International Telecommunication Union

ITU-R

Radiocommunication Sector of ITU

Recommendation ITU-R BT.1908-0

(01/2012)

Objective video quality measurement techniques for broadcasting applications using HDTV in the presence of a reduced reference signal

BT Series

Broadcasting service (television)

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from http://www.itu.int/ITU-R/go/patents/en where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Recommendations

(Also available online at http://www.itu.int/publ/R-REC/en)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	**Broadcasting service (television)**
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management
SNG	Satellite news gathering
TF	Time signals and frequency standards emissions
V	Vocabulary and related subjects

Note: *This ITU-R Recommendation was approved in English under the procedure detailed in Resolution ITU-R 1.*

Electronic Publication
Geneva, 2020

RECOMMENDATION ITU-R BT.1908-0*

Objective video quality measurement techniques for broadcasting applications using HDTV in the presence of a reduced reference signal

(2012)

Scope

This Recommendation specifies methods for estimating the perceived video quality of broadcasting applications using HDTV when a reduced reference signal is available.

The ITU Radiocommunication Assembly,

considering

a) that the ability to automatically measure the quality of broadcast video has long been recognized as a valuable asset to the industry;

b) that Recommendation ITU-R BT.1683 describes objective methods for measuring the perceived video quality of standard definition digital broadcast television in the presence of a reduced reference;

c) that Recommendation ITU-R BT.709 describes parameter values for the HDTV standards for production and international programme exchange and Recommendation ITU-R BT.500 describes subjective assessment methods for image quality including high-definition television;

d) that HDTV is becoming widely used in broadcasting;

e) that ITU-T Study Group 9, based on the results of the HDTV report sent by VQEG, has produced Recommendation ITU-T J.342, which specified objective video quality measurement of HDTV in the presence of a reduced reference;

f) that objective measurement of the perceived video quality of HDTV may complement subjective assessment methods,

recommends

1 that the objective video quality model given in Annex 1 should be used for objective measurement of perceived video quality for broadcasting applications using HDTV in the presence of a reduced reference signal.

* Radiocommunication Study Group 6 made editorial amendments to this Recommendation in February 2020 in accordance with Resolution ITU-R 1.

Annex 1

1 Introduction

This Recommendation provides a video quality measurement method for use in high definition television (HDTV) non-interactive applications when the reduced reference (RR) measurement method can be used. The model was compared to subjective quality scores obtained using Recommendation ITU-R BT.500. Analyses showed that the accuracy of this model was equivalent to that of PSNR.

For the RR model to operate correctly, the unimpaired source video should be available for the model to extract parameters. These extracted parameters as well as the degraded video sequence are the inputs to the RR model. The estimation method performs both calibration (i.e. gain/offset and spatial/temporal registration) and objective video quality estimation.

The validation test material contained both ITU-T H.264 and MPEG-2 coding degradations and various transmission error conditions (e.g. bit errors, dropped packets). The model in this Recommendation may be used to monitor the quality of deployed networks to ensure their operational readiness. The visual effects of the degradations may include spatial as well as temporal degradations. The model in this Recommendation can also be used for lab testing of video systems. When used to compare different video systems, it is advisable to use a quantitative method (such as that in Recommendation ITU-T J.149) to determine the model's accuracy for that particular context.

This Recommendation is deemed appropriate for broadcasting services delivered between 1 Mbit/s and 30 Mbit/s. The following resolutions and frame rates were considered in the validation test:

- 1080/59.94/I
- 1080/25/P
- 1080/50/I
- 1080/29.97/P.

The following conditions were allowed in the validation test for each resolution:

Test factors
Video resolution: 1 920 × 1 080 interlaced and progressive
Video frame rates 29.97 and 25 frames per second
Video bit rates: 1 to 30 Mbit/s
Temporal frame freezing (pausing with skipping) of maximum 2 seconds
Transmission errors with packet loss
Conversion of the SRC from 1080 to 720/P, compression, transmission, decompression, and then conversion back to 1080
Coding technologies
H.264/AVC (MPEG-4 Part 10)
MPEG-2

Note that 720/P was considered in the validation test plan as part of the test condition (HRC). Because currently 720/P is commonly up-scaled as part of the display, it was felt that 720/P HRCs would more appropriately address this format.

1.1 Applications

The applications for the estimation models described in this Recommendation include, but are not limited, to:

1) Video quality monitoring at the receiver when side-channels are available.

2) Video quality monitoring at measurement nodes located between the point of transmission and the point of reception.

The model described in this Recommendation provides a statistically similar performance to PSNR, yet it can be used for video quality assessment when the reduced reference signal is available at the point of measurement.

1.2 Limitations

The video quality estimation model described in this Recommendation cannot be used to replace subjective testing. Correlation values between two carefully designed and executed subjective tests (i.e. in two different laboratories) normally fall within the range 0.95 to 0.98. This Recommendation cannot be used to make video system comparisons (e.g. comparing two codecs, comparing two different implementations of the same compression algorithm). The performance of the video quality estimation model described in this Recommendation is not statistically better than PSNR.

When frame freezing was present, the test conditions typically had frame freezing durations less than 2 seconds. The model in this Recommendation was not validated for measuring video quality in a re-buffering condition (i.e. video that has a steadily increasing delay or freezing without skipping). The model was not tested on other frame rates than those used in TV systems (i.e. 29.97 frames per second and 25 frames per second, in interlaced or progressive mode).

It should be noted that in case of new coding and transmission technologies producing artefacts which were not included in this evaluation, the objective model may produce erroneous results. Here, a subjective evaluation is required.

Note that the model in this Recommendation was not evaluated on talking-head content typical of video-conferencing scenarios.

2 References

The following ITU Recommendations and other references contain provisions, which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU Recommendations is regularly published.

The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

Recommendation ITU-T J.244 (2008), *Full reference and reduced reference calibration methods for video transmission systems with constant misalignment of spatial and temporal domains with constant gain and offset*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 Subjective assessment (picture): The determination of the quality or impairment of programme-like pictures presented to a panel of human assessors in viewing sessions.

3.1.2 Objective perceptual measurement (picture): The measurement of the performance of a programme chain by the use of programme-like pictures and objective (instrumental) measurement methods to obtain an indication that approximates the rating that would be obtained from a subjective assessment test.

3.1.3 Proponent: An organization or company that proposes a video quality model for validation testing and possible inclusion in an ITU Recommendation.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 Frame rate: The number of unique frames (i.e. total frames – repeated frames) per second.

3.2.2 Simulated transmission errors: Errors imposed upon the digital video bit stream in a highly controlled environment. Examples include simulated packet loss rates and simulated bit errors. Parameters used to control simulated transmission errors are well defined.

