Note to reader: As this paper was simple the summary of activities in a working group of the ISO, it was difficult to carry out an analysis in the traditional sense as there were no experimental processes or results. As a result this summary is merely an extraction of key components of MPEG.

Paper title: MPEG: A Video compression standard for multimedia applications – Didier Le Gall

Background and context

Note: The International Telegraph and Telephone Consultative Committee (CCITT) became the Telecommunication Standardization Sector (ITU-T) in 1993

This paper was published in 1991 in a high impact (ISI impact score 2.646) ACM monthly magazine publication which had been in circulation since 1958.

At the time of publication, the only standard for video compression in existence was CMTT/2 being used for compression of television signals at 34 Mbps and 45 Mbps. There was also a failed standard called H.120 which was published by the CCITT in 1984 for 1544 kbps rate NTSC and 2048 kbps rate PAL video. This used a bit by bit codec which had good spacial performance (content with reduced picture size) but poor temporal performance (content with reduced frame rate).

Another expert group in existence at the time was the CCITT expert group on visual telephony (study group XV) which looked at video compression in the context of teleconferencing and video-telephony. This group was specifically introduced because of the availability of Integrated Service Digital Network (ISDN) lines from 1986 which were capable of rates of px64kbps depending on the number of 64kbps channels being used. The result was recommendation H.261: “Video Codec for Audiovisual Services at px64 kbps” published in 1990 which was capable of less than 150ms delay, low overhead and and low bit-rate operation at 64 kbps. It was also the first video compression standard to use superior block based codecs and gained wide spread support.

The activities of the Joint Photographic Experts Group (JPEG) also overlapped somewhat with MPEG as they were both dealing with compression of images, the only difference being that a video was a sequence of still images. However compressed video could take advantage of the extensive frame-to-frame redundancy present and MPEG needed to focus on how to exploit this.

What was to follow was a series of standards between 1991 and 1998 (MPEG-1, MPEG-2, H.262, MPEG-3, H.263, MPEG-4, H.264) coming out of the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) which often collaborated when their objectives overlapped.

Overview

The goal of the paper was to communicate the video compression key requirements and algorithm characteristics of the Moving Picture Experts group (MPEG) – working group 11 of the ISO as they were in 1991.

At the time of publication video compression was needed for an ever increasing range of digital communication and storage media. The standards that were available at the time were not sufficient to address these needs. The MPEG standard being developed was targeting applications that ranged from digital storage media such as CD-ROM's, DAT, writeable optical drives to ISDN lines and local area networks. The activities of MPEG covered compression of video signals (MPEG-Video), compression of audio signals associated with the video (MPEG-audio) as well as synchronization between video and audio (MPEG-System), but the paper only focused on the activities of MPEG-Video. The aim was to compress video and its associated audio to a bit rate of about 1.5 Mbps with an acceptable level of quality.

It was amusing to read the comment "There is much talk of convergence, which does not mean that a computer workstation and a television receiver are about to become the same thing" and realise that this is precisely what has happened. We have even gone beyond this to the point where smart phones are the convergence of a computer workstation, a television, a global positioning system and a mobile phone.

Detailed discussion

The paper is essentially divided into an MPEG standardization process section, MPEG-video requirements section and an MPEG compression algorithm section in order to show how the algorithm meets the requirements envisaged.

The standardization process used a competition in which companies or institutions submitted a document describing their algorithm, a video for testing the algorithm and a collection of computer files to verify the compression claim by an impartial evaluator. This was followed by a convergence process in which the best ideas were integrated into one solution. The method to converge the ideas was in the form of a simulation model and a set of core experiments to resolve which alternative gave the best quality.

