Int J Digit Libr (2005) 5: 275–286 / Digital Object Identifier (DOI) 10.1007/s00799-004-0097-5

Digital imagery for significant cultural and historical materials

An emerging research field bridging people, culture, and technologies*

Ching-chih Chen[1], **Howard D. Wactlar**[2], **James Z. Wang**[3], **Kevin Kiernan**[4]

[1] Graduate School of Library and Information Science, Simmons College, Boston, MA, USA
e-mail: chen@simmons.edu
[2] School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA
e-mail: wactlar@cmu.edu
[3] School of Information Sciences and Technology, The Pennsylvania State University, University Park, PA, USA
e-mail: jwang@ist.psu.edu
[4] College of Arts and Sciences, University of Kentucky, Lexington, KY, USA
e-mail: kiernan@uky.edu

Published online: 22 July 2005 –

Abstract. Digital imagery for significant cultural and historical materials is an emerging research field that bridges people, culture, and technologies. In this paper, we first discuss the great importance of this field. Then we focus on its four interrelated subareas: (1) creation and preservation, (2) retrieval, (3) presentation and usability, and (4) applications and use. We propose several mechanisms to encourage collaboration and argue that the field has high potential impact on our digital society. Finally, we make specific recommendations on what to pursue in this field.

Keywords: Imaging of cultural heritages – Physical and descriptive content-based retrieval – Presentation and usability – Creation and preservation – International collaboration

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* [31]. 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 on Digital Imagery for Significant Cultural and Historical Materials was organized to bring together specialists who study these priceless materials (Fig. 1) with technologists who have the expertise to help mine them in new ways to make them universally available to the world's population.

1.1 Importance of the research

Our world is rich with relatively inaccessible and increasingly vulnerable repositories of unique paintings, sculptures, and other works of art and fragile handwritten records in a plethora of styles 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 or keeping them from public access. For example, for the past 30 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

* This article results from the work of the DELOS/NSF Working Group on Digital Imagery for Significant Cultural and Historical Materials. `http://www.delosnsf-imagewg.unifi.it`