3.2.3 Transmission errors: Any error imposed on the video transmission. Example types of errors include simulated transmission errors and live network conditions.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ACR	Absolute Category Rating (see Recommendation ITU-R BT.500)
ACR-HR	Absolute Category Rating with Hidden Reference (see Recommendation ITU-T P.910)
AVI	Audio Video Interleave
DMOS	Difference Mean Opinion Score
FR	Full Reference
FRTV	Full Reference TeleVision
HRC	Hypothetical Reference Circuit
ILG	VQEG's Independent Laboratory Group
MOS	Mean Opinion Score
MOSp	Mean Opinion Score, predicted
NR	No (or Zero) Reference
PSNR	Peak Signal-to-Noise Ratio
PVS	Processed Video Sequence
RMSE	Root Mean Square Error
RR	Reduced Reference

SFR Source Frame Rate

SRC Source Reference Channel or Circuit

VQEG Video Quality Experts Group

YUV Colour Space and file format

5 Conventions

None.

6 Description of the reduced reference measurement methods

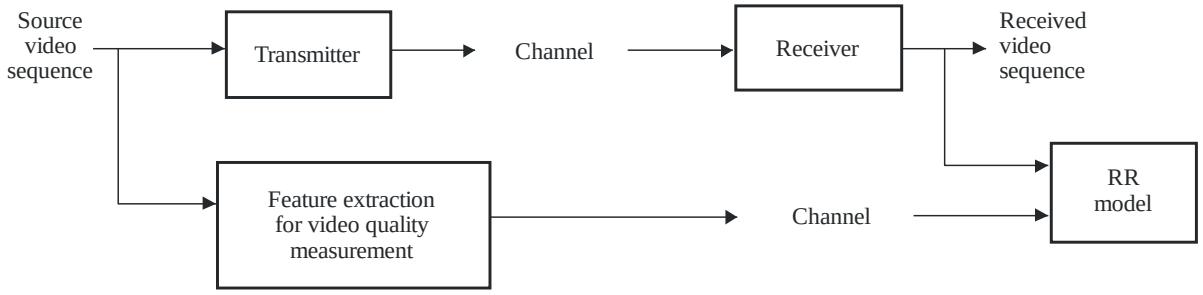
6.1 Introduction

Although PSNR has been widely used as an objective video quality measure, it is also reported that it does not well represent perceptual video quality. By analysing how humans perceive video quality, it is observed that the human visual system is sensitive to degradation around the edges. In other words, when the edge pixels of a video are blurred, evaluators tend to give low scores to the video even though the PSNR is high. Based on this observation, the reduced reference models which mainly measure edge degradations have been developed.

Figure 1 illustrates how a reduced-reference model works. Features which will be used to measure video quality at a monitoring point are extracted from the source video sequence and transmitted. Table 1 shows the side-channel bandwidths for the features, which have been tested in the VQEG HDTV test.

FIGURE 1

Block diagram of reduced reference model



BT.1908-01

TABLE 1

Side-channel bandwidths

Video format	Tested bandwidths
1080/60 Hz (29.97 fps) 1080/30Pp (29.97 fps)	56 kbit/s, 128 kbit/s, 256 kbit/s
1080/25Pp (25 fps) 1080/50I Hz (25 fps)	56 kbit/s, 128 kbit/s, 256 kbit/s

6.2 The EPSNR reduced-reference model

6.2.1 Edge PSNR (EPSNR)

The RR models mainly measure on-edge degradations. In the models, an edge detection algorithm is first applied to the source video sequence to locate the edge pixels. Then, the degradation of those edge pixels is measured by computing the mean squared error. From this mean squared error, the edge PSNR is computed.

One can use any edge detection algorithm, though there may be minor differences in the results. For example, one can use any gradient operator to locate edge pixels. A number of gradient operators have been proposed. In many edge detection algorithms, the horizontal gradient image $g_{horizontal}(m,n)$ and the vertical gradient image $g_{vertical}(m,n)$ are first computed using gradient operators. Then, the magnitude gradient image $g(m,n)$ may be computed as follows:

$$g(m,n) = \left|g_{horizontal}(m,n)\right| + \left|g_{vertical}(m,n)\right|$$

Finally, a thresholding operation is applied to the magnitude gradient image to find edge pixels. In other words, pixels whose magnitude gradients exceed a threshold value are considered as edge pixels.

Figures 2-6 illustrate the procedure. Figure 2 shows a source image. Figure 3 shows a horizontal gradient image $g_{horizontal}(m,n)$, which is obtained by applying a horizontal gradient operator to the source image of Fig. 2. Figure 4 shows a vertical gradient image $g_{vertical}(m,n)$, which is obtained by applying a vertical gradient operator to the source image of Fig. 2. Figure 5 shows the magnitude gradient image (edge image) and Fig. 6 shows the binary edge image (mask image) obtained by applying thresholding to the magnitude gradient image of Fig. 5.

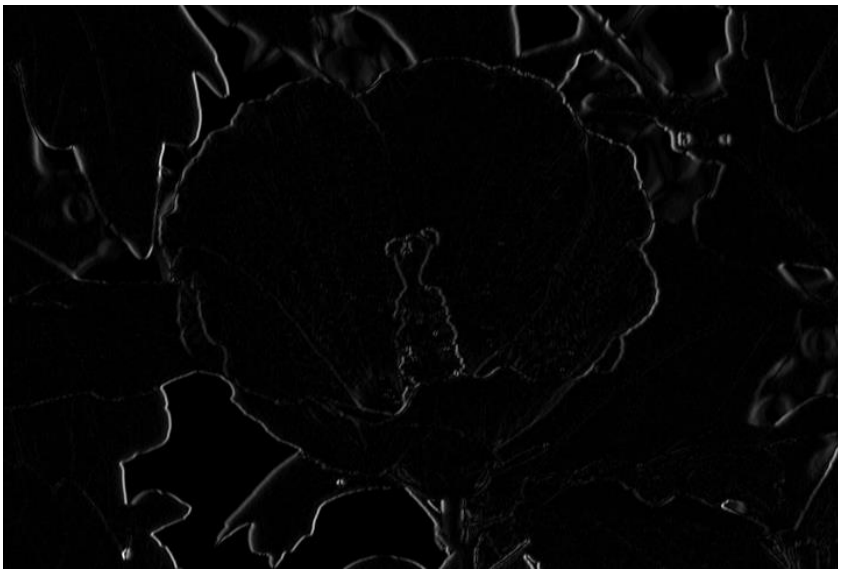
FIGURE 2

A source image (original image)

