Mathematics Publications

Mathematics

1-1-2005

Semantic Notation and Retrieval in Art and Architecture Image Collections

Boyan N. Dimitrov
Kettering University, bdimitro@kettering.edu

Follow this and additional works at: https://digitalcommons.kettering.edu/
mathematics_facultypubs

Part of the Mathematics Commons

Semantic Notation and Retrieval in Art and Architecture Image Collections

Peter L. Stanchev David Green Jr.,
Boyan Dimitrov
Kettering University, Flint, MI 48504. USA
{pstanche,dgreen,bdimitro}@kettering.edu

ABSTRACT: *In this paper, we analyze various methods used for semantic annotation and search in a collection of art and architecture images. We discuss the Art and Architecture Thesaurus, WordNet, ULAN and Iconclass ontology. Systems for searching and retrieval art and architecture image collections are presented. We explore if the MPEG 7 descriptors are useful for art and architecture image annotations. For illustrations we use images from Antoni Gaudi architecture and Claude Monet paintings.*

Categories and Subject Descriptors
H.2.8[Database Applications]: Image Databases ; H.3.1 [Content Analysis and Indexing]; H.5.1 [Multimedia Information System]; I.4 [Image Processing and Computer Vision]

General Terms
Semantic Notation, Architecture Images, MPEG 7 descriptors, ontology

1. Introduction

Recent advances in computing, communications, and storage technology have made multimedia data prevalent. Many museums offer their collections on the web. The rich content of multimedia data built through the synergies of the information contained in different modalities calls for innovative methods for modeling, processing, mining, organizing, and indexing these data. **Content-based image retrieval** and **Content-based video retrieval** are two research areas in multimedia systems that have been particularly popular in the last years. The MPEG-7 standard gives a set of descriptions that have been used to facilitate this research.

Over the past decades, many researchers from the image processing, computer vision and database communities have investigated possible ways of retrieving visual information based solely on its content. Instead of being manually annotated using keywords, images and video clips could be indexed by their own visual content, such as color, texture, and objects' shape and movement. Many research groups in leading universities, research institutes, and companies are actively working in this field. Their ultimate goal is to enable users to retrieve the desired image or video clip from massive amounts of visual data in a fast, efficient, semantically meaningful, friendly, and location-independent environment.

There is evidence that different image features work with different levels of effectiveness depending on the characteristics of the specific image data set. For example, color layout, like color structure, perform badly on monochrome images, while dominant color descriptor performs equally well on several of data sets.

Existing systems are limited by the fact that they can only operate at the primitive feature level, while users operate at a higher semantic level. This mismatch is often referred to as *semantic gap*. It is possible to increase the retrieval effectiveness by a proper choice of image features from the MPEG-7 standard [16].

There are many works devoted to the art and architecture image semantics. Koivunen and Swick [8] presented it from the prospectives of shared collaborations. Handschuh and Staab [5] provide manual and semi-automatic techniques for semantic annotation. Hyvonen et al [6] proposed ontology-based image retrieval. The word "ontology" originates from the Greek words ontos = "being" and logos = "knowledge", and means "knowledge of being". It was first used in the 17th century, from Christian Wolff, for the branch of metaphysics of existing. The most acceptable definition of ontology according to [10] is the Gruber [4] definition:

"*formal* specification of a conceptualization", and is shared within a specific domain. When we ask of existing works of art or architecture we are asking about the ontology of artworks. Ontologies can be used for annotation and search in image collections. In this paper various ontologies used for art and architecture image collections such as AAT, WordNet, Iconclass and ULAN are reviewed in Section 2. Ways of using metadata in these kinds of collections is presented in Section 3.

Art collections, based on ontologoly are discussed in Section 4. Section 5 provides analysis on art and architecture images using MPEG 7 descriptors. We are attempting to analyze the significance of color, shape and texture descriptions for art and architecture image collections. In Section 6 we have given the conclusion of this work.

2 Ontologies used for art and architecture image collection

Some of the most popular ontologies for art and architecture images are: AAT, WordNet, Iconclass and ULAN.

