Visual Quality of Current Coding Technologies at High Definition IPTV Bitrates

Christian Keimel, Julian Habigt, Tim Habigt, Martin Rothbucher and Klaus Diepold

Institute for Data Processing, Technische Universität München
Arcisstr. 21, 80333 München, Germany
{christian.keimel, jh, tim, martin.rothbucher, kldi}@tum.de

***Abstract*—High definition video over IP based networks (IPTV) has become a mainstay in today's consumer environment. In most applications, encoders conforming to the H.264/AVC standard are used. But even within one standard, often a wide range of coding tools are available that can deliver a vastly different visual quality. Therefore we evaluate in this contribution different coding technologies, using different encoder settings of H.264/AVC, but also a completely different encoder like Dirac. We cover a wide range of different bitrates from ADSL to VDSL and different content, with low and high demand on the encoders. As PSNR is not well suited to describe the perceived visual quality, we conducted extensive subject tests to determine the visual quality. Our results show that for currently common bitrates, the visual quality can be more than doubled, if the same coding technology, but different coding tools are used.**

I. Introduction

Delivery of video content over IP based networks (IPTV) has changed the consumers' TV experience in the last few years. It not only offers on-demand services, but also in many cases an increased choice of content. In particular, high definition content (HDTV) becomes ever more popular, as new network services with higher bitrates allow not only higher quality, but also the simultaneous transmission of multiple channels.

The most common coding technology used for high definition IPTV is the well known *H.264/AVC* [1] standard. Yet, even within this standard, a wide range of visual quality can be achieved at the same bitrate.

Therefore we evaluate in this contribution the visual quality achieved by different coding technologies at bitrates from 5.4 Mbit/s to 30 Mbit/s, representative of current and near future high definition IPTV services over ADSL, ADSL2 and VDSL. Exemplary for all current high definition formats, we used progressive 1080p25 video sequences with a resolution of 1920×1080 pixels and a frame rate of 25 frames per second (fps). We do not consider transmission errors in this contribution, assuming that the video payload is transported error free over the transport network and distortions in the videos are only caused by the encoding process.

MMSP'10, October 4-6, 2010, Saint-Malo, France.
???-?-????-????-?/10/$??.?? ©*2010 IEEE.*

Besides the target bitrate, the visual quality is also determined by the selection of different profiles and levels as described in Annex A of H.264/AVC. Moreover, the chosen coding tools in the encoder can have a significant influence on the visual quality with respect to a fixed bitrate. Hence, we consider two H.264/AVC settings in this contribution representing different levels of encoder complexity. Also, we take alternative coding technologies into account by including the wavelet based *Dirac* [2], [3] encoder into our evaluation.

As the peak signal to noise ratio (PSNR) describes the visual quality only poorly [4], we decided to conduct extensive subjective testing to determine the visual quality of the different encoders at different bitrates and for different video sequences.

Subjective tests were also conducted in related works, but only one bitrate (12MBit/s) and interlaced material was considered [5], [6]. Also only one setting for H.264/AVC was used in these contributions and no alternative encoders were included. In another contribution, MPEG-2 was used to encode the high definition video [7].

This contribution is organized as follows: firstly, we will introduce the different encoders used in this contribution. Then we introduce the encoder scenarios before describing the setup of the subjective tests. Finally, we will discuss the results and conclude with a short summary.

II. Coding Technologies

In order to take into account the performance of different coding technologies for high definition IPTV with respect to the visual quality, we selected two different encoders representing current coding technologies: *H.264/AVC* [1] and *Dirac* [2], [3].

While H.264/AVC is the latest representative of the successful MPEG and ITU-T standards, Dirac is an alternative, wavelet based video codec. Its development was initiated by the British Broadcasting Cooperation (BBC) and was originally targeted at high definition resolution video material. Wheras it follows the common hybrid coding paradigm, it utilizes the wavelet transform instead of the usual block-based transforms e.g. DCT. Hence, it is not necessary for the transform step itself to divide the frame into separate blocks, but the complete frame can be mapped into the wavelet domain in one piece. This fundamental difference to the

(a) CrowdRun (b) ParkJoy (c) InToTree (d) OldTownCross

Fig. 1: Test sequences from the SVT high definition multi format test set.

popular standards of the MPEG-familiy was also the main reason we chose Dirac as the representative of alternative coding technologies in this contribution. Overlapped block-based motion compensation is used in order to avoid block edge artifacts, which due to their high frequency components are problematic for the wavelet transform. Unlike the H.264/AVC reference software, the Dirac reference software version 0.7 used in this contribution offers a simplified selection of settings by just specifying the resolution and frame rate, instead of specific coding tools. Therefore, only the bitrate was varied. Special care was taken by its developers to ensure that Dirac could be implemented royalty free, by avoiding state-of-the-art, but patented algorithms particularly in the area of motion compensation. Still, we will see that it is competitive to H.264/AVC.

