

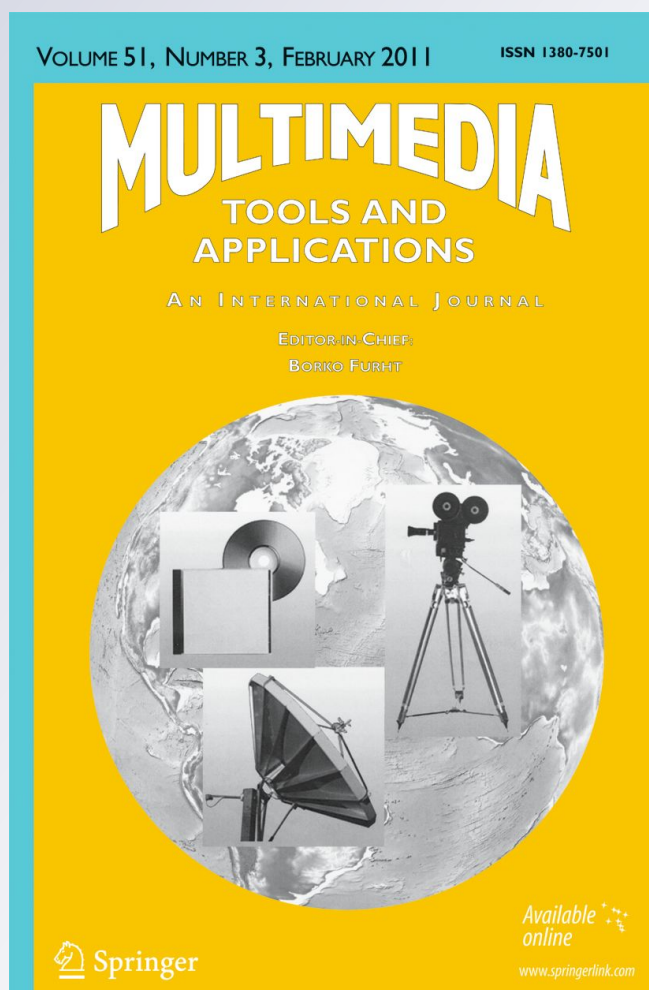
Advanced delivery of sensitive multimedia content for better serving user expectations in Virtual Collaboration applications

**Maria Teresa Andrade, Safak Dogan,
Anna Carreras, Vitor Barbosa,
Hemantha Kodikara Arachchi, Jaime
Delgado & Ahmet M. Kondo**

Multimedia Tools and Applications
An International Journal

ISSN 1380-7501
Volume 58
Number 3

Multimed Tools Appl (2012) 58:633-661
DOI 10.1007/s11042-011-0749-6



Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media, LLC. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.

Advanced delivery of sensitive multimedia content for better serving user expectations in Virtual Collaboration applications

**Maria Teresa Andrade • Safak Dogan • Anna Carreras •
Vitor Barbosa • Hemantha Kodikara Arachchi •
Jaime Delgado • Ahmet M. Kondo**

Published online: 23 February 2011
© Springer Science+Business Media, LLC 2011

Abstract A major challenge when accessing protected multimedia content in heterogeneous usage environments is the ability to provide acceptable levels of quality of experience to all involving users. Additionally, different levels of protection should be possible to be addressed when manipulating the content towards the quality of experience maximization. This paper describes the use of a context-aware and Digital Rights Management (DRM)-enabled content

M. T. Andrade • V. Barbosa
INESC Porto, Rua Dr. Roberto Frias 378, 4200-465 Porto, Portugal
e-mail: maria.andrade@inescporto.pt

V. Barbosa
e-mail: vhsb@inescporto.pt

S. Dogan • H. Kodikara Arachchi • A. M. Kondo
I-Lab Multimedia Communications Research, Faculty of Engineering and Physical Sciences,
University of Surrey, Guildford, GU2 7XH Surrey, UK

S. Dogan
e-mail: S.Dogan@surrey.ac.uk

H. Kodikara Arachchi
e-mail: H.Kodikaraarachchi@surrey.ac.uk

A. M. Kondo
e-mail: A.Kondo@surrey.ac.uk

A. Carreras • J. Delgado
Universitat Politècnica de Catalunya, Departament d'Arquitectura de Computadors, Campus Nord,
Modul D6, Jordi Girona 1-3, 08034 Barcelona, Spain

A. Carreras
e-mail: annac@ac.upc.edu

J. Delgado
e-mail: jaime.delgado@ac.upc.edu

M. T. Andrade (✉)
Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal
e-mail: mandrade@fe.up.pt

adaptation platform towards meeting these challenges. The platform was conceived to deliver advanced content adaptation within different application scenarios, among which Virtual Collaboration (VC) was central. Descriptions of use cases implemented by the platform in heterogeneous VC environments are provided. Conducted experiments have highlighted the benefits to users when compared to an operation without the platform. Results of different adaptations suitable to sensed context conditions are also provided and analyzed. A brief description of the platform functionality is included together with pointers to additional information.

Keywords Context · Adaptation · DRM · Virtual collaborations · Ontology reasoning

Abbreviations

AA	Adaptation authorizer
ADE	Adaptation decision engine
AE	Adaptation engine
AES	Adaptation engine suite
AQoS	Adaptation quality of Service
CAO	Context aware ontology
CS	Classification scheme
CxP	Context provider
DIA	Digital item adaptation
DL	Description logic
DRM	Digital rights management
IROI	Interactive region of interest
GPS	Global positioning system
MDS	Multimedia description schemes
MOS	Mean opinion score
MULTICAO	Multimedia context aware ontology
OWL	Web ontology language
REL	Rights expression language
ROI	Region of interest
SNR	Signal-to-noise ratio
SWRL	Semantic web rule language
UED	Usage environment description
VC	Virtual collaboration

1 Introduction

The audiovisual media landscape is facing important changes driven by the digital convergence phenomena together with the enthusiastic acceptance of emergent technologies and of the “on-line” paradigm by the general public. In parallel, technology has continued to evolve and new multimedia-enabled devices and gadgets have been made available, largely contributing to increase heterogeneity. As a consequence, a growing number of consumers, equipped with different devices and having diverse interests and preferences, navigate through some network looking for different types of multimedia information every day.

An access to content that meets users’ expectations must take into account all the different aspects of this heterogeneous scenario. Additionally, some applications exchange sensitive or confidential content, to which different levels of protection may need to be applied. Dynamic monitoring of the context of usage conditions should be observed for the

whole duration of a service to allow for reacting to changes and thus maintaining an acceptable level of quality of the service while assuring the required level of protection. One form of reaction is to adapt the content in a smart way, i.e., to automatically sense contextual conditions, including user preferences and accordingly modify the content characteristics while minimizing the quality degradation. Figure 1 illustrates this situation assuming a Virtual Collaboration (VC) environment with multiple heterogeneous users participating in a VC session and introduces the need for an intelligent control of such a session.

With this goal in view, within the framework of the VISNET II Network of Excellence, the context-aware and Digital Rights Management (DRM)-enabled content adaptation platform illustrated in Fig. 2 was designed. The platform encompasses four main building blocks that enable to deliver the functionality suitable to monitor and understand the constraints of the consumption environment (or context) of each VC session participant, as depicted in Fig. 1, which is then used to control the access to protected or sensitive content and to adapt the content according to those constraints and permissions.

Although applicable to a range of different networked multimedia applications, the platform incorporates mechanisms customized to VC environments. These impose specific requirements while providing specific set of characteristics. Typically, in these environments, the type of exchanged video content consists of a mixture of views of persons with drawings or written material in a whiteboard for example, with close plans of pieces of equipment, documents or pictures, etc. Also, some of this content may be sensitive, which means that different levels of permission may need to be applied depending on the user or on the physical location the user is in when receiving it. Moreover, it is common to find multiple heterogeneous participants in each VC session, meaning that each one is likely to assume a distinct role and may present different constraints, capabilities and access rights.

The main focus of this paper is to describe the utilization of the developed platform in use cases within those typical VC environments. These use cases impose specific requirements such as simultaneous users with diverse context constraints (many-to-many heterogeneous communications) and distribution of content with different levels of confidentiality. Accordingly, a detailed description is made of specific usage scenarios where the platform is utilized, within VC applications, to exchange sensitive multimedia content in an advanced way. The goal is to illustrate how the content adaptation platform

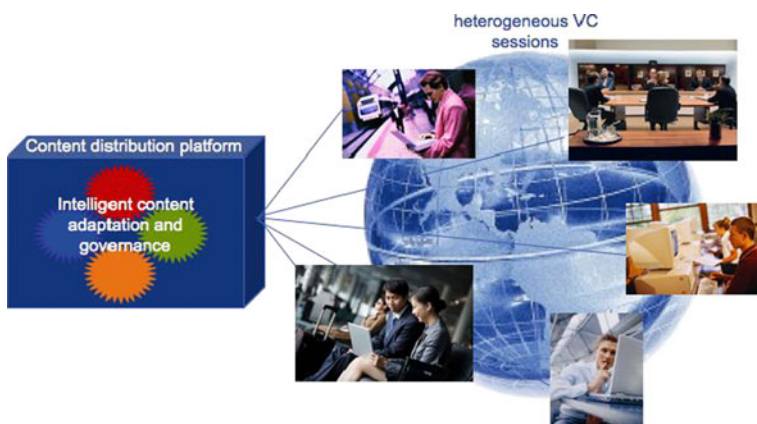


Fig. 1 Virtual Collaboration (VC) applications in heterogeneous environments

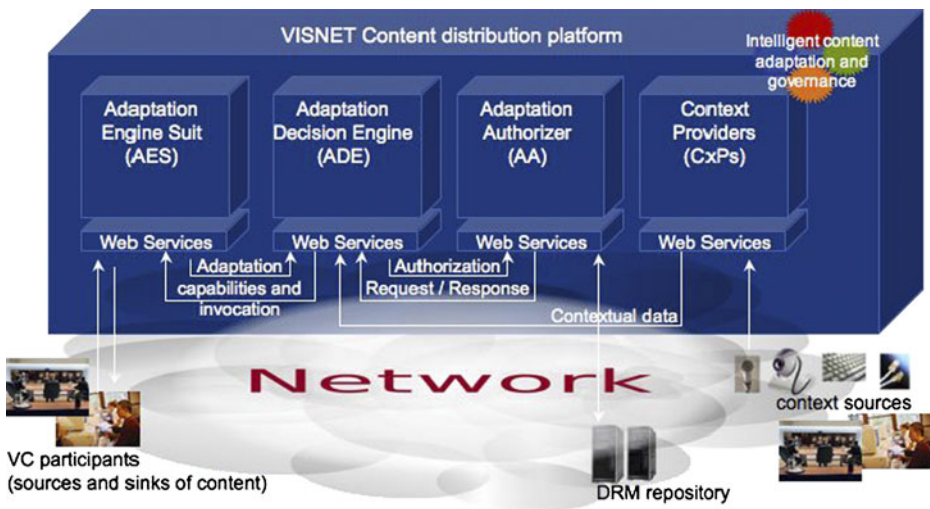


Fig. 2 Functional architecture of the context-aware and DRM-enabled content adaptation platform

