Pham, Binh L. and Smith, Robert (2007) A Metadata Augmentation for Semantic and Context-Based Retrieval of Digital Cultural Objects. In *Proceedings 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications*, pages pp. 515-522, Glenelg, South Australia.

A Metadata Augmentation for Semantic- and Context-based Retrieval of Digital Cultural Objects

Binh Pham and Robert Smith
Faculty of Information Technology, Queensland University of Technology
GPO Box 2434, Brisbane Qld 4001 Australia
E-mail: b.pham@qut.edu.au

Abstract

Cultural objects are increasingly stored and generated in digital form, yet effective methods for their indexing and retrieval still remain an open area of research. The main problem arises from the disconnection between the content-based indexing approach used by computer scientists and the description-based approach used by information scientists. There is also a lack of representational schemes that allow the alignment of the semantics and context with keywords and low-level features that can be automatically extracted from the content of these cultural objects. This paper presents an integrated approach to address these problems, taking advantage of both computer science and information science approaches. The focus is on the rationale and conceptual design of the system and its various components. In particular, we discuss techniques for augmenting commonly used metadata with visual features and domain knowledge to generate high-level abstract metadata which in turn can be used for semantic and context-based indexing and retrieval. We use a sample collection of Vietnamese traditional woodcuts to demonstrate the usefulness of this approach.

.

1. Introduction

Cultural objects are intellectual or artistic creations which contain imagery and non-imagery elements. As cultural objects are increasingly stored in digital forms, there is an impetus to develop intelligent software systems and languages to improve their representation and management, and thus widen their access. Users can vary from the general public (e.g. to find a suitable image or a text passage for inclusion in a document) to professionals (e.g. art historians, critics, archivists, librarians). For a deeper analysis of cultural works such as the study of aesthetics, art appreciation, history, and cultural and societal influences, many facets of information are required beyond what is offered via keywords and common basic metadata. For example, it is desirable to allow search by category, search by association, and search for symbolic meanings. Category search allows a thematic search to explore a concept in depth, e.g. finding collections of cultural objects which depict "romance" in the 18th century in Europe and in America. Search by association is useful for comparative analysis and exploration of further relationships, e.g. finding cultural objects belonged to a certain creator and his or her students. Symbolic meanings convey the essence of each culture and are often woven in the works. For example, in Chinese and Vietnamese culture, "a peach" is a symbol for "longevity" and "red colour" implies "luck". These symbolic meanings are usually not obvious and knowledge about them needs to be stored separately and used in conjunction with other information. In addition to symbolism, works might be distinguished by higher abstract concepts (e.g. works that evoke a particular mood such as happiness or despair, or some aesthetic values such as dynamic or balance). Furthermore, there is a need to incorporate information from other elements of the cultural objects such as narratives, experts' and creators' notes and letters.

To date, researchers from two distinct disciplines – computer science and information science – mostly work separately and follow different paths. This disconnection of research efforts has significantly hindered the progress to seek for more effective indexing schemes which closely satisfy the needs, workflow and tasks of users, especially in arts and humanity domains. Chu [3] has confirmed this problem through his study of over 800 articles published during the last 30 years. What needed is an integrated approach that takes advantage of both of these approaches to allow high-level context be used with automatically extracted features (both textual and

0-7695-3067-2/07 $25.00 © 2007 IEEE
DOI 10.1109/DICTA.2007.27

visual) from the content in order to advance the effectiveness of indexing, retrieving and querying.

The aim of this paper is to present a holistic approach to resolve these issues. To this end, we need to develop semantic and context-based models for the cultural objects to represent their essence. This paper focuses on the rationale and design of the whole system based on this underlying model and demonstrates how the system components link together to produce the desired outcomes. It also discusses the process of metadata augmentation and provides the results of a case study.

Section 2 briefly describes how cultural objects are currently catalogued and also provides an analysis of the requirements, viewing from a number of perspectives: user, task and system. Section 3 presents our integrated approach and an overview of our system and its components. Techniques for metadata transform and augmentation are covered in Section 4. To demonstrate this approach, we show in section 5 how the retrieval of Vietnamese traditional woodcuts based on semantics, context and symbolism can be achieved. Conclusion and future work are given in section 6.

