An integrated multimedia approach to cultural heritage e-documents

Arnold W.M. Smeulders, University of Amsterdam, smeulders@science.uva.nl,
Lynda Hardman, CWI & TU/e, Lynda.Hardman@cwi.nl,
Guus Schreiber, University of Amsterdam, schreiber@swi.uva.nl, and
Jan-Mark Geusebroek, University of Amsterdam, mark@science.uva.nl
www.MultimediaNonline.com

Abstract: We discuss access to e-documents from three different perspectives beyond the plain keyword web-search of the entire document. The first one is situation-depending delivery of multimedia documents adapting the preferred form (picture, text, speech) to the available information capacity or need exemplified by documents from the annotated media database of the Rijksmuseum. It goes beyond Quality of Service methods which insist on delivering information in the same form even if that is no longer effective. Secondly, we discuss the use of ontologies to provide access across diverse library categorizations as part of the W3C semantic web. The system translates codes in the one catalogue system into a set of codes in another expanding the potential access to digital heritage knowledge across all library systems in the ontology, such as AAT, WordNet and IconClass. Thirdly, we discuss access to the pictorial contents of paintings by computer vision techniques, here showing in examples of Pieter de Hoogh and Johannes Vermeer which one of the two consistently painting photometric realistic in addition to adhering to the geometric realism as they both did. It is concluded access is the key issue in digital cultural heritage - be it access by situational delivery of e-document of cultural heritage, be it access to diverse knowledge systems, or be it access to the pictorial content of the picture.

1. Access and access is real access

Eventually, all disciplines of society will be touched by the e-thing. It is generally recognized that the cultural heritage we are preserving for future generations will profit considerably from passing over to the digital world. The digital e-world has several advantages over a paper-based world.

The potential access to cultural e-documents is broader as the number of people on-line quickly approaches a billion. So the beauty of cultural heritage documentation can be shared and compared among a much wider audience. In a second level of access does not have to be linear plus a key-word index - as the common structure of a library and the books in the library - but may be much more complicated than that. E-documents holding knowledge may hide a complex referential system behind the document in the form of hyper-links. To do that in a systematic way also thinking document maintenance and extensibility into account requires storage in multimedia document standards. To provide plain access is one thing, but the digital world also provides the opportunity of a third level of access by personalizing and localizing delivery while it is dynamically adapted to the needs of the person, the machine, the task and its context. We discuss an example of creating cultural heritage access adapting itself to the display and channel capacity in the amount of bytes being provided. Such an access is one way to avoid an information overload of the system and potentially of the consumer of cultural heritage documentation.

As well as access to a single, possibly on-line, possibly hyper-linked, possibly adaptive document, the digital world also brings archives much closer to one another. Where libraries grew in relative isolation with a denominative system formed in a local tradition, the Internet world breaks the locality of coding systems. From the interconnectedness, the need arises to move from the one categorization to the other. In the context of the semantic web [Schreiber], we discuss an example of access to more than one archive by forming an ontology on top of several well-known categorizing systems in pictorial cultural heritage description.

Having discussed flexible access to e-documents as well as e-libraries, we now move on to the third level of access, namely to the pictorial contents of paintings. Computer vision provides the means to access the semantics of pictures. The trouble is that humans are so good in assigning semantics to images they deem understanding a scene as a trivial task. Nevertheless, one-third of the brain is dedicated to viewing, so the task cannot be completely trivial. Standard computer vision tools can assist in interpreting digital paintings in a number of issues. They can score the palette of the paints used by the artist and they can score the histogram of simple shape elements like points, lines and patches. For physically realistic paintings, more advanced computer vision techniques can typify the types of edges [Gevers02] as well as

the truthfulness to natural patterns of shading [Geusebroek01]. We give examples of such quantitative interpretations of the scene in the painting. These techniques may be used to advance the literal pictorial search engines beyond the search in large archives of pictures for an exact copy of an image detail. To detect copies this might be a question of interest, when searching for information it hardly can be informative as a detail of a painting is usually provided when the name of the remainder of the painting is known. A different matter in the development of a quantitative theory of art history apart from the digital image statistics on shape and color is the classification of textures by the materials they represent. This is ongoing research [Geusebroek03] potentially useful in the access to large collections of paintings to provide art historians advanced access to cultural heritage.