BT.1908-02

FIGURE 3

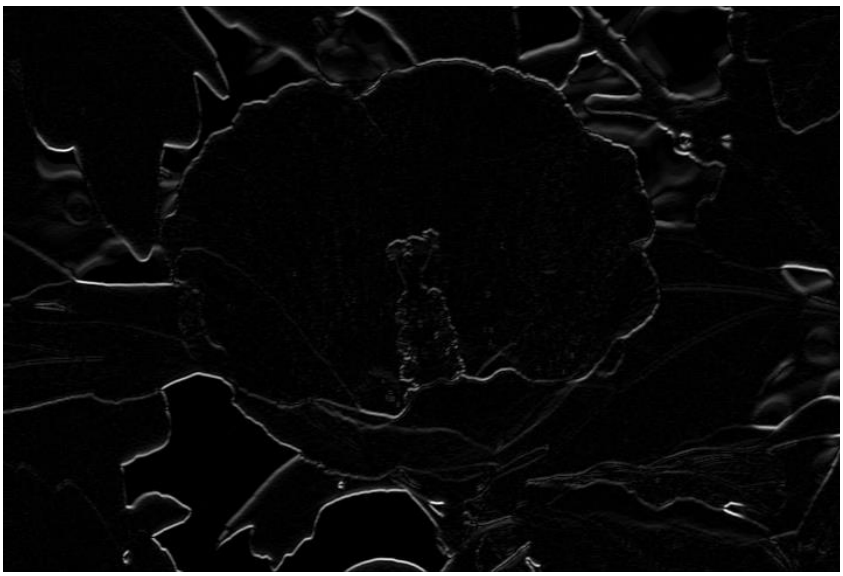
A horizontal gradient image, which is obtained by applying a horizontal gradient operator to the source image of Fig. 2



BT.1908-03

FIGURE 4

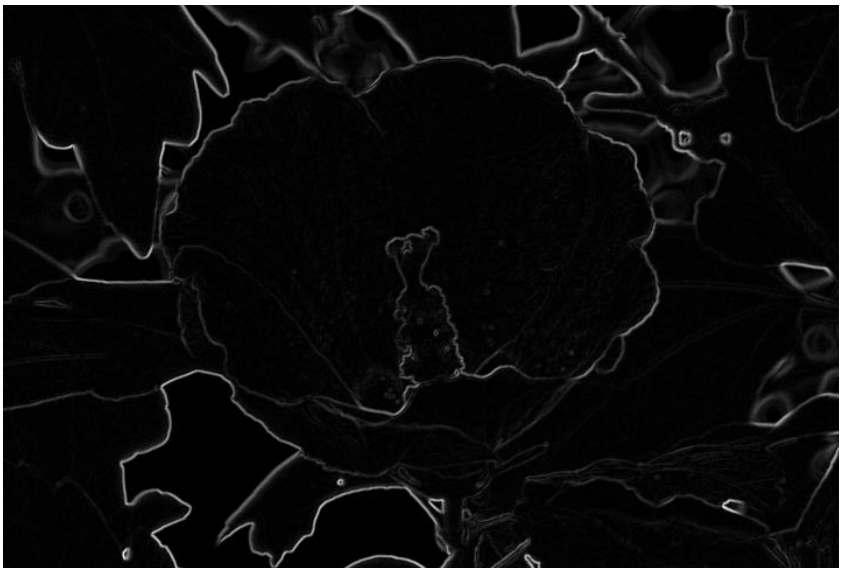
A vertical gradient image, which is obtained by applying a vertical gradient operator to the source image of Fig. 2



BT.1908-04

FIGURE 5

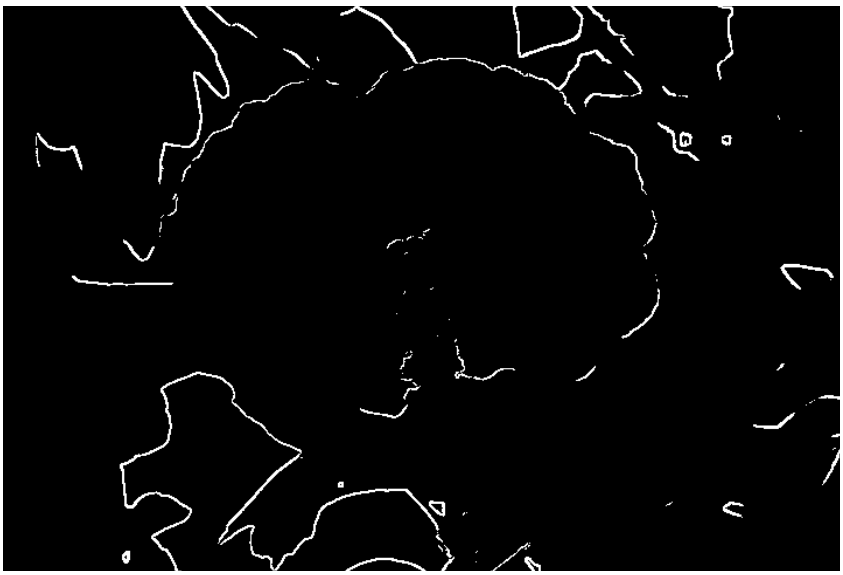
A magnitude gradient image



BT.1908-05

FIGURE 6

A binary edge image (mask image) obtained by applying thresholding to the magnitude gradient image of Fig. 5



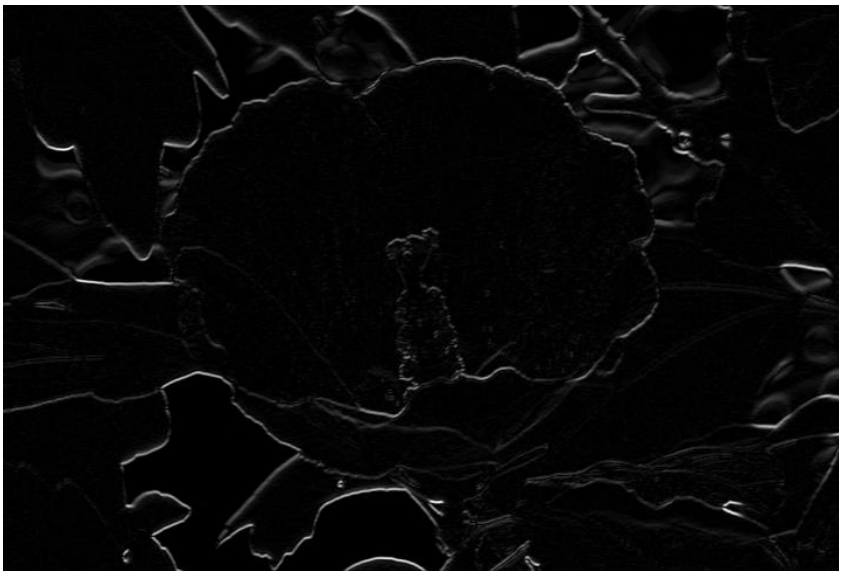
BT.1908-06

Alternatively, one may use a modified procedure to find edge pixels. For instance, one may first apply a vertical gradient operator to the source image, producing a vertical gradient image. Then, a horizontal gradient operator is applied to the vertical gradient image, producing a modified successive gradient image (horizontal and vertical gradient image). Finally, a thresholding operation may be applied to the modified successive gradient image to find edge pixels. In other words, pixels of the modified successive gradient image, which exceed a threshold value, are considered as edge

pixels. Figures 7-9 illustrate the modified procedure. Figure 7 shows a vertical gradient image $g_{vertical}(m,n)$, which is obtained by applying a vertical gradient operator to the source image of Fig. 2. Figure 8 shows a modified successive gradient image (horizontal and vertical gradient image), which is obtained by applying a horizontal gradient operator to the vertical gradient image of Fig. 7. Figure 9 shows the binary edge image (mask image) obtained by applying thresholding to the modified successive gradient image of Fig. 8.

