INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE NORMALISATION
ISO/IEC JTC 1/SC 29/WG 11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC 1/SC 29/WG 11
MPEG2002/M9024
Shanghai, China
October, 2002

Source:	**University of Brescia**
Title:	**CE on Metadata Adaptation – Integration of multiple descriptions**
Editors:	**Nicola Adami, Marzia Corvaglia, Riccardo Leonardi**
Status:	**Draft**

Abstract

In this document, a sub part of the CE on Adaptation Hint DS defined in document [4] is presented. Four partners are involved in this CE: Mitsubishi Electric Corp., Fujitsu, Denso Corporation and University of Brescia. While the first three are focusing in the use of AdaptationHint DS for the adaptation of a single DS (see document m8962), the objective of University of Brescia is to show how integration of multiple descriptions of a given content can be useful for the reduction of the number and dimension of metadata and DI. In this context the AdaptationHint DS has been tested to support the integration phase, concluding that it can be useful in order to speed up the process and to reduce the memory occupancy. Additional comment about metadata integration have been included in this report, considering also the use of two new adaptation tools. The first tool is related to the problem of to establish which is elementary unit that has to be used in the comparison of different description while the second aims to support the preferences of a user about metadata transcoding.

In the following, an introduction to the metadata integration problem will be given in order to point out the problems related to it. For the same reason a use case is presented before the description of the experiment setup and results. At the end, final conclusions and remarks are included.

Introduction

The Consumer Electronic (CE) devices, nowadays available, are characterized by increasing storage capacity, processing power and interconnection/communication capabilities. An example is given by some set-top boxes that allow the user to visualize DVB stream from satellite or cable channels, to memorize the content in a hard disk or DVD and to exchange data through an internet connection. The new challenge is now to enable such devices to interact building a multimedia devices home network. This will allow the user to virtually use services and access the content present in any point of the system. Having in mind that we are dealing with CE users, flexibility and ease of use are two key objectives that have to be assumed. To spend hours to plug into the network and configure a new device is not an appealing feature. Similarly, one would access a particular content, stored for example in a Digital Video Camera, from a laptop connected to the network but not directly to the DVC.

Other important features are related to the possibility to use new added-value services, given by the use of content descriptions. It is well known that to have low/high level description of a content helps in speeding up the time required to browse and understand the content, to organize it and to retrieve relevant information from a large distributed or centralized repository.

Objectives

The objective of the project we are involved in is to find innovative solutions for the movement, organization and retrieval of information in such a heterogeneous home system. Three major technical issues are to be solved:

1. New Meta-data computing methods are needed to support advanced retrieval methods. This means, we need to solve how to generate meta-data by analyzing the content, how to combine meta-data from various sources and how to transform meta-data for use by different devices.
2. We will realize new services providing meta-data, applications and UIs to make retrieval of information easier for non-IT-expert users.
3. Standards for inter-storage communication need to be extended in the area of handheld devices, meta-data storage and services.

Description of work

The project consists of three phases. During the first phase usage scenarios will be defined and available network technology will be evaluated. The work performed in the second stage includes the following important tasks:

- Communication between handheld device and stationary device will be set up.
- Finding solutions for distributed information and defining APIs for accessing information.
- Algorithms for the generation, combination and transformation of meta-data will be developed.

- Results of content analysis will be used for development of applications showing visualization, summarization and retrieval functionality.
- User interface designs will be defined, and validated by user testing.
- Applications targeted at services that enrich TV-content by means of meta-data will be developed.

The last phase deals with implementation of a final demonstrator. It will show movement, organization and retrieval of information in a heterogeneous home system consisting, at least, of three different stationary devices and handheld devices. The final demonstrator will be targeted at presentation at a relevant conference or tradeshow.

A use case

In Figure 1, a simple but enough complex system, that will be used as an example, is depicted. As it can be seen, the multimedia network is composed by a handheld and a Personal Computer communicating by mean of a wireless connection.

