Additional Results for Live Video Frame Rate Control Experiments

Ivan V. Bajić and Xiaonan Ma

***Abstract*—This supplement provides Peak Signal-to-Noise-Ratio (PSNR) results from our live video frame rate control test.**

***Index Terms*—Variable frame rate control, live video, Peak Signal-to-Noise-Ratio (PSNR) results.**

I. Introduction

THIS supplement provides additional results for the frame rate control tests described in [2] in the form of Peak Signal-to-Noise-Ratio (PSNR). As described in [2], the frame rate control testbed consists of four computers (labeled A-D) as shown in Fig. 1. The same analog video is being captured end encoded at two computers (A and C), which implement two frame rate control methods: one described in [1], the other proposed in [2]. Both A and C encode the incoming video using the same SPIHT-based encoder from [3]. However, since A and C implement different frame rate control policies, they capture frames at different times.

In such a configuration, it is possible to compute PSNR for each frame rate control method separately, since each of them has its own reference video. But, in our opinion, these PSNR values are not directly comparable, because they are not computed against a *common* reference video. The two methods capture frames at different times, so there is no *common* reference against which both can be compared. This is also the reason why extensive subjective evaluations were performed to evaluate video quality in [2]. To give an example, let's say one of the methods captures frames at 0 ms, 40 ms, 75ms, etc., while the other captures them at 0ms, 35ms, 90ms, .etc. If we were to compare both videos against the first set of frames, this would penalize the second video, because the frames of the second video are offset in time with respect to that reference. These frames would receive lower PSNR even though they may be visually just as good or even better than the frames of the first video. Similar argument holds if we choose the second set of frames as the reference. If, on the other hand, we choose for each frame the analog (or very finely quantized)

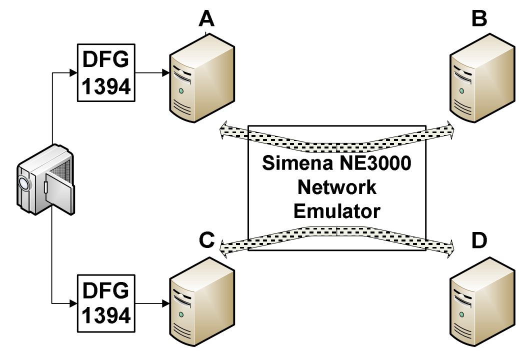

Figure 1. Experimental testbed.

scan lines from the analog video at that time, then neither set of frames is favored (which is good), but the PSNR values are computed against different reference frames, which, in our opinion, would make them incomparable. Nonetheless, due to a request from one of the reviewers and for completeness purposes, we provide these PSNR results below.

In particular, we have computed the average PSNR values in the following way. First, the Mean Squared Error (MSE) between the n-th video frame captured at A (before encoding) and the corresponding frame decoded at B is computed as

$$\mathrm{MSE}_{n,c}^{\mathrm{A,B}} = \frac{1}{N_{pixel}} \sum_{i=1}^{N_{pixel}} \left(X_{n,c}^{\mathrm{A}}[i] - \hat{X}_{n,c}^{\mathrm{B}}[i] \right)^2 ,$$

where N_{pixel} is the number of pixels in a frame, $X_{n,c}^{\mathrm{A}}$ is the c-component (c $\in$ {Red, Green, Blue}) of the n-th frame captured at A, and $\hat{X}_{n,c}^{\mathrm{B}}$ is the c-component of the corresponding frame decoded at B. The PSNR value (in dB) of the c-component between these two frames is

$$\mathrm{PSNR}_{n,c}^{\mathrm{A,B}} = 10\log_{10} \frac{255^2}{\mathrm{MSE}_{n,c}^{\mathrm{A,B}}} .$$

Finally, the average PSNR for the video between A and B is

$$\mathrm{PSNR}^{\mathrm{A,B}} = \frac{1}{3N^{\mathrm{A,B}}} \sum_{c\in\{R,G,B\}} \sum_{n=1}^{N^{\mathrm{A,B}}} \mathrm{PSNR}_{n,c}^{\mathrm{A,B}} ,$$

where $N^{\mathrm{A,B}}$ is the number of frames in video from A to B. Analogous PSNR value is computed for video C→D. Note that because A and C use different frame rate control policies, the video C→D generally contains a different number of frames from the video A→B. That is, $N^{\mathrm{A,B}} \neq N^{\mathrm{C,D}}$ in general.

This work was supported by the NSERC/CCA Grant STPGP 350740.