FIGURE 7

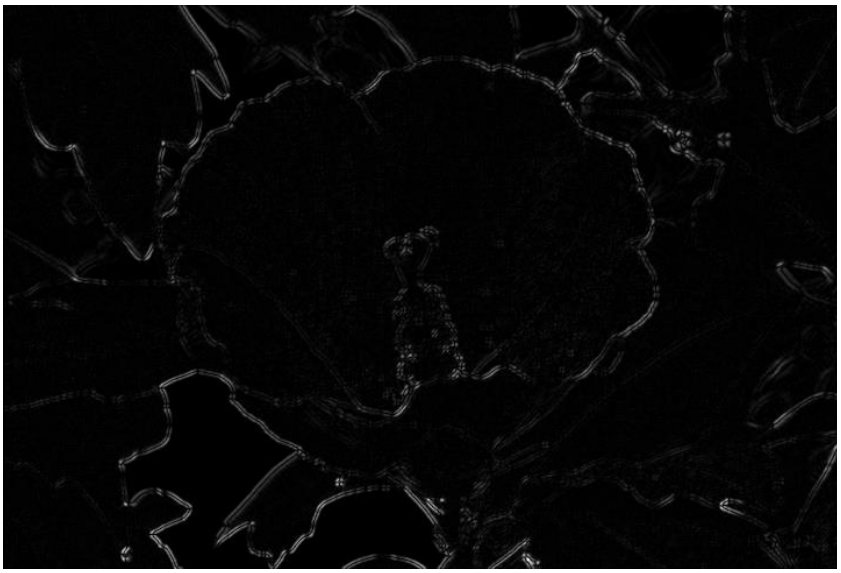
A vertical gradient image, which is obtained by applying a vertical gradient operator to the source image of Fig. 2



BT.1908-07

FIGURE 8

A modified successive gradient image (horizontal and vertical gradient image), which is obtained by applying a horizontal gradient operator to the vertical gradient image of Fig. 7



BT.1908-08

FIGURE 9

A binary edge image (mask image) obtained by applying thresholding to the modified successive gradient image of Fig. 8

BT.1908-09

It is noted that both methods can be understood as edge detection algorithms. One may choose any edge detection algorithm depending on the nature of videos and compression algorithms. However, some methods may outperform other methods.

Thus, in the model, an edge detection operator is first applied, producing edge images (Figs 5 and 8). Then, a mask image (binary edge image) is produced by applying thresholding to the edge image (Figs 6 and 9). In other words, pixels of the edge image whose value is smaller than threshold t_e are set to zero and pixels whose value is equal to or larger than the threshold are set to a non-zero value. Figures 6 and 9 show some mask images. Since a video can be viewed as a sequence of

frames or fields, the above-stated procedure can be applied to each frame or field of videos. Since the model can be used for field-based videos or frame-based videos, the terminology “image” will be used to indicate a field or frame.

6.2.2 Selecting features from source video sequences

Since the model is a reduced-reference (RR) model, a set of features need to be extracted from each image of a source video sequence. In the EPSNR RR model, a certain number of edge pixels are selected from each image. Then, the locations and pixel values are encoded and transmitted. However, for some video sequences, the number of edge pixels can be very small when a fixed threshold value is used. In the worst scenario, it can be zero (blank images or very low frequency images). In order to address this problem, if the number of edge pixels of an image is smaller than a given value, the user may reduce the threshold value until the number of edge pixels is larger than a given value. Alternatively, one can select edge pixels which correspond to the largest values of the horizontal and vertical gradient image. When there are no edge pixels (e.g. blank images) in a frame, one can randomly select the required number of pixels or skip the frame. For instance, if ten edge pixels are to be selected from each frame, one can sort the pixels of the horizontal and vertical gradient image according to their values and select the largest ten values. However, this procedure may produce multiple edge pixels at identical locations. To address this problem, one can first select several times the desired number of pixels of the horizontal and vertical gradient image and then randomly choose the desired number of edge pixels among the selected pixels of the horizontal and vertical gradient image. In the models tested in the VQEG HDTV test, the desired number of edge pixels is randomly selected among a large pool of edge pixels. The pool of edge pixels is obtained by applying a thresholding operation to the gradient image.

In the EPSNR RR models, the locations and edge pixel values are encoded after a Gaussian low pass filter is applied to the selected pixel locations. Although the Gaussian LPF (7 × 3) was used in the VQEG HDTV test, different low pass filters may be used depending on the video formats. It is noted that during the encoding process, cropping may be applied. In order to avoid selecting edge pixels in the cropped areas, the model selects edge pixels in the middle area (Fig. 10). Table 2 shows the sizes after cropping. Table 2 also shows the number of bits required to encode the location and pixel value of an edge pixel.

TABLE 2

Bits requirement per edge pixel

Video format	Size	Size after cropping	Bits for location	Bits for pixel value	Total bits per pixel
HD progressive	1 920 × 1 080	1 856 × 1032	21	8	29
HD interlaced	1 920 × 540	1 856 × 516	20	8	28

FIGURE 10

An example of cropping and the middle area

BT.1908-10

The model selects edge pixels from each frame in accordance with the allowed bandwidth (Table 1). Table 3 shows the number of edge pixels per frame which can be transmitted for the tested bandwidths.

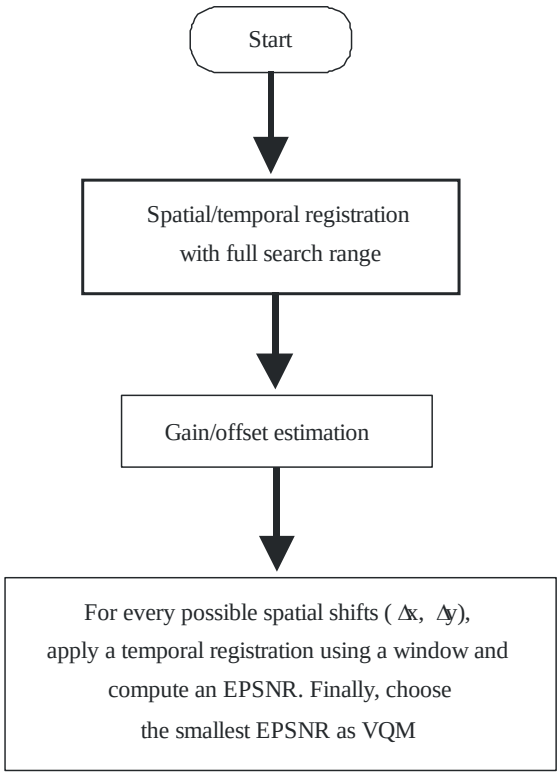
TABLE 3

Number of edge pixels per frame/field

Video format	56 kbit/s	128 kbit/s	256 kbit/s
HD progressive	46	105	211
HD interlaced	24	54	109

