INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND ASSOCIATED AUDIO

Title: **A Possible Extension of the Generic AV DS to Incorporate Highlighting and Ordering Functionalities**

Status: **Proposal**

Source: **University of Brescia**

Authors: **Riccardo Leonardi and Lorenzo Rossi**

1   Introduction

In this document, we propose an extension for the MPEG-7 Generic DS [2]. We believe that some important functionalities for MPEG-7, not yet addressed by the actual DS, should deal with the capability to (1) highlight description items (e.g. images, sounds, events, objects etc.) most relevant to the purpose for which a certain content description of a multimedia (MM) document has been created and (2) the capability of description information ordering. In other words, due to a possible large amount of description items, an entity who will create descriptions of multimedia (MM) documents, according to MPEG-7 specification (i.e. a **description provider**), shall highlight certain items most representative for the kind of document being described in order to facilitate user queries. Besides we consider the need of providing users with ordering mechanisms a very relevant issue for MPEG-7. Such ordering mechanisms are derivable from descriptors (e.g. a set of key − frames ordered on the basis a color descriptor or a set of sounds ordered by means of a loudness D). However a possible large variety in the types of descriptors composing a description could lead to a consequent high number of ordering criteria to arrange description items. Therefore we propose that the description provider should also select a reduced set of descriptors allowing to order a subset of description elements (e.g. key frames, events etc.) pertinent to the MM document being described.

Our proposal consists in the incorporation in the current Generic Ds of two DSs covering the aforementioned functionalities. The document is organized as follows: in Section 2, we explain, after a quick overview of the current Generic DS, the motivation behind our proposals and in Section 3, we show the detailed structure of the DSs.

2 Motivation

The MPEG-7 Generic DS was generated as a video DS during the MPEG Seoul Meeting (March 1999) [2]. On the MPEG-7 DS reflector, it has being extended to describe the content of audio-visual documents [1]. It is composed of three main components: the syntactic structure DS, the semantic structure DS, and the syntactic-semantic links DS. The syntactic structure DS consists of region trees, segment trees, and segment/region relation graphs. The semantic structure DS is composed of object trees, event trees, and object/event relation graphs. The syntactic-semantic links DS presents the links between the syntactic elements (regions, segments, and segment/region relations) and the semantic elements (objects, events, and object/event relations).

The DS substantially reflects the structure of a book where the Syntactic DS aims to represent the table of contents of an audio-visual document (i.e. its temporal structure). The Semantic DS aims to represent a set of indexes of the MM document with the locations of the items being provided by the syntactic-semantic links DS. The elements forming the semantic DS are the events and the objects. However there can be descriptions with an arbitrarily high number of events, objects, key-frames, mosaics, etc. leading to some difficulties in order to be exploited by users. It does not seem sensible to propose a limitation in the size of the description since a high level of details could be useful in several cases. We believe that such a problem could be overcome instead by introducing a DS to highlight the description items relevant to a particular document, given the purpose for which the description has been created or used. In this way, users can be facilitated in their queries by knowing, e.g., the most relevant topics of the documents being described while preserving at the same time all the necessary level of details. The kinds of highlighted items (i.e. **key-items**) could vary according to the category of document being described. For instance, in a documentary about flowers, the visual key items, defined as key-images, could be represented by images of the flower being contained in the programme.

The concept of key-items for MM documents can be seen as a generalization of the concept of key-words for text documents. As in a book or a scientific paper, the key-words are highlighted by the editor and/or the author (rarely by the readers themselves), for MM documents as well key-items should be highlighted by the description provider according to the category of the described document and the purpose for which the description is used. A criterion for selecting the key-items could be based on a clustering of the corresponding description items, according to the semantic they carry. In other words, a subset of description items having the same semantics, can be represented by one key-item (linked to all its represented items).