can serve the user expectations by delivering the content adapted, in the best way that is permissible, to the dynamic conditions of the usage environment. Results of adapting the content are presented, highlighting the benefits to users when compared to an operation without the use of the platform. The types of admissible content adaptation operations offered by the system are especially convenient for typical VC application scenarios, taking into consideration the type of content exchanged. Likewise, the decision-making mechanisms use knowledge specific to this kind of applications in addition to general knowledge capable of characterizing generic usage contexts. Moreover, in relation to other published work on content adaptation, our system addresses many-to-many communications (such as the ones in VC sessions) where the context of usage of each participant can be very different, dealing also with different levels of protection of content. Finally, our system incorporates entities designated as “Context Providers (CxPs)” which dynamically sense modifications in the usage context conditions and accordingly notify the decision-making mechanisms of the platform. The CxPs encapsulate the sensed context into standardized descriptions, namely MPEG-21 [20] and MPEG-7 Multimedia Description Scheme (MDS) [21], with the aim of promoting interoperability. The decision mechanism collects this information and uses it to populate an ontology, which enables to capture the semantics and enrich the knowledge about the situation the user is in.

An overview of the functionality of the platform modules is also provided in this paper. However, interested readers are kindly invited to find further in-depth information in previous publications [4, 5, 32]. As referred above, the main goal of this paper is to describe the use and benefits of the platform under real-world situations, providing a detailed and quantitative characterization of contextual conditions that would affect the quality of experience of the user. To that aim, even though not being the central objective of the paper, the description of the platform architecture, more specifically its ontology-based context aware decision module, provides additional details when comparing to those previous publications and to the other platform modules. It thus provides deeper insights on the mechanisms that specifically enable the system to gain consciousness of contextual conditions that characterize the described use cases, which in turn empowers the system to

adapt the content in the most satisfying way. The utilization and advantages of the platform in those situations in clear contrast with when the proposed platform is not used are therefore described and analyzed in this paper.

The remainder of this paper is organized as follows. A brief review of relevant research work is presented in Section 2, highlighting the advances and benefits brought in by the work described here. Section 3 provides an overview of the functional architecture of the context-aware adaptation platform, briefly describing the role of each module. Section 4 describes the selected use cases, providing a set of walks-through together with the sequence of interactions to illustrate the use of the developed platform in typical VC scenarios. Section 5 presents results of performing a number of different content adaptations according to varying contextual conditions. Finally, Section 6 draws the conclusions.

2 Current research aiming at context-aware content adaptation

Content adaptation has gained considerable importance in today's multimedia communications. This has been fostered by the information explosion triggered by the emergence of the World Wide Web and by the continuing advances in technology, which are steadily emphasizing the great heterogeneity that exists in devices, systems, services and applications today. The implementation of meaningful content adaptation operations that meet users' expectations and satisfy their usage environment constraints, requires the use of contextual information, and thus the ability to take decisions on how to adapt the content based on that information. Dey [7] provides a good generic definition of context. This is probably the most quoted definition of context:

Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves.

This definition matches the purpose of the work described here very well. In fact, to achieve its objectives, the developed platform needs to obtain a characterization of the users of the service, as well as of all technological and natural elements involved in the delivery of the service: the network interface; the core network; the terminal; the surrounding environment, including the proximity of other persons; and the type of content being delivered.

Different approaches are being studied worldwide to develop decision algorithms that provide the indication of adequate content adaptation given the restrictions imposed by the context of usage. The contextual metadata required by the decision algorithms, in line with the above referred definition, describes characteristics and conditions of the context of usage: networks (bandwidth availability, error rates, jitter, etc.); terminals (screen size, CPU availability, etc.); natural surrounding environment (lightning conditions, noise level, etc.); and users (disabilities such as hard-to-hear or color perception, preferences such as language, text over voice, images over video, etc.).

The decision algorithm is seen as the intelligent part of content adaptation systems. It collects required metadata (context-related, as referred above and content-related to describe the characteristics of the content being consumed) and selects service parameters that best suit the conditions of the consumption environment. Generally speaking, the adaptation decision process is seen as a problem of resource allocation in a generalized rate-distortion framework: given a set of available resources or restrictions (network bandwidth,

computer processing power, display spatial dimensions, etc.), it selects the set of service parameters that leads to a variation of the content satisfying those restrictions, while maximizing a given utility (the quality of the content, the price the user has to pay, etc.). Early work and still many of the existing state-of-the-art implementations [11, 12, 16, 30, 31] rely on the use of only low-level metadata, usually values directly acquired from sensors (low-level contextual metadata) and technical parameters of audio and video encoding schemes (low-level content-related metadata) indicating the possible variations of the content associated to a given utility. Some of these frameworks are built on top of the MPEG-21 standard [3, 20], whereby the decision taking operation is focused on solving a constraints matching problem. The constraints, imposed by the conditions of the context of usage, are treated individually and no relations are established between them, or with the characteristics of the application or service. Still, the full meaning of context-awareness is to make use of sensed low-level context to build understanding at higher levels of conceptual abstractions, the way humans do. These systems fail in providing the required support to build this kind of understanding or knowledge [1, 14]. Reasoning about the sensed context usually requires the use of models with rules and relations, to allow combining and relating sensed context. Accordingly, researchers have proposed the use of ontologies to build context-aware systems. In this field of research, the ontology provides a common framework for context representation and information consistency checking, enabling also to infer additional knowledge from the acquired context, thus capturing the semantics of the current situation.

The work more recently undertaken by a number of research groups thus addresses the use of ontologies to obtain high-level semantic descriptions of the context of usage [2, 6, 10, 13–15, 17, 27–29]. Although some combine standards with ontologies [10, 17], a great part of the published work relies on proprietary models to acquire and represent context. Moreover, they usually concentrate on a limited number of adaptations and most often upon Web content only or entertainment video and try to explore context to improve usability aspects. Generally, they do not coherently explore the interrelations among different types of low-level contextual information, customized to the specificities of different applications. Likewise, the aspects concerning security of content and ensuring the privacy of the user are also generally overlooked.

The work presented in this paper combines the use of open standards, notably MPEG-21 [3, 20] and MPEG-7 MDS [21], with ontologies [25] to drive a context-aware content adaptation decision. It clearly incorporates the privacy and security dimension and is able to cope with dynamically varying contextual situations. It foresees the dynamic gathering and use of low-level context regardless of its origin, as long as it is represented according to the selected open standard representation. It incorporates mechanisms to use, customize and infer higher-level knowledge through formally-defined ontologies, both generic as well as specific to concrete application domains. Moreover, it is suitable to be used in many-to-many communications and is built in a complete modular approach, where functionality specific to different applications can be easily plugged and used as needed. This feature applies to the type of content adaptations available as well as to the knowledge mechanisms incorporated. These features are briefly explained in Section 3. Details can be found in previous publications and Web repository [4, 5, 32].

Standardization bodies such as the World Wide Web Consortium (W3C) [33] and Motion Picture Experts Group (MPEG) [19] have started working on specifications to represent and exchange context. MPEG-21 standard is presently the most complete standard, and thus the ideal candidate for exploiting in this work. Part 7 of the standard (MPEG-21 Digital Item Adaptation, DIA) [23] provides a full set of contextual information

that can be applied to any type of multimedia system, as it assures device-independence. It specifies appropriate eXtensible Markup Language (XML) schemas for the description of terminal capabilities and network characteristics as well as User characteristics and preferences, and natural environment conditions. These XML-based descriptions are designated of Usage Environment Descriptions (UEDs). Furthermore, MPEG-21 also provides the support for DRM [34], thus enabling to define users' rights to act on a digital content. Nevertheless, the mechanisms that are provided in these specifications for establishing relationships among acquired contextual information and constraints are still very limited. For this reason, ontologies have been incorporated in our work as explained above. Although kept simple, our ontology is easily extendable and it is powerful as it already supports security aspects, as well as new types of contextual information. Such is the case of the "hasFloor" descriptor, useful for VC applications, which is further detailed in Section 4 and in [4].

3 The context-aware and DRM-enabled content adaptation platform

The developed platform, illustrated above in Fig. 2, consists of four modules: Adaptation Decision Engine (ADE), Adaptation Authorizer (AA), Context Providers (CxPs), and Adaptation Engine Suites (AESs), which comprise a set of Adaptation Engines (AEs). It delivers audiovisual content to end-users, adapted to their context of usage (characteristics of terminals, networks, environment, and user preferences), while satisfying DRM restrictions. Contextual information is acquired both at service initiation as well as in a dynamic way throughout the service lifetime being refreshed whenever context conditions change.

