17-10-2007

Intelligent Multimedia Delivery? It's a question of semantics

Joseph Thomas-Kerr
University of Wollongong, jak09@uow.edu.au

I. Burnett
University of Wollongong, ianb@uow.edu.au

C. Ritz
University of Wollongong, critz@uow.edu.au

Intelligent Multimedia Delivery? It's a question of semantics

Abstract

Intelligent multimedia delivery uses semantic information about content to enhance the delivery process. This paper proposes a model for intelligent multimedia delivery that advances the state of the art by incorporating a concept of semantic distortion into the delivery optimization process. Furthermore, the model combines format-independence with rate-distortion optimization to provide a flexible framework for intelligent delivery of multimedia in existing and future formats. The paper provides a review of the existing work covering components of the model: scalable coding formats, rate-distortion optimization, format-independent adaptation, and semantic adaptation. It then details the model, and identifies open problems and opportunities for further research.

Disciplines

Physical Sciences and Mathematics

Publication Details

This conference paper was originally published as Thomas-Kerr, J, Burnett, I, Ritz, C, Intelligent Multimedia Delivery? It's a question of semantics, International Symposium on Communications and Information Technologies ISCIT 2007, 17-19 Oct, 473-478.

Intelligent Multimedia Delivery? It's a question of semantics

Joseph Thomas-Kerr, Ian Burnett and Christian Ritz
School of Electrical, Computer and Telecommucations Engineering
University of Wollongong,Wollongong NSW 2522 Australia
Tel: +61 2 4221 3065, Fax: +61 2 4221 3236
Email: {joetk,i.burnett,chritz}@elec.uow.edu.au

***Abstract*— Intelligent multimedia delivery uses semantic information about content to enhance the delivery process. This paper proposes a model for intelligent multimedia delivery that advances the state of the art by incorporating a concept of *semantic distortion* into the delivery optimization process. Furthermore, the model combines format-independence with rate-distortion optimization to provide a flexible framework for intelligent delivery of multimedia in existing and future formats.**

The paper provides a review of the existing work covering components of the model: scalable coding formats, rate-distortion optimization, format-independent adaptation, and semantic adaptation. It then details the model, and identifies open problems and opportunities for further research.

I. Introduction

DESPITE great strides in semantic representation of multimedia [1], computers still regard media content largely as a collection of bits. When it comes to content delivery, however, many of the pieces are in place to incorporate semantic information into the process. This would mean that elements in a multimedia delivery chain (Figure 1) make intelligent decisions about which media packets to forward (and which to drop) when inevitably faced with the situation where bandwidth demand exceeds supply, or where significant error rates are encountered (such as on a wireless connection). Instead of making such decisions purely on low-level considerations such as buffer fullness and ordering, intelligent delivery would also consider packet interdependency [2], the contribution of the packet to the *meaning* of the content [3; 4], or other semantic information.

Intelligent delivery targets numerous application scenarios, three of which are depicted in Figure 1. The first is a traditional client-server arrangement, where for example the server at (a) is to stream content to a client (b). Secondly, if the client and server are located on separate networks (such as would be the case for a mobile client, c) a gateway (d) may be interposed between the two. In this case, rate-distortion optimization may occur at both the server (a) and gateway (d). In a third scenario, peer-to-peer multimedia delivery [5] presents very different challenges, because local rate-distortion optimization at each peer participating in the delivery (e,f) may not result in the minimum distortion at the receiver (b).

This paper will provide a review of the state of the art for the technologies relevant to intelligent multimedia delivery, and then propose a model for intelligent delivery which incorporates the existing work and identifies areas for further research.

The first component that is essential for intelligent multimedia delivery is encoding formats that support efficient *scalability*. Such encoding formats are organised in a way that permits certain subsets of the complete bitstream to be decoded, albeit at a lower quality, or spatial or temporal resolution. This increases the flexibility a delivery framework has to vary the transmitted bit-rate when required without interrupting playback at the receiver. Section II considers scalable coding in further detail.

