

COMPARISON OF LOSSLESS VIDEO CODECS FOR CROWD-BASED QUALITY ASSESSMENT ON TABLETS

Christian Keimel, Christopher Pangerl and Klaus Diepold

christian.keimel@tum.de, christopher.pangerl@mytum.de, kldi@tum.de

Institute for Data Processing, Technische Universität München

Arcisstr. 21, 80333 München

ABSTRACT

Video quality evaluation with subjective testing is both time consuming and expensive. A promising new approach to traditional testing is the so-called crowdsourcing, moving the testing effort into the Internet. The advantages of this approach are not only the access to a larger and more diverse pool of test subjects, but also the significant reduction of the financial burden. Extending this approach to tablets, allows us not only to assess video quality in a realistic environment for an ever more important use case, but also provides us with a well-defined hardware platform, eliminating on of the main drawbacks of crowdsourced video quality assessment. One prerequisite, however, is the support of lossless coding on the used tablets. We therefore examine in this contribution the performance of lossless video codecs on the iPad platform. Our results show, that crowd-based video testing is already feasible for CIF-sized videos on tablets, but also that there may be limits for higher resolution videos.

1. INTRODUCTION

Video quality is usually evaluated with subjective testing, as no universally accepted objective quality metrics exist, yet. Subjective testing, however, is both time consuming and expensive. On the one hand this is caused by the limited capacity of the laboratories due to both the hardware and the requirements of the relevant standards, e.g. ITU-R BT.500, on the other hand by the reimbursement of the test subjects that needs to be competitive to the general wage level at the laboratories' locations in order to be able to hire enough qualified subjects.

Crowdsourcing is an alternative to the classical approach to subjective testing that has received increased attention recently. It uses the Internet to assign simple tasks to a group of online workers. Hence tests are no longer performed in a standard conforming laboratory, but conducted via the Internet with participants from all over the world. This not only allows us to recruit the subjects from a larger, more

diverse group, but also to reduce the financial expenditures significantly.

Comparisons between the results from classic and crowd-sourced subjective testing in our previous contributions [1, 2] show a good correlation for some methodologies similar to usual inter-lab correlations. There are, however, still some challenges [3]. In particular, the hardware variation and the necessity for lossless coding.

Tablet-based video quality assessment in combination with crowdsourcing offers not only a more realistic test setup for a ever more important use case of video codecs, but can also provide a uniform hardware and software platform with well-defined parameters. In this contribution we therefore examine a key issue for this application of crowd-based video quality assessment to tablets, the support, but also performance of lossless coding technologies on tablets. A similar concept for tablet-based quality assessment was proposed by Rasmussen in [4] for still images, but not for video. To the best of our knowledge, this is the first contribution focusing on lossless coding performance of video codecs on tablet platforms in the context of crowd-based video quality assessment.

This contribution is organized as follows: after a short introduction into the concept of crowdsourcing and crowd-based video quality assessment, we discuss the extension of the concept to tablets and the lossless coding technologies available for current tablets. After presenting the results of the performance comparison, we conclude with a short summary.

2. CROWDSOURCING

The term Crowdsourcing is a neologism from the words *crowd* and *outsourcing* and describes the transfer of services from professionals to the public via the Internet. These services often consist of tasks which cannot or not efficiently be solved by computers but are simple enough to be performed by non-trained workers, e.g. tagging photos with meaningful key words. However, even rather complex services can be crowdsourced, like creative tasks such as the

generation of new business ideas [5], all kinds of professional design work [5] or financial services via crowd-funding [6]. There are many examples where such services are performed by volunteers, the most prominent one may be Wikipedia, but by now there also exist a number of professional platforms that connect businesses with workers willing to collaborate for a small payment.

The first and still most prominent platform was created in 2005 by Amazon Inc. under the name *Mechanical Turk* where a *requester* can define and place so called *Human Intelligence Tasks (HITs)*. These *HITs* are small tasks which can be performed independently of each other. Any worker who is registered at the platform may choose to perform any *HIT* for the amount of payment which has been assigned to this *HIT* by the requester. There are, however, means to further limit the workforce based on age, nationality, or via a qualification test [1].

