Impact of Encoding Configurations on the Perceived Quality of High Definition Videoconference Sequences

Alexandre Ciancio*, José F. L. de Oliveira*, Cassius D. Estrada*, Eduardo A. B. da Silva* and Amir Said†
*PEE/COPPE and DEL/Poli, Universidade Federal do Rio de Janeiro
C. P. 68504 - Rio de Janeiro - RJ, 21941-972, Brazil
Email: [ciancio,jleite,cassius,eduardo]@lps.ufrj.br
†Multimedia Communications and Networking Lab, HP Labs
1501 Page Mill Road, MS 1811, USA
Email: amir_said@hp.com

Abstract— **In this paper we evaluate the impact of different encoding configurations such as compression ratios, frame rates and resolution on the perceived quality of high definition videoconference applications. After generating a high quality video database, degraded sequences had their quality assessed by state-of-the-art automatic metrics. Results have shown that, for low rates, it is preferable to decrease the frame rate and resolution than to increase the compression ratio. Also, a higher frame rate tends to be more important than spatial resolution in terms of perceived quality.**

I. Introduction

High-quality multimedia systems are increasingly present in today's life in the form of HDTV, home-theater, professional telepresence systems, telemedicine etc. In such scenarios, special attention is given to frameworks that allow real-time human interaction such as videoconference applications. Commercial top-notch systems must attain high levels of user satisfaction to deliver a truly realistic multimedia experience. Such practical systems must face the trade-offs between system resources and the quality delivered to users.

Subjective evaluation of multimedia quality is essential and ubiquitous in the operation, maintenance, calibration, optimization, and design of multimedia equipment and systems. However, as communication systems get more powerful, complex and geographically distributed, the costs of human intervention can become quite significant. While expert human analysis definitely cannot be replaced, a number of automatic quality assessment tools which are able to predict the human level of satisfaction in practical systems have been proposed in the literature [1], [2].

Still, little is known regarding the best operating points for high definition encoders in applications other than broadcasting. Although the impact of different configurations on the encoded video quality has been discussed in [3], the authors focus their research on low bit rate, low resolution applications. In this work we use state-of-the-art objective quality metrics to evaluate how different encoding configurations affect the final quality in high definition videoconference applications. We start by generating a high quality, high definition database (1080p, 30fps), containing a number of videoconference sequences. We then encode such sequences using several combinations of frame rates, compression ratios (bit rates) and resolutions. Based on the assessments obtained by the quality metrics we analyze these encoding configurations and suggest general guidelines for choosing the best combination of frame rate and resolution to be used for a given bit rate in order to achieve the best visual quality.

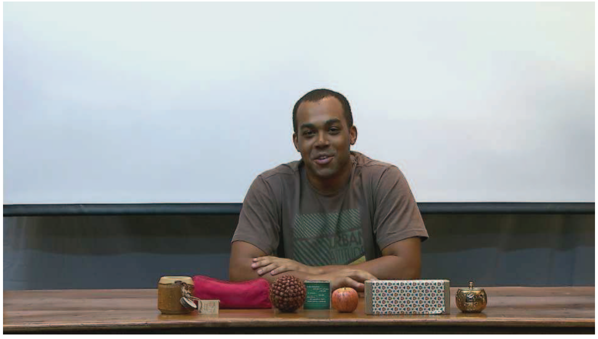
II. Objective Quality Metrics

In this work, two state-of-the-art full reference metrics were selected from the literature due to their excellent correlation to subjective quality assessments. A full reference metric has access to the original, non-degraded video sequence, and generates its quality grade based on a comparison between the degraded and original data.

The Video Quality Metric (VQM), proposed by the NTIA was extensively tested by the Video Quality Experts Group (VQEG) [4], and is part of the ITU Recommendation J.144 [5], that deals with objective quality evaluation of digital video. This metric was designed and tested for a variety of encoding configurations and transmission rates ranging from 10Kbps to 45Mbps. The VQM metric assessment consists in measuring the perceptual effects of a number of video impairments such as blur, unnatural motion, noise, block and color distortion, as well as channel errors (block loss). The individual measurements are pooled over space and time to generate a final grade.

While the VQM metric is based on measuring features related to perceptual models of the human vision, other metrics have been proposed which rely their assessment on the fact that the human visual system (HVS) is specialized in perceiving image structure. The widely used Structural SIMilatity Index

978-1-4673-0219-7/12/$31.00 ©2012 IEEE



(a) Videoconference sequence with plain background.