Secondly, an intelligent delivery framework requires algorithms that are able to identify the optimum choice of packets given a (dynamic) bandwidth constraint, so as to minimize the distortion of the content at the receiver. Such rate-distortion optimization algorithms are discussed in section III.

Section IV presents several frameworks for multimedia *adaptation* based on the MPEG-21 Multimedia Framework [8]. Adaptation is a more general problem than delivery, encompassing arbitrary transformations of multimedia content, which may or may not possess timing constraints. For example, adaptation may involve transcoding or even transmoding the content to provide so-called Universal Multimedia Access [8].

Finally, various techniques for adaptive delivery based on semantic data are discussed in section V. Such techniques are particular examples of the myriad types of metadata that carry semantic information which is significant to the delivery process. In order to support the given techniques, as well as

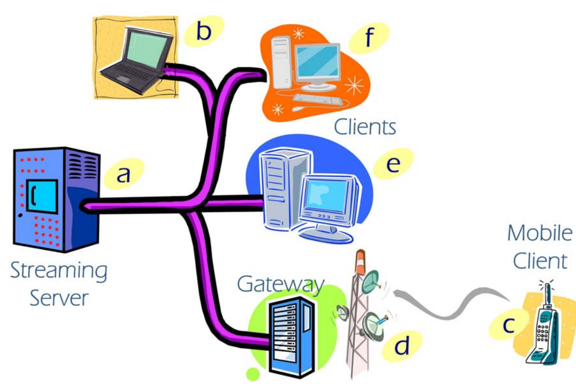

Fig. 1. Components of a multimedia delivery network (adapted from [6])

1-4244-0977-2/07/$25.00 © 2007 IEEE

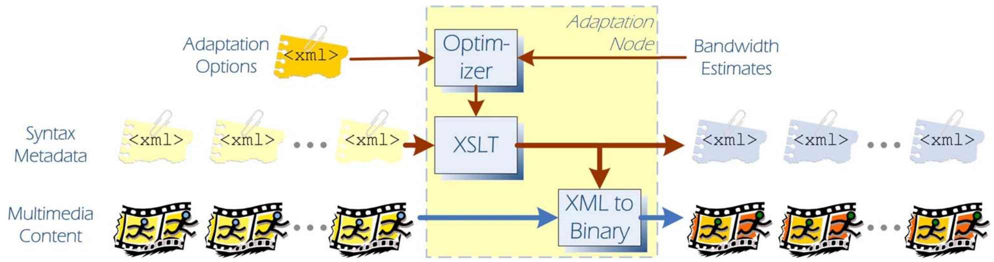

Fig. 2. Dynamic Adaptation using MPEG-21 (adapted from [7])

others as they are identified, a general framework is necessary for intelligent multimedia delivery. This framework is proposed in section VI.

II. Scalable Content Formats

Scalability refers to the ability to encode multimedia content in such a way that the bit-rate used to transmit it may be scaled downward by one or more steps, without re-encoding. This allows a trade-off to be made between bandwidth and quality[1] (or one of several other dimensions, including spatial, temporal, or chromatic resolution). This ability greatly enhances the flexibility of multimedia delivery systems, by (among other things):

- allowing a server to store a single version of a piece of content, yet deliver it to multiple heterogeneous clients. Each client may receive the content at the best resolution and frame-rate for their particular device [9];
- making it possible to transfer a streaming session that is in progress from one terminal to another without interruption [10]. For example, a user may be watching content on a mobile terminal and wish to transfer the content to their TV set when they arrive home; and
- improving the ability of a delivery chain to adapt to dynamically changing conditions by choosing the best available trade-off between bit-rate and quality at each point in time [11].

It is the latter application that is the primary focus of this paper, and as such it is left to [9; 10] to explore the other applications further. There are numerous approaches to scalability presented in the literature, of which the following are some of the more important.

A. Scalable Video

Early work in video scaling focused on providing feedback of network conditions to the encoder, so that it may adjust various parameters to alter the output bit-rate [12]. This approach targets real-time applications such as video-conferencing where encoding and delivery operate concurrently, but is unsuitable for delivery of pre-encoded content. Other work of a similar vintage proposes so-called “dynamic rate shaping” by selective truncation of DCT coefficients from encoded pictures[13]. While this allows adaptive delivery of stored media, it was found to yield poor results [12] and lead to significant coding artifacts under certain conditions.