3. CROWD-BASED VIDEO QUALITY ASSESSMENT

In crowd-based video quality assessment we utilize these crowdsourcing platforms to perform subjective testing with a global worker pool, usually with a web-based application, that can be accessed via common web browsers, e.g. Firefox or Internet Explorer. Examples of web-based audio-visual quality assessment applications include [1, 2, 7–10].

Videos under test are losslessly compressed and then provided to the test subjects via a web interface in their browser. In order to avoid frame freezes or similar distortions due to limitations in the available bandwidth of the transmission network, the videos are locally cached before the playback starts. Subjects then assess the visual quality and the corresponding judgements are provided to the test manager. The aim is to keep the methodology as close as possible to the methodology used for subjective tests in a lab environment.

The motivation to use lossless compression is based on the fact that in general the worker's web-browser and plugins cannot be assumed to support the original encoding format of the videos under test, as this would necessarily limit our research to already widely adopted coding standards and their profiles. Therefore the videos need to be delivered either uncompressed or only using lossless compression to the workers. This enables us to consider also new coding technologies or other processing algorithms. One could of course re-encode the videos for the delivery with common lossy coding techniques, but then we would move further from the ideal lab setup, as we then also implicitly assess the artefacts introduced by this additional compression.

4. EXTENDING THE CROWD-BASED APPROACH TO TABLETS

Extending the crowd-sourcing approach to tablets, the videos under test are no longer presented using a web browser, but rather with a native application for the tablets' operating system. Additionally, this allows us to assess video quality in a realistic environment and use case that is becoming ever more important.

In this contribution, we chose the Apple iPad family, as it currently has a significant market share in the tablet category and can therefore be considered representative of tablet devices. Also the large market share provides a huge pool of potential workers for the quality assessment tasks. Compared to tablets using the Android operating system, the limited range of devices in the iPad family provides us also with a well-defined set of hardware capabilities.

This addresses one of the main challenges mentioned in [3]: the large variation of different hardware in crowd-based video quality assessment and its potential influence on the quality assessment. In particular, only the brightness can be modified by the user, whereas all other options regarding the presentation e.g. colour rendition of the videos on the iPad are controlled by the operating system and can not be changed.

5. LOSSLESS CODING

The Apple multimedia frame work available on iOS devices supports numerous video coding standards including H.264/AVC, that supports lossless coding in its *High 4:4:4 Predictive* profile. This built-in support for H.264/AVC, however, is limited to the *High* profile, thus excluding native support for the decoding of losslessly encoded H.264/AVC videos.

An alternative is the open source multimedia framework *FFmpeg* [11]. It consists of a collection of different software libraries that provide encoding and decoding functionality for a wide range of video codecs. Also *FFmpeg* is available not only for desktop operating systems, but also tablet operating system including iOS. In the version used in this contribution, the following codecs supporting lossless video compression are included: *H.264/AVC*, *Motion JPEG2000*, *HuffYUV*, *FFvHuff*, *FFv1* and *Lagarith*. H.264/AVC is a well-known ITU/ISO standard and we therefore refer to the existing literature e.g. [12, 13] for further information. Compared to these two standards, the *HuffYUV*, *FFvHuff*, *FFv1* and *Lagarith* codecs were developed as alternatives to uncompressed video within the open-source community.

HuffYUV was proposed by Ben Rudiak-Gould [14] as an alternative to uncompressed YC_bC_r video. It combines an intra-frame prediction with a consecutive Huffman entropy coding of the residuals. The intra-frame prediction



Fig. 1: Video Sequences used in the performance comparison. In the left column from top to bottom: *City*, *Crew*, *Football*, *Foreman* and *Stephan*. In the right column from top to bottom: *MobCal*, *ParkRun*, *Shields* and *Stockholm*

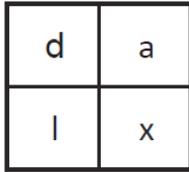


Fig. 2: Intra-frame prediction for HuffYUV

selects between three different prediction models: *left*, *gradient* and *median*. The first model, *left*, only uses the pixel l to predict the pixel x as $x = l$, the second model, *gradient*, predicts x as $x = l + a - d$ and the median model selects the median value from the model *left*, the model *gradient* and from the pixel a above x as illustrated in Fig. 2. Note, that the model selection is not adaptive, but rather one model is selected a-priori to the encoding and then used for the complete video sequence. Also only a fixed table for the Huffman code entropy encoding is used.

