

# Subjective Assessment of BitDetect—A Mechanism for Energy-Aware Multimedia Content Adaptation

Arghir-Nicolae Moldovan and Cristina Hava Muntean

**Abstract**—As mobile devices are becoming more compact and powerful and as they start to be increasingly used for accessing power-hungry multimedia streaming applications, there is an increasing need for mechanisms to efficiently manage the limited battery power resources. This is especially important as the battery capacity has not kept up with the power requirements of an increasing number of mobile device features and “always on” connected users. Adaptive multimedia-based power-saving mechanisms often decrease the clip bitrate to increase the mobile device battery life, without considering the effect of these degradations on the user-perceived quality. This paper proposes BitDetect, a mechanism that uses objective video quality assessment metrics to detect content-specific video bitrate levels that enable saving battery power while maintaining good user perceived quality. Results from a subjective study indicate that the recommended bitrate offers good user-perceived quality across different multimedia clips. Furthermore, experimental tests indicate that significant battery power can be saved by using the recommended bitrates when streaming multimedia clips to a mobile device.

**Index Terms**—Battery power saving, bitrate estimation, objective assessment metrics, subjective video quality assessment.

## I. INTRODUCTION

MOBILE devices such as smartphones and tablets are becoming increasingly popular among Internet users, with smartphone sales alone being estimated to reach 850 million worldwide by 2015 [1]. As these devices come with larger and crisper screens, high-speed connectivity, faster processors and improved graphics, they started to be used more and more for accessing multimedia rich applications such as video-on-demand (VoD), IPTV, video conferencing, mobile learning, etc. In this context, multimedia content production and consumption is increasing fast, industry forecasts estimating that by 2015 mobile video will account for 66% of the total mobile data traffic [2].

Multimedia applications usually require medium to high bitrates to achieve high quality levels, and the fast increase in video traffic puts significant pressure on the wireless networks, with congestion problems expected to continue over the coming years, despite significant advances in terms of network capacity [3]. Additionally, a number of other challenges are faced when delivering multimedia content to mobile devices over wireless networks. These are mainly caused by the high variety of mobile devices and their characteristics (e.g. screen size, processing power, battery life, etc.), as well as by the wireless networks

whose performance is highly influenced by the environment and by other factors such as, for instance the number of connections at a time and the traffic load.

In particular, the limited battery life still represents an important issue for mobile devices [4], despite the significant effort that is put into creating low-power device components, more energy efficient applications and protocols, as well as new battery technologies. Mobile devices are becoming smaller and more compact, while incorporating new features, larger screens, faster processors and multiple broadband connectivity such as WiFi, 3G and 4G. All these pay their price in terms of battery life, as even the latest, top of the line devices can barely last for a couple of days of moderate usage. For example, the iPhone 4S smartphone has a battery rated at 6 hours of internet usage over 3G and 10 hours of video playback [5]. However, the battery life will be considerably reduced below these values when streaming online video, as many users would do nowadays.

Given this wide variety of issues with multimedia content delivery over wireless networks, much research effort has concentrated on proposing solutions to provide users with a better service and help them get the most out of their devices. Adaptive multimedia strategies were shown to be viable solutions to overcome issues with the variable network resources and the mobile device characteristics. Adaptive multimedia can also be used to extend mobile devices’ battery life [6], for example by converting the multimedia content to lower quality versions. While significantly less power is required to receive, decode and display these versions, finding the right balance between quality degradation and power saving is a challenging task, especially when dealing with quality-aware users.

This paper addresses this last challenge and proposes BitDetect, a quality-aware mechanism to decrease the video bitrate of multimedia content in order to save the mobile device battery power. BitDetect uses objective video quality assessment metrics to detect a threshold up to which the video bitrate can be decreased while maintaining a good level of user-perceived quality. A subjective study was conducted to assess the performance of BitDetect across different clips with various characteristics and across different resolutions. The subjective results show that BitDetect offers good user-perceived quality level, while significantly decreasing the video bitrate, across different categories of educational multimedia clips. Results from an educational questionnaire also indicate that the recommended bitrates offer a good-enough quality to support the learning process. Moreover, experimental results show that significant mobile device battery power can be saved by using the proposed method when streaming multimedia content to a mobile device.

The paper is structured as follows. Section II presents previous research work related to this paper, on multimedia content adaptation and quality assessment. Section III details the principle of BitDetect whereas Sections IV and V present the setup and the

Manuscript received April 08, 2011 revised March 01, 2012; accepted March 06, 2012. Date of publication May 07, 2012; date of current version August 17, 2012. This work was supported by IRCSET (Irish Research Council for Science, Engineering, and Technology) Embark Postgraduate Scholarship Scheme.

The authors are with the School of Computing, National College of Ireland, IFSC, Dublin 1, Ireland (e-mail: amoldovan@student.ncirl.ie; cmuntean@ncirl.ie).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TBC.2012.2191688

results of the subjective study respectively. The final section concludes the paper and indicates several future work directions.

## II. RELATED WORK

### A. Multimedia Content Adaptation

Multimedia adaptation aims to address issues that affect end-user Quality of Experience (QoE) when delivering multimedia content over wired and wireless networks. Content adaptation may be driven by network conditions (e.g. available bandwidth) [7], device characteristics (e.g. screen size, processing capabilities, battery life [8], etc.), user expectations [9], preferences, context [10], etc.

Various adaptive solutions have been proposed to overcome the issues with wireless networks and mobile devices. Some of these may address specific problems, while others are more generic and can be used to overcome issues with both mobile devices and wireless networks.

1) *Generic Methods*: Most adaptive solutions addressing wireless network-related issues adjust the throughput of the multimedia stream to the available bandwidth, by adapting the transmission rate of the streamed packet flow, or by adapting the quality of the multimedia clip. These solutions can be roughly classified in: rate control, rate shaping and rate adaptive encoding [11].

Rate control mechanisms reduce the data loss and the buffering associated with network congestion, by adapting the packet sending rate or by reducing the number of packets to be delivered. Examples include the TCP Friendly Rate Control (TFRC) [12], and the Video Transport Protocol (VTP) [13]. In case of TFRC the server estimates the rate based on the packet size, the packet loss rate and the time until receiving the ACK packets. VTP enables a smoother rate adaptation by continuously monitoring the existing rate and by introducing an algorithm to discriminate between transmission-based loss and congestion-induced loss.

Rate shaping techniques adapt the rate by dropping part of the information such as entire frames, groups of pixels from specific frames, bit-planes in pixels, etc. [14].

Rate adaptive encoding solutions adjust the rate by modifying the values of the encoding parameters. Parameters that can be adapted include the spatial and temporal resolutions, the bitrate, the number of B-frames, etc. [11]. In general, these solutions adapt equally the entire frame area of the multimedia clip. For example QOAS [15], reduces the video bitrate of the multimedia clip to decrease the loss caused by increased traffic in the network. An example of a different solution that selectively reduces the quality of the multimedia clip is ROIAS [16]. This involves detecting the areas of maximum user interest (AMUI), and when necessary, affecting the quality of the regions users are less interested in. Two versions were proposed, linear-ROIAS and logarithmic-ROIAS, which were shown to improve the quality on the AMUI when compared to QOAS [17].

Rate control, rate shaping and rate adaptive solutions can also be used to overcome issues with the mobile devices due to different screen sizes, processing capability, battery capacity, etc. For example, the SVC (Scalable Video Codec) [18] extension of

the H.264/AVC standard is an adaptive solution that can be used to address both network and device related issues. SVC enables temporal, spatial and fidelity scalability, as well as ROI (Region of Interest) scalability.

2) *Power-Aware Methods*: The battery capacity still represents one of the key limitations of mobile devices, especially when considering the power hungry nature of multimedia applications. Therefore, various researchers have explored possibilities for adapting the multimedia streaming in order to extend mobile devices' battery life. The use of rate control and adaptive encoding techniques for decreasing the battery power needed to receive, decode and display the multimedia content were considered.

Rate control-based solutions mainly focus on reducing the power consumption of the device wireless interface(s), by maintaining it in a lower power state for longer. The IEEE 802.11 standard integrates a mechanism called Power Saving Model (PSM) which allows the wireless card to go in a sleep state during the inactive periods [19]. Mechanisms to conserve the device battery power are also implemented in next generation wireless cellular networks such as IEEE 802.16m WiMAX and 3GPP LTE [20]. IEEE 802.16m provides advanced power saving mechanisms based on the enhanced versions of the legacy IEEE 802.16 sleep and idle modes, while LTE adopts a discontinuous reception mechanism for power conservation.

Since PSM was shown to not be very effective for real-time multimedia streaming due to the continuous packets flow [21], researchers have proposed to reshape the traffic flow and send the packets in bursts, thus allowing the WNIC to sleep between data reception intervals. Solutions include buffering the incoming data on the mobile device, or buffering the data corresponding to several beacon intervals at an intermediate node and releasing it only after the mobile device wakes up several times to check for incoming data [22]. To maintain the WLAN interface in sleep mode for significant periods of time during VoIP calls, the GreenCall algorithm uses sleep and wake-up schedules, where the sleep periods are computed based on the maximum delay that a user can tolerate during a conversation [23].

