BRIDGET: AN APPROACH AT SUSTAINABLE AND EFFICIENT PRODUCTION OF SECOND SCREEN MEDIA APPLICATIONS

M. Bober[1], I. Feldmann[2], S. García Lobo[3] , A. Messina[4], S. Paschalakis[5], G. Perrone[6], V. Scurtu[7], G. Vavalà[8]

[1]University of Surrey, [2]Fraunhofer Heinrich Hertz Institute, [3]Univ. Politécnica Madrid, [4]RAI - Radiotelevisione Italiana, [5]Visual Atoms, [6]Telecom Italia, [7]Institut Mines Telecom, [8]CEDEO.net

ABSTRACT

This paper provides insights of an advanced architecture for authoring and consumption of second screen content aiming to develop innovative functionalities for enjoying multimedia content by connecting it to other related content, augmenting it with virtual information of interest, and allowing navigation of the 3D reconstruction of the scene. Innovative technology and the underlying architecture for efficient production of second screen applications is described, which includes novel front-end authoring tools as well as back-end enabling technologies such as visual search, media structure analysis and 3D A/V reconstruction to support new editorial workflows. A strong commitment on standardisation by the project's partners ensures the future-proof utility of the results.

INTRODUCTION

Current second screen applications let user enjoying broadcast programmes access related information on other – typically internet-connected – devices. Although several services are already available in the consumer domain the problem of ensuring sustainability of such applications in the media production domain and for generic programme formats still remains vastly unsolved. The most relevant challenges to be tackled in this area are: 1) making production costs related to enriched content affordable for broadcasters; 2) second screen presentation must be ergonomic and functional, and links should include rich and exciting content (e.g., 3D models) to ensure a differentiating factor for the quality of user experience; 3) authors and editors must be supported by efficient authoring tools, which must also be easy to integrate in existing workflows.

To address these challenges BRIDGET, an FP7 European Commission's collaborative project, has developed the namesake notion of a bridget: a link from the programme being watched to (combinations of) external interactive media elements such as web pages, images, audio clips, different types of video (2D, multi-view, 3D free viewpoint) and synthetic 3D models. Bridget are links which represent some inherent semantic relationship between content items. As such, they can be products of an editorial decision, taken by someone inspecting (manually or automatically) content items, or they can be objects of a workflow which involving different roles taking care of finding, organizing and crafting the data that constitute them. In order to facilitate the bridget creation, the project develops: a) advanced visual search tools that help locate semantically related images and video segments in archives; b) advanced media structure analysis and content annotation tools that help identify candidate content segments to be enriched and offer quick and efficient navigation through content; c) advanced A/V 3D reconstruction tools to provide new content types to enrich the main content; d) efficient authoring tools integrating the above technologies in the production workflow.

RELATED WORK

Second screen applications for TV shows have a long history, and there are many existing applications which are relevant in the context of this work. Amazon's X-Ray service, Google Play Movies & TV are examples of general purpose applications that provide additional content related to a reference main content e.g. in-depth information about actors playing in a movie which does not necessarily need to be watched at the same time. Another class of applications are the live sync applications, i.e. applications that provide additional content on a time-synchronous way by exploiting some audio-based or event-based synchronisation mean. These normally provide a more detailed and interactive experience than general purpose apps, e.g. for episodes of fiction series. Examples are the AMC's Story Sync and Team Coco. An application based on a similar model to BRIDGET's is the SyFy Sync application, which uses audio-based synchronisation to allow second screen experience independently of the actual broadcast time. Despite the liveliness of the developments in the reference area of BRIDGET, there is still a lack of systems and approaches capable of making second screen application sustainable in production, which is exactly where the novelty of BRIDGET lies. This novelty resides in the capability released for broadcasters and media companies to efficiently and rapidly produce such applications for a wide range and variety of programmes, including archived ones, to retain editorial control over the additional content, and to exploit the vast amount of relevant content (e.g., outtake) which remains available downstream of a production for a further innovative usage. This is ensured by the employment of a flexible architecture integrating a set of cutting edge backend technologies and an authoring tool specifically designed for the purpose. These features are explained in the next two sections.

