Using Context to Assist the Adaptation of Protected Multimedia Content in Virtual Collaboration Applications

Maria Teresa Andrade
INESC Porto, Rua Dr. Roberto Frias 378, 4200-465 Porto, Portugal
maria.andrade@inescporto.pt

Hemantha Kodikara Arachchi, Sabih Nasir, Safak Dogan, Halil Uzuner, Ahmet M. Kondoz
I-Lab, CCSR, University of Surrey, Guildford, UK
{H.Kodikaraarachchi, S.Nasir, S.Dogan, H.Uzuner, A.Kondoz}@surrey.ac.uk

Jaime Delgado, Eva Rodríguez, Anna Carreras
Universitat Politècnica de Catalunya, Barcelona, Spain
{jaime.delgado, evar, annac}@ac.upc.edu

Tim Masterton, Rachel Craddock
Thales Research and Technology (UK) Ltd., Reading, UK
{Tim.Masterton, Rachel.Craddock}@thalesgroup.com

Abstract**—This paper proposes a framework for a virtual classroom application based on a Virtual Collaboration System (VCS), which is being developed under the VISNET II Network of Excellence (NoE)[1], and discusses adaptation technologies that enable seamless access to classroom sessions while intellectual property and digital rights are managed. The proposed virtual classroom framework enables academic institutions to conduct their collaborative lecture series, to which registered students will be able to attend remotely and interactively over the Internet. Furthermore, the general public may also follow the classroom sessions under certain restrictions imposed by the participating institutions. In order to facilitate seamless access to a heterogeneous audience that is composed of users with various preferences and privileges accessing the classroom sessions over different network infrastructures and using terminal devices with diverse capabilities, context-aware content adaptation is required to meet constraints imposed by the usage context and enhance the quality of the user experience. Thus, this paper describes the concepts and functionalities of a context-aware content adaptation platform that suits the requirements of such multimedia application scenarios. This platform is able to consume low-level contextual information to infer higher-level contexts, and thus decide the need for and type of adaptation operations to be performed upon the content. In this way, it is aimed to meet usage constraints while also satisfying restrictions imposed by the Digital Rights Management (DRM) to govern the use of protected content.**

Keywords-Virtual Collaboration; Virtual Classroom; Digital Rights Management; Digital Item Adaptation; MPEG-21

I. INTRODUCTION

Over many decades, academic institutions around the world have offered remote learning facilities for those who are unable to attend regular lectures. These institutions are turning to state-of-the-art multimedia technologies to offer improved learning experiences for those external students. Currently, video-conferencing allows limited two-way face to face interaction between two groups of tutors or students, and streaming media distribution allows one-way presentation of audio-visual material to a remote audience. However, video-conferencing technologies have serious limitations in terms of the richness of interactivity between the parties involved and one-way presentation is not conducive for clarifying requests or debates. Current systems are also inefficient in terms of their use of network resources, in that they take no account of the content or activity present on the broadband audio and video channels required to provide high quality presentation material to a large audience.

Within the context of this paper, we propose a framework for a virtual classroom application using the components of a Virtual Collaboration System (VCS) with context extraction from the media streams and adaptation of the delivered content. We also propose the use of rights management on the content to allow controlled dissemination to a heterogeneous audience. In combination, these features will enable academic institutions to conduct a series of collaborative lectures and classes with which remotely located students can interact more efficiently. It is envisaged that this infrastructure will enhance the traditional advantages of remote learning, namely:

- Experienced teaching staff can be shared more widely
- Both staff and students can spend less time and travelling expenses
- Scheduling constraints are eased
- Debate and discussion quality is improved

by improving their efficiency. In addition, there will be new benefits:

- More interaction options are available for students.

---

[1] VISNET II is a European Commission funded project under the IST FP6 programme (http://www.visnet-noe.org/).

- There is the possibility of having one or more document tracks: one for the presentation and another for interactivity.
- Small, securely contained, sub-groups can be convened on a single system.
- Quality of the presented material can be tailored to the content at a finer granularity.

Success of the proposed application largely depends on the users' ability to comprehend the materials delivered over the Virtual Collaboration (VC) platform. Ideally, the remote audience should have the same comfort as the local audience in terms of listening to the speaker (lecturer or anyone from the audience) and viewing the speaker's expressions, gestures, presentation materials and the white board. Even though the audience in a remote lecture theatre with full VCS functionality may be able to experience this sensation, the challenge is to also provide it to the individuals accessing through terminal devices with limited capabilities. Therefore, in order to facilitate seamless access of the heterogeneous audience with various preferences and privileges over various network infrastructures using a vast range of terminal devices with various levels of capability, context-aware content adaptation is a key technology needed for the proposed virtual classroom application.

The use of contextual information is instrumental for the successful implementation of useful and meaningful content adaptation operations. These, in turn, are becoming extremely important for the implementation of systems and platforms that deliver multimedia services to the end-user. The continuing advances in technology are steadily emphasizing the great heterogeneity that exists in devices, systems, services and applications today. Likewise, those advances will also bring out the desires of consumers for more choices, better quality and more personalisation options. Implementation of meaningful content adaptation operations that meet users' expectations and satisfy their usage environment constraints requires the use of contextual information to take decisions on how and when to adapt the content. Figure 1 illustrates the present scenario of heterogeneity in the multimedia content value chain, highlighting the role of context-aware content mediation systems.

Adaptation is an operation to be performed upon the content. Accordingly, and as long as the content is governed or protected, any kind of content adaptation operation should also be subjected to rules and rights. Adaptation operations can thus be subjected to restrictions from the view point of the content owner's rights, *i.e.*, content adaptation can be governed, in addition to the constraints imposed by terminals, networks, natural environments and users. This framework thus brings out a new concept, the one of adaptation authorisations, which can be seen as a new form of contextual information.

This paper presents intermediate results obtained within the framework of the VISNET II NoE. It describes the concepts and functionalities of a context-aware content adaptation platform that suits the requirements of multimedia application scenarios where adaptation of content to meet constraints imposed by the usage context would be required to enhance the quality of the user experience. This platform is able to consume low-level contextual information to infer higher-level contexts to assist adaptation decision operations. Inferring higher-level concepts provides the system with a better understanding of the real-world situations. In addition, the system processes DRM information to obtain the authorisation for applying the desired adaptation. Accordingly, the decision on the need and type of the adaptation operations to endorse not only meets the usage environment constraints, but also satisfies the restrictions imposed by the system governing the use of protected content.