Figure 3 provides an overview of the contextual model implemented by the platform. To accurately describe the consumption scenario, we have defined three different kinds of knowledge in our model: 1) related to the multimedia content and obtained from MPEG-7 content-related metadata; 2) related to the usage environment and extracted from contextual information expressed in the form of MPEG-21 UEDs; and 3) associated to specific domain

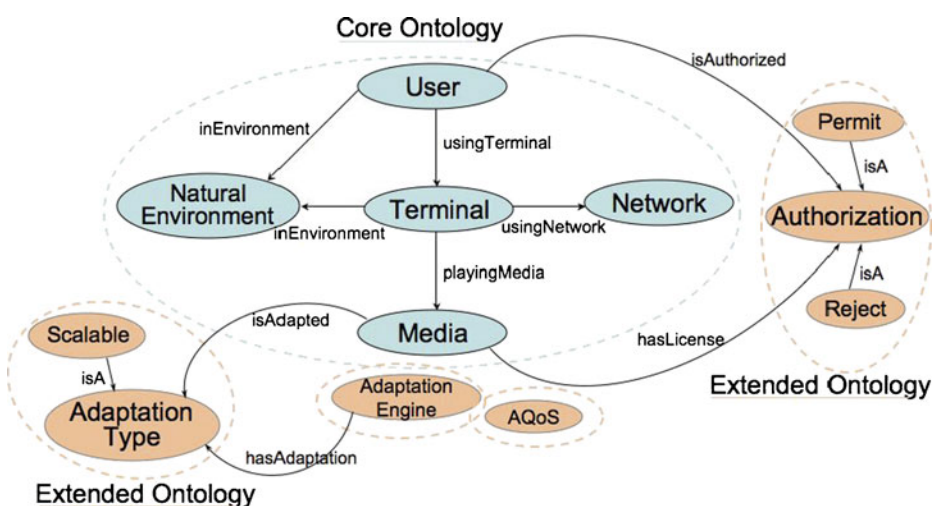


Fig. 3 Overview of the contextual model

of application, comprising related rules and adaptation authorizations. This contextual knowledge-based model has been established as a combination of core concepts and extended concepts. The former set includes concepts such as user's preferences, terminal capabilities, or content characteristics and encapsulates the two first types of knowledge referred above. They are based on the MPEG-21 and MPEG-7 MDS standards and constitute the core ontology of the platform. The latter includes concepts related to adaptation authorizations, types of available adaptation operations adequate to VC applications and rules/statements that apply specifically to VC environments. The former delivers information about persons and entities participating in the collaboration session while the latter provides knowledge specific to the application in view. The advantage of using an ontology to represent the standardized descriptions is the possibility of knowledge sharing between different domains, promoting interoperability and allowing to infer new knowledge. The Classification Schemes (CSs) used to generate the MPEG-21 UEDs and MPEG-7 MDS are also captured in the core ontology. The core ontology needs to be aligned with the MPEG-7 and MPEG-21 specifications. However, we consider this to be an advantage, given that MPEG-7 and MPEG-21 are open standards and can thus be implemented by anyone.

Table 1 lists the different alternatives of content adaptation operations that are possible to be implemented by the proposed platform. They are grouped under three main classes, notably: 1) *scalable*, whereby the AE is capable of delivering a scalable video bit stream with multiple layers of spatial dimensions, of temporal resolutions, of Signal-to-Noise Ratio (SNR) granularities or cropped to focus on or reveal in different Regions of Interest (ROIs); 2) *non-scalable*, whereby the AE is able to adapt a non-scalable video bit stream, delivering different spatial, temporal and SNR resolutions as well as cropping it into different ROIs; and 3) *summary*, whereby the AE adapts the video bit stream by summarizing it. The contextual model, depicted in Fig. 3, together with the combination of the set of content adaptation alternatives currently listed in Table 1, empowers the platform to support a larger number of applications and real-world situations of consumption of protected multimedia content in heterogeneous environments, delivering the content adapted in a dynamic way to the varying conditions of the context of usage.

The adaptation authorization enriches the adaptation decision, not only because it controls the access of users to content by means of MPEG-21 Rights Expression Language (REL) [22], but also it acts as a CxP. In fact, the AA supplies information concerning the permitted adaptations to the ADE, so that it can take the most suitable and DRM-enabled decision.

3.1 Adaptation engine suite (AES)

The AES comprises a set of AEs, which offer capability to perform a number of different content adaptation operations upon request. The AEs are independent software tools that implement different encoding and/or transcoding algorithms on top of the audiovisual

Table 1 Possible content adaptation alternatives supported by the content adaptation platform

Adaptation type	Adaptation operations
Scalable	Spatial, temporal, and SNR scaling; Interactive Region Of Interest (IROI) cropping
Non-scalable	Spatial, temporal, and SNR scaling; ROI cropping
Summary	Summarization

information exchanged between the VC participants. Normally, the audiovisual information is a signal captured by cameras and microphones placed at each VC site. The AES can be centrally located, together with the VC management facility, or otherwise can also be deployed in a distributed manner within each participant's equipment. The proposed AES has been organized in a 3-layer architecture as represented in Fig. 4. This layered architecture is detailed in [4, 5, 32]. Table 1 has listed the major types of adaptation operations offered by the AES, which is envisaged to reduce the processing latency for performing multiple adaptation operations in a sequence. For example, if both cropping and scaling operations need to be performed on a given non-scalable video stream, those operations can occur together in a cascaded fashion.

During its operation, the AES can be in one of the five states depicted in Fig. 5, as follows:

- **System initialization:** during this state, the Service Initialization Agent initializes the system, and subsequently the Registering Agent communicates the AES's capabilities to the ADE. Whenever a new AE is introduced to the AES, the system enters into the same state.
- **Idle:** once the system initialization process is completed, the AES enters into the idle state, waiting for adaptation requests from the ADE.
- **Service initialization:** when the AES receives an adaptation request, it enters into the service initialization state. The Adaptation Decision Interpreter identifies the required AEs to be deployed, which are accordingly configured.
- **In-service:** during this state, which follows the service initialization state, the adaptation is performed on the input media content and the adapted content is forwarded to the user. The AE Monitoring Service also monitors the progress of the adaptation operations during this state.
- **Service re-negotiation:** the AES enters into this state if the ADE signals a need for changing the on-going adaptation. If the AES does not have the adequate AE, it transfers the service to another AES. Otherwise, it changes the service parameters of the currently operating AEs or launches the new AEs to continue to service.

When the system is switched on, the AES automatically enters into the “System initialization” state. During normal operation of the platform, if its services have not been requested, the AES stays in the “Idle” state. This means that, normally, the AES will find

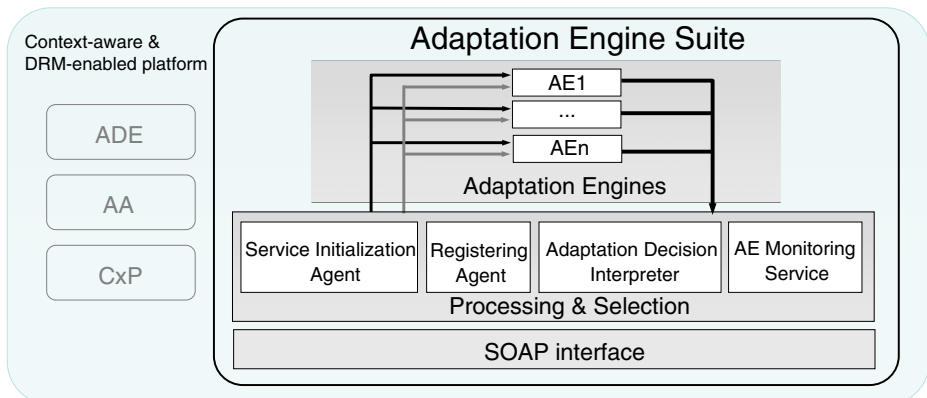
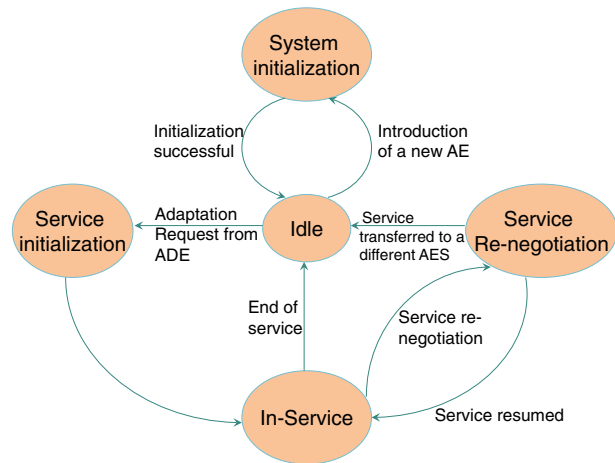


Fig. 4 AES architecture

Fig. 5 State diagram of the AES operation



itself in the “Idle” state whenever it receives a request for adaptation, entering then into the “Service initialization” state.

3.2 Adaptation authorizer (AA)

The main role of an AA in a governed system is to allow (or disallow) adaptation operations based on whether they violate any conditions expressed in licenses. Implementation of these licenses, based on MPEG-21 REL and DIA [23], has been described in [4, 5, 32]. The use of these licenses brings innovation to the AA in two dimensions: on one hand, fine-grained descriptions about adaptations expressed in MPEG-21 DIA allow governing these adaptation operations in a flexible way; on the other hand, the use of MPEG-21 REL guarantees the reliability of the system and its interoperability with other DRM standards.

The authorization process can be divided into three main phases:

- Processing of the authorization request message: an XML message with information about the user, the resource and the context.
- Verification of the associated license: MPEG-21 REL constraints are verified by comparing the information contained in the authorization request with the corresponding fields of the licenses.
- Creation of the authorization response: if all MPEG-21 REL constraints are positively verified, an XML authorization response is created, containing a Boolean element with the value “true” and the associated adaptation restrictions. If any of the MPEG-21 REL constraints is not satisfied, the authorization response message contains only the Boolean element with the value “false”.

