Video Coding and a Mobile Augmented Reality Approach

A. Vlachos, V. Fotopoulos, A. N. Skodras
Digital Systems & Media Computing Laboratory,
School of Science and Technology, Hellenic Open University,
13-15 Tsamadou st., GR-26222, Patras, Greece
{a.vlachos, vfotop1, skodras}@eap.gr

Abstract

The main purpose of this work is to give a brief overview of video coding, from the main principles to the available standards and future trends, and define augmented reality applications that make use of video. Augmented reality (AR) is a novel scientific field that may benefit by video transmission, however special constraints do apply. An attempt will be made to define these constrains, the application field and give an answer to the following question: is one of the existing video standards suitable for mobile AR systems, or a new one should be developed?

Keywords: mobile augmented reality, video coding, codecs, application, position tracking, orientation.

1. Introduction

Augmented reality (AR) is a novel research field, which deals with the combination of real world and computer generated/ provided data [1]. For mobile AR systems, specially designed equipment is required, such as translucent goggles, head-mounted displays etc [2-4], usually in conjusction with some computational equipment, a laptop or other affordable consumer electronics such as mobile phone or PDA device. In most applications, computer graphics or tags are imposed over the real world view, with the assistance of the above mentioned device. That is a major difference from virtual reality (VR) systems which completely immerses the user into a 100% virtual, computer-generated environment. In AR systems, part of the real world view is always directly viewable.

Video compression is one of the most important and by far the most demanding process of processing digital data. The first standard for video compression, H.261, was introduced in 1989, almost two decades ago. Although media storage capacity has grown up very rapidly as well as network bandwidth, newer video standards keep

coming out. From the earlier H.261 and MPEG-1 to the newest H.264/MPEG-4 Part 10 (AVC), which is considered to be the most modern and technologically advanced standard, the need for better compression, with the best possible visual quality, is constant.

This work is focused towards the mixing of video with the real environment for mobile augmented reality applications. To analyze this task, first a brief introduction of video coding standards and algorithms is given in section 2, along with an overview of corresponding standards, challenges and future trends in the field. Some mobile augmented reality (MAR) application scenarios are given in section 3, while section 4 defines the restrictions that MAR imposes to video coding applications. Finally in section 5, discussion follows and conclusions are drawn.

2. Video Coding Algorithms and Standards

A video coding algorithm generally, consists of 4 different stages (Figure 1). In the first stage, redundant information that concerns human eye perception is removed. By prefiltering, quantization and chroma subsampling, part of the chrominance information is removed since the human eye is more sensitive to luminance than chrominance. In the next stage spatial redundancy is removed, by using a common transform, such as the DCT (Discrete Cosine Transform) or some other integer or variable block size transform. In H.264/MPEG-4 Part 10 (AVC) standard, together with the transform used (Integer transform based on DCT) a new method of reducing spatial redundancy is used, called *intraprediction*.

In the next stage the removal of temporal redundancy is performed. This is the most time and computations expensive stage, in the coding procedure. It is based on the remark that the time distance between two successive frames, is a fraction of a second (from 1/15 to 1/30 of a second). Consequently there is very little difference between the two of them. Through block motion estimation with some reference frame (or frames) the difference of the frames is computed and block motion vectors and residuals are assembled for transmission instead of the whole frame. Finally the previous output is fed to the last stage in order to remove statistical redundancy by means of entropy coding. Two of the most well known techniques are Huffman coding and arithmetic coding. In H.264/MPEG-4 Part 10 (AVC) standard a new

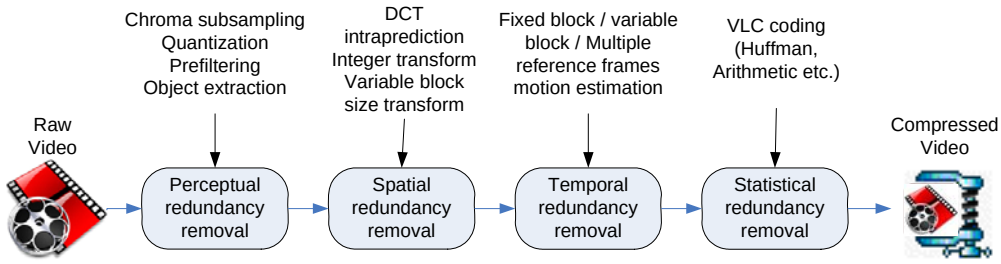

Figure 1. *Stages of video encoding algorithm*

approach is made with the use of context-adaptive coding (CABAC and CAVLC).