ARCHITECTURE

The BRIDGET architecture enables a number of actors (Broadcasters, Application or Service Providers but also users) to create BRIDGET applications and complex services coordinating multiple BRIDGET applications. Applications use functionalities provided by a set of defined BRIDGET components and exposed through APIs exposed by a middleware layer to create particular applications. Applications can access these APIs locally or remotely in a distributed environment; they can be locally installed or delivered through a network to the actual user devices and then executed. Depending on the context, an application may be developed to interact with a Professional User (e.g. in a broadcaster's post production environment) or an end user.

Figure 1 shows the described BRIDGET Architectural model.

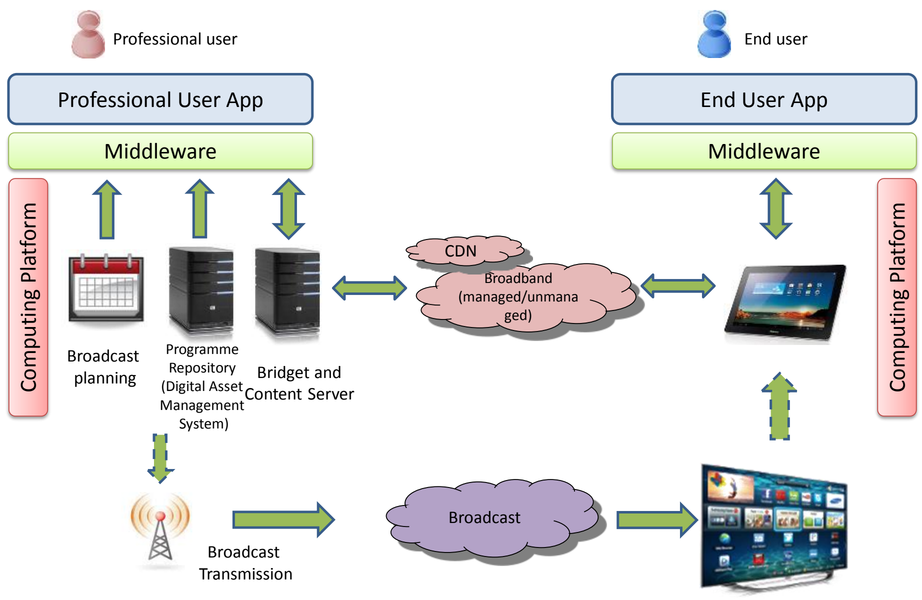

Figure 1. BRIDGET Architectural model.

Figure 2 shows the BRIDGET functional ecosystem built on top of the architectural model. The ecosystem includes two main blocks: the Professional Authoring Environment and the End user's Application. The former includes tools and subsystems supporting the bridget authoring phase, i.e. the process by which an editor actually selects where in the programme timeline to attach enrichment content and what content to present and with which graphical layout. The latter is the actual end-user application, which presents the resulting synchronised experience to the user. A functional detail of each of these two blocks is presented in Figure 3. The Professional Authoring Tool is developed as a GUI integrated with backend APIs connecting it to a set of subsystems providing advanced capabilities supporting the authoring work.

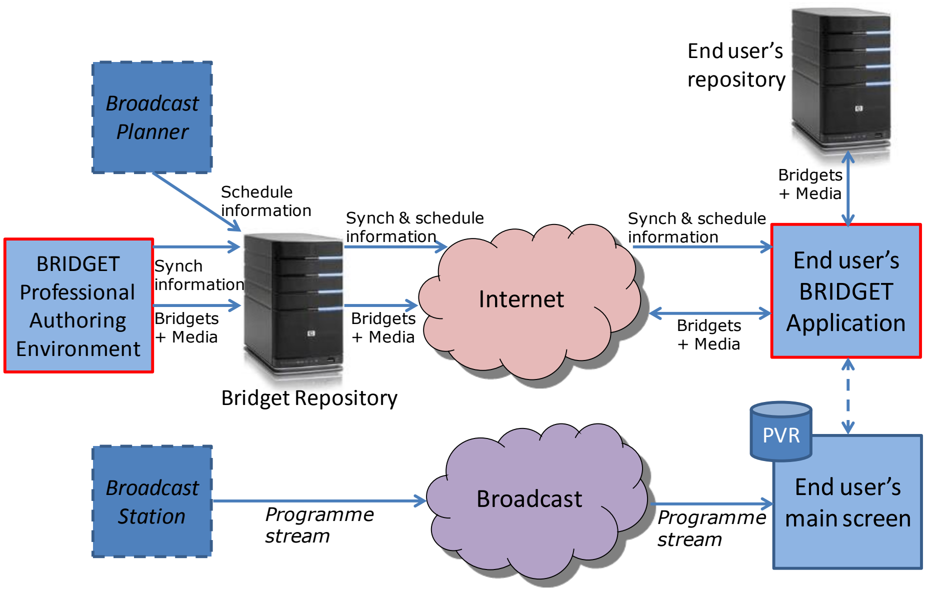

Figure 2 – The BRIDGET ecosystem.