2. Requirement Analysis

Traditionally, cultural objects collections are manually classified and annotated using either free text or controlled vocabularies. Much research has been carried out to find effective ways to catalogue these collections in order to facilitate their classification and retrieval. A recent effort of note is the EPOCH program (European Research Networks of Excellence in Open Cultural Heritage) which establishes a network of more than 100 European cultural institutions to explore ways to improve the quality and effectiveness of deploying technology for cultural heritage [5]. Another good initiative is the Cataloguing Cultural Objects (CCO) project, sponsored by the Visual Resources Association, which provides data content standards and guidelines for describing and cataloguing cultural works and their associated images [1]. The goal of this initiative is to promote cataloguing best practices for the cultural heritage community. We comply with the CCO data content standards for describing cultural objects for our own work. In CCO, a work may contain multiple parts or be created in a series, and an image is a visual representation of a work or a part of a work. Related works have important conceptual relationships with each other. These relationships may be intrinsic (e.g. whole-part, group collection, series) or extrinsic (e.g. same period (temporal), seen together (spatial), same theme (conceptual)). To ensure consistency in the use of information between item records, the CCO also recommends the use of authority files and controlled vocabularies. The authority files contain ancillary information about things and concepts related to the works. A controlled vocabulary contains preferred and variant terms with limited scope or specific domain. Some examples of controlled vocabularies are: taxonomy, thesaurus, subject headings, and synonyms.

Most retrieval schemes currently deployed are keyword-based. One disadvantage of this approach is its adhoc nature. While it is a useful tool for expert users who know why they seek the works and how to judge the usefulness of the retrieved works with respect to their goals, the situation is not the same for novice users who need a more systematic classification and interpretation of these cultural objects. Hence, it is desirable to develop appropriate tools to guide users through specific paths of query, or suggested related paths of query.

Digital cultural images are rich in both attributes and meanings. Users may be interested in their intrinsic characteristics such as attributes that are contained in the works and may be visually perceived, e.g colour, object features, shape, texture, spatial composition, viewing angle and object category. It would be useful if such information can be extracted automatically.

Depending on their tasks, users may also be interested in extrinsic information such as art and historical information (e.g. creator, medium, style and period); physical properties of the work (dimension, type of cultural objects e.g. painting, drawing, photos and material e.g. canvas, silk); administrative information (e.g. storage location, people and organization having relationships with the work); technical information (e.g. digital format, resolution, digitized process); transactions (e.g. time of purchase and people involved); or legalistic information (e.g. copyrights, ownership, preservation). At a more abstract level, affective factors such as emotion (e.g. warm, romantic, sad) and aesthetic (e.g. harmony, balance, expressive, dynamics) are also relevant. Such information can be described in terms of metadata. Hence, appropriate metadata categories need to be defined in order to capture the essence of the works.

Besides aesthetic expression and appreciation, the works might also reflect cultural changes and public values. They help to identify social trends,

relationships and influence. An effective system for indexing and retrieval of these works should be user-centered, allowing users to perceive, recognize, interpret and analyse the meanings of the contents as well as the factors that provoke emotional, aesthetic and cognitive responses. It should also allow users to explore, navigate from different view points, and to refine the search according to specific thematic view points.

The association between information objects - both intrinsic and extrinsic - and visual features are essential for art and cultural study. For example, certain colour schemes belong to a certain period or school of art; certain symbols or emblems have specific meanings and should be interpreted together. The cross references between different cultural objects that are related in certain ways also bring more insight into the works. This would contribute at a deeper level to the task of thinking and create new knowledge using these works (one such example is described by Weedman [12] who analysed how a social scientist used images to create a new understanding of how canals were built in the 17th century).