Solutions based on adaptive encoding techniques aim to decrease the power consumption by providing the end-user device with a version of the multimedia clip that is less energy expensive to be received and/or processed by the mobile device. These versions are created by changing the video compression to a less computational intensive one, or by decreasing encoding parameters such as the bitrate, the resolution or the frame rate [24], [25]. Various versions can be created off-line, in which case the adaptive mechanism selects the most suitable version to be displayed based on the available device battery power. Alternatively, the multimedia clip can be converted on the fly using real time transcoding on the server or at an intermediate proxy node [26].

Adaptive multimedia solutions to facilitate power saving through Dynamic Voltage and Frequency Scaling (DVFS) on various device components, have also been proposed. Baker *et al.* [27] proposed a solution to generate H.264 compressed streams with low priority macroblocks that can be ignored by the decoder based on the user preferences, in order to reduce the CPU workload and the associated power consumption through

DVFS. Amiri *et al.* [28] proposed a transcoding scheme to generate H.264 streams that are tolerant to defective memory cells of the decoding buffer, enabling power saving through voltage scaling on the memories. This solution has good power saving potential, as memories increase their share of the mobile device power consumption with the shift towards System-on-Chips (SoCs). Shin *et al.* [29] proposed to reduce the power consumption of OLED displays, by scaling down the voltage supply, and applying a solution based on human perceived color space to compensate the degradations occurring for bright clips.

Literature review shows that multimedia content adaptation can be used to significantly reduce mobile devices' battery power consumption, in all the stages of multimedia streaming: reception, decoding and displaying.

### B. Multimedia Quality Assessment

Various factors influence the quality of the multimedia content that is being delivered to an end-user. These factors are specific to the media encoding, to the network transmission and to the content itself [30]. Multimedia quality is limited by parameters such as the resolution, the frame rate, the codec and the level of compression used, and is prone to further degradations during transmission, due to variable network conditions. Furthermore, the quality may also be influenced by the performance of the device and the application displaying the content, as well as by factors related to the viewers such as their visual acuity, expectations, capacity to assimilate information, etc.

Multimedia quality assessment aims to address both the Quality of Service (QoS) and the broader aspect of user QoE. While QoS is a measure of service performance, being mainly concerned with network performance and data transmission, QoE was recently adopted by ITU-T to describe "the overall acceptability of an application or service as perceived subjectively by the end-user" [31].

Various methods for assessing the video and audio quality of multimedia applications have been proposed. However, since the BitDetect mechanism presented in this paper involves reducing the video bitrate of a multimedia clip, only methods for video quality assessment are presented. The video quality can be evaluated using subjective methods and objective metrics.

1) *Subjective Assessment Methods*: Subjective methods are considered the most accurate and reliable way for assessing the video quality. Several methods are standardized by ITU in the recommendations ITU-R Rec. BT.500 [32] for television and ITU-T Rec. P.910 [33] for multimedia applications. These standards provide useful guidelines and instructions regarding the selection of the subjects and of the test material, the setup of the test environment, the rating scales to be used for assessment, as well as the methods for analysing the data.

The various methods that are presented in the two recommendations can be classified in two categories: double *stimulus* (DS) and *single stimulus* (SS) methods. In case of DS methods such as the Degradation Category Rating (DCR) [33], viewers are presented with pairs of sequences and are asked to rate either each sequence individually or to rate the difference between them. SS methods such as the Absolute Category Rating (ACR) [33], aim to enable for a higher number of test sequences to be rated in

the same testing duration. Various discrete (e.g. 1-Bad to 5-Excellent) or continuous (e.g. 1 to 100) scales are used for rating purposes.

Another subjective method for assessing the video quality of multimedia applications that is currently standardized by ITU is SAMVIQ - Subjective Assessment Methodology for Video Quality [34]. SAMVIQ differs from other existing subjective methods through its multi-stimulus form of presenting the test sequences. To enable more accurate ratings, subjects are provided with access to multiple random sequences at the same time. Furthermore, the subjects can choose to start, stop and replay the sequences as they wish, and to change their ratings.

Subjective quality scores, represented by the Mean Opinion Score (MOS), are obtained by averaging the ratings across all viewers. The standard rating scales used to express the MOS scores, are often considered to not accurately indicate users' acceptability as defined by QoE. Therefore new research has concentrated to deploy subjective methods based on acceptability and to map the MOS scores to the latter. Acceptability is defined as a binary measure for which subjects have to state if they find the quality acceptable or unacceptable [35].

Although subjective methods are more accurate, they present many disadvantages. The need for human subjects and for a controlled environment makes them time and resource expensive. Furthermore these cannot be used repeatedly in various scenarios such as real time quality monitoring in multimedia applications. Therefore much effort has been made both by the industry and the research community, resulting in a significant number of metrics that try to quantify objectively the video quality as perceived by human subjects.

2) *Objective Assessment Metrics*: Objective metrics present a number of advantages that make them suitable for a wide range of applications such as quality monitoring of transmission services, as well as performance evaluation and optimization of video processing systems [36]. The objective metrics are evaluated depending on how well their values correlate with the subjective MOS scores. Statistical measures such as the Spearman's rank ordered correlation coefficient (SROCC) or root-mean-squared error (RMSE) are used for this purpose [37].

There are various approaches for classifying these metrics. Relative to the presence of the original reference signal, these metrics can be classified in: *no-reference* (NR), *reduced-reference* (RR) and *full-reference* (FR) metrics [38].

NR metrics have a high flexibility since they do not require the presence of the original video, unaffected by the factors under test. The majority of the metrics in this category estimate the blockiness, a distortion common to block-based compression algorithms such as H.264. Other, so called bitstream-based metrics rely mostly on indicators that can be extracted from the packet stream, without decoding it [39]. Using pertinent factors such as motion, the authors in [40] have proposed a solution to predict if a lost packet will be visible in the MPEG-2 video and with what probability. For a survey of NR metrics based on factors such as blockiness, blur, jerkiness, ringing, motion patterns, etc., the reader is directed to [41].

RR metrics use only some information from the reference video and aim to provide a compromise between the measurement accuracy and flexibility of use. Such information may be

represented by the amount of motion or spatial detail [39]. The features extracted have significantly lower bitrates than the reference signals, being more feasible to be transmitted on a separate channel or as an extra header [42].

FR metrics require the presence of the reference video, as well as precise spatial and temporal synchronizations, and luminance and color calibration between the original and the impaired videos. However, with these additional requirements comes higher quality estimation accuracy. Therefore FR metrics are suitable in a laboratory setup for applications such as for example video codecs benchmarking.

Mean Squared Error (MSE) often converted to the Peak Signal-to-Noise Ratio (PSNR), is the most widely used FR method due to its simplicity, even though is often criticized for having a poor correlation with the subjective tests [43]. To overcome its limitations, significant research work has concentrated on proposing objective metrics that aim to model the Human Visual System (HVS) [44]. These metrics incorporate factors such as color perception, contrast sensitivity or pattern masking, and were shown to better correlate with the subjective scores [45]. Examples of such metrics are the Structural Similarity Index (SSIM) [46] and the Video Quality Metric (VQM) [47]. SSIM is based on the idea that the human visual perception is adapted to extract structural information, while VQM is a metric that measures the effect of multiple factors such as blurring and blockiness. For a comprehensive review of RR and FR metrics the reader is directed to [48].

Since the evaluation of objective metrics is usually limited to indicating how well they correlate with the subjective scores, clear mappings between the objective values and the equivalent MOS scores are usually not available in the literature. This makes difficult to interpret how the values of the objective metrics reflect the subjective user-perceived video quality. Furthermore, many of the most accurate and popular metrics are based on the visual statistics and features of natural scenes [48], and these may not be very accurate for clips that present non-natural scenes or texts such as for example animations and screencasts. While there are other non-natural scene-based metrics, it is difficult to identify from the literature which metrics are better to assess the quality of non-natural clips. This is because non-natural clips are considered to complicate the evaluation of objective metrics and are usually avoided in the performance evaluation studies (e.g. [49]). Therefore, more research studies are needed to link the objective values to the equivalent MOS scores, and to assess the performance of different metrics across both natural and non-natural videos with different characteristics.

### III. BITDETECT—OBJECTIVE METRICS-BASED BITRATE ESTIMATION MECHANISM

#### A. Overview

Multimedia streaming applications make use of various video compression techniques in order to reduce the video bitrate to values that are feasible to be transmitted over the existing wired and wireless networks. Since the video quality is also influenced by the content characteristics, different clips encoded with the same bitrate may have different perceived quality levels. Therefore, for a particular video compression, resolution and frame

rate, more complex and dynamic videos will require higher bitrates for the same quality level (e.g. bad, good, or excellent).