3.3 Context providers (CxPs)

Complex multimedia consumption scenarios encompass a diverse range of entities that can provide useful contextual data. A few of the examples are network operators (through the network equipment); content providers (through media repositories or encoders); equipment manufacturers (through terminal devices or sensors, such as cameras, microphones, etc.); and users (via the terminal device or databases holding user profiles). In our work, this

information is represented using the MPEG-21 DIA UED [23]. This low-level contextual information is generated by the aforementioned entities, referred to as CxPs, typically in the form of numeric values, and supplied to the ADE.

3.4 Adaptation decision engine (ADE)

The ADE decides when and how to adapt the content to suit restrictions of the context of usage. It was conceived as a collection of four distinct subsystems: Context Service, Ontology Service and Adaptation Decision Managers, and ADE Interface Layer. Figure 6 provides a detailed functional view of the ADE. Globally, it fulfills the following list of requirements:

- Dynamic gathering of contextual and content information coming from different CxPs in a standardized format (supervised by the “Context Service Manager”).
- Generation of additional knowledge from the collected low-level context (through the “Ontology Service Manager”)
- Identification of the capacities of available AESs; selection of service parameters that meet the contextual constraints without violating users’ rights, using all the acquired knowledge; and activation of an appropriate AES (under the control of the “Adaptation Decision Manager”).

Instead of directly using the contextual information obtained from the CxPs, the ADE combines it with additional knowledge concerning the specifications of the application in view. The use of an ontology-based approach to model context and build the referred additional knowledge fulfils this goal. CxPs detect changes in the context of usage and accordingly encode relevant contextual values into MPEG-21 UEDs, which are sent to the ADE. This module extracts values from these standardized files (i.e., the MPEG-21 UEDs) to populate the core ontology. This standardized information, together with content metadata expressed in MPEG-7 MDS and an authorization response sent by the AA, is thus programmatically instantiated in the data ontology and aligned with the conceptual model. The additional knowledge, specific to the application (in this case a VC application), is either directly built in the ADE or again provided by external CxPs during the service

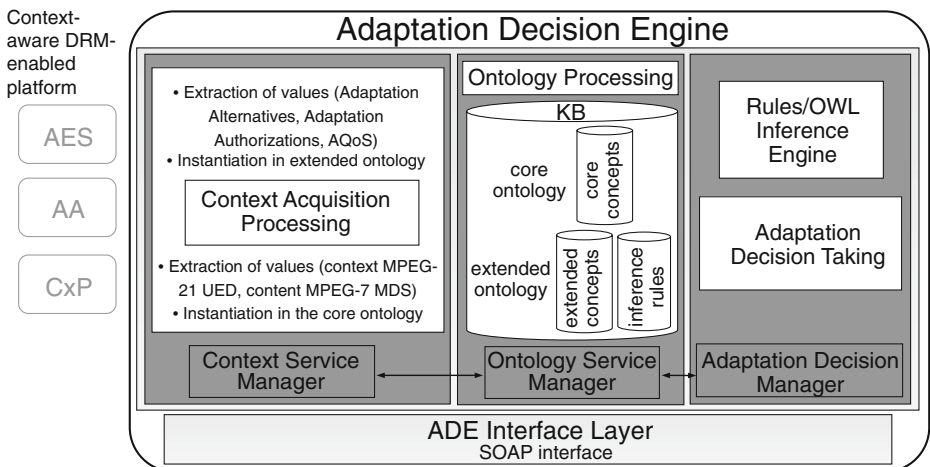


Fig. 6 ADE functional architecture

lifetime. The former comprises specific rules on how to combine contextual information as well as concepts specific to the VC application, such as the location of the speaker. The latter encompasses DRM information to allow governing the type and extent of adaptation operations that can be performed upon a given content. The complete acquired knowledge is then used to formulate an adaptation decision and accordingly invoke the AES module.

The proposed ontology, designated as Multimedia Context Aware Ontology (MULTICAO) [24], comprises two layers (core and extended layers) and was developed using the Web Ontology Language (OWL) Description Logic (DL) sublanguage [18, 26]. OWL-DL provides decidability and complete reasoning mechanisms support to make automatic inferences over the knowledge base. MULTICAO is a set of context ontologies designed to provide the support for context-aware and DRM-enabled multimedia content adaptation decisions in a variety of networked multimedia applications. Its core ontology describes the common knowledge denominator under different real-world situations of multimedia content consumption. The extended layer is made up of multiple ontologies, each one incorporating domain-specific knowledge, relevant to the application in view. With this approach, several extended ontologies can be created, used and merged as required by the domain application. In our case, the extended layer includes an ontology for representing content adaptation authorization mechanisms and an ontology providing the necessary Semantic Web Rule Language (SWRL) rules [9] specific to VC applications. SWRL adds rules to OWL ontologies while also providing an extra layer of expressiveness and the capability to infer additional information from an OWL knowledge base. Thus, using a set of specific rules in the extended ontology enables the inference of real-world situations that are indeed likely to occur in specific applications, such as those in VC environments. The complete set of files for the MULTICAO ontology is available online [32]. The core layer models the context of usage independently of the application, whereas the extended layer adds further concepts and provides rules to interrelate the low-level context according to the specific application in consideration. Figures 7, 8 and 9 provide illustrations of excerpts of the developed ontologies, showing the main OWL classes, individuals, object properties linking classes, and subclasses and datatype properties linking individuals to values that characterize them. The developed core ontology uses in total 74 concepts, 47 object properties, 127 datatype properties, and 804 individuals in its current version. In those illustrations, we have used different colors to differentiate elements of the core ontology (light blue) from those of the extended ontologies (in orange).

Figure 9 shows excerpts of two extended ontologies specifically developed for VC applications where protected content is exchanged. The top left ontology provides information concerning the region where the speaker is located, thus allowing for selecting a ROI adaptation. This kind of adaptation is very useful in VC applications where often the most important part of the visual scene is located around the participant who is speaking (i.e., the participant who “has the floor”). The lower right ontology carries information concerning the possibility of performing adaptation on protected content. It enables the ADE to deduce the type of constraints that apply, thus being able to select an adaptation that is authorized.

An example of specific rules that are used by the Inference Engine, in this case to deduce whether a user is located in a public or private area, is provided in Table 2.

4 Specific use cases

Our work envisions a VC application scenario, which allows remotely located partners to meet in a virtual environment using state-of-the-art communication and audiovisual

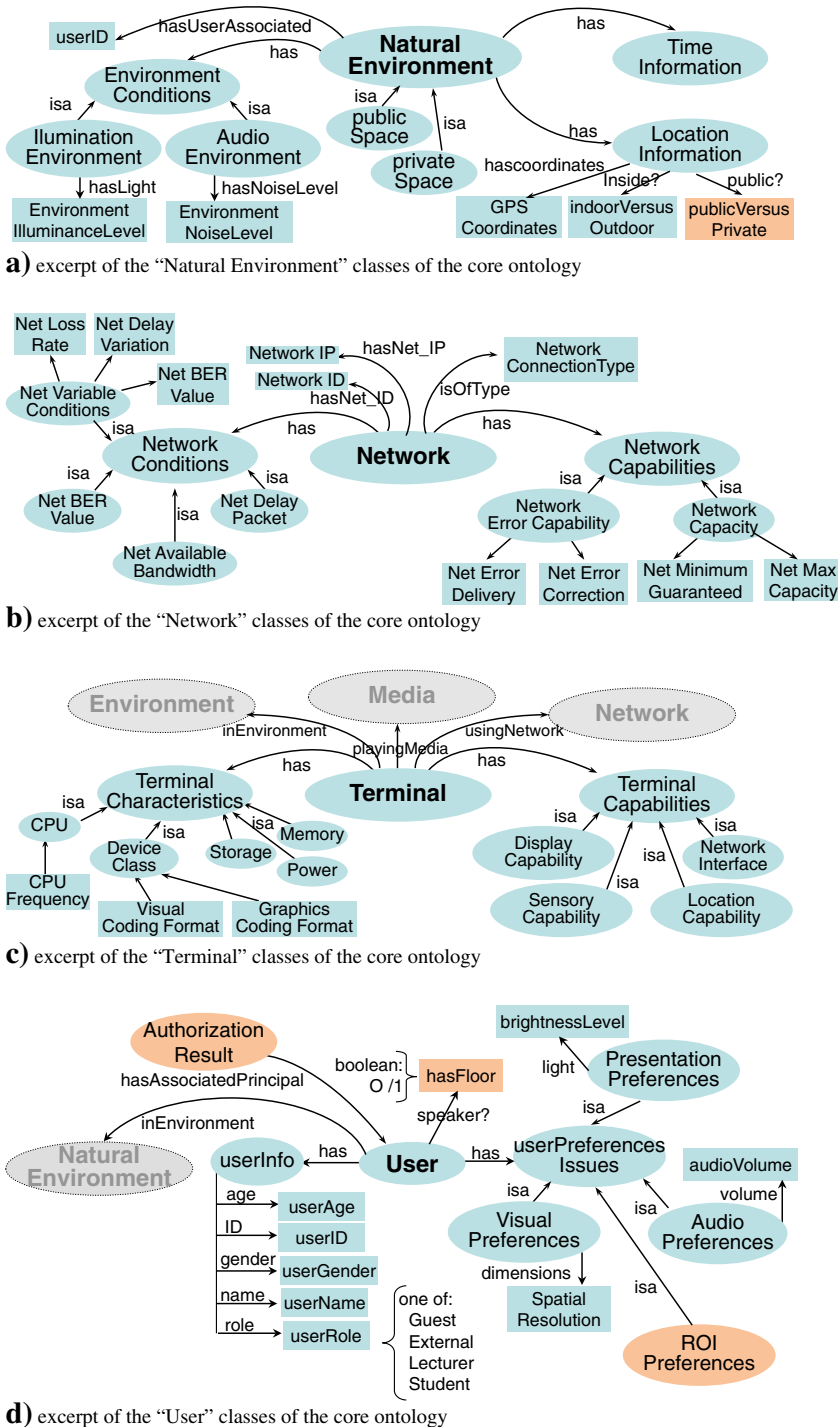
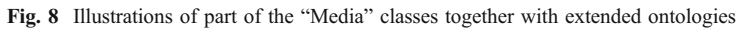


