International Telecommunication Union

ITU-R

Radiocommunication Sector of ITU

Recommendation ITU-R BT.1867

(03/2010)

Objective perceptual visual quality measurement techniques for broadcasting applications using low definition television in the presence of a reduced bandwidth reference

BT Series

Broadcasting service (television)

International Telecommunication Union

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 Annex 1 of 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, 2010

© ITU 2010

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without written permission of ITU.

RECOMMENDATION ITU-R BT.1867

Objective perceptual visual quality measurement techniques for broadcasting applications using low definition television* in the presence of a reduced bandwidth reference**

(2010)

Scope

This Recommendation specifies methods for estimating the perceived video quality of broadcasting applications using low definition television (LDTV) when a reduced reference (RR) signal can be made available, e.g. through an ancillary data channel, watermark, metadata, and so on.

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 full reference;

c) that Recommendation ITU-R BT.1833 describes multimedia systems for broadcasting of multimedia and data applications for mobile reception by handheld receivers;

d) that low definition television (LDTV) is becoming widely used in the broadcasting of multimedia and data applications for mobile reception;

e) that ITU-T Recommendation J.246[1] specifies objective measurement techniques of perceptual video quality applicable to LDTV applications in the presence of a reduced reference;

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

recognizing

a) that the use of LDTV is mainly intended for viewing on small screens, such as those available on handheld and mobile receivers,

* Low definition television (LDTV) refers to video resolutions having less number of pixels than the ones defined in Recommendation ITU-R BT.601. A pertinent ITU-R Recommendation on LDTV is under consideration.

** The measurement method with reduced reference, for objective measurement of perceptual video quality, evaluates the performance of systems by making a comparison between features extracted from the undistorted input, or reference, video signal at the input of the system, and the degraded signal at the output of the system.

[1] ITU-T Recommendation J.246 is available at <http://www.itu.int/rec/T-REC-J.246-200808-P/en>.

recommends

1 that the guidelines, scope, and limitations given in Annex 1 should be used in the application of the objective video quality measurement models identified in *recommends* 2;

2 that the objective perceptual video quality measurement model given in Annex 2 should be used for broadcasting applications using LDTV when a reduced reference signal, as described in Annex 2, is available.

Annex 1

1 Introduction

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

The reduced reference measurement method can be used when the features extracted from the reference video signal is readily available at the measurement point, as may be the case of measurements on individual equipment or a chain in the laboratory or in a closed environment. The estimation methods are based on processing video in VGA, CIF, and QCIF resolution.

The validation test material contained both multiple coding degradations and various transmission error conditions (e.g. bit errors, dropped packets). In the case where coding distortions are considered in the video signals, the encoder can utilize various compression methods (e.g. MPEG-2, H.264, etc.). The models 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 (e.g. frame repeats, frame skips, frame rate reduction). The models 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 ITU-T Recommendation J.149) to determine the model's accuracy for that particular context. This Recommendation is deemed appropriate for services delivered at 4 Mbit/s or less presented on mobile receivers. The following conditions were allowed in the validation test for each resolution:

– QCIF (quarter common intermediate format (176 × 144 pixels)): 16 kbit/s to 320 kbit/s.

– CIF (common intermediate format (352 × 288 pixels)): 64 kbit/s – 2 Mbit/s.

– VGA (video graphics array (640 × 480 pixels)): 128 kbit/s – 6 Mbit/s.

TABLE 1

Factors used in the evaluation of models

Test factors
Transmission errors with packet loss
Video resolution QCIF, CIF and VGA
Video bitrates – QCIF: 16 kbit/s to 320 kbit/s – CIF: 64 kbit/s – 2 Mbit/s – VGA: 128 kbit/s – 4 Mbit/s
Temporal errors (pausing with skipping) of maximum 2 s

TABLE 1 (*end*)

Test factors
Video frame rates from 5 fps to 30 fps
Coding schemes
H.264/AVC (MPEG-4 Part 10), MPEG-4 Part 2, and three other proprietary coding schemes. (See Note 1.)
Applications
Real-time, in-service quality monitoring at the source
Remote destination quality monitoring when side-channels are available for features extracted from source video sequences
Quality measurement for monitoring of a storage or transmission system that utilizes video compression and decompression techniques, either a single pass or a concatenation of such techniques
Lab testing of video systems