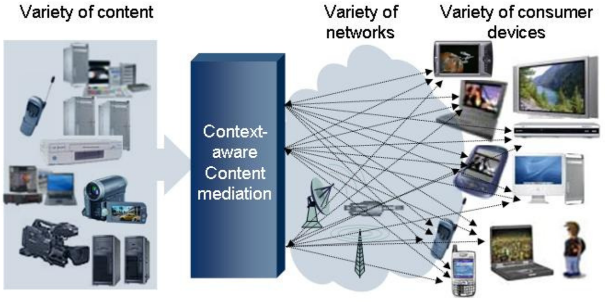

Figure 1. The present heterogeneous contnt generation, access and consumption scenario

The paper is organised as follows: Section II describes the virtual classroom application scenario that is used as a driver for the definition of potential use cases for context-based adaptation issues addressed in this work. Section III provides a brief overview of the state-of-the-art in context-aware content adaptation. Aspects of DRM related with the adaptation of content are analysed in Section IV. The defined system architecture is presented in Section V. Finally, Section VI draws the conclusions and outlines the future work.

II. VIRTUAL CLASSROOM

A VCS allows remotely located partners to meet in a virtual environment using state-of-the-art communication and audio-visual technologies. All remotely located partners should feel the sensation of being in a single room regardless of their true geographical location. The proposed virtual classroom can be treated as an application scenario of the VCS, as illustrated with a context diagram in Figure 2.

In order to realise the proposed application, the lecture theatres of each participating academic institution will be equipped with the necessary VC infrastructure, which includes:

(i) Three sets of projectors and screens for: (a) displaying the presentation, (b) the virtual whiteboard that can be used by the lecturer as well as the audience (local and/or remote), and (c) video feeds of the lecturer and the audience.
(ii) Set of cameras to capture the lecturer.
(iii) A microphone and a video camera fitted to each seat to capture the occupant. These would ultimately be replaced by a smaller number of optical pan-tilt-zoom cameras and a steerable beam microphone.
(iv) An input device (possibly a tablet) fitted to each seat, so that the occupant can write/draw on the virtual whiteboard and alert the audience to his willingness to add his view.

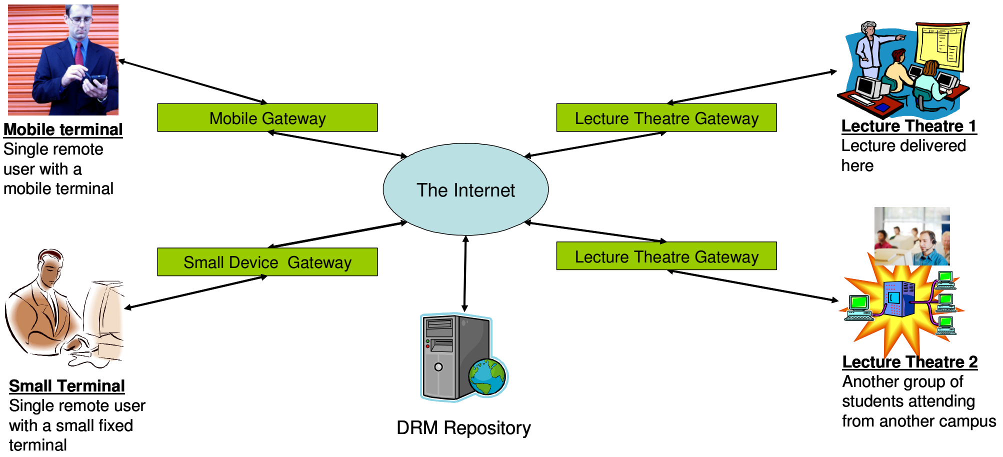

Figure 2. Virtual classroom context diagram

When one institute conducts the lecture in one of their lecture theatres for the enrolled students, those from other institutions who have also enrolled to the same course can attend the same lecture remotely being in another lecture theatre. Unlike a conventional classroom session, these lecture series can also be followed by external students, those who have been unable to be in the classroom as well as the general public over a wired or wireless link using their home PC or a mobile terminal, such as a smart phone. Enrolled remote students can then interact with the lecturer and audiences using the VC platform that facilitates interaction not only through audio and video, but also through a virtual white board. Any Internet-enabled device on which the VC platform is installed can be used as the terminal to enjoy the full functionality of the system. However, devices on which the VC platform cannot be installed, but offer video communication functionality, may be able to present the document tracks also as a video. The video is adapted to match individuals' preferences while also considering other related constraints. However, the general public have rights to view only a low resolution version of the video and do not have the privileges to view any of the adapted versions. Neither can they interact with the classroom sessions.

This raises the prospect of a whole new concept of searching the Internet for scheduled live seminars on a topic of the user's choice, just as seminar notes can be found online today. Live seminars are already possible, but will only achieve mass uptake when users become accustomed to seamless audio and video access, because all the devices they use regularly can provide it (at some level of adaptation).

Technical challenges associated with this scenario are:

(i) Integration and synchronization of audio feeds of individual students and lecture theatres.
(ii) Managing, synchronizing and presenting interactions (audio, visual and white board) from the distributed audience.
(iii) Automatically tracking the movements of the presenter.
(iv) Customizing the presentation.
(v) Detecting requests for interaction from the audience and adapting their audio and video content preferably without the expense of individual microphones and cameras at each audience seat.
(vi) Acquiring low-level context and inferring higher-level concepts. Inferring the state of the user, such as location, activity, intentions, resources, and capabilities in the past, present, and future.
(vii) Adapting content (audio, video and document tracks) for small terminals with limited capabilities and to specific context situations.

Within the context of this paper, we discuss the underlying adaptation technologies that enable seamless access to the classroom while managing the participating institutions' rights on the contents. For this purpose, the aforementioned VC system needs to be equipped with a context-aware content adaptation subsystem. This subsystem is implemented within the gateways shown in Figure 2. The following sections describe this subsystem in further detail.

III. STATE-OF-THE-ART IN CONTEXT-AWARE CONTENT ADAPTATION

Most of the earlier research in context-awareness has been conducted within the mobile services application area [1]. The reported work mostly concentrated on exploring context to improve usability aspects by sensing how the available devices were being used. Generally, the proposed systems reacted directly to the sensed low-level context. They usually lacked flexibility as they neither explored the inter-relations among different types of low-level contextual information nor made use of ontologies to infer higher-level contexts. They did not sufficiently address the aspects of interoperability, scalability and security/privacy either. In fact, earlier research was typically application-centric, overlooking aspects of gathering different types of contextual information from different spaces and interoperability in heterogeneous environments. Likewise, the aspects concerning security of content and context metadata, and ensuring the privacy of the user have only recently started to be addressed. The new generation of projects are now focusing on these aspects relying on the use of standard ontology specifications, context representation frameworks and middleware layers [2]. The goal is to be able to infer more complex implicit contexts using ontologies taking interoperability issues into consideration. Accordingly, current research is also looking at the aspects of context representation through the development of formal and standardised frameworks. Common ontologies are also a vehicle to enable interoperability from the semantic point of view.