Video coding algorithms are hybrid because they combine spatial and time compression. They are also asymmetrical. The encoder is more complex than the decoder, meaning that more computations and consequently more time are needed from the encoder's side. Many video coding algorithms have been standardized by ITU (H.26x family) or the ISO Moving Picture Experts Group, also known as MPEG (from which the corresponding family inherits its name). The bit stream that ends up to the decoder should be formatted according to the standard, so there is some degree of freedom on the creation of the encoder.

A brief overview of the video coding standards follows [5-7]:

H.261

- The first video coding standard which affected very much all successive standards [8].
- Introduced in 1990 by ITU organization.
- Its main application was videoconference through ISDN lines, data rate varying from 64kbit/s to 2Mbit/s.
- Technical characteristics: YCbCr color representation, 16x16 macroblock, block-based motion compensation, 8x8 DCT transform, scalar quantization, entropy coding VLC (variable length code), integer pixel resolution motion vector, frame types supported I frames (intra-coding) and P frames (predictive coding) and deblocking filter.

MPEG-1 Part 2

- Produced by Moving Picture Experts Group (MPEG).
- Its applications were video CD (VCD) and online video in some cases with supported bitrate 1.5Mbit/s.
- Technical characteristics different from H.261, a new type of frame B frame (bidirectional coding) and ½ pixel resolution motion vector.

MPEG-2 Part 2

- Its main applications were cable TV, High Definition Television (HDTV), also used as the coding format for DVD's.
- At data rates higher than 3Mbit/s it outperforms all previous standards.
- From technical point of view it supports interlaced video, while all previous standards supported only progressive video [9].

H.263

- Its targeted application was videoconference at low bitrates of transmission.
- It improved by far the quality of interlaced video even in low bitrates [10].

MPEG-4 Part 2

- Introduced by MPEG group in 1998.
- Main applications: video transmission through internet and videophone.
- It has improved quality compared to MPEG-2 and H.263.

- Its technical characteristics different from previous standards were ¼ pixel resolution motion vector and object-oriented compression characteristics [11].

MPEG-4 Part 10/H.264 AVC (Advanced Video Coding)

- Developed by ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG), which worked together under the name of Joint Video Team (JVT).
- It is considered the state-of-the-art in video coding standards and it comprises features that have been used or suggested in previous standards.
- It is different from previous standards in the following technical characteristics: use of deblocking filter, context-adaptive techniques like CAVLC (Context-Adaptive Variable Length Code) and CABAC (Context-Adaptive Binary Arithmetic Coding), new and advanced quantization techniques, 4x4 and 8x8 integer transform based on DCT, use of block size from 4x4 to 16x16 and spatial intra prediction. It can function in various bitrates from 64kbit/s to 960Mbit/s, and various screen resolutions from 128x96 to 4096x2304. It is used commercially even by other standards like HD DVD and Blu-ray Disc [12-15].

Windows Media Video 9 (WMV9)/ VC-1

- Accepted as a standard in 2006 by the Society of Motion Picture and Television Engineers (SMPTE).
- Its applications vary from low resolution video in low bitrate connections to High Definition Television (HDTV).
- Its technical characteristics are similar to that of MPEG-4 Part 2, with the addition of deblocking filter, use of 4x4 to 8x8 integer transform based on DCT, overlap smoothing and uniform and non-uniform quantization [16, 17].

In Figures 2 and 3 the block diagrams of H.261 and H.264/MPEG-4 Part 10 (AVC) encoders are given. It is easily deduced that from the first standard to the most recent one, the process remains the same. What really changes are the techniques used in parts of the coding chain. For example, instead of plain DCT transform, Integer (based on DCT) transform and intraprediction are used in the most recent version. In Table 1, video compression standards along with some of their characteristics are listed.

