VIDEO CONTENT ADAPTATION USING TRANSCODING FOR ENABLING UMA OVER UMTS

Safak Dogan, Sertac Eminsoy, Abdul H. Sadka, and Ahmet M. Kondoz

CCSR, University of Surrey, Guildford GU2 7XH, Surrey, UK
e-mail: {S.Dogan, S.Eminsoy, A.Sadka, A.Kondoz@eim.surrey.ac.uk}

ABSTRACT

The growing heterogeneity of access networks, multimedia terminals, user profiles as well as applications and services has troubled the notion of accessing the multimedia content anywhere, anytime, and with any device. Several strategies have been adopted to alleviate this problem, and the universal multimedia access (UMA) concept has been developed for this purpose. UMA aims at enabling users to access the multimedia content freely by allowing adaptation of the media content to diverse network characteristics, end-terminal capabilities, and user preferences. In light of these facts, this paper presents a real-time error-resilient video transcoding mechanism, so as to constitute a UMA enabler. The performance of this transcoder is assessed over a simulated universal mobile telecommunication system (UMTS) network.

1. INTRODUCTION

Rapidly increasing volume of multimedia content has imposed new applications and services to be deployed for some time. These applications and services show very diverse characteristics which need accurate mapping to communication links or access terminals to enable content delivery with acceptable service quality. Video is one of the most prominent applications of multimedia communications, as it directly targets the perceptual experience of individuals thorough the human visual system (HVS). However, it is also a delicate application due to its inherent sensitivity to several factors including varying transmission medium and access network characteristics, end-terminal capabilities, and users' preferences. These factors have greatly been diversified in recent years to form a complex heterogeneous multimedia communications environment, as:

- worldwide success of the fixed-Internet, and its extension to wireless scenarios;
- growing amount of content produced by numerous content publishers or producers;
- diversity of the natural environment in which users are present whilst accessing media content (e.g. type of location, time of day);
- availability of increasing bandwidth for enhanced multimedia experience both over wired (e.g. broadband ISDNs, ADSLs) and wireless networks (e.g. WLANs, 2.5/3G, 4G);
- significant regulatory efforts of various international bodies and organisations (e.g. ISO, ITU, W3C, IETF) to provide several open multimedia communication standards (e.g. JPEG, MPEG, MHEG, H./G. families, SMIL) and protocols (e.g. TCP/RTP/IP, SIP);
- proliferation of a wide range of access devices (e.g. PCs, PDAs, laptops, other fixed/mobile terminals) which have already penetrated into the marketplace

have significantly contributed to heterogeneity of today's numerous multimedia communication scenarios. The overall resulting effect of such evolution has thus produced a serious problem for accessing the multimedia data with any device over any access network, and at anytime. Nowadays, this problem is perceived as an integral part of the UMA issue. A number of techniques have been proposed to mitigate several problems coupled with the increasing heterogeneity imposed by the UMA concept. A primary approach, which is widely adopted, is the adaptation of the media content to enable the widespread dissemination of the multimedia data as well as its personalisation to different user preferences.

Video content adaptation is a key strategy to enable UMA. There are a number of methods to achieve the adaptation of video to varying network characteristics, device capabilities and end-user preferences. In this paper, these different methods are discussed following a brief overview of the UMA scenario and the relevant part of the recent MPEG-21 standard, namely Part-7. One of the video adaptation methods, namely video transcoding, is further elaborated to allow for user access to robust real-time video services over a simulated error-prone UMTS network. This is an extension of our previous research where an error-resilient video transcoder was developed to provide enhanced video qualities over 2.5G GPRS and EGPRS access networks [1-2]. Computer simulation results and their discussions are given for the performance evaluation of the transcoder. Finally, concluding remarks are drawn whilst setting pointers to future research directions.

2. UMA SCENARIO AND MPEG-21 PART-7 (DIA)

UMA is designed to bring the notion of "user access to rich multimedia content anywhere, anytime and using any access terminal" into life. The access is specifically aimed at any type of content that is available for consumption. However, the vast heterogeneity in existing communication technologies has resulted in major mismatches among the produced content, consumers and ways the content is consumed. These mismatches have significantly hampered the true UMA experience. The UMA concept (the "big picture") is depicted in Fig.1.