FIGURE 11

Flowchart of the model



BT.1908-11

6.2.3 Spatial/temporal registration and gain/offset adjustment

Before computing the difference between the edge pixels of the source video sequence and those of the processed video sequence, which is the received video sequence at the receiver, the model first applies a spatial/temporal registration and gain/offset adjustment. The calibration method (Annex B) of Recommendation ITU-T J.244 was used. To transmit the gain and offset features of Recommendation ITU-T J.244 (Annex B), 30% of the available bandwidths was used in the VQEG HDTV test. When the video sequence is interlaced, the calibration method is applied three times: the even fields, odd fields and combined frames, while the calibration method is applied to frames in progressive video sequences. If the difference between the even field error (PSNR) and the odd field error is greater than a threshold, the registration results (x-shift, y-shift) with the smaller PSNR were used. Otherwise, the registration results with the combined frames were used. In the VQEG HDTV test, the threshold was set to 2 dB.

At the monitoring point, the processed video sequence should be aligned with the edge pixels extracted from the source video sequence. However, if the side-channel bandwidth is small, only a few edge pixels of the source video sequence are available (Fig. 12). Consequently, the temporal registration can be inaccurate if the temporal registration is performed using a single frame (Fig. 13). To address this problem, the model uses a window for temporal registration. Instead of using a single frame of the processed video sequence, the model builds a window which consists of a number of adjacent frames to find the optimal temporal shift. Figure 14 illustrates the procedure. The mean squared error within the window is computed as follows:

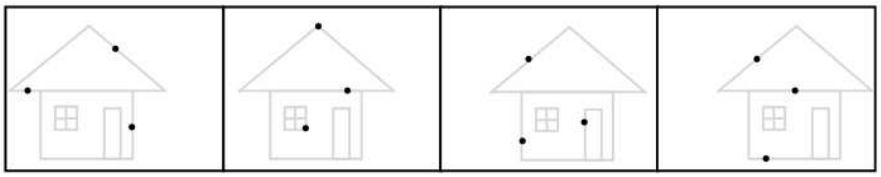
$$MSE_{window} = \frac{1}{N_{win}} \sum (E_{SRC}(i) - E_{PVS}(i))^2$$

where MSE_{window} is the window mean squared error, $E_{SRC}(i)$ is an edge pixel within the window which has a corresponding pixel in the processed video sequence, $E_{PVS}(i)$ is a pixel of the processed video sequence corresponding to the edge pixel, and N_{win} is the total number of edge pixels used to compute MSE_{window}. This window mean squared error is used as the difference between a frame of the processed video sequence and the corresponding frame of the source video sequence.

The window size can be determined by considering the nature of the processed video sequence. For a typical application, a window corresponding to two seconds is recommended. Alternatively, various sizes of windows can be applied and the best one which provides the smallest mean squared error can be used. Furthermore, different window centres can be used to consider frame skipping due to transmission errors (Fig. 18).

FIGURE 12

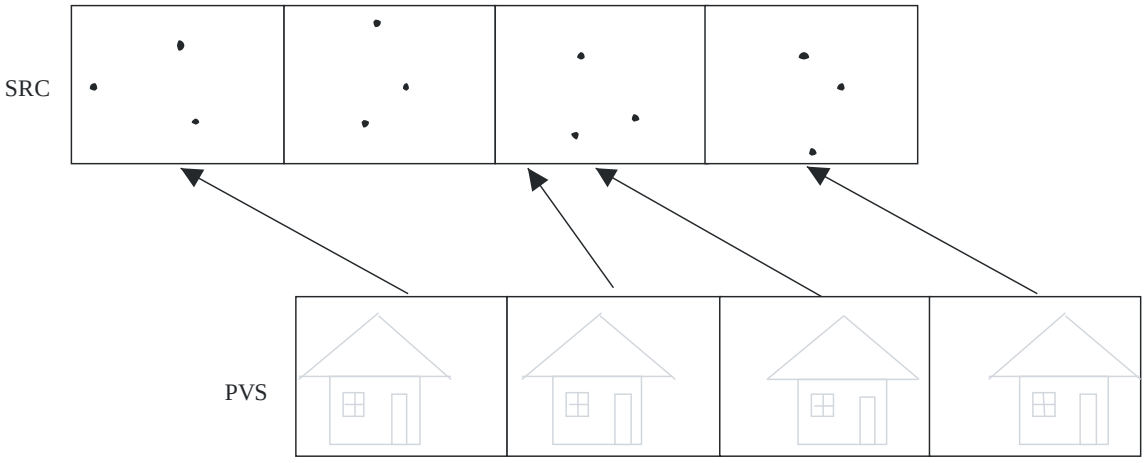
Edge pixel selection of the source video sequence



BT.1908-12

FIGURE 13

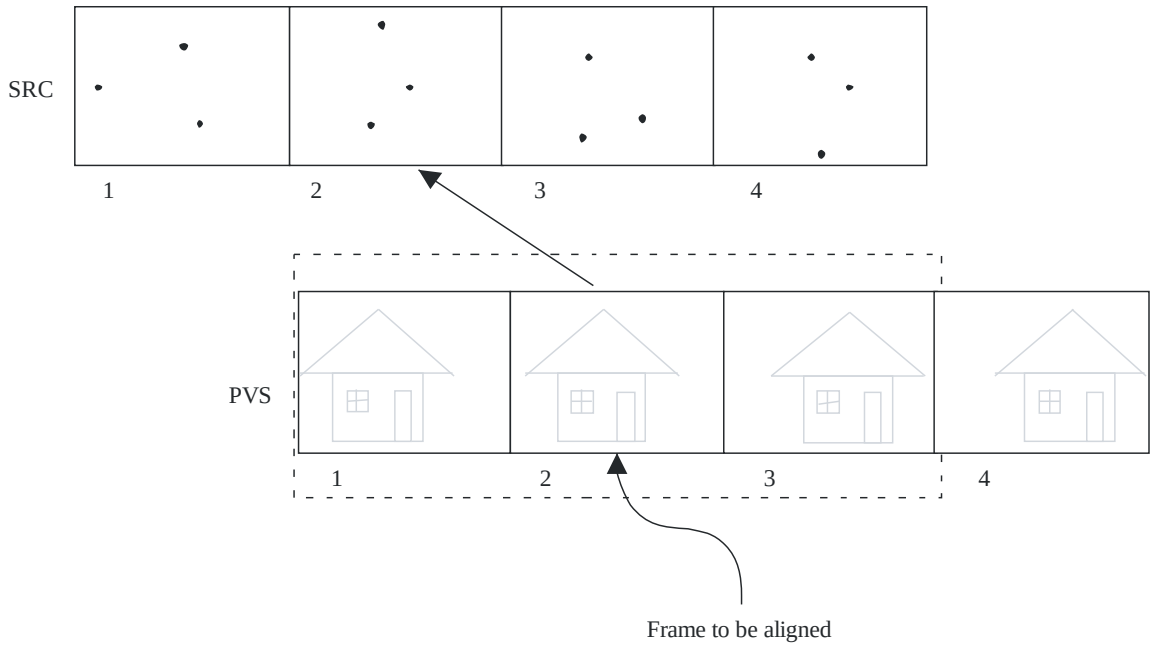
Aligning the processed video sequence to the edge pixels of the source video sequence



