The imaging process should be calibrated so that artifacts can be digitized at multiple instances of time to study degradation or to return an artifact to its original condition in digital form. For example, the digitization of paintings requires the recording of illumination and color for later calibration of the exact appearances of the paintings at a particular point in time. The recording process should document technical metadata such as time, date, equipment, lighting, and calibration parameters. Technologies to automate the recording of technical metadata can be developed for fast digitization of large collections, as well as accruing descriptive metadata by experts in the various subject domains.

Open standards for capturing categories of artifacts can be developed to facilitate inter-operable systems. At the same time, established standards such as MPEG, DICOM, and LOC, developed for other domains, should be studied to determine suitability for digitizing artifacts.

Digital preservation and archiving activities in the area of arts should not focus only on "traditional" works of art (paintings, sculptures, architectural monuments, works of decorative art, etc) but be broad enough (and prepared to) include new forms of art that are also in need of preservation. Thus, archiving and preservation of computer-generated art (computer graphics, animations, web art), video art, movies, artistic installations, landscape art, art performances should also be considered. The limited lifecycle of some of the above forms of art (web-art, installations, performances, landscape art) is an additional reason for preserving and archiving them using digital imagery. Preservation procedures that take into account the particularities of these works of art should be defined.

While "copyright" and "intellectual property" issues are not addressed in this report, we are mindful of the great importance and complexity of these problems and issues. Thus, it is also significant to explore possibilities for creating a corpus of copyright-free image and video documents for research and evaluation.

4.2. RETRIEVAL

Computer Science and humanities disciplines often use the same terms in quite different ways. For example, in content-based retrieval computer scientists use the word "content" to refer to measurable visual properties, such as color, shape, texture, spatial relations, features that we will here call physical content. For non-computer scientists "content" normally refers to meaning. For example, the content of the image of a manuscript page is not its color, shape, etc., but the meaning of the text in the manuscript. In the case of works of art, the content of a painting like the Mona Lisa is for a computer scientist a given combination of color distribution and shape, while for the art historian or visitor to the Louvre the content might be instead painting techniques, historical models, iconographic styles, the representation of women, the study of mood, ambivalent expression, Leonardo DaVinci, and any other features that a computer cannot retrieve without descriptive markup by specialists in art. For productive collaboration between computer scientists and humanities scholars it is necessary to understand and make provision for these differences between physical content and meaningful content. With the potential ambiguity concerning "content" in mind, we discuss a variety of strictly computer science image-based retrieval topics below:

1) **Automatic feature "extraction" and combination**

   Still image attributes such as color distribution, shape, texture, and descriptors and invariant descriptors for scale, light, or point of view, are obtained by statistical image analysis. Automatic generation of categories (clustering) of these attributes enables visual overview of discrete image collections. Dynamic video attributes, such as motion field, scene activities, and camera motion, are extracted from the temporal imagery of animation or motion pictures. Other more complex extractable features include automatic transcript generation by speech recognition and geometric 3D-model description. Physical features are automatically generated metadata, as distinct from descriptive metadata supplied by experts.

2) **Searching 2D and 3D images**

   Machine-based image similarity search is computed by comparing automatically extracted features. Similarity measures must be defined with the specialist according to the feature set pertaining to the specific domain. In performing similarity searches on large repositories, scientists have investigated efficiency and scalability issues. They are developing algorithms and defining access methods that will allow highly efficient search processes for increasingly larger image and video collections. Global searches like these can be usefully narrowed by reference to descriptive metadata supplied by the domain specialists in the course of assembling and editing the image archives.