In the context of this paper we call a *reference bitrate*  $X(RefBR_X)$ , a compressed video bitrate value that offers an excellent visual quality level for the wide majority of the users, even for a video content with a high level of complexity. This bitrate corresponds to a particular *reference resolution*  $X(RefRES_X)$ , and depends on the compression technique used. Multiple *video profiles* ( $Profile_X$ ) consisting of pairs of reference values ( $RefRES_X, RefBR_X$ ), can be predefined and associated to different mobile devices with different screen resolutions. Alternatively, mobile devices can be grouped in classes based on their screen resolutions, and a video profile associated to each class. The multitude of video profiles can be represented in an abstract form, as in (1).

$$Profile_X \in \{(RefRES_X, RefBR_X)\}, X = \overline{1, M} \quad (1)$$

where  $X$  indexes  $M$  video profiles corresponding to  $M$  different mobile devices or classes of devices.

Having a video clip  $Mov^i$  with a particular video resolution  $MovRES^i$ , and a video bitrate  $MovBR^i$ , a set of *reference versions*  $\{MM_X^i\}$  targeting different mobile devices or classes of devices can be created by transcoding the clip at different reference resolutions and the corresponding reference bitrates. Equations (2) and (3) represent the original clip  $Mov^i$ , and a particular reference version  $MM_X^i$ .

$$Mov^i = (MovRES^i, MovBR^i) \quad (2)$$

$$MM_X^i = (RefRES_X^i, RefBR_X^i) \quad (3)$$

Furthermore, the process of creating the reference versions from the original clip is represented in abstract form, as in (4).

$$Mov^i \xrightarrow{\text{Transcoding}} \{MM_X^i\}, X = \overline{1, N}, N \leq M \quad (4)$$

where  $X$  indexes  $N$  reference versions created from the video clip  $Mov^i$  that correspond to  $N$  different video profiles. The number of versions  $N$  is limited by the number of video profiles defined  $M$ , and by the quality of the original clip  $Mov^i$ . As indicated by (5), reference versions are created only for the video profiles with a reference resolution lower than the resolution of the original clip.

$$RefRES_X^i = RefRES_X \text{ and } RefRES_X \leq MovRES^i \quad (5)$$

Furthermore, depending on the content complexity and of the compression technique used, the bitrates of the reference versions created are limited both by the bitrates predefined for the profiles and the bitrate of the original clip, as in (6).

$$RefBR_X^i \leq RefBR_X \text{ and } RefBR_X^i \leq MovBR^i \quad (6)$$

Ideally, high-resolution uncompressed videos are desired in order to create high-quality reference versions covering a wide spectrum of devices. In practice however, only professional cameras offer uncompressed video recording, while most of the consumer HD cameras compress the videos to overcome storage space limitations. Due to additional distortions, the quality of a reference version created from a compressed

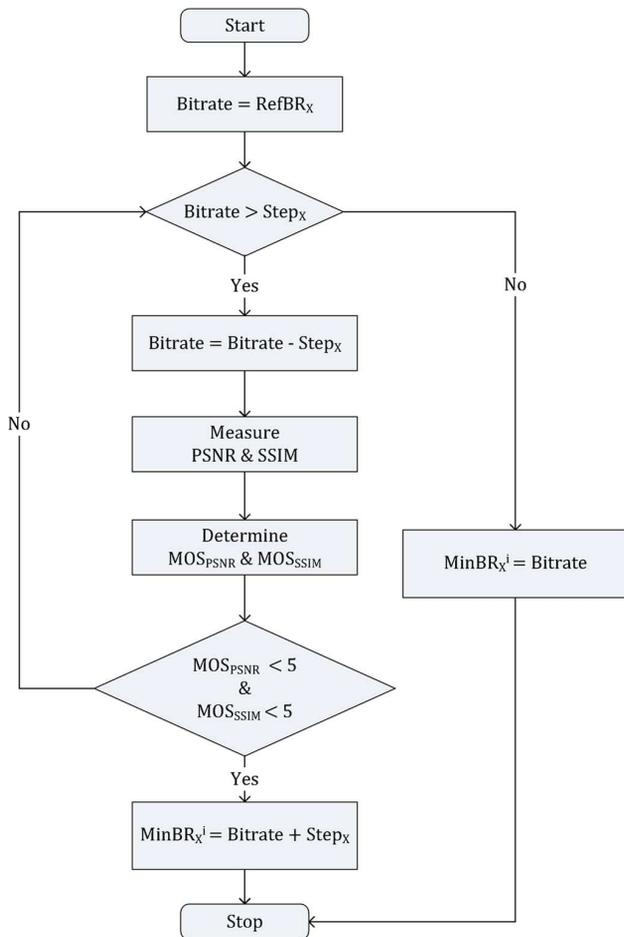


Fig. 1. BitDetect's objective metrics-based procedure for detecting the  $MinBR_X^i$  bitrate threshold.

high-quality source may be significantly lower than the quality of the same version created from an uncompressed source.

When a user mobile device requests a multimedia clip  $Mov^i$ , the reference version  $MM_X^i$  with the most suitable resolution is selected. The BitDetect mechanism decreases the video bitrate from the reference excellent quality value, in order to save battery power. The bitrate reduction is done in a quality-aware way, in the sense that the bitrate is not decreased below a threshold  $MinBR_X^i$  that still offers a good level of visual quality at the corresponding resolution  $RefRES_X^i$ . The procedure used by the BitDetect mechanism to detect the  $MinBR_X^i$  threshold for multimedia clips with different content characteristics and different resolutions is presented next.

### B. $MinBR_X^i$ Threshold Detection

For a particular reference resolution  $RefRES_X$ , the bitrate value that provides a good visual quality level may vary depending on the content characteristics. Therefore, a bitrate threshold  $MinBR_X^i$  has to be detected separately for each particular reference version  $MM_X^i$  of a particular multimedia clip  $Mov^i$ . This bitrate threshold has to provide at least a good user perceived quality level. The 1–5 scale was used when defining perceived quality levels.

Fig. 1 presents the procedure used by the BitDetect mechanism in order to detect the bitrate threshold  $MinBR_X^i$  for a

reference version  $MM_X^i$  of a multimedia clip  $Mov^i$ . The mechanism consists in creating multiple versions of that multimedia clip, having the same resolution  $RefRES_X$ , but lower video bitrate values than the reference bitrate  $RefBR_X^i$ , and estimating their quality using objective video quality assessment metrics.

The lower bitrate values are generated by gradually decreasing the reference bitrate  $RefBR_X^i$  with a fixed step  $Step_X$ . The step can be predefined for each particular reference, depending on the desired granularity (e.g. 32 Kbps, 64 Kbps, 128 Kbps, etc.). The multimedia clip versions ( $MM_{X_j}^i$ ) created by the BitDetect mechanism are represented in an abstract form, as in (7).

$$MM_{X_j}^i \in \{(RefRES_X, Bitrate_{X_j}^i)\}, j = \overline{1, P} \quad (7)$$

where  $j$  indexes  $P$  versions of the multimedia clip  $Mov^i$  with the resolution  $RefRES_X$  created by BitDetect mechanism. The video bitrate of these versions takes values in a collection, as in (8).

$$Bitrate_{X_j}^i \in \{RefBR_X^i - Step_X, RefBR_X^i - 2 \cdot Step_X, \dots, MinBR_X^i\} \quad (8)$$

The quality of the lower bitrate versions  $MM_{X_j}^i$  is estimated as compared to the quality of their corresponding reference version  $MM_X^i$  using the PSNR [43] and SSIM [46] full-reference video quality assessment metrics. PSNR provides a baseline for video quality assessment, and continues to be widely used due to its reduced complexity and computation load. Since PSNR is often criticized for being a poor quality metric, this is combined with SSIM, another popular metric that was shown to have a higher correlation with the subjective tests.

The subjective video quality is usually expressed using discrete MOS scales such as for example a 5-point scale (1—Bad to 5—Excellent). As opposed, the objective metrics return quality values on various specific continuous scales such as for example 1–100 for PSNR or 0–1 for SSIM. Since the video quality is usually a non-linear measure, it is difficult to interpret the objective quality values in terms of MOS scores. While other metrics were shown to more accurately predict the subjective scores, PSNR and SSIM were preferred due to the availability of clear interpretations of their quality values in terms of subjective MOS scores. However, the BitDetect mechanism can use any other objective metric as long as its values are mapped to the MOS scores.

