Future Media Internet Research Challenges and the Road Ahead

April 2010

Created by the "Future Media Internet-Task Force"

Coordinated by the nextMEDIA Coordination Action and supported by the Networked Media Unit of the DG Information Society and Media of the European Commission

European Commission
Information Society and Media

1 Executive Summary

The Media Internet supports professional and novice content prosumers and is at the crossroads of digital multimedia content and Internet technologies. It encompasses two main aspects: Media being delivered through Internet networking technologies (including hybrid technologies) and Media being generated, consumed, shared and experienced on the web. The Media Internet is evolving to support novel user experiences such as immersive environments including sensorial experiences beyond video and audio (engaging all the human senses) that are adaptable to the user, the networks and the services.

This White Paper reflects the consolidated opinion of 25 experts from the EU, USA and Korea on aspects related to the Future Media Internet (FMI), under the guidance of the Networked Media Systems Unit of the Information Society and Media Directorate General of the European Commission. More specifically, this paper describes the research challenges provisioned by the experts for the upcoming years, concerning the FMI, along with the potential impact these challenges might have.

The White Paper focuses more on the Media side of FMI without neglecting, however, the importance of the novel network-related characteristics that Future Internet will embed. At a glance, the experts believe that the FMI should involve high quality research concerning:

- Scalable multimedia compression, transmission, concealment;
- Network coding and streaming;
- Content & context fusion for improved multimedia access;
- 3D content generation leveraging emerging acquisition channels.
- Immersive multimedia experiences;
- Multimedia, multimodal & deformable objects search;
- Content with memory and behaviour.

The aforementioned research challenges are believed to be the main pillars towards exploring new possibilities for boosting individual and social creativity and productivity in a Future Media Internet environment. The ultimate result of such a research is expected to enhance our communication experiences, the way we work and as a consequence, the way we live.

2 Introduction

Internet is nowadays by far the most important vehicle for data exchange and the world-wide-web the most important source of information as it is reflected by recent statistics, such as: i) the Web processes 100 billion clicks per day and offers 55 trillion links between Web pages, ii) 2 million of emails are exchanged per second and iii) 1 million instant messages are exchanged per second[1]. Also, there is a growing penetration of Internet connectivity in terms of geographical size and connected users, while the users' habits also change rapidly, shifting from "occasionally connected" to "always connected". Last but not least, the Internet infrastructure itself is growing in geographical distribution, number of network elements and heterogeneity of physical media connectivity. It is expected that, by 2011, about 3 billion hosts[2] will be connected to the Internet from the 570M of hosts connected in July 2008[3]. All these facts show that the Internet has become the core communication environment not only for business relations but also for social and human interactions. Internet is providing the society the mechanisms to create new forms of social, political and economical activities.

As such the Internet plays a crucial role in the ability of humans to communicate, but at the same time opens new challenging problems. As the Internet grows beyond its original expectations (as a result of increasing demand for performance, availability, security, and reliability) and beyond its original design objectives, it progressively reaches a set of fundamental technological limits and is impacted by operational limitations. Considerable effort has already been devoted world-wide to defining options and concepts, which could form the basis of Future Internet to support a sustainable society. The dimensions of the Internet by and for People, the Internet of Media and Content, the Internet of Services, the Internet of Things supported by advances in Network Infrastructure will form the Future Internet.

The 'Future Media Internet - Task Force' (FMI-TF) is currently engaging in a brainstorming session in order to provide experiences and the vision to pave the way towards the FMI. The vision of the FMI-TF is that the Future Media Internet requires revolutionary steps. Yet, also due to the world-wide economic crisis, backwards compatibility is crucial. Thus, in this white paper, we aim to identify research challenges

[1] Kevin Kelly, "The Future of the Internet," Wired Magazine

[2] Devices that use the communications infrastructure, including smart phones, mobile phones and other types of handheld devices

[3] "Future Internet," The Cross-ETP Vision Document, Version 1.0 - Date: 8/01/2009

and the way ahead towards providing media and rich multimedia and multimodal content via the Internet in the next decade[4].

This white paper is the result of the FMI-TF meeting that was organised by the coordination action nextMEDIA (ICT-249065) in Brussels on 5 February 2010. During the meeting, scientists and professionals from Europe, USA and Korea, presented their vision on the Future Media Internet, the research challenges that they considered as requiring the highest attention by the research and professional community and the steps that should be followed. After the presentations, each participant voted for the 1st, 2nd and 3rd most important aspect. The work continued off-line with contributions and editing of the white paper.

The white paper is organised as follows: Chapter 2 presents the state-of-the art on the identified research challenges and explains the limitations of the current Internet. In Chapter 3 the provisional research challenges are deeply analysed. Chapter 5 presents potential FMI applications and expected impact to the community. Finally, conclusions are drawn in the last Chapter of the white paper.

3 Current situation in selected areas of Media Internet

In this chapter we describe the current situation in selected areas that have been highlighted by the FMI-TF members. It should be noticed that the purpose of this chapter and the white paper in general, is not to give an exhaustive and thorough state-of-the art in each and every area. Other areas should not be excluded. Yet, we highlight the areas that have been voted by the FMI-TF as the most important ones and give the interested readers 'food for thought' for their further research.

3.1 Scalable multimedia compression, transmission, concealment

Multimedia applications involving the transmission of video over communication networks are rapidly increasing in popularity. These applications include, but are not limited to multimedia messaging, wireless and wired Internet video streaming, video telephony and video conferencing, and cable and satellite TV broadcasting. In general, the communication networks supporting these applications are characterised by a wide variability in packet loss, delay, and throughput. Furthermore, a variety of receiving devices with different resources and capabilities are commonly connected to a network. In this context, coding and transmission technology, able to engineer the content to meet demanding application requirements, is critical. As a consequence,

[4] This white paper can be considered as a continuation of previous work of the FMI-TF: "Research on Future Media Internet," January 2009, Future Media Internet Task Force, ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/netmedia/research-on-future-media-internet-2009-4072_en.pdf

methods and models for scalable media compression, transmission and error concealment, play key roles. These aspects are outlined in this section.

Scalable Video Coding

Scalable Video Coding (SVC) is a highly suitable video storage and transmission system designed to deal with the heterogeneity of the modern communication networks. A video bit stream is called scalable when parts of it can be removed in such a way that the resulting substream forms a valid bit stream representing the content of the original with lower resolution and/or quality.

Traditionally, providing scalability has coincided with significant coding efficiency loss and decoder complexity increase. Primarily due to this reason, the scalable profile of most prior international coding standards such as H.262 MPEG-2 Video, H.263, and MPEG-4 Visual has rarely been used. Designed by taking into account the experience with the past scalable coding tools, the newly developed Scalable Extension of H.264/AVC provides superb coding efficiency, high bit rate adaptability, and low decoder complexity. In this report the term SVC is used interchangeably for both the concept of scalable coding in general and for the particular design of the scalable extension of the H.264/AVC standard.

Scalable Video Transmission

Most modern communications channels (e.g., the Internet or wireless channels) exhibit wide fluctuations in throughput and packet loss rates. Bit stream adaptation in such environments is critical in determining the video quality perceived by the end user. It is attained in SVC by deliberately discarding a number of packets at the transmitter or in the network before reaching the decoder, such that a particular average bit rate and/or resolution is reached. A bit stream extractor is required for determining the relative importance of the bits or packets.