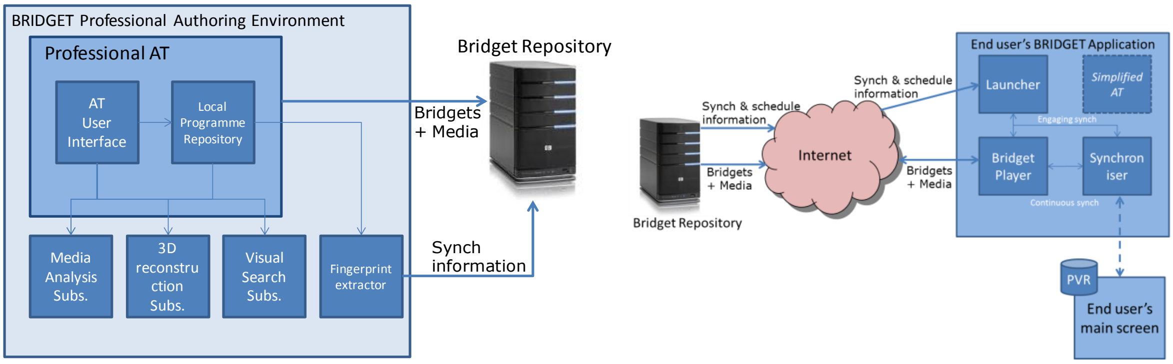

Figure 3 - BRIDGET's Professional Authoring Tool and End user Application.

The tools integrated in each of these subsystems are illustrated in the following Section.

ENABLING TECHNOLOGIES

Visual Search Tools

During the bridget authoring phase, bridget creators can pose a visual query by selecting an object of interest (e.g. a building or a book cover), entire frame (e.g. a scene) or a video segment (e.g. a shot of an interior of a museum). The search process will list all available content in the database that is linked by visual similarities to the query. The author can then choose to utilize the results directly as enrichment content or reuse the metadata associated with them. To reach acceptable performances both in terms of indexing and retrieval time and to obtain relevant search results, the project developed sophisticated technology by extending the CDVS standard 1with new state of the art algorithms. The BRIDGET visual descriptor extraction and search pipeline is presented in Figure 4.

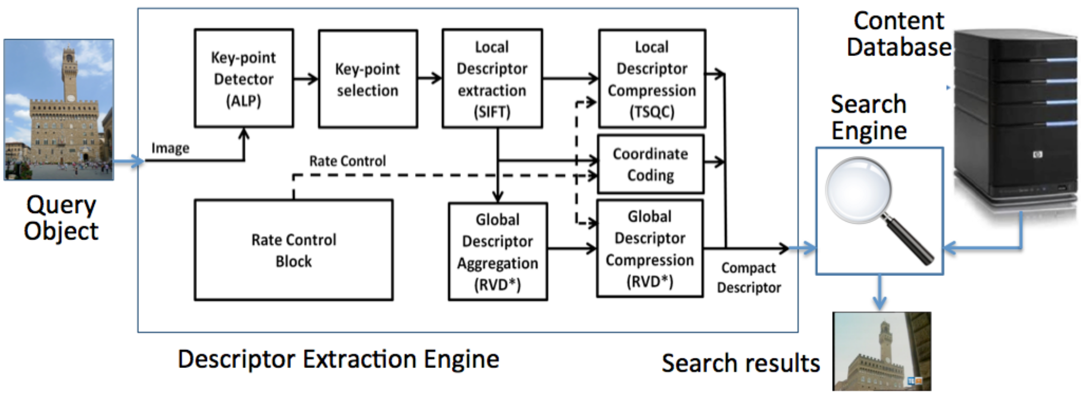

Figure 4 - BRIDGET descriptor extraction and search engine.

Given the very large volumes of images and videos available in archives, particular focus was placed on efficiency, sensitivity and robustness of the visual search engine. New algorithms were developed for fast key-point detection (ALP), global-descriptor aggregation (RVD*) and fast geometric consistency verification (DISTRAT), advancing significantly the state-of-the-art.