From the perspective of a content provider and manager, the system should facilitate the capture or generation of keywords, metadata and information on themes and context that would be useful for query specification and search. As the preparation of such information is tedious and costly if done completely by manual means, a systematic and structured way for extracting such information from existing metadata, controlled vocabularies and narratives is desirable. Conversely, if such structures are well-defined, new narratives may also be generated from visual features, keywords and metadata extracted from the digital artworks. In addition, users need to be able to communicate queries to the system and the system needs to be able to return the search outcomes in an intuitive and meaningful way. Thus, a simple, scalable and structured language based on natural language should be designed to serve these needs.

3. An Integrated Approach for Semantic and Context-based Retrieval

To integrate many facets of information required for semantic and context-based retrieval, we construct a representation model for cultural objects which broadly has three levels of information (as shown in Fig.1).

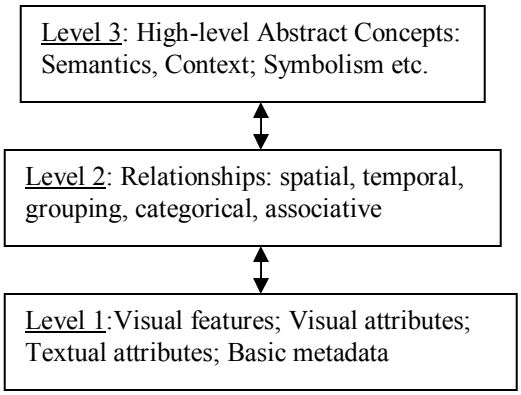


Fig.1. Information levels for cultural objects

The lowest level contains visual features, visual attributes, textual attributes, and basic metadata. They provide information about the *content* of the object. Visual information forms the basic elements of images associated with cultural objects. They consist not only of colour, object shape features, texture, illumination, but also other information that can be discerned visually such as medium, brush stroke type, and material. To distinguish these two types, for convenience, we call the first type "visual features" which may be automatically extracted using computer vision techniques, and the second type "visual attributes" may be expressed in text as keywords or metadata. Basic metadata covers different information aspects:

- Intrinsic: e.g. physical properties – colour, texture, shapes, arrangement, composition, viewing angle, dimension, type (canvas, glass frame, booklet); type of cultural object (painting, drawing, letter, note).
- Extrinsic:
    - Information concerning who (e.g. artist), where (e.g. source), when (e.g. period), how (e.g. medium, style).
    - Administrative information: e.g. location of storage, people having relationships with the works
    - Transactions: e.g. time of purchase, delivery, participating parties.
    - Legalistic information: e.g. preservation, copyrights, ownership, IP rights.
    - Technical information: e.g. process used to capture digital objects, digital storage format, encoding, resolution, camera specification, colour, illumination.

The second level deals with various types of relationships; spatial, temporal, categorical and associative. The spatial relationships allows the formation of an object from its component. Temporal relationships connect work created in the same period, or in periods which have a close relationship (e.g. the revival of design styles of the 60s in 21st century). Categorical relationships link works belonging to the same theme or subjects. Association rules connect basic metadata to form concepts at a higher level (e.g. the co-occurrence of ink, brush, and silk implies a certain type of painting). This level corresponds to the concept of "Related Works" defined by the CCO.

The third level deals with high level abstract concepts which can be deduced by integrating visual features, visual and textual attributes, object categories, with appropriate types of metadata and relationships. For example, a happy, peaceful country lifestyle scene is depicted by children playing and peasants resting surrounded by farm animals. Three main high level abstract concepts often found in cultural objects are: *Semantics*, *Context* and *Symbolism*. Semantics describe the meaning of the content, while Context provides the extrinsic information that would influence or change the meaning of the cultural object. For example, the meaning of a cultural object would be interpreted differently under different circumstances such as the period under which it was created or which country the creator came from. Symbolism provides other hidden meanings or messages intended by the creator. Symbolic meanings often come from cultural traditions or customs.

