Scope of validity of PSNR in image/video quality assessment

Q. Huynh-Thu and M. Ghanbari

Experimental data are presented that clearly demonstrate the scope of application of peak signal-to-noise ratio (PSNR) as a video quality metric. It is shown that as long as the video content and the codec type are not changed, PSNR is a valid quality measure. However, when the content is changed, correlation between subjective quality and PSNR is highly reduced. Hence PSNR cannot be a reliable method for assessing the video quality across different video contents.

Introduction: Peak signal-to-noise ratio (PSNR) has traditionally been used in analogue systems as a consistent quality metric. However, digital video technology has exposed some limitations of PSNR. Yet, because of its low complexity, PSNR is still used as a video quality metric for evaluating image processing algorithms (e.g. video de-noising methods) and is considered to be a reference benchmark for developing perceptual video quality metrics. However, some studies have shown that PSNR poorly correlates with subjective quality [1]. PSNR is also still widely used in evaluating codec performance (e.g. as a measure of gain in quality for a specified target encoding bit rate), video codec optimisation or as a comparison method between different video codecs [2], despite the fact that objective perceptual quality metrics have been shown to outperform PSNR in predicting subjective video quality [3]. PSNR is at the centre of a constantly fuelled debate: is PSNR completely irrelevant for evaluating the quality of digitally compressed video or the performance of video codec? Is PSNR wrongly used as a quality metric? This Letter presents some experimental data that demonstrate where and why PSNR can or cannot be used as a quality metric.

Experiment: Ten source (reference) video contents of 8 s duration at CIF resolution covering a wide range of spatio-temporal characteristics were selected and encoded with H.264 at several bit rates from 24 to 800 kbit/s. The bit rates were chosen to produce compressed (degraded) videos with various levels of quality for each video content. A subjective test was conducted with 40 test sequences using the ACR method and with an experimental setup following international standards [4].

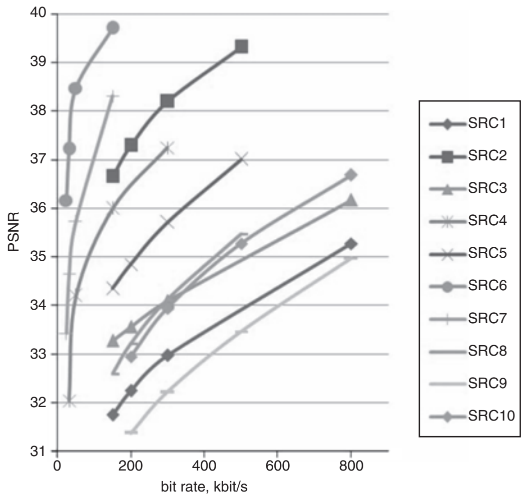

Fig. 1 *PSNR variation with bit rate, per content*

Analysis: Fig. 1 shows the PSNR variation with the bit rate for each content. For each sequence, PSNR is measured as $1/N \sum_{i=1}^{N} 10 \log\left(255^2/\varepsilon^2(i)\right)$, where $\varepsilon^2(i)$ is the pixel luminance mean-squared error between corresponding frame i in the reference and compressed videos, and N is the number of frames in the degraded video. When each content is considered individually, PSNR always increases monotonically with bit rate for that content, and this is consistent for all contents. The subjective quality variation with the bit rate for each content is shown in Fig. 2. This Figure shows that, when each content is considered individually, subjective quality always increases monotonically with bit rate for that content, and this is again consistent for all contents. Combination of Figs. 1 and 2 indicates that, for a specified content, PSNR always monotonically increases with subjective quality as the bit rate increases. The implication is that, within a specified codec and fixed content, the variation of PSNR is a reliable indicator of the variation of quality. Hence, in the context of codec optimisation, PSNR can therefore be used as a performance metric as it correlates highly with subjective quality when the content is fixed. For example, PSNR can be used for testing different codec optimisation strategies designed to maximise the subjective quality of a specified content (i.e. the content remains the same between the optimisations). Tables 1 and 2 provide PSNR and quality values for two examples of content (SRC3 and SRC9). In each case, the correlation between PSNR and quality is well above 0.9.