In addition to bit rate adaptation, packets may be lost in the channel (due, for example, to excessive delay or buffer overflow) or arrive erroneous at the receiver and therefore have to be discarded by the receiver. A direct approach in dealing with excessive channel losses is to employ error control techniques. The optimum video quality, however, is obtained when a combination of source optimization techniques well-integrated with error control techniques are considered in a cross-layer framework. The benefits of a cross-layer design are deemed to be more substantial for scalably encoded video since it usually contains various parts with significantly different impact on the quality of the decoded video.

This property can be used in conjunction with unequal error protection (UEP) for efficient transmission in communication systems with limited resources and/or relatively high packet loss rates. By using stronger protection for the more important information, error resilience with graceful degradation can be achieved up to a certain level of transmission errors. The problem of assigning UEP to scalable video is more complex than the corresponding one for non-scalable video. In this context, SVC emerges as the natural choice for highly efficient Joint Source Channel Coding with UEP.

Error Concealment

Video error concealment is a topic that has been studied extensively. It has been treated as a post-processing step and therefore outside the specifications of the standards. SVC non-normatively adopts four picture-loss concealment methods: picture copy, temporal direct motion vector generation, motion and residual up sampling, and reconstruction base layer up sampling. Since each method exploits information from different parts of the video sequence, the quality of the concealed picture varies accordingly. It is advantageous when the concealment strategy is taken into account by the sender during resource allocation and scheduling.

Although significant advances have been made in all of the above topics, critical problems remain open as discussed later in this report.

3.2 Network coding and streaming

Network Coding is an emerging paradigm for media and generic data communication in complex networked environment as the Internet. It deals with the optimization of information flows in a network. Its implication in the design of the *Future Media Internet* is crucial since it has the potential to dramatically increase network transmission capability. In conventional communication networks such as the Internet, media delivery is achieved through routing. Network coding offers a more general approach by assuming that, in advanced future media networks, nodes can process and code media streams. Network coding complements source and channel media coding. While source coding aims at compressing media by exploiting information redundancy, channel coding adds patterns of redundancy into the media stream in order to lower the error rate during transmission. On the other hand, Network Coding allows nodes to perform non-trivial operations on the media stream advancing the node capability from simple routing to “Network Coding”. Thus, Network Coding complements source and channel coding to overcome current practical and theoretical limits of media communication networks.

In the current Internet, media is delivered or transmitted from a network node to another according to optimized routing matrices. However, routers are very simple network components which are able to just store and forward data. Network coding can be seen as a generalization of routing in which nodes can encode media streams by computing given functions of previously received input data. In a sentence, it allows information to be processed at the nodes and by doing so it brings significant benefits in terms of throughput, reliability and fault tolerance.

Recent work in Network Coding includes performance analysis in peer-to-peer communications and complex linear and nonlinear coding at the nodes. The latter assumes greater computational power at all of the nodes. More practical schemas for dynamically changing networks are also emerging. Here networks with changing structure, topology and demands change at short periods of time. All these studies are bringing Network Coding closer to practical applications and increasing the prospect of using underlying approaches in core technology for Future Media Internet communications.

3.3 Content & context fusion for improved multimedia access

Content and context fusion technologies have been used in several environments such as TV, mobile media, media mobile search, etc. In the general term we can divide this section into two different parts: technologies to capture the user behavior and profile, and the systems to perform content and context fusion.

User profile and context capturing

Contextualization of the content has been focused on user profile and behavior modeling and context capturing techniques. In the user profiling and behavior modeling there are several techniques and algorithms in the literature. Automatic electronic data gathering and analysis techniques have been used in Internet environments normally for Internet usage measurement, which allow to create a user's profile through feedback analysing the user's behavior.

Other widely deployed systems are the recommendation systems for context-aware content which are based on content filtering tools. Techniques such as content-based filtering, collaborative filtering and hybrid methods use the information on content consumption as the input and feedback for their system characterization. Collaborative filtering is interesting for its applicability to different media delivery scenarios such as delivery over the Internet.

In addition there are several ways for capturing the user context. In mobile environments, the use of geo-localization techniques is perhaps the most widely deployed. Research on annotation of the different dimensions of the contextual information (time, space, culture, language, implicit and explicit personal experience, education, etc), as well as their temporal evolution, is one of the focus of the research in context services, although the research is still mainly in the labs than in commercial services.

Systems for content and context fusion

One relevant advance for content and context fusion is the interaction using HCI devices allowing for a more immersive TV experience by capturing the user context by means of gesture recognition (e.g. the Keanu solution developed as partnership between Orange and Softkinetic, and which is expected to be integrated in STBs). Other approaches are focusing on the analysis of TV remote control usage, rather than gesture analysis, and aim to detect the different usage patterns and to distinguish between household users.

RFID (Radio Frequency Identification) readers in STBs have also been proposed in order to personalize users' services. Each user's rights would be stored in an RFID tag to be hold by the user, while that RFID tag would authenticate each user with respect to his/her authorized and personalized services (e.g., parental control and payment services). This approach could help acquiring consumer's feedback and inputs from each user.

In MyNews, a broadcast TV news personalisation project, users express their preferences over a sample of news taxonomy by rating taxonomy concepts in. The taxonomy is divided into three sections, categories (sport, economic, politic, etc.), locations (America, Africa, Europe, etc.), and people (Bill Gates, Barack Obama, etc.). This approach could build the user's profile, and then content recommendation and services' adaptation could take place based on the profile.

Approaches for services' adaptation to user's preferences and contexts have also been proposed for adapting a videoconference service by choosing the TV or the laptop as a communication device according to user context and policies (e.g. use TV when no one is using it, otherwise the laptop). We notice that profiling and personalisation technologies rely on explicit or implicit identification of the user or the group profile to offer them a personalised experience (like for content recommendation). However, with the diversity of services offered by an IPTV environment (EPG 'Electronic Program Guide', VoD, IM 'Instant Messaging', etc.), a single user or group may need/want to define different identities and associated profiles to use with the various services and circumstances while preserving their privacy.

An alternative environment where content and context fusion is readily available for the user is within mobile phones. Such devices are nowadays truly multimedia devices, incorporating many sensors like cameras, accelerometers, microphone, GPS (for geo-localization) and compass, and are connected to their environment via GSM/3G, wifi or bluetooth. Mobile phones are always carried around by their users and can help them prosume (consume and produce) multimedia by taking into account their context (location, activity, who the user is next to, etc.) and their content (what they are capturing or consuming).

3.4 3D content generation leveraging emerging acquisition channels

Content creators always look for new forms and ways for improving their content and adding new sensations to the viewer experience. High Definition video has been the latest innovation in the area of content enrichment. 3D is the next single greatest innovation in film-making. Recent film releases such as 'Avatar' have revolutionalised cinema by the extensive use of 3D technology and 3D content production along with real actors creating a new genre at the outset of the 2010s. The box office tickets show that audiences have very quickly embraced this by making Avatar the most successful in cinema ticket sales film in the history of digital cinema (beating even the Titanic).

However, today's young society is becoming increasingly content art and design "literate" as a result of technological advances and lower costs in photography, cinematography, 2D/3D graphics design and animation technologies and as a result of increased emphasis on media design in education. As a result, novel forms of 3D content, should also find its way into small and medium size content creation companies, moving the experience from cinema halls and cinema projectors to the everyday household environments and computers, providing increased number of audiences with a taste of the versatility and power of 3D as both consumers and producers. 3D Internet is a concept that has recently come into the spotlight in the R&D arena, catching the attention of many people, and leading to a lot of discussions. Several research challenges such as: visualisation and representation of information, and creation and transportation of information, among others, will need to be investigated and solution found for 3D internet to become a reality.