NOTE 1 – The validation testing of models included video sequences encoded using 15 different video codecs. The five codecs listed in Table 1 were most commonly applied to encode test sequences and any recommended models may be considered appropriate for evaluating these codecs. In addition to these five codecs a smaller proportion of test sequences were created using the following codecs: H.261, H.263, H.263+[2], JPEG-2000, MPEG-1, MPEG-2, H.264 SVC, and other proprietary systems. It can be noted that some of these codecs were used only for CIF and QCIF resolutions because they are expected to be used in the field mostly for these resolutions.

Before applying a model to sequences encoded using one of these codecs the user should carefully examine its predictive performance to determine whether the model reaches acceptable predictive performance.

2 Application

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

1 codec evaluation, specification, and acceptance testing, consistent with the limited accuracy as described below;

2 real-time, in-service quality monitoring;

3 remote destination quality monitoring when side channels are available for features extracted from source video sequences;

4 quality measurement for monitoring of a storage or transmission system that utilizes video compression and decompression techniques, either a single pass or a concatenation of such techniques;

5 lab testing of video systems.

3 Limitations

The estimation models 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. If this Recommendation is

[2] H.263+ is a particular configuration of H.263 (1998).

utilized to make video system comparisons (e.g. comparing two codecs), it is advisable to use a quantitative method (such as that in ITU-T Recommendation J.149) to determine the model's accuracy for that particular context.

The models in this Recommendation were validated by measuring video that exhibits frame freezes up to 2 s.

The models in this Recommendation were not validated for measuring video that has a steadily increasing delay (e.g. video which does not discard missing frames after a frame freeze).

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

4 Model descriptions

The following models are described in Annex 2:

Model A (Annex 2) − VQEG Proponent Yonsei University, Korea (Republic of).

Appendix 1
to Annex 1

Findings of the Video Quality Experts Group (VQEG)

Studies of perceptual video quality measurements are conducted in an informal group, called VQEG, which reports to ITU-T Study Groups 9 and 12 and Radiocommunication Study Group 6. The recently completed Multimedia Phase I test of VQEG assessed the performance of proposed reduced reference perceptual video quality measurement algorithms for QCIF, CIF, and VGA formats.

Based on present evidence, the following method can be recommended by ITU-R at this time:

Model A (Annex 2) − VQEG Proponent Yonsei University, Korea (Republic of).

Tables 2, 3 and 4 provide informative details on the model's performances in the VQEG Multimedia Phase I test.

TABLE 2

VGA resolution: Informative description on the model's performances in the VQEG Multimedia Phase I test: Averages over 13 subjective tests

Statistic	Yonsei RR10k	Yonsei RR64k	Yonsei RR128k	PSNR[1]
Correlation	0.803	0.803	0.803	0.713
RMSE[2]	0.599	0.599	0.598	0.714
Outlier ratio	0.556	0.553	0.552	0.615

[1] PSNR: peak signal-to-noise ratio.

[2] RMSE: root mean square error.

TABLE 3

CIF resolution: Informative description on the model's performances in the VQEG Multimedia Phase I test: Averages over 14 subjective tests

Statistic	Yonsei RR10k	Yonsei RR64k	PSNR
Correlation	0.780	0.782	0.656
RMSE	0.593	0.590	0.720
Outlier ratio	0.519	0.511	0.632

TABLE 4

QCIF resolution: Informative description on the model's performances in the VQEG Multimedia Phase I test: Averages over 14 subjective tests

Statistic	Yonsei RR1k	Yonsei RR10k	PSNR
Correlation	0.771	0.791	0. 662
RMSE	0.604	0.578	0.721
Outlier ratio	0.505	0.486	0.596