3D Media Tools

BRIDGET's 3D media tools aim to enrich the multimedia content by providing additional 3D audio and video information. Based on this the user is enabled to change the 3D view point of a given object of interest (see examples in Figure 5) or to virtually navigate in a 3D scene. The 3D audio experience can be adapted accordingly.

Figure 5 –Results for 3D reconstruction, left) splat based representation, right) depth based 3D surface reconstruction

In BRIDGET the 3D Media tools target at the service provider's side multi-view 3D A/V media generation as well as the possibility of 3D reconstruction updates and refinement at the end user's side. In this way the content creator can run a semi-automatic extraction of 3D scene information from archived (e.g., previously broadcast) 2D/3D A/V content captured from several viewpoints. After its efficient transmission, the virtual or augmented 3D scene is decoded and rendered according to the user-selected viewpoint.

From a technical point of view, 3D reconstruction workflow consists of an initial structure from motion step with subsequent AKAZE feature based refinement of the 3D point cloud (Figure 6). Depending on the 3D scene complexity various types of 3D reconstruction and visualization techniques can be then applied, such as splat based 3D surface modelling or dense depth based 3D surface reconstruction.

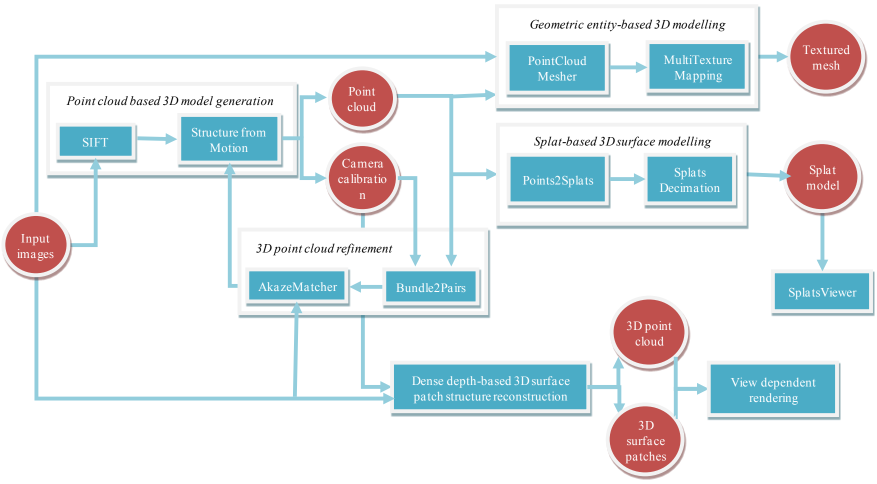

Figure 6 – Work flow for 3D reconstruction tools

Media Analysis Tools

The division of long programmes into lower level structures aims to increase the efficiency the quality of the content produced by the authoring tools by helping editors quickly segment content. The specific challenge in this area is to specify media structure analysis tools that allow implementations which are much faster than real-time or any existing systems so that they may be effective in enhancing the operation of the visual search and authoring tools. Media structure analysis results make it possible to quickly navigate bridget source and destination content through multi-modal storyboards, organized around a hierarchical shot/scene structure, or around the people which appear in different segments, and so on (see example of Figure 7).

The operator can then apply the same action, e.g. creating a bridget and associating content to it, to many portions of the programme in one single action.

Another key area is that of automated content annotation and scene classification, to provide meaningful labels to portions of content in order to support efficient content filtering on the authoring side. BRIDGET content annotation tools are based on a novel aggregation mechanism which uses dense SIFT descriptors as input features and Support Vector Machines as classification technology and reaches state of the art performance (**Error! Reference source not found.**).