FFvYUV [15] is an extension of HuffYUV developed by the FFmpeg project to address some of HuffYUV's shortcomings: instead of a fixed Huffman table, context-adaptive Huffman tables are used for the entropy encoding.

FFvI [16] was also developed by the FFmpeg project and is derived from HuffYUV. The main difference is that the intra-frame coding model is limited to the median model as described above and instead of a fixed Huffman code table, two different options for the entropy coding are available: the first option is a Golomb-Rice variable run length code, the second option an arithmetic code based on [17].

Similar to FFvYUV and FFvI, *Lagarith* [18] is also based on HuffYUV. It combines the median intra-frame coding model with a combination of Huffman variable run length coding followed by arithmetic coding. Unique to *Lagarith* is the support of so-called *null frames*: if the previous frame is identical to the current frame, the current frame is skipped and the decoder will use the previous frame.

6. PERFORMANCE COMPARISON

In order to determine if lossless coding and therefore crowd-based video quality assessment is feasible on a tablet, we compare H.264/AVC, FFvYUV, FFvI and *Lagarith* with respect to the achieved compression ratio and frame rate. As FFvYUV is identical to HuffYUV except for the improvement in the entropy encoding, we will only consider FFvYUV in the performance comparison.

The test device was an iPad2 with iOS 5.1.1 and for encoding the video sequences FFmpeg version 0.7.12 was used. FFmpeg was compiled with the inline assembler option to increase the decoding performance. For H.264/AVC, the High 4:4:4 Predictive profile, with both CABAC and CAVLAC entropy encoding was used, for all other codecs the FFmpeg default settings were used.

We considered in total four different video formats: *CIF*, *576i50*, *720p50* and *1080i50*. For *CIF*, we used *City*, *Crew*, *Football*, *Foreman* and *Stephan*, all at a frame rate of 30 frames per second (fps). The *576i50* and *1080i50* video sequences are *MobCal*, *ParkRun*, *Shields* and *Stockholm*. Lastly, for the *720p50* format we used the sequences *MobCal*, *ParkRun* and *Shields*. All video sequences were provided in $YCbCr$, 4:2:0 format to the different encoders and