2. User-tailored access to digital repositories

Users seeking access to cultural repositories want immediate access to information relevant to their task, suited for display on their device, able to be transmitted over the available bandwidth and using media suited to their current situation.

Figure 1: Based on the annotated media database of the Rijksmuseum in Amsterdam, the Cuypers engine automatically generates multimedia presentations in SMIL in response to user queries. The figure illustrates a presentation about the painting technique "chiaroscuro" in the context of the work of the painter Rembrandt van Rijn. The presentation consists of a slide show with example paintings by Rembrandt, using the chiaroscuro technique, alongside a textual explanation of the technique itself. While consisting of a number of individual media items returned by the database, the presentation is intended to convey the semantic relations among them. For example, domain independent layout knowledge is used to place the presentation title centrally, because it applies to both the slideshow and the textual explanation. In contrast, the title of the textual explanation is left-aligned with the body text to indicate that it refers only to that paragraph. The close proximity of the text "Self Portrait (1661)" to the image, is intended to convey that the text is a label referring to the painting. In addition, explicit knowledge of domain-specific presentation conventions allows the engine to display it in a way that the user can interpret it as being the painting's title and the year of creation.

Consider the different requirements for an art history student walking through a museum, a quiet environment providing access to the original artifacts themselves, compared with an interested lay person driving home in the evening curious about the emergence of a particular art genre. In the first case the user needs no explanation of the artifact itself – although a description of how to reach it would be useful – but rather contextual information on the creator and the period in which it was created. In the second case, a spoken commentary, presenting a high-level synopsis of the different art schools in vogue just before the emergence of the genre in question, would be more appropriate. To construct a system able to create such a variety of presentations requires expertise in many different aspects. We discuss the annotation, analysis and retrieval problems below.

Here, we concentrate on how some selection of relevant items of different media can be appropriately displayed to the user. The semantic relations among the selected media items need to be communicated explicitly by choosing appropriate layout and style. Creating well-designed multimedia presentations requires an understanding of both the presentation's global discourse and interaction structure as well as details of multimedia graphic design. In addition, knowledge of a number of other factors is needed. A domain description allows the relationships among domain concepts to influence the layout and links within the presentation. Knowledge about the user's task or environment allow appropriate choices of media to be made. If information is also known about the user, then presentations can be tailored to, e.g., skip things the user already knows and explain new concepts in terms of already known concepts. A description of the characteristics of the end-user platform (such as screen resolution, bandwidth, ability to display colour, audio capabilities) allows optimal use to be made of the capabilities of the device.

At CWI, the Cuypers presentation generation engine is being developed [Van Ossenbruggen01]. This allows the specification of different information types and incorporates them within the overall process of generating a presentation. The current system is able to generate presentations tailored to device characteristics, in particular for different screen sizes and aspect ratios and bandwidth availability. In addition, rudimentary user tailoring can be carried out.

An important aspect of future work is the realization that the assumption underlying current style sheet technology, e.g. XSLT and CSS, stating that content and presentation are independent, is often an oversimplification for multimedia presentations [Van Ossenbruggen02].

3. Ontology-based access to heterogeneous documentation

There is an increasing interest in using domain knowledge corpora (ontologies[1]) to aid multimedia annotation and search. This work is in line with recent efforts to arrive at a semantic web in which distributed information can be found and processed with the help of semantic annotations. We show one typical application scenario for using semantic annotations in the cultural-heritage domain.

Fig. 2 shows part of the interface of a tool that can be used for semantic annotation and search of art images [Schreiber00]. The tool contains a set of ontologies, including the Art and Architecture Thesaurus, AAT, [Peterson94], WordNet and IconClass. The knowledge corpora are all represented in RDF Schema, a recent W3C standard for semantic annotation. See www.w3.org for more information. Each ontology is represented as a subclass hierarchy of RDFS classes. For annotation purposes the tool provides the user with an annotation template derived from the VRA 3.0 Core Categories [VRA00]. The VRA template provides a specialization of the Dublin Core set of meta-data elements, tailored to the needs of art images. In the tool each slot in the VRA annotation template is, whenever possible, bound to one or more subtrees of the ontologies. For example, the VRA slot *style period* is bound to two trees in AAT containing the appropriate style and period concepts.