Fig. 7 Illustrations of parts of the contextual ontology



```

graph TD
    VCSession([VC Session]) -- hasAvailableAES --> CAOaesAES([CAO:aes:AES])
    CAOaesAES -- needsAdaptationType --> CAOMedia([CAO:Media])
    CAOMedia -- hasROI --> RegionofInterest([RegionofInterest])
    RegionofInterest -- isLocatedinROI --> CAOUser([CAO:User])
    CAOUser -- hasFloor --> PersonAttentiveness[Person Attentiveness]
    PersonAttentiveness -- hasFloor --> OriginX[Origin X]
    PersonAttentiveness -- hasFloor --> OriginY[Origin Y]
    PersonAttentiveness -- hasFloor --> Height[Height]
    PersonAttentiveness -- hasFloor --> Width[Width]
    CAOMedia -- hasResource --> AdaptationAuthorizationResult([Adaptation AuthorizationResult])
    AdaptationAuthorizationResult -- hasPrincipal --> AdaptationType([AdaptationType])
    AdaptationAuthorizationResult -- isA --> PermitAdaptation([PermitAdaptation])
    AdaptationAuthorizationResult -- isA --> RejectAdaptation([RejectAdaptation])
    PermitAdaptation -- hasChangeConstraints --> ChangeConstraints([ChangeConstraints])
    PermitAdaptation -- hasPermittedChanges --> PermittedChanges([PermittedChanges])
    PermitAdaptation -- hasAssociatedContentOperations --> ContentOperationRights([ContentOperation Rights])
    ChangeConstraints -- hasAssociatedChangeConstraints --> PermittedChanges

```

Extended Concepts Related with specific requirements of a VC session

Extended concepts related with specific requirements of protected content

Fig. 9 Excerpts of two extended ontologies with knowledge specific to the VC application

Table 2 Example of rules to determine whether the user is in a public space or not

$\text{AudioEnvironment}(\text{?userA}) \wedge \text{hasnoiseLevel}(\text{?userA}, \text{?nl}) \wedge \text{swrl:greaterThan}(\text{?nl}, 20)$ $\wedge \text{IlluminationEnviron}(\text{?userA}) \wedge \text{hasIllumination}(\text{?userA}, \text{?il}) \wedge \text{swrl:greaterThan}(\text{?il}, 50) \wedge$ $\text{naturalEnvironment}(\text{?userA}) \wedge \text{personDetected}(\text{?userA}) \Rightarrow \text{publicSpace}(\text{?userA})$

exchanging pervasive yet protected and trusted contents. Accordingly, different use cases within VC applications have been defined, thoroughly characterized and simulated using the content adaptation capabilities provided by he developed platform. This section thus describes these specific use cases, listing the possible adaptation operations that can be performed, the associated DRM information and the low-level contextual information required to trigger the adaptation decision as well as the actual content adaptation processes. The sequence of messages exchanged between the different modules of the platform during the execution of those use cases is equally detailed.

Figure 10 illustrates a generic VC business/office scenario, which has a number of active parties, both local and remote office workers and/or clients of an organization, and offers a Virtual Collaboration System (VCS) facility.

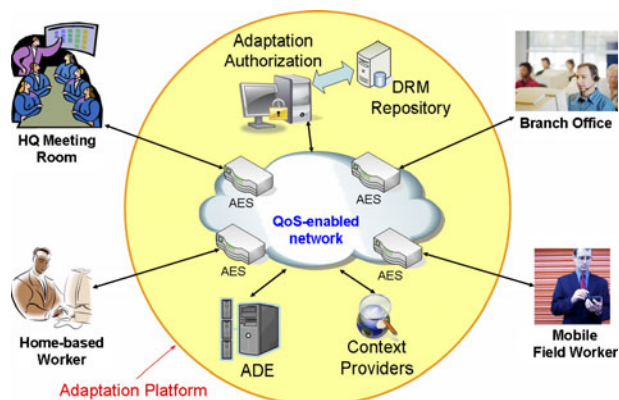
Parties use heterogeneous terminal types, communicating and exchanging media with each other through heterogeneous communications networks:

- Large terminals, supporting multiple users, which have fixed locations and high data rate wired connections to the VCS (e.g., VCS terminals at a Headquarters (HQ) Meeting Room).
- Small terminals, which can be portable yet have a fixed location during the collaboration session with wireless or wired connection to the VCS, as represented by the Home-based Worker in the figure.
- Mobile terminals, which can be moved around while in use and are designed for a single user, as for example the one used by the Mobile Field Worker in the same figure.

All of these result in a highly heterogeneous scenario, in which context-aware content adaptation is required. In the following subsections, a set of walks-through of different use cases are described and the results of using the platform are presented.

Three detailed uses cases involving three collaborating parties were designed. In these use cases, conditions in the context of usage of the involved parties are made to vary along the lifetime of the VC session. The first party (P1) is an individual field worker, who is

Fig. 10 Virtual Office for collaboration



using his smart phone over a mobile wireless network in a public area. The second party (P2) is another individual collaborator, who is in her office, sitting at her desk and using her tablet Personal Computer (PC) with a WiFi connection (through a wireless access point shared with other users). The third party (P3) comprises four individuals (P3I1, P3I2, P3I3, and P3I4) in a company head office, who use a large-scale VC terminal with a big display unit and a wired connection to a Gigabit Ethernet network.

The VC session starts when all three parties are connected to the system. The content adaptation platform gathers contextual information from each party and from the available access networks. It also collects information concerning the multimedia content to be exchanged among the parties to evaluate the required resources. In particular, given the limited capabilities of the terminal device and network connection of P1, the system realizes that it will be necessary to downscale the content to be delivered to this party. Based on all of this information, it selects adequate service parameters, which are passed to the AES, and the VC session then starts (instant t_0). The contextual information required to allow the system to decide and actually perform adequate adaptation operations is listed in Table 3. The characteristics of the original multimedia content are provided in the form of a MPEG-7 MDS, and are presented in Table 4. As can be seen, the video codec supports spatial, temporal SNR scalability features.

Table 3 Contextual information exchanged during the VC session at t_0

Time	Party	Contextual information		
		Source	Parameter	Metadata
t_0	P1	User context	User static preferences	Volume control = 0.85 Number of channels = 2 Audio sampling frequency = 44,100 Hz Bits per sample = 16 Name = John Doe
		Terminal context	Terminal capabilities	Display size = 352×288 Maximum frame rate = 25 fps Audio codecs supported = MP3 Video codecs supported = MPEG-4, H.264/AVC Device class = smart phone
		Network context	Network conditions	Available average bandwidth = 120 kbit/s Delay two way = 330 ms Packet loss rate = 0.001
		Natural environment context	Natural environment characteristic	Location = public Noise level = 25 dB Illuminance = 700 lx
		AES context	AE capabilities	Supported adaptation operations = Spatial scaling, temporal scaling, SNR scaling, ROI cropping Maximum temporal resolution constraint = 100 fps Maximum spatial resolution constraint = 1,440×1,080
		Adaptation authorization context	Adaptation authorization response	Permitted adaptation operation = Spatial resolution scaling

Table 4 Technical characteristics of the original content

Technical characteristic	Parameters
Content	Content type = audiovisual
Media format	File format = AVI
Bit rate	Variable = true; average = 948 kbit/s, maximum = 1,223 kbit/s, minimum = 13 kbit/s
Scalability	Scalability type = spatial, temporal, and SNR
Visual coding	Visual coding format = H.264/AVC scalable extension (H.264/SVC) Frame height = 960, frame width = 512, and frame rate = 25
Audio coding	Audio coding format = MPEG-1 audio layer III Audio channels = 2 Sample rate = 44,100 Hz

4.1 Use case 1—access to protected content in public areas

This use case comprises the initial period of the VC session, immediately after selection of initial service parameters and service initialization, as explained above. In this use case, protected content is exchanged between the VC participants. The participant P1 using a smart phone with a mobile wireless network connection is located in a public area. The system will use low-level contextual information concerning the type of access network being used by the participants, the surrounding noise level and nearby person activity to infer if the participant P1 is located in a public or private area. High noise level and/or activity of persons detected, combined with the identification of the location of the mobile terminal through the Global Positioning System (GPS) assistance, enable the system to deduce that the participant is in a public space.

At one point during the collaboration session (i.e., at instant t_1), P3 shows a company-confidential diagram related to the topic of meeting discussion. Since P1 cannot see visual objects in detail on the small display of the smart phone, he wishes to zoom into the diagram in the video provided to his mobile terminal. This request is passed to the content adaptation platform to confirm if this operation is possible to be performed, given the confidential nature of the information. Based on the contextual information, the system realizes that P1 is in a public area, and therefore it does not authorize the adaptation for further zooming into the objects of interest. Accordingly, upon indication of this event, P1 moves into a private area, and requests the same adaptation once again. In return, the system authorizes the adaptation this time, and the zoom operation is processed for presenting the requested area of attention (i.e., the company-confidential diagram, in this particular example) in the input video. Table 5 represents the variations of the contextual information that drive the system to react and perform the adequate adaptation operation.

