Research Article

Multimedia in Cultural Heritage Manuscripts: Integrating Description, Transcription, and Image Content

Catalin Calistru,[1] Cristina Ribeiro,[1, 2] and Gabriel David[1, 2]

[1] *Depatment of Telecommunications and Multimedia, INESC Porto, Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal*
[2] *Departamento de Informática, Faculdade de Engenharia, Universidade do Porto, 4200-465 Porto, Portugal*

Correspondence should be addressed to Catalin Calistru, cmc@inescporto.pt

Cultural heritage documents are often subject to digitization processes resulting in image material, even for textual contents. It is therefore common, in collections of valuable documents, to have descriptive information generated by the institutions, along with digitized images, transcriptions created by scholars, translations and even miscellaneous annotations. To offer a faceted access to the collection it is necessary to explore these diverse materials, integrate them according to a model that accounts for both metadata and the content and provide a comprehensive retrieval environment. In this work we have applied the MetaMedia multimedia database framework to a collection of ancient documents, processed the documents in their descriptive, textual, and image content and produced a browsing and searching system. The main challenges are the integrated management of metadata and content, the indexing of the image content, and the design of the browsing and searching interface where various views on the data are kept together.

1. Introduction

The management of and access to large digital repositories are current concerns in commercial and cultural organizations. Results obtained in diverse areas can contribute to solve the main problems involved. In the area of database management, recent research on hybrid models exploring both structured and semistructured information are valuable for complex or heterogeneous collections [1, 2]. Advanced visualization techniques are required for presenting large answer sets of multimodal documents [3]. Digital libraries research has focused both on the conceptual aspects of digital collections as information systems [4] and on the development of operational platforms to support the organization and access to digital collections [5–7]. The success of text-based retrieval has raised the expectations of users concerning the possibilities of search on multimedia collections. Recent results on multimedia retrieval are being tested on large datasets originated in news services or broadcasters [8–10].

Historic cultural heritage repositories, where reproductions of the actual documents are being provided by means of digitization, are better regarded as multimedia collections, while the gathering of current materials intrinsically requires multimedia facilities. In cultural heritage collections, where document curation typically involves some kind of expert analysis, an operational system must satisfy at least three requirements. The first is to allow rigorous descriptive metadata to be handled and associated to documents. The second is to support the management of the collection. The third is to offer content-based search to specialized and lay users with state-of-the-art technologies [11]. Our goal is to make search based on low-level visual features integrate seamlessly into a retrieval system supporting high-level descriptive metadata.

This paper presents a platform for the integration of functionalities for the representation of documents, the management of the collections and the multifaceted access to their contents. The underlying multimedia data model identifies the main concepts in standards from the archival,

and the audiovisual areas. The main features of the model are the integration of descriptive and content analysis metadata, the association of metadata to collections as well as to documents, the extensibility with respect to the inclusion of new descriptors and the support to several retrieval modes.

The multimedia model has been designed to support database applications, and its scope is distinct from the models underlying current standards, and from the reference models adopted in several communities. The former [12–14], due to their specificity, prescribe strict definitions and formats for the elements of a description. The latter [15–17] are more complex models encompassing the documents and their processes in an organization. A model designed to support a multimedia database can be restricted to a set of core features and still allows the incorporation of data from different standards and supports the storage and retrieval of individual documents and collections.

The paper is organized as follows. Section 2 introduces the concepts adopted to describe and manage multimedia documents. Section 3 outlines the proposed repository model, and how it supports the representation of document features. The following sections present the model for the repository and the associated workflow, the application of the MetaMedia platform to a case study, the document views in the user interface and the retrieval methods.

2. Describing Documents

To deal with multimedia collections, it is necessary to handle the content itself, which may require specific storage and presentation devices, and to manage the associated metadata that may be of different nature and generated according to a variety of standards. Metadata covers aspects of media such as its description, content analysis, technical details, terms of use, and administrative aspects and it can be automatically generated or manually associated to the documents. In cultural heritage collections, there is usually a great concern with descriptive metadata, aimed at retrieval, and with record-keeping metadata used to manage the collections.

Depending on the documents, there are several features that can be automatically extracted to generate content descriptors. For a document with textual content, text indexing can be viewed as the extraction of descriptors (words) which are organized in specialized structures for retrieval. The analysis of image or video generates feature representations that will be referred to as low-level descriptors. For example, color analysis generates descriptors such as DominantColor, ColorStructure, and ScalableColor descriptors [18]. Similarly, texture analysis generates descriptors such as HomogeneousTexture and EdgeHistogram [18]. These low-level descriptors consist predominantly of vectorial data. The DominantColor descriptor, for example, consists of an RGB-tuple, and the color histogram descriptors are vectors of 128 or 256 values. The entire set of descriptors can be viewed as a vectorial space, called feature space, that can have hundreds or thousands of dimensions. Searching in feature spaces requires the use of custom-designed similarity metrics [19] and indexes.