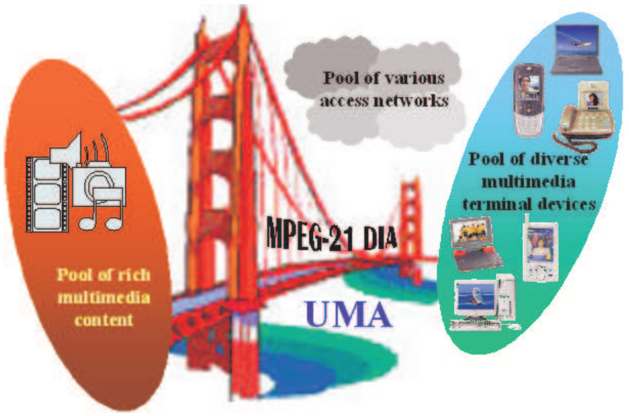

Fig.1: The “big picture”: UMA and its enabler MPEG-21 Part-7 (DIA)

UMA involves key strategies for widespread dissemination of various media resources. These strategies mainly comprise content adaptation systems, which adjust content and context-centric attributes to network, device and user-centric limitations. Consequently, this leads to building a *bridge* between resources and consumers that are connected via various access networks. Content adaptation consists of tailoring the media resources to specific requirements dictated by different:

- network characteristics (e.g. delay, bit error rate, packet loss rate, bandwidth);
- end-terminal capabilities (e.g. type of codec, memory storage, processing power, display size);
- user preferences (e.g. level of interactivity with content, personalisation);
- natural environment of users (e.g. location, time);
- content representations (e.g. encoded at various bit/frame rates, spatial resolutions, syntax formats).

Due to the existing heterogeneity in communication technologies discussed previously, there are a great number of multimedia representation methods, codecs, transport protocols, and presentation schemes. This is particularly complicated by the absence of a standardised framework for interoperability. From this point of view, UMA can be perceived as the “big picture” for interoperability, and MPEG-21 can then be realised as its enabling worldwide standard in this big picture with its inherent content adaptation capability. MPEG-21 standard is aimed to define a multimedia framework with a comprehensive set of tools and descriptors to enable transparent and augmented use of multimedia resources across a wide range of networks and devices. MPEG-21 contains a large suite of functionalities classified in several parts. The standard defines a fundamental multimedia content unit, i.e. Digital Item (DI), for distribution and transactions. Part 7 of the standard, which is specifically dedicated to the Digital Item Adaptation (DIA), is the main concern of UMA, as it addresses the necessary requirements, tools and descriptors for the adaptation of DIs [3-6].

3. STRATEGIES TO ENABLE UMA

There are several strategies presented in literature to enable UMA within a heterogeneous communications environment [7-8]. The three foremost of those strategies are namely the simulcast distribution model, scalable media model and transcoding model. Simulcast is the most straightforward method for enabling UMA, as it produces several independently encoded copies of the same video content at varying features, such as different bit/frame rates, and spatial resolutions. It then provides the delivery of these copies to serve clients with different connection speeds, allowing them to decide and select which stream to use. Despite being the simplest method, the simulcast strategy ends up clogging communication links up with numerous variations of the same content, which results in over utilising the available transmission bandwidth. In order to cover a wide range of bandwidth and user preferences, the content provider has to generate excessive number of versions of video streams. This procedure inevitably needs many streams to be stored on the server for one video sequence. Therefore, it does not provide an efficient solution to various personalisation demands of users (e.g. interaction with the content) due to the provision of a pre-determined number of copies of video without *a priori* knowledge of user preferences and terminal capabilities.

On the other hand, scalable media has long been perceived as an effective solution to the drawbacks mentioned above, as it provides a base layer for minimum requirements, and one or more enhancement layers to offer improved qualities at increasing bit/frame rates and resolutions. Thus, it gives the content provider the opportunity to generate only one basic stream and its interdependent enhancements to cover a wide range of client bandwidth, terminal requirements, and preferences. This method therefore significantly decreases the storage costs of the content provider. However, personalised video delivery is sometimes limited with this technique, as the interaction capabilities of users with the distributed video content are restricted to connection speeds, and clients’ terminal and network-related characteristics. This is due to the *all-or-none* enhancement nature of scalability, as interdependency between the layers of scalable video streams requires the full reception of enhancement layer or layers by the receiver for increased service functionalities and qualities. To overcome this weakness, fine granular scalability (FGS) scheme has been developed and adopted by the MPEG-4 video coding standard, which provides progressive scalability depending on whether bandwidth requirements are adequately met. FGS is primarily aimed at providing scalability for wireless video transmissions, as it offers enhanced qualities adaptive to varying network conditions [8]. Nevertheless, scalability has a major inherent disadvantage: the overall video quality degrades significantly with the increased level of scalability, particularly when the base layer is encoded at a low bit rate. To avoid this problem, the base layer is usually encoded at high rates and resolutions. However, this causes even more severe problems when the available channel bandwidth is not large enough to transmit even the base layer video stream. Furthermore, the scalable media strategy also needs layered encoding and decoding capabilities at the server and receiver-sides, respectively. In most cases, low-power mobile terminals are not equipped with such functionality, as it requires increased device complexity.