3) **Bridging the semantic gap**

   For the subject specialists preparing the collections for the universal user the term "semantic" most likely relates to signification or meaning. The *semantic gap* for imaging

scientists is the space between low-level features such as color, shape, and texture, and the high-level queries such as objects and concepts. Precise machine search allows the user to focus interest on selected objects or parts of an image, such as a small detail in a complex landscape. Machine learning techniques hold the promise to further bridge the semantic gap by generalizing from manually generated descriptive markup. For example, computer algorithms can potentially learn to classify paintings of different styles. The fact that a digitized work of art is not the work itself but an image (instance) of this work, acquired at a certain time, under specific conditions (size, resolution, camera position, light, physical condition of the work, e.g. before or after a restoration operation) makes semantic-based indexing and retrieval an absolute necessity in this area. For example, a query on "Mona Lisa" should retrieve all images of the painting regardless of size, view angle, restoration procedures applied on the painting, etc. Alternatively, image fingerprinting (or robust hashing) which deals with extracting unique image identifiers that are robust to image deformations (cropping, resizing, illumination changes, rotations etc) might be used along with query by example techniques to partially deal with this task. However, this area is still in the early stages of research & development. In the meantime, descriptive markup provided by subject specialists remains the most precise and reliable recourse and will continue to be an invaluable guide to any develop of automated search strategies.

4) Integrated access to digital repositories

After decades of research, integration of image, text-based retrieval, and other information retrieval techniques have improved search effectiveness. For example, text encoding provides semantic description of content. Automatic search can begin with user-defined constraints on the search domain. We can also define semantic structures as relationships of concepts allowing high-level content-based retrieval, which can be integrated with existing retrieval techniques to better facilitate user access. Automatic image analysis can furnish additional text annotation relating to physical features.

The availability of a huge amount of digital material, both images and videos, of cultural heritage requires the investigation of new cost-saving, and effective methods for annotation and retrieval that are easy to use for most users. Image analysis processing techniques provide a powerful means to extract useful information from pixels, and provide automatic description of image and video content. These relate to "syntactic" information (like color and texture, video editing effects), low-level primitives (like corners, shapes and spatial relationships) and higher-level information (like objects, scene content, subject description, even associated emotions...) as well as to invariance under different aspects. Content annotation based on pixels can be used to perform search operations from objective measures and descriptors of the visual content. Effective descriptors that agree with human perception and feeling are required, with particular attention paid to the computer science "semantics" of images and scenes, among other things. Obviously, some of these processes cannot work alone for images of hand-written documents, such as ancient manuscripts and cuneiform tablets, nor for some prominent 20th-century painting, such as cubism, impressionism, and abstract impressionism. In these cases, descriptive metadata is essential first.

4.3. PRESENTATION AND USABILITY

Although it does not present a complex computer science challenge, we recognize the basic need to integrate standard specialist markup into any solutions for presentation and usability for universal users. It is accordingly essential to capture commentary and annotation from past, present, and future users of digital archives of cultural and historical materials. This capability must be embedded at every level of a system and should be part of its overall design. Along with collecting and progressively adding specialist metadata, the system must be capable of tracking, filtering, and quantifying all of this information.

There is a great challenge to developers of computer systems with respect to making large collections of digital imagery available and meaningfully accessible to investigators interested in cultural and historical subject matter. In general this large and diverse audience generates demands on computer systems for simple and intuitive interfaces that stress almost every existing mode of presentation and usability, and demand substantially more sophistication on the underlying systems to make such new interfaces possible. We see the need for development in five key areas including the 1) design of advanced multimedia interfaces, 2) display and delivery technologies, 3) three-dimensional issues, 4) presentation/exploration of multimedia from multiple perspectives, and 5) visualization and summarization of cultural material/ collections and potential relations between collections.

1) Design of advanced multimedia interfaces

We need new interfaces both for expressing queries of cultural materials and for presenting results of cultural material so they can be exploited meaningfully in multiple contexts (e.g. research, teaching, public exhibition). In particular, for non-specialists (and for searching vast collections) new graphical user interfaces (GUI's) are needed for expressing both verbal and non-verbal queries. To meet this need several things must be done:

- New query and browse paradigms must be envisioned that permit iterative refinement for the investigator
- An abstract layer must be devised for posing queries in a way that is independent of data modality and language independent way. These new query types include objects to deal with either low level features (texture, color, shape) of an item or with a high level concept such as indoor/outdoor, portrait/landscape, smile/frown or even metaphors, while operations might include such relationships as logical, temporal, or spatial operations.
- Given the abstract query layer, a translation layer must be developed to present results of the query in an intelligible way to the user
- These query possibilities must be translated into flexible and informative interfaces (running the gamut from natural language to completely non-verbal queries) for the widest possible audience.