In a system aimed at providing access and navigation on complex documents, it is necessary to use existing metadata associated with the documents when they are incorporated in the collections, and also to export sets of documents with their associated descriptions using the standards for the domain. Several specific standards have been established and standards from other domains can also be applied.

2.1. Standards. Standards tend to address different aspects of the digital content, the ones most relevant for the intended use of the documents. While it may be essential for a library to have detailed information on a scientific journal, title and author descriptors may be enough for documents in a web site; a film distribution corporation may require media- and genre-specific descriptors to provide search on the movie catalog, but the same documents when used by a broadcaster require information on the days and times when they will be programmed.

In the library area, there are well-established standards that support applications and metadata sharing [12, 20]. Recent work concerns conformance to the XML language syntax, inclusion of sound and image documents, and stronger networking.

ISAD and ISAAR [13] are standards for archival description. Their basic principles are multilevel structure and uniform description, and they handle both the documents and the people and organizations involved in their creation.

Dublin Core (DC) [21] has appeared to solve the problem of the lack of description of documents on the web. It consists of a set of basic descriptors such as title, creator and date, intentionally kept at a basic nonspecialist level. DC is being widely adopted as part of other web-related standards, such as those for the Semantic Web initiatives [22].

MPEG-7 [14] comes from the audiovisual signal processing community and aims at creating metadata for complex multimedia items. The emphasis is on descriptors that can be automatically extracted from audiovisual content, leaving descriptive metadata for other standards such as DC. MPEG-21 [23] originated in the same community and concerns metadata for handling the multimedia delivery chain rather than item or collection descriptions.

3. The Multimedia Repository

Our target application is a multimedia repository offering functionality to several actors. The repository managers will be able to design a structure for their documents, choose the descriptors that will be associated to the documents and the general features of the retrieval interface. The archivists or curators will create descriptive metadata, and in some cases also semantic annotations according to a tag vocabulary. The visitors will have several possibilities of exploring the repository: browsing the structure of the collection or the documents, viewing the details for a single document, searching on any of the available facets. These functionalities are in the line of those offered by existing repository and content management systems such as DSpace,

Fedora, Greenstone or CONTENTdm [5–7, 24] aimed at the repository and digital library managers.

Our repository model has two additional goals: to abstract the underlying technologies and to deal with the advanced features required by multimedia content. The motivations for the first are the support for applications in diverse domains and the ease of maintenance as technologies evolve. The second treats the specificities of multimedia content, both on the required metadata and on the possibilities for content indexing.

Each of the standards mentioned in Section 2.1 relies on its own model. Metadata reference models are also available in several domains [15, 16, 25]. Adopting the model for one of the standards leads to a specialization in its application domain. Reference models, on the other hand, are not focused on the repository, having a much broader scope within the organizational workflow.

We have chosen to design a compact model which captures concepts from several relevant standards, accounts for structured documents and is amenable for implementation. This has required the identification of the core concepts, the inclusion of descriptors from diverse standards and the design of a generic hierarchy for documents. The model integrates descriptive and content analysis metadata for multimedia documents, to support the automatic indexing and retrieval tasks. It is aimed at state-of-the-art technologies and can be supported in any current database management system.

3.1. The Concepts. The multimedia model is organized around four main principles. The first one is that multimedia documents are usually organized in a *part-of hierarchy*. To each level one can associate an attribute set characterizing them. These attributes typically cover aspects related to the creation context, format, support and access conditions.

The second principle is that of *uniform description*, whereby the same set of attributes is used for an individual document, for composite documents, and for sets. This principle has been followed in the standards for archival description such as ISAD [13] and is very useful in the representation of large collections: metadata is frequently available for sets of documents rather than individual ones, and inheritance can make it useful further down the hierarchy.

The third principle applies to actual multimedia data. It is concerned with the internal structure of individual documents content. The actual content is stored in one or more *segments* which are parts of multimedia documents.

The segments can be analyzed by appropriate tools, which generate specialized *descriptors*. For example, the video track segment of a specific multimedia document may have associated motion activity and color descriptors, while the audio track segment of the same document may be connected to a melody contour descriptor. The fourth principle states that the descriptor resulting from the analysis of some feature of a segment is expressed as an XML file of an appropriate format.

According to these principles, four main concepts were selected. The first one, the Description Unit (DU), is already present in archival description [13], corresponds to the concept of Digital Item in the audiovisual standards and captures the notion of a multimedia document or collection of documents with an associated context.

