INVENIO: An MPEG-7 image indexing platform for content re-use within audio-visual production chains

Titus Zaharia, Alain Vaucelle, Thomas Laquet and Françoise Prêteux
Institut TELECOM; TELECOM SudParis; ARTEMIS Department; UMR CNRS 8145 MAP5
{titus.zaharia, alain.vaucelle, thomas.laquet, francoise.preteux}@it-sudparis.eu;

Abstract

In this paper, we propose a novel image indexing platform, so-called INVENIO (INdexing Visual ENvironment for multimedia Items and Objects). Entirely based on the ISO/MPEG-7 normative specification, the INVENIO platform offers, within an integrated system, visual metadata extraction engine, annotation tools, image databases management tools, as well as appropriated, ergonomic user interfaces. In order to validate the INVENIO platform, we have considered an industrial application related to the issue of content re-use within an audio-visual production chain, including both natural and synthetic (i.e. cartoons) image content. The proposed solutions demonstrated that the exploitation of the MPEG-7 visual descriptors makes it possible to obtain significant savings in terms of production time/cost, while ensuring an optimal re-use of content.

The INVENIO platform has been validated within the framework of the HD3D-IIO structural project of the French CapDigital competitiveness cluster.

1. Introduction

Today, content-based indexing and retrieval methods [1] offer alternative solutions for indexing multimedia material. The principle consists of associating with the content, instead of textual descriptions, mathematical representations of the visual data, related to features such as color, shape, motion or texture. Such representations make it possible to obtain objective image descriptions, which can be automatically calculated.

The principle of the content querying process changes then drastically. Instead of presenting as input to the search engine a set of key-words, as in the case of traditional search engines, the user presents an example (*e.g.* an image) or a manually designed sketch (*e.g.* drawing). Then, the similarity measures associated with visual descriptors makes it possible to automatically perform queries and retrieve similar results from an image database.

The field of content-based retrieval has known a spectacular break-through starting from early 1990's, as testifies the important volume of scientific literature dedicated to this issue. The reader is invited to refer to [1, 2] for comprehensive surveys of this research field. Within this context, an important step has been reached at the beginning of the 2000 years, with the publication of the ISO/MPEG-7 standard [3, 4]. Officially called *Multimedia Content Description Interface*, the MPEG-7 standard proposes a rich set of multimedia description technologies, including both textual and content-based approaches.

In particular, MPEG-7 proposes a set of visual descriptors [4, 5] and description schemes, expressed in a XML Schema-based data description language.

However, the relevance of the MPEG-7 technologies for real-life industrial applications still needs to be demonstrated. In order to achieve such an objective, it is mandatory to elaborate, validate and deploy efficient and ergonomic tools that can facilitate the processes of multimedia annotation, content querying, browsing and navigation in visual databases.

The proposed INVENIO (*INdexing Visual ENvironment for multimedia Items and Objects*) platform notably tackles this issue. INVENION exploits a set of MPEG-7 visual descriptors and integrates, within a unified system, extraction utilities, search engine and ergonomic user interfaces.

In order to validate the INVENIO platform, we have considered an industrial application related to

image indexing with texture/color descriptors for the audio-visual content production. The main issue of interest here concerns the re-use of existing visual content within the production chain. This application was considered and treated by the HD3D-IIO structuring project of the CapDigital (www.capdigital.com) French competitiveness cluster.

The rest of the paper is organized as follows. In section 2, we enounce the generic objectives of the HD3D-IIO project, while underlying the importance of content re-use aspects in term of time/cost savings in the audio-visual production process. Section 4 describes the INVENIO platform, with MPET-7 visual descriptors retained, user interface design and the associated functionalities. Finally, section 5 concludes the paper and opens perspectives of future work.

2. The HD3D-IIO context

The HD3D-IIO project concerns the creation and production of audio-visual content for the television and cinematographic industries. HD3D-IIO major ambition is to provide a novel and unified technological framework, dedicated to an industrial sector in permanent transition and with global, world-wide challenges. Such a technological framework has to be designed collectively, in the perspective of an open industry in terms of numerical exchanges between the industrial actors involved.