The third content adaptation strategy is the transcoding of video streams. It is a process to convert one video format into another with different features, and is used to tailor the video content to network characteristics, terminal capabilities, and user preferences. Several video transcoding schemes have been developed in the past few years, which provide a wide range of services, such as bit/frame rate conversions, resolution scaling, cross-syntax changes, error robustness insertion [9]. As opposed to the previous strategies, this method provides very flexible solutions, as transcoding is a middleware operation, and thus is performed on compressed video streams at the edges of different networks. Video gateways are deployed to perform this operation to customise, personalise, and summarise the content as required [10]. Particularly, personalisation of the video information is improved since the gateway is located much closer to users allowing for rapid responses [11]. Similarly, dynamic, and hence more efficient adjustments of the content in response to varying network conditions are also facilitated with the use of transcoders. Adaptation and personalisation via transcoding are possible both on-line (real-time) and off-line [12-13]. This strategy helps the content provider keep only one high-quality copy of a video stream, reducing the storage costs significantly. Likewise, the operational complexity associated with layered encoding and decoding is not a matter of concern with this technique. This makes transcoding a viable option for content adaptation in heterogeneous networking, particularly for a wide range of mobile terminals with limited processing and battery power.

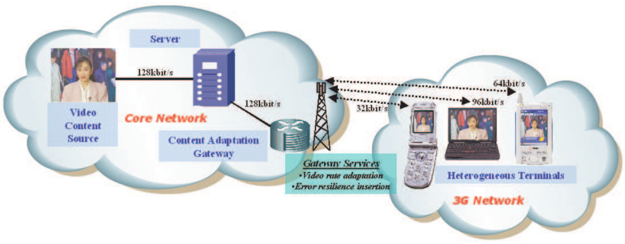

Fig.2: The overall transcoding scenario for UMTS access

4. ERROR-ROBUST VIDEO TRANSCODING PERFORMANCE OVER UMTS

In [1], an error-resilient video transcoding algorithm was developed for robust GPRS communications, and in [2] it was improved to provide enhanced error resilience services operating in real-time over GPRS/EDGE networks. In this work, this video transcoding strategy has been further experimented to accomplish the error-resilient adaptation of input video data to a number of bandwidth requirements of wideband code division multiple access (WCDMA)-based UMTS networks. Thus, in this section, the resulting observations from these tests are presented and discussed in detail.

As shown in the scenario illustrated in Fig.2, the video transcoding is performed in real-time at a content adaptation gateway at the edge between the core and 3G networks. The input to the transcoder is therefore a relatively high bit rate video stream, and the output is TM5 rate-controlled video transmission. The 3G network parameters, such as the spreading factor (SF), existence of power control and strength of the channel coding that is provided, are configured according to the user quality of service (QoS) profiles or terminal capabilities. The gateway is responsible for this configuration, and it performs necessary operations on non-resilient variable bit rate (VBR) streams to provide robust video transmission over the wireless access network. The error resilience tools that are used are namely data partitioning (DP), insertion of video packet resynchronisation (VPR) markers, and adaptive intra refresh (AIR) algorithms. The video transcoder produces MPEG-4 standard-compliant [14] error-resilient video outputs from non-resilient inputs at fixed transmission bit rates.

Fig.3 demonstrates the performance results of error-resilient video transcoding algorithm using an in-house produced 40-second long QCIF size video sequence with moderate motion. The input video to the transcoder was a variable rate sequence at 128kbit/s on average whilst the transcoder output was at fixed rate of 64kbit/s. The frame rate was set to 10fr/s throughout the experiments. The video packet size and number of AIR macroblocks (MBs) were selected as 700 bits and 10 MBs, respectively, for resilience experiments [2]. The error-prone 3G channel conditions were simulated for a scenario where 1/3-rate convolutional codes were used without power control for a vehicular speed of 50km/h. The operating point was chosen as Eb/No=8dB for an SF of 32 [15]. Each plot of Fig.3 was obtained averaging the resulting peak signal-to-noise ratios (PSNRs) of 15 simulations run with randomly varying seeds. The results show the quality improvement between the error-resilient and non-resilient sequences, which is around 7dB on average. Moreover, it can also be noticed that the addition of AIR to DP and VPR has produced better performance compared to the DP+VPR only resilient transcoding.