The success of 3D cinema has led to several major consumer electronics manufacturers and broadcasters to announce plans to launch 3D-capable TVs and offer 3D content in 2010. 3DTV will require the integration of a diversity of key technologies from computing to graphics, imaging to display, and signal processing to communications. There are a number of competing 3D technologies available, and the decision to support 3D will require an understanding of the relative merits of each in the context of the home. The provision of 3D content into the home will require significant cooperation between content providers, service providers and consumer electronics manufacturers to ensure consumer confidence in the technology and avoid a repeat of the confusion surrounding the introduction of HD technologies.

Today's 3DTV technology is based on stereo vision where left and right eye images are presented to the viewer through temporal or spatial multiplexing by wearing a pair of glasses. Usually the content is captured using two cameras mounted on a rig. Recently there are few consumer electronics manufacturers that provide a single camera set up for the capture of the left and right eye images. The next step in the 3D TV development could be the multiview autostereoscopic imaging system, where a large number of pairs of video signals are recorded and presented on a display that does not require glasses for viewing. Although, several autostereoscopic displays have been reported, there are still limitations on resolution and viewing position. Furthermore, stereo and multiview technologies rely upon the brain to fuse the two disparate images to create the 3D sensation. As a result such systems tend to cause eye strain, fatigue and headaches after prolonged viewing as users are required to focus to the screen plane but converge their eyes to a point in space, producing unnatural viewing. With recent advances in digital technology, some human factors which result in eye fatigue have been eliminated. However, some intrinsic eye fatigue factors will always exist in stereoscopic 3D technology.

Creating a truly realistic 3D real-time viewing experience in an ergonomic and cost effective manner is a fundamental engineering challenge. Future 3D technology should seek to advance the current existing technologies not only in capturing and manipulating 3D content but also in creating a new 3D content format which offers fatigue free viewing with more than one person independently of the viewer's position. 3D holoscopic and holography are two technologies that overcome the shortcomings of stereoscopic imaging, but their adoptions for 3D TV and 3D cinema are still in their infancy. Holographic recording requires coherent light but offer the ultimate 3D viewing experience. Holoscopic video uses microlens arrays to recording a 3D scene and can operate under incoherent illumination, which is in contrast with holography, and hence it allows more conventional live capture and display procedures to be adopted. Future 3D video could use different technologies for 3D content creation and display.

3.5 Immersive multimedia experiences

Traditionally, research on multimedia has provided information to the user primarily through just two sensory modalities - sound and vision. Hence up to now, research on different aspects involving multimedia from coding, to transmission to evaluation of the quality of experience and interaction has mainly been focused on these topics. Although there is a growing interest in other types of sensory interaction such as haptics, sound and in general more immersive experiences of multimedia, the topic is still under researched. This statement is even more valid for research on aspects related to the integration, display and transmission of multisensory information enriching the multimedia experience.

Research on immersive multi-sensory environments has proven that taking into account multisensory data such as vibro-sensory (e.g. floor vibration) and low-frequency subsonic effects could improve distance communication in applications like remote music performance and telemedicine. However, still no advanced strategies for data compression and transmission of these alternatives modalities have been adopted. Other research is focusing on alternative and novel ways for interacting with media, e.g. through tangible interactions and use of natural elements.

Regarding the development of new user interfaces, recent research on immersive displays has made relevant progress related to instruments for surround-view and high definition cinema for designing more engaging immersive experiences. Other approaches of immersion consider a shift from highly structured settings of interfaces (like virtual reality and large screens) to more portable ones such as mobile devices augmenting the real world. An important research aspect in all these new interactive systems is pushing interaction designers to exploit the intrinsic characteristics of immersiveness to improve intuitiveness in interaction to enhance user experiences.

Nowadays, the research on immersive interactions has provided a series of applications at the level of the state-of-the-art with high potential of providing social benefit to users and better life quality. Examples of these are in the field of therapeutic systems and rehabilitation, exertion games, connectedness, locative media for education and digital TV among others. To not underestimate is the important issue that is rising among social scientist regarding immersive environments is the perception and relevance of aesthetics in such environments or more critical, the social implications of these new applications amongst which potential addictions to technology-based experiences.

Other relevant work relies on the concept of Social Immersive Media. The notion of this concept has been formalized recently[5] in the context of multimedia immersive

[5] S. Snibbe, H. Raffle, 2009. Social immersive media: pursuing best practices for multi-user interactive camera/projector exhibits. In Proceedings of the 27th international Conference on Human Factors in Computing Systems. Boston, MA, USA, April 04 - 09, 2009.

museum exhibits, by extrapolating a trend enabled by advances in the technology of interacting games, direct manipulation interfaces, and interactive arts.

3.6 Multimedia, multimodal and deformable objects search

Multimedia content, which is available over the Internet, is increasing at a rate faster than the respective increase of computational power and storage capabilities. Internet capacity will approach the amount of yota (10^{24}) bytes in 2010. Such a tremendous amount of content cannot be processed and indexed by the current computational power unless personalised and user-centric mechanisms are implemented so that only the content of interest is delivered to the end-users. This growth of popularity of media is not accompanied by the rapid development of media search technologies and the existing solutions still lack several important features, which could guarantee high-quality search services and improved end-user experience. These features are listed below:

Multimodal content search and retrieval in a unified manner

Currently, information is perceived, stored and processed in various forms leading to vast amounts of heterogeneous multimodal data (ranging from pure audiovisual data, to fully enriched media information associating also data originating from real world sensors monitoring the environment, etc.). User perception and interpretation of the information is in most cases on a conceptual level, independently of the form this content is available. Assuming the availability of an optimal, user-centric, search and retrieval engine, when users search for content they should be able to:

- express their query in any form most suitable for them;
- retrieve content in various forms providing the user with a complete view of the retrieved information;
- interact with the content using the most suitable modality for the particular user and under the specific context each time.

Sophisticated mechanisms for interaction with content

Secondly, social and collaborative behaviour of users interacting with the content should be exploited at best, which will enable them to better express what they want to retrieve. We need to pay a great research attention on increasing the content utilization efficiency, measured as the fraction of the relevant delivered content (i.e., content which satisfies their information needs and preferences) over the total amount of delivered content. Towards the direction of delivering relevant multimedia content to users, another barrier that needs to be overcome is that the vast amount of information is not actually annotated.

Efficient presentation of the retrieved results

Search results suffer in most cases from the sequential presentation and the information overload, i.e. the presentation of huge amounts of information which is in most cases irrelevant to the query or not optimally presented to the user. This becomes significantly more important when information search and retrieval is performed from mobile or notebook devices with limited presentation capabilities. Appropriate filtering of the retrieved results is needed combined also with advanced visualisation techniques to compress information utilising the visual space and novel interfaces for fast and easy information access, based on context aware information, such as location and device specific performance indicators.

Efficient methods for deformable objects search and retrieval

In most of the multimedia object retrieval approaches presented so far, search is performed by using as query a similar object. Low-level feature extraction methods are applied to the object, providing a description of the object's global shape. These methods are not sufficient for Future Media Internet applications for three main reasons:

- An input multimedia object is not always available and it cannot be created from scratch by a non-expert user. On the other hand, using as query an image/video or a hand-drawn sketch is more convenient (e.g. a 2D photo can be taken from the mobile device's digital camera; a sketch can be drawn by using a PDA touch screen interface, a low-quality audio excerpt can be recorded from the mobile phone, etc.).

- Neither partial matching nor articulation invariance is supported by methods applied to the (3D) objects' global shape. Articulation invariance is essential for applications where deformable objects exist (usually fashion applications).

