Picture Appraisal Rating

(PAR) - A Single-ended Picture Quality Measure for MPEG-2

Mike Knee

Snell & Wilcox, UK

Abstract: A method of estimating the picture quality of an MPEG-2 bitstream without requiring reference to the source picture is presented. The method uses information readily available from the early stages of decoding the bitstream. It is therefore highly suited to applications in which it is required to monitor several video channels and provide an automatic warning of possible picture quality problems. A theoretical treatment of the subject is presented, followed by a description of how the PAR algorithm for the PAR measure was obtained. Test results are given which show that the PAR measure correlates well with the conventional peak signal-to-noise measure (PSNR), which itself remains largely unsurpassed as a method of picture quality measurement.

snellwilcox.com

Putting Pictures to Work™ SNELL & WILCOX™

Introduction

This paper describes a method of estimating the picture quality of an MPEG-2 video bitstream as deined by ISO/IEC (1). The method uses information readily available from the early stages of decoding the bitstream. It is therefore highly suited to applications in which it is required to monitor several video channels and provide an automatic warning of possible picture quality problems.

The first section of this paper discusses the problem of picture quality monitoring in compressed systems and surveys some of the solutions that have already been proposed. The following section then provides a theoretical basis for a single-ended measure based on parameters readily available from an MPEG-2 bitstream. Then, experimental results on the accuracy of the resulting Picture Appraisal Rating (PAR) measure are presented. The final section briefly presents a practical implementation of the method, available commercially as the PAR™ measure.

Measuring picture quality in compressed systems

Double-ended measurement

The natural approach to measuring the effect on picture quality of a compression process, or indeed any process that might be expected to degrade the quality, is to make a comparison between the processed image and the unprocessed source, as shown in Figure 1.

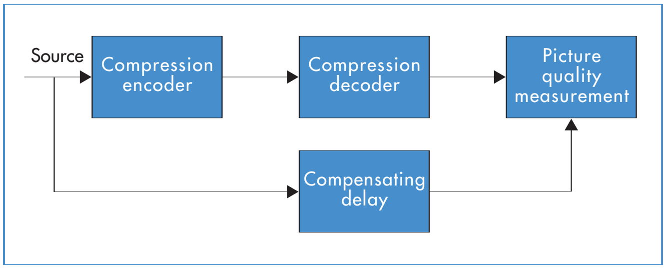


Figure 1 Double-ended picture quality measurement

This approach is known as *double-ended* measurement because access is required to the image at both ends of the chain. The compensating delay shown must precisely align the two pictures to be compared. This can be done by incorporating rugged, easily identifiable regions into the source images or sequences, from which spatial and temporal alignment information can be retrieved even when the sequence has undergone severe degradation. Double-ended measurement is not well suited to monitoring applications where the source is neither available nor controllable, but it is useful as a laboratory tool where those restrictions pose less of a problem.

The most commonly accepted double-ended measure is the peak signal-to-noise ratio (PSNR). There are slight variations in the exact definition of this measure, but for the purposes of this paper it is defined as follows:

1)

$$PSNR(dB) = 10\log_{10}\frac{255^2}{\left\langle (x_i - y_i)^2 \right\rangle}$$

where x_i are the (8-bit) samples of the source sequence, y_i are the corresponding samples at the decoder output, and the angled brackets denote averaging over some suitable time interval, which may be a single picture or group of pictures (GOP) and which may be taken blockwise or as a rolling average.

In this paper, the *noise power* is also used. This is the denominator of the expression in equation 1. It follows that estimates of PSNR and of noise power are interchangeable.

Can PSNR be improved upon?

Several attempts have been made to replace the PSNR measure with something that models more closely the behaviour of the human visual system. The problem is typically attacked in two ways. First, we note that the visibility of a given distortion depends on the local content of the source picture. For example, distortions are usually more objectionable in plain areas and on clearly defined edges than they are in 'busy' areas of the picture. Second, it is possible to model the visual effect of the distortion itself in a more sophisticated way than simply measuring its energy, as in the case of PSNR. For example, a weighting function may be applied in the frequency domain, giving more weight to the lower-frequency components of the error than to the higher-frequency components.