Except from video coding standards, some other commercial or experimental codecs exist, that demonstrate excellent (or at least promising) results. Real video 10 and WMV9 codecs are such examples, achieving similar visual quality to H.264 with 15% and 30% lower bitrates respectively. Real video uses a two-pass encoding scheme which first analyze the video before compressing the content [18]. Another codec that uses two-pass compression is TrueMotion VP7 Video Codec, being 2 to 3 times less complex than H.264/MPEG-4 Part 10 (AVC). Using the two-pass analysis, it generates statistics in the first pass and in the second it uses these to choose better key frames and to allocate more bits to tougher sections [19].

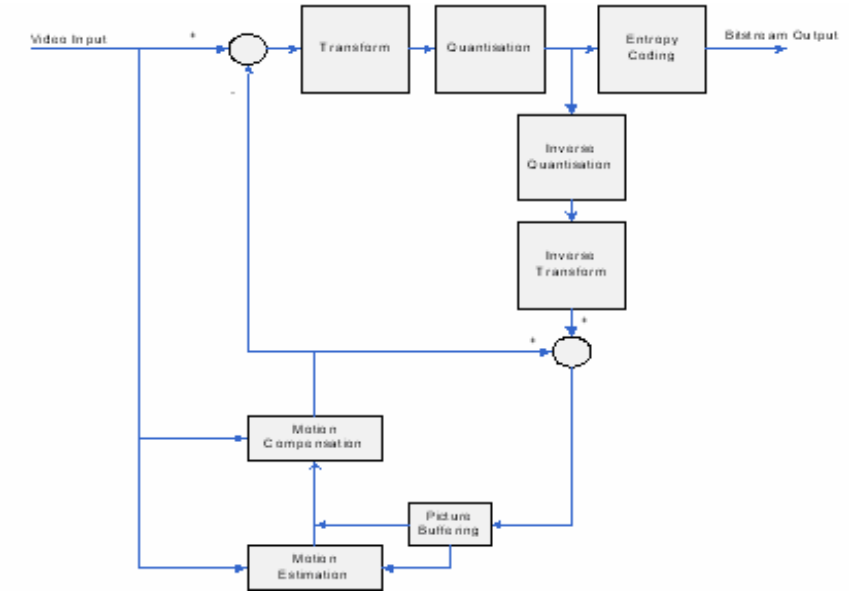

Figure 2. *Block diagram of the H.261 video encoder*

Table 1. Video Compression Standards Characteristics

Compression Standard	**Main Application**	**Video Format**	**Bitrate transmission**
H.261	Videoconference (ISDN)	CIF/QCIF	384-64 kbps
H.263	Videoconference (Internet)	4CIF/CIF/QCIF	64 kbps
H.263	Videoconference (wireless)	QCIF	18 kbps
MPEG-1	Video Distribution (CD/www)	CIF	1,5 Mbps
MPEG-2	Video Distribution (DVD/ DTV)	CIF/QCIF	3 – 10 Mbps
MPEG-4	Distribution (interactive) multimedia	CIF/QCIF	28 – 1024 kbps
MPEG-4 AVC/H.264	Multimedia applications	CIF/QCIF	All the range
WMV9/VC1	Video Distribution (DVD/www)	CIF/QCIF	96 – 384 kbps