To enable users to retrieve cultural objects based on these abstract concepts, we develop an abstraction transformation engine which can generate more complex and abstract metadata and use them as indices for retrieval. A schematic diagram for the whole system is shown in Figure 2. To automatically extract prominent visual features from the content of the imagery component of the cultural objects, computer vision and image processing can be used. We have successfully extended the approach for representation and detection of deformable shapes by Felzenszwalb [6] for this purpose. In particular, we have advanced his work further by integrating heuristic knowledge to the deformable templates for objects in order to improve the performance deficiency and accuracy of the search. Further details on this work may be found in [10]. Metadata from existing databases can be ingested into the system and sorted into Level 1 and Level 2 types of metadata. The Abstract Transformation Engine (ATE) will combine these two types of metadata, visual features together with other information provided by creators or experts (e.g. curators, librarians, information cataloguers) to infer Level 3 type of metadata. The ABT contains 2 main modules: the Transformation Module and the Inference Module. The Transformation Module generates high-level metadata from the combination of low-level metadata with appropriate ontology, thesaurus, and information from other sources such as annotations and notes. The Inference Module takes care of the reasoning, resolves conflicts, vagueness and uncertainty whenever required. It works out the likelihood that the semantics, context , or symbolism of an item might belong to one type or another. The Transformation Module then assigns the output according to the results of the Inference Module.

4. Metadata Transform and Augmentation

While Dublin Core [4,9] can adequately cater for simple and precise information required for cataloguing such as basic technical and administrative data, it does not provide ways to represent the more abstract elements of artworks. Later schemes such as METS [7], MPEG7 and MPEG21 [8] provide richer structural frameworks for metadata to ensure that digital objects in library collections are preserved, but they still have shortcomings when describing abstract concepts. Furthermore, as these schemes were designed for specific purposes (e.g. METS for archiving), they have many mandatory fields that are irrelevant and that also make it cumbersome and inefficient to implement for other purposes.

To overcome these problems, we have designed a lightweight metadata wrapper schema digital works in various domains (museums, home photos, films etc.), that has a minimum number of mandatory fields and reduced depth of hierarchical levels. This schema allows the ingestion of metadata from existing metadata schemas. Such metadata once ingested will be sorted according to a core ontology provided by the system. The metadata wrapper also allows plug-ins for specific applications that deploy the ingested metadata, thus offers extensible capability for future development. We have demonstrated its effective use for a digital film collection and presented this work in a previous paper [2].

In order to provide a convenient and comprehensive means to describe a digital cultural object, a simple and structured language is also required. This language can be used for communication between users and the system through the expression of the input and output of queries. In addition, the language can be

conveniently used in the generation process of high-level abstract data from low-level metadata and other associated information. To this end, we have extended our previous work on a Digital Artworks Expression Language (DAEL) [11] by opening up the definition of digital objects to include other types of cultural objects besides artworks.

The Transformation Module was implemented using XSLT style sheets [13] which provide a convenient way of transforming XML documents. To date, we have tested three types of abstraction on a small collection of Vietnamese traditional woodcuts to determine the Content, Context and Symbolism of these works and subsequently to use such abstract concepts for more meaningful retrieval. Figure 3 shows an example of a Vietnamese traditional woodcut with narrative, visual features and the mapping of these features to symbolic meanings.

The Content can be extracted from a number of sources: notes from creators or curators, basic metadata, or automated extraction from an image using computer vision techniques. In the example woodcut, the Content includes a rooster, a child and a chrysanthemum. The Semantics of this work is a child holding a rooster and a chrysanthemum. The Context is then determined by combining the Content with other available metadata which provide information on when and where the work was created, its creator and style. In this case, the woodcut was produced during the preparation for Lunar New Year celebration with the intention to convey good wishes to the receiver. The Symbolism is then determined by matching the Content, Context with a Symbolism Thesaurus. The symbols produced for the woodcut in Figure 3 show that it conveys good wishes for fertility, wealth and longevity. Figure 4 shows the results of a query on woodcuts which depict "Rural Life", where "Rural Life" was portrayed as the inclusion or presence of animals, rice fields and planting activities.