A notable attempt to develop an objective picture quality measure that improves on PSNR is described by Lubin (2). This measure is based on the concept of Just Noticeable Differences (JND). Equipment that measures both PSNR and the JND-based measure, using off-line processing based on specially generated source sequences, is available.

The reader should note the recent work of the ITU study group VQEG (Visual Quality Experts' Group) (3), in which several sophisticated double-ended quality measures have been compared with each other and with PSNR. The study concludes that, for MPEG-2 coding at bit rates of around 0.8 Mbit/s and higher, none of the measures is significantly more accurate than PSNR itself. We may therefore conclude, for the time being at least, that PSNR may safely be used as a benchmark for any picture quality measure we might wish to develop or to evaluate.

Single-ended measurement

In order to overcome the problems of unavailability of the source and of alignment between the source and processed sequences, methods of *single-ended* measurement have been proposed. Single-ended measurement has actually been in existence for a number of years in the form of noise measurement, a good example of which is described by Drewery et al (4).

More recently, attention has been paid to the specific single-ended measurement of impairments introduced by compression coding. Methods of doing this can be divided into two classes, (i) those that look at the decoded image alone without reference to the bitstream, and (ii) those that make use of information from the bitstream which is normally discarded once it has been used to produce the decoded picture. Class (ii) is the more general case, and a top-level diagram of its architecture is given in Figure 2.

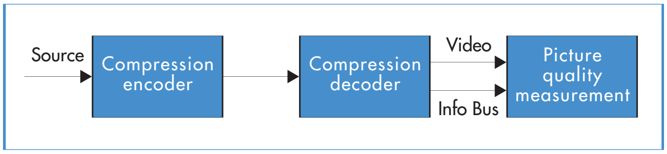


Figure 2 Single-ended picture quality measurement

Here, the picture quality measurement device makes use not only of the decoded video but also of information from the bitstream, which is conveniently formatted into a signal synchronous with the video, known as the Information Bus, described by Knee et al (5). The MPEG Re-coding Data Set with its transport mechanism, which are about to become SMPTE standards (6, 7), is a practical way in which the Information Bus data can be conveyed to a picture quality measurement device.

An example of single-ended quality measurement for MPEG-2 bitstreams is given by Lauterjung (8). This exploits the fact that blocking artefacts are among the most objectionable visual impairments in a decoded MPEG-2 picture, and makes direct measurements of the 'blockiness' based on pixel differences across block boundaries and pixel differences within blocks. Equipment that makes blockiness and other single-ended measurements on decoded pictures is available.

The method presented in this paper differs from those discussed so far in that it makes use *only* of parameters available from the early stages of decoding a bitstream. It does not require access to a decoded picture, nor does it require processing capability at decoded video rates. It is thus well suited to multi-channel monitoring where processing simplicity is important, and would also be ideal for HDTV bitstreams where the cost of generating and processing a decoded picture is relatively high.

Theory of bitstream-based quality measurement

In this section, we give a theoretical background to picture quality measurement based on MPEG bitstream parameters, in particular quantisation parameters. The basic assumption is that the impairment being estimated is the noise due to DCT coefficient quantisation in the MPEG-2 coding process.

Measuring noise in the DCT coefficient domain

We wish to estimate the noise power in the pixel domain. The first point to establish is that, for a transform-based compression scheme such as MPEG-2, this is equivalent to estimating the noise power in the transformed coefficient domain. Because the transform is orthogonal, we may invoke Parseval's Theorem to show that the noise power in one domain is directly proportional to the noise power in the other domain, provided factors such as the quantiser weighting matrices are taken into account correctly.