[1] typically referred to as signal-to-noise ratio (SNR) scalability

More recently, MPEG published extensions to the MPEG-4 video format, known as Fine Granular Scalability (FGS) [14]. FGS aims to provide SNR and temporal scalability without significantly increasing the computational complexity of either decoding or streaming. This is achieved by encoding video content into a base layer and a single *enhancement* layer. The latter performs bitplane DCT coding on the base layer residual to achieve fine-grained scalability.

The H.264 video codec [15] introduced a new structure for motion compensation reference picture organisation, providing significantly more control over frame interdependency and thus temporal scalability. More recently, amendments have been introduced to this standard that provide fine-grained SNR as well as spatial scalability [16] (often referred to as SVC - Scalable Video Coding). In addition to providing fine-grained SNR scalability via a similar means to FGS, SVC performs coarse-grained SNR and spatial scaling using separate coding layers and inter-layer prediction techniques. Also, to overcome the poor coding efficiency typically observed in FGS, motion compensation prediction for the enhancement layer is performed using complete reference pictures, rather than only their base layer.

B. Scalable Speech and Audio

Audio and speech generally have much lower bit rates than video. However, for low bit rate mobile terminals, the audio or speech signal becomes a significant component in the perception of audio-visual content [17]. Earlier approaches to scalable audio coding focused on multi-rate audio coders, such as MPEG-4 AAC [18], while more recent scalable coders aim to provide for fine-grained scalability, such as the recently standardised MPEG scalable to lossless coder [19].

Approaches to scalable speech coding include the multirate coders, such as the recent adaptive and variable multirate wideband speech coders of [20; 21], with recent research focusing on fine granular scalable coders, such as described in [22]. Many of these scalable coders are designed to maximise the perceptual quality at each bit rate. More recent research has focused on scalable coders that are resilient to packet losses in networks, such as the multiple description based coders described in [23; 24].

III. Rate-Distortion Optimized Packet Scheduling

The scalable coding formats described in the preceding section provide components of the delivery process with substantial flexibility to vary the rate of transmission according to the capability of the channel. However, making use of

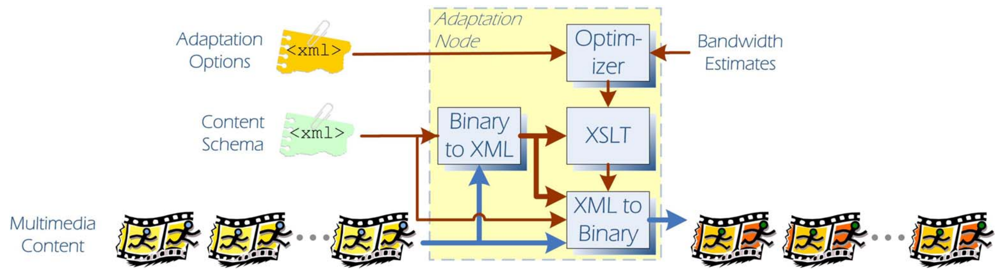

Fig. 3. Bandwidth-efficient Dynamic Adaptation using MPEG-21 (adapted from [6])

this flexibility is not a trivial excercise. In [25], Chou et al. observe that the process of deciding if, and when, to send each atomic unit of scalably coded media is a problem in rate-distortion optimization. Put differently, given a (dynamically changing) channel bandwidth, which parts of the media should be delivered to minimize the total distortion of the content at the receiver? It turns out that optimal distortion-rate performance $D(R)$ may be achieved by minimizing the Lagrangian $D+\lambda R$ for some λ [26]. This is in turn based on error probability-cost functions for each unit of data, where errors are characterised in such a way as to encompass bit error rate, packet loss, and delay (such that the packet is too late to be useful).