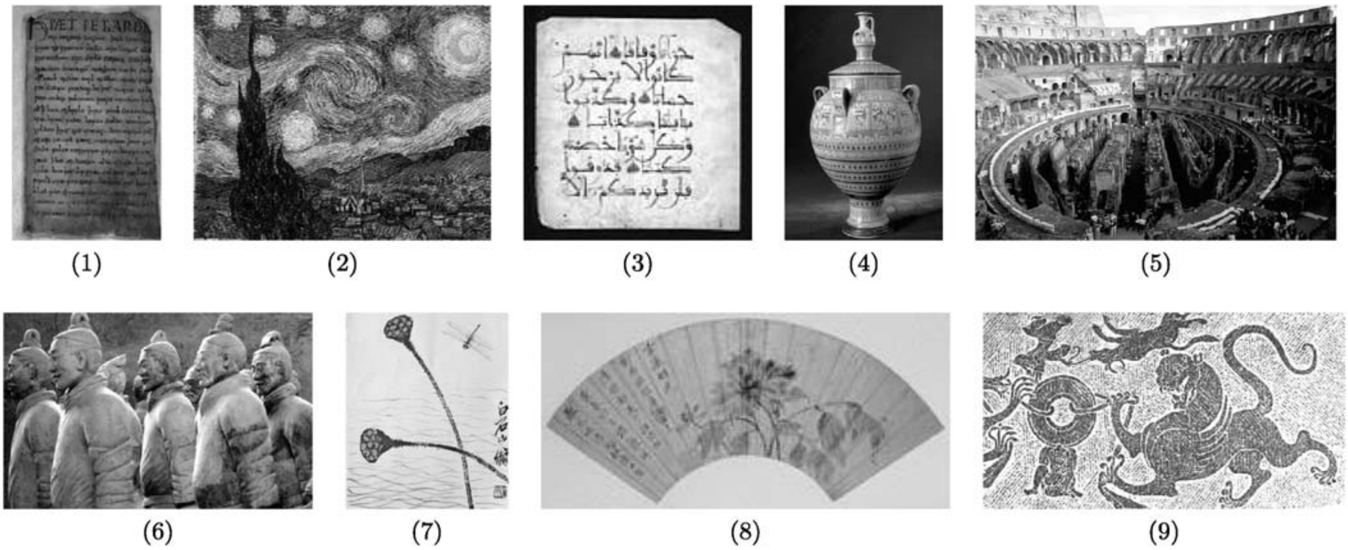

Fig. 1. Digital images of some significant cultural and historical materials. (1) English manuscripts; (2) Western painting; (3) page from a manuscript of the Qur'an; (4) ancient Greek pottery; (5) Roman Colosseum; (6) Qin terracotta warriors; (7) Chinese painting; (8) Chinese artifact; (9) stone rubbings

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 2000-year-old monumental Buddhas in the Bamiyan valley in Afghanistan. Indeed, as we were drafting the initial report, two priceless paintings by the Dutch master Vincent van Gogh were stolen from the Van Gogh Museum in Amsterdam [14]. On September 11, 2001 – widely known as "9/11" – 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. During and after the second Gulf war, countless Mesopotamian artifacts and archives, spanning a time from before 9000 B.C. into the Islamic period, have been lost. 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. We believe that it is the lack of understanding and sympathy for cultural heritage that is responsible for so much of the destruction of cultural objects that we have recently witnessed. Therefore, research on significant cultural and historical materials is important not only for preserving them but for preserving an interest in and respect for them.

Research on digital imagery for significant cultural and historical materials is imperative because of (1) its role in promoting cultural understanding, (2) its relevance to education at all levels, (3) its potential impact on many related sciences and engineering, (4) its potential to improve consolidation of such works of art, and (5) its contribution to improving our understanding of works of art themselves by knowing more about them. 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 its own right for research on computational intelligence. Applying modern computing techniques to analyze these materials will yield insights for general-purpose archiving, distribution, and intelligent automatic extraction of information from images.

Education at 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 greater access to the cultural heritages of all times.

While other researchers 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 cuneiform tablets, for example, can potentially be read and understood by people without foreign language capabilities, with the help of intelligent graphical user interfaces. Interfaces can be developed to instruct users how to compose these special handwritten artifacts or to 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 misunderstanding, or unwillingness to understand. Cultural materials are nonviolent, 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 paper, 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.

1.2 Goals and conceptual framework

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.

Our conceptual model (Fig. 2) attempts to illustrate the relationships among *people, cultural content, and technologies* in our proposed research agenda. Our interdisciplinary research will develop technologies to enhance the way people create, manage, 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. Cultural content is the vast array of significant cultural and historical materials throughout the world (Fig. 1). Technologies are the

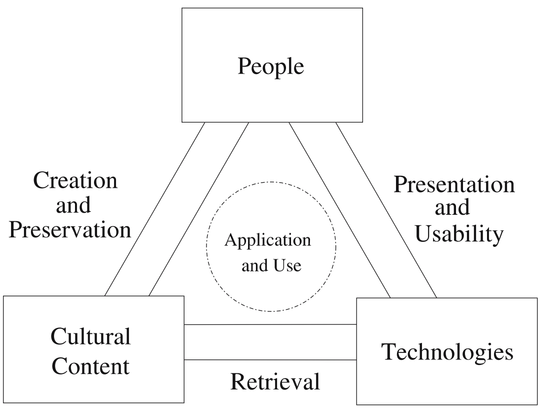

Fig. 2. Conceptual model of the proposed research directions

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 cultural content is the area of creation and preservation of digital imagery. 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.