Quantisation noise power

It is well known that the quantisation noise power from a linear quantiser of spacing q in a uniformly distributed signal is given by the expression

2)

$$\frac{1}{q}\int_{-q/2}^{q/2} x^2 dx = \frac{q^2}{12}$$

In MPEG-2 coding, the signal consists of DCT coefficients and (apart from the intra DC coefficient) has a highly non-uniform probability distribution. The simple expression for the quantisation noise power therefore has to be replaced by the following general expression for each coefficient:

3)

$$\sum_{i=1}^{M} \frac{1}{d_{i+1} - d_i} \int_{d_i}^{d_{i+1}} p(x)\left(x - r_i\right)^2 dx$$

where $d_i, i=1 \ldots M$ are the quantiser decision levels, r_i are the corresponding reconstruction levels (with $d_i \; d \le r_i \le d_{i+1}$) and $p(x)$ is the probability density function of the input signal x, here assumed to be a continuous function.

If the quantiser has *2N+1* uniformly spaced reconstruction levels centred on zero (as defined, for example, in the MPEG-2 specification for intra coefficients) and has a decision threshold offset parameter l (defined by Werner et al (9); λ = 0 corresponds to truncation and λ = 1 to rounding), then the expression for the quantisation noise power becomes

4)

$$2\left[\int_0^{(1-\lambda/2)q} p(x)x\,dx + \sum_{i=1}^{N} \frac{1}{q} \int_{(i-\lambda/2)q}^{(i+1-\lambda/2)q} p(x)\left(x - iq\right)^2 dx\right]$$

This expression depends only on three quantities:

- the quantiser level spacing q, which is known from the quantiser scale code and q_scale_type parameters received in the MPEG bitstream
- the decision threshold offset parameter λ. This is not known but it is reasonable to suppose that it takes the compromise value 0.75 in the case of intra coding, as defined in the MPEG Test Model (10)
- the probability distribution $p(x)$ of the input signal

It follows that the quantisation noise power can be estimated if the shape of the probability distribution of the input signal can be found. This information is of course not directly available, but it is in principle possible to estimate it from the reconstructed signal at the output of the inverse quantiser. This approach has some disadvantages:

- it is complicated
- if the quantiser is coarse, the granularity of the measurements becomes correspondingly coarse and it becomes difficult or impossible to obtain a reliable approximation to the source distribution

In the following paragraphs we therefore develop a simpler approach.

In image processing theory it is common to try to model the probability distribution of a signal such as a particular DCT coefficient, by a function specified by one or two parameters. For example, the Laplacian, or two-sided exponential distribution, is often used:

5)

$$p(x) = \frac{\alpha}{2} e^{-\alpha|x|}$$

The shape of the distribution is determined by the single parameter a. If this parameter can be estimated from the decoded DCT coefficients, the general expression (equation 4) for the quantisation noise power can be calculated exactly; in fact it is given by the expression

6)

$$\frac{2}{\alpha^2} - \frac{q e^{\alpha q\left(\frac{\lambda}{2} - 1\right)}}{\alpha\left(1 - e^{-\alpha q}\right)}\left[q\alpha\left(1 - \lambda\right) + 2\right]$$

The parameter a depends on the activity of the original picture or prediction error signal at the spatial frequency represented by the particular DCT coefficient under consideration. In principle, source material might exhibit different levels of activity at different frequencies. In practice, however, we find that there is a degree of interdependence between the activity levels across the spectrum. This is not surprising if we observe that features that contribute to high activity, such as edges, have energy over a wide range of spatial frequencies. It is therefore reasonable to suppose that the huge variations possible in the shapes of the probability distributions for all the DCT coefficients can be parametrised by a single quantity which describes the overall activity of the source. Then, the quantisation noise power in the DCT domain can be estimated from this single 'activity' parameter together with the quantiser scale parameter q.