2.1. The Art and Architecture Thesaurus (AAT)

The AAT [14] is a structured vocabulary currently containing nearly 128,000 terms and other information about concepts. The terms are organized in a single hierarchy. Terms in AAT may be used to describe art, architecture, decorative arts, material culture, and archival materials. Terms for any concept may include the plural form of the term singular form, natural order, inverted order, spelling variants, various forms of speech, and synonyms that have a variety of etymological roots. Among these terms, one is flagged as the preferred term, or descriptor. The focus of each AAT record is a concept. Currently there are about 34,000 concepts in the AAT. In the database, each concept's record is identified by a unique numeric ID. Linked to each concept record are terms, related concepts, a parent (that is, a position in the hierarchy), sources for the data, and notes. The temporal coverage of the AAT ranges from Antiquity to the present and the scope is global. Here is an example of a note for "architecture" in the AAT in XML form.

<?xml version="1.0" encoding="US-ASCII" ?>

<Vocabulary xmlns="http://localhost/namespace" xmlns:vp="http://localhost/namespace" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://localhost/namespace Getty_Vocab_Export_AAT.xsd"
Title="Art & Architecture Thesaurus" Date="2004-08-05">
…..

<Note_Text>The Styles and Periods hierarchy contains the names of art and architecture styles, historical periods, and art movements. Names of peoples, cultures, individuals, and sites are included only if they designate distinct styles or periods (e.g., Yoruba, Louis XIV). Geographic descriptors are included only for broad cultural regions and nations. Relation to Other Hierarchies: Descriptors for genres of art, including all the arts not specific to a given people or period (e.g., amateur art, pattern poetry) are found in the Associated Concepts hierarchy, as are descriptors for general approaches to art (e.g., realism) while specific movements named after such approaches are found here (e.g., Realist). </Note_Text>

2.2. WordNet

WordNet® [12] is an online lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. WordNet® was developed by the Cognitive Science Laboratory at

Princeton University under the direction of Professor George A. Miller. Concepts of the words could be used to describe the content of the image. For example for architecture in the system we found:

1. architecture - (an architectural product or work)

2. architecture - (the discipline dealing with the principles of design and construction and ornamentation of fine buildings; "architecture and eloquence are mixed arts whose end is sometimes beauty and sometimes use")

3. architecture - (the profession of designing buildings and environments with consideration for their esthetic effect)

4. computer architecture, architecture — ((computer science) the structure and organization of a computer's hardware or system software; "the architecture of a computer's system software")

2.3 Iconclass

The Iconclass system [1, 19] was developed by Henri van de Waal (1910- 1972), Professor of Art History at the University of Leiden. His ideas for a systematic overview of subjects, themes and motives in Western art, which later became the Iconclass System, took preliminary form in the early 1950's. It is a subject-specific classification system. It is a hierarchically ordered collection of definitions of objects, persons, events and abstract ideas that can be the subject of an image. It can be used to describe, classify and examine images represented in various media such as paintings, manuscripts, posters, photographs and even newspaper clippings. It has 28,000 hierarchically ordered definitions divided into ten main subject divisions. 14,000 keywords are used for locating the notation and its textual correlate needed to describe and/or index an image. There also 40,000 references to books and articles of iconographical and cultural historical interest. For example the story for Claude Monet starts with:

Claude Monet

[French Impressionist Painter, 1840-1926]

• Also known as: Claude Oscar Monet

• Relationships: Studied under Charles Gleyre.

Monet's students included American Impressionists Theodore Butler, Lilla Cabot Perry and Theodore Robinson.

2.4. The Union list of Artist Names

The ULAN [18] - Union List of Artist Names® is a structured vocabulary that can be used to improve access to information about art, architecture, and material culture. It contains information about nearly 220,000 artists. The information includes name variants and some limited biographical information (dates, locations, artist type). A subset of 30,000 artists, representing painters, is incorporated in the tool. The ULAN is a compiled resource; it is not comprehensive. The ULAN grows through contributions. Information in the ULAN was compiled by the Getty Vocabulary Program in collaboration with many institutions. Implementers should keep in mind that the vocabularies grow and change over time. For example, in ULAN we can find the following text for Antoni Gaudi.