In addition to abstracting and generalizing query paradigms, systems must support both multilingual and language-independent retrieval. For example, one must be able to pose

queries related to abstract concepts such as "deity," "truth," "beauty," and "style," all of which exist in different languages, different scripts, and different cultures with dramatically different semantic meanings. The search results of such queries must permit the presentation of cultural and historical images and videos available from sources throughout the world. This interface challenge is related to the abstract query layer but requires cultural and language dependent ontology management to drive the abstract query layer so that it can perform the necessary translations in multilingual contexts. Computers cannot possibly achieve this goal without pervasive assistance from specialists.

In a system offering rich arrays of material there are also new challenges simply to render the range of query possibilities comprehensible.

Evolving digital collections must support multi-modal (heterogeneous) data handling and their integrated presentation (photographic images, UV, X-Ray etc.). All categories of users must be able to browse simultaneously along manifold axes. In presenting query results, new retrieval software must dynamically adapt to various devices and bandwidth as well as to support personalized formatting of content. The issue of personalization requires meta-tagging of the content: it must permit a characterization of what is important in the content. A current example of this can be seen in the trans-coding hints supported in MPEG-7. This permits progressive visualization of content based on relative prioritization of the content components. This automatic metadata must also integrate material based on characteristics of continually changing user demands.

Looking at a system from its ability to learn to present material in a hierarchy related to what is important, it will be desirable to attempt to monitor the actions of the user (areas of visual concentration) to add attributes/annotations to the database. If a system can accumulate these over the lifetime of the entry it can record both what the users express interest in as well as what the domain specialists may tag as valuable. Finally, this value tagging can be used to enable the display system to dynamically select between different/ multiple resolution representations or between 2D/3D representations of the content. They must also use already manually encoded metadata to narrow searches to specific realms of inquiry, just as museums organize displays, or clients themselves organize their individual visits to museums.

The design of advanced multimedia interfaces should always come hand-in-hand with the objective of making the digital repositories available to as many people as possible. Since the Internet is currently the most widely used means of information sharing & retrieval, care should be taken to construct new interfaces that integrate easily with standard Internet browsers (e.g. as plug-ins) and do not require the installation of application-specific search/browsing software.

2) Display and delivery technologies

Existing systems are workstation based. Future systems designers will need to investigate such technologies for many forms of digital imagery in varying technical and personal contexts as:

- New display solutions (e.g. Research shows that Large Scale Displays generate different impact on users and enable greater understanding of complex data.)
- Seamless interfaces which integrate information about an entire collection with recent queries against the collection. This notion is embodied in projects such as a DL-I project, "Concept Space" [Chen et al., 1997; Schatz and Chen, 1996]. In describing this project researchers have articulated the need to pose a query and to present the results in the same display space, which makes it easier to refine a query.
- Systems which support interaction based on gesture recognition. We need to build on the experience of museum guides and art historians and encourage their interaction with computer scientists. These systems can digitally capture a gesture by monitoring it directly. Gestures may include eye, hand, head, position movement, attention span tracking and even mood (based on brain wave detection.) The collection of such gestures may also help to create a natural way to interact with the viewer by permitting the simulation of virtual guides.
- Different display spaces and different users. There are two degrees of freedom, the device and the user: Currently, systems are mostly designed with single device and discrete user roles (such as curator, preservationist, and general public) in mind. What is needed is a framework to allow continuous personalization of a common interface, so that the same interface can be adjusted for different user roles using various devices.

3) Three dimensional issues

Archives holding cultural and historical materials will contain data about many artifacts with three dimensional attributes. There are two main purposes for 3D representations: (1) the representation of existing objects and (2) the reconstruction of objects that no longer exist.