If all information required for each transformation is available and there are no ambiguities or conflicts, then the process of generating these high-level abstract metadata to augment existing low-level metadata is straightforward. However, this is often not the case in practice, where information is either missing, or ambiguous, or simply contradicts each other. In addition, many vague queries are inherent for cultural works, for example, *"searching for cultural objects that depict romance"*. For these cases, we construct a Bayesian Inference Network (BIN) to find the best matches to the query. The BIN is initially constructed based on experts' beliefs on the category or categories that a particular cultural object should belong to. For example, the presence of a temple provides a strong evidence that the work is a religious one, while the presence of a farm animal indicates that the work might be related to country scene. The experts' beliefs can be deduced and extracted from narratives or notes associated with the work. In addition to expert knowledge, the BIN incorporates relationships between data sets, where causal linkages are assigned in order to record probability relationships between data sets. The BIN also allow learning capability via relevant feedbacks.

5. Conclusion and Future Work

We have presented an integrated approach for semantic and context-based retrieval of cultural objects via a metadata augmentation process to produce high-level abstract metadata from low-level metadata and associated information. A number of components of this system have been implemented and tested on a small sample collection of woodcuts. We are currently preparing data from a real collection of digital cultural objects for further evaluation. A Bayesian Inference Network will also be constructed based on these data sets.

Acknowledgement

The authors wish to acknowledge the support of the ARC Centre of Excellence for Creative Industries and Innovation (CCI).

References

[1] M. Baca, P. Harping, E. Lanzi, L. McRae, A. Whiteside, *Cataloging Cultural Objects: A `Guide to describe cultural works and their images*, ALA Version, 2006. http://vraweb.org/ccoweb/cco/index.html.

[2] S. Choudhury, B. Pham, R. Smith, P. Higgs, A Metadata Wrapper for Digital Motion Pictures, *Proc. Internet and Multimedia Systems and Applications (IMSA2007 Conference),* 2007, Honolulu, Hawaii, USA.

[3] H. Chu, Research in Image Indexing and Retrieval as Reflected in the Literature, *Journal of the American Society for Information Science and Technology,* 52 (12), 2001, 1011-1018.

[4] DCMI. *Dublin Core Metadata Initiative*. 2007 [cited 2007 March]; Available from: http://dublincore.org/documents/1999/07/02/dces/.

[5] EPOCH, http://www.epoch-net.org.

[6] P.F. Felzenszwalb, Representation and Detection of Deformable Shapes, *IEEE Trans. Pattern Anal. Mach. Intell.* 27(2), 208-220, 2005.

[7] *METS: Metadata Encoding & Transmission Standard*. 2007 [cited 2007 March]; Available from: http://www.loc.gov/standards/mets/.
[8] Moving Pictures Expert Group, *The MPEG Home Page*. 2007 [cited 2007 March]; Available from: http://www.chiariglione.org/mpeg/.
[9] J. Riley and A. Hutt, *Semantics and syntax of dublin core usage in open archives initiative data providers of cultural heritage materials*, in *Digital Libraries, 2005. JCDL '05. Proceedings of the 5th ACM/IEEE-CS Joint Conference on*. 2005. p. 262--270.
[10] R. Smith, B. Pham, A Robust Object Category Detection System using Deformable Shapes, *Journal of. Machine Vision and Applications,* 2007, in print.
[11] R. Smith, B. Pham, S. Choudhury, A Digital Artwork Expression Language (DAEL), *Proceeding of Internet and Multimedia Systems and Applications (IMSA2007 Conference),* 2007, Honolulu, Hawaii, USA.
[12] J. Weedman, Thinking with Images: An Exploration into Information Retrieval and Knowledge Generation, *Proceeding of ASIST 2002 (2002),* 376-382.
[13] W3C, The Extensible Stylesheet Language Family, http://www.w3.org/Style/XSL (cited March 2007).