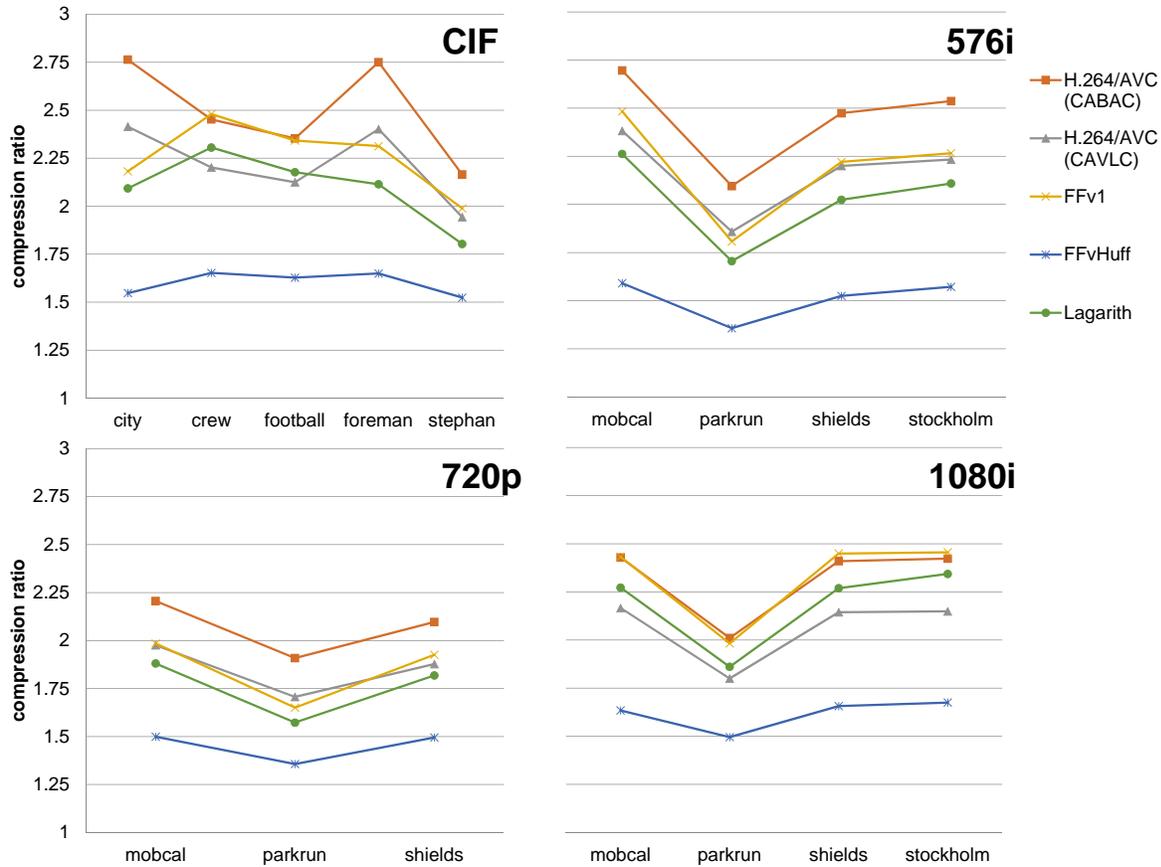


Fig. 3: Compression ratio for different codecs and different video formats

for both interlaced formats the videos were de-interlaced to 50 fps before encoding. All video sequences have a length of 10 s and are shown in Fig. 1.

7. RESULTS AND DISCUSSION

The compression ratios achieved for the different video sequences and codecs are presented in Fig. 3, the achieved frame rate in Fig. 4. Considering the compression ratios, we can see from Fig. 3 that H.264/AVC with activated CABAC outperforms all other codecs for most of the CIF, 576i50 and 720p50 video sequences. Only for the 1080i50 sequences, another codec, FFv1, comes close. Not surprisingly, the lower the complexity of the codec, the worse the compression performance. In particular, we can also notice the difference between CABAC and CAVLC for H.264/AVC. But even H.264/AVC with CABAC as the best performing codec leads to rather large bitrates and corresponding file sizes for the encoded video sequences: for CIF and 576i50, the average bitrate is 1.8 MByte/s and 6.5 MByte/s, respectively, and for 720p50 and 1080i50 the average bitrate is 34 MByte/s for both formats.

If we consider the achieved frame rates in Fig. 4, however, we can see that unfortunately the good compression ratio seems to correspond to more decoding complexity and slower frame rates. For the CIF format, all codecs are able to achieve a frame rate of at least 30 fps, the necessary frame rate to display the video sequences without judder. But for the higher resolution formats, none of the video codecs considered in this contribution is able to achieve the required frame rate for judder-free playback and the frame rate drops nearly linearly with increasing spatial resolution.