3D is one way to examine cultural artifacts in a more real context. To fully support the presentation of three-dimensional artifacts developers will need to create a solid and sustainable 3D representation and presentation platform with associated queries. There is a need to develop a formal extension of existing concepts, such as VRML and its extensions in MPEG4. In addition, in a culturally aware user-world, curators and others will want access to 3D texture information as well as to information that is not related to the visual appearance of the work but is of high importance for restoration and research (material properties & condition, structural information, etc). Here too there is a need to represent and query in this domain. Finally, research is needed to determine which modes of virtual world presentation are meaningful for cultural and historical investigations such as the relative value of immersive versus non-immersive environments.

4) Presentation/Exploration of multimedia from multiple perspectives

A single cultural heritage investigation (e.g., Rome Revisited, the The First Emperor of China project, the Perseus project, Uffizi, St Petersburg, Dunhuang, Sutton Hoo, etc.) may involve artifacts of very different media types (a building, statues, a cave, a ship, textiles,

gold-working, inscriptions, pictographs) and draw on information from the realms of history, art history, archeology and the like. This multidisciplinary environment necessitates investigating a new level of abstraction across these axes so support of multiple perspectives is not uniquely defined for each different case. Research is needed to define the special common attributes of each of these kinds of perspectives.

5) Visualization and summarization of cultural material/collections and potential relations between collections

As the quantity and range of cultural archives expands, the summarization of collections themselves will be required in addition to summarizations of like items within a single collection. What kind of relations can be envisioned among collections? We must imagine and make available new taxonomies for ever-emerging interrelationships. Evolutions in history, expansion of technologies, wars etc. all provide context for a collection that will need to be queried, displayed and analyzed. This constantly evolving situation will require new techniques to view and represent the attributes of an entire collection and to relate these to the attributes of individual items. Additionally one may want to define the attributes of a collection relative to known items represented in the collection. In particular in the realm of visual collections one can imagine the Mona Lisa appearing in many collections. This notion has implications for complementary content-based image retrieval (CBIR) as well.

4.4. APPLICATIONS AND USE

The critical issue is providing access to digital archives in appropriate forms for widely different user needs. Digital image and video archives together with technologies to access and present them will provide a resource for both general education at all levels as well as for specialized research. Potential user groups include historians, curators, educators, students, and members of the general public, with their use running the gamut from curiosity to research to analysis.

Instructional technologies are necessary for teaching and research in educational institutions, which utilize digital archives to illustrate concepts and allow users to search for relationships among many collections of artifacts. This advanced range of use requires discipline-specific advice, supervision, and extensive descriptive metadata.   Unified access of multiple archives from different sources is required to support queries across heterogeneous documents of historical materials.

Museum installations for interaction with digital reproductions of cultural materials enhance the educational value. Presentation of historical artifacts together with digital enhancements can be used to illustrate the context. Display of multi-modal imagery can reveal structures or information not visible in the original. Virtual representations of artifacts allow the user to interact and explore objects. Natural mechanisms such as gesture analysis are desirable to enhance interaction with visual representations.

Extension of digital collections online is necessary to share archives among sites of different geographic locations. Sharing our cultural heritages will promote productive interchange of

knowledge and establish common ground to reach the greatest possible audience. The technological challenge of online access is to achieve high quality and high speed at low cost to the broadest possible user base.

There is a need to establish benchmark datasets and related queries among user groups to facilitate evaluation of the research progress. Technology development is an iterative process requiring continuous evaluation and improvements. Scalable deployment to archive, search, and analyze very large collections of artifacts is the ultimate challenge. Lifecycle management of content capturing, cleansing, normalization, indexing and retrieval is crucial for scalable deployment.

The lifecycle of "content" into a digital library includes the following major stages:

1) Ingestion/creation - also known as capturing or digitizing of those physical objects (painting, artifacts) into digital representation,
2) Editing - includes the normalization, standardization and cleansing of the captured data, including color and brightness adjustment,
3) Analysis - includes various metadata extraction such as low-level features (color histogram, textures, shapes, geometries, etc.), high-level features, and potentially correlating with other "related" content,
4) Management - includes the management of both metadata and content, such as developing indices for faster retrieval, addressing issues such as data integrity, consistency, and versioning,
5) Distribution - addressing issues related to content dissemination for the consumption by the end user and require I/T infrastructures (such as caching), copyrights management, etc.