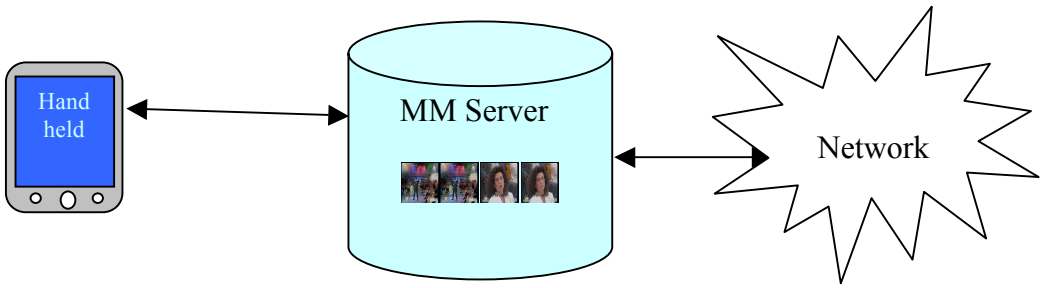

Figure 1 A simple MultiMedia sub-network composed by a handheld and a content/descriptions server.

Assuming the PC is hosting a multimedia content/descriptions repository, we want to access this information from the handheld. More specifically, the objective is to understand the content of a specific document, like a video sequence, using the associated MPEG-7 descriptions.

As originally proposed in the Universal Multimedia Access project, the access to any multimedia resource from any terminal and network has to be easy and transparent to the User. In the use case here described, we focus the attention on the word “easy”. This means that, due to the limited visualization capabilities of the handheld, some adaptation of the content/descriptions has to be performed.

If, for example, the user wants to explore the content by using the k-frames associated to the shots of the sequences, at least we need to adapt the k-frames size and color unitization in order to fit the display characteristics. However, in general more than one kind of description can be available for a given sequence, for example, obtained by using different extraction methods to create a different temporal segmentation of the AV document. In this case, the user has to select which part of the description he/she wants to use. However the user may not have enough information to do the best choice or he/she does not want to choose. Consequently, one may require to adapt the descriptions in order to obtain a unique one, generally, richer than the initial ones and lighter than the description obtained by simple aggregation. For example, if several k-frame series, relative to a specific video, are present in the network a unique description can be generated by using the technique described in [2].

Assuming now that we are referring to the above mentioned MPEG-7 descriptions by means of Digital Items it would be useful to have mechanism to support their integration.

CE description

Considering the above mentioned scenario, we will now define a specific case where integration can be useful.
Our user would like to watch the goals and the replay of goals of the soccer game Spain-Sweden. He is lying down on his bed and he does not want to go downstairs and use the PC to browse the video. However, he is holding a handheld connected to the domestic network that, through a gateway, can provide an access to Internet. So he decide to do a query using a search engine in order to retrieve MPEG-7 description of the soccer game, in which he is interested. After a query across the multimedia network, four MPEG-7 descriptions of the considered soccer game are retrieved. In this situation, four description should be downloaded and browsed by the user that, in principle, has no a priori knowledge of the overall metadata content. As mentioned in the previous section, before to send the retrieved information to the user, it can be useful to process them in order to achieve the following objective:

- Reduce the metadata memory occupation in order to better fit network and user terminal characteristics.
- Reduce, whenever it is possible, the metadata information redundancy.

This operation is relevant especially in the considered case where the user terminal has reduced capabilities.

In the following simulated experiment it will be shown how these objectives can be supported by using AdaptationHint DS as metadata adaptation tool.

Description of Metadata

A video sequence, concerning the soccer game Spain-Sweden from the MPEG-7 Content Set (CD#18), has been used in the experiments. The sequence has been temporally segmented and a VideoSegment DS instance has been generated on the basis of the decomposition information. For each shot, camera motion (pan) and audio volume values have been extracted in order to help the instantiation of three descriptions regarding salient event such as “goal” and “goal replay”.
The events have been detected using three different methods:

1. manual extraction of the “goal” and “goal replay”;
2. algorithm based on audio loudness: the first ten shots characterized by the higher audio loudness are labeled as “goal” (Ordering Key concept [1]);
3. algorithm based on audio loudness and pan (camera motion): the first ten shots characterized by the higher loudness and pan are labeled as “goal”.