BT.1908-13

FIGURE 14

Aligning the processed video sequence to the edge pixels using a window

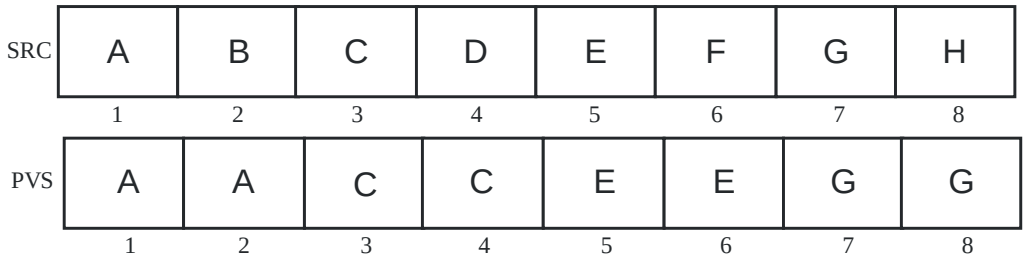


BT.1908-14

When the source video sequence is encoded at high compression ratios, the encoder may reduce the number of frames per second and the processed video sequence has repeated frames (Fig. 15). In Fig. 15, the processed video sequence does not have frames corresponding to some frames of the source video sequence (2, 4, 6, 8th frames). In this case, the model does not use repeated frames in computing the mean squared error. In other words, the model performs temporal registration using the first frame (valid frame) of each repeated block. Thus, in Fig. 16, only three frames (3, 5, 7th frames) within the window are used for temporal registration.

FIGURE 15

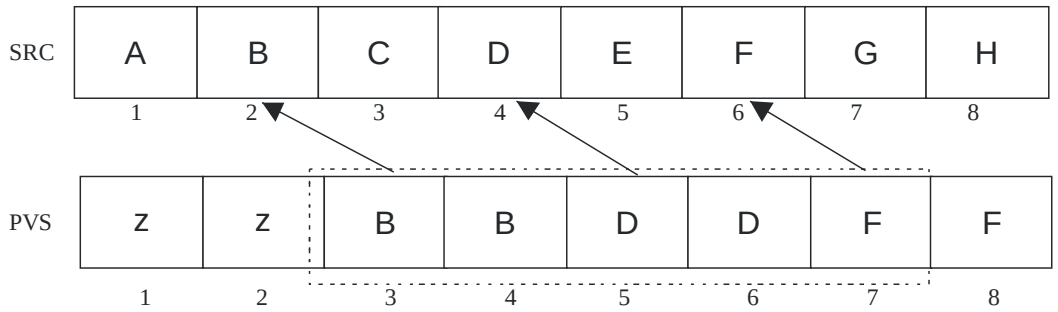
Example of repeated frames



BT.1908-15

FIGURE 16

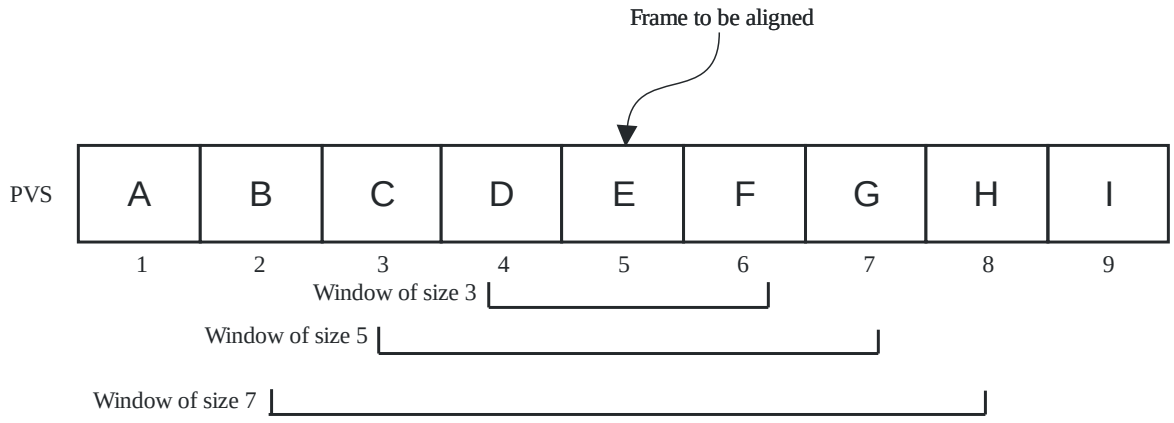
Handing repeated frames



BT.1908-16

FIGURE 17

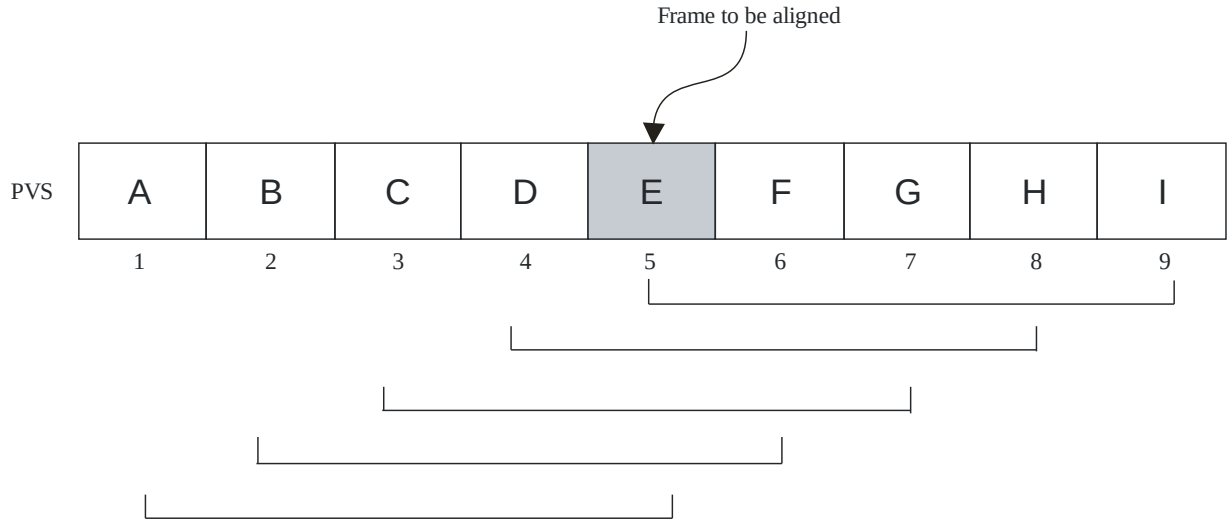
Windows of various sizes



BT.1908-17

FIGURE 18

Window centres



BT.1908-18

6.2.4 Computing EPSNR and post-processing

After temporal registration is performed, the average of the differences between the edge pixels of the source video sequence and the corresponding pixels of the processed video sequence is computed, which can be understood as the edge mean squared error of the processed video sequence (MSE_{edge}). Finally, the EPSNR (edge PSNR) is computed as follows:

$$EPSNR = 10\log_{10}(\frac{P^2}{MSE_{edge}})$$

where p is the peak value of the image.