The average correlations of the primary analysis for the RR VGA models were all 0.80, and PSNR was 0.71. Individual model correlations for some experiments were as high as 0.93. The average RMSE for the RR VGA models were all 0.60, and PSNR was 0.71. The average outlier ratio for the RR VGA models ranged from 0.55 to 0.56, and PSNR was 0.62. All proposed models performed statistically better than PSNR for 7 of the 13 experiments. Based on each metric, each RR VGA model was in the group of top performing models the following number of times:

Statistic	Yonsei RR10k	Yonsei RR64k	Yonsei RR128k	PSNR
Correlation	13	13	13	7
RMSE	13	13	13	6
Outlier ratio	13	13	13	10

The average correlations of the primary analysis for the RR CIF models were 0.78, and PSNR was 0.66. Individual model correlations for some experiments were as high as 0.90. The average RMSE for the RR CIF models were all 0.59, and PSNR was 0.72. The average outlier ratio for the RR CIF models were 0.51 and 0.52, and PSNR was 0.63. All proposed models performed statistically better than PSNR for 10 of the 14 experiments. Based on each metric, each RR CIF model was in the group of top performing models the following number of times:

Statistic	Yonsei RR 10k	Yonsei RR64k	PSNR
Correlation	14	14	5
RMSE	14	14	4
Outlier ratio	14	14	5

The average correlations of the primary analysis for the RR QCIF models were 0.77 and 0.79, and PSNR was 0.66. Individual model correlations for some experiments were as high as 0.89.

The average RMSE for the RR QCIF models were 0.58 and 0.60, and PSNR was 0.72. The average outlier ratio for the RR QCIF models were 0.49 and 0.51, and PSNR was 0.60. All proposed models performed statistically better than PSNR for at least 9 of the 14 experiments. Based on each metric, each RR QCIF model was in the group of top performing models the following number of times:

Statistic	Yonsei RR1k	Yonsei RR10k	PSNR
Correlation	14	14	5
RMSE	14	14	4
Outlier ratio	12	13	4

Annex 2

Model A reduced reference methods[***]

TABLE OF CONTENTS

Page

[***] This model is identical to that specified in Annex A of ITU-T Recommendation J.246.

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, 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 5 shows the side-channel bandwidths for the features, which have been tested in the VQEG MM test.

FIGURE 1

Block diagram of reduced reference model

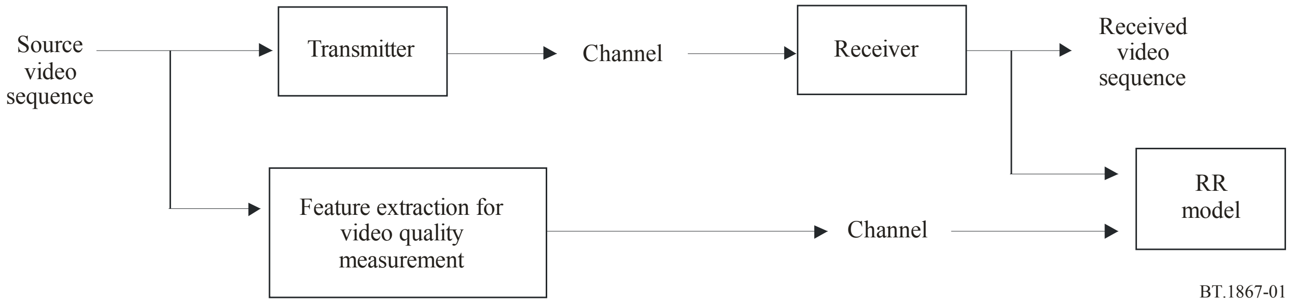


TABLE 5

Side-channel bandwidths

Video format	Tested bandwidths
QCIF	1 kbps, 10 kbps
CIF	10 kbps, 64 kbps
VGA	10 kbps, 64 kbps, 128 kbps

2 The EPSNR reduced-reference models

2.1 Edge PSNR

The reduced-reference (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 (EPSNR) 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:

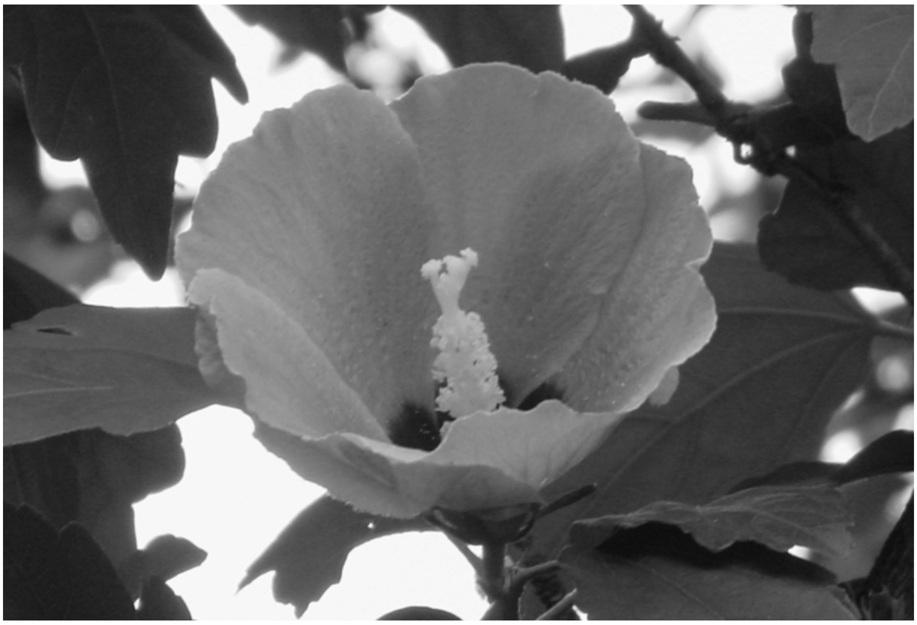
$$g(m,n) = |g_{horizontal}(m,n)| + |g_{vertical}(m,n)|$$

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

Figures 2 to 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

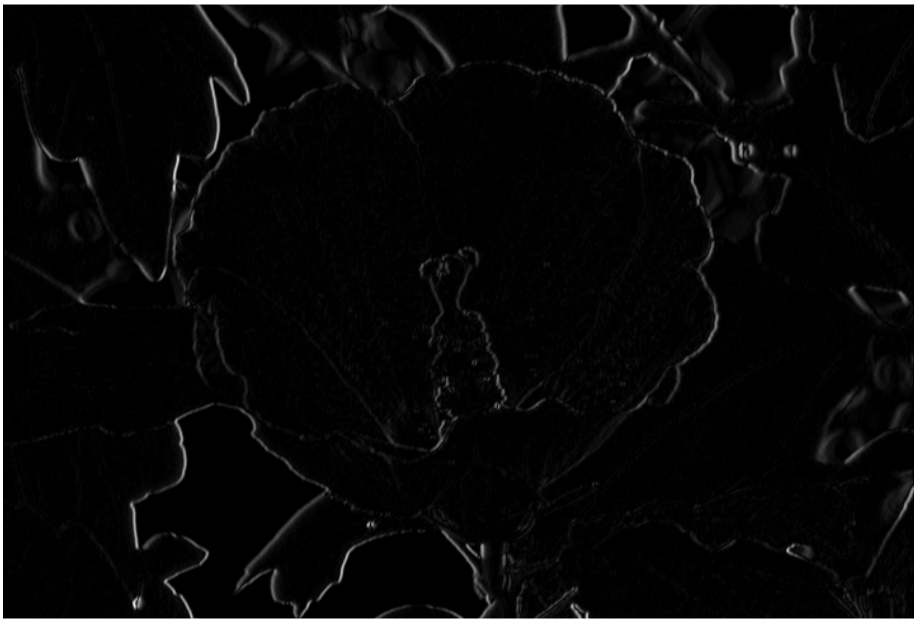
A source image (original image)