In the simulation three Event DSs instances have been considered, labeled as E1,E2,E3. While the segment decomposition information is common to all the three descriptions, the positions where events have been detected are different depending on the used algorithm. The resulting events are:

- E1 (description n° 1) detects: “goal” at shots #10, #38 and “replay of a goal” #16, #44, #45, #46, #53.
- E2: “goal” at shots #1, #2, #9, #10, #12, #16, #25, #32, #39, #43.
- E3: “goal” al shots #14, #31, #35, #39, #43, #54, #58, #61, #62, #68.

All the descriptions are composed by a list of events as shown in Example 1. An event has been associated to each shot where an id equal to: “goal#N” (N is the shot number) has been used to identify a goal, “replay#N” a replay of a goal and “not-classified#N” for a shot with a not classified content.

```
  ..
<Event id="goal#02">
     <Label>
         <Name> Goal n.2 </Name>
     </Label>
     <SemanticTime id="shot#02">
         <Time>
             <TimePoint>T00:07:48</TimePoint>
             <Duration>PT0M9S</Duration>
         </Time>
     </SemanticTime>
 </Event>

 <Event id="not-classified#03">
     <SemanticTime id="shot#03">
         <Time>
             <TimePoint>T00:07:57</TimePoint>
             <Duration>PT0M15S</Duration>
         </Time>
     </SemanticTime>
 </Event>
   ..
```

Example 1 : Structure of E1, E2 and E3.

Metadata Integration

The integration of E1, E2 and E3 has been realized in two different way considering the “intersection” and the “union” of the events detected. In both cases, two kind of information have been used the “Event id” and “Semantic Time id”.