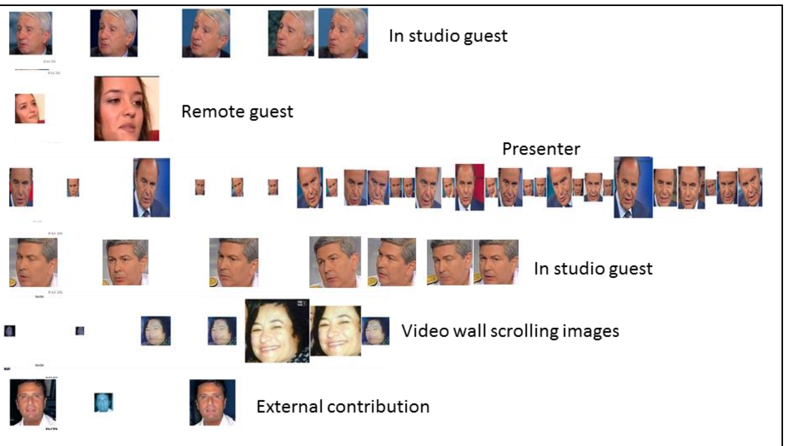

Figure 7 – Example of media structuring results based on automated face clustering.

Presentation Technologies

In BRIDGET the second screen is not only used for presenting additional content to the end user but also to offer an interactive and personalized experience. Therefore, the presentation technology should address requirements such as multi-media representation (text, image, video, audio, 2D and 3D graphics), dynamic behaviour (through scripting), classic and natural interaction (touch, motion, vibration), connection to various sensors (cameras, microphones, accelerometers). A straightforward implementation of a second screen presentation layer is the development of native applications, the mobile devices programming eco-system being nowadays mature in terms of functionalities and available expertise. However, the BRIDGET vision on this particular matter is to not follow this straightforward approach and to separate the software development from the content design. This follows the main trend in Web sites design where non programmers, by using advanced authoring tools, are able to produce web pages faster and as good as the ones produced by programmers using directly HTML5 and JavaScript. BRIDGET specificity is the development of a dedicated presentation format in order to offer a mechanism to parameterise, in a easy manner, the second screen presentation experience. At a lower level, BRIDGET is reusing BIFS 3and HTML5 5.

THE PROFESSIONAL AUTHORING ENVIRONMENT

Professional authors' goal is to present on the second screen content that has semantic connections with the first screen content: additional information about people, places, events, etc. Together with including classical text-based search mechanisms the BRIDGET Authoring Tool goes beyond this paradigm allowing also visual links performed by using Visual Search Tools described previously: by using an image as query, the professional author obtains a set of similar images from the repository, images that may be already connected to other metadata (text or other media). He can then chose the ones he want to include in the bridget and select what kind of information will be presented (the media itself, metadata that comes with the media, …) and how this will be presented (the layout). All those functionalities are made available through the Professional Authoring Tool, implemented as a web tool based on the MyMultimediaWorld 2platform. This framework manages multimedia content and the associated descriptions in a cloud-friendly manner and is built around a set of services which are exposed to the frontend through a REST API as presented in Figure 8.

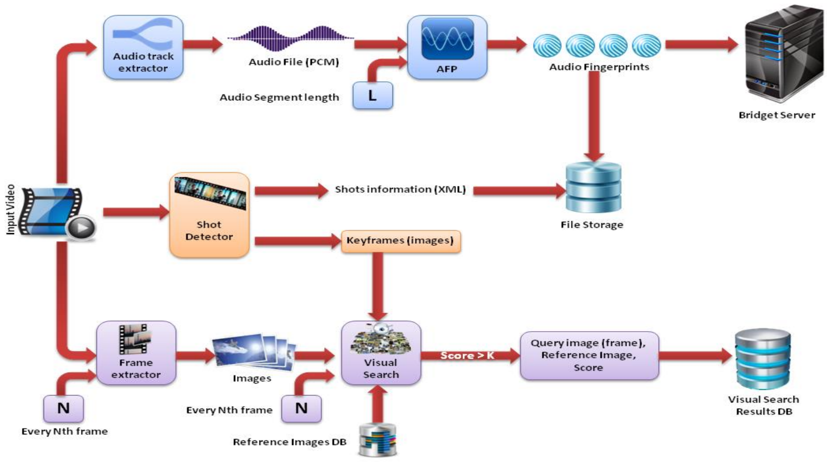

Figure 8 - Back-end components and their connections.

The Programme Player allows the bridget author to select shots candidate for bridget enrichment, using results provided by the Media Analysis Tools. When the bridget editor decides that a shot should be enriched he starts the creation of a bridget. Destination content can then be selected manually or by calling the visual search subsystem. The author can select one or more items and include them as bridget destination content and view the bridgets list of an enriched programme. See Figure 9 - Figure 12.