For H.264/AVC, we used two significantly different encoder settings, each representing the complexity of various devices and services. The first setting is chosen to simulate a low complexity (LC) H.264/AVC encoder representative of standard devices: many tools that account for the high compression efficiency are disabled. In contrast to this, we also used a high complexity (HC) setting that aims at getting the maximum possible quality out of this coding technology, representing sophisticated broadcasting encoders. We used the reference software [8] of H.264/AVC, version 12.4. The difference in computational complexity is also shown by the average encoding time per frame: 34 and 201 seconds per frame for the LC and the HC H.264/AVC version, respectively. Selected encoding settings for H.264/AVC are given in Table I.

III. Encoder Scenarios

We selected four bitrates from 5.4 Mbit/s to 30 Mbit/s representing real life high definition IPTV applications from the lower end on the bitrate scale e.g. ADSL services at 6 MBit/s, to the upper end on the bitrate scale e.g. 50 MBit/s or higher VDSL services.

The test sequences were chosen from the SVT high definition multi format test set [9] with a spatial resolution of 1920×1080 pixels and a frame rate of 25 fps. The used sequences are shown in Fig.1: *CrowdRun(CR)*, *ParkJoy(PJ)*, *InToTree(ITT)* and *OldTownCross(OTC)*. They represent different coding complexity: InToTree and OldTownCross are encoded rather easily, CrowdRun and ParkJoy are encoded more difficultly due to the large number of moving objects in them. Each of those videos was encoded at the selected four different bitrates. This results in a quality range from 'not acceptable' to 'perfect', corresponding to mean opinion scores (MOS) between 0.19 and 0.96 on a scale ranging from 0 to 1.

The artifacts introduced into the videos by this encoding scheme include pumping effects i.e. periodically changing quality, a typical result of rate control problems, obviously visible blocking, blurring or ringing artifacts, flicker and similar effects. An overview of the sequences and bitrates is given in Table II.

IV. Subjective Testing

The tests were performed in the video quality evaluation laboratory of the Institute for Data Processing at the Technische Universität München in a room compliant with recommendation ITU-R BT.500 [10].

A professional 24 inch LCD reference display with a native resolution of 1920×1080 pixels was used (Cine-tal Cinemagé 2022). The decoded videos were converted to 4:2:2 YC_BC_R for output to the reference display via a HD-SDI link. Due to the screen size, only two viewers took part in the test at the same time to allow stable viewing conditions for all participants. All test subjects were screened for visual acuity and color blindness. The distance between the screen and the observers was three times the picture height.

We used a variation of the standard DSCQS test method as proposed in [11]. This Double Stimulus Unknown Reference (DSUR) test method differs from the standard DSCQS test method, as it splits a single basic test cell in two parts: the first repetition of the reference and the processed video is intended to allow the test subjects to identify the reference video. Only

TABLE I: Selected encoder settings for H.264/AVC

	LC	HC
Encoder	JM 12.4	
Profile&Level	Main, 4.0	High, 5.0
Reference Frames	2	5
R/D Optimization	Fast Mode	On
Search Range	32	128
B-Frames	2	5
Hierarchical Encoding	On	On
Temporal Levels	2	4
Intra Period	1 second	
Deblocking	On	On
8x8 Transform	Off	On

TABLE II: Tested video sequences

Sequence	Frame Rate	Bitrate [MBit/s]
CrowdRun	25 fps	8.4 / 12.7 / 19.2 / 28.5
IntoTree	25 fps	5.7 / 10.4 / 13.1 / 17.1
OldtownCross	25 fps	5.4 / 9.6 / 13.7 / 19.0
ParkJoy	25 fps	9.0 / 12.6 / 20.1 / 30.9

the repetition is used by the viewers to judge the quality of the processed video in comparison to the reference. To allow the test subjects to differentiate between relatively small quality differences, a discrete voting scale with eleven grades ranging from 0 to 10 was used, later rescaled to a range from 0 to 1. Before the test itself, a short training was conducted with three sequences of different content to the test at similiar rate points to the test, resulting in a training session of ten sequences. During this training the test subjects had the opportunity to ask questions regarding the testing procedure. In order to verify if the test subjects were able to produce stable results, a small number of test cases were repeated during the test.