Another issue deals with the ordering mechanisms for the key-items (but not only). It is easy to understand that a suitable ordering of items can help user queries and that the ordering can be based on the values of several descriptors (e.g. color histogram, audio loudness etc.). However in MPEG-7 descriptions, there can be so many kinds of Ds making quite difficult the selection of ordering criteria for the items to be arranged, also depending on the kind of key-items itself. Hence we believe that in this case also a DS presenting a set of suitable **ordering keys** should be incorporated in the Generic DS so that to facilitate selection processes.

3 Structure of the proposed DS

We have grouped the two DSs addressing the aforementioned functionalities in a DS called **Analytical Index DS**. Figure 1 shows the UML diagram of the Generic DS. In Figure 2, is represented the Generic DS with the proposed integration.
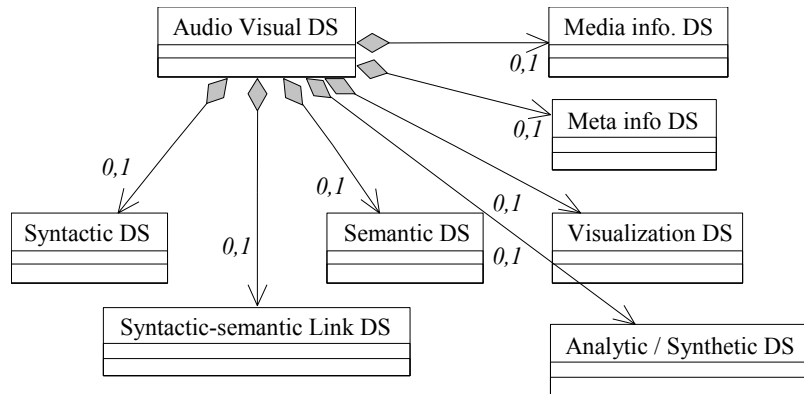
Figure 1: The Generic Audio-Visual DS.

The analytical index DS consists of two DSs: the **Key-Items DS** and the **Ordering Key DS** (see Figure 3). The key items DS can be seen like a generalization of the concept of key-words in the context of MM documents. It is composed of several DSs (key-events DS, key-objects DS, key-images DS, key-sounds DS etc.), each of them consisting in a set of key-items representative for certain type of description element (e.g., events, objects etc.). The ordering keys DS presents a list of applicable ordering mechanisms for key-items (e.g., a set of key-images of a violent movie being ordered according to the level of underlying audio loudness descriptor). Obviously these ordering keys can be applied also to order description items (i.e. to all key-frames and mosaics, for example).

Every key-items DS must include also the links to its represented items and the links to its corresponding representation (e.g., a set images). In Figure 4 is shown the structure of the key-images DS. The other key-items DS (key events etc.) reflects the represented structure. Thus by means of the Syntactic-Semantic Links DS, the key-items are linked to the ToC of the MM document.
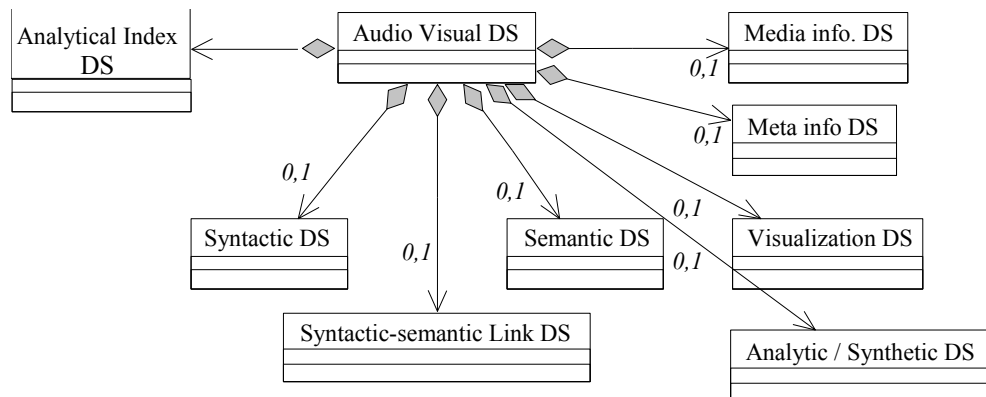