The HD3D-IIO consortium gathers major professional actors within the field of audio-visual production (advertisement, 2D/3D cartoon, special effects…) from the Ile de France region and includes companies such as Duran Duboi, Eclair, LTC, Mac Guff, Mikros Images, Teamto, or 2 Minutes.

Concerning the technological bottle-necks identified by the project, the major issues are related to the specification of appropriated exchange formats for both 2D and 3D content, to the elaboration of collaborative production platforms, to the specification of content re-use methodologies, and to the elaboration of efficient data protection techniques, able to ensure a secure content transmission.

Within this framework, one of the major objectives of the HD3D-IIO project concern the image content re-use issues. This issue is stated as follows. Each audio-visual production leads to the creation of dozens of thousands of images. Among them, numerous elements (*e.g.* decorum of a film, characters/accessories in a cartoon…) can be potentially re-exploited within novel productions. The sole condition to be satisfied is to be able to found them efficiently within the collections of images obtained from previous productions.

Today, there is no efficient tool allowing users to efficiently access such image databases in order to facilitate/accelerate the multimedia production process by exploiting previous creations. The unique solution is based on the knowledge, skills and memory of the professionals and artists involved in the production chain. However, such an approach is time-consuming and requires important human resources.

In addition, the existing solutions for content re-use present strong limitations :

– manual keyword indexing is time consuming. For this reasons, content creators are poorly using such an approach because of time constraints,

– most often, the keywords are absent. Moreover, the terms and names associated to image may strongly vary from a production to another, which makes it difficult to retrieve the useful information,

– there is no systematic approach for defining appropriate ontologies with adequate concepts and relations between concepts.

Setting up automatic, content-based indexing solutions for describing the image collections involved in the audio-visual production chain can significantly speed-up the production process. However, the image retrieval techniques should satisfy a certain number of constraints in order to guarantee a tractable solution. The first one concerns the time of response to the queries: interactive rates are here required. The second one relates to the pertinence of the query results, in particular when performing image retrieval in large databases. The use of visual descriptors with high discriminative power is mandatory. Finally, the third constraint concerns the elaboration and development of appropriate search engines, with user-friendly interfaces, supporting a large range of image formats that can be easily integrated within existing production chains.

The INVENIO platform proposed in this paper and described in the next section aims at responding to such various constraints, in order to ensure an optimal re-use of image content during the audio-visual production process.

3. The INVENIO platform

INVENIO exploits the MPEG-7 visual descriptors, briefly presented in the next section.

3.1. The retained MPEG-7 descriptors

The MPEG-7 standard offers a rich technological framework for multimedia content description, with a large set of descriptors and descriptions schemes, and with both content-based and textual approaches.

Among the different parts of the standard, the part 3, so-called MPEG-7 Visual [6] proposes a set of descriptors adapted to content-based images retrieval and associated with visual features such as color, texture, shape, and motion.

Within the HD3D-IIO framework, solely the texture and color attributes have been considered pertinent by content creation experts. For this reason, uniquely color and texture descriptors have been integrated within the INVENIO platform. Table 1 summarizes the retained MPEG-7 descriptors.

For a detailed description of MPEG-7 visual descriptors, the reader is invited to refer to [5, 7].

Let us now describe the functionalities and user interface proposed by the INVENIO platform.

Table 1. MPEG-7 color and texture descriptors.

Color descriptors	
Color space	Specification of the color representation space (*e.g.* RGB, HSV, Luv, HMMD…)
Scalable color	Multi-resolution, color histogram representation based on the Haar transform
Dominant color	Representation of an image with a reduced number (maximum 8) of so-called dominant colors.
Color-structure	Color histogram enriched with spatial information based on a structuring element
Color layout	Description of the spatial color distribution of an image, based on a DCT transform
Texture descriptors	
Edge histogram	Coarse classification of edge orientations into five classes and construction of global and semi-global histograms with respect to such classes.
Homogeneous texture	Multiresolution representation with the Gabor transform, with 6 orientations and 5 resolution levels

3.2. Supported functionalities

Being dedicated to content creators, the integration of INVENIO into a graphical and ergonomic environment is determinant for its successful deployment within the content production chain.