1.3 Outline of the paper

The remainder of the paper is organized as follows. Sections 2–5 discuss the four main areas: creation and preservation, retrieval, presentation and usability, and applications and use. Section 6 suggests mechanisms to encourage collaborations. We discuss the potential impacts of the research in Sect. 7 and conclude in Sect. 8.

2 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.

2.1 Imaging modalities

Digital imaging modalities encompass visual appearance, texture, surface shape, and subsurface hidden structure. Multimodal 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, 3D 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 and 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 multimodal rendering and presentation are required to support different audiences and applications. Techniques are required for measurements of degradation and support of restoration [2]. Here, we refer to physical restoration rather than digital restoration [17]. Considering a digital repository of cultural artifacts not only as an educational and 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 or metadata information, a wealth of other research or restoration data should be stored in the repository [29]. 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.

2.2 Data recording, compression, standards

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 degrade 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, longest-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 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 or speed restrictions. Thus, existing lossy or lossless compression schemes [26] should be incorporated, and new compression techniques that take into account the special requirements of the repositories of digital artifacts must be investigated. Some good candidates for this purpose are wavelet-based multiresolution schemes that can adapt to the requirements of resolution and transmission speed of a specific application [27, 33].

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 interoperable 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.

2.3 New forms of art, copyright

Digital preservation [20, 21] and archiving activities in the area of the fine arts should not only focus on "traditional" works of art (e.g., paintings, sculptures, architectural monuments, works of decorative art) but be open and ready to include new forms of art that are also in need of preservation. Thus, archiving and preservation of computer-generated art (e.g., computer graphics, animations, Web art), video art, movies, artistic installations, landscape art, and performances should also be included in our research agenda. 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. Here the need may extend to archiving not just the data or running code, but the environment within which it executed, in order to capture and reproduce the temporal *performance experience*. How these can be done efficiently opens up a number of challenging research questions.

While "copyright" and "intellectual property (IP)" issues are not addressed in this paper, 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.

3 Retrieval

Computer science and humanities disciplines often use the same terms in quite different ways. For example, in content-based image retrieval (CBIR) [15, 28, 35], computer scientists use the word "content" to refer to measurable visual properties such as color, shape, edges [9], texture [11, 19], spatial relations, and other features that we will here call physical content. For non-computer scientists "content" normally refers to meaning [12]. As an 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.

3.1 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 [22]. Dynamic video attributes, such as motion field, scene activities, and camera motion, are extracted from the temporal imagery of animation or motion pictures [16, 39–41]. Other more complex extractable features include automatic transcript generation by speech recognition [3, 32] and geometric 3D-model description [5]. Physical features are automatically generated metadata, as distinct from descriptive metadata supplied by experts.

Over the years, researchers have noticed the significant gaps between the features we can extract from the images and the meanings of the images. According to Smeulders et al. [35], the *sensory gap* is the gap between the object in the world and the information in a (computational) description derived from a recording of that scene. The *semantic gap* is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation. Research to bridge these gaps is extremely important.

3.2 Searching 2D and 3D images

Machine-based image-similarity search is computed by comparing automatically extracted features [23, 36, 38]. 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 [1, 10, 18, 37, 42]. 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.

3.3 Bridging the semantic gap

For subject specialists preparing 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 the features or information extracted from the visual data and the user interpretation of the same data. 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 [7]. Semantic-sensitive retrieval methods integrate image classification with feature-based retrieval. Machine learning techniques hold the promise of further bridging the semantic gap by generalizing from manually generated descriptive markup or keyword annotation [4, 24]. For example, computer algorithms can potentially learn to classify paintings of different styles or of different painters [25]. New search paradigms such as mental image search with visual words [8] enhance the user expression of visual target without a starting example.

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 and 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 development of automated search strategies.

3.4 Integrated access to digital repositories

After decades of research, integration of image, text-based retrieval, and other information-management and 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 of extracting 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, etc.) 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 handwritten documents, such as ancient manuscripts and cuneiform tablets, or for some prominent 20th-century paintings, such as cubism, impressionism, and abstract impressionism. In these cases, descriptive metadata are essential first.