Furthermore, the formulation of distortion must consider the interdependencies between data units, since descendent packets (e.g. enhancement layers in FGS or SVC, or any motion-compensated frame) generally cannot be decoded if their ancestors are not received. This dependency is modelled as a directed acyclic graph [26] and is used to compute the distortion D of a set of packets π with error probabilities ϵ [2]:

$$D(\pi) = D_0 - \sum_{l} \Delta D_l \prod_{l' \prec l} (1 - \epsilon(\pi_{l'})) \qquad (1)$$

Using this model, the optimal set of transmission decisions is completely characterised by

- ➢ the dependence graph;
- ➢ the set of distortion increments ΔD_l;
- ➢ the set of packet sizes; and
- ➢ the error-cost functions.

All but the final parameter are features of the media, and error-cost is a characteristic of the channel. The dependence relations l' complicate the subsequent optimization process, but it may be solved using an iterative descent algorithm [26].

IV. Adaptation

Hutter et al. [7] propose a framework for multimedia adaptation based largely on MPEG-21 [27]. This adaptation framework is format-independent; that is, it allows adaptation of *any* content, regardless of its encoding. This feature is crucial, since the number of possible formats for multimedia content is large and growing. To be widely useful, an adaptation framework must be able to cope with content irrespective of the format in which it is encoded.

Hutter's framework also supports both distributed and dynamic adaptation. The former denotes that the adaptation operation can occur at any (and possibly multiple) points in the delivery network. Dynamic adaptation, on the other hand, is where the adaptation process can react to dynamically changing parameters in real time. In [7], the scope of such parameters is limited to network bandwidth estimates.

Figure 2 depicts the architecture of this framework. Fragments of metadata are delivered alongside content fragments, to describe the high-level structure of the multimedia. This allows an XSLT transformation to selectively prune parts of the structure to effect the adaptation, as well as adjust any header fields affected by the pruning. Feasible adaptation operations are, for example, downsampling by removing non-reference frames, or removal of scalability layers (see section II).

The XSLT transformation can be parameterized to provide multiple adaptation operations, in which case additional metadata is delivered with the content to detail the available options, their effects, and constraints. Such options may be static, or may be updated over the course of the delivery. An optimization engine is used to compute the optimum adaptation(s) to perform upon the bitstream to meet dynamic bandwidth constraints, as well as static constraints on the network, terminal device and/or user. Static constraints are provided to the adaptation node via other metadata.

The transformed metadata is subsequently fed to an *XML to Binary* module, which enacts the transformation on the content itself. Performing the transformation in the XML domain in this way allows all of the modules within the adaptation node may be generic (format-independant).

There are a number of issues with the architecture as it stands. Firstly, the syntax metadata is reported at between 15 and 200 percent overhead[3] [28], although De Schrijver et al suggest that compression may be used to significantly reduce this amount. Compressing the metadata in this way effectively trades space for time, adding complexity to the adaptation process. An alternative approach [6] recognises that the high-level structure of content fragments is in fact very regular, and conforms to a schema that is part of the normative description of the content format. Rather than transmitting the structural metadata and incurring the large overhead, the schema for the format as a whole may be delivered, and the structural metadata generated on-the fly in the adaptation node (Figure 3). Like compression, this approach trades time for space, but reduces the space requirement to a negligible amount.

Secondly, the architecture uses a wide range of metadata about the usage and delivery environments, but its examination of the content itself is limited to syntactic characteristics:

[2] where D_0 is the distortion when zero packets are received. ($l' \prec l$) indicates that l' ranges over the set of ancestors of l