DU's are organized in hierarchies that may have various topologies and different semantics for their levels. Such hierarchies can be created for new collections and can be extracted from existing ones. In both cases they capture the nature of the collections. The second concept is a Scheme that defines the possible levels, their semantics and their interconnections is proposed. Figure 1 shows a sample from a part-of hierarchy and the corresponding scheme, adopted for a collection of historic documents. In this case, the archive manager designs the levels of the hierarchy in the "Scheme Level" and the archivists create the actual DU's, each with an assigned level.

The third concept is the Segment, following the MPEG-7 vocabulary, that captures the notion of some part of an actual multimedia document, such as a video sequence reused in a new documentary work. A segment has no context of its own, getting it from the DU of the document it belongs to.

The fourth concept is that of a Descriptor. The sense in which Descriptor is used is the one established by the MPEG-7 standard—a representation of a feature [14]. A Descriptor results from the analysis of a Segment. An image Segment, for example, can be associated to its corresponding instances of the *DominantColor* and *NumberOfFaces* descriptors, a video Segment can be associated to its *MotionActivity* descriptor and an audio Segment to its *MelodyContour*.

3.2. The Data Model. Figure 2 shows a simplified version of the data model that evolved from an early version [26, 27]. The concepts in the model are associated with the main classes. Control of the hierarchy is provided by the Scheme Level class. Each Description Unit is of a specified level; the structure of the levels, omitted in the simplified model, allows the application development platform to enforce creation and structure of the DU's according to the repository schema. In Figure 1, for instance, a Description Unit at the level of Document can be a direct descendant of instances of Collection or Series, but not of Fonds. Attributes such as title, author, date, and copyright apply at the fine-grained level of the document as well as at the coarser grain of the whole collection. They are captured as attributes in the Description Unit and appear uniformly at all levels.

The Segments class embodies the corresponding concept, modeling documents facets for the corresponding Description Units. The association between Description Units and Segments is visible as the Contents class. Segments are further specialized as text, image, video, and audio.

Descriptors such as *MotionActivity*, *ColorLayout* and *MelodyContour*, from the MPEG-7 set, are likely to be used only for specific kinds of Segments. The Descriptor class represents them and the association of segments to descriptors is modeled with the Descriptor Instances class.
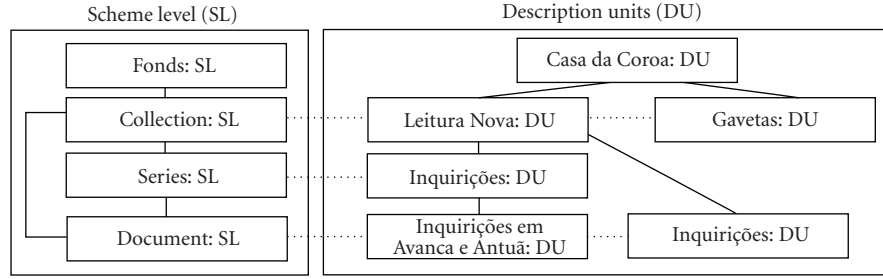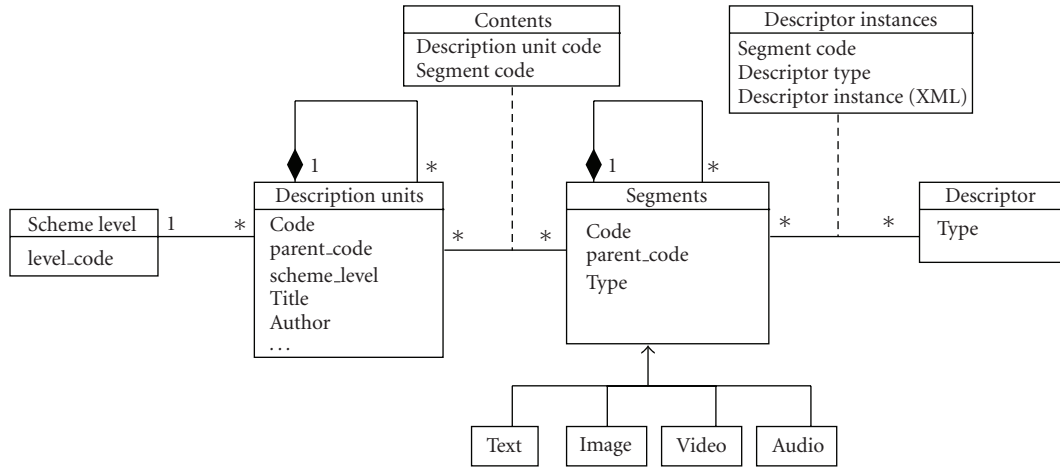
FIGURE 1: Scheme example.

FIGURE 2: The MetaMedia model.