Gaudí, Antoni (Spanish architect, landscape architect, and furniture designer, 1852-1926)

Note: Gaudí was influenced by Catalonia's medieval history and architecture. His works display a respect for craftsmanship and structural logic. He was also inspired by forms in nature, using it in structure and ornament, creating a highly personal, organic style. His work is characterized by sculptural plasticity, the manipulation of light, and the use of mosaics and polychromy. His later style is classified as Catalan Modernisme, a style related to Art Nouveau.

3 Art and architecture image descriptions with the help of metadata

The most popular description for the metadata is the semantic web. Berners-lee [2] proposed the architecture of the Semantic Web in the following three layers:

- The *metadata layer*. The data model at this layer contains just the concepts of resource and properties.
- The *schema layer*. Web ontology languages are introduced at this layer to define a hierarchical description of concepts (is-a hierarchy) and properties.
- The *logical layer*. More powerful web ontology languages are introduced at this layer.

The RDF (Resource Description Framework) [9] is the most used data model for the metadata layer. The metadata RDF schema is separated into three different schemas:

Dublin Core schema. The Dublin Core scheme is a general scheme for identifying original works, typically books and articles, but also films, paintings or photos. It contains such properties as creator, editor, title, date of publishing and publisher. It is being developed by the Dublin Core Metadata Initiative. For photo image the following information could be collected: title, subject, description, creator, publisher, contributor, date, type, format, identifier, source, language, relation, location, rights.

Technical schema. This scheme captures the technical data. For photo images the following data could be collected: the type of camera, the type of film, the date the film was developed and the scanner and software used for digitizing.

Content schema. It contains the keywords we use in the "subject" property of the Dublin Core scheme. That property should contain as many of the following keywords as are applicable. For the photo image storage system the keywords could have the following meaning: Portrait, Group-portrait, Landscape, Baby, Architecture, Wedding, Macro, Graphic, Panorama, Animal.

An example for indexing art multimedia data at multiple levels is given in [7]. They proposed a pyramid with 10 levels. The *first level* is the image type: photograph, digital image, drawing, painting, animation, medium, style. The *second level* is the global description: color type, global color, global color quality, global shape, and global texture. The *third level* is the local structure: local color, local color quality, local color placement, local shape, local shape placement, local texture, local texture placement, object placement, shape, texture, size, number, color, living placement, living size, and living number. The *fourth level* is the global composition point of view. The *fifth level* describes the generic objects: generic objects category, generic objects living, object type, object living gender, object living age. The *sixth level* is describing the specific objects, specific object symbolism, emotions states, relationships, status. The *seven level* is about abstract objects description such as symbolism, emotions, relationship. The *eight level* describes the generic scene characteristics such as: when, where, genre, category, general event, activity, pose/action. The *ninth level* is connected with specific characteristics such as: when, specific where, specific event. The last *tenth level* is for abstract scene subject, symbolism, emotions/ mental states, relationships atmosphere.

4 Ontology based systems for art and architecture collections

The distribution of collection of objects in museums at different locations creates an obstacle to information retrieval, for both the public and for researchers. It should be possible to use the collections as if they were in a single database. Some systems which can be used for this purpose are:

4.1. The Helsinki museum system

Using semantic web technologies at the Helsinki University Computer Science Department and the Helsinki Institute for Information Technology (HIIT) an art retrieval system is developed [6]. The systems are used in the Espoo City Museum and Musketti (the National Museum). The system shows multiple ontologies used in annotating collection data, such as *ObjectType*, *Material*, etc. By selecting ontological classes from these hierarchies, the user can express the search profile easily in the right terminology. For example, by selecting *ObjectType* = "carpet" and *Material* = "silk", silk carpets are found.