Towards a practical PSNR measure

Having accepted that a single parameter may suffice to account for the dependence of quantisation noise power on the probability distribution of the source data, we observe that it is unnecessary to impose a particular shape such as a Laplacian on the data and then attempt to fit the data to the model. Instead, we could take a representative set of source data and directly observe the dependence of quantisation noise power on q and a measure of picture activity.

The 'Picture Appraisal Rating' (PAR') measure was derived using such a representative set of source data covering sport, drama, effects, film etc., each encoded using a reference encoder with several different quantiser_scale values.

The steps taken in deriving the functions used in the PAR measure were as follows:

- A picture-dependent linear relation between the directly measured PSNR and the logarithm of quantiser_scale was established.
- A further linear relation between the slope of the above PSNR function and a picture activity measure was established
- Finally, a quadratic function was derived to estimate the picture activity measure from the number of coefficient bits and the quantiser_scale value

Thus, the basic PAR PSNR estimate depends on two quantities only: the quantiser_scale value and the number of coefficient bits. Because of the variability in quantiser_scale value from macroblock to macroblock, we found that the PAR estimate was more accurate if calculated on a macroblock basis.

The actual PAR algorithm, which is not described in detail in this paper, is somewhat more complex than what we have shown so far, in the following respects:

- the practical PAR algorithm takes account of non-standard quantiser weighting matrices in the bitstream, which affects the relation between PSNR calculated in the DCT coefficient domain and the actual pixel-based PSNR
- the PSNR values for P and B pictures are expressed in different terms than for terms of the I-frames. PSNR estimate with a correction to account for the values of quantiser_scale and of quantiser weighting matrices encountered. This has been done because the direct modelling of picture activity in terms of quantiser_scale values and bit counts is not as reliable for predictive coding as it is for intra coding.

More information is available in the patent application of Knee and Diggins (11).

Comparison between PAR™ and PSNR

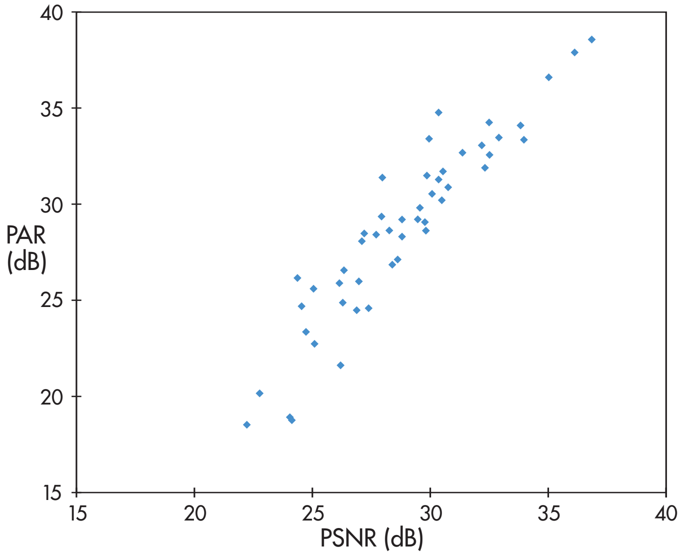
We now describe some comparisons between the PAR estimate and directly measured PSNR. Fifty coded sequences were used. They can be divided into two sets.

The first set, which we call the *broad* set, contains twenty sequences selected to cover an extremely wide range of coding parameters and source characteristics, including:

- source material: ten source sequences including film, sport, slow-moving detail
- coded bit rate: 0.9 Mbit/s to 12.0 Mbit/s
- constant bit rate and variable bit rate
- coding algorithms: the MPEG SSG encoder (10) and two proprietary algorithms
- quantiser weighting matrices: four different intra matrices and seven different non-intra matrices
- GOP length: 1 to 18
- coded picture sizes: 720 ¥ 576 to 480 ¥ 288
- temporal sampling rates: 50, 59.94 and 25 fields/sec
- sequence length: 96 to 499 pictures

The second set, known as the *narrow* set, contains thirty 36-frame sequences, all generated by the same coding algorithm with fairly standard parameters, from ten source sequences each coded at 2.0, 4.5 and 6.5 Mbit/s.