In the lower window in Fig. 2 we see a fragment of these subtrees. The user is selecting the concept *baroque*. The ontology makes sure that the user is aware she could have used a more specific term (e.g. *high baroque*) if she wanted. The upper window of Fig. 2 shows seven of the VRA data elements. The others can be found at the three tabs that are not shown. The slots with a *magnifier* make use of the ontology The domain knowledge could have been extended to cover more slots. For example, the "creator" slot could takes values from the Getty thesaurus ULAN (Universal List of Artist Names).

[1] In the context of this paper we simply use the term ontology to refer to shared knowledge, which is not specific for one particular application. See the ontology literature for a more detailed discussion of the notions concerned.

For annotation purposes the ontologies serve two purposes. Firstly, the user is immediately provided with the right context for finding an adequate index term. This ensures quicker and more precise indexing. Also, the hierarchical presentation helps to disambiguate terms. For example, if the user would type the term *bed* to describe the scene, the tool will provide the user with a choice of concepts in AAT and WordNet that are denoted with this term (piece of furniture, land depression, etc.). From the placement of the terms in the respective hierarchies, it is usually immediately clear to the indexer which meaning of the term she wants.

Semantic annotations are especially useful for search purposes. For example, suppose the image is annotated with the concept *Venus*, sub-concept of *Roman deity*. Now, searching for *Aphrodite* enables the tool to find this picture. Semantic annotation allows us to make use of concept search instead of pure syntactic search. It paves also the way for much more advanced search strategies. For example, users may be specializing or generalizing a query with the help of the concept hierarchy when too many or too few hits are found.