4 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) 3D issues, (4) presentation and exploration of multimedia from multiple perspectives, and (5) visualization and summarization of cultural material collections and potential relations between collections.

4.1 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 nonspecialists (and for searching vast collections) new graphical user interfaces (GUIs) are needed for expressing both verbal and nonverbal 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 or language. 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 nonverbal 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 multimodal (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 metatagging 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 transcoding hints supported in MPEG-7. This permits progressive visualization of content based on relative prioritization of the content components. These automatic metadata must also integrate material based on characteristics of continually changing user demands.

Looking at a system based on 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 or 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 or 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.

4.2 Display and delivery technologies

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

- New display solutions. For example, large-scale displays may generate different impacts on users and enable greater understanding of complex data. Handheld and possibly wireless devices allow users to access the data from virtually any place and at any time.
- Seamless interfaces that integrate information about an entire collection with recent queries against the collection. This notion is embodied in projects such as the NSF project "Concept Space" [13, 34]. 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 that support interaction based on gesture recognition [30] or voice-and-gesture recognition [6]. 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 of interacting 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.

4.3 3D representation

Archives holding cultural and historical materials will contain data about many artifacts with 3D 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 3D 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 MPEG-4. 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 that is of high importance for restoration and research (material properties and 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 and nonimmersive environments.

4.4 Presentation and exploration of multimedia from multiple perspectives

A single cultural heritage investigation (e.g., Rome Revisited, 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. The process of conservation is greatly helped by electronic files tied together, like the electronic patient records in health care, as these are all multimedia files.

4.5 Visualization and summarization of cultural materials or 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 as well.

4.6 Practical issues

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 and retrieval, care should be taken to construct new interfaces that integrate easily with standard Internet browsers. Plug-ins or Java applets should be provided for normal users to install quickly. Advanced application-specific search or browsing software can be developed for trained users.

5 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.

5.1 Applications

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 their educational value. Presentation of historical artifacts together with digital enhancements can be used to illustrate the context. Display of multimodal 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 for sharing 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.

5.2 Scalability and lifecycle management

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, preservation, cleansing, normalization, indexing and retrieval, and disposition 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, as well as the encoding of specialist commentary;
3. Analysis – includes various metadata extraction such as low-level features (color histogram, textures, geometries, shapes, etc.), high-level features (specialist commentary, etc.), 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; and
5. Distribution – addresses issues related to content dissemination for the consumption by the end user and requiring information technology infrastructures (such as caching), copyright 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.

6 Mechanisms to encourage collaborations

Perhaps the most difficult aspect of our paper 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 for people from these areas 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" (Sect. 1.2). We conclude that the following mechanisms are the most promising ways to institute meaningful and productive collaborations.

6.1 Workshops in specific domains

We believe workshops in specific domains can help establish interdisciplinary collaboration. Such domains include works of art, the handwritten record, film, sculpture, architecture, archeology, etc. Workshops can provide domain training to technologists. We should focus on narrow technology domain problem assessment and idea exchange (i.e., "this is what I've tried on this problem and this is what works or not") and technological needs of domain specialists (i.e., "this is what I need, can't you help me?"). Some of the workshops may require pre- and postwork on some large common datasets (which may have limited distribution).

6.2 International exchange programs for student and researcher

We advocate more exchange programs between like-minded computer scientists and domain specialists to establish effective collaborative teams. US federal agencies like NSF can provide stipends and travel expenses for foreign researchers while in the US and American researchers while in other countries (for work in direct support of NSF grants, agreements, or contracts). US institutions can provide such programs with government support. If feasible, foreign countries can also provide stipends.