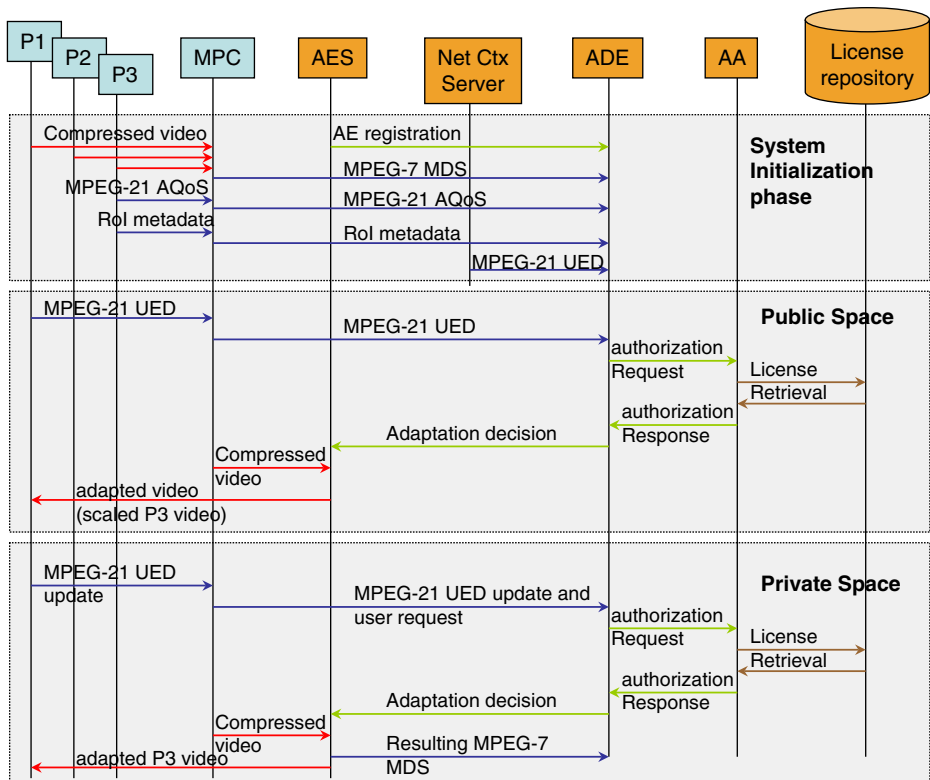
The complete sequence diagram of this first use case is shown in Fig. 11. In this figure, we have identified: 1) the content flow with the red arrows; 2) the content-related metadata flow with the blue arrows; 3) the contextual information flow with the green arrows; and 4) the message exchange between each module with the pink arrows. Multi-Party Control (MPC) unit is a VCS-specific block, which collects all the audiovisual information as well as related content and context based metadata from the participating VC parties. It then provides a composite media feed to the relevant blocks of the content adaptation platform,

Table 5 Contextual information exchanged during the VC session at t_1

Time	Party	Contextual information		
		Source	Parameter	Metadata
t_1	P1	User context	User dynamic preferences	Focus of attention = (464,224;639,367)
		Natural environment context	Natural environment characteristic	Location = private
		Adaptation authorization context	Adaptation authorization response	Permitted adaptation operations = Spatial resolution scaling (Scalable and Non-scalable), ROI cropping (Scalable and Non-scalable)

consisting of the media components and associated metadata. The Network Context Server provides the necessary network context information to the adaptation platform, and thus acts as one of the many CxPs.

Figure 12 presents excerpts of the Authorization Request and Response messages for this particular use case. The messages use a combination of elements from MPEG-21 REL (mx, sx, and r namespaces) and MPEG-21 DIA (dia namespace).

**Fig. 11** The sequence of actions caused by P1 during the VC session in use case 1


```

<r:keyHolder>
  <r:info>
    <dsig:KeyName>urn:user:p1</dsig:KeyName>    <!-- Principal-->
  </r:info>
</r:keyHolder>
<mx:play/>
<mx:diReference>
  <mx:identifier>urn:video:p3</mx:identifier>    <!--Resource-->
</mx:diReference>
<sx:territory>
  <sx:location>
    <sx:country>SP</sx:country>
    <sx:postalCode>private place postal code</sx:postalCode>
    <sx:street>private place address</sx:street>
  </sx:location>
</sx:territory>

<authorisation>true</authorisation>
<permittedDiaChanges>
  <!-- ROI Cropping -->
  <dia:ConversionDescription xsi:type="dia:ConversionUriType">
    <dia:ConversionActUri uri="urn:visnet:AdaptationTypeCS-NS:1.4"/>
    <!-- 1.4 refers to Scalable ROI Cropping -->
  </dia:ConversionDescription>
  <dia:ConversionDescription xsi:type="dia:ConversionUriType">
    <dia:ConversionActUri uri="urn:visnet:AdaptationTypeCS-NS:2.4"/>
    <!-- 2.4 refers to Non-Scalable ROI Cropping -->
  </dia:ConversionDescription>
  <!--Adaptation of the Spatial Resolution -->
  <dia:ConversionDescription xsi:type="dia:ConversionUriType">
    <dia:ConversionActUri uri="urn:visnet:AdaptationTypeCS-NS:1.2"/>
    <!-- 1.2 refers to ScalableSpatialResolutionScaling -->
  </dia:ConversionDescription>
  <dia:ConversionDescription xsi:type="dia:ConversionUriType">
    <dia:ConversionActUri uri="urn:visnet:AdaptationTypeCS-NS:2.2"/>
    <!-- 2.2 refers to Non-ScalableSpatialResolutionScaling -->
  </dia:ConversionDescription>
</permittedDiaChanges>

```

Fig. 12 Excerpts of authorization request and response messages exchanged between the ADE and AA when P1 is in a private place (use case 1)

4.2 Use case 2—poor terminal connectivity

In the meantime, while still engaged in the VC session, P2 starts downloading some video sequences to perform later, simulations of her new object tracking algorithm on her PC (instant t_2). Consequently, she starts experiencing degradation in the image quality of her VC session. The system receives updated information concerning the network conditions, and realizes that the network to which P2's terminal is connected experiences a period of congestion, which leads to video packet losses (i.e., the packet loss rate = 0.20, as indicated in Table 6). Therefore, the system decides switching to a video support with lower resolution and bit rate. When the simulation ends and conditions are set back to normal, it makes a decision for switching back to the original resolution and bit rate. The new contextual information collected by the system at instant t_2 , reflecting the new usage conditions, is presented in Table 6.

Table 6 Contextual information exchanged during the VC session at t_2

Time	Party	Contextual Information		
		Source	Parameter	Metadata
t_2	P2	Network context	Network conditions	Available average bandwidth = 250 kbit/s Delay two way = 550 ms Packet loss rate = 0.20 Congestion max. burst length = 12
			Network capabilities	Maximum capacity = 4,096 kbit/s Minimum guaranteed = 128 kbit/s
			Adaptation	Permitted adaptation operation = Spatial resolution scaling (Scalable and Non-scalable), temporal resolution scaling (Scalable and Non-scalable)
			Authorization	Minimum frame width = 100 Minimum frame height = 50
		Adaptation Authorization	Adaptation authorization response	

4.3 Use case 3—region of interest (ROI) selection

A few minutes later into the VC session (instant t_3), the system detects an update on the user's preferences and decides to select a ROI based adaptation. This decision is based on the fact that P1 had indicated a preference for a viewing mode focusing on the active participant (i.e., the speaker) among his peers (i.e., the four individuals from the P3 video received from the company HQ). Therefore, the system exploits the available speaker identification and tracking support for attaining a VC-specific metadata, named “hasFloor” [4], which is the virtual assignment of a tag to determine active (i.e., talking, animated, etc.) person or persons in the discussion. In this way, P1 is provided with the speaker's head and shoulders view, which happens to switch between the four participants in the output video from P3 depending on who is actively contributing to the collaboration session in progress. Similarly, when P2 starts to lead the discussion, P1's display then switches to show the P2 output instead of P3's. Table 7 represents the variations of the contextual information that drive the system to react and perform the adequate adaptation operation, by delivering an automatic ROI cropping.

Table 7 Contextual information exchanged during the VC session at t_3

Time	Party	Contextual information		
		Source	Parameter	Metadata
t_3	P1	User context	User dynamic preferences	Automatic ROI cropping = true
		Multimedia content analysis	ROIs extraction	ROI_1 = (0,144;175,303), hasFloor = false ROI_2 = (288,78;463,221), hasFloor = false ROI_3 = (512,96;687,239), hasFloor = false ROI_4 = (720,80;895,223), hasFloor = true
		Adaptation authorization context	Adaptation authorization response	Permitted adaptation operations = Spatial resolution scaling, ROI cropping

5 Results

In this section, the results of performing the different selected adaptation operations, as described in the previous section, are compared against those of not performing any adaptation. This allows assessing the benefits in terms of serving user expectations through utilizing the platform in VC applications. During the described excerpt of this VC session, context conditions have suffered from variations, which have been sensed by the system, triggering the necessary content adaptation operations. In these specific use cases, four distinct adaptation operations have been performed at instants t_0 , t_1 , t_2 and t_3 . The corresponding adaptation alternatives and associated service parameters passed to the AEs are indicated in Table 8.

Figures 13, 14, 15 and 16 allow comparison of the results of performing the listed adaptations against not performing any content adaptation at all, at each of the instants when the context of usage conditions change. Thus, the results focus on presenting the adapted and non-adapted videos, which were obtained using the common VISNET II audiovisual database created during the lifetime of the project. The adapted versions in Figures 13, 14 and 15 represent the signal as visualized by participant P1, who is using the smart phone, whereas the adapted version of Fig. 16 is the image as experienced by participant P2 on her tablet PC.

Figures 13 (a) and 16 (a) present a selected frame of the original video sequence. This frame has been scaled down to the same real spatial dimensions of the adapted versions for presentation purposes only. In these figures, the average bit rate of the original sequence is 948 kbit/s, as indicated in Table 4. As such, Figs. 13 and 16 provide a comparison between the high-quality original frame and adapted versions of it, suitable to the sensed context constraints.

Figures 14 and 15 present a comparison between different adaptation possibilities, all of which match the prevailing contextual constraints.

Figure 13 illustrates the adaptation that is performed at instant t_0 (temporal and spatial scaling, as indicated in Table 8). With this adaptation, the bit rate of the bit stream delivered to P1 is decreased by a factor of 1:43.