At the standardisation level, one of the most complete standards is the MPEG-21 Part 7 (Digital Item Adaptation – DIA) specification [3]. It is a part of the MPEG-21 ISO/IEC standard currently under its final phase of development [4]-[6]. MPEG-21 DIA provides a set of tools in the form of XML schemas to describe the characteristics and capabilities of networks, terminals and environments as well as preferences of users. It also provides the definition of the operations that can be performed upon the content and the result that can be expected. Figure 3 provides an illustration of the available tools of MPEG-21 DIA and its use for adaptation purposes.

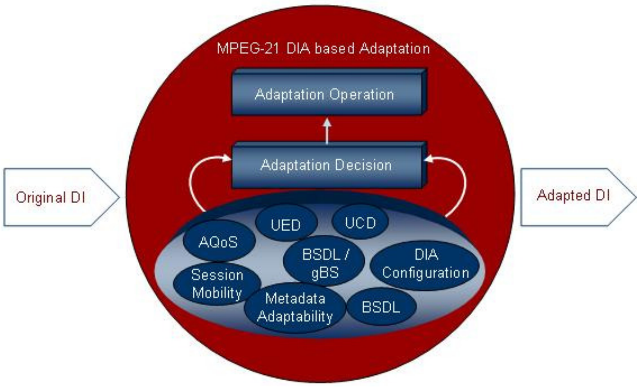

Figure 3. Adaptation within MPEG-21 DIA

The World-Wide-Web Consortium (W3C) has also been quite active in this area with its Composite Capability/Preference Profiles (CC/PP) specification [7], and a set of standards has been developed that are closely related to the Semantic Web paradigm, notably the Resource Description Framework (RDF) [8] and the Web Ontology Language (OWL) [9] specifications.

A. MPEG-21: DRM and Adaptation

Even if content adaptation and DRM are two big lines of research, their work has always been performed separately. The only standardization initiative trying to integrate both research areas together comes from the Moving Picture Experts Group (MPEG), which includes specifications for DRM and content adaptation under one of its standards, the MPEG-21.

Rights Expression Languages (RELs) have been proposed to express conditions of use of digital content. Parts 5 and 6 of MPEG-21 standardize a Rights Expression Language (REL) [10] and a Rights Data Dictionary (RDD) [11] used by providers to protect the distribution and the use of their contents in governed networks. Specifically, Part 5 of the MPEG-21 standard defines the syntax and semantics of a REL. MPEG chose the eXtensible rights Markup Language (XrML) [12] as the basis for the development of the MPEG-21 REL. The most important concept of this REL is the license that grants the permission to an entity to exercise a right against a resource if the conditions specified within the license have been previously satisfied.

Initially, the principals, rights, resources and conditions of the MPEG-21 REL were organized in three main groups. The first one, the Core, specifies structural elements and types and how they are related. The standard extension defines terms to extend the usability of the core schema. It defines conditions to restrict the use of the digital content, for example in the interval of time or the number of times that it can be used, the fees that must be paid, the territory in which it can be used etc. Finally, the multimedia extension expands the core schema by specifying terms that relate to digital works. It describes rights, conditions and metadata for digital works, which includes rights, such as modify, enlarge, reduce, move, adapt, play, print, execute etc.

MPEG-21 REL can be profiled to facilitate the interoperability with other RELs. Currently, two profiles were defined for the MPEG-21 REL, the Mobile And optical Media (MAM) [13] and the Dissemination And Capture (DAC) profile [14]. The MAM profile was defined to facilitate the interoperability with OMA DRM REL v2. In order to support the requirements of this profile, the “multimedia extension one” was defined with new rights and conditions for the pre-recorded optical media and mobile domain. On the other hand, the DAC profile was designed to facilitate the interoperability with the TV-Anytime Rights Management and Protection information [15]. For this profile the “multimedia extension two” was defined with rights and conditions for the broadcast domain. Furthermore, Part 7 of the standard (DIA) provides a set of tools for digital item adaptation as mentioned before, and the first Amendment of DIA [16] focuses on the adaptation authorisation.

B. Adaptation Engines

A Universal Multimedia Access (UMA)-ready content delivery system is not complete without Adaptation Engines (AEs) that carry out the actual adaptation operation. Most of the video AEs that have been discussed in literature address network constraints. Bit rate adaptation of MPEG-4 visual coded bit streams by Frame Dropping (FD), AC DCT Coefficient Dropping (CD) and their combinations (FD-CD) have been discussed in [17]. A utility function has been used to model the video entity, adaptation, resource, utility, and the relations among them. Each video clip is classified into one of a number of distinct categories, and then local regression is used to accurately predict the utility value. Techniques reported in [18]-[21] consider only frame dropping as a means of bit rate adaptation. A more intelligent frame skipping technique has been presented in [22]. This technique determines the best set of frames (key frames) that represents the sequence. The proposed technique utilises a neural network model that is capable of predicting the indices of the most appropriate subsequent key frames. In addition, a combined spatial and temporal technique for multidimensional scalable video adaptation has been discussed in [23].

A framework for video adaptation based on content recomposition is presented in [24]. The objective of the proposed technique is to provide effective small size videos, which emphasize the important aspects of a scene while faithfully retaining the background context. This is achieved by explicitly separating different video objects based on a generic video attention model that extracts objects of interest to the user. Three types of visual attention features, namely intensity, colour, and motion, have been used in the attention model. Subsequently, these objects are integrated with the direct-resized background to optimally match the specific screen sizes.

The aforementioned techniques, however, do not consider user's preferences for determining the nature of adaptation. The technique presented in [25] extracts highlights in a sports video according to the user's preferences. The system is able to extract the highlights, such as shots at goal, free kicks etc. in soccer, and start, arrival, and turning moments of swimmers in swimming scenes. Objects and events are assigned to classes of relevance. The user can assign a degree of preference to each class, in order to have the best quality-cost trade-off for the most relevant classes at the price of a lower quality for the least relevant ones. The adaptation module performs content-based video adaptation according to the bandwidth requirements and the weights of the classes of relevance. Xu *et al.* has considered users' preferences together with the network characteristics for adapting sports videos [26]. Events are detected by audio/video analysis and annotated by the Description Schemes (DSs) provided by MPEG-7 Multimedia Description Schemes (MDSs). Subsequently, users' preferences of events and network characteristics have been considered to adapt video by event selection and frame dropping.