- Presently introduced multimedia search engines commonly do not allow the user to form complex and multi-modal queries. While the multimedia object retrieval is becoming popular is of utmost importance to allow the user to formulate query which consist not only of a multimedia object, but at the same time –a textual or voice description.

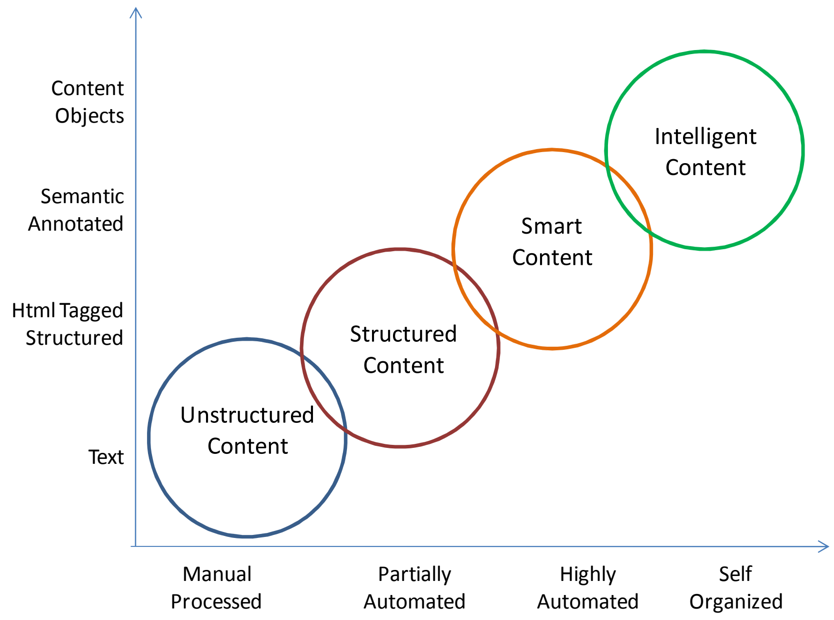
Methods for measuring Quality of Experience

The introduction of more complex search methods than the users are used to create a new problem – the user perceived satisfaction and the method of its measurement. When issuing a multimedia query it is hard to judge whether the result is relevant, and, furthermore whether the user will be satisfied with such result. The situation encountered here is much more complex than with traditional textual queries, as the opinion regarding the relevance of the query result will vary from user to user depending on the query context.

The information on the level of satisfaction of the user with a search service is critical, both during the development and deployment stages. A common methodology and toolbox for measurement of the Quality of Experience of search services would be also an aid for the academic community allowing for comparison of results achieved by different search engines.

3.7 Content with memory and behaviour

In the area of 3D, virtual worlds and gaming, advances are needed to increase the level of realism and interactivity. By adding to virtual characters and virtual objects memory and behavior will lead to the transition from the “smart content” to the “intelligent content”. The positioning and the advantages of the intelligent content may be seen in the following figure[6]



Intelligent content positioning

We define as *Content Objects* polymorphic/holistic containers, which may consist of media, rules, behaviour, relations and characteristics or any combination of the above. *Media* can be anything that a human can perceive/experience with his/her senses, *characteristics* are meaningfully descriptions of the object, *Rules* refer to the way an object is treated and manipulated by other objects or the environment (discovered, retrieved, casted, adapted, delivered, transformed, presented), *behaviour* can refer to the way the object affects other objects or the environment, *relations* between an object with other objects can refer to time, space and synchronisation issues.

As can be seen the integration of memory and behavior to content objects can in some sense be said to be the equivalent of personality for virtual characters. The behaviour rules their interactions, while memory provides them with the capacity to remember past interactions and to learn from their experiences. As a result these virtual characters can interact socially and serve as companions for people and even potentially amongst each other. While the level of physical realism of these characters has

[6] This figure is an extension of the original figure provided by Dale Waldt in the article: “What is Smart Content?” available on-line at http://gilbane.com/xml/2009/11/what-is-smart-content.html#ixzz0kKNBPF6h

vastly improved (improved movement, more realistic facial gestures, etc.), the biggest challenge not only making them look like humans but make them behave as natural as humans for which they must have social and cognitive intelligence, emotions and memory.

To date this level of realism has not been achieved in any virtual worlds or games. The only characters which, today, have such a high level of social and cognitive skills are those which have a human being behind them. But providing emotions and personality for virtual characters isn't just a challenge for virtual worlds and gaming – it can have real world benefits in areas such as education and health care.

4 Future Media Internet - Research Challenges

This section presents major challenges in the area of Future Media Internet as they were identified by the experts. It should be noted that these were voted by the experts as the most popular ones.

4.1 Scalable multimedia compression, transmission and concealment

Despite substantial progress in scalable coding technology as outlined in Section 3.1 there are still several key areas of research that require a deeper, systematic and focused approach. It will be important to leverage added value fundamental research to solve open issues, specifically in the context of scalable media coding. Important research areas relate to high granularity scalable extensions of available coding methods and models. A major research direction in this area is content and context adaptive cross-layer optimization of resources for controlling the rate versus quality of experience trade-off. In addition, the development of scalable (and non-scalable) algorithms and standards for compressing, transmitting, and concealing multisensory information is a completely unexplored research area. Some of the major research directions are outlined below.

Scalable multimedia compression

Developing scalable multimedia compression algorithms and standards is an important topic for research and development for the Future Internet. Such algorithms and standards should be compatible and competitive with their non-scalable versions. They should also be computationally efficient, guaranteeing their wide adoption and proliferation. The scalable compression of other than video media and modalities, such as, multi-view video, 3D meshes, haptic information and multisensory data, is in its infancy. Once an algorithm has been established, the efficient and accurate bit-stream extraction (i.e., given a scalably encoded bit stream provide the substream that will meet a rate constraint and provide the best possible quality) is an important research and development topic. Such an extraction provides a prioritization of the bits in the medium, which is essential for their content aware transmission.

Scalable multimedia transmission

To enable seamless streaming, one of the major challenges is the bit rate adaptation when the bandwidth gap between two different networks is large. In heterogeneous networks, the available channel bandwidth usually fluctuates in a wide range according to the type of network. Thus, a bit stream that inherently has bit rate scalability is essential. Content and context aware transmission of scalable multimedia is essential in guaranteeing the maximization of the utility or the experience for the user. Optimization across the layer stack is required in order to optimally allocate all the available resources. The inherent coupling of the control parameters across transmission layers, as well as, the coupling of the control parameters across enhancement layers makes this a very challenging problem. The bit-rate extraction problem is also challenging and crucial in determining the priority of the packets, based on their contribution to the overall quality. As already mentioned, the (scalable) compression and transmission of media other than video and audio requires special attention if we are to materialize some of our overall objectives, such as content and context fusion of multimedia for improved access and immersive multimedia experience. Finally, the problem of scalable multimedia transmission needs to be considered for various network topologies, such as, broadcast, multicast and P2P networks.

Scalable multimedia concealment

There are numerous video concealment techniques in the literature. Many error concealment techniques have also been proposed for HVC. Basically, these methods have been used to primarily reconstruct corrupted frames of non-scalable bitstreams. Although error concealment can be applied directly to a scalable bitstream for each layer separately, it may produce unsatisfactory results since the interlayer correlation is not considered. In the case of a scalable bitstream, a corrupted frame can be recovered using information from the lower layers as well as the current layer. Thus, error concealment needs to be performed by considering this inter-layer correlation. The development of effective methods for solving the error concealment problem in SVC is one of the challenges in this area. In addition concealing losses in a scalably encoded bitstream for other media (e.g., multiview video, 3D meshes, haptic information) is a wide open research problem. Finally, it is worth considering scenarios for which the encoder has knowledge of the concealment strategy followed by the decoder. This information is quite valuable in determining the priority of each packet by considering its content (i.e., a packet that is “easily” concealable should be given low priority, or might even be skipped).