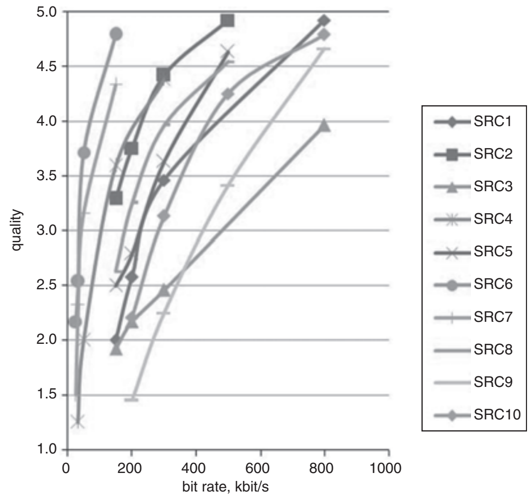

Fig. 2 *Quality variation with bit rate, per content*

Table 1: PSNR and quality for SRC3

Bit rate	PSNR	Quality
150	33.2712	1.92
200	33.5519	2.17
300	34.0897	2.46
800	36.1635	3.96

Table 2: PSNR and quality for SRC9

Bit rate	PSNR	Quality
200	31.3712	1.33
300	32.2077	2.13
500	33.4570	3.29
800	34.9695	4.54

On the other hand, Figs. 3 and 4 illustrate the problem of using PSNR as a quality metric across different contents. Fig. 3 shows the variation of PSNR with subjective quality for content SRC3 and SRC9 (Fig. 3*a*), the variation of PSNR with bit rate (Fig. 3*b*) and the corresponding variation of subjective quality with bit rate (Fig. 3*c*). We can observe in Figs. 3*a* and *b* that the PSNR curve for SRC3 is always on top of the PSNR curve for SRC9. In other words, if PSNR were used as a quality predictor, it would indicate that the quality for SRC3 is always higher than the quality of SRC9 at all bit rates. However, the subjective data in Fig. 3*c* show that this is not true. This Figure shows that the subjective quality is lower for SRC9 at the lower bit rates but the trend is inverted for higher bit rates where the quality of SRC9 is higher than the quality of SRC3, indicating that, although the monotonic relationship between PSNR and subjective quality exists separately per content, it does not exist anymore across contents. Fig. 4 shows that, if PSNR is used across contents, it loses its prediction accuracy. The correlation between PSNR and subjective quality for the data in Fig. 4 is only 0.71. The data in Table 3 show that, when each content is considered separately, the Pearson correlation between PSNR and subjective quality is well above 0.9, but the value dramatically drops to 0.71 when all data points from the ten contents are considered together in the test set. If more sources of content were jointly assessed, the Pearson correlation would be much lower than this. Furthermore,

Fig. 4 clearly shows that different video contents with the same PSNR have in fact a very different subjective quality. For example, SRC6 at 24 kbits/s and SRC3 at 800 kbits/s both have a PSNR = 36.1 dB but are characterised by a subjective quality score of 2.17 and 3.96, respectively, for SRC6 and SRC3. Our results show that PSNR is therefore unreliable as an objective metric for predicting subjective quality. It is worth mentioning that, although the test was conducted with CIF-resolution video, the same results can be expected for other resolutions as codec behaviour will be identical at other resolutions. Furthermore, the measured correlation coefficient of 0.71 between the PSNR and subjective quality shown in Table 3 is only for ten video contents. Adding more contents can only reduce this correlation coefficient further, whereas the individual correlation coefficient for each content remains high. This reiterates our conclusion that PSNR is not a reliable measure of quality across various video contents, but it is reliable within the content itself.