In the model, the Description Units and Segments classes are very similar in structure: both offer a hierarchical organization of their documents. The main distinction lies in the nature of the documents and in the kind of associated metadata. An instance of Description Units captures a document for which a well-established description is available, and which has been related to other documents according to the repository hierarchy. An instance of segment is appropriate for representing an image whose description is in the associated DU and for which some automatic low-level descriptors have been produced. The part-of structure for the segments has no predefined structure and is intended to follow the granularity of the existing analysis tools. A Description Unit for a document may have one segment for each document page, and one of these segments may have a subsegment for a page detail which has been analyzed in detail.

For interoperability reasons, the MetaMedia system was intended to be MPEG-7 compliant. To some extent this is already a fact. For example, we import MPEG-7 as XML descriptors and export MetaMedia descriptors in MPEG-7 format if required. Note that currently most MPEG-7 materials result from low-level feature extraction, and have the form of descriptors. In MPEG-7 terminology, that means we have to deal mostly with the Content Description part of the standard, and especially with the visual features. The MPEG-7 Content Description artifacts that are captured in our model are Segment, subSegment, and Descriptor.

The other MPEG-7 parts, such as Content Organization, Content Management, Navigation and Access, and User Interaction, have not been extensively adopted in current multimedia description. Therefore, we had no opportunity to test their use in our model, nor to export multimedia items in full MPEG-7 format. However, these parts of MPEG-7 are also covered in standards such as ISAD and ISAAR, which we also support in our model. Classes such as Description Units, Scheme, and Scheme Level are especially designed for such purposes.

4. Managing the Repository Workflow

The MetaMedia model builds on the standards to provide a structure for the documents, their descriptors. Descriptive metadata such as the title and copyright for a document are accessible in the archivist's interface. Content metadata such as a color descriptor or a set of textual tags are obtained with extraction tools which require an offline processing of all or part of the repository collection.

Content analysis tools create different kinds of descriptors. Our approach has been to store the descriptors in XML in the Descriptor class in the model. This allows the incremental addition of descriptors with arbitrary structure without changing the model. Automatically extracted descriptors are not visible in the user interface: most of them would be meaningless as they concern specialized aspects

of the content or contain numeric structures. Depending on their nature, they are handled by specialized indexes manipulated by the browse and retrieval modules.

A collection of text documents may have extracted text, kept as a segment for each document, and also automatically extracted textual descriptors based on a controlled vocabulary, stored in another descriptor for content annotations. These textual segments are indexed with a text indexing tool accessible in the retrieval interface.

Non-text documents may have multidimensional descriptors such as *ColorLayout* or *SIFT* [28], automatically obtained for some of the video segments. These multidimensional descriptors are indexed with a specialized signature index designed to handle multiple high-dimensional descriptors [29]. Search based on the visual content is also integrated in the retrieval interface.

Content-based retrieval of textual documents has succeeded in providing useful answers to common queries based on automatic indexing. Content-based retrieval of audiovisual materials has to deal with the "semantic gap" [30], the mismatch between the low-level content descriptors and the high-level concepts required by the search tasks. The approaches to bridge it fall into two categories. The first one is to perform query-by-example, which implies stating the information need as an image or a set of images, and using them as examples. The answer is computed in the domain of low-level descriptors, in the form of nearest-neighbor queries over high-dimensional spaces. In a way, this approach avoids the semantic gap by making queries fit more closely the document content.

The second approach is to have keyword queries that have to be matched to documents represented by low-level descriptors. The matching can be based on the derivation of high-level features from the low-level descriptors and the use of text-based techniques. It can also be based on more indirect methods where high- and low-level descriptors are used together.

Our approach is hybrid and consists of two steps. First, we use the high-level features to guide the search at the coarser level. The resulting documents can be used as query examples in a second, low-level based step, for purposes such as query refinement or focus change. This approach helps bridging the semantic gap because the set of documents obtained after the first search step can be used as good query examples for the second search step.

However, such an approach is highly dependent on the quality of the first search step, based on the high-level features. It is already known that the high-level features can be obtained by means of either automatic or manual annotations. The automatic annotations are cheap to obtain when trained concept detectors already exist. However, such detectors is available only for a very small number of concepts and often provide low accuracy rates. On the other hand, the manual annotations are considered subjective and expensive to obtain, but if domain experts validate the annotations, they can be accurate [3].

Considering that, for our dataset, the source of the high-level features is expert-based manual annotation, thus accurate, it is appropriate to start video search based on them.

5. The "Terra De Santa Maria" Documentation Center

The "Terra de Santa Maria" historic documentation center [31] has been a case study for the MetaMedia platform. The collection is a virtual archive of medieval documents, for which there are transcriptions in either Latin or archaic Portuguese. Information for a document in the repository includes three parts: the digitized image, the document transcription and the archival description according to the ISAD/ISAAR standards.

The interface is available in English and in Portuguese, and the archival descriptions have been generated in Portuguese. There are several views on the document, intended for different kinds of users.