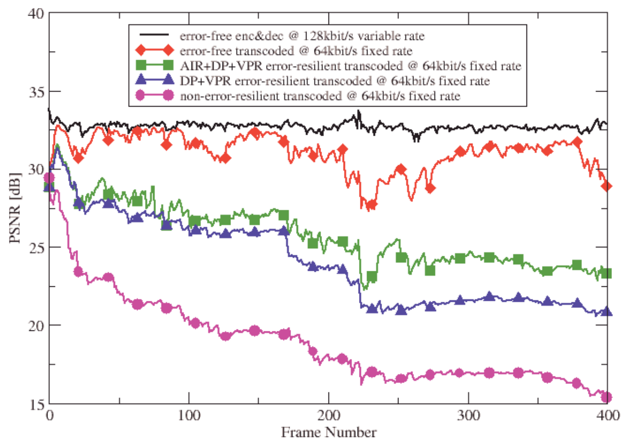

Fig.3: Error-resilient transcoding test results over a simulated 3G access network

Fig.4 shows a set of error-resilient video transcoding performances against a number of bit rates, and their associated SFs. Video sequences were transcoded from 128kbit/s to 32, 64, and 96kbit/s, as in the scenario depicted in Fig.2. The varying SFs, which correspond to the three transcoded rates, are also mentioned within the graph. The selection of these SFs was made in line with the 3G network requirements, as actual information data rate is a function of SF. The test conditions remained the same as before, and PSNR values were averaged over 15 simulations with random seeds. The results demonstrate that the added resilience improved the overall video quality as opposed to that of non-resilient video transmission. In error resilience experiments, the effects of the addition of combined AIR with DP+VPR are evaluated only, as the previous test results proved the superior performance of this combination compared to the DP+VPR only resilience method.

In addition, it can be observed that the quality performance increases as the transcoding rate increases (i.e. PSNRs at 96kbit/s are better than those at 64 and 32kbit/s, and similarly PSNRs of 64kbit/s are better than those of 32kbit/s). This is due to the more efficient allocation of increased number of bits per frame for transcoded video data with finer quantisation. In error-prone conditions, the more the video streams are compressed the more susceptible they become to channel error effects, as they comprise less redundancy. This is also demonstrated in Fig.4. Moreover, SF=16 provided better video qualities compared to SF=32 at Eb/No=8dB although it is a lower spreading factor than 32. This is due to the fact that better channel protection capabilities of higher spreading factors, such as SF=32, can only be realised at poorer channel conditions (i.e. Eb/No<7.5dB) [15].

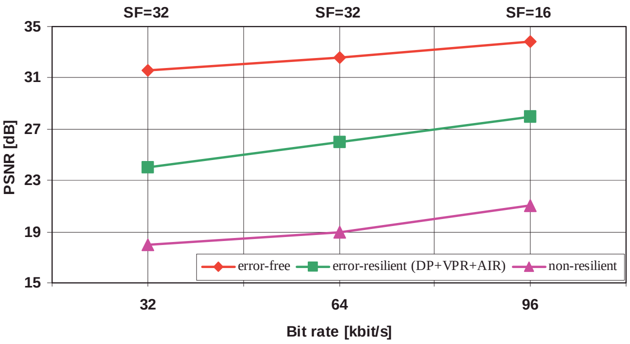

Fig.4: Video transcoding performances versus bit rates and associated SFs

The performances of transcoded video qualities were also assessed perceptually using subjective tests. The quality of non-resilient video clips was observed to significantly deteriorate with multiple full video frame losses resulting from error propagation. Further tests were also performed using typical ITU video test sequences, and similar results were obtained.