[3] as a proportion of the size of the content

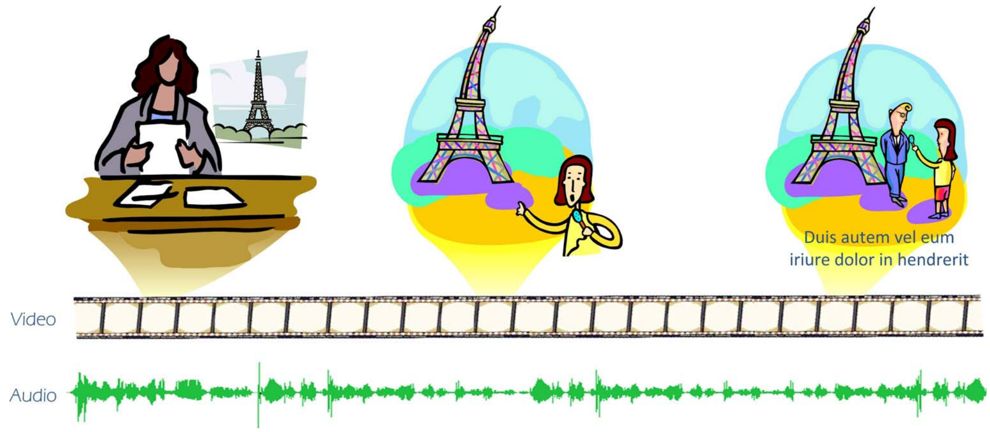

Fig. 4. The structure of a news story contains semantics that can improve the adaptation process

scalability layers and bit-rate [7]. Clearly, these parameters are fundamental to the adaptation process. However, as we shall see in the subsequent section, semantic knowledge of the content can lead to greatly improved adaptation performance.

Leopold et al. [29] present an alternative view of the MPEG-21 adaptation framework. Instead of utilizing format-independent processors, they envision a web services approach to adaptation, where nodes communicate the adaptation operations that they are able to perform (including all necessary details about supported formats and constraints). For example, one node could support an operation that reduces the resolution of JPEG, GIF and TIF images, while another node can convert from color to greyscale. A path-finding algorithm is then used across the state-space of all of the atomic operations, to find the sequence of steps necessary to complete the required adaptation. In [30] this concept is extended to use Semantic Web Services, using semantic markup languages (such as RDFS and OWL) to describe the supported operations. This work makes use of semantic tools to describe adaptation operations, but not to describe the content itself.

V. Adaptive Delivery using Semantics

Several approaches have been proposed to adapt or deliver content according to certain semantics. Bertini et al. [3], Xu et al. [4], and Baba et al. [31] all argue that applying the same adaptation operation to different parts of a multimedia presentation will have differing effects in the perceptual quality of the presentation as a whole. More specifically, Bertini and Xu both propose adaptation on the basis of semantic classification of sporting events into categories such as Shot on Goal, Corner (for soccer) [3], or Shot, Foul, Penalty (for Basketball) [4], among others. User preferences are used to prioritize the categories, and this priority information is then used to guide the adaptation. In other words, given a bandwidth constraint such that the full content can not be delivered, the adaptation engine will reduce the bit rate of lower priority sections before those with higher priority.

This semantic metadata can be considered as very high-level, and coarse-grained. In other words, it identifies relatively large segments of content, using concepts with a high level of abstraction from the digital representation of the content. In the first case, heuristic methods are proposed to automatically classify content segments, with a precision of 83% to 96% [3].

Baba et al. [31] propose adaptation of speech signals on the basis of a much lower-level semantic concept: sound level. They argue that regions of (relative) silence within a speech signal carry no *semantic information*, and as such may be truncated during playback. In fact, this feature of speech[4] may be used to guide adaptation, allowing regions of silence to be constrained to a zero bit-rate (or as close as the scalable codec will allow) with no perceptible loss of fidelity.

These approaches to delivery clearly demonstrate the potential for semantics to improve the process, and indeed merely scratch the surface in terms of the semantic concepts that may be useful in guiding delivery. As a further example, consider a news report (Figure 4). In general, news reports have a relatively consistent structure, beginning with a studio introduction, then footage of the event (often with commentary overlaid on audio from the event), using subtitles where subjects speak in a foreign language, and concluding with further studio footage. As a report proceeds through these various stages, the relative importance of the audio and video varies. For example, in the studio introduction, virtually all of the semantic content of the presentation is carried in the audio. On a low-bandwidth (e.g. mobile) channel, reduction of the frame-rate to zero in this region would have little impact on perceptual quality. When the report cuts to on-site footage, a much greater proportion of the semantic content is carried by the visuals, though the amount would vary from one report to another. If subtitles are used, then virtually all of the semantic content is conveyed by the video, and bits spent transmitting audio in this section will contribute much less to perceptual quality.