(b) Videoconference sequence with simple background.

(c) Videoconference sequence with simple background and two subjects.

(d) Videoconference sequence with complex background.

Fig. 1. Snapshots of representative videoconference sequences used during the simulations.

(SSIM), was introduced by Wang and Bovik [6], [7], and is based on the idea that natural images are highly structured and that pixel dependencies carry important information about the image structure. In order to estimate similarity, the algorithm compares the luminosity, the contrast and the structure of a degraded image and its corresponding original, mesuring the loss of structure correlation and the distortions in contrast and luminance. Although the SSIM metric was developed to assess the perceptual quality of static images, it can easily be extended to video sequences by simply computing its average over frames. The SSIM is part of the reference software of the H.264 video standard [8]. Even though it does not take into account temporal features of the sequence, as the VQM does, the SSIM has a much lower computational complexity. The following simulations used the implementation of SSIM available in [9].

III. Test Sequences

31 high quality, high definition (1080i) uncompressed sequences were generated using professional equipment (a Sony PDW-700 camera) at a local TV studio. These sequences included features such as hand movement, face rotation, partial occlusion, high frequency patterns (striped shirts), male and female subjects, all recorded in plain (single color), simple (low frequency) and complex (high frequency) backgrounds. Snapshots of four representative sequences can be seen in Fig. 1. The sequences have a duration of 10 seconds each, and can be freely downloaded for noncommecial use from [10].

IV. Simulations and Results

For the simulations, each of the 31 video sequences was encoded with the configurations shown in Table I (bit rate ranges are given in Mbps), in a total of 6878 degraded sequences. For comparison purposes, the encoded sequences were converted back to their native displaying conditions (1080p/i resolution and 30Hz frame rate) using spatial and temporal interpolation before being evaluated by the metrics. The final obtained grades were averaged over all the 31 sequences.

The results for the VQM metric are plotted in Fig 2 (for better visualization purposes, a zoom of the region between 0 and 3 Mbps and above VQM grade of 0.5 can be seen in Fig. 3). Since the VQM measures distortion, lower grades indicate better performance. It can be seen that the encoding configuration leading to the best quality varies with the bit rate.

Since a similar behavior was obtained for the SSIM metric, the plots for this metric will be ommited, but a tabulation of some of its numerical results is presented later in this section.

The obtained results indicate that, at low transmission rates, better overall quality results are obtained by reducing both the spatial resolution and the temporal resolution of the original video sequences. In other words, it is preferable to decrease