4.2 Network Coding

Routing in networks operates in a similar manner as the transportation problem. Here the aim is to transport a given commodity in a cost-efficient fashion. In the case of media communication over the Internet the goods to be transported are packets of bits representing multimedia information. Conventionally, data is compressed and recovered at the edges. Cost is defined according to a given cost of routes or by adjusting to the flows. More importantly this conventional model does not generally make use of the fact that the commodity being transported is made up of bits representing chunks of media. The latter is a key aspect since bits can be manipulated, coded and transmitted in a fashion that is very different from other less operable commodities. The fact that in current Networked Electronic Media approaches, routing in networks operates in a manner akin to the transportation problem opens a very promising research field.

Here, the properties of the media and the fact that bits can be operated on, is exploited to achieve significant improvements in network throughput, reliability and fault tolerance. Current nodes or routers just store and forward packets. Network Coding aims at adding "*intelligence*" and computational power to the nodes. It enables nodes to execute coding operations on the actual content. Thus, it facilitates the development of novel advanced communication system architectures and transmission models. These models are more suitable for a Future Media Internet where network capacity and better performance is required. Indeed, Network Coding has the potential to unleash technologies to help re-designing of future networks and substantially improve their performance, since it allows to trade-off communication capacity against computational costs.

Although the potential of the Network Coding paradigm is great, there are still many open problems that prevent a successful immediate deployment of derived technology in real-world applications. The following four examples of selected key research challenges outline the need for focused research in Network Coding to fully exploit its potential.

Join Source-Channel-Network Coding

In the same way source and channel coding can be designed to support each other, a third aspect can be considered (Network Coding). Though, the relationship between channel and Network Coding appears to be obvious, there are many open problems to overcome successful implementation in an optimized triple synergetic approach involving source, channel and network coding.

Error Resilience

Network coding techniques are based on performing coding operations at intermediate nodes. If each receiver node knows the coding functions and the network topology, then error-free decoding is possible. However, this assumes that the transmission is error-free too. However, the assumption of error-free networks does not hold in real networks and practical applications. Thus, error-correcting techniques are needed since even a single error has the potential to endanger decoding success at receiver nodes. Though significant progress has been made in theoretical aspects of Network Coding and error resilience, additional effort is needed to fully exploit theoretical results in practical applications.

Synchronization and Network Coding

The basic Network Coding example assumes that there is synchronization between the network nodes, and each node performs fixed encoding operations. Both, sender and receivers are assumed to know these operations, and use this knowledge to de-code the streams. However, in a practical sensor network, this assumption does not hold or it is hard to implement. The network structure changes continuously due to varying channel conditions, nodes moving, or nodes dying. Each network change results in a redesign of the coding operations to be carried out at network nodes. The critical problem is that distributing information regarding the overall network structure and coding operations is costly. Consequently network coding cannot be exploited unless it can be implemented in a decentralized manner. This is by large an open problem that requires focused attention.

Security

Network coding is also vulnerable to security attacks since it requires complex operations to be conducted in the network. This has the potential to increase network exposure to eavesdropping, breach of privacy and other security attacks. Designing Network Coding techniques that include robust countermeasures to prevent eavesdropping and other attacks is critically needed. Sometimes countermeasures against security attacks may make the aforementioned three key research issues to be difficult. On the other hand, deep packet inspection can be used to enhance network security. Thus, there are new open problems considering both the security issue with other issues.

4.3 Content & context fusion for improved multimedia access

With the evolution of multimedia and audio-visual services, the “Context Awareness” principle becomes increasingly promising in improving the dynamic interactions between the connectivity and application layers. Context-aware systems gather and manage different types of context information related to users, including his/her access networks, his/her terminals and his/her services; and therefore enable: i) Content adaptation: allowing dynamic adaptation of content according to different users, their QoE and their contexts as well as the network, terminals and services contexts, and ii) Extending the QoS/QoE scope to tackle the user satisfaction, the resource level (access network capabilities, terminal(s) capabilities), and the semantic level (content and service metadata, user profile and preferences). Consequently, content distribution would be achieved in a network and service aware manner.

The user’s content consumption style and context, including his/her profile, are core components of personalisation and recommendation engines. Consequently, the fusion of both content and context is necessary in future multimedia access systems, where the multimedia content is analyzed and combined with contextual information, such as location, tags added to the multimedia, recommendations (dependent or not on the location) and status of the user when creating/consuming the multimedia.

Most of the work in academia and industry in context awareness has traditionally focused on location-based services. However, a rapidly emerging field for innovative services is linked to personal sensors, with the advent of embedded physical sensors such as mobile phone accelerometers. This allows for adapting the media to human behaviors and needs, and consequently providing user-centric personalized media with content intelligent adaptation (taking the form of recommendation for individuals’ and groups, selection and customization) according to the user's and his environment contexts. Another dimension on context awareness is the social domain, where empowering users with context information about members of their social networks appears as a promising field, receiving attention both from academia and industry. Finally, it is important to mention the contextual information coming from the personal experience of the user, which relates to different dimensions such as age, language, cultural area, education, etc. This information has a potentially strong impact on the way of perceiving, describing and approaching contents and media.

The challenge of combining content and context entails many open research problems, including:

- Content's context, including, but not limited to, content localization with respect to the user through proximity awareness and management and the efficient ontology through studying the relationship between different content information;
- User experience, both implicit and explicit, considering the terminals and network contexts, seamless content access across multiple screens, new models of users' interaction, users' satisfaction and acceptability;
- Effective ways/interfaces for the user to add context to the content (through crowd sourcing tools, social tagging, etc.).
- Research on automatic creation of meaningful contextual description of the content (tag expansion, automatic tagging, noise detection and reduction, etc.).
- Research on recommendation engines that take into account the other contextual and content information to generate localized recommendations to the user.
- Research on fusion techniques that can group together all content and context sources to improve the user's search/retrieval of the exact requested material and improve their consumption (QoS).
- Users' profile sharing among different applications and different actors, mainly through a convergent standard architecture across TV (open/managed), mobile and Internet with metadata existence and among different actors in the media domain (operators, service providers, content providers, etc).
- Adapted business models, with multi-party agreements among the different actors (mainly network providers and service providers).
- Automatic adaptation of the annotation jointly taking into account different dimension of the context information (time, space, culture, language, personal experience, education, etc), as well as their temporal evolution.

4.4 Multimedia, multimodal and deformable objects search

Defining the next generation of media search technologies requires major research efforts in multimedia, multimodal and deformable objects search. Indeed, the continuous and rapid growth of multimedia content available on the Internet has not been accompanied by a similar development of efficient, multimodal and intuitive cross-media search capabilities.