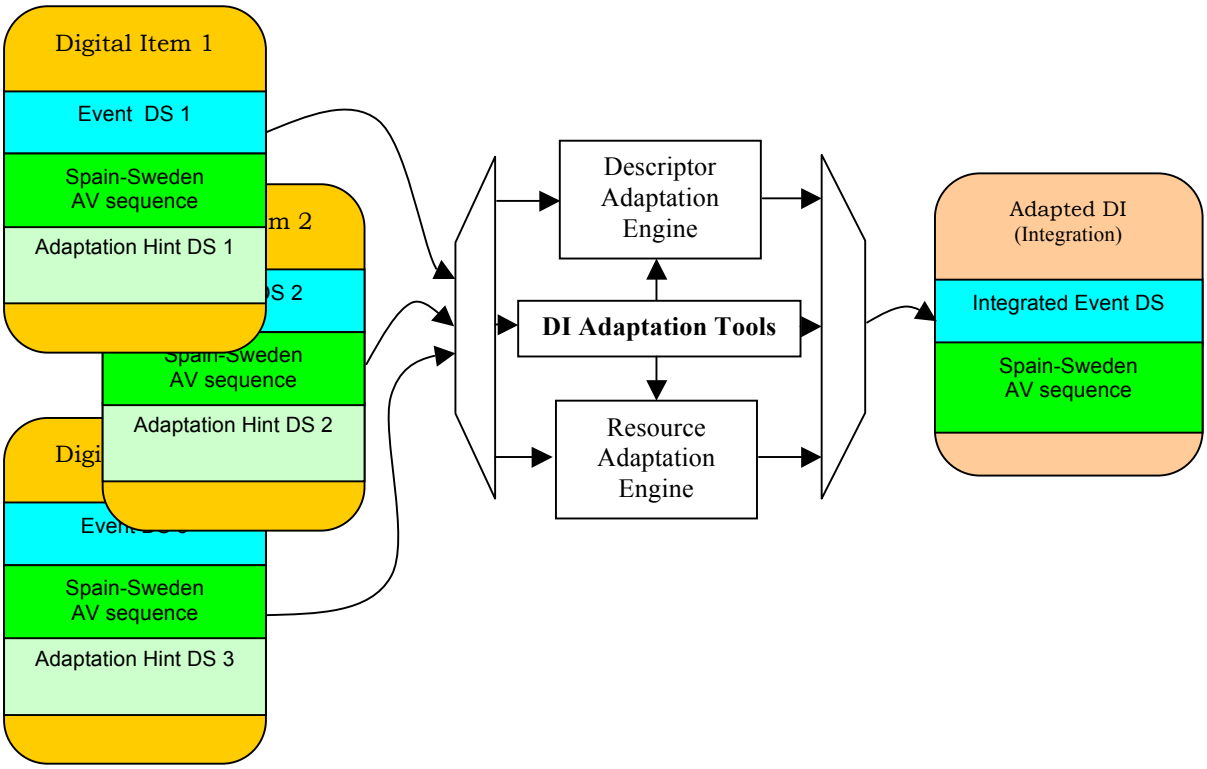

Figure 2 Metadata scaling and Integration by using AdaptationHint DS and EventIntegration DS.

Event Intersection

In this modality events are integrated separately according to the Event id "goal" or "replay". The following rules have been set for the integration:

- In order to do not loose information, for a given description an event with specific id is intersected only with elements of the same type in the other descriptions. This means that if an event with id="my_id.." is present only in one description nothing is done and the intersection output is the list of events with id="my_id.." coming from the initially considered description.
- Event of the same type are intersected considering a temporal window of three shot.

Applying the above rules we obtained the following results:

- "goal" at shot #39.
- "goal replay" at shot #16, #44, #45, #46 and #53.

Union method

The union method considers the common and not common events ("goal" and "goal replay") given by the three algorithm.

- "goal" at shot #1, #2, #9, #10, #12, #14, #16, #25, #31, #32, #35, #38, #39, #43, #54, #58, #61, #62 and #68
- "goal replay" at shot #16, #44, #45, #46 and #53

It has to be mentioned that this is one of the possible choice about integration methods that can be implemented and more studies has to be conduced in this area [1].

Results evaluation

In Table 1, the comparison between size of descriptions before and after the integration process is presented. It can be noticed how integration can simultaneously reduce the number of description instances (consequently the number of DI) and the total memory occupation.

Adapted size (byte) **(E1+E2+E3)**	**Integration** (byte) **Union**	**Integration** (byte) **Intersection**
63819	22604	21200

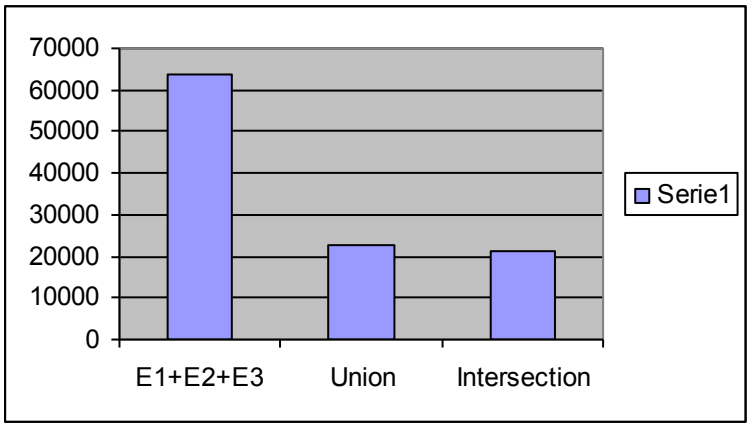

Table 1 Memory occupation comparison(byte): before and after integration.