Figure 3 shows one of the application views, namely the *Tour* tab. It provides a view of the collection at the document level, browsing through the documents while viewing their images, descriptions and position in the hierarchy. On the *Archive* tab, it is possible to browse the structure of the archive, designed by archivists, and to edit the descriptions. The *Creators* tab has detailed information on the creators for the current unit. Under the *Documents* tab users may explore the contents of the documents. A medievalist studying one of the parchments uses this mode to observe the digitized image and its transcription side by side and to possibly upload his own analysis of the text.

The document transcriptions, captured as textual segments, have been manually analyzed by the history specialists who have made annotations using a dedicated annotation tool [32] and a custom-designed thesaurus. The thesaurus has been used to assists the transcription annotation process, thus enriching transcriptions with labels such as "person", "place", "institution" or "date". Figure 4 illustrates some highlighted regions of a fragment of a text document that was subject to the process of annotation. The resulting markup identifies key concepts that would be hard to spot on the original Latin documents and is mainly intended for retrieval purposes. Some of the highlighted words in the figure are marked as *name*s, where the *name* concept comes from a locally defined controlled vocabulary. Similarly, other highlighted words were marked as *locations*.

The collection has been processed for content-based search and retrieval. Textual materials from the descriptive metadata, from the transcription segments and from their annotations were processed to build a text index.

6. Visiting the Collection

Interesting documents are essential when we want to offer access to a digital repository, and the available metadata may allow rich views on their contents. The MetaMedia platform has been implemented as a web portal supported on a database system. The interface supports a contextual view on collections and documents.

The most straightforward look at the collection is obtained on the *Tour* tab, as shown on Figure 3. Each document in the presented sequence is viewed together
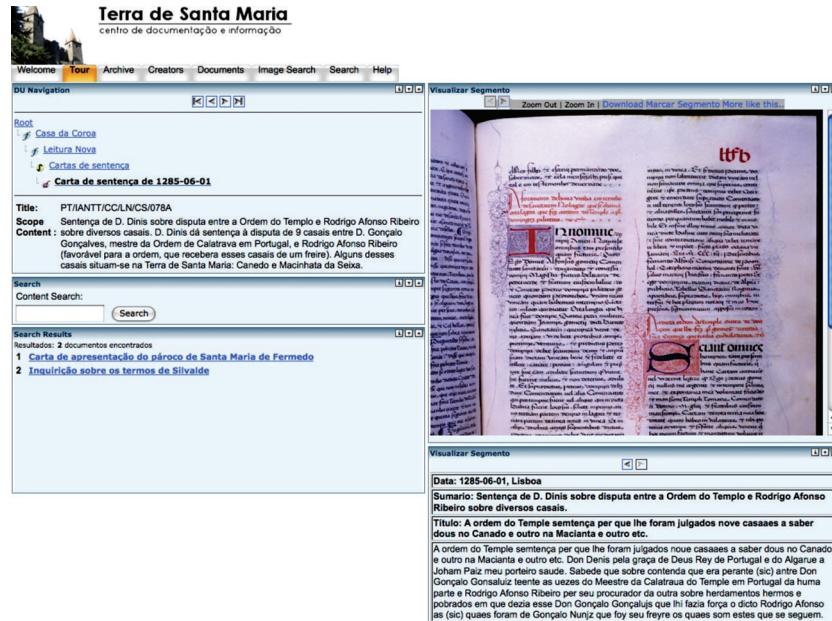
Figure 3: Interface of the documentation center.

Figure 4: The annotation process.

with its position in the tree, the contextual metadata and the transcription text. The archival view is obtained in the *Archive* tab, where the complete set of descriptive metadata is provided, as shown in Figure 5.

The complementary view, using the contents, is available in the *Documents* tab, where the segments pertaining to the current Description Unit are visible as shown in Figure 6. It is possible to browse through the segments and view the sequence of pages for a document. An expert analyzing the image can compare it with the transcribed text or even download a high-resolution version for a closer examination.

In any of the above mentioned contexts, there is a *More like this* link taking the visit to the *Image Search* tab where segments with image features similar to the current one are searched, as shown in Figure 7.

Any switch between the various document facets maintains the context of the current document, and we can therefore locate a document while browsing the tour, looking at its complete description in the *Archive*, hiting the *More like this* contextual search and launching a visual similarity search that may take us to new documents.

7. Searching and Browsing

A multimedia repository must offer several retrieval modes. The user may search on textual content for textual segments of the documents, on the structured contextual metadata available as descriptors in the Description Units, and on the visual features for image segments of the documents.

Query by keyword is the most straightforward and requires a full-text indexing system. The platform uses the Apache Lucene technology [33], configured to index selected parts of the available textual information: the textual content itself and parts of the descriptive information. A structured query interface is intended for specialized users, who are aware of the meaning of the descriptive metadata in the Description Units. Search on the contextual metadata is handled by the built-in indexes of the relational database management system.