From a functional point of view, INVENIO supports the following features:

- query by example: the user can browse an image database, select an query, specify a sketch or select an image sub-part,
- query by pre-defined keywords: a mini-ontology (Figure 1) related to texture specification has been set up with the collaboration of graphical experts. It includes basic texture motifs, materials, shapes, colors, frequency of repetition…
- query by tags: a free text annotation mechanism has been included in order to make possible the use of personalized tags during the production process.

In addition to such individual query and annotation functionalities, the INVENIO system also makes it possible to combine multiple search and description criteria. Such a feature is particularly useful for taking into account the intrinsically poly-semantic nature of the image content, by jointly exploiting different search modalities, including content-based and semantic driven.

The proposed system offers a effective relationship between natural language (*i.e.* tags), structured representations (*i.e.* keywords from the established ontology) and content-based descriptions (*i.e.* MPEG-7 visual representations).

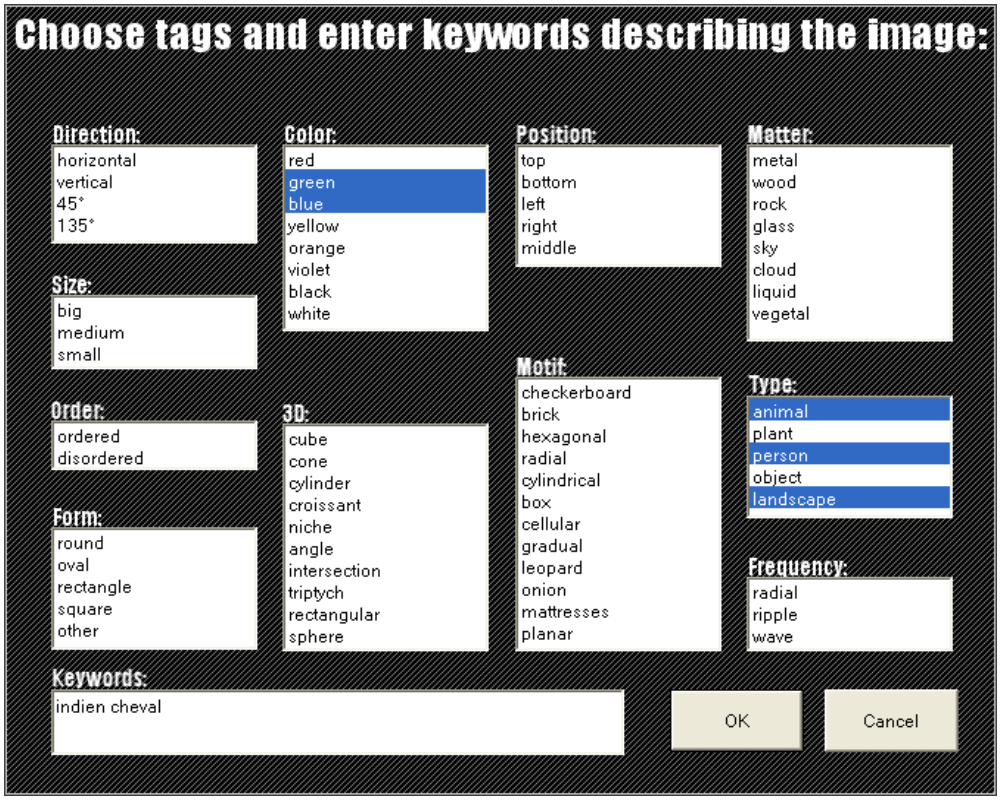

Figure 1. Mini-ontology and tag annotation with the INVENIO platform.

3.3. Graphical user interface

The design of appropriate and ergonomic graphical user interfaces is a key element for the adoption of an image indexing platform by graphical experts. In our work, we have considered a 3D approach, implemented in OpenGL.