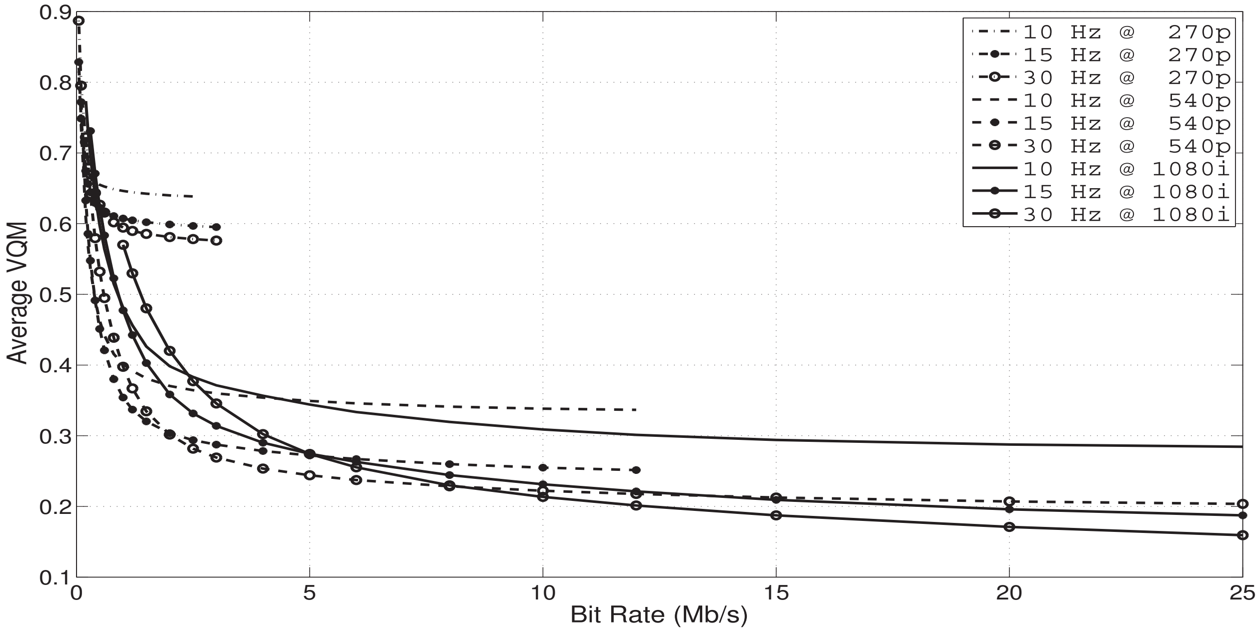


Fig. 2. VQM Average grades Vs. Bit Rates (Mbps).

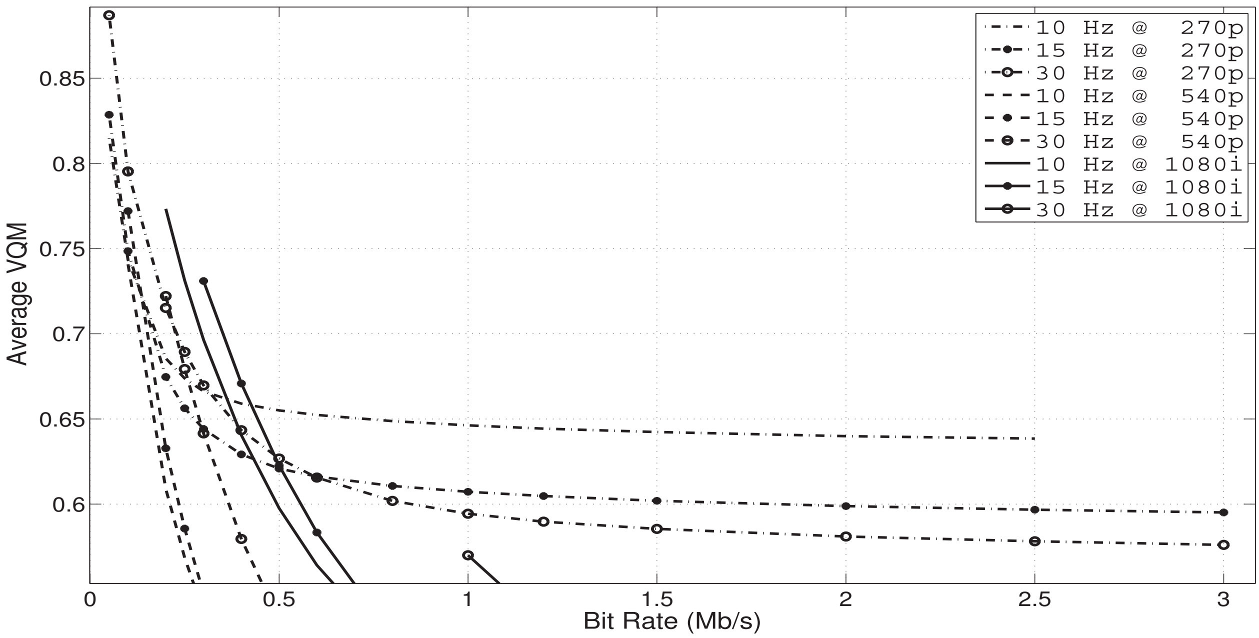


Fig. 3. Zoom of the upper left region of Fig. 2.

temporal and spatial resolutions of the video sequences, instead of simply increasing the compression ratio at the full spatial resolution and frame rate. In addition, the data also suggests that, as bit rate increases, it is better to keep the resolution at 540 lines and gradually increase the frame rate from 10 to 30 frames per second to only then increase the spatial resolution to 1080 lines. It is important to note that the resolution 270p should be avoided in this application, since it is not the best one in any case. Also, by analyzing Figs. 2 and 3, we notice that, when the quality is low, for a given quality level and resolution, the smaller frame rates tend to give smaller bit rates. As the quality increases, there is a tendency to the higher frame rates gradually give better results. In other words, for increasing quality, the frame rates,

TABLE I

ENCODING CONFIGURATIONS USED IN THE SIMULATIONS. BIT RATES ARE GIVEN IN MBPS.

	Horiz. Resolution	Min Bit Rate	Max Bit Rate
10Hz	270p	0.05	2.5
	540p	0.05	12
	1080i	0.1	25
15Hz	270p	0.05	3
	540p	0.05	12
	1080i	0.3	25
30Hz	270p	0.05	3
	540p	0.05	25
	1080i	0.25	25

TABLE II

BEST CONFIGURATIONS BASED ON VQM GRADES.