Query by visual features is integrated with the keyword search. In a collection where documents have textual segments, keyword search provides an initial answer which can be expanded by visual similarity. A collection with just image or video documents can also be searched with textual queries, provided that some minimal concept annotations are present. Concepts in the query are extracted and matched against the annotations, resulting in an initial set of documents. Image and audio similarities are then used to expand the answer and a relevance feedback interface is offered to refine the query. The list of low-level descriptors includes the MPEG-7 color and texture descriptors, namely ColorLayout, ColorStructure, ScalableColor, ColorMoments,
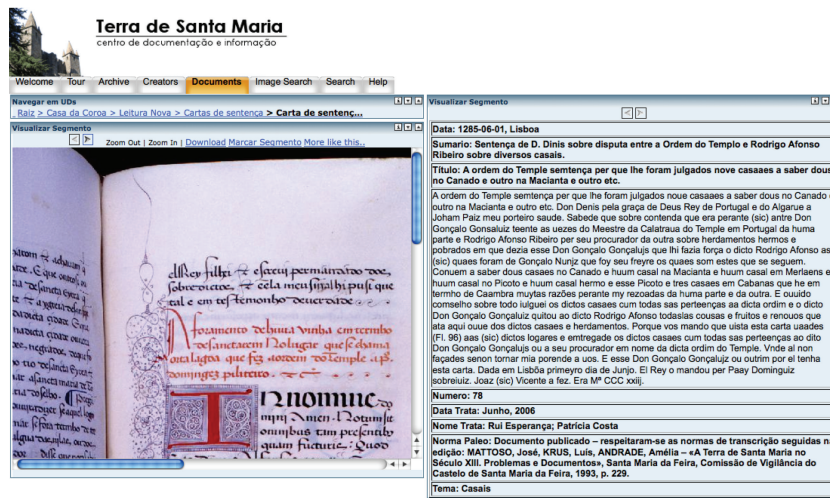
FIGURE 5: Viewing descriptive metadata.

FIGURE 6: Viewing document contents.

EdgeHistogram, and Homogeneous Texture. The "BitMatrix" multidimensional indexing technique [29] is used to ease the computation of image similarity and to allow the tuning of the set of descriptors to the nature of the collection.

When collections have substantial structure, browsing can be an effective way of guiding a retrieval process. The MetaMedia platform keeps the context of documents being visited, and it is therefore possible to browse the hierarchy, locate an interesting subcollection, view its description, travel down to a document and analyze its visual or textual content. From a selected image or video segment it is also possible to start a new search using low-level similarity.

The effectiveness of our retrieval system, MetaMedida, depends on the performances of each retrieval method, mainly the query by keyword, and query by visual features. The first one, as a text-based search, is expected to perform at the level of the current state-of-the-art text indexing systems. Its performance depends on the accuracy of the high-level features used to describe the documents contents. Considering that in our case the high-level features are manually produced by experts, it is appropriate to search based on them. The second search modality is based on low-level features. The retrieval quality of such approaches has already been evaluated [29, 34]. The experiments carried on the TRECVID dataset [34] show that, although the effectiveness depends on query examples, features, and similarity metrics, query by visual features brings significant improvements to the retrieval experience. These promising results encouraged us to apply the same approach on our collection of digitized historic documents, without having a specific evaluation benchmark.

8. Conclusions

Cultural heritage collections may include very diverse materials, with descriptions that conform to different standards. On the other hand collections are becoming more dynamic, and the rate at which new materials appear may not be matched by the available resources for description. It is therefore necessary to organize, manage, and search heterogeneous collections with possibly incomplete metadata and document organization. Our proposal is to integrate standardized description and organization with automatic content-base browsing and searching to maximize collection accessibility. In a collection structured with subcollections, it is possible for a user to browse the structure, locate
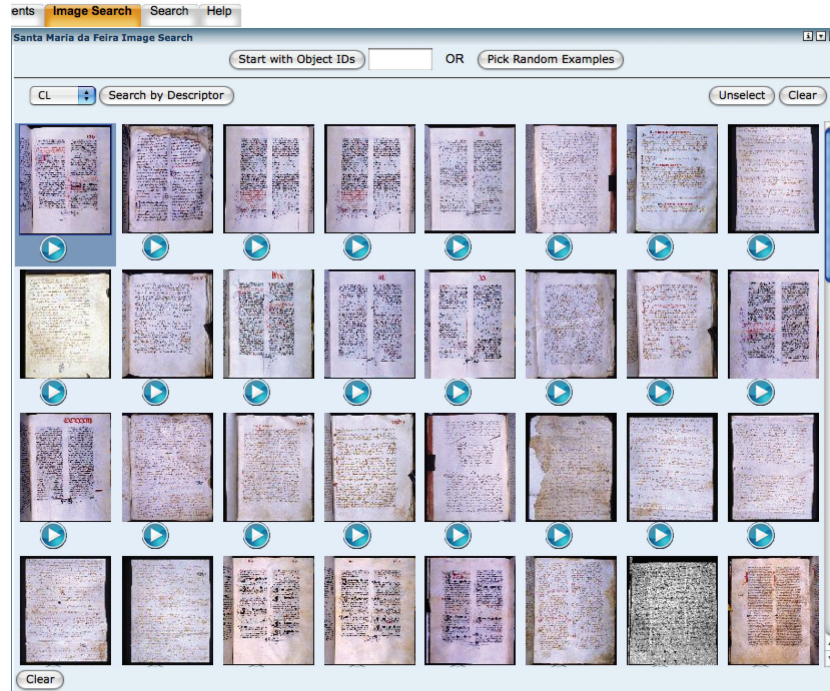
FIGURE 7: Search by image similarity.