4.2. The IBM Marvel system

The IBM Marvel system [13] is designed to automatically categorize (and subsequently retrieve) clips using modifiers like "outdoor,"

"indoor," "cityscape" or "engine noise" that describe the action in the clip. The Marvel research team, which is working on the project with libraries and a few select news organizations, such as CNN, showed off the first prototype at a conference at Cambridge University. The prototype system can scan through a database of more than 200 hours of broadcast news videos and uses 100 different descriptive terms to classify and identify scenes. A query takes about two to three seconds. Marvel is based on the MPEG-7 data format, but it can search on any standard video format.

4.3. Photo system

In the photo system [15] two groups of definitions are used:

- Structure of a photo annotation. This ontology provides the description template for annotation construction. How, when, and why was the photo made? A specify metadata about the circumstances related to the photo such as the photographer or the vantage point (for example, a close-up or an aerial view). Other elements are photo characteristics of the medium feature. This represents metadata such as the storage format (e.g. jpeg) or photo resolution.

- Subject matter vocabulary. For this, they used the notion of structured annotation from Audrey Tam and Clement Leung [17]. The description template consists of four elements:

1. An agent, for example, "an ape." An agent can have modifiers such as "color = orange."

2. An action, for example, "eating."

3. An object, for example, "a banana." Objects can also have modifiers (color = "green").

4. A setting, for example, "in a forest at dawn."

5 Analysis on the art and architecture images using *MPEG 7* descriptors

Are the MPEG 7 descriptors good for art and architecture image? To answer this question we made lots of experiments. In this paper we present some of them.

Let's start with Claude Monet paintings. We use photographs from[11]. We have a photograph of him in his garden from 1924 . We

extract the flowers from the right down corner (image 1) and compare with the flower part from "Corner of the Garden with Dahlias", 1873

(image 2). We use also an otter photography of Monet's garden,

(image 3) and measure the similarity with the paintings "Water pond symphony in green", 1899.

(image 4). Texture similarity we try to find in picture of his garden

(image 5) and part of the painting "Main path through the Garden at Giverny", 1902

(image 6). Other experiment was done with Antoni Gaudi architecture [3]. The measure of the distance between this image extract from a flower

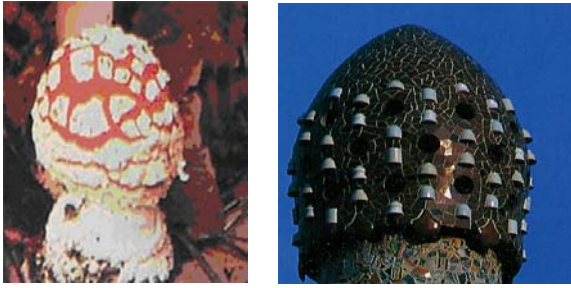(image 7) and this segment in Gaudi architecture

(image 8). The both images were extracted automatically from the images

(image 9) and

(image 10). We measure also similarity between mushroom image

(image 11) and part of Gaudi work (image 12).

Images	Scalable color	Dominant color	Color structure	Color layout	Combination color descriptors	Edge histogram	Homogeneous texture	Region shape
1/2	0.172	0.023	0.472	0.042	0.278	0.274	0.238	-
3/4	0.502	0.084	1.143	0.090	0.613	0.289	0.277	-
5/6	0.690	0.041	1.923	0.063	0.958	0.283	0.168	-
7/8	-	-	-	-	-	-	-	0.285
9/10	0.555	0.040	0.563	0.100	0.390	0.336	1.196	-
11/12	0.510	0.157	0.708	0.218	0.475	0.320	1.238	-

We found that the color descriptors are giving the best match. Dominant color and color layout give the best results. Combination of color descriptors gives worse result. Shape descriptors are difficult to use, because the automatic object segmentation do not give in many cases satisfactory results. Edge histograms are giving better results that the homogeneous texture. The reason was the different background.

6. Conclusions

The painter translates the reality onto a picture using a vision. The viewer of the picture translates it back to reality in his/her own vision. The current technology is still not usable for majority of people who need to describe their view of the search object. If some objects are "very near" in the human eyes they are not always so near according to the MPEG 7 descriptors. It is not possible to tell which descriptors will best match the human ranking without analyzing the image. Making our experiments we found that it is very important to study the image data set in regards to the descriptors and their combinations before using them in the indexing and retrieval process.