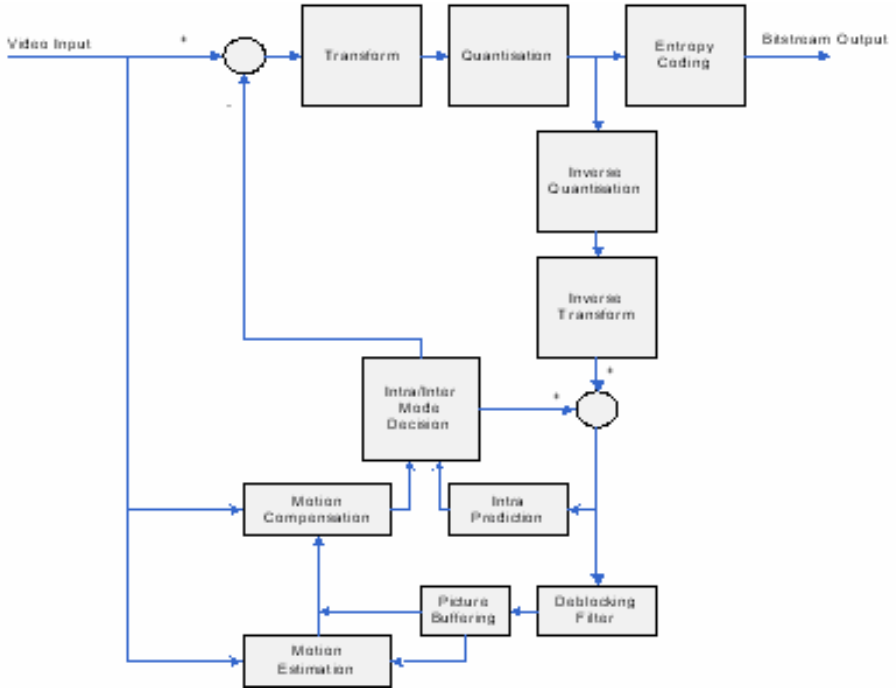

Figure 3. *Block diagram of the H.264 video encoder*

Dirac Codec [20] is a revolutionary approach to video coding; it uses wavelet transform applied to the whole frame instead of block based transform that all the other codecs use. Wavelets have been used with great success in JPEG2000, the still image compression standard, but not in video. Another breakthrough of Dirac is that it is not designed to maximize PSNR (Peak Signal to Noise Ratio), but it tries to estimate the subjective quality of the video by weighting large errors more and de-emphasizing high frequency errors.

3. MAR Application Scenarios

There is a general notion that AR is only reality augmentation by computer graphics, something that is only partially true. Reality can be augmented also by text or even video, imposed over the real world's image. Usage of digital video is also possible, as can easily be understood by the following application scenarios.

The user of a mobile device (a mobile phone or a PDA) approaches a building e.g. a museum, mall or a public service, and wants to learn more about what's happening inside. To be able to see more about the interior of the specific building, the user needs to connect to some digital content providing service. The server that

will provide the user with the feedback may serve him/her with prerecorded video, concerning the building activity, or even live video streamed from locally installed video equipment. For example, if the building is part of university campus, prerecorded video of lectures or lab experiments could be fed to the user. If on the other hand the building is a mall, the user could request a video concerning specific store windows to check if there's something interesting for him, or a shopping traffic view in order to decide if it is the right time to go shopping or not.

A key role to the success of the service belongs to localization. The service providing system should be able to identify the user's position in order to send him/her the right video stream. There are three different possible ways to achieve such a goal. First one is using a GPS device (Global Positioning System). This could be accurate, within a range of 3 meters (best case). GPS is nowadays an affordable capability of many handheld devices; however the system would also need orientation feedback. If however, the database contains data for selected buildings that are not in close distance, then this should not be a real problem. One should note of course that a GPS based system should have direct sky-contact while this type of signal can be jammed (as any other type of electronic communication signal).

Another way to achieve localization could be through the use of CBIR (Content Based Image Retrieval). In this method, the user could take a picture shot of the building (most modern phones and PDAs include camera features) and then transmit it to the information providing system. In that system, an image database would exist and for each of the images, a predefined video file, or a video-stream address is associated with. By CBIR methods, the system will try to identify the building and provide the user with the existing video or the address of the video-stream in order to give him/her the ability of watching live video from the building's interior. By using such a method, there is no need for special GPS equipment, thus lowering even more the application's equipment cost. However there's great concern about the successful identification (CBIR methods, are not 100% reliable).

The third method is a combination of the previous two. Combining GPS with CBIR, localization becomes very reliable since the deficiency of any method alone, is anticipated by the use of the other one. If the GPS fails even by 15 meters let's say, this is enough to give preprocessing data to the CBIR application. This means that the system will filter the available images in order to keep only those belonging to a small area around the position given by the GPS. This means that the possibilities for reliable recognition are increased significantly. Also, the orientation problem is meaningless in that case since the building's picture completely identifies what the user is facing. Another problem that is eliminated with such a tactic is that of the direct sky view. The last GPS position will be used by the system for filtering the image database and increase the method's reliability.