As a result, the lifecycle management for a digital library includes potentially the capturing and tracking of the workflow associated with all aspects of content so that the processing steps of the content from creation to dissemination can be better automated with richer, more intelligible, results.

5. MECHANISMS FOR COLLABORATIVE RESEARCH

Perhaps the most difficult aspect of our report is devising agreeable ways to bring about meaningful collaboration between subject specialists and computer scientists. The sharp disciplinary divide and consequent disciplinary isolation of these two areas in the course of the previous century in academia and in modern society in general make it almost impossible to work together on projects of common interest and importance. We recognize, however, that this collaboration is critical to the success of the overarching goal, "to establish sustainable and enduring digital archives of the world's cultural heritage and to provide universal and ubiquitous online access for advanced research as well as for all levels of formal and informal public education." Our Working Group believes that the following mechanisms are the most promising ways to institute meaningful and productive collaborations:

1) Workshops in specific domains (works of art, the hand-written record, film, sculpture, architecture, archeology, etc.)

- Domain training of technologists
- Focus on narrow technology domain problem assessment and idea exchange (i.e., "what I've tried on this problem and this is what works or not")
- Focus on technological needs of domain specialists (i.e., "this is what I need, can't you help me?")
- Requires pre and post work on some large common data sets (which may have limited distribution)

2) International exchange programs for student and researcher

- Exchanges between like computer scientists and domain specialists to establish effective collaborative teams
- NSF stipends (including transportation) for EU people while in US and US people while in EU (for work in direct support of NSF grants, agreements, or contracts). Payments to US institutions required.
- Similar EU stipends if feasible
- Proposal: Grants for student/researcher exchange program (with simple rules regarding IP ownership developed by exchangee)
  o Periods of 3 months to 1 year
  o Family travel and living allowances
  o Some common use of code or data a result, if only in experiments and papers

3) Shared testbeds of significant cultural and historical materials

- Focus on common (i) technology, (ii) corpus or (iii) application in each of the joint testbeds
- Partnerships may be from same or different of the above categories
- Most preferable are those between domain specialists (corpus, application) and technologists
- Testbeds may be used for:
  o Applying new content analysis or CBIR technology to domain-specific corpus; validating them "at scale"
  o Usage and user studies
  o Evaluation/testing of data exchange and display standards
- Proposal: Joint/common testbeds to be central theme for cooperation
  o Common corpus with agreed upon mutual IP access (which may be otherwise limited)
  o Common technology might be analysis, descriptive metadata, etc.

4) Bilateral researcher-to-researcher projects

- Funding under single grant (NSF or EU) *or* common proposal, independently funded by NSF and EU
- Domain specialist-to-technologist (preferred) or technologist-to-technologist
- IP problems similar to joint projects (institutional), for research and practical development of sophisticated GUI's.

5) International benchmarking competitions

- National Institute of Standards and Technology (NIST) Text REtrieval Conference (TREC) – team entries
- US/EU group pairing (preferred) or national/regional team competition
- Proposal: Grants for collaborative technologist/specialist participation in NIST TREC (or comparable benchmarking) with travel grants for cooperating researcher visits and attendance at yearly meetings

Among the possible opportunities listed above for joint activity between the cultural and computer science communities, a primary focus should be to stimulate the maximum amount of interaction as possible between the communities. To this end two specific mechanisms stand out. First a set of image testbeds is very much needed that can drive computer science research in Content Based Image Retrieval that is specifically directed to meaningful applications in the humanities. A set of workshops to explicitly design and gather these datasets would make an enormous contribution towards furthering research of digital imagery of significant cultural and historical materials.