From the user side, the most evident benefit is the possibility to browse only one segment decomposition instead of three with an immediate access to all the information he is interested in.

```
<AdaptationHint instanceSchema=.......>
  <instanceFileSize> 19920 </instanceFileSize>
  <totalNumOfElements> 78 </totalNumOfElements>
  <Component name="EventDS" number=7>
     <Location type="listOfID"> goal#1, goal#2, replay#1a, replay#2a,
                                 replay#2b, replay#2c, replay#2d
     </Location>
  </Component>
</AdaptationHint>
```

Example 2: AdaptationHint DS for Event DS 1 (E1).

```
<AdaptationHint instanceSchema=.......>
  <instanceFileSize> XXXX </instanceFileSize>
  <totalNumOfElements> 78 </totalNumOfElements>
  <Component name="EventDS" number=10>
     <Location type="listOfID"> goal#1, goal#2, goal#9, goal#10, goal#12,
                                goal#16, goal#25, goal#32, goal#39, goal#43
     </Location>
  </Component>
</AdaptationHint>
```

Example 3: AdaptationHint DS for Event DS 2 (E2).

For what concern the use of Adaptation Hint DS we can say that it is useful in order to speed up the integration process if the "intersection" modality. Looking at Adaptation Hint DSs associated to the descriptions see Example 2 and Example 3 it can be noticed that it is immediate to identify that no "replay" events are in E2 and E3. This eliminate the need to perform the intersection step for this type of event.
Even if Adaptation Hint DS can be useful to speed up description integration more adaptation tools can be identified in order to support a metadata adaptation engine as explained in the following sections.

Proposal for Integration Tools

One of the main problem for the integration of metadata concern the elementary unit that has to be considered during the process. In the experiment above described we considered the integration of entity with the same granularity (an event associated to a shot) . What happen if, for example, we want to compare the two Video Segment instances, associated to the same video, described in Figure 3. As can be seen each of them have a different number of levels and it is not trivial for an adaptation engine to automatically establish which are the elements that have to be compared during the integration process.
In this case the to assign to each level the ratio given by the number of segment over the total length of the video can help to decide which level have a similar granularity.

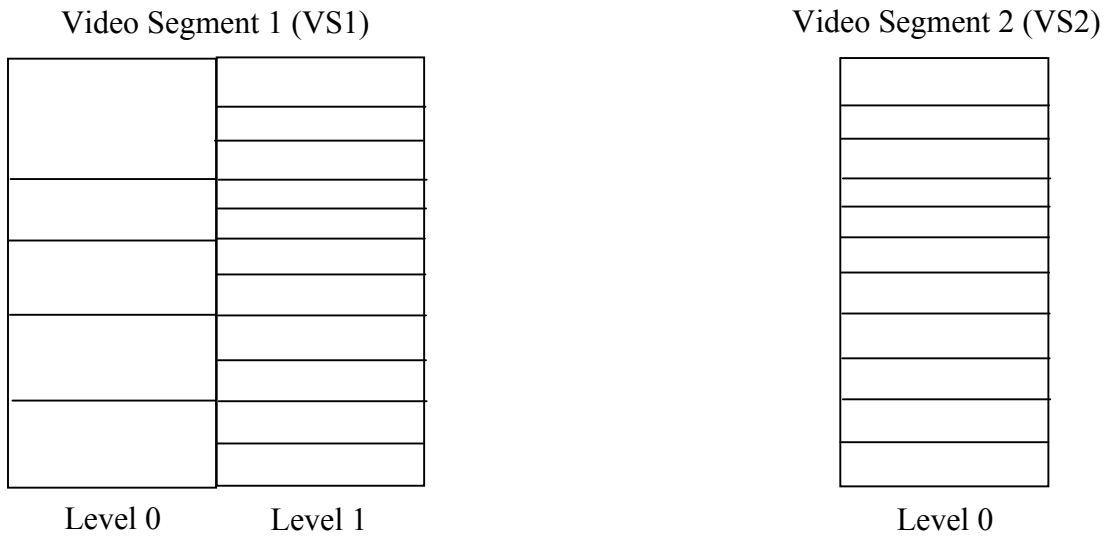

Figure 3: Segment comparison; segments of VS2 have to be compared with segment at level 1 of VS1.

Proposal for Metadata Adaptation Preferences

In the previous sections it has been shown how metadata integration can be supported by the use of Adaptation Hint DS. While audiovisual content transcoding it is sometimes necessary to fit for example the size of the user display, metadata adaptation can be an option. Considering for example a video tree structured video segment decomposition it is possible that a user does not want to get a scaled version, even if he has limited display capabilities. In this case to scale the Video Segment according to a depth constrain can be a waste of adaptation resources (memory occupation and scaling time) and also it can make the end User not satisfied of what he got.

