Multimedia Streaming On Mobile Phones

Stephan Brumme
Hasso-Plattner-Institute for Software Systems Engineering, Potsdam, Germany
currently at the University of Technology, Sydney, Australia

Abstract In this paper, I evaluate the opportunities and challenges of streaming multimedia content on mobile phones from the technical, commercial and social point of view. A summary of the used MPEG technology and its modifications for mobile phones, such as real-time requirements, power consumption and enhanced error detection are presented. Finally, I examine major problems of user interface on small devices and how to overcome them.

1 Introduction

The recent availability of public *UMTS* (Universal Mobile Telecommunications System) networks prepares the ground for the introduction of numerous new high data rate communication services on mobile phones. One of the most promising services is *streaming multimedia content* due to its widespread uses in the fields of visually enhanced telephone conferences, entertainment, product presentations [Varshney02], private talks, and many more. Each of these services gives the opportunity of a richer and more efficient information interchange among its users that leads to a higher willingness of the customers to actually pay an additional charge for the add-on service.

In fact, the telecommunication industry considers multimedia streaming as the driving force behind the launch of the third generation of mobile networks, often abbreviated by *3G*, and to introduce UMTS. Beside some changes of the underlying protocol layer, the UMTS technology promises a far higher data rate, sometimes 100 times higher than *GSM* (Groupe Spécial Mobile, now often Global System for Mobile Communications) [GSM] which operates at up to 14.4Kbps. A raw UMTS bit rate of up to 384Kbps for wide-area coverage and even up to 2Mbps for local-area coverage is comparable to nowadays internet broadband connections.

Major consulting corporations predict a rapidly growing demand for 3G networks. NTT DoCoMo already serves more than three million UMTS customers (March 2004).

This paper was written as part of the lecture *Advanced Data Communications* at the University of Technology, Sydney, in 2004. The author can be contacted via email: stephan.brumme@hpi.uni-potsdam.de, stephan.brumme@uts.edu.au, or info@stephan-brumme.com

Multimedia streaming on 3G mobile phones shares many aspects with multimedia streaming over the internet. The underlying protocol bases on *TCP* (transmission control protocol) [Alonso95], i.e. is packet switched. Transmission errors may occur; also they are more likely to happen on a wireless connection because of the higher number of randomly influencing factors.

The requirements for multimedia streaming slightly differ from pure downloading of data. Whereas the latter a 100% error-free transport without any time constraints expects, tolerates the former minor errors as long as some recovering techniques like retransmission, client-based interpolation or switching to a backup server apply. The user does not want to face noticeable delays, mostly experienced as frame drop-outs or a pausing audio signal, but accepts small artefacts if they do not significantly change the overall impression of the stream.

Although most technical problems in the field of mobile data transmission are solved, the acceptance among customers is still very low. Beside high costs, the insufficient ergonomic design of mobile phones is responsible for

Almost all recently sold mobile phones are able to connect to the mobile internet using various technologies such as *GPRS* (General Packet Radio Service) which is sometimes called 2.5G to indicate its position between pure voice based GSM (2G) and the more data driven UMTS (3G). The latest figures evidently show that only a fraction of the already available data services is actually used.

2 MPEG

2.1 Goals

The Motion Picture Experts Group (*MPEG*) was founded to define a standard for compressed movies. The need for compression becomes quite obvious

when looking at a simple example: a single frame of the PAL (Phase Alternating Line) broadcast television format consists of 576 lines, each subdivided into 720 samples which means the whole frame is composed of approx. 415K samples. Each point in turn is usually represented in the common RGB (Red-Green-Blue) format. To achieve a high quality, the RGB format requires 8 bits per colour channel, e.g. 3*8=24 bits per sample. In conclusion, a single frame equals 24bit*415K≈10Mbit.

Considering that there are 25 frames per second, an uncompressed movie needs a bandwidth of 250Mbps – that astonishing number does not even include the audio signal and exceeds even the highest achievable UMTS bandwidth by far. It is clear that a certain compression scheme has to be applied in order to reduce that data volume by magnitude.

Unfortunately, the best lossless compression algorithms available do not a very good job when shrinking multimedia content. Even in good cases, they allow for a maximum compression by a factor of 10.

Hence a lot of so-called "lossy" compression algorithm evolved over the past decade. The main idea is to remove redundancy and to concentrate on details the human senses are very perceptive to. For example, the eye can distinguish brightness better than colours. A distinct colour space, derived from RGB, is better suited for lossy algorithms. So the precision used to store the colour value can be quite low. The resulting quantization error is hardly noticeable but leads to a better compression. Another concept is to re-use past frames or signals and analyse them to predict the near future.