Figures 14 and 15 present a comparison of different adaptation possibilities, suitable to the contextual constraints experienced at time instants t_1 and t_3 . Figures 14 (a) and 15 (a) show simple bit rate adaptation, enabling to decrease the bit rate of a factor of approximately 1:9, suitable to the network constraints observed at instants t_1 and t_3 respectively. Their corresponding versions depicted in (b) figures show alternative adaptation, clearly more suitable to be applied in VC applications, namely ROI-based adaptation. The different bit rates of all of the adapted versions are listed in Table 8.

Figure 16 illustrates the case when there is network congestion occurring at time instant t_2 experienced by participant P2. Figure 16 (a) presents a frame of the original video sequence at the output of the encoder, whereas (c) shows the same frame as visualized by participant P2 after it has traversed the congested network. Figure 16 (b) presents a spatially and SNR adapted version of the original bit stream at the output of the AES, whereas (d) shows the corresponding adapted frame as received by P2 after transmission through the congested network. This experiment and the associated figures provide the means to demonstrate the negative impact of packet losses on the former set-up, which in turn leads to significant image quality degradation. A direct comparison between Fig. 16 (c) and (d) clearly highlights the benefits of performing adaptation from the user's perspective.

Table 8 Content adaptation operations requested during the VC session

Time	Party	Required content adaptation operations	
		Parameter	Metadata
t_0	P1	Spatial resolution scaling	Spatial resolution before adaptation = 960×512
			Spatial resolution after adaptation = 240×128
		Temporal resolution scaling	Temporal resolution before adaptation = 25 fps
			Temporal resolution after adaptation = 12.5 fps
			Number of Quality Layers before adaptation = 4
			Number of Quality Layers after adaptation = 3
			Expected bit rate after adaptation = 22.4 kbit/s
t_1	P1	User requested ROI cropping	Crop Origin X = 464; Crop Origin Y = 224 Crop Width = 176; Crop Height = 144 Target bit rate = 120 kbit/s
t_2	P2	Spatial resolution scaling	Spatial resolution before adaptation = 960×512
			Spatial resolution after adaptation = 480×256
		Temporal resolution scaling	Temporal resolution before adaptation = 25 fps
			Temporal resolution after adaptation = 25 fps
			Number of Quality Layers before adaptation = 4
			Number of Quality Layers after adaptation = 3
			Expected bit rate after adaptation = 167 kbit/s
t_3	P1	Automatic ROI cropping	Crop Origin X = 720; Crop Origin Y = 80 Crop Width = 176; Crop Height = 144
			Temporal resolution before adaptation = 25 fps
			Temporal resolution after adaptation = 12.5 fps
		Temporal resolution scaling	Target bit rate = 120 kbit/s



Fig. 13 Adaptation results for party P1 at time t_0 : **a**) 25% scaled version before adaptation **b**) non-scaled version of the frame after adaptation

Since the Joint Scalable Video Model (JSVM) decoder is not capable of decoding a scalable bit stream with four spatial resolution layers when there are packet losses, the highest resolution layer of the lossy bit stream (i.e., 960×512 pixels) is discarded before decoding it. Hence the resulting image shown in Fig. 16 (c) has a spatial resolution of 480×256 pixels. Moreover, it should be noted that the congestion is effectively reduced when the adaptation operation is performed, since less data is transmitted through the network. Accordingly, in order to obtain Fig. 16 (d), it is assumed that the packet drop rate is reduced to 0.10 (from 0.20 originally). To perform the above simulations, an Internet Protocol (IP) channel model was implemented using the Advanced Video Coding/Scalable Video Coding (AVC/SVC) loss simulator described in [8] and the ITU-T Video Coding Experts Group (VCEG) [35].

Subjective evaluation has been carried out with the assistance of a heterogeneous audience of 20 people, including both expert and non-expert male/female volunteer viewers. Double stimulus technique was used in these subjective experiments. We started by showing the reference material (original sequences) to the viewers, followed by one of the video sequences as received by the participants (adapted and non-adapted versions). Each viewer compared the processed one with the reference and provided his/her comparative opinion score according to the Mean Opinion Score (MOS) scale (an opinion score with 5 points, where 1 corresponds to the lowest quality and 5 to the highest quality when compared to the original). The results of this evaluation to assess the usefulness of the bit rate adaptation in use case 2 (“Poor terminal connectivity”) and the ROI cropping adaptation in use case 3 (“ROI selection”) are presented in Tables IX and X, respectively. Table 9 provides the average score obtained for the relative quality of non-adapted and bit rate adapted versions of the original sequence in a lossy communication channel condition. Table 10 shows the average scores for a simple bit rate

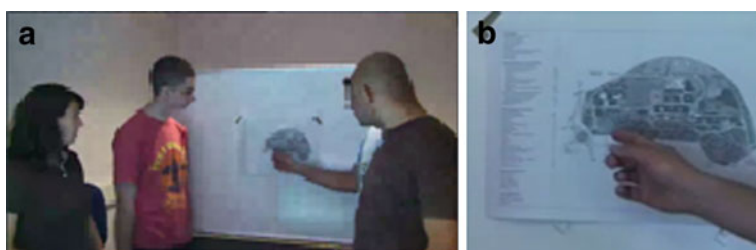


Fig. 14 Adaptation results for party P1 at time t_1 : **a**) 25% scaled version after encoding the sequence for 109 kbit/s bit rate **b**) non-scaled version of the frame after adaptation



Fig. 15 Adaptation results for party P1 at time t_3 : **a**) 25% scaled version after encoding the sequence for 109 kbit/s bit rate **b**) non-scaled version of the frame after adaptation

adapted version and an ROI cropped version. According to the experimental results presented, it is clear that the viewers in this sample tend to prefer the adapted versions and, among the adapted, a version that is adapted in a way specific to the VC application. Assessing the quality improvement from the objective perspective would not be very straightforward, since the existing objective quality metrics only compare contents of similar spatio-temporal resolutions. In addition its value would be minimal, given that in some cases we would be comparing different sequences (particularly while performing a ROI-based or a spatial adaptation).

6 Conclusions

The work described in this paper has illustrated a series of walks-through of the context-aware and DRM-enabled content adaptation platform, highlighting the benefits of

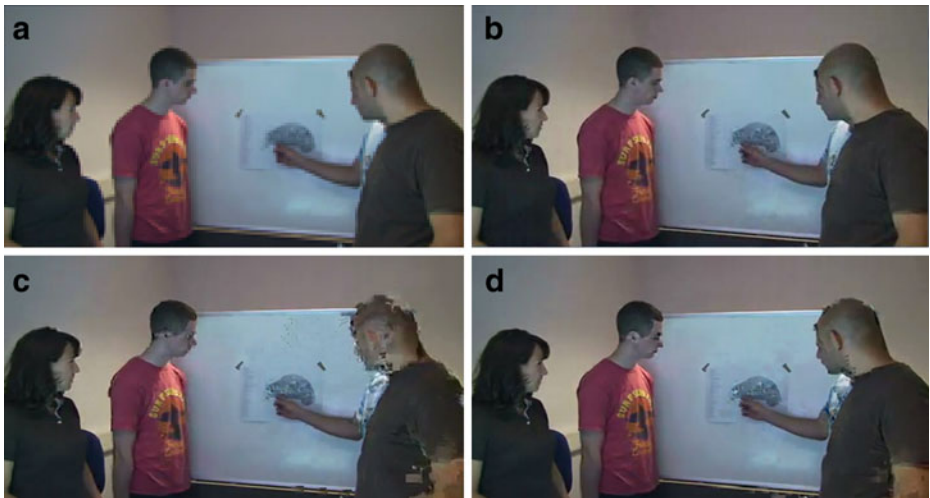


Fig. 16 Adaptation results for party P2 at time t_2 : **a**) 25% scaled version of a selected frame before performing any adaptation **b**) 50% scaled version of the frame as seen at the AE output after adaptation **c**) 50% scaled version of the frame as seen by P2 if non-adapted version of the bit stream is received through the congested network **d**) 50% scaled version of the frame as seen by P2 if the adapted version of the bit stream is received through the network

Table 9 Subjective assessment results for use case 2

Adaptation	MOS out of 5
No adaptation	2.4
Bit rate adaptation	3.5

introducing content adaptation within VC scenarios in a number of different situations. Accordingly, it has demonstrated the applicability and usefulness of the developed platform in multimedia networked applications. The results presented here, which were obtained using typical VC video test sequences, clearly show the improvement in received video quality when conditions deteriorate.

The use of a context ontological model to drive the adaptation decision taking process enables the platform to selectively decide specific types of adaptations adequate to specific conditions of the context of usage taking into consideration the kind of application. The integration of adaptation authorization and licenses' management allows supporting this kind of operations while accessing particularly the protected content. A suite of adaptation algorithms, which are able to respond to the adaptation decisions made in a very flexible and comprehensive way, completes the platform. Globally, this platform is able to fulfill the objective of offering a service without disruption and with a minimum guaranteed level of quality, regardless of the conditions of the consumption environment, while also ultimately enhancing the users' satisfaction of the multimedia services.

The main advantage of the platform comparing to the existing solutions are best described by its ability to combine a number of unique features: 1) adaptations are dynamically selected in a transparent way according to the varying usage conditions; 2) manipulation of different levels of content protection, which directly influence the adaptation decisions; 3) use of emergent standards with ontologies, thus promoting interoperability and enriching the decision process with knowledge-based mechanisms; 4) built in a modular way such that common requirements usually found in all types of networked multimedia applications constitute the core of the system, whereas specific requirements of different applications can be incorporated or plugged as needed. This modular concept applies both to context and content descriptions and related knowledge through the use of different sets of MPEG-7 and MPEG-21 and of the core and extended ontologies, as well as to admissible content adaptation operations: new extended ontologies and new types of content adaptation suiting specific requirements of other applications can be easily plugged-in; 5) suitable for operation in many-to-many communications.

Table 10 Subjective assessment results for use case 3 (ROI cropping concept)

Adaptation	MOS out of 5
Bit rate adaptation	2.5
Cropping	3.8

Acknowledgments The work presented was partially developed within VISNET II, a European Network of Excellence (<http://www.visnet-noe.org>), funded under the European Commission IST FP6 programme.