Mutual training workshops and a common set of image testbeds could stimulate specific efforts to write the kind of code necessary to develop truly accessible digital libraries of cultural and historical materials. These efforts might usefully be set up as search-and-retrieval competitions using multi-cultural image datasets and judged by a joint team of citizen user, cultural and historical scholar, and computer and information scientist. The joint evaluation team would inspire the computer scientists to develop solid automatic retrieval algorithms; the cultural/ historical scholar would insure that meaningful cultural questions are tested; the citizen user would provide a necessary dose of serendipity and a measure of broad and unpredictable applicability.

6. SUMMARY

The long-term outcomes of a successful program of collaborative digital imaging research and collection development will well serve three communities of interest: the citizen user, the cultural and historical scholar, and the computer and information scientist.

For the universal citizen user it will provide:

- The ability to recreate the experience of getting to know an historic artifact in the simulated environment of its original place at the time from the convenience of a desktop.
- Remote accessibility that enables one to see the great creative works in settings, detail and perspective unavailable even to the local viewer; and understand its history, context and relevance.
- Information facilities that enrich education, enhance cross-cultural understanding and sustain one's heritage and cultural diversity.

For the cultural and historical scholar it will enable:

- Routine use of machine image understanding technology. Automated classification of content, similarity search, and semantic retrieval will be standard database functions, useable by domain specialists without the aide of computer scientists and programmers.
- Capability and functionality for imagery search and summarization at least as rich and easy to use as that for textual sources.
- Understanding of art and artifacts will be technically deepened by sub-visual analysis and simulated reconstructions.
- Historical understanding and relevance will be enriched by a broader visual context and electronic visualization.
- Integration of descriptive markup by domain specialist

For the computer scientist, the grand challenge of determining image semantics and automatically verbalizing it will be significantly addressed. With useful accuracy in a range of domains, systems will be able to

- Describe objectively what a picture appears to say or mean,
- Describe what actions happen or what events occur in a video scene,
- Elaborate on its context by linking to the global information space through automatically generated visualizations in time and space.

The development and application of digital imaging technologies, combined with the increased information content understanding and accessibility to unique kinds and sources of visual data, will prove relevant to the many applications that serve the multiplicity of our society's needs, including those crucial to our health, security and economy.

7. RECOMMENDATIONS

Recommendation 1: Investigate issues, through collaborative workshops, common testbeds, and joint projects, to improve substantially the value of retrieval for subject specialist and general public:

- Effectiveness in bridging the gap between physical content and meaningful content.

- Fidelity of automatic physical content description/extraction and semantic structure generation.

- Efficiency of search methods to deal with large or highly complex repositories.

Recommendation 2: Develop a common system that integrates descriptive metadata supplied by subject specialists with automatically generated metadata for visual content with search techniques readily deployed to different sites.

Recommendation 3: Develop a people-centric collaborative system through collecting user annotations and commentaries about the use of digital archives and establish domain-specific benchmark datasets and related queries to facilitate evaluation of the research progress.

REFERENCES

CNN.com. "Amsterdam, Netherlands -- A pair of paintings by the Dutch master Vincent van Gogh have been stolen from the Van Gogh Museum in Amsterdam," Saturday, December 7, 2002.

Chen, H., J. Martinez, T. D. Ng, and B. R. Schatz, "A concept space approach to addressing the vocabulary problem in scientific information retrieval: An experiment on the worm community system," *Journal of the American Society for Information Science,* 48 (1):17-31, 1997.

DELOS-NSF Working Group on Digital Imagery for Significant Cultural and Historical Materials - http://www.delosnsf-imagewg.unifi.it/activities.htm

Pappas, M., G. Angelopoulos, A. Kadoglou and I. Pitas, "A database management system for digital archiving of paintings and works of art," *Computers and the History of Art*, 8 (2): 15-35, 1999.

Schatz, B. R. and H. Chen, "Building large-scale digital libraries," *IEEE Computer*, Special Issue on *Building Large-scale Digital Libraries*, 29 (5): 22-27, 1996.

U.S. President's Information Technology Advisory Committee. Panel on Digital Libraries. *Digital Libraries: Universal Access to human Knowledge. Report to the President.* Washington, DC: US National Coordination Office for Information Technology Research and Development, February 2001.