Since various impairments can reduce video quality, the EPSNR value is adjusted by considering these effects which are quantified in the next sub-subsections.

1) Blocking metric I

To consider blocking effects, average column differences are computed. Assuming modulo 8, the blocking score for the i-th frame is computed as follows:

$$Blk[i] = \frac{\text{largest column difference}}{\text{second largest column difference}}$$

The final blocking score (*Blocking*) is computed by averaging the frame blocking scores.

$$Blocking = \frac{1}{number of\ frames}\sum_i Blk[i]$$

Finally, the following equations are used:

IF(BLOCKING > 12 and 25≤ EPSNR<30) adjust_EPSNR_blk1=3
IF(BLOCKING > 5 and 30≤ EPSNR<35) adjust_EPSNR_blk1=5

2) Blocking metric II

Assuming that blocking impairments may occur in every 8-th column (e.g. in MPEG2), a second blocking metric is also used. To compute the second blocking metric, the absolute horizontal difference is first computed as follows (Fig. 19):

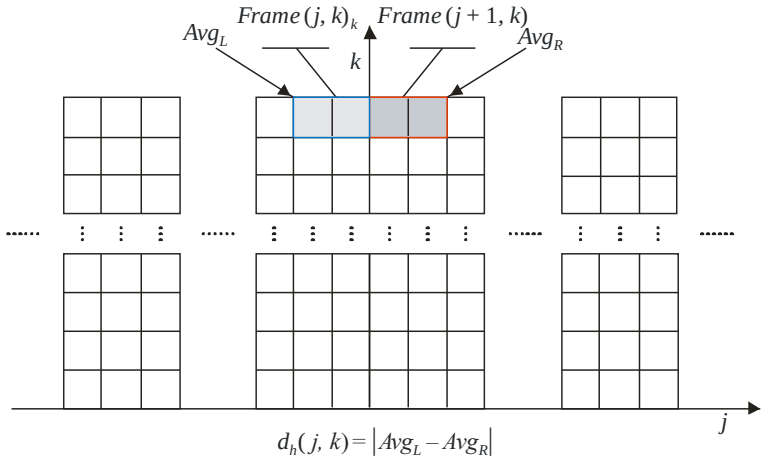
$$d_h(j,k) = |Avg_L - Avg_R|$$

where:

$$Avg_L = \frac{1}{2}\sum_{p=-1}^{0} Frame(j+p,k)\,,\ Avg_R = \frac{1}{2}\sum_{p=1}^{2} Frame(j+p,k)$$

FIGURE 19

Computing the absolute horizontal difference ($d_h(j, k)$)



BT.1908-19

Then, the sum of horizontal blockiness (SB_h) at position j is defined as follows:

$$SB_h[j] = \left(\sum_{1 \le k \le height} \left(|Frame(j,k) - Frame(j+1,k)| \times u\left(d_h(j,k) - \Phi(Avg_L)\right) \right) \right)^2$$

where $u(\cdot)$ represents the unit step function and:

$$\Phi(s) = \begin{cases} 17\left(1 - \sqrt{s/127}\right) + 3 & \text{if} \quad s \le 127 \\ 3(s-127)/128 + 3 & \text{else} \end{cases}$$

After repeating the procedure for the entire frames, the frame horizontal blockiness (FB_h) is computed as follows:

$$FB_h = \left(\sum_{\substack{1 \le j \le width \\ j \equiv 0 (\mathrm{mod} 8)}} SB_h(j) \right)^{1/2}$$

For each frame, the column difference (NFB_h) excluding every 8th column is computed as follows:

$$NFB_h = \frac{1}{7}\sum_{l=1}^{7}\left(\sum_{\substack{1\le j\le width \\ j\equiv l\,(\mathrm{mod}8)}}\left(\sum_{1\le k\le height}\left(\left|Frame(j,k)-Frame(j+1,k)\right|\times u\left(d_h(j,k)-\Phi(Avg_L)\right)\right)\right)^2\right)^{1/2}.$$

Then, the final horizontal blocking feature, BLK_H, is computed as follows:

$$BLK_H = \ln\left(FB_h / NFB_h\right)$$

The vertical blocking feature BLK_V was similarly computed. For interlaced video sequences, the vertical blocking feature is computed in the field sequence. The *i*-th frame blocking score is computed as follows:

$$FrameBLK(i) = 0.5\times BLK_H + 0.5\times BLK_V$$

The final blocking score (*BLOCKING2*) is computed by averaging the upper 10% frame blocking scores.

Finally, the following equations are used:

IF(BLOCKING2 > 1.5 and 25 ≤ EPSNR<30)	adjust_EPSNR_blk2=2
IF(BLOCKING2 > 1.3 and 30 ≤ EPSNR<35)	adjust_EPSNR_blk2=2
IF(BLOCKING2 > 1.5 and 35 ≤ EPSNR<40)	adjust_EPSNR_blk2=2
IF(BLOCKING2 > 1 and 40 ≤ EPSNR<45)	adjust_EPSNR_blk2=2
IF(BLOCKING2 > 0.5 and 45 ≤ EPSNR<55)	adjust_EPSNR_blk2=2

As can be seen in the above equations, this adjustment has minor effects on the final EPSNR value. If blocking artefacts do not occur in every 8th column, one may skip this adjustment or first find the blocking locations. Also, one may use a different function for $\Phi(s)$.

3) Maximum freezed frames and total freezed frames

Transmission errors may cause long freezed frames. To consider long freezed frames, the following equations are used:

IF(MAX_FREEZE ≥ 8 and 25 ≤ EPSNR<30)	adjust_EPSNR_max_freeze=3
IF(MAX_FREEZE ≥ 6 and 30 ≤ EPSNR<35)	adjust_EPSNR_max_freeze=3
IF(MAX_FREEZE ≥ 3 and 35 ≤ EPSNR<40)	adjust_EPSNR_max_freeze=3
IF(MAX_FREEZE ≥ 1.5 and 40 ≤ EPSNR<45)	adjust_EPSNR_max_freeze=2
IF(MAX_FREEZE ≥ 1 and 45 ≤ EPSNR<95)	adjust_EPSNR_max_freeze=2

where MAX_FREEZE is the largest duration of freezed frames. It is noted that if the video sequence is not 10 seconds, different thresholds should be used.