This lack in performance could be caused by two different issues: FFmpeg's lack of optimisation for the iOS platform and/or the limited hardware capabilities of tablets in general and the iPad 2 in particular. Although the current iPad platform allows for GPU acceleration using OpenGL ES, the available FFmpeg versions have so far not been adapted and therefore perform all operations on the iPad's CPU. Secondly, due to the high bitrates of up to 34 MByte/s or roughly 272 MBit/s required by the lossless coding, the utilized flash memory may not be able to achieve the necessary data rate for judder-free playback. An indication for this is a periodicity of the judder, typically also observed on desk-

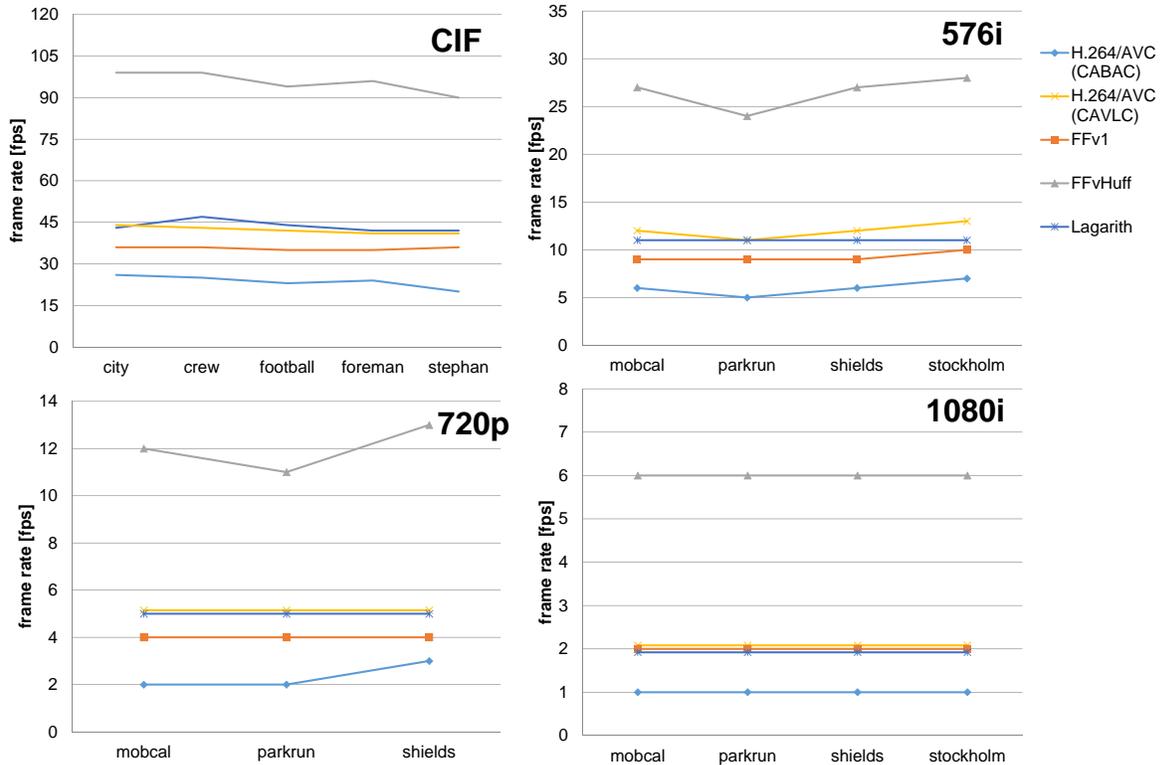


Fig. 4: Achieved frame rate for different codecs and different video formats

top systems in the case of insufficient hard disk data transfer rate. These two points, however, have not been confirmed, yet, and further studies are necessary to determine the exact performance bottle-neck.

8. CONCLUSION

In this contribution we discussed the extension of crowd-based video quality assessment to tablets. After shortly discussing the crowdsourcing principle and its approach to video quality assessment, we extended the concept to tablets, exploiting the uniform hardware and realistic testing environment.

We then examined an important part of this framework, lossless video coding, and compared available lossless codecs for the Apple iPad platform for the complete range of currently used resolutions from CIF to HDTV:

The results show that at least for video sequences in CIF resolution and a CIF-typical frame rate of 30 fps, a tablet based video quality assessment is feasible. For higher resolutions, however, further studies regarding possible hardware limitations of current generation tablets and the evaluation of encoder optimizations on the tablet platform are necessary.