The PSNR for a video sequence is determined based on the average MSE across all the frames. Since SSIM is an image quality assessment metric, this is applied to videos by computing it frame-by-frame and averaging the values across all the frames. After computing the PSNR and SSIM metrics, the corresponding quality values are translated in discrete quality levels on the MOS scale ( $MOS_{PSNR}$  and  $MOS_{SSIM}$ ), using the mapping table proposed by Zinner *et al.* [50] (see Table I). The  $MinBR_X^i$  threshold determined by the BitDetect mechanism represents the lowest bitrate value at which at least one of the  $MOS_{PSNR}$  and  $MOS_{SSIM}$  scores is still 5 (Excellent). The detection mechanism stops when at least one of the  $MOS_{PSNR}$  and  $MOS_{SSIM}$  scores become for the first time equal to 4, after a

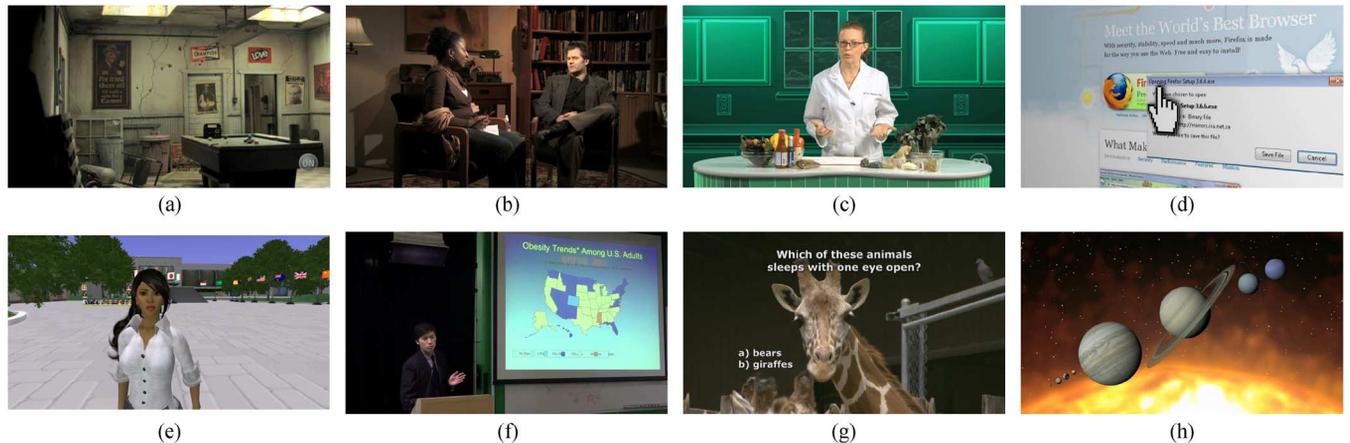


Fig. 2. Representative frames for the eight multimedia clips. (a) Arts. (b) Dubus. (c) Hotness. (d) Hulu. (e) Languagelab. (f) Obesity. (g) Sleep. (h) Sol.

TABLE I  
MAPPING OF OBJECTIVE VIDEO QUALITY TO SUBJECTIVE  
PERCEIVED QUALITY [50]

<i>MOS</i>	<i>PSNR [dB]</i>	<i>SSIM</i>
5 (Excellent)	$\geq 45$	$\geq 0.99$
4 (Good)	$\geq 33$ & $< 45$	$\geq 0.95$ & $< 0.99$
3 (Fair)	$\geq 27.4$ & $< 33$	$\geq 0.88$ & $< 0.95$
2 (Poor)	$\geq 18.7$ & $< 27.4$	$\geq 0.5$ & $< 0.88$
1 (Bad)	$< 18.7$	$< 0.5$

$Step_X$  reduction in bitrate was applied. Alternatively, the mechanism stops when the bitrate of the last version is lower than the  $Step_X$ , and thus cannot be decreased any further.

While, other mappings such as for example the one proposed for PSNR in [51] can be used as well, the mapping from [50] was preferred as it has higher threshold for the Excellent quality level (i.e. PSNR  $> 45$  dB vs. PSNR  $> 37$  dB). Since in practice the original clips  $Mov^i$  may be already compressed, quality degradations will be passed to the reference version  $MM_X^i$  and further to  $MM_{X_j}^i$ . Therefore, using a higher quality threshold can partially compensate for these degradations.

The BitDetect mechanism can be easily tuned to stop at a lower quality level, in order to achieve higher bitrate reductions. This can be done when additional battery power needs to be saved, or to reduce the impact of transmission related factors (e.g. loss, jitter, etc.) on the user perceived quality. However, the subjective study presented in this paper did not address the impact of network transmission factors.

#### IV. TESTING SETUP AND METHODOLOGY

A subjective study was conducted to assess if the bitrate threshold provided by the BitDetect mechanism offers at least a good user perceived quality level for different multimedia clips with various characteristics. Lately, multimedia content has started to be increasingly used in mobile learning. Since educational multimedia clips usually have knowledge transfer purposes, decreasing the quality can impact the learning process. Therefore, the study has concentrated on educational multimedia clips having in mind to test also whether the quality

is good-enough for learners to achieve knowledge from multimedia clips encoded at the recommended bitrates. However, the BitDetect mechanism may be used with any type of multimedia clips.

Two mobile devices with different screen sizes were used in order to assess the performance of the proposed mechanism for various resolutions. To assess if the recommended bitrate thresholds enable battery power saving, an experimental test in which the test sequences were streamed to one of the devices was also conducted.

#### A. Multimedia Clips

Eight high-quality H.264 compressed clips were used for the subjective study. The clips were selected from large collections of educational clips that are available freely for download on iTunes U [52] and Miro Guide [53].

The eight clips are representative for the broad category of educational clips, covering a large spectrum in terms of content type and dynamicity. The clips have different characteristics and correspond to various categories of educational clips such as screencasts, slideshows, presentations, animations, etc. Fig. 2 shows representative frames from each multimedia clip. A short description of the multimedia clips is provided below.

- *Arts*—slideshow, consisting of sequences of images with high level of details, and low dynamicity, represented by the transitions between the images (e.g. fade-in/out, zoom in/out, etc.);
- *Dubus*—interview, presenting low level of dynamicity due to almost static characters on a static background;
- *Hotness*—live demonstration, with medium-low dynamicity, presenting a demonstrator that shows how to perform different tasks;
- *Hulu*—computer screen recording (screencast), with high level of details (e.g. text, buttons, etc.) and medium-low dynamicity (e.g. pop-up windows, text typing, cursor moving, etc.);
- *Languagelab*—video presenting a computer generated 3D Virtual Learning Environment, with low-contrast scenes and low dynamicity (mostly static characters);

TABLE II  
CHARACTERISTICS OF THE ORIGINAL HIGH-QUALITY MULTIMEDIA CLIPS

Multimedia Clip	Educational Clips Category	Video Encoding Characteristics					
		Resolution [pixels]	Frame Rate [fps]	Bitrate [Kbps]	H.264 Compression		
					Profile	B frames	CABAC
Arts	Slideshow	1280x720	24	3700	Main	Yes	No
Dubus	Interview	1280x720	30	5000	High	Yes	No
Hotness	Lab Demo	1280x720	24	4800	Main	Yes	No
Hulu	Screencast	1280x720	30	2000	High	No	Yes
Languagelab	Graphics	1280x720	30	2000	High	No	Yes
Obesity	Presentation	1280x720	30	2000	High	No	Yes
Sleep	Documentary	1280x720	30	5200	Main	Yes	No
Sol	Animation	960x540	30	3400	Main	Yes	No

- *Obesity*—clip characteristic to lecture recordings and presentations, with medium-low dynamicity, consisting of a static presenter and a slideshow projection on the background;
- *Sleep*—natural scenes with different levels of dynamicity characteristic to documentaries; textual information in the form of questions and answers are present in the video;
- *Sol*—computer generated animations of presenting concepts such as i.e. the planets orbiting the Sun, with medium-low dynamicity.

Table II presents the encoding characteristics of the 8 multimedia clips. MediaInfo [54] software was used for extracting this information. As already mentioned they were originally compressed using H.264/MPEG-4 AVC video compression. The Sol multimedia clip had a resolution of  $960 \times 540$  pixels, while the rest of the clips have a resolution of  $1280 \times 720$  pixels. Arts and Hotness clips had a frame rate of 24 frames per second, whereas the rest of the clips had a frame rate of 30 frames per second. The video bitrate of the eight original clips varies between 2000 Kbps and 5200 Kbps. H.264 is a complex codec with a multitude of settings that can impact the quality and the bitrate reduction [55]. In order to provide a better idea about the quality of the original clips, Table II presents also additional information that was possible to be extracted about the H.264 compression: the profile used, and the presence of B-frames and CABAC (Context-adaptive binary arithmetic coding). Four clips (Arts, Hotness, Sleep and Sol), were compressed using the Main profile, while the other four (Dubus, Hulu, Languagelab and Obesity) used the High profile. B-frames and CABAC can provide additional bitrate saving or a higher quality for the same bitrate, at the expense of higher encoding/decoding complexity. The three clips with the lowest bitrate (2000 Kbps), did not present B-frames but present CABAC, while the other five clips, with higher bitrates, present B frames but did not present CABAC.

The different video encoding schemes used for compressing the eight clips may have an impact on the results of the study. Unfortunately, while a number of Web sites such as CDVL [56] and Xiph.org [57] provide some uncompressed or high-quality clips for research or testing purposes, it was not possible to find high-quality educational clips compressed with the same settings for all the categories identified.