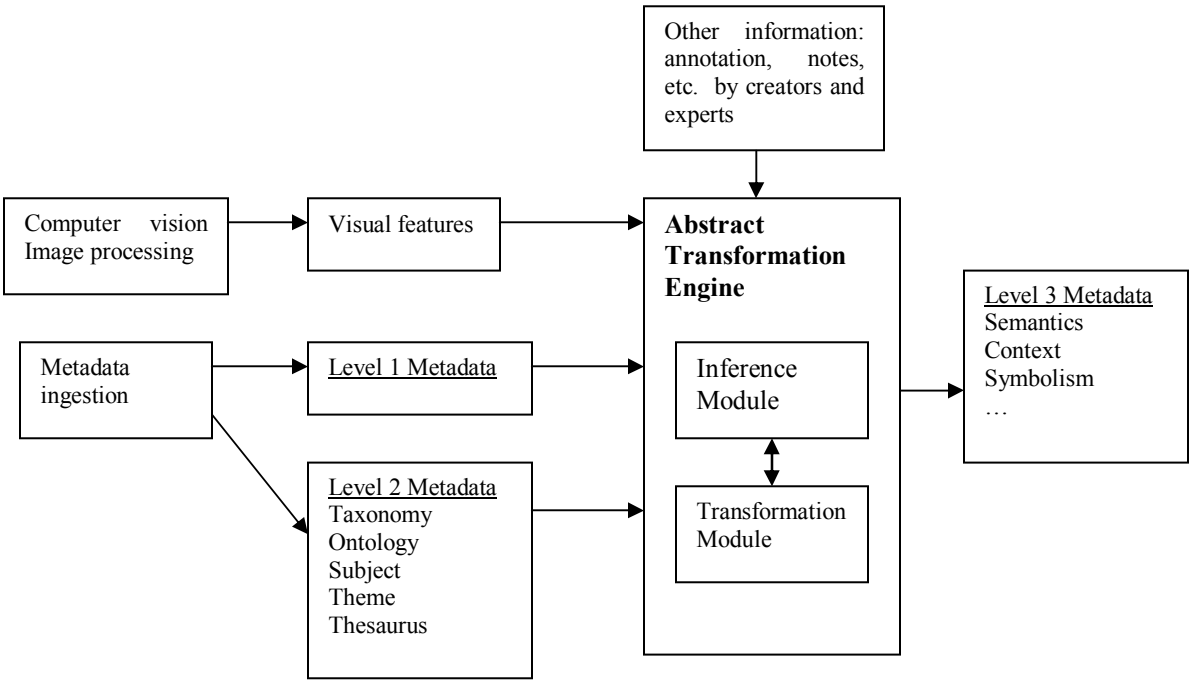


Figure 2. Schematic diagram of the system.

Narrative:
A rooster depicts strength and is viewed as a talisman which can exorcise ghosts and evil spirits. Chrysanthemum has brightly coloured flowers which are popular for display during Tet festival (Vietnamese New Year).

Click image to display any visual features

Visual Features:
rooster, child, chrysanthemum

Symbolism:
rooster >> strength
child >> wish for numerous offspring
chrysanthemum >> autumn, serene old age, permanence

User Tags:

Figure 3. A sample woodcut with symbolic meanings

Results (showing 1 to 3 of 3)

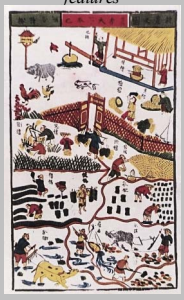
1. *"Farming"*
Style: Dong Ho
Media: Woodcut
Category: Rural Life

2. *"A Short Rest"*
Style: Dong Ho
Category: Rural Life

3. *"Rooster and the Chickens"*
Style: Dong Ho
Media: Woodcut
Category: Rural Life

Narrative:
This woodcut depicts typical traditional activities in the rice growing process.

Click image to display any visual features



Culture:
Farmers prepared soil either by hand or with the help of a water buffalo pulling a wide rake. Rice shoots were planted by hand in soil submerged in water.
Visual Features:

Symbolism:

User Tags:

Figure 4. Retrieval of woodcuts depicting "Rural Life"