Assuming that operation involving metadata adaptation such as scaling, integration, etc. are relevant in the above described context, adaptation tools have to be defined in order to establish if the metadata transcoding has to be performed or not according to the specific User preferences.
Here after it is proposed a preliminary definition of an Adaptation tool useful to determine User preferences about metadata adaptation. Whenever an adaptation engine can support functionalities as metadata scaling, integration, etc. it can be established if this time consuming operations have to be performed or not.

```
<!-- ##################################################-->
<!--  Definition of UserCharacteristics              -->
<!-- ##################################################-->
<complexType name="UserCharacteristicsType">
      <choice>
       <element ref="Reference"/>
         <sequence>
          ..
        <element name="MetaAdaptationPreferences"
                 type="dia: ContentAdaptationPreferencesType"
                 minOccurs="0" maxOccurs="1"/>
              ..
            </sequence>
      </choice>
      <attribute name="id" type="ID" use="optional"/>
</complexType>
```

New UserCharacteristics semantics

Semantics of the UserCharacteristicsType:

Name	*Definition*
UserCharacteristicsType	Tool for describing characteristics of a User.
Reference	References an external description of User characteristics. Either this reference is used as a description, or the User characteristics are specified inline according to the specified elements.
User	Describes general characteristics of a User such as name and contact information. A User can be a person, a group of persons, or an organization.
ContentPreferences	Describes preferences of an End User related to the type and content of Digital Items.
PresentationPreferences	Describes preferences of an End User related to the presentation of Digital Items.
Accessibility	Describes accessibility-related characteristics of a User.
MetaAdaptationPreferences	Describes Preferences of a User concerning metadata adaptation such as: scaling integration ..

MetaAdaptationPreferences

Introduction

This subclause specifies preferences of a end User about the metadata adaptation operation that can be performed by a metadata adaptation engine.

MetaAdaptationPreferences

```
<!-- ##################################################-->
<!--  Definition of MetaAdaptationPreferences                    -->
<!-- ##################################################-->

<complexType name=" MetaAdaptationPreferencesType">
   <element name="Scaling" type="boolean"/>
   <element name="Integration" type="boolean"/>
    ..
</complexType>
```

MetaAdaptationPreferences semantics

Semantics of the MetaAdaptationPreferencesType:

Name	*Definition*
MetaAdaptationPreferencesType	Tools that describes preferences about metadata adaptation
Scaling	Allows scaling of a description
Integration	Allows Integration of the content descriptions of a given document

Consideration and conclusion

The main object of this part of the CE has been the evaluation of the benefit, in terms of information redundancy reduction that can be obtained considering the integration of the metadata concerning a specific multimedia document. In this case, the AdaptationHint DS can still help the integration process because it immediately identifies if an integration algorithm has to be applied or information have just to be propagated in the final description (intersection).
In addition two new series of adaptation tools have been mentioned in the document to support user preferences about metadata transcoding and to support metadata integration.

References

[1] N. Adami, A. Bugatti, R. Leonardi, P. Migliorati, & L.A. Rossi: "The ToCAI Description Scheme for Indexing and Retrieval of Multimedia Documents", *Multimedia Tools and Application*, Vol. 14 N. 2, 153-173, Kluwer Academic Press, 2001.

[2] N. Adami, M. Corvaglia and R. Leonardi, "Comparing descriptions of multimedia data for simplification and integration", *In Proc. of IPMU 2002*, Annecy, France, July 2002.

[3] H.Nishikawa, et al, "Description for Metadata Adaptation Hint", ISO/IEC JTC1/SC29/WG11/M8324, Fairfax, USA, May 2002.

[4] MDS: "Workplan for CE on Metadata Adaptation", ISO/IEC JTC1/SC29/WG11 N4953, Klagenfurt, July 2002