All this information is sent to the mobile device in a wireless way, through wi-fi or GPRS. This information is projected on part of the screen of the mobile device; specifically it is superimposed to the real world image of the building the user is

facing. So what the user finally sees on the screen of his device is the real world image combined with extra, useful information about the building he is looking at.

4. Restrictions imposed to video coding by MAR

Video coding is one of the most (or the most) demanding tasks for a digital system. This is probably the reason behind the great interest of the research community along with the huge commercial interest around the subject. Another reason is the growing capabilities of mobile devices in terms of computing power, storage capacity, and of course the constant evolution of networks. However, some restrictions apply over the video exchange for AR applications and these are the following:

- Power consumption: although power cells have improved significantly over the latest years, power consumption is still a critical application parameter. The simpler the video coding algorithm is the lower is the consumption too.
- Computing power & memory: most of these devices have limited computational and memory resources, a serious constrain for every video-coding standard.
- Real time service: video decoding that is performed on the handheld platform should be fast enough to provide real time video to the end user.
- Need for transcoding: original video capture could be high quality and high resolution, however it is required that the video stream should be easily transcoded to lower resolution (since handheld displays are typically low-resolution) and arbitrary bitrates (depending on the available bandwidth).

5. Discussion and Conclusions

Many of the handheld devices already support one or more from the earlier mentioned video coding standards. However it should be noted that video players for such devices are full purpose applications that consume all of the system's resources and can be easily characterized as cell drainers. In an AR scenario, video coding is only a component of the runtime application, thus only a limited fraction of the system resources will be available for this purpose. For that reason we propose building a new video codec that will combine features from various available experimental and standardized codec's, taking also into account the restrictions imposed by the MAR application.

This codec should be very simple, since fast video decoding on the handheld side with only part of the CPU power is a must. Low complexity, speed and small memory footprint will be the main targets. That means that this work can not be based on modern, extremely complex standards, such as H.264. After all, the latest (and others of its generation) requires significant memory resources. Only full pixel accuracy will be used, along with suboptimal motion estimation algorithms. Full search algorithm will also be tested but for a reduced number of the test block pixels, in order to reduce

calculation to acceptable levels. Two pass algorithms are not acceptable because of the real time transmission. Transcoding capabilities will be of prime concern, both in terms of resolution as well as bit rate. There is no standard resolution for handheld displays and so the new standard should be able to deliver a great diversity of different video resolutions to the various devices. In order to easily change the stream bit rate, wavelets will be used for intra frame coding as well as for the encoding of differences. Points for future work are depicted in grey in Figure 4.