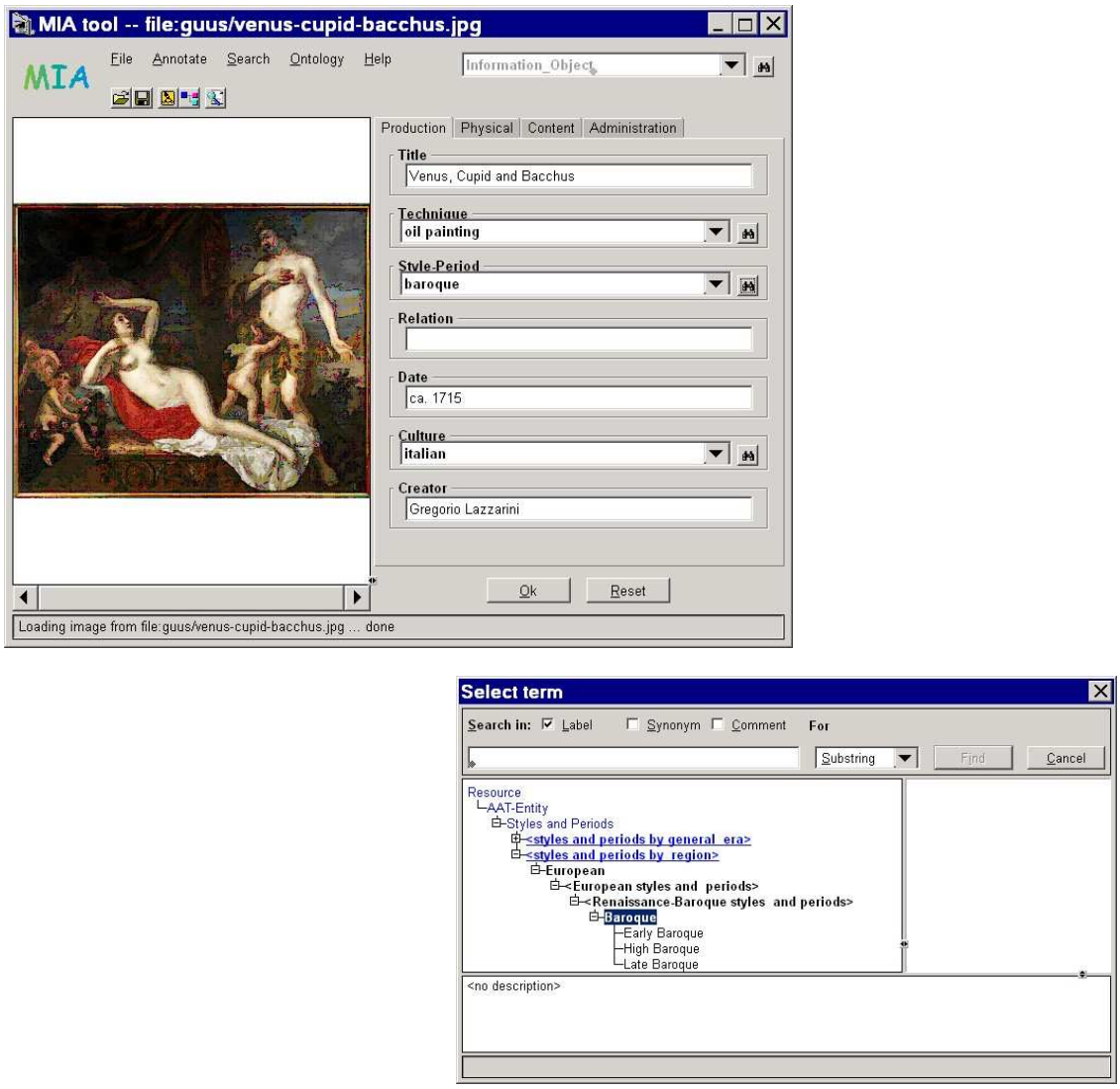

Fig. 2: Snapshot of a semantic annotation and search tool for art images. The figure shows a fragment of the annotation window (upper part) with one tab of VRA-data elements for describing the image, here the production-related descriptors). The slots associated with a magnifier button are linked to one or more subparts of the underlying ontologies, which provide the concepts for this part of the annotation. The lower window shows part of the AAT-hierarchy containing terms relevant for the *style/period* slot.

4. Access to the image content by invariant computer vision

Of all information forms, the pictorial information form is farthest away from a complete semantic interpretation.

Key obstacles are what is known as the sensory data gap and the semantic gap [Smeulders00]. The *sensory data gap* gives a name to the fact that there are a million technically, bit-wise different data arrays which would be immediately associated by the human observer or listener with the same object. Differences in lightning, scene or shadow make no difference in the interpretation for humans. This is radically different from coded and numerical information where one bit-representation stands for one interpretation. The sensory gap does not play a role in the analysis of pictures when studied at the facsimile, the literal level. As pictures of arts are always recorded in frontal view with white-light illumination, there is a standard representation. When studying the content of the scene, the sensory gap is present in full glory as each object is painted under different illumination and different pose. So if one wishes to address the objects in a (physically realistic) painting, one has to use an invariant representation taking away the accidental illumination conditions. Examination if a painter in detail draws the shadows and highlights may be assessed by evaluation of the painted colors to conform to the physical laws of light reflection. Edge classification based on the physical laws of light reflection [Gevers99] highlights realistic shadow edges, see Fig. 3, and ignores or wrongly classifies shadow edges which are not realistic in physical sense.

In addition to the sensory gap, there is the *semantic gap*. The semantic gap is the difference between the immediate interpretation of pictorial information in all its different forms and the interpretation that follows from a formal description of the object. As formal descriptions in the end are the only commodity the computer can handle, the semantic gap is to be bridged when the desire is to address the pictorial contents of an image by a set of lingual codes such as ICONCLASS. The semantic gap is to be approached by a combination of top-down term translation and bottom-up feature description. The top-down part represents the index terms describing the contents of an image in ontologies, see paragraph 4. In an interpretation the image is interpreted by complete feature sets characterizing the textures in the picture. We are currently engaged in a research program to connect the two approaches: top-down term translation into pictorial features and image analysis yielding invariant texture descriptions [Geusebroek03] that hopefully connect to one another.