All the eight original clips presented stereo sound (2 channels), compressed using the LC (Low-Complexity) profile of the AAC (Advanced Audio Codec) audio standard. The clip Arts

had an audio bitrate of 160 Kbps and a sampling rate of 48 KHz, while the other seven clips had an audio bitrate of 128 Kbps and a sampling rate of 44.1 KHz.

### B. Generating Reference Test Sequences for Mobile Devices

Mobile devices have different screen resolutions and video playback capabilities that need to be taken into account when encoding a multimedia clip. The eight high resolution multimedia clips are not suitable to be streamed and displayed on the wide majority of mobile devices. However, they were selected in order to be able to create reference versions (see Section III-A) that offer excellent quality level and are suitable for mobile devices.

The perceived visual quality of a multimedia clip may be dependent on the device screen size and its resolution. Therefore, the subjective testing was conducted for two different resolutions,  $640 \times 360$  and  $856 \times 480$ , which are suitable for mobile devices such as smartphones, tablets and netbooks. The original aspect ratio of 16:9 was maintained.

Reference video bitrates of 2000 Kbps and 1200 Kbps were associated for the two video resolutions respectively. The values were selected based on recommendations found in [58]–[60], assuming that the above values offer an excellent visual quality level even for the clips with the highest level of dynamicity. The reference bitrates correspond to the H.264 / MPEG-4 AVC video compression.

Two test sequences (part A and part B) were extracted from each of the eight original multimedia clips. The sixteen sequences correspond to the eight categories of educational clips, two sequences per category. Their length varies between 20 and 30 seconds long (see Table III), because each test sequence was selected to present a learning concept. In this way, educational questions can be asked in order to assess if the bitrate thresholds recommended by the BitDetect mechanism provide a good quality level that supports the learning process.

Eight of the test sequences (Arts A, Dubus A, etc.) were downsampled to create reference versions for the  $640 \times 360$  resolution, while the other eight sequences (Arts B, Dubus B, etc.), were downsampled to create the reference versions for the  $856 \times 480$  resolution. This was performed using XMedia Recode [61] media converter software, with  $\times 264$  [62] video encoder and FAAC [63] audio encoder. The resolution scaling-down was done using bicubic interpolation in order to minimize the distortion artifacts due to aliasing.

TABLE III  
CHARACTERISTICS OF THE 16 REFERENCE TEST SEQUENCES

Test Sequence	Length	Resolution	Video Bitrate	Frame
	A/B [sec]	A/B [pixels]	A/B [Kbps]	Rate [fps]
Arts A/B	20 / 27			24
Dubus A/B	30 / 25			30
Hotness A/B	30 / 30			24
Hulu A/B	27 / 21	640x360 /	1200 / 2000	30
LanguageLab A/B	30 / 30	856x480		30
Obesity A/B	30 / 30			30
Sleep A/B	26 / 20			30
Sol A/B	30 / 30			30

TABLE IV  
ENCODING SETTINGS FOR THE REFERENCE TEST SEQUENCES

Parameter	Value
Container	.mp4
<i>Video Settings</i>	
Compression	H.264 / MPEG-4 AVC, Main Profile, No B-Frames, No CABAC
Resolution	856x480 / 640x360
Bitrate	2000 Kbps / 1200 Kbps
Bitrate Mode	2-Pass Average Bitrate
Frame Rate	Original
<i>Audio Settings</i>	
Compression	AAC-LC
Channels	2 (Stereo)
Bitrate	128 Kbps
Sample Rate	44.1KHz

Apart from the video bitrate and the resolution, all other encoding parameters were maintained constant across the sixteen reference-quality sequences. A summary of the encoding settings are presented in Table IV. 2-Pass Average Bitrate video encoding mode was chosen, because of the better quality offered as compared to 1-Pass encoding mode. Using this mode, the compressed video sequences have a bitrate that is identical to the reference bitrate in all the cases. B-frames and CABAC were not used, due to the low decoding complexity required by mobile devices.

The sound was maintained during the subjective testing in order to provide a more natural viewing experience. Since the presence of audio may influence the subjective video quality scores, additional care was taken to reduce this impact. Audio impairments, as well as a good audio quality were shown to have an impact on viewers' ability to identify video artifacts [64], [65]. However, the impact of audio quality on perceived video quality was shown to be significantly lower than the impact of video quality on perceived audio quality of an audio-visual sequence (0.2 points versus 1.2 points on a 9-point scale) [66]. In order to focus users' attention to the video quality assessment, the audio components at high quality and with the same settings were included in all the test sequences. AAC sound compressed at 64 Kbps was shown to provide comparable audio quality with uncompressed sound in subjective listening tests [67]. Therefore, the 128 Kbps audio bitrate used for this experiment can be considered to provide an excellent audio quality level.

The decision to use different sequences for the two devices was made in order to avoid influencing the answers to the educational questions. If the same sequence is displayed on the

TABLE V  
CHARACTERISTICS OF THE MOBILE DEVICES

Device	Characteristics
HP iPAQ 214	OS: MS Windows Mobile 6
	Memory: 128 MB SDRAM
	CPU: Marvell PXA310, 624MHz
	Screen Size: 4"
	Screen Resolution: 640x480 (VGA)
Dell Inspiron Mini 10	OS: MS Windows XP Home Edition
	Memory: 1 GB DDR2
	CPU: Intel Atom N270, 1.6 GHz
	Screen Size: 10.1"
	Screen Resolution: 1024x576 (WSVGA)

two devices, learners would answer the first educational question after seeing only once the sequence on the first device, and the second question after viewing the sequence twice, on both devices.

### C. $MinBR_X^i$ Threshold Detection for the Test Sequences

After the reference versions were created for the sixteen test sequences, their corresponding  $MinBR_X^i$  bitrate thresholds were detected by applying the BitDetect mechanism. For each test sequence multiple versions with lower bitrate values than the reference bitrate were created and their quality was measured using the SSIM and PSNR metrics. Apart from the lower video bitrate, these versions are created using the same encoding settings as for the corresponding references (see Table IV). MSU Video Quality Measurement Tool [68] software was used for measuring the two metrics. The bitrate step between successive lower bitrate versions was selected equal to 64 Kbps for the  $640 \times 360$  resolution and 128 Kbps for the  $856 \times 480$  resolution.

Therefore, for sequences with a  $640 \times 360$  resolution, the reference video bitrate of 1200 Kbps was reduced gradually with the step of 64 Kbps, and the quality of the resulting versions was assessed based on the SSIM and PSNR metrics. Using the BitDetect mechanism, the  $MinBR_X^i$  bitrate threshold was detected as 512 Kbps for the Sleep A sequence and 256 Kbps for the other seven sequences. Due to the higher level of motion content, the Sleep A sequence (documentary) required double video bitrate for the same quality level.

Following the same procedure for the sequences with an  $856 \times 480$  resolution, the reference video bitrate of 2000 Kbps was reduced gradually with a step of 128 Kbps. The  $MinBR_X^i$  bitrate threshold was detected as 768 Kbps for the Sleep B test sequence and 384 Kbps for the other seven test sequences.

### D. Test Setup

Two mobile devices were used during the subjective testing. The first device, a HP iPAQ 214 PDA, displayed the test sequences with a  $640 \times 360$  resolution. The second device, a Dell Inspiron Mini 10 Netbook, displayed the test sequences with an  $856 \times 480$  resolution. The characteristics of the two devices are summarized in Table V.

In order to maintain a constant luminance level across all participants the test room was shielded from natural light. The screen brightness was set to 100% on both devices in order to offer a maximum visibility for the subjects. To further control the noise level, the subjects were provided with headphones.

TABLE VI  
RESULTS FOR  $640 \times 360$  SEQUENCES

<i>Test Sequence</i>	Bitrate [Kbps]	MOS	STDEV	PSNR [dB]	$MOS_{PSNR}$	SSIM	$MOS_{SSIM}$
Arts A	256	4.0	0.89	43	4	0.99	5
Dubus A	256	4.5	0.68	45	5	0.98	4
Hotness A	256	4.5	0.75	43	4	0.99	5
Hulu A	256	4.0	0.95	44	4	0.99	5
Languagelab A	256	3.6	1.12	43	4	0.99	5
Obesity A	256	3.9	0.62	46	5	0.99	5
Sleep A	512	4.0	0.84	45	5	0.98	4
Sol A	256	4.5	0.68	45	5	0.99	5
Average		4.1	0.82	44.3	4.5	0.988	4.8

TABLE VII  
RESULTS FOR  $856 \times 480$  SEQUENCES

<i>Test Sequence</i>	Bitrate [Kbps]	MOS	STDEV	PSNR [dB]	$MOS_{PSNR}$	SSIM	$MOS_{SSIM}$
Arts B	384	3.7	1.10	45	5	0.98	4
Dubus B	384	4.5	0.75	44	4	0.99	5
Hotness B	384	4.3	0.72	43	4	0.99	5
Hulu B	384	4.2	0.89	46	5	0.99	5
Languagelab B	384	3.9	1.01	47	5	0.99	5
Obesity B	384	3.8	0.77	44	4	0.99	5
Sleep B	768	4.0	0.86	44	4	0.99	5
Sol B	384	4.5	0.60	47	5	0.99	5
Average		4.1	0.84	45.0	4.5	0.989	4.9

Only one participant was allowed to do the test at any moment in time. The test sequences were stored and played locally on the mobile devices. This was done in order to reduce to minimum the influence of other factors such as the variable network conditions on the perceived quality.