7. Acknowledgement

References

[1] Berg J. van den (1995). Subject retrieval in pictorial information systems. In: Electronic Filing, Registration, and Communication of Visual Historical Data. Abstracts for Round Table no 34 of the 18th International Congress of Historical Sciences. Copenhagen, 21–28, also http://www.iconclass.nl

[2] Berners-lee, T. (1998). Semantic web road map, W3C Design Issues. Cambridge, MA, also: http://www.w3.org/DesignIssues/Semantic.html

[3] Cirlot J. Visas P, Pla R (2001). Gaudi. An introduction to his architecture, Triangle postals.

[4] Gruber, T (1993). A translation approach to portable ontology specifications, *Knowledge Acquisition* 6(2) 199-221.

[5] Handschuh, S. Staab, S (2003). Annotation of the shallow and the deep web, *In*: S. Handschuh, S. Staab, editors, Annotation for the Semantic Web, volume 96 of Frontiers in Artificial Intelligence and Applications, IOS Press, Amsterdam, 25–45.

[6] Hyvonen, E. Kettula, S. Raatikka, V. Saarela, S. Viljanen, K (2003). Finnish museums on the semantic web. *In*: Proceedings of WWW 2003, Budapest.

[7] Jaimes A. Jorgensen, C. Benitez, A. B. Chang, S.-F. (1999). Experiments for Multiple Level Classification of Visual Descriptors, ISO/IEC JTC1/SC29/WG11 MPEG99/M5593, Maui, Hawaii, USA, Dec.

[8] Koivunen M-R. Swick. R. R (2003). Collaboration through annotation on the semantic web, *In*: S. Handschuh and S. Staab, editors, Annotation for the Semantic Web, volume 96 of Frontiers in Artificial Intelligence and Applications, IOS Press, Amsterdam, 46–60.

[9] Lassila, O. Swick, R (1999). Resource Description Framework (RDF) Model and Syntax Specification, W3C Recommendation 22 February Cambridge, MA: W3C, also: http://www.w3.org/TR/REC-rdf-syntax.

[10] Lu, S. Dong, M. Fotouhi, (2002). The Semantic Web: opportunities and challenges for next-generation Web applications, *Information Research* 7(4), also: http://InformationR.net/ir/7-4/paper134.html

[11] Mancoff D (2001). Monet's Garden in art, Viking Studio.

[12] Miller G (1995). WordNet: A lexical database for English, *Communications of the ACM* 38(11) November. also http://wordnet.princeton.edu/

[13] Natsev A. Naphade M. Smith, J. R. (2004). Over-complete Representation and Fusion for Semantic Concept Detection, Special session on Content Understanding for Home Image and Videos, Proc. IEEE Intl. Conf. on Image Processing (ICIP) Oct.

[14] Peterson T (1994). Introduction to the Art and Architecture Thesaurus, Oxford University Press, also http://www.getty.edu/research/tools/vocabulary/aat/

[15] Schreiber A. Th. Dubbeldam, B. Wielemaker, J. Wielinga, B. J. (2001). Ontology-based photo annotation, *IEEE Intelligent Systems*, 16(3) 66–74, May/June.

[16] Stanchev, P. Amato, G . Falchi, F. Gennaro, C. Rabitti, F. Savino P (2004). Selection of MPEG-7 Image Features for Improving Image Similarity Search on Specific Data Sets, 7-th IASTED International Conference on Computer Graphics and Imaging, CGIM 2004, Kauai, Hawaii, 395-400.

[17] Tam, A.M. Leung, C.H.C (2005). Structured Natural-Language Description for Semantic Content Retrieval, J. American Soc. Information Science, to be published Sept.

[18] The Getty Foundation (2000). ULAN: Union List of Artist Names also http://www.getty.edu/research/tools/vocabulary/aat/

[19] Waal H, van der (1985). CONCLASS: An iconographic classification system, Technical report, Royal Dutch Academy of Sciences (KNAW). also: http://www.iconclass.nl