I. V. Bajić is with the School of Engineering Science, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada. (Tel: 1-778-782-7159; Fax: 1-778-782-4951; e-mail: ibajic@ensc.sfu.ca)

X. Ma is with the School of Engineering Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. (e-mail: xma2@sfu.ca)

II. RESULTS

TABLE I
PSNR (dB) RESULTS FOR $\alpha = 0.25$

Movement	Bit rate (kbps)	Ref. [1]	Ours
Talking head	400	33.9	27.3
	850	40.8	34.7
	1000	39.2	37.3
	1500	41.3	40.4
Walking	400	38.8	32.4
	850	43.2	35.7
	1000	43.0	37.1
	1500	41.9	37.5
Handwave	400	26.1	25.2
	850	42.3	34.7
	1000	42.3	36.6
	1500	42.1	38.1
Camera pan	400	36.8	31.9
	850	43.0	36.7
	1000	43.2	38.5
	1500	43.8	40.0

TABLE II
PSNR (dB) RESULTS FOR $\alpha = 0.35$

Movement	Bit rate (kbps)	Ref. [1]	Ours
Talking head	400	32.8	28.6
	850	40.6	32.3
	1000	38.3	32.8
	1500	40.3	38.0
Walking	400	33.7	29.2
	850	42.4	34.9
	1000	42.5	34.3
	1500	43.6	40.1
Handwave	400	25.6	25.4
	850	41.4	26.1
	1000	41.1	30.1
	1500	41.0	37.0
Camera pan	400	31.0	26.5
	850	43.2	29.1
	1000	44.2	28.5
	1500	40.8	30.2

TABLE III
PSNR (dB) RESULTS FOR $\alpha = 0.5$

Movement	Bit rate (kbps)	Ref. [1]	Ours
Talking head	400	36.4	33.5
	850	41.9	38.1
	1000	40.1	38.2
	1500	42.3	40.9
Walking	400	38.4	32.0
	850	42.3	37.2
	1000	41.8	37.7
	1500	40.6	39.9
Handwave	400	33.6	34.3
	850	40.2	34.9
	1000	41.9	35.2
	1500	41.6	37.6
Camera pan	400	36.4	34.6
	850	41.5	38.5
	1000	43.8	38.3
	1500	43.2	41.3

As in [2], results are sorted by α values. Please note that some PSNR values at higher bit rates seem lower than the PSNR of the corresponding movement type at lower bit rates. Although these PSNR values are not directly comparable (see below), we are currently trying to determine whether this was partially caused by misalignment of the frames between sender and receiver, or possibly by skipped frames.

Note that videos of the same motion type (e.g., talking head) are not the same for each bandwidth constraint, because they were captured at different times, with the subjects trying their best to reproduce the same motion in each recording. If we wanted to capture the same analog video with two frame rate control policies at four bandwidths and three alpha values, we would have needed 2×4×3 = 24 transmitting computers (instead of 2, A and C in Fig. 1), 24 receivers (instead of 2), and at least 4 network emulators (instead of 1). Hence, comparisons cannot be made between different rows, or across different tables. Within the same row, two videos do come from the same analog original, but their individual frames are captured at different times. Hence, because their PSNRs are computed with respect to different "originals," they are not comparable in the same way as they would be if they were computed against the same original video.

III. DISCUSSION AND CONCLUSIONS

As discussed in [2], the method from [1] allocates lower frame rate (i.e., produces fewer frames) in general, so for the same bandwidth constraint, the average PSNR is higher. Since the PSNR computed as described in Section I essentially measures the fineness of quantization, the frames arising from the rate control policy [1] tend to be more finely quantized, and tend to have higher PSNR than the frames arising from our frame control policy.

But this is only part of the story, because PSNR does not take frame rate into account. Consider the following example. If one wishes high PSNR (and only high PSNR), one could simply capture a single frame and give it all the bits that would have been assigned to the entire sequence, thereby creating an extremely finely quantized (possibly lossless) image with an extremely high PSNR. Of course, one would then have a hard time convincing anyone that what was produced is a video; indeed, it would be a *still image*, not video.

Therefore, to judge the quality of video, one must take into account frame rate in addition to frame quality. The issue is not producing high PSNR, but producing the best tradeoff between frame rate and frame quality. Once a quality metric that takes both frame rate and frame quality into account, and accurately matches human notion of video quality is developed, we will be able to compute that quality from the video. Until such time, the best we can do is compare the video quality subjectively. That is what was done in [2]. As discussed there, observers showed statistically significant preference for the video produced by our method, despite the fact that the quality of individual frames was lower on average. Evidently, what the observers cared about is not the quality of individual frames, but the tradeoff between frame rate and frame quality. Our method, with its rapid increase in frame rate at the onset of motion, apparently produces a better tradeoff between these video parameters. The reasons are discussed in more detail in [2].

REFERENCES

[1] H. Song and C.-C. Jay Kuo, “Rate control for low-bit-rate video via variable-encoding frame rates,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 4, pp. 512-521, Apr. 2001.

[2] I. V. Bajić and X. Ma, “A testbed and methodology for comparing live video frame rate control methods,” accepted with minor revisions for publication in *IEEE Signal Processing Letters*, October 2010.

[3] I. V. Bajić and X. Ma, “MCL.JIT library for scalable live video in Max/MSP/Jitter,” *Proc. IEEE CCECE'10*, Calgary, AB, May 2010.