A total of 19 subjects participated in the test, all students with no or little experience in video coding aged 20-30. Processing of outlier votes was done according to [10] and the votes of one test subject were removed based on this procedure. The 95% confidence intervals of the subjective votes are below 0.07 on a scale between 0 and 1 for all single test cases, the mean 95% confidence interval is 0.04. We deteremined the mean opinion score (MOS) by averaging all valid votes for each test case. The results were also used partly in developing the visual quality metrics proposed in [12], [13].

V. Results

In Fig. 2 the results of our conducted subjective tests for the different video sequences and encoders are shown. For comparison we also included the corresponding PSNR.

Firstly, we can notice that the high complexity setting for H.264/AVC, *AVC HC*, outperforms the low complexity setting, *AVC LC*, in all cases. Especially at lower, but currently common bitrates of up to 12 MBit/s in the more demanding sequences CrowdRun and ParkJoy, the visual quality of the high complexity setting is nearly double that of the low complexity setting. At the upper end of the bitrate scale, the difference is not as pronounced, but still the low complexity setting is outperformed by the high complexity setting at a statistically significant margin. This shows clearly that while the use of more sophisticated coding tools increases the computational complexity, the gain in visual quality can be significant.

Secondly, we see that Dirac delivers a better visual quality than the H.264/AVC low complexity setting in most test cases and for high bitrates performs as well as the high complexity setting of H.264/AVC. Thus its use may be sensible in some application scenarios where royalty and patent free standards are needed e.g. in open source projects. In general, this indicates that depending on the requirements, the use of alternative coding technologies instead of predominant standards like H.264/AVC may be a viable option without sacrificing too much performance.

Note that Fig. 2 clearly shows that the PSNR does not correspond very well to the perceived visual quality. Particularly for Dirac, we can observe that while its PSNR is always worse than the two H.264/AVC encoder settings, it still delivers equal or even better visual quality. In fact, if we calculate the Pearson correlation coefficient between PSNR and the visual quality, we achieve only a correlation of 0.69 between both.

VI. Conclusion

We conducted extensive subjective tests to compare different coding technologies at current and near future high definition IPTV bitrates from 5.4 Mbit/s to 30 Mbit/s.

Our results show that even within the same standard, the careful selection of appropriate settings can increase the visual quality significantly, especially at lower, but currently common bitrates. Moreover, we can see clearly that PSNR is not very well suited to determine the visual quality and thus should not be the only criteria in the selection of specific coding technologies or encoder settings.

Even tough we have not included transmission errors in this contribution, it nevertheless provides a good indication for the use of different coding technologies and their performance in real-life applications.

References

[1] *ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG4-AVC), Advanced Video Coding for Generic Audiovisual Services*, ITU, ISO Std., Rev. 4, Jul. 2005.

[2] T. Borer, T. Davies, and A. Suraparaju, "Dirac video compression," BBC Research & Development, Tech. Rep. WHP 124, Sep. 2005.

[3] T. Borer and T. Davies, "Dirac - video compression using open technology," BBC Research & Development, Tech. Rep. WHP 117, Jul. 2005.

[4] B. Girod, *Digital images and human vision*. Cambridge, MA, USA: MIT Press, 1993, ch. What's wrong with mean-squared error?, pp. 207–220.

[5] O. Issa, W. Li, H. Liu, F. Speranza, and R. Renaud, "Quality assessment of high definition TV distribution over IP networks," in *Broadband Multimedia Systems and Broadcasting, 2009. BMSB '09. IEEE International Symposium on*, May 2009, pp. 1–6.

[6] W. Li, O. Issa, H. Liu, F. Speranza, and R. Renaud, "Quality assessment of video content for HD IPTV applications," in *Multimedia, 2009. ISM '09. 11th IEEE International Symposium on*, Dec. 2009, pp. 517–522.

[7] S. Pompei, L. Rea, C. Zema, F. Matera, G. Pierri, E. Binnella, and R. Iacchetti, "Quality of service investigation on high definition IPTV services in a multivendor optical network with adsl access," in *Transparent Optical Networks, 2007. ICTON '07. 9th International Conference on*, vol. 4, Jul. 2007, pp. 32–34.

[8] K. Sühring. (2007) H.264/AVC software coordination. [Online]. Available: http://iphome.hhi.de/suehring/tml/index.htm

[9] SVT. (2006, Feb.) The SVT high definition multi format test set. [Online]. Available: http://www.ldv.ei.tum.de/lehrstuhl/team/Members/tobias/sequences

[10] *ITU-R BT.500 Methodology for the Subjective Assessment of the Quality for Television Pictures*, ITU-R Std., Rev. 11, Jun. 2002.

