

Use of the MPEG-7 standard as metadata framework for a location scouting system – An evaluation study

Valérie Gouaillier and Langis Gagnon
Computer Research Institute of Montreal
Tel: +514-840-1235; Fax: +514-840-1244
valerie.gouaillier@crim.ca
langis.gagnon@crim.ca

Sylvain Paquette and Philippe Poullaouec-Gonidec
Chaire en Paysage et Environnement, Montreal Univ.
Tel: +514-343-7500; Fax: +514-343-6771
sylvain.paquette@umontreal.ca
philippe.poullaouec-gonidec@umontreal.ca

Abstract:

The paper presents an evaluation study about the application of the MPEG-7 standard to the metadata associated with a database supporting identification, characterization, classification and referencing of film shooting locations. The metadata describes photos taken at each location, as well as site characteristics, mainly according to a standard taxonomy and vocabulary in the domain of landscape and environment. Not only should the metadata schema allow conventional queries, but it must also be suited for advanced search functionalities such as similarity-based image retrieval. We also discuss the advantages and shortcomings of MPEG-7 for this application.

Keywords:

MPEG-7, location scouting, visual content description, metadata, landscape

1. Introduction

In western countries, major issues are related to landscape concerns within the land planning process as landscapes offer great social, cultural and economical potential [1-2]. Better knowledge of territorial qualitative attributes, supported by suitable tools, is instrumental for efficient landscape preservation, enhancement and monitoring. The aim of the paper is to present an evaluation study about the application of MPEG-7 to the management of the metadata associated with a location scouting application. This work is a

phase of a larger research project called SIMIPE-Ciné. The goal of SIMIPE-Ciné is to develop a Web-enabled software prototype for search through geographical and visual content criteria. Various types of data about the site will be integrated to the database: annotations, photos, panoramic views, audio clips, maps and geomatic information.

The targeted application is location scouting for film and television production. The system will assist location managers in retrieving more efficiently the sites fulfilling visual and logistical requirements of a particular film shooting. The system architecture needs to be open and flexible enough to be adapted to other domains such as tourism and land planning/development. One of the major challenges of this project is the design of a suitable data and metadata structure that supports both syntactic and semantic queries, among which some are based on mathematical features for low-level visual content description of images. Tools we have developed will be used for content-based retrieval [3-4]. Moreover, the metadata schema should include mechanisms to link all the material pertaining to a site to its description. We have chosen to explore the usability of MPEG-7 schema for this application because it offers a comprehensive set of normative tools for the description of both structure and semantics of audio-visual content. The goal is to evaluate if the standard can adjust to the landscape description taxonomy established by experts in the field. Although this work does not use the Dublin Core (DC) standard, we think

the experiences and conclusions from this work are of interest as it could provide some insights to DC experts regarding the limitations/advantages of MPEG-7 for metadata description of text-visual documents. The paper is organized as follows. Section 2 presents the type and structure of metadata needed for the management of our database. A brief overview of the MPEG-7 standard is then presented in Section 3, followed by the implementation we have made of its schema. The paper concludes with a discussion on the suitability of MPEG-7 for this application.

2. System description

2.1 Database

Our database is composed of about 3000 visual and audio elements (photos, panoramic views and audio clips of ambient sound) of over 500 sites of the Monteregian region in Quebec (Canada) which present potential interest for the cinema and tourist industry. A data sheet listing the territorial, geophysical and infrastructural characteristics of each location has been filled. All photos taken on a site was georeferenced. Structural aspects of the picture, as well as aesthetic quality assessment of the landscape scene depicted, were manually recorded from an expert-based evaluation, completed from standard categories together with indicators specifically designed for this study.

2.2 Queries

A user will be able to retrieve a site corresponding to various criteria: description of the site, related infrastructures, visual features of corresponding stored photographs, climatic and environmental context, type of ambient noise and geomatic properties. It will also be possible to provide a photo of scenery as a query to the system and to retrieve pictures of the database presenting the most visual similarities. The comparison will be based on low-level features, as color and texture, which will be automatically extracted with image processing algorithms. Geomatic queries will be taken in charge by a Web engine developed with Push'n'See [5] and linked to the principal interface. Combination of queries will also be possible. For example, a user could search for sites in the database visually similar to a given picture of a New England garden, adjacent to a church of traditional architecture and within less than 10 km from an urban center providing all facilities for the movie crew (hotels, restaurants, etc.).