Furthermore, as well as comparing relative semantic importance along the modal dimension, similar comparisons could be made on the temporal axis. Here, temporal segments of the news report would be annotated with an indication of the relative importance of the segment to the story as a whole. This would allow users to receive a short "digest", the full story, or something in between. A similar approach could work for coverage of sporting events.

[4]or audio, although silence is less prevalent there

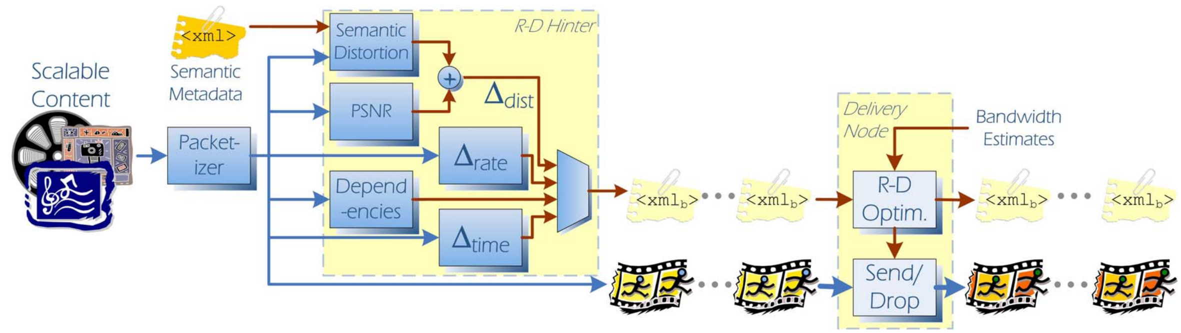

Fig. 5. A model for intelligent multimedia delivery

VI. A Model for Intelligent Multimedia Delivery

To recap: there are many elements of an adaptive, intelligent multimedia delivery system. First, scalable content formats are needed to make it easy for delivery nodes to adjust to dynamically changing conditions. Secondly, algorithms are required for (near-)optimal rate-distortion optimization to perform this adjustment. Ideally, this process should operate in a format-independent manner, so as to be able to cope with the large and growing range of content types.

We propose that the concept of distortion [26] be broadened to encompass *semantic distortion*, in addition to the current objective measures (typically peak-SNR). This so-called semantic distortion would be defined as a measure of the "SNR" between the intended semantic (meaning) of the content prior to delivery, as compared to the semantics conveyed by the content that arrives at the client. Such a measure is clearly highly subjective, as is much of the semantics associated with any given piece of content. However, even approximations of semantic distortion as perceived by parties on the server-side of the process possess substantial value for delivery, as discussed in section V.

It is likely that semantic distortion is non-associative: for example, values of the semantic distortion associated with the segments of a news report do not easily compare (or combine) with values that represent sound level semantics. Nevertheless, it may be desirable to use both at once to improve the R-D optimization. These and other issues associated with quantifying semantic distortion are currently open questions.

Figure 5 depicts a model for intelligent multimedia delivery that encompasses the components described in the preceding sections. It allows delivery nodes (streaming servers, gateways, or peers—see Figure 1) to optimize distortion according to dynamic rate (bandwidth) constraints, in a format-independent manner. However, the model does not target the full generality of the adaptation framework presented in section IV. Because *delivery* is a much more specific problem, the overhead of XSLT (or STX—see [32]) along with Binary-to-XML is not necessary to achieve format-independence.

Instead, a delivery node simply decides whether to forward or drop each packet[5]. That decision is made on the basis of the rate-distortion optimization algorithm, which along with channel feedback requires metadata about each packet. There are several options for the serialization of this metadata. On the one hand, a binary syntax could be specified, in order to maximize space efficiency over-the-wire. However, most recent metadata prefers XML over binary syntax, because of the ease with which it is processed and parsed. As it turns out, it is possible to achieve most of the benefits of both, using the so-called Binary format for Metadata (BiM) [33]. BiM uses XML Schema to provide efficient binary encodings of XML data. This means that the R-D metadata can be created and processed as XML, but if it must be transmitted, it may be compressed using BiM to close to the size of a dedicated binary syntax. Furthermore, at the downstream node, the binary representation may be parsed directly, without decompression, avoiding any additional time complexity.