5. Concluding remarks

In the paper we have illustrated several aspects of accessibility in a digital world. We believe cultural heritage is an excellent platform to demonstrate concepts since very soon the need will be felt in society at large.

The next step in our approach is to bring pair-wise integration among the three approaches to accessibility of cultural heritage e-documents:

The integration of personalized and location-based delivery integrated with access to heterogeneous libraries. This requires techniques for localized delivery without the perfect background information that is assumed in the current systems. It requires ontology-based access to diverse sources to be extended to delivery sensitive information. For example, a user may compare one painting on a mobile display while standing in front of another. To be more specific on this issue, current style sheets operate on the syntactic XML-level and are unaware of the new generation Semantic Web languages, such as RDF(S) and OWL [Van Ossenbruggen02]. The emphasis of the current work at CWI is on creating a more realistic user model, developing a graphic design model and investigating the requirements for a discourse model.

The integration of semantic understanding of pictures with personalized delivery implies the on-line interrogation of the picture with questions that come to mind. Such an open-ended toolbox requires a general semantic understanding and visualization of computer vision tools. We aim to achieve such a coupling but deem it more feasible to concentrate on the remaining integration of semantic understanding and heterogeneous library access first.

The integrated approach to semantic understanding of pictures with heterogeneous library access is happening by extending the formal systems, as described in this paper, to formalize the contents of pictures with computer vision features, as described in this paper. We are currently working on this combination. For example, a user may immediately access the pictorial content from the digitized painting in addition to the denominative efforts of art historians' cataloguing.

6. References

[Geusebroek01] J. M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders and H. Geerts. Color invariants. IEEE trans PAMI, pages 1338 -- 1350, 2001.

[Geusebroek03] J. M. Geusebroek and A. W. M. Smeulders: A theory for stochastic texture understanding, in preparation.

[Gevers99] T. Gevers and A. W. M. Smeulders: Color based object recognition. Pattern Recognition, 32 - 3:453 -- 464, 1999.

[Gevers02] T. Gevers and A. W. M. Smeulders: Color constant ratio gradients for image segmentation and similarity of textured objects. In Proceedings Computer Vision and Pattern Recognition, 18 - 21. IEEE Press, 2001.

[Van Ossenbruggen01] J. van Ossenbruggen, J. Geurts, F. Cornelissen, L. Rutledge, and L. Hardman: Towards Second and Third Generation Web-Based Multimedia. In: The Tenth International World Wide Web Conference, Hong Kong pp. 479-488, ACM Press, May 1-5, 2001. Available at http://www10.org/cdrom/papers/423/index.html, http://www.cwi.nl/~media/publications/www10.pdf, http://www.cwi.nl/~media/publications/www10/index.html

[Van Ossenbruggen02] J. van Ossenbruggen and L. Hardman: Smart Style on the Semantic Web. In: Semantic Web Workshop, WWW2002 May 2002. Available at http://www.cwi.nl/~media/publications/www2002-semwebworkshop.pdf, http://semanticweb2002.aifb.uni-karlsruhe.de/proceedings/Research/ossenbruggen.pdf

[Peterson94] Peterson, T. *Introduction to the Art and Architecture Thesaurus.* Oxford University Press 1994. See also: http://shiva.pub.getty.edu.

[Schreiber01] Schreiber, A.T., Dubbeldam, B., Wielemaker, J., Wielinga, B.J. Ontology-based photo annotation. *IEEE Intelligent Systems,* **16**:66-74, 2001.

[Smeulders00] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain. Content-based image retrieval: the end of the early years. IEEE trans. PAMI, 22 - 12:1349 -- 1380, 2000.

[VRA00] VRA core categories, version 3.0. Technical report, Visual Resources Association, 2000. URL: http://www.gsd.harvard.edu/~staffaw3/vra/vracore3.htm.