CoreCodec TCPMP [69] open source media player was used for playing the test sequences on the PDA device, and Google Chrome [70] web browser with HTML5 video support for the H.264 codec, was used for playing the sequences on the Netbook device. The sequences were displayed in full-screen on the PDA, since the screen resolution width was equal with the video resolution width at 640 pixels. On the Netbook, the test sequences were displayed on a 50% level of gray background in the middle of the screen, at their native resolution of  $856 \times 480$  pixels, to avoid any distortions caused by the scaling-up process.

### E. Subjective Assessment Methodology

The Absolute Category Rating (ACR) [33] method for assessing the user perceived video quality of multimedia sequences was used during the subjective study. The subjects visualized the sixteen test sequences on the two devices and they rated their perceived quality independently on the five level quality scale (i.e. 1—Bad to 5—Excellent). Each sequence had a bitrate equal with the  $MinBR_X^i$  level detected by the BitDetect mechanism (see Section IV-C).

First, the test sequences with a  $640 \times 360$  resolution were displayed in a random order on the PDA device and rated by the subjects. After that the test sequences with an  $856 \times 480$  resolution were displayed in a random order on the Netbook device and rated by the subjects.

The voting time was not limited to a fixed duration, the participants deciding when they were ready to view the next sequence. The subjects were not able to pause, stop, or review the sequences.

Each subject was explained the testing procedure before starting the test. Personal information and preferences were collected using a questionnaire. Furthermore, a training session with different video sequences was provided in order to make the subjects familiar with the assessment procedure.

Twenty-one (thirteen male, eight female) non-expert subjects participated in the study. They ranged from different domains: undergraduate students, postgraduate students and professionals. Their age ranged from 21 to 37 years old, where the average age was 27.14 years old. All the subjects have reported that they had normal vision or have corrected to normal vision (they were wearing glasses).

## V. EXPERIMENTAL RESULTS ANALYSIS

### A. Subjective Video Quality Assessment Results

To assess if the bitrate values recommended by the BitDetect mechanism offer a good quality, the subjects were asked to watch the test sequences and rate their perceived quality on the 1 (Bad) to 5 (Excellent) scale. For each sequence, the mean value represented as the MOS score, and the standard deviation (STDEV) of the statistical distribution of the assessment grades were computed. The subjective video quality assessment results are presented in Tables VI and VII for the two resolutions  $640 \times 360$  and  $856 \times 480$  respectively. The two tables also present the objective PSNR and SSIM values, as well as the equivalent  $MOS_{PSNR}$  and  $MOS_{SSIM}$  scores.

Subjective results show that the majority of the sixteen sequences have scored close to level 4 (Good) with an average of

4.1. One sequence with a  $640 \times 360$  resolution (Languagelab A), and two of the sequences with an  $856 \times 480$  resolution (Arts B and Obesity B) have scored slightly lower than 4, but higher than 3.5. Pearson correlation coefficient ( $r$ ) was computed in order to assess if the compression rate of the original multimedia clips influenced the MOS scores achieved by the test sequences. The results show that the correlation between the bitrate of the original clips from which the sequences were extracted and the subjective MOS scores, is significant only when this is computed across all sixteen sequences ( $r = .515, p < .05$ ). Individually for the two resolutions there is no significant correlation ( $r = .634$  for  $640 \times 360$  sequences,  $r = .387$  for  $856 \times 480$  sequences;  $p < .05$ ). While it may be important to have high quality original clips for a more accurate detection of the bitrate threshold, the low correlation values indicate that the subjective MOS scores were not significantly impacted by the different compression rates of the original clips.

When assessing individually the two resolutions tested, Tables VI and VII show that the average MOS score across the two resolutions is the same (4.1). A two sample T-Test analysis on these means values, indicate with a 99% confidence level, that statistically there is no difference in the final scores for the two video resolutions being tested ( $\alpha = 0.01, t = 0.16, t - \text{critical} = 2.99, p(t) = 0.43$ ). The Pearson correlation between the MOS scores corresponding to the two resolutions is  $r = .81$ . Despite different test sequences from the same clip were used for the two resolutions, the results indicate that the BitDetect mechanism has similar performance across different resolutions characteristic to mobile devices.

The average MOS scores of 4.1 indicating a good quality level is close to the quality level indicated by the objective metrics. The average  $MOS_{PSNR}$  is equal with 4.5 for both resolutions. The SSIM metric, which measures more accurately the quality as perceived by the human subjects, indicates a slightly higher quality. The average  $MOS_{SSIM}$  is equal with 4.8 for the sequences with a  $640 \times 360$  resolution and 4.9 for the sequences with an  $856 \times 480$  resolution.

The conclusion that can be drawn from this study is that the bitrate values recommended by the BitDetect mechanism for different clips with different characteristics offer a good level of user perceived-quality. These values are characteristic to the H.264 video compression. Since, video quality is a complex measure that can be influenced by a multitude of factors, there may be situations when clips with lower bitrates than the ones recommended by BitDetect to be perceived as still having a good quality. While the proposed mechanism provides a more general method for detecting content-dependant bitrate thresholds, this could be further improved by considering other factors such as for example the user preferences.

Table VIII shows the bandwidth saving performance when BitDetect mechanism is applied in order to decrease the video bitrate, for the same test sequences. The results show that the video bitrate can be decreased with up to 80% for most of the educational type multimedia clips while still maintaining a good quality level. It is worth mentioning here that 1200 Kbps and 2000 Kbps are video bitrate levels that are commonly used for encoding H.264 multimedia content at these video resolutions.

TABLE VIII  
BITDETECT BANDWIDTH SAVING

Sequence	Reference Bitrate [Kbps]	Recommended Bitrate [Kbps]	Bandwidth Saving [%]	
Arts A	1200	256	78.6 %	
Dubus A		256	78.6 %	
Hotness A		256	78.6 %	
Hulu A		256	78.6 %	
Languagelab A		256	78.6 %	
Obesity A		256	78.6 %	
Sleep A		512	57.3 %	
Sol A		256	78.6 %	
Arts B		2000	384	80.8 %
Dubus B			384	80.8 %
Hotness B			384	80.8 %
Hulu B			384	80.8 %
Languagelab B			384	80.8 %
Obesity B			384	80.8 %
Sleep B			768	61.6 %
Sol B			384	80.8 %

TABLE IX  
EDUCATIONAL QUESTIONNAIRE RESULTS

Sequence	CRR	Sequence	CRR
Arts A	95 %	Arts B	100 %
Dubus A	100 %	Dubus B	95 %
Hotness A	95 %	Hotness B	95 %
Hulu A	86 %	Hulu B	100 %
Languagelab A	100 %	Languagelab B	100 %
Obesity A	95 %	Obesity B	24 %
Sleep A	90 %	Sleep B	100 %
Sol A	67 %	Sol B	62 %
Average	91 %		85 %

### B. Learning Process Related Assessment Results

In order to assess if the bitrate thresholds recommended by the BitDetect mechanism are suitable for multimedia-based learning, the subjects were also asked to answer an educational question after watching each of the sixteen test sequences [71]. Since the audio was maintained in order to replicate a normal viewing experience, the questions were carefully selected to make sure that their answers can be found only in the visual information being presented. The assumption made was that, if subjects are able to answer the questions after watching each clip a single time at the bitrate recommended by the BitDetect mechanism, than the quality of the clips is good-enough for learning.

The results for the educational questionnaire are presented in Table IX. The correct response rate (CRR), computed as the percentage of correct answers across the 21 subjects is provided for each of the 16 test sequences. The results show that the majority of the subjects were able to answer correctly the educational questions for 15 out of the 16 test sequences. The average correct response rate was 91% for the test sequences with a  $640 \times 360$  resolution, and 85% for the test sequences with an  $856 \times 480$  resolution.

### C. Battery Power Saving Results

An experimental testing was also conducted in order to assess the performance of the proposed BitDetect mechanism in terms of mobile device battery power saving [72]. Each of test

TABLE X  
BITDETECT BATTERY POWER SAVING PERFORMANCE

Sequence	Streaming Time [min]		Time Difference	
	256 / 512* Kbps	2000 Kbps	[min]	[%]
Arts A	36.33	31.37	4.97	15.83%
Dubus A	36.27	31.87	4.40	13.81%
Hotness A	37.07	32.27	4.80	14.88%
Hulu A	37.07	31.77	5.30	16.68%
Languagelab A	36.03	31.82	4.22	13.25%
Obesity A	36.55	31.80	4.75	14.94%
Sleep A*	34.75	30.75	4.00	13.01%
Sol A	36.47	31.63	4.83	15.28%
MIN	34.75	30.75	4.00	13.01%
MAX	37.07	32.27	5.30	16.68%
AVG	36.32	31.66	4.66	14.71%
STDEV	0.73	0.44	0.42	1.28%

sequences with the  $640 \times 360$  resolution was streamed wirelessly to the PDA device in two situations. In the first trial the sequences had the reference video bitrate of 1200 Kbps. For the second trial the test sequences had the video bitrate equal with the recommended values of 512 Kbps for the Sleep A sequence and 256 Kbps for the other seven sequences. Each sequence was streamed continuously in a loop for one hour discharging the battery from its full charge of 100%.