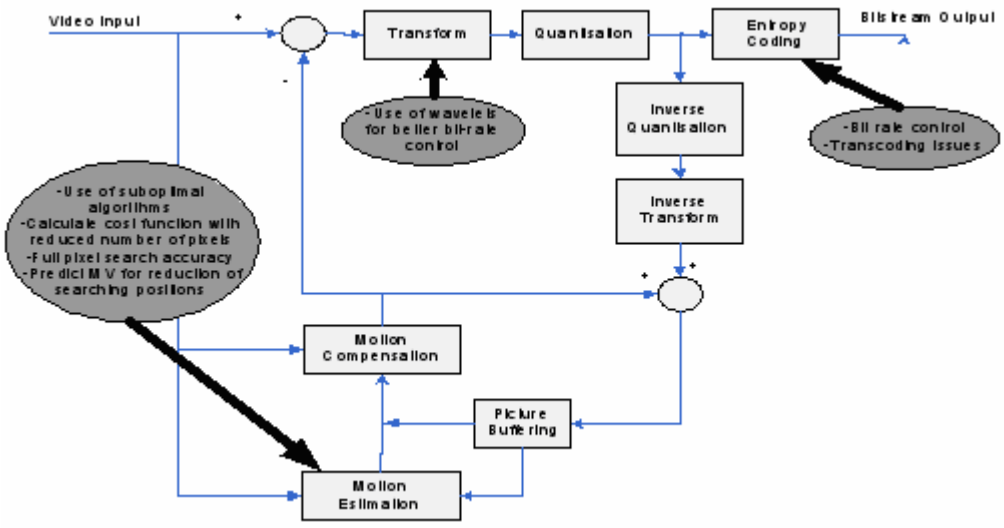

Figure 4. *Points of interest and experimentation for the desired codec*

Acknowledgements

This work was funded by the European Union - European Social Fund (75%), the Greek Government - Ministry of Development - General Secretariat of Research and Technology (25%) and the Private Sector in the frames of the European Competitiveness Programme (Third Community Support Framework - Measure 8.3 - programme ΠΕΝΕΔ - contract no.03ΕΔ832).

References

[1] http://en.wikipedia.org/wiki/Augmented_reality
[2] http://www.liteye.com
[3] http://www.microopticalcorp.com
[4] http://www.icuiti.com
[5] Iain E.G. Richardson (2002). *Video Codec Design (Developing Image and Video Compression Systems*), John Wiley & Sons, Printed in Great Britain, ISBN 0-471-48553-5

[6] Peter Symes (2004). *Digital Video Compression*, McGraw & Hill, Printed in the U.S.A., ISBN 0-07-142487-3

[7] Abdul H. Sadka (2002). *Compressed Video Communications*, John Wiley & Sons, Printed in Great Britain, ISBN 0-470-84312-8

[8] "Video Codec for Audiovisual Services at p x 64 kbit/s", Int. Telecommun. Union-Telecommun. (ITU-T), Geneva, Switzerland, Recommendation H.261, 1993

[9] MPEG-2: ISO/IEC JTC1/SC29/WG11 and ITU-T, "ISO/IEC 13818-2:Information Technology- Generic Coding of Moving Pictures and Associated Audio Information: Video," ISO/IEC and ITU-T, 1994.

[10] Karel Rijkse (December 1996), *H.263: Video coding for low-bit-rate communication*, IEEE Communications Magazine, pp. 42 – 45

[11] MPEG-4: ISO/IEC JTCI/SC29/WG11, "ISO/IEC 14 496:2000-2: Information on Technology-Coding of Audio-Visual Objects-Part 2: Visual," ISO/IEC, 2000.

[12] Gary J. Sullivan, Pankaj Topiwala, Ajay Luthra (August 2004*), The H.264/AVC advanced video coding standard: Overview and introduction to the fidelity range extensions*, SPIE conference on applications of digital image processing

[13] Iain E.G. Richardson (2003) *H.264 and MPEG-4 Video Compression*, John Wiley & Sons, Printed in Great Britain, ISBN 0-470-84837-5

[14] Gary Sullivan and Thomas Wiegand (January 2005), *Video Compression - From Concepts to the H.264/AVC Standard*, Proc. of the IEEE, Special Issue on Advances in Video Coding and Delivery, Vol. 93, No. 1, pp. 18-31

[15] Detlev Marpe, Heiko Schwarz, and Thomas Wiegand (July 2003), *Context-Based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No. 7, pp. 620-636

[16] Sridhar Srinivasan, Shankar L. Regunathan (2005), *An Overview of VC-1,* Visual Communications and Image Processing 2005, Proceedings of SPIE, Vol. 5960

[17] Sridhar Srinivasan, Pohsiang (John) Hsu, Tom Holcomb, Kunal Mukerjee, Shankar L. Regunathan, Bruce Lin, Jie Liang, Ming-Chieh Lee, Jordi Ribas-Corbera (2007), *Windows Media Video 9: overview and applications*, Signal Processing: *Image Communication*, special issue on technologies enabling movies on Internet HD DVD and DCinema 2004, vol. 19, no. 9, pp. 851-875.

[18] Real Networks (2003), *Real Video 10 Technical Overview*, version 1.0, available online at http://www.real.com

[19] On2 Technologies, Inc. (2005), *White Paper TrueMotion VP7 Video Codec, Document Version 1.0,* available online at www.on2.com

[20] T. Borer, T. Davies, A. Suraparaju (2005), *Dirac Video Compression*, Research & Development, British Broadcasting Corporation (BBC), available online at http://dirac.sourceforge.net