Figure 3 Correlation between PAR and PSNR



The PSNR figures for all fifty sequences are plotted against the PAR estimates in Figure 3. In each case, the value plotted for each sequence is the first quartile of the set of frame-by-frame values. This choice of value was made from several possibilities including the mean and other percentile values, to maximize the correlation between PSNR and subjective quality in informal subjective tests.

What is the correlation between PAR and PSNR for the fifty sequences? Two quantities were calculated:

- The *Pearson product-moment correlation coefficient.* This takes the value 0.933
- The *coefficient of determination,* which is effectively the correlation coefficient when the straight line through the data is forced to have the equation y=x. This takes the slightly lower value of 0.907

When the results are plotted for the broad and narrow data sets separately, it is interesting to note that there is very little difference in spread between the scatter plots obtained in each case. This reinforces the impression that PAR is a rugged algorithm for PSNR estimation.

A practical implementation of the PAR measure

A practical, real-time implementation of the PAR measure is available from Snell & Wilcox Ltd. in the form of the MVA200, a real-time MPEG bitstream analyser. The analyser's PAR function provides a rolling graph of systems performance which may be used to report on key events where PAR values fall below a defined level. Applications so far identified for the MVA200 and its PAR function include

- systems performance optimization
- encoding mode debugging
- noise reduction assessment
- performance monitoring of statistical multiplexing

Further details on the MVA200 and associated products may be found on the company's Internet site (12).

More compact, multichannel, HDTV-capable and software-only versions of the PAR measurement system are being plannedhave also been developed.

Conclusions

We have presented a single-ended coded MPEG picture quality measure known asas available commercially as the PAR™ measure. the Picture Appraisal Rating (PAR). It This measure uses information readily available from the early stages of decoding an MPEG video bitstream, namely quantiser_scale, quantiser weighting matrices and bit counts. It does not require access to the source picture, nor to the decoded picture. It is therefore ideally suited to multichannel monitoring applications, including those involving HDTV bitstreams.

A comparison was made between the PAR measure and PSNR, which itself remains largely unsurpassed as a double-ended picture quality measure. A very wide variety of source material, coding algorithms, bit rates, quantisation parameters, GOP structures and subsampling structures was used. The results show a coefficient of determination between the PAR measure and PSNR of better than 90%, or a correlation coefficient better than 93%.

These results confirm that the PAR measure is a reliable indicator of PSNR and hence of picture quality for monitoring applications. Equipment implementing the PAR measure in real time is available.

References

1. ISO/IEC TR 13818-2 : 1997. Information technology – generic coding of moving pictures and associated audio information. Part 2. Video.
2. Lubin, J. 1995. A visual discrimination model for imaging system design and evaluation. David Sarnoff Research Center, February 1995.
3. ITU Study Group VQEG (Visual Quality Experts' Group). See http://www.itu.int and ftp://ftp.its.bldrdoc.gov/dist/ituvidq
4. Drewery, J.O., Storey, R. and Tanton, N.E. 1984. Video noise reduction. BBC Research Department Report no. BBC RD 1984/7, July 1984.
5. Knee, M.J., Wells, N.D. and Devlin, B.F. Video compression. European Patent Application EP 0 845 908 A2. European Patent Office, 3 June 1998.
6. SMPTE 327M-1999. The MPEG-2 re-coding data set. Draft SMPTE Standard, 1999, available for public comment (see http://www.smpte.org)
7. SMPTE 319M-1999. Transporting MPEG-2 re-coding information through 4:2:2 component digital interfaces. Draft SMPTE Standard, 1999, available for public comment (see http://www.smpte.org)
8. Lauterjung, J. Picture quality measurement. In: International Broadcasting Convention, Amsterdam, 11-15 September 1998. IEE, London, 1998. pp. 413-417.
9. Werner, O.H., Wells, N.D and Knee, M.J. 1998. Digital signal compression encoding with improved quantisation. International Patent Application WO 98/38800. World Intellectual Property Bureau, 3 September 1998.
10. ISO/IEC TR 13818-5 : 1997. Information technology – generic coding of moving pictures and associated audio information. Part 5. Software simulation.
11. Knee, M.J. Diggiens J. Improvements in data compression International Patent Application WO 00/22834. World Intellectual Property Bureau, 20 April 2000.
12. Snell & Wilcox Internet site: http://www.snellwilcox.com