BIBLIOGRAPHY

USEFUL URLs ON MULTIPLE PROJECTS

URL on computer vision projects:
   http://www-2.cs.cmu.edu/afs/cs/project/cil/ftp/html/v-demos.html

URL on image retrieval projects:
   http://viper.unige.ch/other_systems/

GENERAL

Chen, C.-c., ed. *Global Digital Library Development in the New Millennium: Fertile Ground for Distributed Cross-Disciplinary Collaboration*. Beijing, China: Tsinghua University Press, 2001. 614 pages.

Kiernan, K., "Digital facsimles in editing: Some guidelines for editors of image-based scholarly editions," in *Electronic Textual Editing*, John Unsworth, Lou Burnard, and Katherine O'Brien O'Keeffe, eds. Forthcoming.

"INTERDISCIPLINARY" COLLABORATIVE RESEARCH ON DIGITAL IMAGERY AND HUMANITIES

Chen, C.-c., "Multimedia and the First Emperor of China: Moving toward a digital knowledge base," *Multimedia Today* (IBM), 2 (2): 68-71 (1994).

Chen, C.-c. and J. Z. Wang, "Large-scale Emperor digital library and semantics-sensitive region-based retrieval," *Proceeding of the International Conference on Digital Library -- IT Opportunities and Challenges in the New Millennium*, National Library of China, Beijing, China, July 9-11, 2002. pp. 454-462.

Kiernan, K., B. Seales, and J. Griffioen, "The reappearances of St. Basil the Great in British Library MS Cotton Otho B. x," *Computers and the Humanities* 36: 7-26, 2002.
  http://www.digitalatheneum.org
  http://www.uky.edu/~kiernan/eBeowulf/guide.htm
  http://beowulf.engl.uky.edu/~kiernan/ARCHway/entrance.htm

Wactlar, H. D. and C.-c. Chen, "Enhanced perspectives for historical and cultural documentaries using InforMedia technologies," *Proceeding of the Second ACM/IEEE Joint Conference on Digital Libraries (JCDL),* 2002, Portland, OR, July 19-22. pp. 338-339.

Wang, J. Z., J. Li, and C.-c. Chen, "Interdisciplinary research to advance digital imagery indexing and retrieval technologies for Asian art and cultural heritages," *Proceedings of the ACM Multimedia, MIR Workshop*, France, December 2002[†]

CONTENT-BASED IMAGE RETRIEVAL

Amato, G., C. Gennaro and P. Savino, "Indexing and retrieving documentary films: Managing metadata in the ECHO system,"

Bergman, L.D., V. Castelli, C.-S. Li, and J. R. Smith, "SPIRE: A digital library for scientific information," *International Journal on Digital Libraries*, 3 (1), July, 2000.

Boujemaa, N., M. Ferecatu, and V. Gouet, "Approximate search vs precise search by visual content in cultural heritage image databases," *4th International Workshop on Multimedia Information Retrieval* in conjunction with ACM Multimedia 2002, Juan les Pins, France, December 6, 2002.

[†] This reference provides a rather extensive list of references on content-based image retrieval from early dates to the present as well. Thus, we have simplified the bibliography listing on this topic in this report.

Chang, Y.-C., L. D. Bergman, V. Castelli, C.-S. Li, M.-L. Lo, and J. R. Smith, "The Onion technique: Indexing for linear optimization queries,'' *ACM SIGMOD 2000*, May, 2000.

Ciaccia, Paolo, Marco Patella and Pavel Zezula, "M-tree: An efficient access method for similarity Search in metric spaces," *VLDB'97, Proceedings of 23rd International Conference on Very Large Data Bases,* Athens, Greece, August 25-29, 1997.

Del Bimbo, A. *Visual Information Retrieval.* San Francisco, CA: Morgan Kauffman, 1999.

Guttman, A., "R-trees: A dynamic index structure for spatial searching," *Proceedings of the 1984 ACM SIGMOD, International Conference on Management of Data,* Boston, MA, 1984, pp. 47-57.