An effective way of performing content adaptation is to utilise adaptation operations in the networks due to advantages, such as transparency and a better network bandwidth resource utilisation. Scalable coding technologies help to simplify the function of the network element that carries out the adaptation operation. A number of scalability options have been discussed in literature namely, spatial, temporal, fidelity, and Interactive Region of Interest (IROI) [27]. If the coded video is featured with one or more of the aforementioned scalability options, the adaptation operation is as simple as letting the set of coding units that define the adapted bit stream through the adaptation engine while discarding the rest.

IV. DRM ISSUES

In the scenario considered in this paper, intellectual property and digital rights are managed during adaptation. In VC environments, two different types of licenses can be used to control the access and usage of protected digital content. The first type of licenses is restrictive licenses that limit the use of digital content, for example establishing the number of times that the content can be rendered or the interval of time in which the content can be played. The second type of licenses used in such a scenario is attribution and non-commercial licenses.

If we consider MPEG-21 technologies to express digital objects, adaptation information etc., then we can choose the MPEG-21 REL for expressing licenses. The standard elements defined in this REL can be used in our scenario to restrict the usage of multimedia content, for example for limiting the number of times that a video can be played. However, currently a new profile for the MPEG-21 REL is under development to support the different types of Creative Commons (CC) [28] licenses. This profile is based on a contribution that we made [29] in the 76th MPEG meeting to facilitate the interoperability with CC licenses. The Open Release Content Profile [30] includes new rights and conditions, such as the *governedAdapt*, embed or *governedAggregate* rights, and the *copyrightNotice* or *nonCommercialUse* conditions. Then, licenses using this profile can express different types of CC licenses which include attribution, non-commercial, no derivatives, share alike etc.

V. SYSTEM ARCHITECTURE OF THE CONTEXT AWARE CONTENT ADAPTATION SUBSYSTEM

As highlighted in Section III, the new generation of context-aware systems is now focusing on the aspects of using multiple sources of sensed or low-level contextual information to infer more complex implicit contexts that will allow them to take wiser or more suitable decisions regarding adaptation. Our research is aligned with these trends while also considering DRM, security and privacy issues as major concerns. Accordingly, we have drawn a common modular system architecture for implementing advanced context-aware services. This is a layered platform that embraces the system interoperability approach proposed by the new generation of systems, and the aspects of combining multiple explicit (sensed or low-level) contexts to build elaborate implicit contexts in an interoperable manner with the ultimate goal of maximizing the quality of the experience of the user by selecting the most adequate type of adaptation. Figure 4 illustrates this layered architecture. The external lower layer can be seen as a middleware layer that abstracts the higher-layers from the actual generation of low-level contextual information. This layer is instrumental to enable interoperability at the system level, as:

- Applications can be developed without needing to be aware of the details of the sensor devices in use, and thus are independent from sensor technology.
- Different applications may use the same sensors, and may make different uses of the low-level sensed information.
- Sensors can be distributed, and thus applications may profit from using explicit contextual information gathered in remote points.

Each layer of this architecture is further divided into functional modules. For example, the low-level context sensing layer incorporates different modules acting as services, offering functionalities to collect different types of low-level context. Likewise, the context reasoning layer will provide various modules that reason about different sets of low-level context. The existence of particular modules is dependent on the type of adaptations offered by the application layer, which in turn is dictated by the application scenario. Accordingly, the generic context-aware platform may present different functionalities based on the application scenario in view. It is still a generic architecture in the sense that it can seamlessly incorporate different functionalities as needed by different applications, whereas a common functionality can be re-used between those applications. The specific application scenario that we are interested in this work involves the gathering of the following set of low-level contextual information:

- Characteristics of the terminal, conditions of the network, user characteristics and interactions.
- Sensed low-level visual and auditory information both related to the user as well as to his/her surrounding environment. This information can be used to reason and conclude on the user's emotional or physical state or in identifying indoor/outdoor situations.
- Security and DRM information (conveyed in licenses).

Accordingly, the functional blocks that compose the VISNET II context-aware architecture for the virtual classroom application are identified in Figure 5.