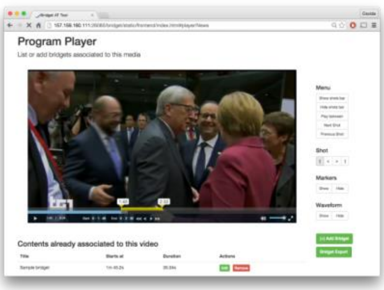

Figure 9 - Find source point for a bridget.

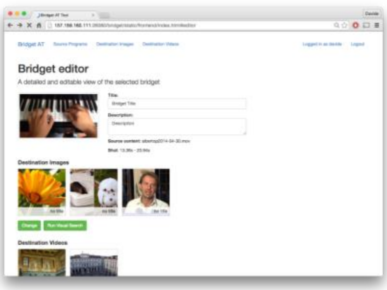

Figure 10 - A bridget is created.

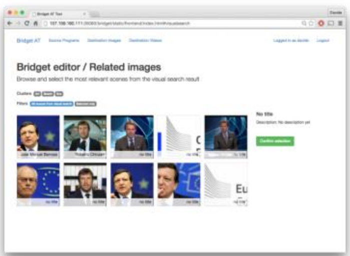

Figure 11- Suggestions by visual search.

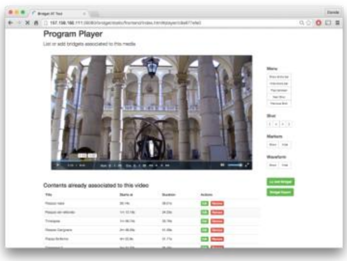

Figure 12 - Viewing all bridgets.

USER EXPERIENCE

The result of the authoring is a data format describing the interactive experience, which currently can be exported in MPEG-4 BIFS 3 and played by a standard MPEG-4 BIFS player. The BRIDGET player, based on GPAC open source 4, is supporting representation of all data types required by rich media applications, composition of synthetic and natural objects, and access to remotely/locally stored audio, video and graphics. The player supports user interaction and server generated scene changes as well as management of the physical context, either captured by a broad range of standard sensors or affected by a broad range of standard actuators. The applications UI design concentrated more on the presentation of BRIDGET functionalities on the second screen, in order to optimize the user experience also depending on each programme's target audience. Three scenarios were elaborated for three kinds of TV programmes: a news report, a documentary on Torino architecture and a TV entertainment show.

Figure 13 shows the UI design for the news report, where bridgets are centred on the subjects discussed in the report. The users can browse at any time the summary of the show, in order to see the list of the discussed subjects. Users don't have access to all the bridgets from the beginning, but only when such bridgets become "available", i.e. when the related subject is presented in the show.

Figure 13 - UI design as presented on the tablet for the news report.

The UI design of the documentary content was conceived around the map of the city, as shown in Figure 14. When a bridget becomes available, a new landmark is presented in the documentary, it is positioned on the map. Restaurants and hotels situated in the landmark's neighbourhood are also shown and at the end of the experience, a quiz related to the presented content is proposed to the user.

Figure 14 - UI design as presented on the tablet for the documentary

The TV entertainment show application (Figure 15) was focused on street style, make up, look and hair styles. At the end of the programme, a dressing game is proposed to the user, consisting of creating different outfits for a character by combining garments from the proposed gallery. The game also provides a feedback for each created outfit regarding the user's skills in matching different clothes.

Figure 15 - UI design as presented on the tablet for the TV entertainment show

CONCLUSIONS

This paper presented an advanced architecture for efficient and sustainable production of second screen applications for broadcasters based on the results of BRIDGET, a collaborative European project of the FP7 framework. The project develops technologies and tools enabling a producer to plan, design and realise second screen rich-content applications including support for innovative content types like 3D models visualisation. The consortium partners are active in the standardisation of their technologies in ISO (MPEG), to make results of the project available and exploitable beyond the project life.

ACKNOWLEDGEMENTS

This work has received funding from the European Commission under Grant Agreement 610691.

REFERENCES

1. ISO/IEC FDIS 15938-13, Information technology – Multimedia content description interface (MPEG 7) – Part 13:Compact descriptors for visual search, ISO, Dec. 2014**.**

2. http://www.mymultimediaworld.com/

3. ISO/IEC 14496-11:2005 - Information technology -- Coding of audio-visual objects -- Part 11: Scene description and application engine.

4. http://gpac.wp.mines-telecom.fr/

5. http://www.w3.org/TR/html5/