BT.1867-02

FIGURE 3

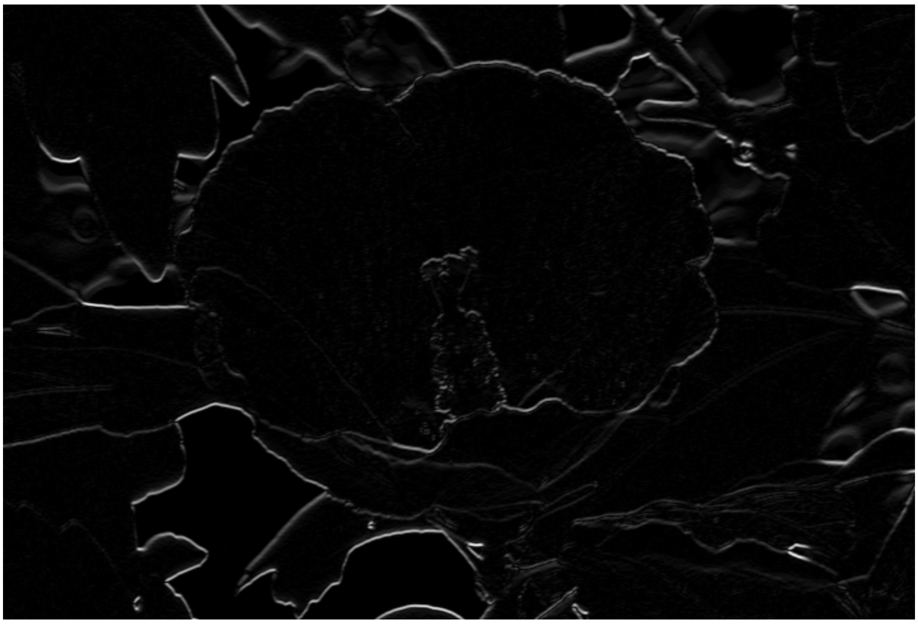
A horizontal gradient image, which is obtained by applying a horizontal gradient operator to the source image of Fig. 2



BT.1867-03

FIGURE 4

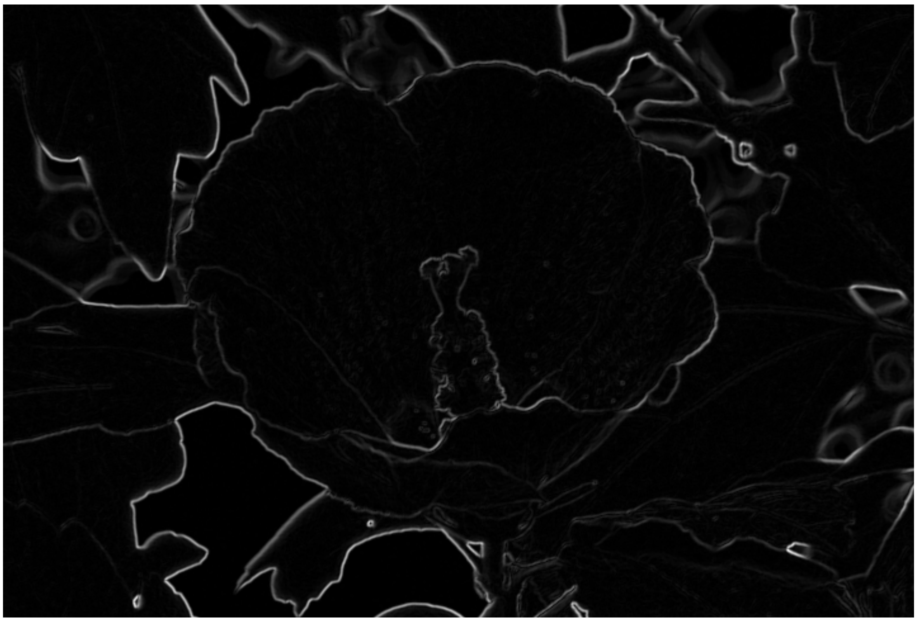
A vertical gradient image, which is obtained by applying a vertical gradient operator to the source image of Fig. 2