Despite the significant achievements, the current technology suffers from a number of strong limitations which prevent the user to efficiently access the desired (and theoretically available) information. This leads to several research challenges which are grouped below in three interlaced categories:

Towards truly multimodal search

A first Research Challenge in this domain consists of the expansion of current schemes to truly *multimodal capabilities* (e.g. exploiting all information of all available modalities when searching for media content). When current search engines are mostly limited to text queries, future search engines will need to use richer and more diverse sources of information including combined data from speech, audio, video, images, 3D, social tags (either automatically or user generated), physiological signals providing information of the emotional state or activity of the user (heart beat, brain waves, etc) or geolocation information. An important aspect of this research challenge is the possibility of performing cross-modality search such as for example audio-based video retrieval (a typical example of the latter would be to retrieve a video –e.g. sequence of images - that semantically matches a given audio content such as music). Other key aspects in multimodal search include the management of diversity and uncertainty in search to achieve richer and more personalized results, and the use of contextual information as a way to create a common ground to model the relationships among different content sources.

This challenge calls for new paradigms for content-based high-level (or semantic) signal representations that would permit cross modal navigation. It also appears obvious that major challenges remain in the field of machine learning and in particular with respect to multiple heterogeneous signals fusion, content and context adaptation using limited training data, and audio source separation. Finally, it is necessary to increase the quality of search by eliminating or grouping all multimedia content that does not add extra information to the user (because it is duplicated or of less quality than existing content) and implementing personalized search systems.

Towards search of multimodal and deformable content

Another important research challenge will target the access of complex, possibly multimodal and *deformable information* from a rudimentary query. Indeed, current search experiences may be very tedious and especially in the case where it is not straightforward to describe the searched information with simple text queries. It is, here, highly desirable to be able to define a query by associating rudimentary and imperfect – possibly cross modal- descriptions of the searched information (e.g. search from hand-drawn sketches, query by humming or low quality audio recordings, search for multiple videos of the same scene - but taken from different angles - search of 3D objects from 2D views or simple drawings or search of 2D projections of an object from the relevant 3D description, possibly considering 3D deformation models). This represents major theoretic research challenges in multi-level hierarchical content representation, in complex multimodal scene analysis and in decomposition models on known or unknown dictionaries or object bases.

Towards efficient user interaction

The third research challenge directly targets the development of new paradigms for *user interaction* (and satisfaction) in a multimodal context. It is, indeed, essential that future generation search tools can propose intuitive and rich query interfaces (multimodal, multi-level – that range from a full picture or only sketch – and intuitive – e.g. by means of, for example, natural language), efficient presentation of the retrieved results and finally efficient means for user interaction with the retrieved information for successive searches. This includes the possibility of multimodal navigation across different media, for instance, jumping from a piece of soundtrack to the relevant movie, and then to other movies involving the same actor or to the e-book version of the book from which the movie originated.

It is critical to develop methods and tools for assessment of the quality of experience of the user. These methods were developed and standardized for voice and video services – now need to be moved forward to 3D content and multimedia services. This will put the user in the center of attention and will forge a true user-centric media environment.

This represents major challenges for the characterization of mono-modal and cross-modal media similarity and its use in the selection of the retrieved information. Other research challenges of the same kind include the presentation of the retrieved information on heterogeneous devices and better user-feedback exploitation for multi-query search in a fully multimodal context.

4.5 3D content generation leveraging emerging acquisition channels

Communicating information using images plays a major role in today's society. There are a significant number of applications where the ability to display and visualize a 3D image comfortably confers real benefits. Examples from the professional domain include medical imaging, scientific visualization, security/defense, education, computer-aided design, and remote inspection. While in consumer markets 3D video games and 3D multimedia offer a rich experience to the consumer.

In very recent years 3D technology has become an extremely hot topic of research and there is a real feeling of excitement surrounding the technology. With the success of 3D movies such as Avatar burned in their memories, content creators, distributors, consumer electronics manufacturers and Hollywood studios have all expressed serious interest in wowing their own audiences. As a result, research in the 3D technology has intensified to progress its introduction to the home consumer be it 3D TV and/or 3D internet. However, for the 3D technology to be fully adopted by the home consumer, solution to several research challenges need to be investigated and solutions are needed to simplify the generation of 3D content and provide the users/producers similar hardware and software facilities as those enjoyed today by 2D video makers and users. Among other research challenges, visualization and representation of information, and transportation of information, remain key despite the numerous advances made in the field of stereo vision.

3D content generation

Today, the capture of 3D images and video relies on stereo vision where a number of cameras located in different positions are used. However, this kind of setup for 3D content production is cumbersome and requires correct calibration of all cameras. To that effect an important research challenge is to develop novel technologies which allow the 3D content to be captured using a single 3D camera and hence simplifies 3D video production and allows adoption of the well known techniques used today in 2D video production.

Due the increasing in computing power, computer generated graphics are becoming more and more an essential part of today visual content. Hence another research challenge is provide repositories of reusable and adaptable 3D assets (animated characters, background/environments, props, etc) that can communicate/interact with each other and systems that share/distribute rendering and processing requirements over intelligent networks. Computer graphics plug-ins have been developed for the generation of stereo images. However, if a new 3D content genre is defined to avoid the shortcomings of stereo imaging, then similarly true 3D graphic tools need to be made available.

3D Content Editing, Authoring and Sharing:

The dramatic explosion of the user-generated content culture on the web has illustrated that any new form of 3D content should be able to take off on non-professionals online resources such as YouTube. This would allow both professional and non-professional authors to jointly develop 3D content with a more realistic sensation. Tools are required for users, whether professionals or not, to be able, via open source authoring environments, to create intelligent content but also to share it across open networks and remix it by using various elements of different distributed content items to produce in turn new content.

There exist a wealth of software tools available today to home consumer for editing and authoring of 2D audio visual content. Hence, it is a requirement that similar tools are made available to the home consumer as well as professionals to allow editing and authoring for seamless compositing of any new form of 3D content. The interconnecting backbone of today's world wide web is still the hypertext. Although a plethora of multimedia content is published on the internet, non-hypertext interactivity is sandboxed on the site. In order to create a real 3D Internet experience, hyperlinks must not be restricted to the hypertext, but extended to any kind of media available on the internet (hypermedia). In 3D video, extraction of objects is potentially easier, as parallax effects clearly separate foreground objects from the background. These objects can be used to link to further 3D video clips, allowing for nonlinear video experiences.

3D Visualisation

Binocular vision feed the human brain at all time with slightly offset views of a scene: the greater the disparity, the closer the object. At the same time, converging nearby or far away gives our brain hints of distances. This principle has been exploited by stereo vision technique where two discrete views are presented to the left and right eye of the viewer via colour, polarisation or time separation techniques requiring special glasses. Over the last few years a number of autostereoscopic multiview 3D displays have been demonstrated where lenticular optical elements or parallax barriers technologies are used to separated the left and right view. A particularly contentious aspect of stereoscopic displays for entertainment applications is the human factors issue. Furthermore, due to the lack of perspective continuity in 2D view systems, objects in the scene often lack solidity (cardboarding) and give rise to an 'unreal' experience. Hence for 3D TV and 3D internet to become a reality, another research challenge that needs investigating and resolving is the development of 3D content format and related technologies which offers fatigue free viewing with more than one person independently of the viewer's position.

4.6 Immersive multimedia experiences

We live in a multimedia world. Users are increasingly looking for new multimedia experiences: new ways to capture, share and consume their multimedia content. In recent times, both research and industry has focused on designing technology that could enable immersive experiences with multimedia content, especially suited for home environments. These environments might enable the users to interact with multimedia content even when they are not co-located.

