

Challenges for Video on the Web

Erik Mannens, Peter Lambert, Davy Van Deursen, Rik Van de Walle

Ghent University – IBBT

Department of Electronics and Information Systems - Multimedia Lab

Gaston Crommenlaan 8/201, B-9050 Ledeborg-Ghent, Belgium

{erik.mannens,peter.lambert,davy.vandeursen,rik.vandewalle}@ugent.be

1. Multimedia Lab Introduction

IBBT, the Interdisciplinary institute for BroadBand Technology, is a research institute founded by the Flemish Government, focusing on information and communication technology (ICT) in general, and applications of broadband technology in particular. IBBT's primary mission as a virtual research centre (over 500 researchers) is to gather highly competent human capital, and to perform multidisciplinary research.

Multimedia Lab (MMLab, multimedialab.elis.ugent.be) is a young research group within Ghent University (Faculty of Engineering, ELIS, www.elis.ugent.be), and a leading partner within IBBT. MMLab was founded in 2001. It accounts for about 25 researchers and it has a portfolio of basic research, applied research, and contract-based research with industrial partners. MMLab is active within MPEG and W3C standardization, via the submission of technical contributions, by chairing several ad-hoc groups, and through the editorship of several specifications. The main areas of expertise of MMLab are advanced video applications, including scalable video compression techniques; adaptation of multimedia with respect to varying usage environments; development of interactive Digital Television applications; mobile applications which includes adaptive coding and decoding, rendering of multimedia data on mobile terminals with limited resources, and transparent handover of multimedia sessions between different devices (session mobility). MMLab has been a partner in more than 30 research projects (FP5, FP6, IWT, FWO, IBBT and bilateral).

2. Motivation for attendance

Within the scope of this workshop it is worth mentioning that we were one of the main contributors to the MPEG-21 specification (DID, DIA, DIP) and we still actively participate in JVT on H.264/AVC and its scalable extension (SVC). We also actively participated within W3C's XG Multimedia Semantics (MMSEM) where, among other things, we proposed a video and media production Use Case.

The focus of our current research and project work is on end-to-end quality assurance for the end-user watching video over the internet. We know there are still a lot of issues to solve on both the producer-end (scalable formats, error concealment/resilience, adaptation, ...) and consumer-end (UI, context awareness, retrieval, ...). Therefore, it is important for us to keep in touch with the ongoing development and future evolutions in this area. This workshop offers a unique opportunity to get a broader view on the topic via the visions of peer organisations and to get an insight into the future of Video on the Web. We welcome the opportunity to attend the workshop and will gladly share our experiences and views on Video on the (mobile) Web.

3. Challenges for Video on the Web

We expect that multimedia content, more specifically (mobile) video content, will emerge itself as the leading content type on the internet. Given the fact that the easiest content to produce for common people is video -all handycams and mobile phones will have native support for H.264/AVC in the near future- and that internet users produce an awful lot of (personal) data to be shared these days, the power of large numbers makes

this former statement -about video being the leading content type- a trivial one. Needless to say that the real visible challenges for the end-users will be the provision of:

- A simple, uniform UI to post to the net & view from the net
 - How codec and/or player agnostic should Browsers be?
 - How to make Devices and/or Browsers independent of Look&Feel?
 - Can we make sure that metadata always remains connected with the video resource?
- An easy way to annotate inside video fragments
 - How to cope with the extra time dimension during tagging for instance?
 - How to put extra semantics into visual data? Can detected objects be tracked throughout the footage?
 - What trade offs are there to make to (semi-) automate the annotation process taking into account context parameters?
- An easy way to search through video fragments (with this extra time dimension)
 - Can (streaming) videos also be identified via unique URI's or at least be addressed in an UMA-compliant way? In particular, how to deal with different versions of the same video?
 - Can/will the extra semantic metadata be incorporated within the same container as the footage itself?
 - How will current search engines cope with this extra time dimension, notwithstanding the fact that another (semantic) search paradigm will also have to be implemented?

For the content producers, on the other hand, it is vital/ideal that their content should be produced only once -preferably in HD quality- and that it could be distributed transparently to multiple destinations. Given the fact that TV, mobile, and internet gradually will blend in one another, this could be both linear broadcasting, mobile streaming or on-demand internet. Some challenges for these (commercial) organisations will be the provision of:

- Scalable formats
 - Is a single scalable (from HD down to QCIF) format desirable? If so, which one?
 - How do we cope with degradable (mobile) networks? Should error resilience/error concealment be taken into account from the start?
 - Will Digital Rights Management (DRM) (still) be an issue?
- Further means of adaptation taking into account some extra context parameters
 - Transcoding proxies could take both end-user and network context parameters into account to maximize end-to-end quality assurance.
 - Will bandwidth be a problem? If so, are edge-servers, proxies, and/or PeerToPeer-nodes the solution? Can semantic web technologies be of any salvation here?
- Augmented level of interactivity (cfr. MHP, MPEG LAsER and the like)
 - How can semantic web technologies help to increase this level of interactivity (object detection, object tracking, 2-way interaction, ...)?
 - Can we deal with any kind of synchronisation problem (audio, video, stills, subtitles, powerpoints, captions, ...)?
 - Are we waiting for 3D-Video or some other form of virtual reality?