References

1. Andrade MT (2007) Architectural support for ubiquitous access to multimedia content. Ph.D. Thesis, Faculty of Engineering of the University of Porto
2. Andrade MT, Kodikara Arachchi H, Nasir S, Dogan S, Uzuner H, Kondoz AM, Delgado J, Rodriguez E, Carreras A, Masterton T, Craddock R (2007) Using context to assist the adaptation of protected multimedia content in virtual collaboration applications. In IEEE/ICT Proceedings of the 3rd International Conference on Collaborative Computing (CollaborateCom 2007). New York
3. Burnett I, Pereira F, Walle R, Koenen R (eds) (2006) The MPEG-21 book. Wiley
4. Carreras A, Andrade MT, Masterton T, Kodikara Arachchi H, Barbosa V, Dogan S, Delgado J, Kondoz AM (2009) Contextual information in virtual collaboration systems beyond current standards. In Proceedings of the 10th Int. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2009). London, pp 209–213
5. Carreras A, Delgado J, Rodriguez E, Barbosa V, Andrade M, Kodikara Arachchi H, Dogan S, Kondoz A (2010) A platform for context-aware and digital rights management-enabled content adaptation. *IEEE Multimedia* 17(2):74–89
6. Chaari T, Laforest F, Celentano A (2006) Adaptation in context-aware pervasive information systems: the SECAS project. *Int J Pervasive Comput Commun* 2(2)
7. Dey AK (2000) Providing architectural support for building context-aware applications. Ph. D. Thesis, College of Computing, Georgia Institute of Technology
8. Guo Y, Li H, Wang Y (2005) SVC/AVC loss simulator donation. JVT input document JVT-Q069
9. Horrocks I, Patel-Schneider PF, Boley H, Tabet S, Grosz B, Dean M (2004) SWRL: a semantic web rule language combining OWL and RuleML, W3C member submission, May 2004. [Online]. Available: <http://www.w3.org/Submission/SWRL/>
10. Jannach D, Leopold K, Timmerer C, Hellwagner H (2006) A knowledge-based framework for multimedia adaptation. *Journal of Applied Intelligence*, Special Issue on Innovations in Applied Artificial Intelligence 24(2):109–125
11. Kofler I, Timmerer C, Hellwagner H, Hutter A, Sanahuja F (2007) Efficient MPEG-21-based adaptation decision-taking for scalable multimedia content. In Proceedings of the 14th Multimedia Computing and Networking Conference (MMCN), San Jose
12. Kofler I, Seidl J, Timmerer C, Hellwagner H, Djama I, Ahmed T (2008) Using MPEG-21 for cross-layer multimedia content adaptation. *J Signal Image Video Process (SIVP)* 2(4):355–370
13. Kosch H, Timmerer C (guest editors) (2009) Multimedia metadata and semantic management. In IEEE Computer Society Computing Now
14. Lee K-Ch, Kim J-H, Lee J-H, Lee K-M (2007) Implementation of ontology based context-awareness framework for ubiquitous environment. In Proceedings of the IEEE International Conference on Multimedia and Ubiquitous Engineering, Seoul, Korea
15. Liu D, Meng XW, Chen JL (2008) A framework for context-aware service recommendation. In Proceedings of the International conference on Advanced Computing Technologies (ICACT '08), Hyderabad, India
16. Lopez F, Jannach D, Martinez JM, Timmerer C, Hellwagner H, Garcia N (2008) Multimedia adaptation decisions modelled as non-deterministic operations. In Proceedings of the 9th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'08), Klagenfurt
17. López F, Martínez JM, García N (2009) CAIN-21: an extensible and metadata-driven multimedia adaptation engine in the MPEG-21 framework. In: Chua T-S et al (eds) *Lectures notes in computer science*. Springer Verlag, vol. 5887, pp 114–125, 4th International Conference on Semantic and Digital Media Technologies (SAMT 2009), Graz, Austria, December 2–4, 2009

18. McGuinness DL, Harmelen FV (2004) Web Ontology Language (OWL): Overview. W3C Recommendation. [Online]. Available: <http://www.w3.org/TR/2004/REC-owl-features-20040210/>
19. Mpeg 2010. [Online]. Available: <http://www.chiariglione.org/mpeg/>
20. Mpeg-21 Part 1 2004. Information Technology—Multimedia Framework (MPEG-21)—Part 1: Vision, Technologies and Strategy, ISO/IEC Standard.
21. Mpeg-7 Part 5 2003. Information Technology—Multimedia content description interface—Part 5: Multimedia description schemes, ISO/IEC Standard.
22. Mpeg-21 Part5 2004. Information Technology—Multimedia Framework (MPEG-21)—Part 5: Rights Expression Language, ISO/IEC Standard
23. Mpeg-21 Part7 2007. Information Technology—Multimedia Framework (MPEG-21)—Part 7: Digital Item Adaptation, ISO/IEC Standard
24. Multicao 2010. [Online]. Available: http://www.visnet-noe.org/dokuwiki/doku.php?id=context-based_adaptation:multicao
25. OWL (2004) Web Ontology Language (OWL): Overview, W3C Recommendation. [Online]. Available: <http://www.w3.org/TR/owl-features/>.
26. Pellet (2010) Pellet: The Open Source OWL DL Reasoner. Clark & Parsia. [Online]. Available: <http://clarkparsia.com/pellet/>
27. Preuveneers D, Vandewoude Y, Rigole P, Ayed D, Berber Y (2006) Context-aware adaptation for component-based pervasive computing system. In Proceedings of the 4th International Conference on Pervasive Computing, Dublin, Ireland, pp 125–128
28. Qin W, Shi Y, Suo Y (2007) Ontology-based context-aware middleware for smart spaces. Tsinghua Sci Technol 12(6):707–713
29. Serrano JM, Serrat J, Strssner J (2007) Ontology-based reasoning for supporting context-aware services on autonomic networks. In Proceedings of the IEEE International Conference on Communications 2007 (ICC-2007), Glasgow
30. Timmerer C, Vetro A, Hellwagner H (2008) MPEG-21 digital item adaptation. In: Furht B (ed) Encyclopedia of multimedia, 2nd edn. Springer, pp 457–463
31. Vetro A, Timmerer C (2005) Digital item adaptation: overview of standardization and research activities. IEEE Trans Multimedia 7(3):418–426
32. VISNET II Wiki (2010) [Online]. Available: <http://www.visnet-noe.org/dokuwiki/>.
33. W3C 2010. World Wide Web Consortium [Online]. Available: <http://www.w3.org/>
34. Wang X, DeMartini T, Wragg B, Paramasivam M, Barlas C (2005) The MPEG-21 rights expression language and rights data dictionary. IEEE Trans Multimedia 7(3):408–416
35. Wenger S (2002) Error patterns for internet experiments. ITU-T VCEG Q15-I-16r1



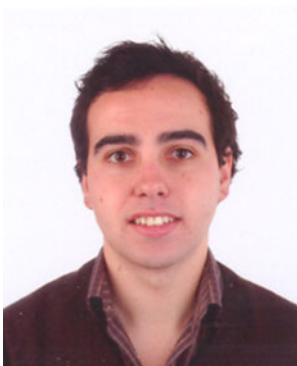
Maria Teresa Andrade is an Assistant Professor at the Electrotechnical and Computing department of the Faculty of Engineering, University of Porto. She coordinates the Multimedia Systems group at INESC Porto. Her main interests are multimedia networking, content adaptation and context-awareness. Author of more than 40 published papers. Received her M.Sc. and Ph.D. degrees from the University of Porto.



Safak Dogan is a Senior Research Fellow at I-Lab, University of Surrey with M.Sc. and Ph.D. degrees in Satellite Communications (1996) and Multimedia Communications (2001), respectively. His research interests include video adaptation, transcoding, context-based content adaptation, low bit-rate video communications, error-resilience, and multi-party conferencing. He is a member of the IEEE and the EC Future Media and 3D Internet-Task Force.



Anna Carreras obtained her B.Sc. degree in Telecommunication Engineering in 2006 and her M.Sc. degree in Computer Architectures, Networks and Systems in 2008, both from Universitat Politècnica de Catalunya (UPC), Barcelona, Spain. Member of the Distributed Multimedia Applications Group (DMAG) since 2005. Her research interests include Digital Rights Management, and the use of context and metadata in distributed multimedia applications.



Vitor Barbosa is a research fellow at INESC Porto in the Telecommunications and Multimedia Unit. He obtained his M.Sc. in Electrical and Computers Engineering from Faculdade de Engenharia da Universidade do Porto (FEUP), Portugal, in 2008. His current research interests include multimedia systems and supporting technologies, multimedia metadata modelling, ontology engineering, Semantic Web, and context-aware content adaptation.



Hemantha Kodikara Arachchi received his B.Sc. (Eng.) degree with honors (1997), M.Phil. degree in Electronic and Telecommunication Engineering from University of Moratuwa, Sri Lanka (2000) and Ph.D. degree in Telecommunications from Asian Institute of Technology, Thailand (2004). Currently, he is a Research Fellow at I-Lab, University of Surrey, UK. His research interests are in video coding and context-aware content adaptation.



Jaime Delgado received his Ph.D. degree in Telecommunication Engineering in 1987. Professor at the Computer Architecture Department of the UPC. Head and founder of the Distributed Multimedia Applications Group (DMAG). Project Manager of European and national research projects. Editor of International Standards. Evaluator and reviewer for the European Commission and several Spanish Ministries. Author of hundreds of papers and books.



Ahmet M. Kondoz received his M.Sc. and Ph.D. degrees in Telematics (1984) and Communication (1986), respectively. He is a Professor in Multimedia Communication Systems at University of Surrey and the Head of I-Lab. He has graduated more than 40 Ph.D. students, and has been a consultant for major technology companies. He is a member of the IEEE and IET B.Eng.