Figure 1 illustrates the INVENIO graphical user interface. The different constitutive elements, numbered from 1 to 8 are explained here-below:

1: The image database is graphically presented as a rolling 3D spiral that allows the navigation through image databases. At each query performed, the results are order by decreasing order of similarity with the query image.

2: Frame presenting the active image. The user has also the possibility to select a region of the active image.

3: Frame for memorization/visualization of the current query.

4: Metadata related to image file format and textual keywords.

5: Help menu.

6: List of available color/texture descriptors.

7: Query lancer button.

8: External image loading.

In terms of image database and query results presentation, the key feature proposed is the 3D rolling spiral, which creates user immersion in a 3D navigation space dynamically peopled with relevant images.

The image retrieval experiments have been carried out on different databases totalizing 10000 images, including both natural and synthetic images and constituted by content creators. They correspond to real-life film productions, such as Yakari, Atomic Betty, Cajou, Step by Step, Faubourg 36.

Experimental results have demonstrated the pertinence of the INVENIO approach and of the

adopted MPEG-7 visual descriptors. The retrieval rates are of 60%-70%, for time of response to the queries inferior to 0.5s. The highest performances in terms of retrieval rates are achieved by the color-structure and dominant color descriptors.

Figures 2 and 3 present some examples of similarity search with the INVENIO platform. In the first case, a region of an image in the database has been used as a query. In the second one, a manually created sketch has been used.

Figure 2. INVENIO user interface and query by example.

Figure 3. Query by sketch with the INVENIO platform.

Let us note that a similar character in the database has been retrieved on the first positions.

4. Conclusion and perspectives

In this paper, we have presented the INVENIO image indexing and retrieval platform.

Based on MPEG-7 technologies for automatic, content-based description of visual content, INVENIO proposes ergonomic user interfaces and tools for browsing, navigation and search in image databases.

The INVENIO platform has been validates within the framework of the Capdigital HD3D-IIO French national project, for applications of content re-use within the audio-visual production chain.

Perspectives of future work concern the integration of advanced descriptors that can further optimize the performances of the system. In particular, an important issue concerns the elaboration of methods dedicated to partial image search, in order to be able to identify items of interest in cluttered scenes. At a longer term, our future work concerns the implementation of the INVENIO platform as a web application, in order to make possible the re-use of content in collaborative environments.

Acknowledgment

This work has been partially supported by the French national project HD3D-IIO, labeled by the French CapDigital competitiveness cluster.

6. References

[1] Gupta A., Jain R., Santini S., Smeulders A. W. M., Worring M., *Content-Based Image Retrieval at the End of the Early Years*, IEEE Trans. on PAMI, Volume 22, Issue 12, Décembre 2000.

[2] Datta R., Li J., Wang J., Wang, J.Z., *Content-based image retrieval: approaches and trends of the new age Proc*, 7th ACM SIGMM international workshop on Multimedia information retrieval, p. 253 – 262, 2005.

[3] ISO/ IEC 15938-5, Information technology - MultimediaContent Description. Interface - Part 5: Multimedia Description Schemes. 2003.

[4] Manjunath B. S., Salembier P., Sikora T., *Introduction to MPEG-7 : Multimedia Content Description Interface*, John Wiley & Sons, 2002.

[5] Zaharia, T., Prêteux F., Descripteurs visuels dans le standard MPEG-7, chapitre dans Gestion des données multimédias (Chapitre 5), *Traité IC2 - Série Informatique et Systèmes d'Information*, Mostefaoui, A., Prêteux, F., Lecuire, V., Moureaux, J.-M. (Eds.), Editions Hermès-Lavoisier, Paris, 2004, p. 85-139.

[6] ISO/IEC 15938-3: 2002, MPEG-7-Visual, Information Technology – Multimedia content description interface – Part 3: Visual, 2002.

[7] Zaharia, T., Prêteux F., Normes de description des contenus multimédias. L'indexation multimédia - description et recherche automatique, *Traité IC2 - Série Traitement du Signal et de l'Image*, P. Gros, Editions Hermès-Lavoisier, Paris, 2007, p. 163-185.