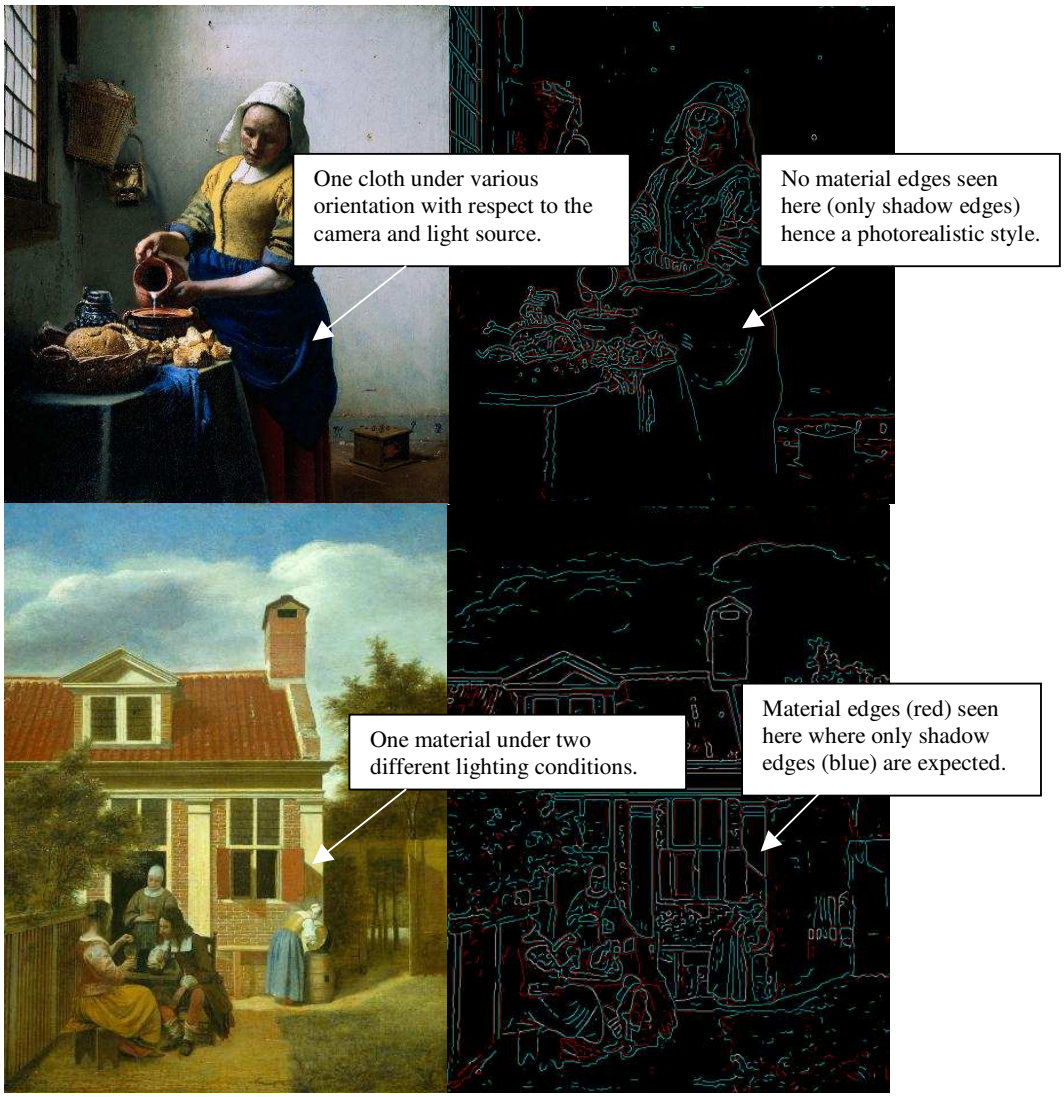

Fig. 3. Edge type classification for a) "De keukenmeid" painted by Johannes Vermeer, and b) "Een gezelschap op de plaats achter het huis" painted by Pieter de Hooch. The photometric edge images as derived from invariant properties are shown on the right, blue indicating color changes, accentuation by red indicating shadow edges. Vermeer was very precise in the photo-realistic painting of the subject, as can be derived from the edge type classification image. The photometric shadow edges precisely follow the painted shadows. De Hooch was less accurate, as many painted shadows do not result in a photometric shadow edge. Hence, these shadows are not photo-realistic. Note that this does not necessarily mean that the perceived shadow effect is inaccurate. Digital pictures courtesy RijksMuseum.