The R-D metadata is generated offline by a hinter and stored with the content, ready for delivery. The hinter uses the syntax metalanguage described in [6] to identify content packets and their associated rate and timing information, and dependencies are described using Eichhorn's model [2]. Calculation of semantic distortion is clearly a large question that warrants further study. However, the model aims to support both semantic metadata developed specifically to calculate distortion, and semantic metadata developed for other purposes that is nonetheless useful for this purpose. To achieve this, a description language is required that can efficiently specify the transformation from the existing semantic metadata, into quantitative semantic distortion values. Whether existing transformation languages (XSLT, STX, XQuery, SWQL [34], or others) are suitable for this task, or what unique requirements the task may possess, are open problem.

[5]This is a slight simplification—with FGS and SVC, packets may also be truncated

VII. Conclusion

This paper proposes a model for intelligent multimedia delivery that advances the state of the art in rate-distortion optimization by incorporating *semantic distortion*. In effect, this alters the conceptual end-points of the multimedia delivery chain. Instead of server-client, using semantics extends the process to (human) creator-consumer, by minimizing distortion of the intended meaning of the content (see for example the news report in Figure 4). While a precise quantitative representation of semantic distortion remains an open problem, the model presented here describes how it may be integrated into a format-independent delivery framework (one that can operate on multimedia content regardless of its encoding). This property of the framework allows semantic R-D optimization for both current and future content formats.

References

[1] G. Stamou and S. Kollias, Eds., *Multimedia content and the semantic Web : methods, standards, and tools*. Chichester :: Wiley, 2005.

[2] A. Eichhorn, "Modelling dependency in multimedia streams," in *Multimedia, 14th ACM intl. conf. on*. ACM Press New York, NY, USA, 2006, pp. 941–950.

[3] M. Bertini *et al.*, "Semantic adaptation of sport videos with user-centred performance analysis," *Multimedia, IEEE trans. on*, vol. 8, no. 3, pp. 433–443, 2006.

[4] M. Xu *et al.*, "Event on demand with MPEG-21 video adaptation system," in *Multimedia, 14th ACM intl. conf. on*, 2006, pp. 921–930.

[5] L. Rong and I. Burnett, "Dynamic resource adaptation in a heterogeneous peer-to-peer environment," in *Consumer Communications and Networking, IEEE Conf. on*, 2005, pp. 416–420.

[6] J. Thomas-Kerr *et al.*, "Is that a fish in your ear? a universal metalanguage for multimedia," *IEEE Multimedia*, vol. 14, no. 2, pp. 72–77, 2007.

[7] A. Hutter *et al.*, "Automatic adaptation of streaming multimedia content in a dynamic and distributed environment," in *Image Processing, IEEE Intl. Conf. on*, vol. 3, 2005, pp. 716–719.

[8] C. Timmerer *et al.*, "Digital item adaptation - coding format independence," in *MPEG-21*, I. Burnett *et al.*, Eds. Chichester, UK.: Wiley, 2006.

[9] A. Vetro and C. Timmerer, "Digital item adaptation: overview of standardization and research activities," *Multimedia, IEEE trans. on*, vol. 7, no. 3, pp. 418–426, 2005.

[10] L. Rong and I. Burnett, "Application level session hand-off management in ubiquitous multimedia environment with MPEG-21," in *E-Business and Telecommunication Networks*. Springer Netherlands, 2006, pp. 298–304.

[11] B. Girod *et al.*, "Advances in channel-adaptive video streaming," in *Image Processing, Intl. Conf. on*, vol. 1, 2002, pp. 9–12.

[12] T. Lakshman *et al.*, "VBR video: tradeoffs and potentials," *proc. of the IEEE*, vol. 86, no. 5, pp. 952–973, 1998.