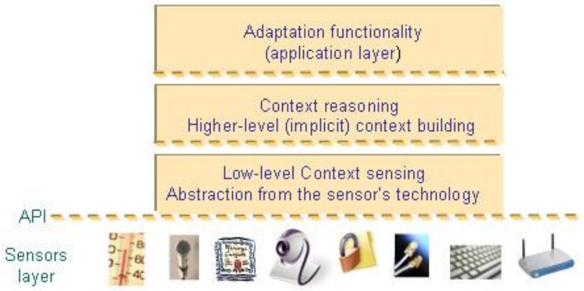

Figure 4. Generic context-aware architecture

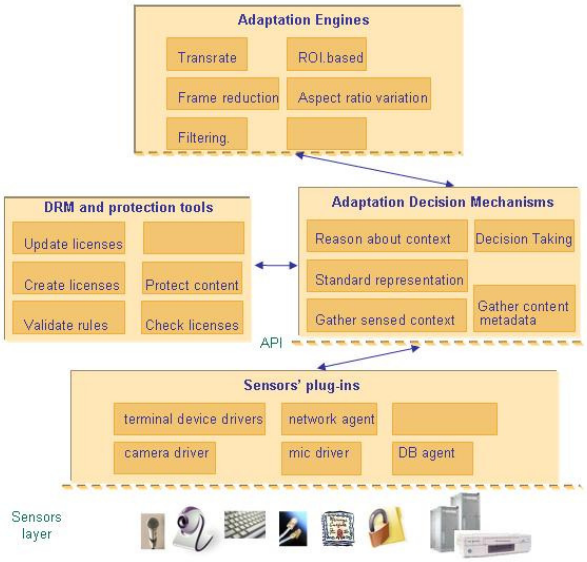

Figure 5. System functional architecture

A. *Context Providers and Profiles*

Contextual information can be any kind of information that characterizes or provides additional information regarding any feature or condition of the complete delivery and consumption environment. This diversity of information can be grouped into four main context classes according to the feature or entity it refers to: *Resources*, *User*, *Physical* and *Time*. There are also two other dimensions of the contextual information that need to be taken into consideration: one referring to the characteristics and nature of the contextual information itself and not to the type of entity that it describes; another one related with the form in which it is obtained. In the former, the aspects to consider include the accuracy, period of validity or inter-dependencies. In the latter, it is usual to consider two classes: low-level, sensed or *explicit context*, which is the contextual information that can be directly acquired by different kinds of sensors; high-level or *implicit context*, which is obtained through reasoning about the low-level context. Entities, either software or hardware, that are able to generate and provide the explicit contextual information are designated as "*Context Providers*". As illustrated in Figure 5 and listed in Section II, in our work, we are interested in a number of different low-level contexts, which means that our system interacts with different *Context Providers*. The information generated by these entities, once acquired and represented according to a standard format, will be used to infer higher-level concepts, and thus assist the adaptation decision operation. More details are provided on the role and use of *Context Providers* in the virtual classroom application in the next subsection. The standardised format used in this work is the MPEG-21 DIA specification. This aspect is considered as instrumental to enable interoperability among systems and applications, and across services. MPEG-21 DIA specifies appropriate XML schemas to represent the low-level contextual information. In particular, the Usage Environment Description (UED) tool of MPEG-21 DIA provides an optimal set of descriptors that can be used to represent any of the four context classes. It is divided into four main blocks, where each block is associated with an entity or concept within the multimedia content chain: User, Terminal, Network and Natural Environment. Based on this division, we have defined profiles. Although the combined use of descriptors from different blocks can offer increased functionality, the identification of profiles inside each group can potentially simplify the use of these standardised descriptions by the different entities involved in the provision of networked context-aware multimedia services, and thus increase its rate of acceptance. One of the resulting advantages will be realized at the level of interoperability among these entities, which operate in distinct domains but need to interact with each other. All of the participating entities should collect and make available useful contextual information. These entities can be designated as *Context Providers*. If they all use the same open format to represent their context, any one entity can use descriptions provided by any other entity. Moreover, the definition of profiles based on the UED block division means that each Context Provider needs only to know and implement his/her own profile to offer contextual information concerning his/her own sphere of action. For example, a Network Provider and a Manufacturer will make available contextual information related to the network dynamics and about the terminal capabilities, respectively. Accordingly, our initial proposal for the definition of profiles is based on the four existing UED groups, each profile being composed of the corresponding DIA elements. This is a quite generic and broad definition of profiles, and can be seen as *Context Provider-centric* approach. Additionally, on top of this initial profiling, further partitioning can be done by adopting an *Application-centric* approach. For the VC application described in Section II, the profiles presented in Figure 6 have been defined.

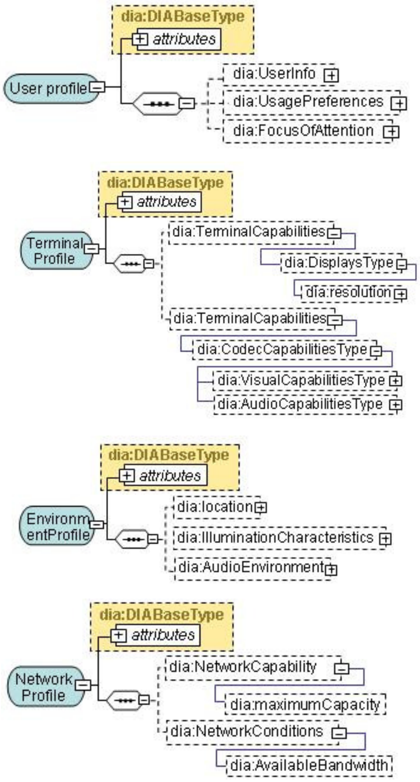

Figure 6. Virtual collaboration context profiles

B. *Adaptation Decision Engine*

The Adaptation Decision Engine (ADE) can be considered as the intelligent part of the content adaptation subsystem in the VC application. Its goal is to take a decision regarding the actions to perform when contextual information is available with the goal of maximizing the quality experienced by the user. The ADE being designed in VISNET II is illustrated in Figure 7. The approach is to have a central coordinator, named *ContextServiceManager*, which interacts with dedicated modules that acquired the sensed low-level context and other required metadata. The former is generated by sensors, either logical, such as software programmes, or physical, located in terminals at the networks or on electronic equipment. The latter will essentially be content-related metadata (media characteristics and MPEG-21 DIA Adaptation Quality of Service – AQoS [3]), and rules for reasoning specific to the application in view. It should be noted that, in our work, we are interested in defining the required low-level context and interoperable forms of representing it, but will not be developing new sensors, except some simple logical sensors to extract capabilities of terminals and network conditions.

These modules are responsible for formatting the acquired context into the identified UED, where appropriate, using the context profiles described in the previous subsection. They work in a "push" model notifying the coordinator when new context is available. The context providers that are required for the virtual classroom application consist of logical and physical sensors. The former are software applications running in the terminal, at the network edge equipment and in Databases (DBs), holding descriptions relative to the content and reasoning rules. The latter are external physical sensors, namely the overall camera and microphone. These context providers pass their information to the ADE lower layer through:

- *Terminal device drivers* to acquire the capabilities and conditions of the user terminal, and also audio and video information captured by the built-in camera and microphone.
- *Network agent* to describe the characteristics and the conditions of networks.
- *Content service agent* in the form of a DB access module to acquire content-related metadata and reasoning rules.
- *Audio and video sensors.*

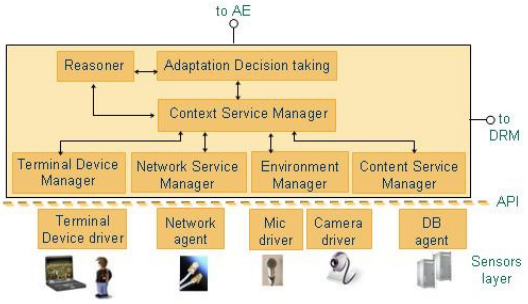

Figure 7. Modular architecture of the ADE

Whenever rules are available, the *Reasoner* is invoked by the *ContextServiceManager* and interacts with the *DecisionTaking* module to select the most appropriate adaptation and corresponding service parameters, maximizing the user quality of experience. The *ContextServiceManager* has also an interface towards the DRM tools subsystem to request information concerning the authorisation of the adaptations.

The big challenge in the design and development of the ADE relates to the *Reasoner*. This module uses the sensed context to infer the state of the user, such as type of user, location or activity he/she is engaged with or degree of satisfaction being experienced, as well as environmental conditions. This will be done through the use of an ontology-based model. We are in the process of developing a two-layer ontology approach using OWL. The basic layer provides descriptions of generic concepts and rules that can be used for any generic application scenario while the second layer provides specific rules for the virtual classroom application scenario. This ontology virtual classroom-specific layer will provide the means for the ADE to reason on how the different possible adaptations would help the user better understand the classroom sessions. This is precisely the big challenge: how to obtain descriptions and sets of relationships in the form of rules that represent as accurately as possible the real-world situations in virtual classroom applications. Accordingly, this will allow the ADE not only to base its decision on the restrictions of the consumption environment, such as terminal screen size, but also on a user satisfaction model consistent with the learning objectives of the virtual classroom application. A two-layer strategy has been proposed in [31] recently. For the implementation of the rule-based inference engine, we are investigating the use of the OWL reasoner supplied with the Jena2 platform [32].