Despite many years of research on Media Spaces, we are still far from developing technologies that would allow people to virtually interact at distance with the same efficiency and ease than when face-to-face. Ethnographical observations from real work settings show that many solutions developed to support collaborative interactions at a distance are flawed as they "fracture the relation between action and the relevant environment". For example, using many video cameras to capture and share different points of view between two remote locations might seem to be an improvement over the use of a single camera. However, users might feel lost in the attempt to understand which view is the partner currently looking at or how to adapt common communication strategies to this multitude of perspectives. We need to find more subtle technological solutions to translate communication mechanisms which are effective in presence but not available when conversational partners are not co-located. These solutions should allow recreating the same functions using different but equivalent strategies. We highlight a number of these mechanisms that might potentially enable unexplored interaction capabilities in media spaces.

Gesture recognition

It is well known that gestures complement verbal interactions and help humans to disambiguate references used during the interaction (e.g., discussing blood test reports from different patients) or to better support comparisons between various information media (e.g., combining a broken leg x-ray result with a plastic leg miniature so that the physician can point specific articulations in the former and manipulate the latter while explaining the injury cause). Supporting and developing new approaches to gesture recognition is of primary importance for immersive multimedia experiences. Advanced Gesture recognition also helps in enhancing the user QoE.

Modeling the focus of attention

One of the challenges that designers of interactive immersive environments are constantly facing is the detection of the users' focus of attention. This roughly corresponds to the point in the shared workspace that is currently looked at by the user and/or the objects s/he is interacting with. Detecting this element is extremely complicated. However, this information is extremely valuable as it can help designers design better support systems for the users' interactions in the system. Future immersive multimedia spaces will benefit from sensors and models able to recognize the users' focus of attention. In this context, the eye contact could be advantageous, where most of the existing means for remote interaction (e.g., video conferencing, web cams, etc.) do not allow the eye contact, where each user should focus on the camera, which does resemble the nature communication.

Context detection and modeling

Future Internet Media devices should react in a dynamic fashion to changes in the user's situation, for instance switch automatically from visually displaying text to reading the text out loud when the user starts another activity that keeps him/her from looking, yet not from listening. Time-saving applications would be very helpful as well. Any modality of content (including 3D, haptic content, games, etc.) and its delivery should be intelligently substitutable at any time, for reasons of convenience, time-saving, filtering for relevance, or improved understanding - not to mention bandwidth bottlenecks or the benefits of people with disabilities.

3D body reconstruction

Advancements in the field of video analysis might allow in the near future to design immersive environments where the 3D model of the body of the user is fully reconstructed from the video feed of several cameras available in the environment. From this information more elaborated task models can be derived and used to allow forms of interactions with computational devices and multimedia content. With the development of 3D human body scanners, it becomes possible to generate hundreds of accurate body measurements as well as an accurate 3D shape (without skeletal knowledge) in the form of point clouds from a specific subject. From this information a complete pipeline will allow capturing the shape of real people with parameterization techniques for modeling to animation. Furthermore, comprehensive biomechanical models can offer accurate mechanical body deformability through an accurate, however complex representation of the biologic materials below the skin (muscles, fat tissues, bones) and how these interact between each other.

Integration of digital and physical space

In media spaces, humans tend to interact with both digital and physical artifacts. However, connections between these two realms have been limited so far. One of the challenges for the future of these environments will be to build better support and easy translation between “bits and atoms”.

Furthermore, the goal would be to not only recreate and emulate face-to-face setting but to design interaction capabilities that can be augmented by technology enabling forms of interactions that are not possible in standard co-located interactions.

Persuasive multimedia experiences

Immersive multimedia environments might not only provide new ways of creating and consuming media for entertainment or work purposes. Combined with persuasive technologies, these environments might stimulate users towards a positive change in their behavior (e.g., embracing a more active lifestyle, etc.).

Multisensory immersive experiences

Today immersive systems are mostly focused on audio-visual applications while other sensorial modalities are still largely uncovered. While for audio and video there exist extensive studies employing models of human perception (e.g. for advanced compression, interaction, etc.), multisensory data is only marginally considered.

Social Immersive Multimedia (SIM)

SIM is a form of computer generated/mediated augmented reality that focuses on the aspects of social (interpersonal) interactions and aims at overcoming the still dominant GUI (Graphic User Interface) metaphor in multimedia. SIM aims at providing immersive experiences (including for example: communication, education, entertaining, training, and socialization) to users by means of body controlled interactivity, using lightweight (even invisible) sensing infrastructures, rather than cumbersome wearable technologies. Ultimately the goal is to design user interactive behavior so to attain highly effective and engaging experiences at a social level. SIM is an emerging research area in the Human Computer Interaction (HCI) research community whose value and impact (economic and social) will be boosted by its incorporation in the framework of the Future Internet Multimedia. In particular the open, scalable, ubiquitous Future Media Internet infrastructure coupled with Social Immersive Multimedia will open the way to the formation of a new generation of Internet based services for consumers and of user-centric social networks.

4.7 Content with memory and behaviour

Future Media on the Internet must build on the new capabilities offered by new web technologies to provide an improved user experience. Content will adapt to users, context and purpose. Such content exposes ‘a behaviour’. It remembers, reacts, interacts and thereby actually becomes bi-directionally immersive, i.e. immerses the user as well as immerses itself into the user's environment.

Recent years brought us neologisms such as blobjects, blogjects, tweetjects and the Spime. All these terms denote a kind of object which is capable to converse with its environment – the real one as well as the virtual one. This is not just the Internet of things. Today, the Simple Object Access Protocol (SOAP) talks to systems, but in the future we will access objects – no matter what systems manage those objects or where they are managed.

Future Media will be composed of such autonomous content objects or content object mash-ups. The autonomous objects will *travel over the network, split and combine* to generate the new service or a ‘virtual world object’.

The above ideas include a number of research challenges that have already been described. For example media encoding, media search etc. Yet, there is a number of research areas related to the content objects themselves:

Ontologies and Semantic Description

This set of challenge includes research on the modeling and semantic description of the content objects as media containers, along with descriptions of their structure, characteristics, behavior, rules and relations.

Decomposition and Reconstruction

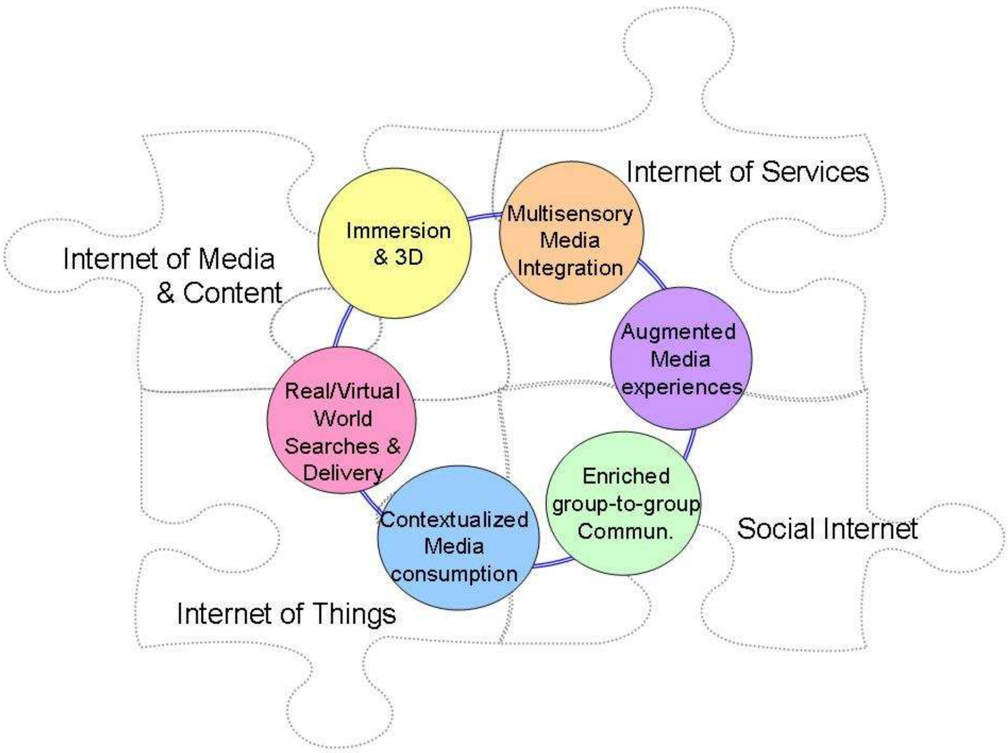
This set of challenges includes research on the decomposition of scenes and scenarios, along with real-time component-preserving capturing technologies: advanced capturing systems using stereoscopic cameras, camera arrays, time-of-flight cameras or 3D scanners preserve three-dimensional information and thereby the components and their spatio-temporal relationships. Moreover, assembly and reconstruction of complex scenes and scenarios as content objects mash-ups.