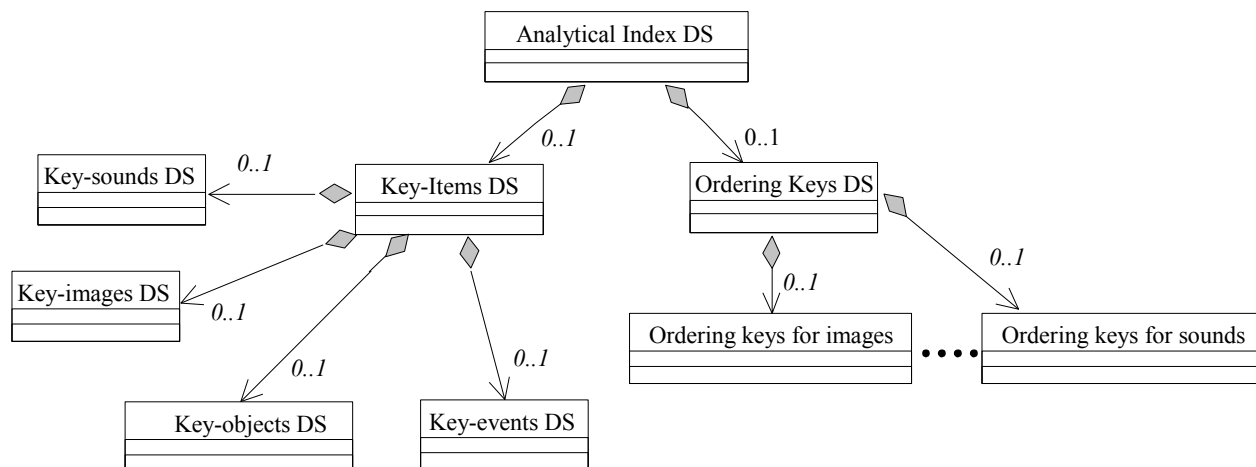**Figure 2: The Generic DS with the proposed integration.**

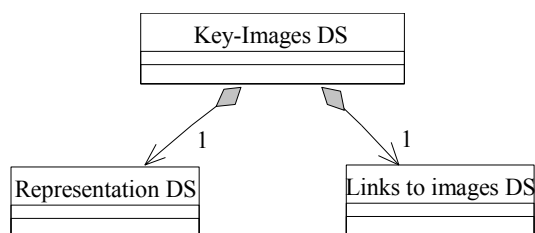**Figure 3: UML representation of the proposed DS**

Figure 4: Key-images DS.

4 Example

In a broadcast news programme description, there can be a key-image linked to all key-frames (or mosaics) associated to shots dealing with a foreign affairs reportage, another one linked to key frames (or mosaics) associated to shots dealing with all sport reportages etc. (see Figure 5). Our example is based on the Portuguese News from CD # of the MPEG-7 Content Set.

5 SUMMARY

In this document, we have presented some open issues linked with the MPEG-7 Generic DS: arbitrarily high number of description items and ordering criteria. Therefore we have proposed two new DSs. The former allows to highlight the more relevant description items (descriptor values) of a MM document. The latter allows to order the description items according to a list of criteria selected according to the application context. The aim of the aforementioned DSs is to facilitate user queries.

6 REFERENCES

[1] AHG on MPEG-7 DS, "MPEG-7 Generic Description Scheme", Proposal to ISO/IEC JTC1/SC29/WG11 MPEG99/M4677, Vancouver, Canada, July 1999.