We propose the establishment of grant mechanisms for student/researcher exchange programs. Some simpler rules regarding IP ownership developed by exchangees should be created. For example, the collaborating institutions may negotiate a joint ownership agreement to the IP before an exchange program starts. This is not always easy but should not be impossible. Periods of exchange can range from 3 months to 1 year. Family travel and living allowances should be provided. Some common use of code or data results, if only in experiments and papers, should be encouraged.

6.3 Shared testbeds of significant cultural and historical materials

In each of the joint testbeds, we may focus on common (1) technology, (2) corpus, or (3) application. Partnerships may be from the same or different of the above categories. Most preferable partnerships are those between domain specialists (corpus, application) and technologists.

Testbeds may be used for applying new content analysis or CBIR technology to domain-specific corpora, validating them "at scale," conducting usage and user studies, and evaluating/testing data exchange and display standards.

We propose that the central theme for cooperation be joint or common testbeds. Such testbeds should include common corpora with agreed-upon mutual IP access (which may be otherwise limited) and common technology (including analysis, descriptive metadata).

6.4 Bilateral researcher-to-researcher projects

International collaborative research projects should be encouraged. Funding mechanisms may include (1) funding under single grant or (2) common proposal, independently funded by individual countries. Domain specialist-to-technologist (preferred) or technologist-to-technologist collaboration is important for such projects. IP problems for such joint projects on technology research and practical development of sophisticated GUIs should be addressed at the institution level.

6.5 International benchmarking competitions

We believe benchmarking competitions, if properly executed, can expedite the technology development. The National Institute of Standards and Technology (NIST) Text REtrieval Conference (TREC) is one example. We should encourage international research group pairing (preferred) or national/regional team competition. We propose the establishment of grants for collaborative technologist and specialist participation in competitions like NIST TREC with travel grants for cooperating researcher visits and attendance at yearly meetings. Grants can also be established to stimulate more participation in such benchmark competitions from the research community.

6.6 Summary of mechanisms

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 CBIR 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 multicultural 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, and the citizen user would provide a necessary dose of serendipity and a measure of broad and unpredictable applicability.

7 Impacts of the research

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 to 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 as automated classification of content, similarity search, and semantic retrieval will be standard database functions, usable by domain specialists without the aid 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 that will be technically deepened by subvisual analysis and simulated reconstructions;
- Historical understanding and relevance that 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 increased understanding of information content 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.

8 Conclusions

In this paper, we introduce the emerging field of digital imagery for significant cultural and historical materials. We believe the research is imperative due to its potential impact on many related sciences and engineering, its relevance to education at all levels, and its role in promoting cultural understanding. As an interdisciplinary field bridging people, culture, and technologies, it has four interrelated subareas: (1) creation and preservation, (2) retrieval, (3) presentation and usability, and (4) applications and use.

Based on the discussions, we make the following specific recommendations:

1. Investigate issues, through collaborative workshops, common testbeds, and joint projects, to improve substantially the value of retrieval for subject specialist and general public: (i) effectiveness in bridging the gap between physical content and meaningful content, (ii) fidelity of automatic physical content description/extraction and semantic structure generation, and (iii) efficiency of search methods to deal with large or highly complex repositories.
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.
3. Develop a peoplecentric 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.

Acknowledgements. This work is supported by the EU Network of Excellence DELOS and the US National Science Foundation. Conversations with Michael Lesk and Stephen Griffin have been helpful. We would like to acknowledge the comments and constructive suggestions from the reviewers. The following members of the DELOS/NSF Working Group on Digital Imagery for Significant Cultural and Historical Materials contributed to this article:
Ching-chih Chen (US Co-Chair), Simmons College, USA
Alberto Del Bimbo (European Cochair), University of Florence, Italy
Kirk Alexander, University of California at Davis, USA
Giuseppe Amato, ISTI-CNR, Italy
Nozha Boujemaa, INRIA, France
Patrick Bouthemy, INSA INRIA – Rennes, France
Kevin Kiernan, University of Kentucky, USA
Joseph Kittler, University of Surrey, UK
Chung-Sheng Li, IBM T.J. Watson Research Center, USA
Ioannis Pitas, Aristotle University of Thessaloniki, Greece
Arnold Smeulders, University of Amsterdam, The Netherlands
Howard D. Wactlar, Carnegie Mellon University, USA
James Z. Wang, The Pennsylvania State University, USA