[11] V. Baroncini, "New tendencies in subjective video quality evaluation," *IEICE Transaction Fundamentals*, vol. E89-A, no. 11, pp. 2933–2937, Nov. 2006.

[12] C. Keimel, T. Oelbaum, and K. Diepold, "Improving the prediction accuracy of video qualtiy metrics." *Acoustics, Speech and Signal Processing, 2010. ICASSP 2010. IEEE International Conference on*, pp. 2442–2445, Mar. 2010.

[13] ——, "No-reference video quality evaluation for high-definition video," *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pp. 1145–1148, Apr. 2009.

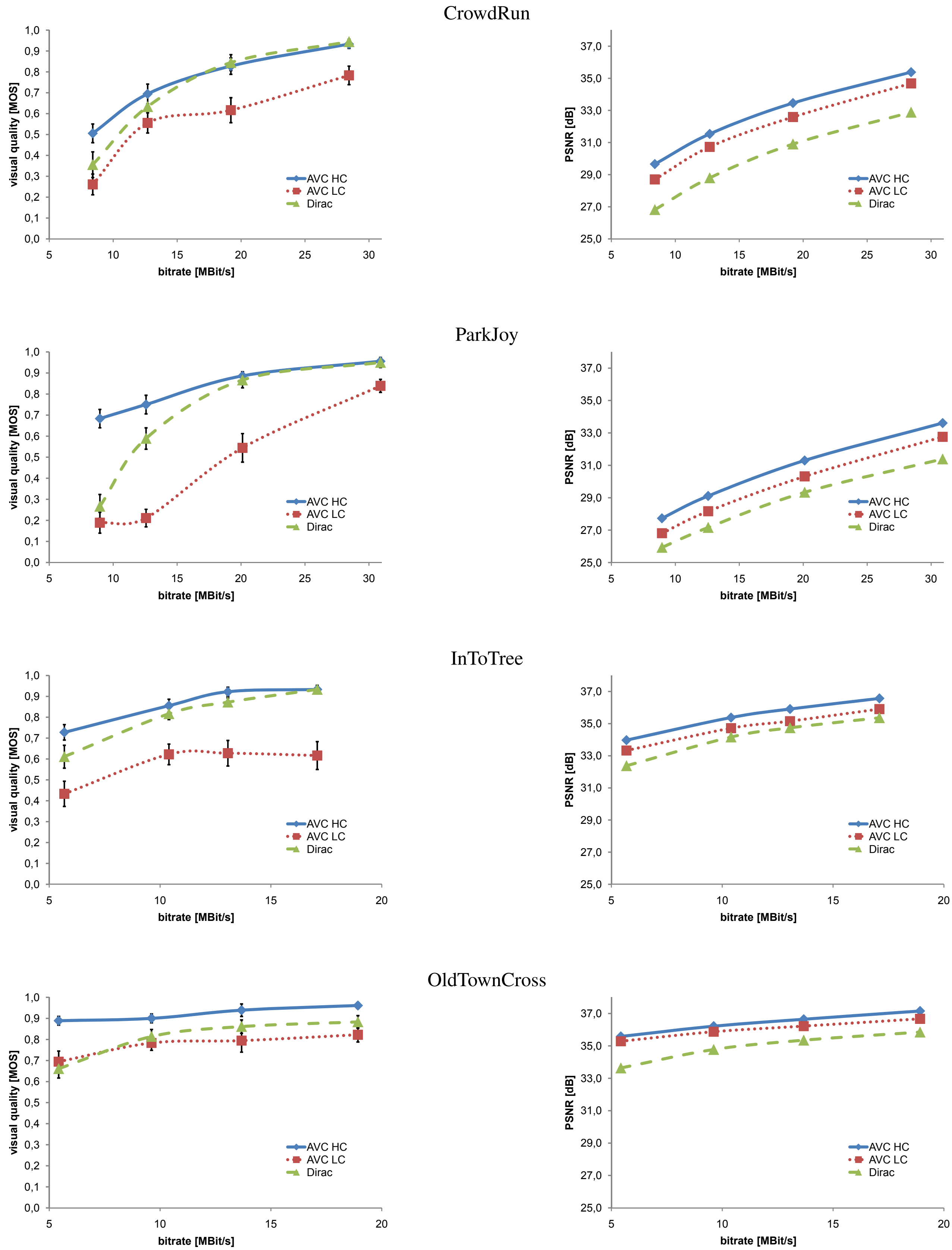

Fig. 2: Visual quality and corresponding PSNR at bitrates from 5.4 Mbit/s to 30 Mbit/s for different video sequences and encoders. The whiskers represent the 95% confidence intervals of the subjective test results for the visual quality.