[2] Video Group, "Generic Visual Description Scheme for MPEG-7", ISO/IEC JTC1/SC29/WG11 MPEG99/N2694, Seoul, Korea, March 1999.
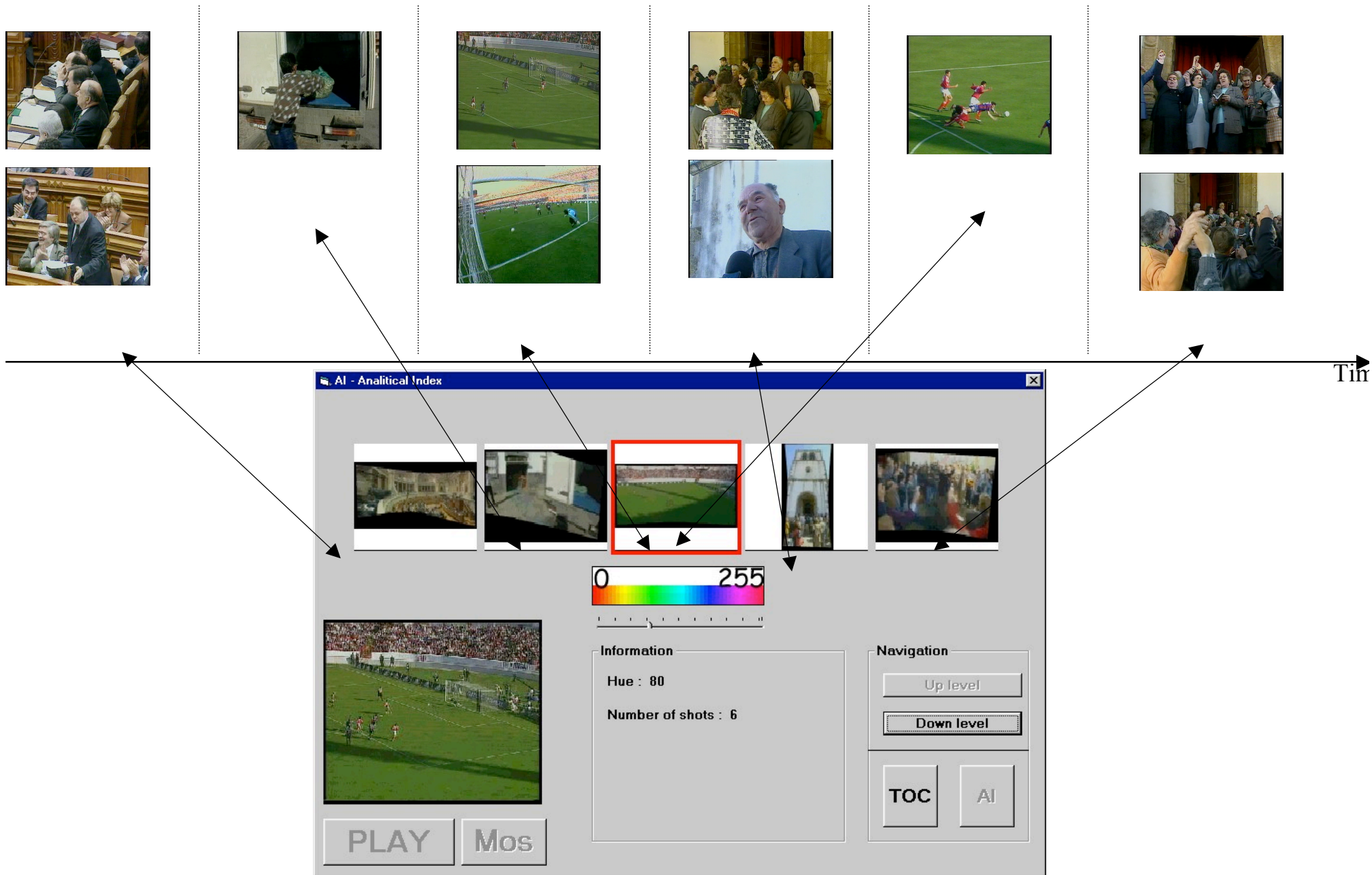
Figure 5: Example of key-images for a set of key-frames of the Portuguese news.