References

1. Amato G, Rabitti F, Savino P, Zezula P (2003) Region proximity in metric spaces and its use for approximate similarity search. ACM Trans Inf Syst 21(2):192–227
2. Baglioni P, Giorgi R, Chen C-c (2003) Nanoparticle technology saves cultural relics: potential for a multimedia digital library. In: Online Proc. of DELOS/NSF workshop on multimedia contents in digital libraries, Crete, Greece
3. Bahl LR, Jelinek F, Mercer RL (1983) A maximum likelihood approach to continuous speech recognition. IEEE Trans Pattern Anal Mach Intell 5(2):179–190
4. Barnard K, Duygulu P, de Freitas N, Forsyth DA, Blei DM, Jordan MI (2003) Matching words and pictures. J Mach Learn Res 3:1107–1135
5. Beardsley P, Torr P, Zisserman A (1996) 3D model acquisition from extended image sequences. In: Proc. European conference on computer vision, 2:683–695
6. Bolt R (1980) Put-that-there: voice and gesture at the graphic interface. Comput Graph 14(3):262–270
7. Boujemaa N, Ferecatu M, Gouet V (2002) Approximate search vs precise search by visual content in cultural heritage image databases. In: Proc. ACM Multimedia, MIR workshop, France
8. Boujemaa N, Fauqueur J, Gouet V (2004) What's beyond query by example? In: Shapiro L, Kriegel HP, Veltkamp R (eds) Trends and advances in content-based image and video retrieval. Lecture notes in computer science, vol . Springer, Berlin Heidelberg New York
9. Canny J (1986) A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 8(6):679–698
10. Chang Y-C, Bergman LD, Castelli V, Li C-S, Lo M-L, Smith JR (2000) The Onion technique: indexing for linear optimization queries. In: Proc. ACM SIGMOD
11. Chellappa R, Chatterjee S (1985) Classification of textures using Gaussian Markov random fields. IEEE Trans Acoust Speech Signal Process 33(4):959–963
12. Chen C-c (2003) Designing multimedia digital libraries with content development and management: technology potential and challenges. In: Online Proc. of DELOS/NSF workshop on multimedia contents in digital libraries, Crete, Greece
13. Chen H, Martinez J, Ng TD, Schatz BR (1997) A concept space approach to addressing the vocabulary problem in scientific information retrieval: an experiment on the worm community system. J Am Soc Inf Sci 48(1):17–31
14. CNN (2002) Amsterdam, Netherlands – A pair of paintings by the Dutch master Vincent van Gogh have been stolen from the Van Gogh Museum in Amsterdam. CNN.com, 7 December 2002
15. Del Bimbo A (1999) Visual information retrieval. Morgan Kauffman, San Francisco
16. Fablet R, Bouthemy P, Perez P (2002) Non-parametric motion characterization using causal probabilistic models for video indexing and retrieval. IEEE Trans Image Process 11(4): 393–407
17. Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell 6(6):721–741
18. Guttman A (1984) R-trees: a dynamic index structure for spatial searching. In: Proc. ACM SIGMOD, Boston, pp 47–57
19. Jain A, Farrokhnia F (1991) Unsupervised texture segmentation using gabor filters. Pattern Recog 24(12):1167–1186
20. Kiernan K, Seales B, Griffioen J (2002) The reappearances of St. Basil the Great in British Library MS Cotton Otho B. x. Comput Humanit 36:7–26. `http://www.digitalatheneum.org`
21. Kiernan K (2004) Digital facsimiles in editing: some guidelines for editors of image-based scholarly editions. In: Unsworth J, Burnard L, O'Keeffe K (eds) Electronic textual editing. `http://www.tei-c.org/Activities/ETE/Kiernan/`