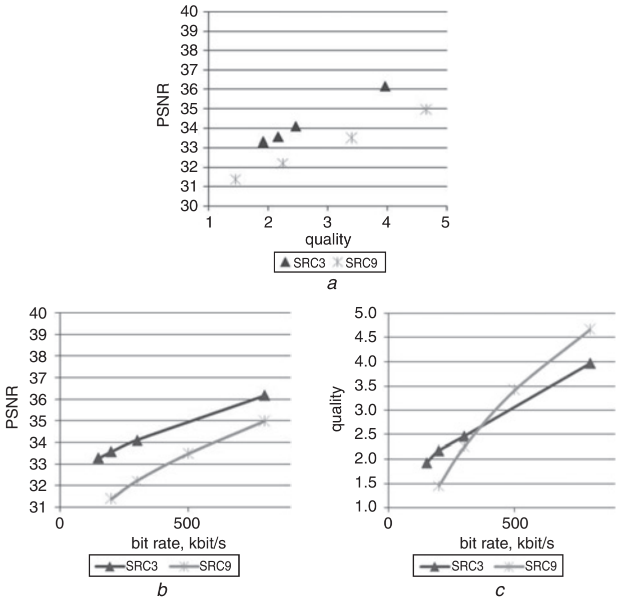

Fig. 3 *Unreliability of PSNR as quality metric*

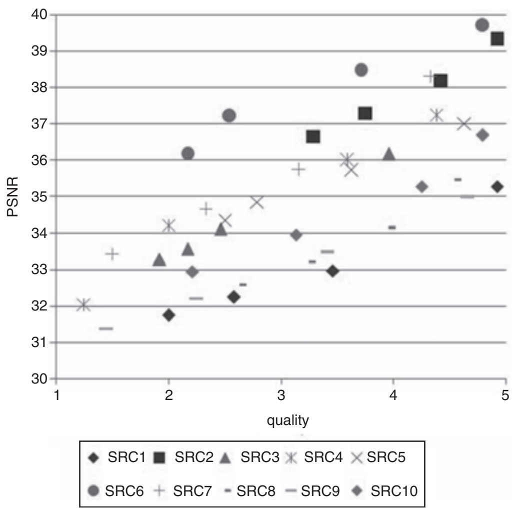

Fig. 4 *Inaccuracy of PSNR as predictor of subjective quality*

Table 3: Correlation between PSNR and quality per content and across contents

Content	All	SRC1	SRC2	SRC3	SRC4	SRC5	SRC6	SRC7	SRC8	SRC9	SRC10
Correlation	0.71	0.98	0.99	0.99	0.98	0.99	0.98	0.99	0.98	0.99	0.98

Conclusions: We have shown that PSNR can be used as a good indicator of the variation of the video quality when the content and codec are fixed across the test conditions, e.g. in comparing codec optimisation settings for a given video content. On the other hand, we have shown that PSNR is an unreliable video quality metric when different contents are considered in the test conditions.

25 February 2008
Electronics Letters online no: 20080522
doi: 10.1049/el:20080522

Q. Huynh-Thu (*Psytechnics Ltd, Fraser House, 23 Museum Street, Ipswich IP1 1HN, United Kingdom*)

E-mail: quan.huynh-thu@psytechnics.com

M. Ghanbari (*Department of Computing and Electronic Systems, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom*)

References

1 Girod, B.: 'What's wrong with mean-squared error?' *in* 'Digital images and human vision' (MIT Press, 1993), pp. 207–220
2 Wiegand, T., Schwarz, H., Joch, A., Kossentini, F., and Sullivan, G.: 'Rate-constrained coder control and comparison of video coding standards', *IEEE Trans. Circuits Syst. Video Technol.*, 2003, **13**, (7), pp. 688–703
3 ITU: 'Objective perceptual video quality measurement techniques for digital cable', ITU-T Recommendation J.144, Geneva, Switzerland, March 2003
4 ITU: 'Subjective video quality assessment methods for multimedia applications', ITU-T Recommendation P.910, Geneva, Switzerland, September 1999