2.3 Metadata structure

The metadata provide descriptive information on

both the geographical site and its photo samples. As defined by experts in landscape and environment, the main categories of metadata relative to a site are the following:

- **Geographic location:** geographic area, region, municipality, street;
- **Site characteristics:** public/private land, function (residential, leisure, etc.);
- **Landscape type:** context (urban, rural, forest, etc.), characteristic of the scene (contracted, wide-opened, etc.), site condition (inhabited, empty, abandoned, etc.);

The metadata associated to each photograph can be divided into three categories:

- **Structural aspects of the picture:**
 - o Visual field characteristics;
 - o Presence of foreground, middle ground and background;
 - o Frame characteristics;
 - o Visual effects (composition, boundary lines, kinetic effects);
- **Semantic description:**
 - o **Biophysic attributes:** topographical, hydrographic, vegetation characteristics;
 - o **Human settlements:** built environment, settlement form; architectural style;
 - o **Agriculture:** crops, land divisions;
 - o **Infrastructures:** transport networks and industrial activity;
 - o **Aesthetic and scenic evocation:** realism, style, atmosphere, feeling evocation;
 - o **Contextual information:** ambient sound, odors, meteorological conditions.
- **Low-level visual descriptors:** color and texture features.

3. MPEG-7 Overview

Formally named "Multimedia Content Description Interface", MPEG-7 (ISO/IEC International Standard 15938) was developed by the Moving Picture Experts Group. This standard defines a normative indexing of audio-visual material at many levels ranging from low-level features to higher semantic description [6-7]. It allows to record information on both content management and content itself (structure and semantics). A MPEG-7 description is a XML file instantiating a subset of predefined normative tools. These tools are of two types: Descriptors (D), which define the XML syntax and the semantics of each feature of the multimedia content, and Description Schemes (DS), that describe the metadata elements' structure and their relationships. These tools are

defined by the Description Definition Language (DDL), which merely adopts XML Schema syntax with only few extensions. MPEG-7 can address different media types in various formats and offers a generic metadata framework to support a broad range of multimedia applications that necessitate the interoperability, interpretation of multimedia content (e.g. content-based retrieval, content management, navigation, filtering, and automated processing).

4. MPEG-7 Application

4.1 Motivations

Since an important part of our metadata concerns the description of photographs, working within MPEG-7 framework appears a good choice. In fact, this standard comprises predefined Descriptors and Description Schemes specifically tailored for describing both structural and semantic aspects of images, as requested by our application. It also provides constructs for classification schemes that define sets of controlled terms forming vocabularies for particular domains or applications. Furthermore, it standardizes a set of low-level visual descriptors that have proven to be efficient for similarity search on images [7-8]. Finally, the MPEG-7's DDL offers all the XML advantages such as platform independence and human readability. However, although apparently suitable, the use of the MPEG-7 standard still raises several practical issues for a film location scouting system.

4.2 Implemented MPEG-7 tools

For our application, we only used the Visual Descriptors [8] and the Multimedia Description Schemes (MDS) belonging to the Content Description category [9]. Figure 1 summarizes the MPEG-7 structure we have implemented. The first type of MDS Content Description tools available addresses the structural aspects of an audio-visual piece. For example, one of these schemata, the StillRegion type, regroups all metadata related to a fixed image and its different sub-regions. This DS can also include the MPEG-7 standard Visual Descriptors. The second type of MDS for content description is the semantic description tools. "These tools can be used to describe real-life concepts or narratives, which includes objects, agent objects, concepts, events, concepts, states, places, times, and narrative worlds, which are depicted by or related to multimedia content.

These tools can also be used to describe semantic attributes and semantic relations" [9]. These DS are sufficiently generic to be used for the annotation of all non structural aspects of the sites and photographs in