an interesting subcollection, identify a document, search for similar ones based on the description, on the textual content or on the visual features, and then focus on the new documents which can also be located in the hierarchy and explored in their relationships with the collection. A document with a complete description has more facets to be explored and ways to be located, but a document which has just been incorporated into the collection can also be accessed using the automatically extracted metadata.

The MetaMedia model and platform have been used to address several of the issues mentioned above. The model accounts for a clear identification and integration of document structure, descriptive information, and the so-called segments containing descriptors that result from content analysis. The platform provides an environment where sets of documents are loaded into a collection, subject to content analysis resulting in text or image indexes and available for the upload of further analysis results. The user browsing the collection can navigate in the collection hierarchy, search—in descriptive metadata or actual content—for documents with specified features, and explore similarity between documents in any of their facets.

The collection used as a case study is a set of medieval documents pertaining to a once powerful region in the north of Portugal and were made available online as a document center. The documents are written in Latin or archaic Portuguese, and the calligraphy used requires paleographic skills. The documents had been studied by scholars who produced document transcriptions. A team of historians and archivists worked on the collection to generate a hierarchic structure for the documents, their descriptions and annotations on the content. All these

sources of document information have been integrated in the MetaMedia platform. The document repository can be managed and enriched by their curators, who may introduce descriptions and add new content descriptors. Generic users can access the document center, browse its structure, search the structured information in the descriptors and annotations or find documents by similarity on their textual or visual features.

Many innovative approaches are being proposed for multimedia navigation. The MetaMedia platform can be improved by plugging in new visualization facilities. The model provides the core concepts for linking documents to the collections and to their descriptors. In cultural heritage applications, one aspect to further explore is the use of collaboration for description and annotation.

The overall performance of our system should ideally be evaluated in user-oriented tests. It has been argued that an important evaluation measure is the user satisfaction [35, 36]. In order to measure it, we need to set up an experimental environment where a large set of users will have their behaviors monitored. This kind of evaluation requires a preliminary study to identify relevant variables for our domain, which we expect to be able to do in the future.

References

[1] K. Beyer, R. J. Cochrane, V. Josifovski, et al., "System RX: one part relational, one part XML," in *Proceedings of the ACM International Conference on Management of Data (SIGMOD '05)*, pp. 347–358, ACM Press, New York, NY, USA, 2005.

[2] M. Stonebraker and D. Moore, *Object Relational DBMSs: The Next Great Wave*, Morgan Kaufmann, San Francisco, Calif, USA, 1995.

[3] B. Shneiderman, B. B. Bederson, and S. M. Drucker, "Find that photo!: interface strategies to annotate, browse, and share," *Communications of the ACM*, vol. 49, no. 4, pp. 69–71, 2006.

[4] M. A. Gonçalves, E. A. Fox, L. T. Watson, and N. A. Kipp, "Streams, structures, spaces, scenarios, societies (5S): a formal model for digital libraries," *ACM Transactions on Information Systems*, vol. 22, no. 2, pp. 270–312, 2004.

[5] DSpace, "The DSpace digital repository system," 2007, http://dspace.org/.

[6] C. Lagoze, S. Payette, E. Shin, and C. Wilper, "Fedora: an architecture for complex objects and their relationships," *International Journal on Digital Libraries*, vol. 6, no. 2, pp. 124–138, 2006.

[7] I. H. Witten and D. Bainbridge, "Building digital library collections with Greenstone," in *Proceedings of the Joint Conference on Digital Libraries (JCDL '05)*, p. 425, 2005.

[8] T. Westerveld, "TRECVID as a re-usable test-collection for video retrieval," in *Proceedings of the Multimedia Information Retrieval Workshop*, 2005.

[9] A. B. Benitez and S. F. Chang, "Multimedia knowledge integration, summarization and evaluation," in *Proceedings of the International Workshop on Multimedia Data Mining (MDM/KDD '02)*, pp. 39–50, 2002.

