A flexible environment for multimedia management and publishing

M. Bertini, A. Del Bimbo
Univ. of Florence - MICC
Firenze, Italy

[bertini,delbimbo]@dsi.unifi.it

G. Ioannidis, A. Stan
IN2 search interfaces development
Bremen, Germany

[gi,as]@in-two.com

E. Bijk
Hogeschool voor de Kunsten Utrecht
Hilversum, The Netherlands

emile.bijk@kmt.hku.nl

ABSTRACT
In this paper, we describe the IM3I system, which provides a flexible approach to managing and publishing collections of images and videos. The system is based on web services that allow automatic and manual annotation, retrieval, browsing and authoring of multimedia. Results of user evaluations, performed by professional archivists and archive managers on a real-world system deployment have confirmed that the system is easy to be used and delivers a complete set of functionalities.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval]**]: Information Search and Retrieval – *Search process.* H.3.5 [**Information Storage and Retrieval**]: Online Information Services – *Web-based services.*

General Terms
Design, Experimentation.

Keywords
Video retrieval, authoring, ontologies, web services.

1. INTRODUCTION
Current state-of-the-art software for digitally capturing, storing, and displaying ever-increasing audio-visual information requires attachment of metadata, keywords, and other information for content, which is highly subjective, difficult to scale, and evolves over time. Automatic video annotation systems are based on large sets of concept classifiers [1], typically based on supervised machine learning techniques such as SVMs; however the current performance of these systems [2] is still not acceptable for professional video archives, that require a semi-automatic system where a human can check and validate the results of an automatic annotation system. Regarding the system interfaces, two surveys conducted within the EU VidiVideo (www.vidivideo.info) and IM3I (www.im3i.eu) projects to gather user requirements from more than 50 professionals working in broadcasters, cultural heritage institutions and media archives organizations, have shown that web based systems are "mandatory" (75% of the interviewees) or "desirable" (20%) [3].

In this demonstrator we will showcase IM3I – Immersive

Multimedia Interfaces. The system puts together a series of innovative interactive applications and services that empower content owners and collection holders in aggregating, developing, and hosting creative content to interact smoothly with multimedia data using advanced mixed multimedia annotation, indexing, search and retrieval tools, and navigation interfaces. These services are enabled by a publishing environment that exploits repository content by the authoring of flexible user interfaces, whereby each interface element can be fully configured regarding its layout, the data element it represents, and the functionality it provides. Based on service-oriented architecture and middleware, modules and tools developed are easily integrated into existing frameworks and systems, reducing integration costs.

2. The System
IM3I can be defined as intelligent and adaptable multimedia repository software. The notion of 'intelligence' is derived from the ability of the platform to derive semantic descriptions from media assets, while 'adaptability' refers to the authoring capabilities of the platform that enables users to define and publish repositories of media artifacts in an easy manner. Both capabilities are derived from and facilitated by an open and flexible three layers architecture composed by: Interface Layer; Architecture layer and Analysis Layer. Communication between analysis and interface layers is routed through the architecture layer, which also takes care of the main repository functions. The IM3I platform is designed in an open and flexible manner following the notions of a services oriented approach, that allows the deployment and execution on a distributed network.

2.1.1 Multimedia search and annotation
Automatic video annotation is performed using a system based on the BoW paradigm [4], that uses a fusion of MSER and SURF interest points and the Pyramid Match Kernel [5]. The system is part of the set of web services of IM3I and can be combined in a processing pipeline with other services like audio track extraction, audio segmentation and media transcoding. Pipelines are created and managed by users through appropriate services of the Analysis layer. Audio concepts are detected using a combination of timbral texture features like ZCR, MFCCs, chroma and spectral features, and SVMs. A web-based tool to validate automatic annotations, add other annotations for which no concept detector has been trained yet and specify additional metadata like geo localization of annotations. This tool supports different access levels to deal with the typical workflow of a video archives. Semantic annotation is complemented by low-level annotation, to allow content-based retrieval, a feature that is of interest for music and photo archives. Images and keyframes are processed to extract a combination of MPEG-7 descriptors (CLD, SCD and EHD), while audio is processed to extract rhythm and pitch features.