22. Le Saux B, Boujemaa N (2002) Unsupervised robust clustering for image database categorization. In: Proc. IEEE-IAPR international conference on pattern recognition, Quebec City, 1:259–262
23. Li C-S, Chang Y-C, Smith JR, Bergman LD, Castelli V (2000) Framework for efficient processing of content-based fuzzy Cartesian queries. In: Proc. SPIE – Storage and Retrieval System
24. Li J, Wang JZ (2003) Automatic linguistic indexing of pictures by a statistical modeling approach. IEEE Trans Pattern Anal Mach Intell 25(9):1075–1088
25. Li J, Wang JZ (2004) Studying digital imagery of ancient paintings by mixtures of stochastic models. IEEE Trans Image Process 13(3):340–353
26. Li J, Gray RM (2000) Image segmentation and compression using hidden Markov models. Kluwer, Dordrecht
27. Mallat SG (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell 11(7):674–693
28. Niblack W, Barber R, Equitz W, Flickner M, Glasman E, Pektovic D, Yanker P, Faloutsos C, Taubin G (1993) The QBIC project: querying images by content using color, texture, and shape. Storage and retrieval for image and video databases. Proc SPIE 1908:173–181
29. Pappas, M, Angelopoulos G, Kadoglou A, Pitas I (1999) A database management system for digital archiving of paintings and works of art. Comput History Art 8(2):15–35
30. Pavlovic VI, Sharma R, Huang TS (1997) Visual interpretation of hand gestures for human computer interaction: a review. IEEE Trans Pattern Anal Mach Intell 19(7):677–695
31. PITAC, US President's Information Technology Advisory Committee – Panel on Digital Libraries (2001) Digital libraries: universal access to human knowledge. Report to the President. US National Coordination Office for Information Technology Research and Development, Washington, DC
32. Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 77(2):257–286
33. Said A, Pearlman WA (1996) A new fast and efficient image codec based on set partitioning in hierarchical trees. IEEE Trans Circuits Syst Video Technol 6(3):243–250
34. Schatz BR, Chen H (1996) Building large-scale digital libraries. IEEE Comput 29(5):22–27
35. Smeulders AWM, Worring M, Santini S, Gupta A, Jain R (2000) Content-based image retrieval: the end of the early years. IEEE Trans Pattern Anal Mach Intell 22(12):1349–1380
36. Smith JR, Li C-S (2000) Adaptive synthesis in progressive retrieval of audio-visual data. In: Proc. international conference on multimedia and expo (ICME), New York
37. Wang JZ, Li J, Wiederhold G (2001) SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. IEEE Trans Pattern Anal Mach Intell 23(9):947–963
38. Wang JZ, Li J, Chen C-c (2002) Interdisciplinary research to advance digital imagery indexing and retrieval technologies for Asian art and cultural heritages. In: Proc. ACM Multimedia, MIR workshop, France

39. Wactlar HD, Chen C-c (2002) Enhanced perspectives for historical and cultural documentaries using InforMedia technologies. In: Proc. ACM/IEEE joint conference on digital libraries. Portland, OR, pp 338–339
40. Wactlar HD, Hauptmann A, Gong Y, Christel M (1999) Lessons learned from the creation and deployment of a terabyte digital video library. IEEE Comput 32(2):66–73
41. Wactlar HD, Olligschlaeger A, Hauptmann A, Christel M (2000) Complementary video and audio analysis for broadcast news archives. Commun ACM 43(2):42–47
42. Zezula P, Savino P, Amato G, Rabitti F (1998) Approximate similarity retrieval with M-trees. Int J Very Large Data Bases 7(4):275–293