A compression factor of up to 1:1000 can be achieved and thus broadcasting via MPEG is able to fit the bandwidth constraints of UMTS at a pretty high quality.

2.2 Video Encoding

MPEG is bases on two approaches to compress the video stream: spatial redundancies (similarity to neighbouring regions) and temporal redundancies (similarities to the future and the past). The compression algorithm works per frame and subdivides each frame into small blocks, often 8x8 pixels. Smaller blocks may increase to compression efficiency but enlarge the administrative overhead and usually lead to higher encoding efforts.

The reduction of spatial redundancies is done by the well-known JPEG (Joint Picture Experts Group) algorithm. A discrete cosine transform (DCT) maps the brightness, colour and intensity to frequencies. This transformation is reversible and nearly lossless. Now the highest frequencies can be removed without significantly changing the result; the frames become just slightly blurry. As most pictures contain smooth transitions and large uniformly coloured areas, high frequencies are rarely found in them and they are perfectly suited for that kind of algorithm. On the other hand, sharp edges and sudden changes in brightness – as observable for example in scanned book texts – cannot fully gain from the JPEG technique because they vitally need high frequencies. The often chosen coding efficiency/quality trade-off is to adaptively change the bit rate according to the desired quality.

Temporal redundancies can be found when a sequence of frames shows the same or a similar content. Then only the changes between frames need to be stored which is far more efficient than a pure spatial compression. Motion is encoded by motion vectors that describe which block's content is most similar to the current one. Unfortunately, the compression using temporal redundancies suffer from accumulation of prediction errors. They become visible after several frames and may significantly influent the visual experience.

To resolve that issue, MPEG defines three kinds of frames: intra frames (I-frames) are compressed only using spatial redundancies and they do not refer to any previous frame. Predictive frames (P-frames) allow temporal based compression as well but only refer to past frames where bidirectional frames (B-frames) may refer to future frames. A sequence of frames looks like this: IBBPBBP. The MPEG standard does not define a lower or an upper limit of I-, P-, or B-frames.

However, I-frames play the role of a recovering point (in addition to their importance for the overall visual quality) and are often placed at least once per second. Losing such an I-frame means that the following P-frames become worthless.

A final post-processing step applies a common lossless Huffman algorithm to the gained data stream to further shrink it.

2.3 Audio Encoding

Three different algorithms, differing in their quality and efficiency, are defined by the MPEG standard. They remove signals from the audio stream that cannot be heard by the human ear based on a psychoacoustic model. For example, a loud signal may suppress a quiet one.

These algorithms are called layers, whereas the layer 3 seems to be the best known abbreviation on

the internet: MP3. Basically it stands for MPEG 1, layer 3 and became the de-facto standard in the (mainly illegal) music community.

Again, the signal is transformed to its frequencies and unimportant ones are removed. Depending on the used psychoacoustic model, the result may vary among audio encoders. The open source LAME encoder has a very good reputation.

Some new audio compression schemes appear at the horizon and promise better spatial reproduction of the original surround signal.

2.4 Modifications For Mobile Phones

The MPEG-2 standard provides a very good quality and is the coding scheme used for Digital Versatile Discs (DVD). Unfortunately, the bit rate of approx. 4 to 9Mbps is too high for mobile phones.

The successor MPEG-4 (there is no MPEG-3) aims at very low bit rates at an acceptable quality. One of the major keys in achieving that goal is an optimized object recognition algorithm aimed at natural scenes. Synthetic scenes like comics often cannot be encoded in a satisfying manner. MPEG-4 is able to carry meta-information, too. That can be the names of actors, timetables etc. or even interactive elements like VRML environments [Puri98].

Error detection and recovering from errors are essential parts of the MPEG-4 specification. The algorithm tries to prioritize objects of the stream. For example, slight losses of image quality are usually more tolerated than losses of audio packets.

The stream is separated into packets that can be independently decoded. So it is always possible to decompress a single packet without knowledge about the past packets.

Mobile phones suffer from limited battery energy supply. A proposed feedback protocol [Choi03] may help to find the optimal multimedia stream quality without wasting energy. The server encodes the stream on demand in a desired quality that perfectly fits the mobile phone's needs. Although the stream providers may need more computing capacity for their servers on first glance, there is a limited number of different mobile phones on the market so streams can be pre-encoded and cached.

3 Modifying The Network Transport Protocol

Usually more than one client subscribes to a live stream. Therefore it seems to be natural to avoid sending separate but identical streams to each client as it happens with point-to-point unicast connections. A more sophisticating solution tries to push just one single stream into the network and ask the network to deliver identical copies to the clients. The result is a significantly reduced traffic saving money both for the server and the network.