[13] A. Eleftheriadis and D. Anastassiou, "Constrained and general dynamic rate shaping of compressed digital video," in *Image Processing, 2nd IEEE Intl. Conf. on*, 1995, pp. 396–399.

[14] H. Radha *et al.*, "The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP," *Multimedia, IEEE trans. on*, vol. 3, no. 1, pp. 53–68, 2001.

[15] ITU-T, "Recommendation H.264: Advanced video coding for generic audiovisual services," 2005.

[16] H. Schwarz *et al.*, "Overview of the scalable H.264/MPEG4-AVC extension," in *Image Processing, IEEE Intl. Conf. on*, Oct 2006, pp. 161–164.

[17] S. Winkler and C. Faller, "Perceived audiovisual quality of low-bitrate multimedia content," *Multimedia, IEEE trans. on*, vol. 8, no. 5, pp. 973–980, 2006.

[18] M. Bosi and R. Goldberg, *Introduction to Digital Audio Coding and Standards*. Springer, 2003.

[19] R. Yu *et al.*, "A fine granular scalable to lossless audio coder," *Audio, Speech and Language Processing, IEEE Trans. on*, vol. 14, no. 4, pp. 1352–1363, 2006.

[20] S. Ahmadi and M. Jelinek, "On the architecture, operation, and applications of VMR-WB: the new CDMA2000 wideband speech coding standard," *Communications Magazine, IEEE*, vol. 44, no. 5, pp. 74–81, 2006.

[21] B. Bessette *et al.*, "The adaptive multirate wideband speech codec (AMR-WB)," *Speech and Audio Processing, IEEE Trans. on*, vol. 10, no. 8, pp. 620–636, 2002.

[22] H. Dong and J. Gibson, "Structures for SNR Scalable Speech Coding," *Audio, Speech and Language Processing, IEEE Trans. on*, vol. 14, no. 2, pp. 545–557, 2006.

[23] G. Schuller *et al.*, "Robust low-delay audio coding using multiple descriptions," *Speech and Audio Processing, IEEE Trans. on*, vol. 13, no. 5, pp. 1063–6676, 2005.

[24] B. Wah and D. Lin, "LSP-based multiple-description coding for real-time low bit-rate voice over IP," *Multimedia, IEEE Transactions on*, vol. 7, no. 1, pp. 167–178, 2005.

[25] P. Chou and Z. Miao, "Rate-distortion optimized sender-driven streaming over best-effort networks," *Multimedia Signal Processing, IEEE Wshop on*, pp. 587–592, 2001.

[26] P. Chou, "Rate-distortion optimized streaming of packetized media," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 390–404, 2006.

[27] I. Burnett *et al.*, Eds., *The MPEG-21 book*. Hoboken, NJ, USA: Wiley, 2006.

[28] D. De Schrijver *et al.*, "An optimized MPEG-21 BSDL framework for the adaptation of scalable bitstreams," *Journal of Visual Communication and Image Representation*, vol. 18, no. 3, pp. 217–239, 2007.

[29] K. Leopold *et al.*, "A knowledge and component based multimedia adaptation framework," in *Multimedia Software Engineering, IEEE Intl. Symp. on*, 2004, pp. 10–17.

[30] D. Jannach and K. Leopold, "A multimedia adaptation framework based on semantic web technology," in *Semantic Web, 2nd European Conf. on*, 2005.

[31] M. Baba *et al.*, "Adaptive multimedia playout method based on semantic structure of media stream," 2004, pp. 269–273.

[32] S. Devillers *et al.*, "Bitstream syntax description-based adaptation in streaming and constrained environments," *Multimedia, IEEE trans. on*, vol. 7, no. 3, pp. 463–470, 2005.

[33] U. Niedermeier *et al.*, "An MPEG-7 tool for compression and streaming of XML data," in *Multimedia and Expo, IEEE Intl. Conf. on*, 2002, pp. 521–524.

[34] P. Lehti and P. Fankhauser, "SWQL: A query language for data integration based on OWL," *Lecture notes in computer science*, vol. 3762, pp. 926–935, 2005.