To reduce the impact of wireless network conditions on the results and enable a smooth playback, the devices were maintained at a fixed distance from the AP and no other traffic was present in the test network. Furthermore, the tests were conducted in an environment with few other wireless networks to reduce the interferences. To reduce the impact of the device settings on the results, the screen brightness level and the audio was maintained constant, the device power saving was disabled, while all applications except the video player and the battery logger were closed.

Battery information such as the battery level, voltage and temperature were logged at a 10 seconds frequency for the entire streaming duration. Since the device does not provide information about the battery current drain, the battery power saving is determined by comparing the streaming duration when discharging the battery between the same battery levels. While measuring the battery discharge time may not be the most accurate method, this provides a good idea about the additional device usage time a user can gain.

Table X presents the experimental results for a 20% battery discharge, from 95% level to 75% level. These values were selected as the battery discharge is almost linear in this range and increasingly non-linear as the battery depletes. The results show that the battery life can be increased by up to 17% when using the BitDetect recommended bitrates for streaming multimedia content to mobile devices. Despite the fact that seven test sequences have the same bitrate, different power saving results are achieved for these sequences. Since the same testing conditions were maintained when streaming each multimedia clip, the results indicate that the mobile device power consumption also depends on the decoding complexity of the various multimedia clips.

## VI. CONCLUSION

The research presented in this paper shows that the video bitrate can be significantly reduced from values that are commonly used for streaming multimedia content to mobile devices, in order to reduce the device power consumption, while maintaining a good user-perceived quality level. The BitDetect mechanism that uses PSNR and SSIM full-reference objective video quality assessment metrics in order to detect content-specific video bitrate levels that enable good user-perceived quality, was presented. Results from a subjective study indicate that full-reference metrics in general and PSNR and SSIM metrics in particular, can be used to detect good quality bitrate thresholds for multimedia clips with different content types and dynamics, and for different video resolutions. Furthermore, results from an educational questionnaire indicate that the bitrates recommended by BitDetect offer a good-enough quality to support the learning process, while experimental results show that significant battery power can be saved by using the recommended bitrates when streaming multimedia content to a mobile device.

Two challenges were faced while conducting this research. The first was represented by the unavailability of uncompressed or high-quality educational multimedia clips. Since high-quality educational clips encoded with the same settings could not be found for all the different categories, clips with different compression rates had to be used. Despite the reduced number of test sequences, the results indicate that the subjective video quality scores were not significantly impacted by the different bitrates of the original clips from which these sequences were extracted. The second challenge was represented by the limited research to clearly map the objective values of different metrics, as well as various bitrate values to subjective user-perceived quality levels (MOS scores). Full-reference objective metrics estimate the video quality of a clip as compared to its reference. Therefore, accurate bitrate threshold detection requires having a good idea about the perceived quality both of the reference clip and of its lower bitrate versions.

Future work will explore performance improvements of the BitDetect mechanism, in order to more accurately detect the bitrate threshold for both natural and non-natural clips. Possible options include using reduced-reference or no-reference objective metrics, as well as applying the principle of BitDetect to the subjective study, in order to compare the performance of various objective metrics with the subjective scores. Additionally, experimental tests with a wider range of devices and multimedia clips will be considered in order to assess the effect of device and coding complexity on the battery power consumption.

## ACKNOWLEDGMENT

The authors would like to thank Dr. G.-M. Muntean and Dr. S. Weibelzahl, for reading the manuscript and providing useful feedback, as well as to the two anonymous reviewers whose thoughtful comments have helped to significantly improve this paper.

## REFERENCES

- [1] In-Stat, "More than half of US handset shipments will be smartphones by 2012, worldwide smartphone shipments move toward 1 billion by 2015," 2011. [Online]. Available: <http://www.in-stat.com/newmk.asp?ID=3012&SourceID=00000501000000000000>

- [2] Cisco Systems, Inc., "Cisco visual networking index: Global mobile data traffic forecast update, 2010-2015," 2011. [Online]. Available: [http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white\\_paper\\_c11-520862.html](http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html)
- [3] G. Fleischman, "The state of 4G: It's all about congestion, not speed," *Ars Technica* 2010 [Online]. Available: <http://arstechnica.com/telecom/news/2010/03/faster-mobile-broadband-driven-by-congestion-not-speed.ars>
- [4] M. Kennedy, H. Venkataraman, and G.-M. Muntean, "Battery and stream-aware adaptive multimedia delivery for wireless devices," presented at the IEEE International Workshop on Performance and Management of Wireless and Mobile Networks (P2MNET), IEEE Local Computer Networks (LCN) Denver, CO, 2010.
- [5] Apple Inc., "Apple—iPhone 4S—Size, weight, battery life, and other specs," [Online]. Available: <http://www.apple.com/iphone/specs.html>
- [6] J. Zhang, D. Wu, S. Ci, H. Wang, and A. K. Katsaggelos, "Power-aware mobile multimedia: A survey (invited paper)," *Journal of Communications*, vol. 4, no. 9, pp. 600–613, 2009.
- [7] G.-M. Muntean and N. Cranley, "Resource efficient quality-oriented wireless broadcasting of adaptive multimedia content," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pp. 362–368, Mar. 2007.
- [8] A.-N. Moldovan and C. H. Muntean, "Personalisation of the multimedia content delivered to mobile device users," presented at the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB'09), Bilbao, Spain, 2009.
- [9] V. H. Muntean and G.-M. Muntean, "A novel adaptive multimedia delivery algorithm for increasing user quality of experience during wireless and mobile e-learning," presented at the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB'09), Bilbao, Spain, 2009.
- [10] H. Luo, S. Ci, D. Wu, and H. Tang, "Adaptive wireless multimedia communications with context-awareness using ontology-based models," presented at the GLOBECOM 2010, 2010 IEEE Global Telecommunications Conference, Miami, FL, 2010.
- [11] N. Cranley and L. Murphy, "Incorporating user perception in adaptive video streaming systems," in *Digital Multimedia Perception and Design*, G. Ghinea and S.Y. Chen, Eds. Hershey, PA: Idea Group Pub., 2006, pp. 242–263.
- [12] S. Floyd, M. Handley, J. Widmer, and J. Padhye, "TCP friendly rate control (TFRC): Protocol specification," 2008.
- [13] G. Yang, T. Sun, M. Gerla, M. Y. Sanadidi, and L. Chen, "Smooth and efficient real-time video transport in the presence of wireless errors," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, vol. 2, pp. 109–126, 2006.
- [14] S. Chang and A. Vetro, "Video adaptation: Concepts, technologies, and open issues," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 148–158, 2005.
- [15] G.-M. Muntean, P. Perry, and L. Murphy, "Objective and subjective evaluation of QOAS video streaming over broadband networks," *IEEE Trans. Network and Service Management*, vol. 2, no. 1, pp. 19–28, 2005.
- [16] G.-M. Muntean, G. Ghinea, and T. N. Sheehan, "Region of interest-based adaptive multimedia streaming scheme," *IEEE Trans. Broadcast.*, vol. 54, no. 2, pp. 296–303, Jun. 2008.
- [17] B. Ciubotaru, G.-M. Muntean, and G. Ghinea, "Objective assessment of region of interest-aware adaptive multimedia streaming quality," *IEEE Trans. Broadcast.*, vol. 55, no. 2, pp. 202–212, Jun. 2009.
- [18] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, 2007.
- [19] IEEE, "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," 2007.
- [20] R. Kim and S. Mohanty, "Advanced power management techniques in next-generation wireless networks," *IEEE Commun. Mag.*, vol. 48, no. 5, pp. 94–102, 2010.
- [21] J. Korhonen and Y. Wang, "Power-efficient streaming for mobile terminals," in *Proceedings of the International Workshop on Network and Operating Systems Support for Digital Audio and Video*, Stevenson, Washington, 2005, pp. 39–44.
- [22] J. Adams and G.-M. Muntean, "Adaptive-buffer power save mechanism for mobile multimedia streaming," in *Proc. IEEE International Conference on Communications 2007 (ICC '07)*, Glasgow, Scotland, U.K., 2007, pp. 4548–4553.
- [23] V. Namboodiri and L. Gao, "Energy-efficient VoIP over wireless LANs," *IEEE Trans. Mobile Computing*, vol. 9, no. 4, pp. 566–581, 2010.
- [24] A.-N. Moldovan and C. H. Muntean, "Assessing power saving techniques and their impact on e-learning users," in *Proceedings of the 2009 China-Ireland International Conference on Information and Communications Technologies (CICT)*, Maynooth, Ireland, 2009, pp. 181–187.
- [25] C. Lin, J. Liu, and C. Liao, "Energy analysis of multimedia video decoding on mobile handheld devices," *Computer Standards & Interfaces*, vol. 32, no. 1, pp. 10–17, 2010.
- [26] M. Tamai, T. Sun, K. Yasumoto, N. Shibata, and M. Ito, "Energy-aware video streaming with QoS control for portable computing devices," in *Proceedings of the 14th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, Cork, Ireland, 2004, pp. 68–73.
- [27] M. A. Baker, V. Parameswaran, K. S. Chatha, and B. Li, "Power reduction via macroblock prioritization for power aware H.264 video applications," in *Proceedings of the 6th IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis*, New York, NY, 2008, pp. 261–266.
- [28] K. Amiri, A. Khajeh, A. Eltawil, and F. Kurdahi, "Process variation aware transcoding for low power H.264 decoding," in *Proc. 8th IEEE Workshop on Embedded Systems for Real-Time Multimedia (ESTIMedia 2010)*, Scottsdale, AZ, 2010, pp. 90–96.
- [29] D. Shin, Y. Kim, and M. Pedram, "Dynamic voltage scaling of OLED displays," presented at the 48th Annual Design Automation Conference, New York, NY, 2011.
- [30] S. R. Gulliver and G. Ghinea, "Defining user perception of distributed multimedia quality," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, vol. 2, pp. 241–257, 2006.
- [31] International Telecommunication Union (ITU-T), "P.10/G.100 Amendment 2: New definitions for inclusion in recommendation ITU-T P.10/G.100—Vocabulary for performance and quality of service," Geneva, Switzerland, 2008.
- [32] International Telecommunication Union (ITU-R), "BT.500-12: Methodology for the subjective assessment of the quality of television pictures," Geneva, Switzerland, 2009.
- [33] International Telecommunication Union (ITU-T), "P.910: Subjective video quality assessment methods for multimedia applications," Geneva, Switzerland, 2008.
- [34] J. L. Blin, "New quality evaluation method suited to multimedia context: Samviq," in *Proceedings of the Second International Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM*, Scottsdale, AZ, 2006, vol. 6.
- [35] T. C. M. de Koning, P. Veldhoven, H. Knoche, and R. E. Kooij, "Of MOS and men: Bridging the gap between objective and subjective quality measurements in mobile TV," in *Proceedings of Multimedia on Mobile Devices 2007, IS&T/SPIE Symposium on Electronic Imaging*, San Jose, CA, 2007, vol. 6507.
- [36] K. Seshadrinathan and A. C. Bovik, "Video quality assessment," in *The Essential Guide to Video Processing*, A. C. Bovik, Ed., 1st ed. New York: Elsevier, 2009.
- [37] A. K. Moorthy and A. C. Bovik, "Visual quality assessment algorithms: What does the future hold?," *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 675–696, 2010.
- [38] S. Winkler, "Perceptual video quality metrics—A review," in *Digital Video Image Quality and Perceptual Coding*, H. Wu and K. Rao, Eds., 1st ed. Boca Raton, FL: CRC Press, 2005, pp. 155–181.
- [39] S. Winkler and P. Mohandas, "The evolution of video quality measurement: From PSNR to hybrid metrics," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 660–668, Sep. 2008.
- [40] S. Kanumuri, P. Cosman, A. Reibman, and V. Vaishampayan, "Modeling packet-loss visibility in MPEG-2 video," *IEEE Trans. Multimedia*, vol. 8, no. 2, pp. 341–355, 2006.
- [41] S. S. Hemami and A. R. Reibman, "No-reference image and video quality estimation: Applications and human-motivated design," *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 469–481, 2010.
- [42] I. Gunawan and M. Ghanbari, "Efficient reduced-reference video quality meter," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 669–679, Sep. 2008.
- [43] Z. Wang and A. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.
- [44] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment and human visual system," in *Proceedings of SPIE, Video Communications and Image Processing Conference 2010*, Huangshan, China, 2010, vol. 7744.