Acknowledgements

The author would like to thank his colleagues at Snell & Wilcox Ltd. for their contributions to this work and the Directors of the company for permission to publish this paper.

snellwilcox.com

For further information

please contact one of our international sales offices:

Americas

New York
Snell & Wilcox Inc.
274 Madison Avenue
Suite 1704
New York City
NY 10016
Tel: +1 212 481 1830
Fax: +1 212 481 2642
info@snellamerica.com

Burbank
Snell & Wilcox Inc.
3519 Pacific Ave.
Burbank, CA 91505
Tel: +1 818 556 2616
Fax: +1 818 556 2626
info@snellamerica.com

Asia Pacific

Hong Kong
Snell & Wilcox (Hong Kong) Ltd.
Room 603, Tai Tung Building
No.8 Fleming Road
Wanchai
Hong Kong
Tel: +852 2356 1660
Fax: +852 2575 1690
swhk@snellwilcox.com.hk

China
Snell & Wilcox China
Unit 17
12th Floor of Tower B
Beijing COFCO Plaza
No.8 Jianguomennei Dajie
Dongcheng District
Beijing 100005
China
Tel: +86 10 6528 2761
Fax: +86 10 6528 7635
swchina@snellwilcox.com

India
Snell & Wilcox (India Liaison Office)
I.B. Centre
N-6 Greater Kailash-1 (Market)
New Delhi 110 048
India
Tel: +91 (0) 11 641 2609
Fax: +91 (0) 11 621 8777
swindia@snellwilcox.com

Japan
Snell & Wilcox Japan KK
3-1-3 Ebisu
Shibuya-ku
Tokyo
150-0013
Japan
Tel: +81 (0) 33 446 3996
Fax: +81 (0) 35 449 7392
swjapan@snellwilcox.com

South East Asia

Snell & Wilcox
South East Asia Sdn. Bhd.
Suite 9-09
Plaza 138
Jalan Ampang
50450 Kuala Lumpur
Malaysia
Tel: +60 (0) 3 2732 5557
Fax: +60 (0) 3 2732 8669
swmalaysia@snellwilcox.com

Europe, Middle East & Africa

UK
Snell & Wilcox Ltd.
Southleigh Park House
Eastleigh Road
Havant
Hampshire
PO9 2PE
United Kingdom
Tel: +44 (0)23 9248 9000
Fax: +44 (0)23 9245 1411
info@snellwilcox.com

France
Snell & Wilcox France
3 rue de Rome
93110 Rosny Sous Bois
France
Tel: +33 (0) 1 45 28 1000
Fax: +33 (0) 1 45 28 6452
swfrance@snellwilcox.com

Germany
Snell & Wilcox GmbH
Senefelderstrasse 3a
65205 Wiesbaden
Germany
Tel: +49 (0) 6122 98430
Fax: +49 (0) 6122 9843 55
swgermany@snellwilcox.de

Russia
Snell & Wilcox ZAO
Room 214
35 Arbat str.
Moscow 121835
Russia
Tel: +7 095 248 3443
Fax: +7 095 248 1104
arbat35@snellwilcox.ru

Code Corp14 03/06 v1