C. Adaptation Authorisation

In Section IV, we have discussed how to protect the use and distribution of the lectures by means of digital rights. Accordingly, adaptation operations should only be performed if they do not violate any condition expressed in the licenses. As adaptation is becoming more and more important in multimedia systems, we have arrived at a point where more detailed descriptions about permissible conversions are need to be able to govern content adaptations.

The first Amendment of MPEG-21 DIA provides the description of fine-grained media conversions by means of the conversion operations' names and parameters, which can be used to define rights expressions to govern adaptation in an interoperable way.

For example, a teacher may want the students to download the lectures with a good resolution exceeding a given minimum, as otherwise they can miss some important details. Figure 8 illustrates an example of such a license that the teacher should issue associated to his lectures.

```
<r:license>
 <r:grant>
  <r:keyHolder licensePartIdRef="student"/>
  <mx:play/>   <-- the student can play the lecture-->
  <mx:diReference licensePartIdRef="lecture"/>
  <r:allConditions>    <-- under these conditions -->
   <dia:permittedDiaChanges> <--Adaptation of the Resolution -->
    <dia:ConversionDescription xsi:type="dia:ConversionUriType">
     <dia:ConversionActUri uri=”urn:mpeg:mpeg21:2003:01-RDD-
NS:CropRectangularBitmapImage”/>
    </dia:ConversionDescription>
   </dia:permittedDiaChanges>
<--constraints apply whether there is adaptation or not-->
   <dia:changeConstraint>
    <dia:constraint>
     <dia:AdaptationUnitConstraints>
      <dia:LimitConstraint> <--min lim horizontal resolution -->
       <dia:Argument xsi:type="dia:SemanticalRefType"
semantics="urn:mpeg:mpeg21:2003:01-DIA-MediaInformationCS-NS:17"/>
       <dia:Argument xsi:type="dia:ConstantDataType">
        <dia:Constant xsi:type="dia:IntegerType">
         <dia:Value>min_value</dia:Value>
        </dia:Constant>
       </dia:Argument>
       <dia:Operation operator="urn:mpeg:mpeg21:2003:01-DIA-
StackFunctionOperatorCS-NS:13"/> <-- 13 refers to operator “>” -->
      </dia:LimitConstraint>
     </dia:AdaptationUnitConstraints/>
    </dia:constraint>
   </dia:changeConstraint>
  </r:allConditions>
 </r:grant>
 <r:issuer>
  <r:keyHolder licensePartIdRef="teacher"/>
 </r:issuer>
</r:license>
```

Figure 8. License expressing conditions upon the video resolution

D. Adaptation Engines

Three types of media content are involved in the virtual classroom scenario described in Section II, namely: (i) audio, (ii) video and (iii) documents. Adaptation of all media types may be necessary to present most relevant information while satisfying the usage environment constraints. Nevertheless, within the context of this paper, we mainly concentrate on video adaptation. The function of an AE is to perform the adaptation operation recommended by the ADE described in Section V.B. For its operation, the following content adaptations have been identified, including: (i) spatial scaling, (ii) temporal scaling, (iii) bit rate (transrate) adaptation, (iv) error resilience adaptation, (v) summarization, (vi) cropping, and (vii) document-to-video conversion.

A Scalable Video Coding (SVC) approach is feasible for most of the adaptation operations required in the virtual classroom application. However, the applicability of transcoding cannot be played down as an option for some of the operations, such as cropping, summarization and error resilience, since the encoder may not provide the required flexibility. For example, even though the IROI scalability option in the SVC extension of H.264/AVC is helpful for performing a cropping operation, the overhead may be too much to allow its use in practical applications. Scalability options have to be selected in such a way that they do not penalize all the users connected to the classroom to provide specific functionalities that only some of the users actually need. Table I summarizes a set of identified adaptation operations to satisfy the objectives of the virtual classroom scenario described in Section II by fulfilling the possible user preferences and usage environment constraints.

TABLE I. ADAPTATION POSSIBILITIES FOR DIFFERENT SCENARIOS IN A VIRTUAL CLASSROOM SESSION

Scenario	Adaptation
Bandwidth scarcity constraint	• Bit rate transcoding, discarding higher SNR scalability layers, spatial scalability layers and temporal scalability layers. • Prioritize bit rates for important regions of the frame.
User is at a low signal reception area	Improve error resilience and/or use stronger error protection.
Receiving device does not support documents	Transmoding documents to a video sequence.
Inadequate display size	• Downscaling to lower resolution. • Cropping a selected region.
Various priorities for displaying different objects	Scaling selected objects according to priority.
A user wants to watch highlights	Summarize the session.
Delay sensitive content transmission (chat, discussion, debate etc.)	• Prioritize audio content by scaling video content. • Separate background from foreground, and prioritize foreground
Remaining terminal battery power is not enough for the full session	Lowering spatial/temporal resolution and/or fidelity of the video to minimize the utilization of the processor

As presented in the Table I, available scalability options and transcoding mechanisms have been used to address more than one constraint. For example, the spatial scalability option in SVC can be used to address constraints such as terminal display size, processing power issues and network bandwidth limitations whereas temporal scalability addresses terminal processing power, power drainage, and network bandwidth issues as well as terminal display refresh rate limitations.

One of the most important adaptation operations required in the virtual classroom scenario is the error resilience adaptation. Adaptation requirements in such a scenario can be extended to error robustness, especially for users connected over error-prone mobile channels. The objective of the error resilience adaptation is to adapt the encoded video stream based on specific requirements of the user and underlying channel conditions. The concept of unequal error protection with joint source and channel coding can be used for protecting sensitive syntax elements, such as packet header and motion information of the video content using data partitioning. Partitions can then be transmitted over different radio bearers of a mobile channel with different channel coding schemes to provide more protection to important parts of the video stream.

Another error resilience adaptation option is to provide higher protection to the segments of a frame that are expected to have the highest distortion during transmission. This is based on modelling distortion and providing higher source coding rates to the regions with higher estimated distortion, and then transmitting it over a radio bearer with a more robust channel protection scheme. The scheme is further extended to apply different levels of importance to certain regions of the content. A student connecting through a mobile phone would struggle to comprehend the content if the whole of the visual screen were scaled to the screen size of the mobile phone. The adaptation engine in such a situation would select a certain region of the screen and then scale it to the screen size of the viewing device.