A set of web services for semantic and syntactic retrieval, with appropriate specialized search and browsing interfaces has been developed. Semantic search and browsing is based on an ontology-based search engine (Figure 1), and the CBIR engine copes with large-scale archives using approximate NN search [6]. All the search tools (implemented in Flex/Flash according to the RIA paradigm) are integrated so that, for example, it is possible to start browsing the concepts of the ontology, select a concept to create a specific and complex query (using Boolean and temporal operators, specifying media and geolocalization metadata) and then select images and videos based on visual similarity.
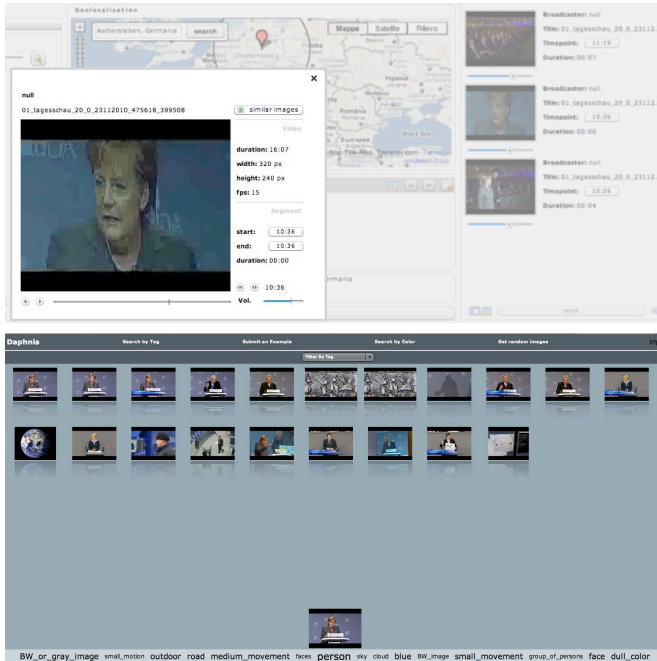

Figure 1 A semantic search (top) used to start a content-based search (bottom)

2.1.2 Authoring and publishing repositories

Flexible search, browse and display interfaces for media collections is enabled through a user-friendly authoring environment, whose versatility permits for a great variety of usage scenarios.

As reflected by the navigation interface of the environment (Figure 2), a typical workflow for authoring IM3I end-user functionality is composed of the following steps: 1) importing an existing repository; 2) extending the associated data model; 3) editing layout and feature 4) editing search and retrieval interfaces 5) embedding the end-user interface in a website. This latter functionality is particularly interesting since it allows to deploy IM3I in an existing media management system of an archive, without requiring to redesign it from scratch.
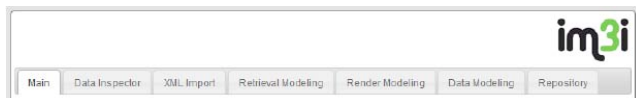

Figure 2 The navigation interface reflecting the workflow of authoring an IM3I user interface

The functionalities provided include: preview and import RSS streams and XML based repositories and use multiple data-objects to store the contents; dynamically create, copy and delete data models, as well as define relations between data models;

managing flexible element-types; create and preview multiple layouts for a given data model (see Figure 3); manipulate views and edit data objects; inspect repositories and limit them for publication; publish and embed a workflow in a corporate website.
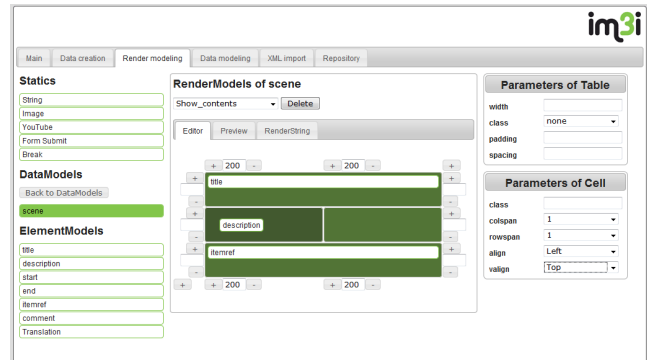

Figure 3 Drag-and-drop repository content to defined layout

The more advanced functionality of repository linking is achieved through the introduction of 'object elements'. This artifact allows for multiple instances of one repository to be shown in another. Navigational components can be added to render models in order to create elaborated sequences of render models. The authoring environment is integrated with the pipelines of the search and annotation tools described in Section 2.1.1. As the generated interfaces are created in real-time, Javascript/AJAX was used for implementation.

3. Demonstrator

User evaluation in field trials with expert professional archivists was very positive (3.8/5 marks for ease of use and usefulness of the combination of tools).

The live demo shows all the tools of the system, from the creation and management of the processing pipelines, the annotation, browsing, semantic and syntactic search tools, to the creation of the flexible interfaces of the authoring environment. ***Acknowledgments*** The IM3I platform (www.im3i.eu) was developed in 2008-2010, co-financed by the European Commission under the FP7 programme.

4. REFERENCES

[1] C. G. M. Snoek, K. E. A. van de Sande, O. de Rooij, *et al.*, "The MediaMill TRECVID 2010 semantic video search engine," *Proc. of TRECVID Workshop*, 2010.

[2] A. Smeaton, P. Over, and W. Kraaij, "High-level feature detection from video in TRECVid: a 5-year retrospective of achievements," *Multimedia Content Analysis, Theory and Applications*, pp. 151–174, 2009.

[3] M. Bertini, G. D'Amico, A. Ferracani, M. Meoni, and G. Serra, "Sirio, Orione and Pan: an integrated web system for ontology-based video search and annotation," *Proc. of ACM MM*, 2010.

[4] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," *Proc. of ICCV*, 2003.

[5] K. Grauman and T. Darrell, "The pyramid match kernel: Efficient learning with sets of features," *Journal of Machine Learning Research (JMLR)*, vol. 8, pp. 725–760, 2007.

[6] G. Amato and P. Savino, "Approximate similarity search in metric spaces using inverted files," *Proc. of InfoScale*, 2008