BT.1867-04

FIGURE 5

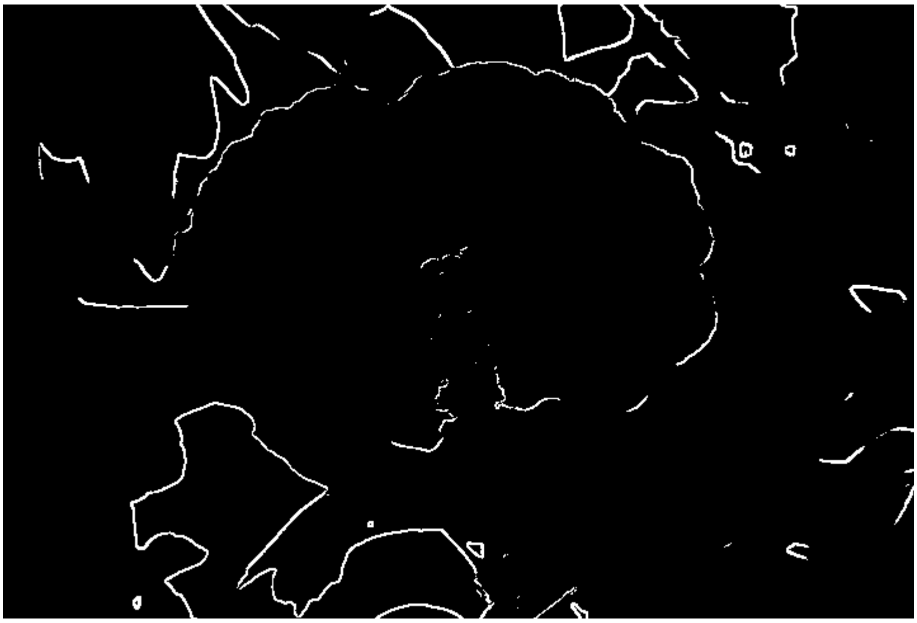
A magnitude gradient image



BT.1867-05

FIGURE 6

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



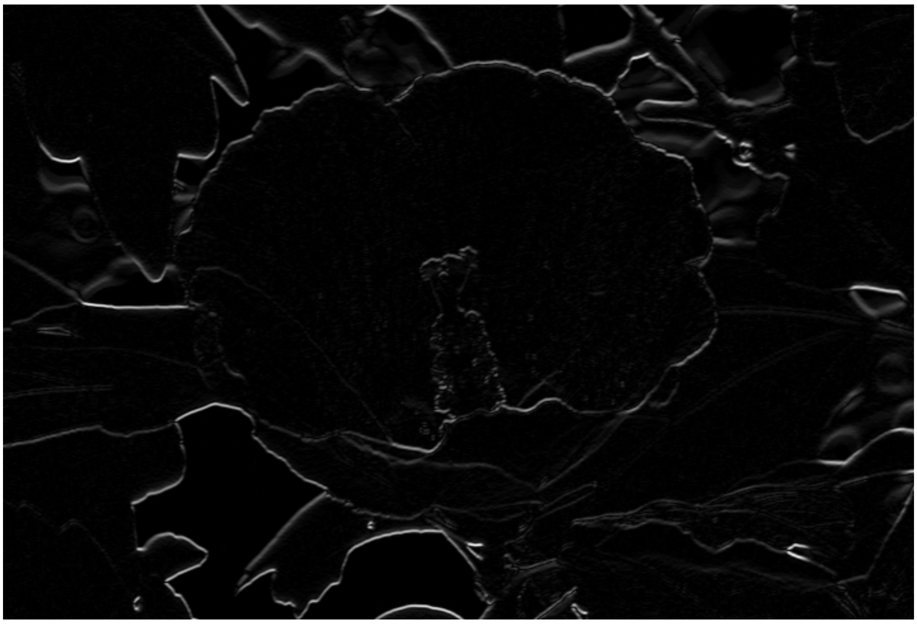
BT.1867-06

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 to 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