The feasibility of using scalability options in the SVC extension of H.264/AVC to achieve some of the required adaptation operations in the virtual classroom scenario have been evaluated using publicly available video sequences [33]. In these evaluations, the JSVM 7.13 [34] and a H.264/AVC transcoder based on the same JSVM version are used as the software platform. Figure 9 compares the objective quality of bit rate adaptation using transcoding and fine grain fidelity scalability for the 'Crowd Run' video test sequence. Four scalable and one non-scalable source bit streams are used as the original video sequences. Scalable bit streams are obtained by varying the number of temporal and spatial scalability levels. Highest spatial and temporal resolutions are 1280×704 pixels (cropped from the original 1280×720 resolution to meet frame size constraints [34] in JSVM software) and 50 Hz (progressive), respectively. Lower resolutions are dyadic subdivisions of the maximum resolution. The adapted bit streams also have the same temporal and spatial resolutions. For comparison purposes, the rate distortion performance that can be achieved by directly encoding the raw video sequence is also shown in the figure.

According to Figure 9, it is clear that fidelity scalability offers flexibility of adjusting the bit rate over a large range. Since fine granular scalability layers can be truncated, a large number of bit rate adaptation points can be achieved with this technology allowing the adaptation subsystem to react to the dynamics of the available network resources more precisely. These experimental results also suggest that the objective quality can be further improved by selecting the most appropriate number of spatial and temporal scalability layers for a given situation. For example, when there is no demand for resolutions below 640×352 pixels, the encoder can omit the 320×176 pixel resolution. Limiting the number of resolution layers to two (*i.e.*, $S = 2$) can achieve an additional objective quality gain of over 0.5 dB. In order to take such a decision dynamically, though, it is necessary to have information regarding the required number of adaptation levels at a given time. Since ADE tracks the dynamics of the prevailing context, a feedback from the ADE can be used to decide the level of adaptation.

The number of temporal scalability levels can be increased only at the expense of delay. However, unlike in the case of the number of spatial scalability levels, increasing the number of temporal levels, in return, increases the compression efficiency as illustrated in Figure 9. The reason behind the improved compression efficiency is the use of more hierarchically predicted bidirectional pictures (B-pictures) to achieve more scalability layers [27]. Consequently, the allowed maximum delay becomes the major factor in selecting the number of temporal scalability layers.

Figure 10 shows rate distortion characteristics of adapting the above mentioned source bit streams to achieve 640×352 pixel and 25 Hz temporal resolution. Source bit streams used for this adaptation are the same as those used in adaptations described in Figure 9. In this case, when $S = 2$, the adaptation performs better than transcoding. This is because when $S = 2$, 640×352 pixel resolution is coded as the base layer which is H.264/AVC compatible.

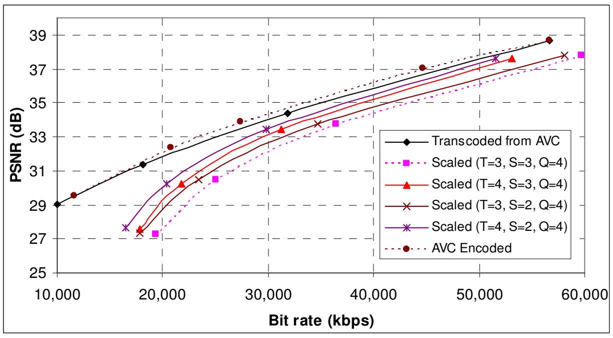

Figure 9. Objective quality comparison of bit rate adaptation using fine grain fidelity scalability (T – number of temporal scalability levels, S – number of spatial scalability levels, and Q – number of fine grain fidelity scalability levels)

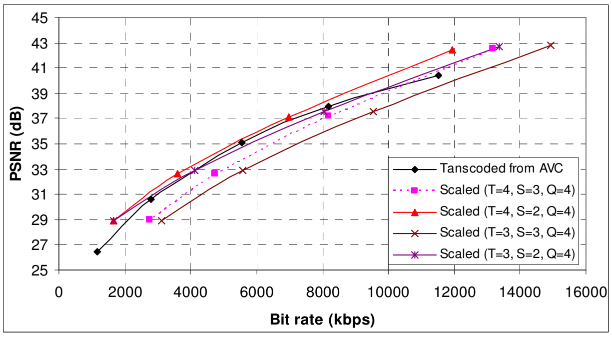

Figure 10. Objective quality comparison of adaptation to 640×352 spatial and 25 Hz using temporal, spatial and fine grain fidelity scalability

Moreover, Figure 9 and Figure 10 demonstrate that the scalability options are only available at a cost of coding efficiency. However, the computational complexity associated with the adaptation operation, which can be achieved by a simple packet selection, is negligible compared to transcoding operations that achieve the same adaptation from a non-scalable bit stream. Therefore, it can be concluded that scalability is a more effective way of assisting spatial and temporal adaptation requirements in the proposed virtual classroom application. Currently, we are investigating the IROI scalability in SVC extension of H.264/AVC for achieving attention area based content adaptation.

VI. CONCLUSIONS

This paper describes a proposed framework for a virtual classroom application, which is being developed in the VISNET II NoE. It combines an existing VCS with context-aware content adaptation of multimedia contents to enhance the quality of the user experience while managing the digital rights. The proposed content adaptation subsystem infers higher-level contexts from low-level contextual information. The ADE considers this information together with the restrictions imposed by the associated DRM to decide the optimum adaptation option that offers the most relevant information to the user while maximizing the level of user satisfaction. Subsequently, the AE carries out the adaptation operation recommended by the ADE. The proposed adaptation subsystem is an important integral component that enables seamless access to virtual classroom sessions overcoming ever increasing heterogeneity related issues, which exist in devices, systems, services and applications as well as user preferences.

ACKNOWLEDGMENT

The work presented was developed within VISNET II, a European Network of Excellence (http://www.visnet-noe.org), funded under the European Commission IST FP6 programme. The authors would like to thank their colleagues of the VISNET II project for their fruitful discussions in the development of the work described here.

REFERENCES

[1] S. Pokraev, P. D. Costa, G. Pereira Filho and M. Zuidweg, “Context-aware services; State-of-the-art,” Telematica Instituut, Tech. Rep., TI/RS/2003/137, 2003.