Of course another option would be to avoid lossless coding altogether and instead encode the videos under test with lossy compression. On the one hand, the bitrate must be low enough for play back with the tablet's native multimedia framework, on the other hand the bitrate must be high enough so that additional coding artefacts introduced by the re-encoding are not perceivable. But it is unclear if such an equilibrium is achievable and additional studies are needed to examine this option in more detail.

9. REFERENCES

- [1] C. Keimel, J. Habigt, C. Horch, and K. Diepold, "QualityCrowd - A Framework for Crowd-based Quality Evaluation," in *Picture Coding Symposium (PCS) 2012*, Krakow, Poland, May 2012, pp. 245–248.
- [2] —, "Video Quality Evaluation in the Cloud," in *Proceedings International Packet Video Workshop (PV) 2012*, Munich, Germany, May 2012, pp. 155 – 160.
- [3] C. Keimel, J. Habigt, and K. Diepold, "Challenges in crowd-based video quality assessment," in *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*, Jul. 2012, pp. 13 –18.

- [4] D. R. Rasmussen, "The mobile image quality survey game," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 8293, Jan. 2012, pp. 82 930I–1 – 82 930I–12.
- [5] D. C. Brabham, "Crowdsourcing as a model for problem solving," *Convergence: The International Journal of Research into New Media Technologies*, vol. 14, no. 1, pp. 75–90, 2008.
- [6] A. Gaggioli and G. Riva, "Working the crowd," *Science*, vol. 321, no. 5895, p. 1443, 2008.
- [7] K.-T. Chen, C.-C. Wu, Y.-C. Chang, and C.-L. Lei, "A crowdsourcable QoE evaluation framework for multimedia content," in *Proceedings of the 17th ACM international conference on Multimedia*, ser. MM '09. ACM, 2009, pp. 491–500.
- [8] K.-T. Chen, C.-J. Chang, C.-C. Wu, Y.-C. Chang, and C.-L. Lei, "Quadrant of euphoria: a crowdsourcing platform for QoE assessment," *Network*, *IEEE*, vol. 24, no. 2, pp. 28–35, Mar. 2010.
- [9] F. Ribeiro, D. Florencio, C. Zhang, and M. Seltzer, "Crowdmos: An approach for crowdsourcing mean opinion score studies," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, May 2011, pp. 2416–2419.
- [10] F. Ribeiro, D. Florencio, and V. Nascimento, "Crowdsourcing subjective image quality evaluation," in *Image Processing (ICIP), 2011 IEEE International Conference on*, Sep. 2011, pp. 3158–3161.
- [11] FFmpeg, "About FFmpeg," Aug. 2012. [Online]. Available: <http://ffmpeg.org/about.html>
- [12] D. Marpe, T. Wiegand, and G. Sullivan, "The H.264/MPEG4 advanced video coding standard and its applications," *Communications Magazine, IEEE*, vol. 44, no. 8, pp. 134–143, August 2006.
- [13] J. Heo, S.-H. Kim, and Y.-S. Ho, "Improved CAVLC for H.264/AVC Lossless Intra-Coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 20, no. 2, pp. 213–222, Februar 2010.
- [14] B. Rudiak-Gould, "Huffyuv," Aug. 2012. [Online]. Available: <http://neuron2.net/www.math.berkeley.edu/benrg/huffyuv.html>
- [15] M. Niedermayer, "FFvHuff," Aug. 2012. [Online]. Available: http://svn.ffmpeg.org/doxygen/trunk/huffyuv_8c-source.html
- [16] —, "FFV1 Video Codec Specification," Aug. 2012. [Online]. Available: <http://www1.mplayerhq.hu/~michael/ffv1.html>
- [17] G. Nigel and N. Martin, "Range encoding: an algorithm for removing redundancy from a digitised message," in *Proceedings of the Conference on Video and Data Recording*, ser. IERE conference proceedings. Institution of Electronic and Radio Engineers, Jul. 1979.
- [18] B. Greenwood, "Lagarith lossless video codec," April 2004. [Online]. Available: <http://lags.leetcode.net/codec.html>