5. CONCLUSION

This paper has highlighted the growing heterogeneity in technologies and preferences to access a vast variety of multimedia content. The UMA concept has been overviewed, and the reasons why a true UMA experience has not yet been realised, have been identified. MPEG-21 Part-7 (DIA) has been introduced as the UMA enabler. Three key content adaptation strategies (i.e. simulcasting, scalability, and transcoding of media streams) have been discussed in detail. One of the strategies, namely transcoding, has been tested within a selected 3G access network scenario. For this purpose, a real-time video transcoder has been developed, which inserts robustness to input non-resilient video streams at the edge of the 3G network. MPEG-4 error resilience tools have been employed to achieve robust transcoding at a number of fixed 3G-compatible bit rates. The resilience tools are DP, insertion of VPR markers, and AIR algorithms. The transcoder thus produces MPEG-4 standard-compliant streams. The error-resilient transcoder has been observed to produce several error-robust video outputs at a number of bit rates. The test results indicated the effectiveness of the resilient video transcoding algorithm in terms of enhanced video qualities in the UMTS networking scenario. The future research will focus on developing an efficient algorithm for a region-of-interest (RoI) selection in the transform domain whilst transcoding in real-time. In this way, it is envisaged to provide more robustness to a particular RoI compared to the rest of video data during transcoding, so as to enhance the overall perceptual quality and user satisfaction with 3G video services.

6. REFERENCES

[1] S. Dogan, A. Cellatoglu, M. Uyguroglu, A.H. Sadka, and A.M. Kondoz, "Error-resilient video transcoding for robust internetwork communications using GPRS", *IEEE Trans. CSVT*, Vol. 12, No. 6, pp. 453-464, June 2002.

[2] S. Eminsoy, S. Dogan, and A.M. Kondoz, "Real-time services for video communications", *Proc. of IEE VIE'2003*, No. 495, Guildford, UK, pp. 262-265, 7-9 July 2003.

[3] MPEG-21: ISO/IEC JTC1/SC29/WG11/N5231, http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm

[4] J. Bormans, J. Gelissen, and A. Perkis, "MPEG-21: The 21st century multimedia framework", *IEEE Sig. Proc. Mag.*, Vol. 20, No. 2, pp. 53-62, Mar. 2003.

[5] I. Burnett, R. Van de Walle, K. Hill, J. Bormans, and F. Pereira, "MPEG-21: Goals and achievements", *IEEE Multimedia*, Vol. 10, No. 4, pp. 60-70, Oct.-Dec. 2003.

[6] H. Sun, A. Vetro, and K. Asai, "Resource adaptation based on MPEG-21 usage environment descriptions", *Proc. of IEEE ISCAS'2003*, Vol. 2, Bangkok, Thailand, pp. 536-539, 25-28 May 2003.

[7] J. Liu, B. Li, and Y.-Q. Zhang, "Adaptive video multicast over the Internet", *IEEE Multimedia*, Vol. 10, No. 1, pp. 22-33, Jan.-Mar. 2003.

[8] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard", *IEEE Trans. CSVT*, Vol. 11, No. 3, pp.301-317, Mar. 2001.

[9] A. Vetro, C. Christopoulos, and H. Sun, "Video transcoding architectures and techniques: An overview", *IEEE Sig. Proc. Mag.*, Vol. 20, No. 2, pp. 18-29, Mar. 2003.

[10]Z. Lei, and N.D. Georganas, "Video transcoding gateway for wireless video access", *Proc. of IEEE CCECE*, Vol. 3, Montreal, Canada, pp. 1775-1778, 4-7 May 2003.

[11]S. Dogan, S. Eminsoy, A.H. Sadka, and A.M. Kondoz, "Personalised multimedia services for real-time video over 3G mobile networks", *Proc. of IEE 3G'2002*, No. 489, London, UK, pp. 366-370, 8-10 May 2002.

[12]O Steiger, T. Ebrahimi, and D.M. Sanjuan, "MPEG-based personalized content delivery", *Proc. of IEEE ICIP'2003*, Vol. 3, Barcelona, Spain, pp. 45-48, 14-17 Sept. 2003.

[13]J.-G. Kim, Y. Wang, and S.-F. Chang, "Content-adaptive utility-based video adaptation", *Proc. of IEEE ICME'2003*, Vol. 3, Baltimore, MD, USA, pp. 281-284, 6-9 July 2003.

[14]MPEG-4: ISO/IEC JTC1/SC29/WG11/N4668, http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm

[15]C. Kodikara, S. Worrall, S.N. Fabri, and A.M. Kondoz, "Performance evaluation of MPEG-4 video telephony over UMTS", *Proc. of IEE 3G'2003*, No. 494, London, UK, pp. 73-77, 25-27 June 2003.