Also, the total freezed frames are considered as follows:

IF(TOTAL_FREEZE ≥ 80 and 25 ≤ EPSNR<30)	adjust_EPSNR_total_freeze=3
IF(TOTAL_FREEZE ≥ 40 and 30 ≤ EPSNR<35)	adjust_EPSNR_total _freeze=4
IF(TOTAL_FREEZE ≥ 10 and 35 ≤ EPSNR<40)	adjust_EPSNR_total _freeze=3.5
IF(TOTAL_FREEZE ≥ 2 and EPSNR ≥ 40)	adjust_EPSNR_total _freeze=1.5

where TOTAL_FREEZE is the total duration of freezed frames. It is noted that if the video sequence is not 10 seconds, different thresholds should be used.

4) Transmission error block

Local freezed blocks may occur due to transmission errors. Also, in static scenes, some blocks are identical to the blocks of the previous frames at the same positions. To consider the local freezed blocks due to transmission errors, the blocks which contain the transmitted edge pixels are classified either as identical blocks (i.e. the blocks are identical to the blocks of the previous frames) or different blocks. Then, two EPSNRs are computed for the identical blocks and the different blocks. If the difference of the two EPSNRs (EPSNR_diff) is large, it indicates that transmission errors might occur. Based on this observation, the EPSNR is adjusted as follows:

IF(8≤ EPSNR_diff≤30 and 25 ≤ EPSNR<30)	adjust_EPSNR_diff= 3
IF(9≤ EPSNR_diff ≤30 and 30 ≤ EPSNR<35)	adjust_EPSNR_diff= 4
IF(10≤ EPSNR _diff ≤30 and 35 ≤ EPSNR<40)	adjust_EPSNR_diff= 6
IF(9≤ EPSNR _diff <10 and 35 ≤ EPSNR<40)	adjust_EPSNR_diff= 2
IF(9≤ EPSNR _diff ≤30 and 40 ≤ EPSNR<45)	adjust_EPSNR_diff= 4

However, if the total number of the identical blocks is smaller than 100, no adjustment is made.

5) Final adjustment of EPSNR

Finally, the EPSNR value is adjusted as follows:

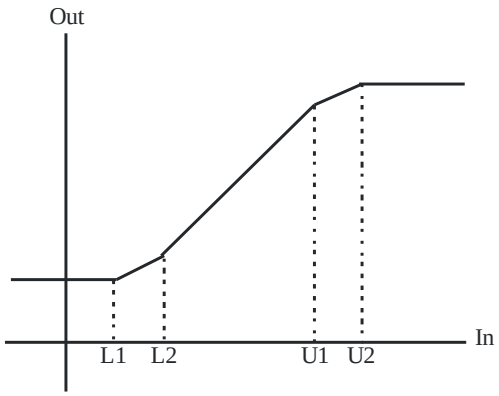
EPSNR <= EPSNR – MAX(adjust_EPSNR_blk1,adjust_EPSNR_blk2,adjust_EPSNR_max_freeze, adjust_EPSNR_total _freeze ,adjust_EPSNR_diff)

6) Piecewise linear fitting

When the EPSNR exceeds a certain value, the perceptual quality becomes saturated. In this case, it is possible to set the upper bound of the EPSNR. Furthermore, when a linear relationship between the EPSNR and DMOS (difference mean opinion score) is desirable, one can apply a piecewise linear function as illustrated in Fig. 20. In the model tested in the VQEG HDTV test, the upper bound was set to 50 and the lower bound to 19.

FIGURE 20

Piecewise linear function for linear relationship between the EPSNR and DMOS



BT.1908-20

Appendix 1
(Informative)

Findings of the Video Quality Experts Group (VQEG)

Studies of perceptual video quality measurements are conducted in an informal group, called the Video Quality Experts Group (VQEG), which reports to ITU-T Study Groups 9 and 12. The recently completed high definition television phase I test of VQEG assessed the performance of proposed full reference perceptual video quality measurement algorithms.

The following statistics are taken from the final VQEG HDTV report (VQEG Report). Note that the body of the VQEG HDTV report includes other metrics including Pearson Correlation and RMSE calculated on individual experiments, confidence intervals, statistical significance testing on individual experiments, analysis on subsets of the data that include specific impairments (e.g. ITU-T H.264 coding-only), scatter plots, and the fit coefficients.

Primary analysis

The performance of the RR model is summarized in Table 4. PSNR is calculated according to Recommendation ITU-T J.340 and included in this analysis for comparison purposes. "Superset RMSE" identifies the primary metric (RMSE) computed on the aggregated superset (i.e. all six experiments mapped onto a single scale). "Top performing group total" identifies the number of experiments (0 to 6) for which this model was either the top performing model or statistically equivalent to the top performing model. "Better than PSNR total" identifies the number of experiments (0 to 6) for which the model was statistically better than PSNR. "Better than superset PSNR" lists whether each model is statistically better than PSNR on the aggregated superset. "Superset correlation" identifies the Pearson correlation computed on the aggregated superset.

TABLE 4

Metric	PSNR	Yonsei56k	Yonsei128k	Yonsei256k
Superset RMSE	0.71	0.73	0.73	0.73
Top performing group total	6	4	4	4
Equivalent to or better than PSNR total	6	4	4	4
Equivalent to superset PSNR	Yes	Yes	Yes	Yes
Superset correlation	0.78	0.77	0.77	0.77

Because the performance of the model is statistically identical for the three bandwidths, it is recommended to use this model with at least a side-channel bandwidth of 56 kbit/s.

Secondary analysis

Table 5 lists the RMSE for the RR model, for subdivisions of the superset. These subdivisions divide the data by coding type (ITU-T H.264 or MPEG-2) as well as by the presence of transmission errors (Errors) or whether the HRC contained coding artefacts only (Coding). Because the experiments were not designed to have these variables evenly span the full range of quality, only RMSE are presented for these subdivisions.

TABLE 5

HRC type	PSNR	Yonsei56k	Yonsei128k	Yonsei256k
H.264 coding	0.75	0.65	0.65	0.65
H.264 error	0.67	0.86	0.85	0.86
MPEG-2 coding	0.78	0.81	0.81	0.80
MPEG-2 error	0.66	0.68	0.68	0.68
Coding	0.75	0.69	0.69	0.69
Error	0.67	0.79	0.78	0.79

Bibliography

Recommendation ITU-T P.910 (2008), *Subjective video quality assessment methods for multimedia applications.*

Recommendation ITU-T P.911 (1998), *Subjective audiovisual quality assessment methods for multimedia applications*.

Recommendation ITU-T J.143 (2000), *User requirements for objective perceptual video quality measurements in digital cable television*.

Recommendation ITU-R BT.500, *Methodology for the subjective assessment of the quality of television pictures*.

Recommendation ITU-T J.340 (2010), *Reference algorithm for computing peak signal to noise ratio of a processed video sequence with compensation for constant spatial shifts, constant temporal shift, and constant luminance gain and offset.*