Multiple protocols have been developed to ensure such behaviour. The best known is IP multicast, a recommended IETF standard. IP multicast is independent of the actual network technology and works on LANs, ATM and wireless networks.

The MBone is a large-scale implementation of the IP Multicast technology. Big institutions like NASA use IP Multicast as well for different purposes (such as anti-virus updates distribution) [Muller98].

4 User Interface

Current mobile phones have inherited much of their design from their stationary counterparts. The basic shape and layout of a mobile phone must ensure that it is big enough to be used for voice communication while being small enough to be carried in pockets. It has to fit to an adult's hands and almost all keys should be accessible by the thumb [Ljungstrand01].

Some vendors experiment with non-standard designs. The customers will have to change their used typing behaviours and thus suffer from an increased rate of mistypings. The same effect can be observed on country-specific computer keyboards.

Beside the antenna, lots of chips, the microphone, the speaker and, most important, the batteries are responsible for far more than 50% of the mobile phone's dimensions. However, there is a clear tendency to shrink their size while maintaining their functionality.

However, the audio and video communication may be enriched by vibrations of the mobile phone. That has been proven to be a highly perceivable way of gaining attention and is available on almost every device. It often serves to notify the user of arriving text messages when switched to silent mode. In general, vibrations are usually closely related to audio signals such as explosions and may be used to underline extraordinary events. When used too frequently, the user tends to ignore vibrations or find them annoying.

Empirical studies on user perception revealed that the frame rate is the most influential factor [Song02]. According to that paper, the suggestion of a fluent stream plays a more important role than the image quality or image size. Consequently, the TFT displays of mobile phones should be able to refresh the screen quickly (about 25 times per second).

A trade-off between screen size and visual details is an interactive zooming features described by Microsoft Research Asia [Fan03].

Some researchers replace certain contents – like football – by 3D scenes where only the geometric descriptions and player movements have to be transmitted as metadata which required just a few Kbps [Wikstrand02].

5 Conclusion

The technological foundations for multimedia streaming on mobile devices exist and are profoundly examined by researchers. As multimedia streaming becomes available as a mainstream technology, one can assume a broad usage over the next years if user interface issues are resolved. Moreover, if the costs are too high, many users may turn down that technology.

It can be expected that the major UMTS network providers strongly push multimedia streaming to establish a significant share in the emerging UMTS market. As there is no other killer application available for UMTS, the main focus of the telecommunication industry will be on multimedia streaming. Right now it remains rather unclear, whether the revenues of the networks providers will be higher than the costs.

I expect multimedia streaming more as an add-on feature that does not fully replace the ordinary voice. My prediction is that the combination of mobile phones with home entertainment equipment can actually lead to a new experience. For example, streaming a video and forwarding it via Bluetooth to a television set and a Dolby-Surround system is an intelligent alternative to renting a DVD.

6 References

[Alonso95] Rafael Alonso, Yuh-Lin Chang, Liviu Iftode and V.S. Mani, *Managing Video Data In A Mobile Environment*, Matsushita Information Technology Laboratory, Princeton, 1995

[Choi03] Kihwan Choi, Kwanho Kim and Massoud Pedram, *Energy-Aware MPEG-4 FGS Streaming*, DAC 2003, Anaheim / USA, 2003

[Fan03] Xin Fan, Xing Xie, He-Qin Zhou and Wei-Ying Ma, *Looking Into Video Frames On Small Displays*, MM '03, Berkeley / USA, 2003

[GSM] http://www.gsmworld.com

[Ljungstrand01] Peter Ljungstrand, *Context Awareness And Mobile Phones*, Personal And Ubiquitous Computing, Volume 5, pp. 58-61, 2001

[Muller98] Nathan J. Muller, *Improving and Managing Multimedia Performance Over TCP/IP Networks*, International Journal Of Network Management, Volume 8, pp. 356-367, 1998

[Puri98] Atul Puri and Alexandros Eleftheriadis, *MPEG-4: An Object-Based Multimedia Coding Standard Supporting Mobile Applications*, Mobile Networks And Applications, Volume 3, pp. 5-32, 1998

[Song02] Seungho Song, Youjip Won and Injae Song, *Empirical Study On User Perception Behaviour For Mobile Streaming*, Multimedia '02, Juan-les-Pins / France, 2002

[Varshney02] Upkar Varshney and Ron Vetter, *Mobile Commerce: Framework, Applications and Networking Support*, Mobile Networks And Applications, Volume 7, pp. 185-198, 2002

[Wikstrand02] Greger Wikstrand and Staffan Eriksson, *Football Animations For Mobile Phones*, NordiCHI '02, Arhus / Denmark, 2002