[10] C. G. M. Snoek, M. Worring, O. D. Rooij, K. E. A. van de Sande, R. Yan, and A. G. Hauptmann, "VideOlympics: real-time evaluation of multimedia retrieval systems," *IEEE Multimedia*, vol. 15, no. 1, pp. 86–91, 2008.

[11] T. Deselaers, D. Keysers, and H. Ney, "Features for image retrieval—a quantitative comparison," in *Proceedings of the 26th DAGM Symposium on Pattern Recognition (DAGM '04)*, Lecture Notes in Computer Science, Tbingen, Germany, 2004.

[12] Library of Congress, "MAchine-Readable Cataloguing (MARC)," 2007, http://www.loc.gov/marc/.

[13] ISAD(G), "General International Standard Archival Description," 2nd edition, 1999, http://www.ica.org/.

[14] J. M. Martínez, R. Koenen, and F. Pereira, "MPEG-7: the generic multimedia content description standard, part 1," *IEEE Multimedia*, vol. 9, no. 2, pp. 78–87, 2002.

[15] International Council of Museums, "CIDOC Conceptual Reference Model," 2006, http://cidoc.ics.forth.gr/.

[16] DELOS Network of Excellence on Digital Libraries, "A Reference Model for Digital Library Management Systems," 2007, http://www.delos.info/.

[17] J. Hunter, "Combining the CIDOC CRM and MPEG-7 to describe multimedia in museums," in *Proceedings of the Conference on Museums and the Web, Archives and Museum Informatics (MW '02)*, 2002.

[18] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 703–715, 2001.

[19] J. Yu, J. Amores, N. Sebe, and Q. Tian, "A new study on distance metrics as similarity measurement," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '06)*, pp. 533–536, 2006.

[20] Library of Congress, "Metadata Encoding and Transmission Standard (METS)," 2007, http://www.loc.gov/standards/mets/.

[21] Dublin Core Requirements Group, "Dublin Core Metadata Initiative," 2007, http://dublincore.org/.

[22] W3C Consortium, "Semantic Web," 2007, http://wwwsw/.w3.org/2001/sw/.

[23] I. Burnett, R. V. de Walle, K. Hill, J. Bormans, and F. Pereira, "MPEG-21: goals and achievements," *IEEE Multimedia*, vol. 10, no. 4, pp. 60–70, 2003.

[24] CONTENTdm, "CONTENTdm—Digital Collection Management Software," 2009.

[25] TEI Consortium, e., "TEI P5: Guidelines for Electronic Text Encoding and Interchange," 2007.

[26] C. Ribeiro and G. David, "A metadata model for multimedia databases," in *Proceedings of the International Cultural Heritage Informatics Meeting, Archives and Museum Informatics (ICHIM '01)*, 2001.

[27] C. Ribeiro, G. David, and C. Calistru, "A multimedia database workbench for content and context retrieval," in *Proceedings of the IEEE International Workshop on Multimedia Signal Processing (MMSP '04)*, IEEE Computer Society Press, 2004.

[28] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[29] C. Calistru, C. Ribeiro, and G. David, "Multidimensional descriptor indexing: exploring the BitMatrix," in *Proceedings of the International Conference on Image and Video Retrieval (CIVR '06)*, H. Sundaram, M. R. Naphade, J. R. Smith, and Y. Rui, Eds., vol. 4071 of *Lecture Notes in Computer Science*, pp. 401–410, Springer, 2006.

[30] R. Zhao and W. I. Grosky, "Narrowing the semantic gap—improved text-based web document retrieval using visual features," *IEEE Transactions on Multimedia*, vol. 4, no. 2, pp. 189–200, 2002.

[31] Comissão de Vigilância do Castelo de Santa Maria da Feira, "Centro de Documentação da Terra de Santa Maria," 2007, http://www.castelodafeira.pt/.

[32] C. Ribeiro, G. David, and A. Barbosa, "XML annotation of historic documents for automatic indexing," in *XML: Aplicações e Tecnologias Associadas (XATA '06)*, pp. 325–336, Universidade do Minho, Portalegre, Portugal, 2006.

[33] Apache Software Foundation, "Apache Lucene 2.2," 2007, http://lucene.apache.org/.

[34] C. Calistru, *Data organization and search in multimedia databases*, Ph.D. thesis, University of Porto, 2009.

[35] M. Sanderson and J. Zobel, "Information retrieval system evaluation: effort, sensitivity, and reliability," in *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '05)*, pp. 162–169, ACM Press, New York, NY, USA, 2005.

[36] J. Tague-Sutcliffe, "The pragmatics of information retrieval experimentation, revisited," *Information Processing and Management*, vol. 28, no. 4, pp. 467–490, 1992.