Le Saux, B. and N. Boujemaa, "Unsupervised robust clustering for image database categori-zation," *IEEE-IAPR International Conference on Pattern Recognition* (ICPR'2002), Quebec City, Canada, August 2002. Vol.1, p.259-262.

Li, C.-S., Y.-C. Chang, J. R. Smith, L. D. Bergman, and V. Castelli, "Framework for efficient processing of content-based fuzzy Cartesian queries," *SPIE Storage and Retrieval System*, January, 2000.

Smeulders, A.W.M., M. Worring, S. Santini, A. Gupta, and R. Jain. "Content-based image retrieval: The end of the early years," *IEEE Transaction on Pattern Analysis And Machine Intelligence*, 22 (12):1349-1380, 2000.[‡]

Smith, J. R. and C.-S. Li, "Adaptive synthesis in progressive retrieval of audio-visual data," *ICME 2000*, New York, July 2000.

Venkatesan, R., S-M. Koon, M. Jakubowski, and P. Moulin, "Robust image hashing," *Proceedings of the International Conference on Image Processing*, 3: 664–666, 2000.

Zezula, P., P. Savino, G. Amato, and F. Rabitti, "Approximate similarity retrieval with M-trees," *The VLDB Journal – The International Journal on Very Large Data Bases*, 7 (4): 275-293, December 1998. Included in the *ACM SIGMOD Antology*.

MACHINE LEARNING AND SEMANTIC-SENSITIVE RETRIEVAL

Barnard, K. and D. Forsyth, "Learning the semantics of words and pictures," *Proceedings of the International Conference on Computer Vision*, 2: 408-415, 2001.

[‡] This reference provides a rather complete list of references on content-based image retrieval from early dates to 2000. Thus, we have simplified the bibliography listing on this topic in this report by including mostly those after 2000.

Duygulu, P., K., N. Barnard and D. Forsyth de Freitas, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary,"' *Proceedings of the European Conference on Computer Vision,* 4: 97-112, 2002.

Li, J., R. M. Gray, and R. A. Olsen, "Multiresolution image classification by hierarchical modeling with two dimensional hidden Markov models," *IEEE Transaction on Information Theory*, 46 (5): 1826-41, August 2000.

Wang, J. Z. *Integrated Region-based Image Retrieval*. Dordrecht: Kluwer Academic Publishers, 2001.

Wang, J. Z. and J. Li, "Learning-based linguistic indexing of pictures with 2-D MHMMs," *4th International Workshop on Multimedia Information Retrieval* in conjunction with ACM Multimedia 2002, Juan les Pins, France, December 6, 2002.

Wang, J. Z., J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-sensitive integrated matching for picture libraries," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 23 (9): 947-963, 2001.

Zhang, Q., S. A. Goldman, W. Yu, and J. E. Fritts, "Content-based image retrieval using multiple-instance learning," *Proceedings of the International Conference on Machine Learning*, 2002.

DIGITAL VIDEO SUMMARIES, INDEXING AND RETRIEVAL

Christel, M., D. Ng, H. Wactlar, and A. Hauptmann, "Collages as dynamic summaries for news video," *Proceedings of ACM Multimedia 2002*, Juan-les-Pins, France, December 1-6, 2002.

Fablet, R., P. Bouthemy, and P. Perez, "Non-parametric motion characterization using causal probabilistic models for video indexing and retrieval," *IEEE Transactions on Image Processing*, 11 (4): 393-407, April 2002.

Fablet, R. and P. Bouthemy, "Non-parametric scene activity analysis for statistical retrieval with partial query," *Journal of Mathematical Imaging and Vision,* 14 (3): 257-270, May 2001.

Wactlar, H., A. Hauptmann, Y. Gong, and M. Christel, "Lessons learned from the creation and deployment of a terabyte digital video lbrary," *IEEE Computer*, 32 (2): 66-73, 1999.

Wactlar, H., A. Olligschlaeger, A. Hauptmann, and M. Christel, "Complementary video and audio analysis for broadcast news archives," *Communications of the ACM*, 43 (2): 42-47, February, 2000.