- [45] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Processing*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.
- [46] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication, Special Issue on "Objective Video Quality Metrics"*, vol. 19, no. 2, pp. 121–132, 2004.
- [47] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312–322, Sep. 2004.
- [48] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 165–182, Jun. 2011.
- [49] A. K. Moorthy and A. C. Bovik, "Efficient motion weighted spatio-temporal video SSIM index," *Proceedings of SPIE*, vol. 7527, no. 1, pp. 752711–752711-9, 2010.
- [50] T. Zinner, O. Abboud, O. Hohlfeld, T. Hossfeld, and P. Tran-Gia, "Towards QoE management for scalable video streaming," in *Proceedings of the 21th ITC Specialist Seminar on Multimedia Applications—Traffic, Performance and QoE*, Miyazaki, Japan, 2010.
- [51] J. Klaue, B. Rathke, and A. Wolisz, "EvalVid—A framework for video transmission and quality evaluation," in *Proc. of the 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, Urbana, IL, 2003, pp. 255–272.
- [52] "Miro guide—Podcast directory, video and audio RSS feeds," [Online]. Available: <http://www.miroguide.com/>
- [53] "Apple—iTunes U—Learn anything, anywhere, anytime," [Online]. Available: <http://www.apple.com/education/itunes-u/>
- [54] "MediaInfo," [Online]. Available: <http://mediainfo.sourceforge.net/en>
- [55] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H. 264/AVC video coding standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.
- [56] "Consumer digital video library," [Online]. Available: <http://www.cdvl.org/>
- [57] "Xiph.org—Test media," [Online]. Available: <http://media.xiph.org/>
- [58] M. O'Connell and J. Smith, "A guide to working with m-learning standards," *Australian Flexible Learning Framework*, pp. 1–38, 2007.
- [59] A. Kapoor, "Dynamic streaming on demand with Flash Media Server 3.5 | Adobe developer connection," 2009. [Online]. Available: [http://www.adobe.com/devnet/flashmediaserver/articles/dyn-stream\\_on\\_demand.html](http://www.adobe.com/devnet/flashmediaserver/articles/dyn-stream_on_demand.html)
- [60] Apple Inc., "Apple—QuickTime—Tutorials—Creating amazing video content with H.264," [Online]. Available: <http://www.apple.com/mx/quicktime/tutorials/h264.html>
- [61] "XMedia recode," [Online]. Available: <http://www.xmedia-recode.de/>
- [62] "VideoLAN—x264, H.264/AVC encoder," [Online]. Available: <http://www.videolan.org/developers/x264.html>
- [63] "AudioCoding.com—FAAC," [Online]. Available: <http://www.audiocoding.com/faac.html>
- [64] L. Gaston, J. Boley, S. Selter, and J. Ratterman, "The influence of individual audio impairments on perceived video quality," presented at the 128th Audio Engineering Society Convention, London, U.K., 2010.
- [65] A. Joly, N. Montard, and M. Buttin, "Audio-visual quality and interactions between television audio and video," in *Proc. Sixth International Symposium on in Signal Processing and its Applications*, 2001, vol. 2, pp. 438–441.
- [66] J. G. Beerends and F. E. De Caluwe, "The influence of video quality on perceived audio quality and vice versa," *Journal of the Audio Engineering Society*, vol. 47, no. 5, pp. 355–362, 1999.
- [67] G. Stoll and F. Kozamernik, "EBU subjective listening tests on internet audio codecs," *EBU Technical Review 2000* [Online]. Available: [http://tech.ebu.ch/docs/techreview/trev\\_283-kozamernik.pdf](http://tech.ebu.ch/docs/techreview/trev_283-kozamernik.pdf)
- [68] "MSU video quality measurement tool (PSNR, MSE, VQM, SSIM)," [Online]. Available: [http://compression.ru/video/quality\\_measure/video\\_measurement\\_tool\\_en.html](http://compression.ru/video/quality_measure/video_measurement_tool_en.html)
- [69] "CoreCodec TCPMP (open source)—HPC factor," [Online]. Available: <http://www.hpcfactor.com/downloads/tcpmp/>
- [70] "Google Chrome," [Online]. Available: <https://www.google.com/chrome>
- [71] A.-N. Moldovan and C. H. Muntean, "Towards personalised and adaptive multimedia in m-learning systems," in *Proceedings of 16th World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (E-Learn 2011)*, Chesapeake, VA, 2011, pp. 782–791.
- [72] A.-N. Moldovan, A. Molnar, and C. H. Muntean, "EcoLearn: Battery power friendly e-learning environment for mobile device users," in *Learning-Oriented Technologies, Devices and Networks—Innovative Case Studies*, A. Lazakidou and I. El Emary, Eds. Saarbrücken, Germany: LAP LAMBERT Academic, 2011, pp. 273–296.