Bit Rate Range (Mbps)	0 → 0.4	0.4 → 2.0	2.0 → 8	8 → 25
Best Setting	540p @ 10Hz	540p @ 15Hz	540p @ 30Hz	1080i @ 30Hz

in decreasing bitrate order, gradually change from [10Hz 15Hz 30Hz] to [15Hz 10Hz 30Hz] to [15Hz 30Hz 10Hz] to [30Hz 15Hz 10Hz]. That is, for high qualities the smaller bitrates are given by higher frame rates. Such conclusions can provide a useful guidance for the design of videoconference applications. The best overall configurations for the generated database based on the VQM are summarized in Table II, and the ones from SSIM in Table III. One can see that although the bitrate ranges shown in the tables differ a little, the plots show a similar behaviour pattern, and therefore, the conclusions one can take from the SSIM are essentially the same as the ones from the VQM. This corroborates with the idea that, at lower bit rates, lower frame rates lead to better quality and, as bit rate increases, we should first increase frame rate to only then increase spatial resolution.

TABLE III

BEST CONFIGURATIONS BASED ON SSIM GRADES.

Bit Rate Range (Mbps)	0 → 0.5	0.5 → 1.0	1.0 → 4.4	4.4 → 25
Best Setting	540p @ 10Hz	540p @ 15Hz	1080i @ 15Hz	1080i @ 30Hz

TABLE IV

BIT RATE RANGES (IN MBPS) FOR WHICH EACH FRAME RATE LEADS TO BETTER QUALITY FOR A GIVEN RESOLUTION.

		Frame Rate		
		10Hz	15Hz	30Hz
Resolution	270p	0 → 0.13	0.13 → 0.6	0.6 → 3
	540p	0 → 0.4	0.4 → 1	1 → 25
	1080i	0 → 0.94	0.94 → 5	5 → 25

It is interesting to notice that the same behavior can be observed when we consider each resolution individually. Table IV lists the bit rate ranges for which each frame rate leads to better quality for a given resolution. This data shows that, for a given resolution, the best qualities at lower bit rates are also achieved by using lower frame rates and, as the bit rate increases, the higher frame rates lead to better qualities.

V. CONCLUSION

In this paper we evaluated the effect of different encoding configuration in the perception of quality in high definition videoconference applications. We have generated a high quality database containing 31 uncompressed sequences acquired with professional equipment at a local TV studio. 1395 encoded sequences were generated and had their quality measured by state-of-the-art objective metrics. Results have shown that when such applications face bandwidth limitations, a better overall quality for the user may be obtained by reducing the spatial and temporal resolution of the sequences and, as bit rate restrictions loosen, by gradually increasing frame rate before increasing spatial resolution.

REFERENCES

[1] Z. Wang, H. R. Sheikh, and A. C. Bovik, "Objective video quality assessment," *in The Handbook of Video Databases: Design and Applications (B. Furht and O. Marqure, eds.)*, pp. 1041–1078, Sept 2003, cRC Press.

[2] H. R. Wu and K. R. Rao, *Digital Video Image Quality and Perceptual Coding*. CRC Press, 2006.

[3] G. Zhai, J. Cai, W. Lin, X. Yang, W. Zhang, and M. Etoh, "Cross-dimensional perceptual quality assessment for low bit-rate videos," *IEEE Trans. Multimedia*, vol. 10, no. 7, November 2008.

[4] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment, Phase II," Tech. Rep., August 2003. [Online]. Available: http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseII

[5] "ITU-T J.144 Recommendation: Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," International Telecommunication Union, Telecommunication Standardization Sector, Tech. Rep., March 2004.

[6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, April 2004.

[7] M. P. Sampat, Z. Wang, S. Gupta, A. Bovik, and M. K. Markey, "Complex wavelet structural similarity: A new image similarity index," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2385 – 2401, November 2009.

[8] Joint Video Team, "H.264/AVC Reference Software Version 18 (JM18)," 2011. [Online]. Available: http://iphome.hhi.de/suehring/tml/download/

[9] Z. Wang, "Structural similarity (SSIM) index software." [Online]. Available: https://ece.uwaterloo.ca/~z70wang/research/ssim/

[10] "High Quality HD Videoconference Database." [Online]. Available: http://www.lps.ufrj.br/~tvdigital/futurasequences/