The MPEG standard requirements needed to be generic and meet the constraints of a wide range of media and communication channels. This included applications that could be asymmetric such as movies or symmetric such as video conferencing. There were some key features of the video compression algorithm to support these applications

- Random access: needed to be able to access a stream from any frame in the video. This requirement was also spurred on by the availability of random access media such as CD-ROM's.
- Fast forward and reverse: Selected pictures had to be displayed to obtain the fast forward or reverse effect.
- Reverse playback: Interactive applications may have require this
- Audio-visual synchronization: This needed to be maintained even if the signals were derived from different clocks
- Robustness to errors: Couldn't assume that the communication channel would be error-free and cause complete collapse of the video playback.
- Coding/decoding delay: For video conferencing this needed to be kept to under 150ms
- Editability: Needed to keep the editable units as small as possible to allow an acceptable level of time-granularity for editing.
- Format flexibility: Needed support for a wide range of raster sizes and frame rates
- Cost tradeoff: The algorithm had to perform in real time on hardware available at the time

The MPEG compression algorithm had the challenge of meeting orthogonal requirements of quality and random access. Random access was best satisfied with intraframe coding and quality requirements, which required a high degree of compression, was best satisfied with interframe coding. To deal with this balance, MPEG used two interframe techniques: predictive and interpolative.

Two core techniques were used by MPEG compression: temporal redundancy reduction using block-based motion compensation and spacial redundancy reduction using transform domain –

Discrete Cosine Transform (DCT) based compression.

For block-based motion compensation, three types of pictures were used, Intrapictures (I), Predicted pictures (P) and Interpolated pictures which used bidirectional predicition (B). The Intrapictures provided random access with moderate compression and bidirectional pictures, which required a past and future picture, had the highest degree of compression relying on motion-compensation compression. The level of random access and compression required dictated the ratio of interpolated pictures to intra or predicted pictures.

Motion compensation prediction exploited the fact that in video, very often the current picture could be modelled using a transition of a picture at a previous time. This transition was modelled as a set of amplitudes and directions and was not uniform across the picture. Interpolation went beyond simple prediction and is was key component of MPEG. It essentially used a sub-signal with low temporal resolution and reconstructed a full resolution signal by adding a correction term to a combination of a past and future references. This had a number of advantages including high compression, reduction of noise and no error propagation. A motion-compensation unit, called a Macroblock, had a resolution of 16x16 and motion vectors operated on these quantized blocks. The motion information consisted of a vector for forward-predicted macroblocks and two vectors for bidirectional macroblocks. This was coded with variable-length code to provide better efficiency. There was some freedom left to implementers on how motion vectors were calculated to allow for creativity.

The spacial redundancy reduction used by MPEG shared a lot in common with that used by JPEG and H.261. It essentially used a Discrete Cosine Transform (DCT) on macroblocks which had many proven advantages such as being orthogonal, fast and producing very little artefacts. Coarser quantization levels could be used on higher frequency information stored in non-intrapicture blocks whereas finer quantization was required for intrapicture blocks containing energy in all frequencies. This was due to the subjective perception of quantization error as “blocking effects” which increased when a course grained quantization was used on pictures with high bandwidth.

An MPEG bit stream always started with a video sequence header which defined key parameters of the video stream such as picture width, aspect ratio, bit rate and buffer size. This told the application decoding the stream what buffer size was required, for example, to correctly decode the stream. MPEG also made use of a layered architecture to support its generic nature. These layers separated the group of pictures layer from the macroblock layer, for example, to allow the random access unit and the motion compensation unit to have quick access to only the data they need. The bit rate, resolution and quality was focused on those commonly seen using Source Input Format (SIF) resolution: 352x240, 352x288, or 320x240 using a bitrate less than 1.5 Mbit/s. This made up the minimum video specifications any decoder would need to be able to handle to be considered MPEG compliant. The MPEG committee at the time was also going to begin exploring compression of video for higher bit rates up to 10Mbps.

Paper analysis

Overall this was a fair summary of the work of the MPEG committee at the time. Being a highly cited work (1544 citations on Google scholar), it obviously sufficiently highlighted the core features of what would become MPEG-1.

To turn the paper into a discussion beyond just a summary of a draft specification, the author could have presented some results which show the trade-offs between random accessibility and quality (in terms of SNR) and analysis of temporal, spacial and quality scalability with different bit rates.