our database. Since MPEG-7 schema is designed for the description of audio-visual material, in our application, we have considered the photographs as the central point of the description. All our metadata can be split between what is relative to the visual field of the picture and what is outside of it in the nearby shooting location. A MPEG-7 file is then associated to each photograph of the database. MPEG-7 allows linking the description file to the image file, as well as all related material, with Uniform Resource Identifiers (URI). The information about a photo is stored in the ContentEntity DS which contains a MultimediaContent element of Image type. The SpatialDecomposition element of the latter allows indicating the presence of the foreground, middle ground and background. In order to achieve similarity searches, we have used the MPEG-7 standard visual descriptors to encode color and texture features of each photo. For all other metadata relative to the description of the photo, even structural aspects, we were constrained to use semantic description tools. For example, MPEG-7 offers no specific D or DS for the visual field characteristics and the visual effects of a picture. In a semantic perspective, the photograph can be considered as a narrative world that is described with the MPEG-7 types deriving from SemanticBase DS. The site where the picture was taken can be considered as another narrative world related to the photograph and partially depicted by it. Each main category of metadata, (site characteristics, type of landscape, infrastructures, biophysics attributes, human settlements, agriculture, visual field, visual effects, context, aesthetic assessment and scenic evocation) has been described within a SemanticType tool. These generic DS contain other DS more specialized for the description of place, objects and concepts. The SemanticPlace DS is used to express the geographic location information. All physical objects belonging or not to the field of view have been identified using the semantic type ObjectType. For example, presence of a mountain or a house in the picture would be indexed as ObjectType included respectively in the "topographical attributes" and "human settlements" DS. Examples of Concept DS may be found in the description of the aesthetic assessment and evocation of the photograph. Quality and attributes of each aspect of the site and its related photographs have been described either as Property of their respective descriptive semantic tool or as SemanticState. In MPEG-7, semantic states are sets of numerical and verbal attributes that can be attached to semantic entities. Their semantic relation can also be specified.

5. Discussion and conclusion

The goal of the presented work was to evaluate

the applicability of MPEG-7 standard to the set of metadata used by a search engine aimed at location scouting. The resulting schema had to respect the taxonomy defined by experts in the field of landscape and environment and integrate their standard vocabulary. We found that the usability of MPEG-7 in this context has advantages but also important drawbacks. One advantage we found in using MPEG-7 is the high modularity of the standard which allowed us to build our own implementation of the schema in a limited time. Furthermore, the combination of highly specialized descriptive tools with more generic ones gave us the necessary flexibility for expressing every type of metadata of our database. However, since the ontology embedded in the MDS targets the description of audio-visual material, parts of our metadata about geographic sites do not fit well into this formal representation. Thus, we were constrained to use generic semantic DS, labeled only with free text, for the description of this information. These description fragments might impair interoperability because their inherent semantics is not specified in the schema. Unfortunately, in order to insure full compliance to MPEG-7, no new D or DS specific to our application can be created. Despite these limitations, representation of both structural and semantic information relative to sites and photographs forming our database has been achieved. Although MPEG-7 MDS are not designed for managing attributes outside of the scope of multimedia content description (e.g. logistical characteristics of geographic sites), we still succeeded in adapting them to all the metadata of location description. The resulting metadata structure respects landscape and environment description taxonomy. Moreover it is suitable for advanced queries such as visual similarity based retrieval of photographs.

Acknowledgements

This work is financially supported by the Conseil régional de développement de la Montérégie (CRDM), CRIM, CPEUM and Natural Science and Engineering Research Council of Canada (NSERC).

References

1. S. Paquette, G. Domon, P. Poullaouec-Gonidec. Les enjeux et les acteurs du paysage au Québec. In M.-O. Trépanier et al., Concept et opérationnalisation du paysage, vol 2, report CPEUM, pp. 5-64, 2003.
2. M. Zins, J. Jacques. La valeur touristique et économique des paysages. *Teoros*, 18 (1) : 48-51, 1999.
3. S. Foucher., V. Gouaillier, L. Gagnon. "Global semantic classification of scenes using ridgelet transform", Proc. conference Human Vision and Electronic Imaging IX, (SPIE #5292), pp 402-413, 2004
4. L. Gagnon, S. Foucher, V. Gouaillier. "ERIC7: An Experimental Tool for Content-Based Image Encoding and Retrieval under the MPEG-7 Standard", Proc. conf. WISICT2004, on CD-ROM
5. <http://www.pushnsee.com/>.
6. <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>.
7. P. Salembier, T. Sikora. Ed. by B.S. Manjunath. Intro. to MPEG-7: Multimedia Content Description Interface. New York (NY, USA) Wiley & Sons, 2002.
8. Information Technology—Multimedia Content Description interface, Part 3: Visual, ISO/IEC JTC 1/SC 29/WG 11/N4358, ISO/IEC, 2001
9. Information Technology—Multimedia Content Description interface, Part 5: Multimedia Description Schemes, ISO/IEC JTC 1/SC 29/WG 11/N4242, ISO/IEC, 2001



Figure 1. MPEG-7 structure for a location scouting database application. In bold are the categories defined by landscape and environment experts with their implemented MPEG-7 types in *italic*.