Network Support for Content Objects

This set of challenges includes research on the network architecture (including distributed network intelligence, network topology and traffic awareness, content distribution including caches), and issues like routing and streaming of the content objects.

5 New application areas of the FMI

The influence of the Internet in today's interpersonal communication and interaction with information has reached a level never imagined by past generations. The penetration of Internet services and applications has reached a stage that is starting to have a profound impact on different dimensions of people's lifestyles, including their everyday habits. The advances in Future Media Internet technologies will enable innovative applications and services, engaging new experiences, where multimedia digital content will be more immersive and closely related to the physical world. In fact, the Future Internet will allow a new generation of online ubiquitous applications sustained by new enablers of the Internet amongst which the Internet of Services, the Internet of Things, the Internet of Media & Content and the Social Internet (see figure below).



New Dimensions of Applications Enablers of Future Internet

On top of these new dimensions of the Internet, the applications that will be created will allow users to enhance human-media and human-human communications as follows:

Immersive and 3D Applications

The envisioned new applications in this field will include revolutionary ways of interacting with media through sophisticated representations of real and virtual worlds. Among the kind of applications enabled by immersive media it is expected the appearance of new physically strong experiences such as exertion games, simulation and training of real life situations in realistic way (e.g. chirurgical interventions), new entertainment experiences (e.g. tele-reality concerts), and personal enriched communications (e.g. sophisticated tele-presence).

Multisensory Media Integration

The Future Internet will enable applications dealing not only with data and information but also will enhance the perception of digital contents by exploring new ways of impacting Human perception. In fact the Future Internet will include a new set of revolutionary applications focusing on enriching Human sensations. This will be achieved by stimulating simultaneously the different senses beyond audiovisual content by including haptics or smell.

Augmented Media Experiences

A set of new applications of the Future Internet is envisioned to impact the way people interact with the physical world. In fact, there are expected applications dealing with augmented media that will enrich elements in the physical world with digital content in a way that will modify the conception of reality. In such context, a new kind of reality will be conceived as the sum of physical and augmented information that will go beyond early augmented reality prototypic developments into a well established discipline rich of applications and services.

Enriched Group to Group Communications

Traditional one-to-one and group-to-group communications will be transformed in the context of the Future Internet. The kind of applications envisioned in this concern, will include synchronous and asynchronous interactions including multi-channel sources to enhance the way people communicate. Typical applications and services in this domain will include engaging of distant family members into end-to-end games and collaborative work with remotely located teams, involving verbal and nonverbal communications and interaction with tangible devices and objects.

Contextualized Media Consumptions

Future Media Internet will enable applications where information not only will be available at any time but moreover its consumption will be adaptive in order to be selective in terms of when, where and how to present it. This will imply a level of intelligence embedded in media and in its composition in order that users will be able to access content in its best possible way and in the right moment when it is needed. This will allow for applications contextualized according to personal profiles, location, types of media, available devices, resources and QoS needs.

Real-Virtual Worlds Searches and Delivery

The Future Internet will include a rich set of applications dealing with real, virtual and mixed information and consequently its identification and retrieval. The envisioned applications in this respect will include searching facilities allowing for cross modal search in both, input and output. This means that the input mechanisms enabled by search applications will allow not only textual queries but also multimodal ones while the output will be given also in combined modalities. Moreover, the searches will not be limited to the virtual world but will retrieve information regarding physical objects, their location, their state and their relationship and enrichment with digital contents.

6 Conclusions

This white paper is the result of the first brainstorming session of a group of experts from the EU, USA and Korea on new research trends of the Future Media Internet, with emphasis on *media*, without neglecting the importance of the network. During the discuss sessions, research challenges were identified and highlighted with the goal of determining the current trend and future perspectives of the research community on the Future Media Internet.

More specifically, the Task Force identified the following major research challenges:

- Scalable multimedia compression, transmission, concealment;
- Network coding and streaming;
- Content & context fusion for improved multimedia access;
- 3D content generation leveraging emerging acquisition channels.
- Immersive multimedia experiences;
- Multimedia, multimodal & deformable objects search;
- Content with memory and behaviour.

Based on the above challenges the experts have identified potential applications and impact that these challenges should have. This white paper is expected to be the basis for a Future Media Internet Book which is the next big challenge of the Task Force.

7 Editors & Contributors

Editors

No	Name	Affiliation	e-mail
1	Isidro Laso-Ballesteros	EC	Isidro.laso@ec.europa.eu
2	Petros Daras	CERTH/ITI	daras@iti.gr
3	Theodore Zahariadis	SYNELIXIS	zahariad@synelixis.com
4	Federico Alvarez	UPM	fag@gatv.ssr.upm.es
5	Paul Moore	ATOS	paul.moore@atosresearch.eu
6	Oscar Mayora	CREATE-NET	oscar.mayora@create-net.org
7	Ebroul Izquierdo	QMUL	ebroul.izquierdo@elec.qmul.ac.uk

Contributors

No	Participant Name	Affiliation	e-mail
1	Tomas Piatrik	QMUL	tomasp@elec.qmul.ac.uk
2	Marco Pellegrini	CNR	marco.pellegrini@iit.cnr.it
3	Hassnaa Moustafa	FT/Orange	hassnaa.moustafa@orange-ftgroup.com
4	Xavier Anguera Miro	Telefonica	xanguera@tid.es
5	Amar Aggoun	Brunel University	amar.aggoun@brunel.ac.uk
6	Noel O'Connor	DCU	Noel.OConnor@dcu.ie
7	Gaël Richard	IT Paris	gael.richard@telecom-

			paristech.fr
8	David Griffin	University College London	d.griffin@ee.ucl.ac.uk
9	Ralph Traphoener	EMPOLIS	Ralph.Traphoener@empolis.com
10	Aggelos Katsaggelos	Northwestern University	aggk@ece.northwestern.edu
11	Dimitris Protopsaltou	University of Geneva	protopsaltou@miralab.unige.ch
12	Hyoung Joong Kim	Korea University	khj-@korea.ac.kr
13	Nuria Oliver Ramirez	Telefonica	nuriao@tid.es
14	Eric Karstens	European Journalism Center	karsten@ejc.net
15	Michał Grega	AGH University	grega@kt.agh.edu.pl
16	Francesco De Natale	University of Trento	denatale@ing.unitn.it
17	Mikołaj Leszczuk	AGH University	leszczuk@agh.edu.pl
18	Julien Masanes	European Archive Foundation	julien@iwaw.net