[2] A. Singh and M. Conway, “Survey of context aware Frameworks – Analysis and Criticism,” University of North Carolina at Chapel Hill, Information Technology Services, Tech. Rep., Jul. 2006.

[3] *Information technology - Multimedia Framework - Part 7: Digital Item Adaptation*, MPEG-21 Part-7: ISO/IEC 21000-7, 2004.

[4] I. Burnett, F. Pereira, R. Walle, and R. Koenen, Eds., *The MPEG-21 Book*, John Wiley and Sons Ltd, 2006.

[5] *Information technology - Multimedia Framework - Part 1: Digital Item*, MPEG-21 Part-1: ISO/IEC 21000-1, 2004.

[6] *Information technology - Multimedia Framework - Part 2: Digital Item Declaration Language*, MPEG-21 Part-2: ISO/IEC 21000-2, 2004.

[7] *Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies 1.0*, W3C Recommendation, 2003. [Online]. Available: http://www.w3.org/TR/CCPP-struct-vocab/

[8] *Resource Description Framework (RDF): Concepts and Abstract Syntax*, W3C Recommendation, 2004. [Online]. Available: http://www.w3.org/TR/rdf-concepts/

[9] *Web Ontology Language (OWL): Overview*, W3C Recommendation, 2004. [Online]. Available: http://www.w3.org/TR/owl-features/

[10] ISO/IEC, ISO/IEC IS 21000-5 – Rights Expression Language.

[11] ISO/IEC, ISO/IEC IS 21000-6 – Rights Data Dictionary.

[12] XrML. [Online]. Available: http:// www.xrml.org/

[13] X. Wang, J. Delgado, and C. Barlas, “ISO/IEC 21000-5/FDAM 1 Rights Expression Language: the MAM profile,” ISO/IEC JTC 1/SC 29/WG 11/N8342, Klagenfurt, Austria, Jul. 2006.

[14] F. Chiariglione, T. Kim, and X. Wang, “ISO/IEC 21000-5/FDAM 2 Rights Expression Language: the DAC profile,” ISO/IEC JTC 1/SC 29/WG 11/N8812 Marrakech, MA, Feb. 2007.

[15] TV-Anytime Forum. [Online]. Available: http://www.tv-anytime.org/

[16] *Information technology – Multimedia framework – Part 7: Digital Item Adaptation, AMENDMENT 1: DIA Conversions and Permissions*, MPEG-21 Part-7: ISO/IEC 21000-7:2004/FDAM 1:2005(E), 2004.

[17] Y. Wang, J.-G. Kim, S.-F. Chang, and H.-M. Kim, “Utility-Based Video Adaptation for Universal Multimedia Access (UMA) and Content-Based Utility Function Prediction for Real-Time Video Transcoding,” *IEEE Trans. Multimedia*, vol. 9, no. 2, pp. 281-284, Feb. 2007.

[18] M. Kropfberger, K. Leopold and H. Hellwagner, “Quality variations of different priority-based temporal video adaptation algorithms,” in *Proc. IEEE 6th Workshop Multimedia Signal Process.*, Siena, Italy, pp. 183-186, Oct. 2004.

[19] M.A. Bonuccelli, F. Lonetti, and F. Martelli, “Temporal Transcoding for Mobile Video Communication,” in *Proc. 2nd Annu. Int. Conf. Mobile Ubiquitous Syst.: Networking Services*, San Diego, California, pp. 502-506, Jul. 2005.

[20] H. Shu and L.-P. Chau. “Frame-skipping transcoding with motion change consideration”, in *Proc. IEEE ISCAS 2004*, vol. 3, pp. 773–776, Vancouver, Canada, May 2004.

[21] J.-N. Hwang, T.-D. Wu, and C.-W. Lin, “Dynamic frame skipping in video transcoding”, in *Proc. IEEE Workshop Multimedia Signal Process.*, pp. 616–621, Redondo Beach, CA, USA, Dec. 1998.

[22] A. Doulamis and G. Tziritas, “Content-based Video Adaptation in Low/Variable Bandwidth Communication Networks Using Adaptable Neural Network Structures,” in *Proc. Int. Joint Conf. Neural Networks*, Vancouver, BC, Canada, Jul. 2006.

[23] Y. Wang, S.-F. Chang and A. C. Loui, “Subjective preference of spatio-temporal rate in video adaptation using multi-dimensional scalable coding,” in *Proc. IEEE Int. Conf. Multimedia Expo*, Taipei, Taiwan, Jun. 2004.

[24] W.-H. Cheng, C.-W. Wang, and J.-L. Wu, “Video Adaptation for Small Display Based on Content Recomposition,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 1, pp. 43-58, Jan. 2007.

[25] M. Bettini, R. Cucchiara, A. Del Bimbo and A. Prati, “Content-based video adaptation with user’s preferences,” in *Proc. IEEE Int. conf. Multimedia Expo*, vol. 3, pp. 1695-1698, Jun. 2004.

[26] M. Xu, J. Li, Y. Hu, L.-T. Chia, B.-S. Lee, D. Rajan, and J. Cai, “An Event-driven Sports Video Adaptation for the MPEG-21 DIA Framework,” in *Proc. IEEE Int. Conf. Multimedia Expo*, pp. 1245-1248, Jul. 2006.

[27] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the Scalable Video Coding Extension of the H.264/AVC Standard,” input document to The 23rd Joint Video Team (JVT) meeting, San Jose, CA, USA, Apr. 2007.

[28] Creative Commons. [Online]. Available: http://www.creativecommons.org/

[29] J. Delgado, and E. Rodríguez, “Towards the interoperability between MPEG-21 REL and Creative Commons licenses,” ISO/IEC JTC 1/SC 29/WG 11/ M13118, Montreux, Switzerland, Apr. 2006.

[30] T. Kim, J. Delgado, F. Schreiner, C. Barlas, and X. Wang, “ISO/IEC 21000-5:2004/FPDAM 3: ORC (Open Release Content) Profile,” ISO/IEC JTC 1/SC 29/WG 11/N9108, San Jose, USA, Apr. 2007.

[31] T. Gu, H. K. Pung, and D. Q. Zhang, “A service-oriented middleware for building context aware services,” *Elsevier J. of Network and Comput. Applicat. (JNCA)*, vol. 28, no. 1, pp. 1-18, Jan. 2005.

[32] [Online]. Available: http://jena.sourceforge.net/inference/

[33] [Online]. Available: ftp://vqeg.its.bldrdoc.gov/HDTV/SVT_MultiFormat

[34] ITU-T and ISO/IEC JTC 1, “JSVM 7 software,” Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 20th Meeting, JVT-T203, Jul. 2006.