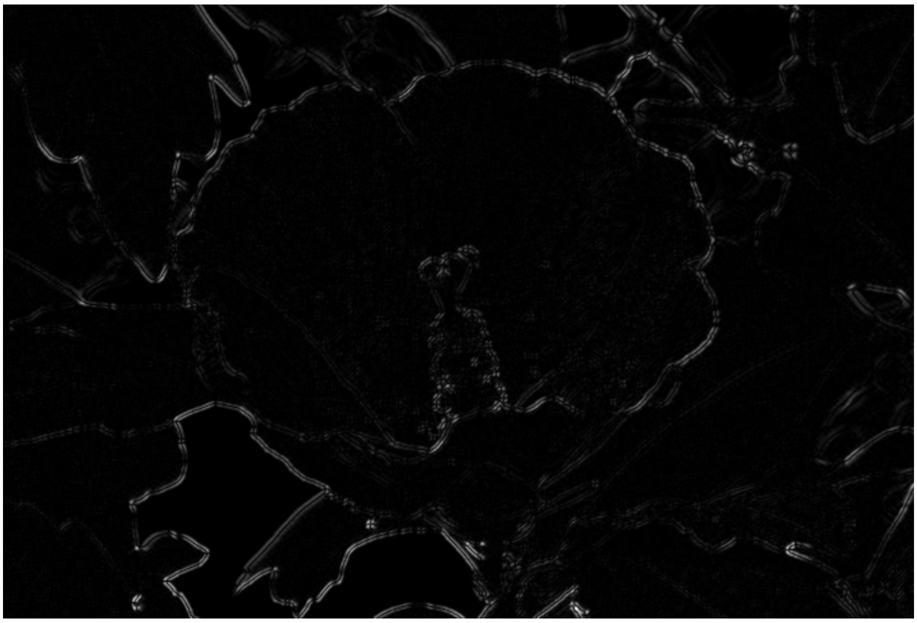
A vertical gradient image, which is obtained by applying a vertical gradient operator to the source image of Fig. 2



BT.1867-07

FIGURE 8

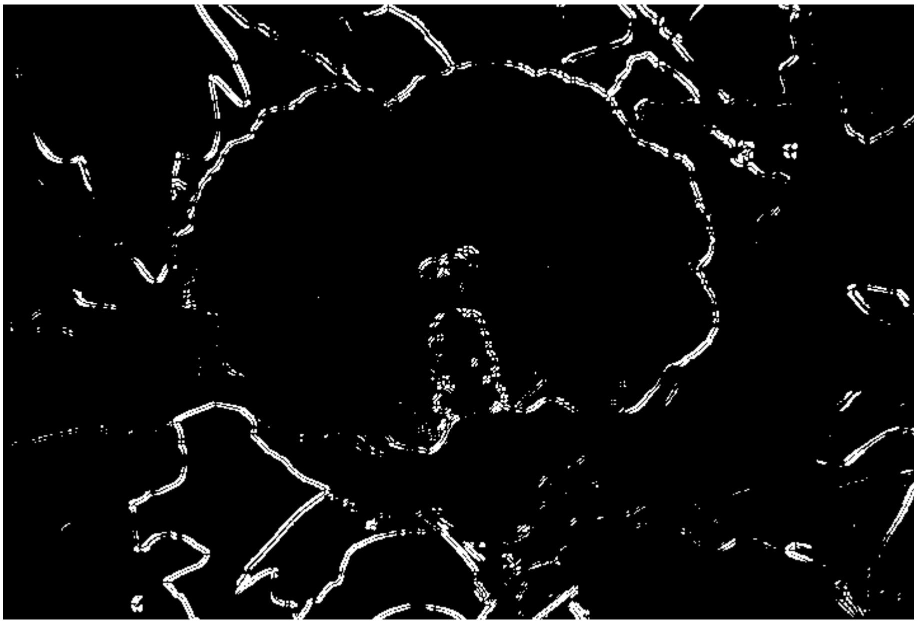
A modified successive gradient image (horizontal and vertical gradient image), which is obtained by applying a horizontal gradient operator to the vertical gradient image of Fig. 7



BT.1867-08

FIGURE 9

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



BT.1867-09

It is noted that both methods can be understood as an edge detection algorithm. 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.

2.2 Selecting features from source video sequences

Since the model is a 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 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 10 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 10 values. However, this procedure may produce multiple edge pixels at the 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 multimedia 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. It is noted that during 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 6 shows the sizes after cropping. Table 6 also shows the number of bits required to encode the location and pixel value of an edge pixel.

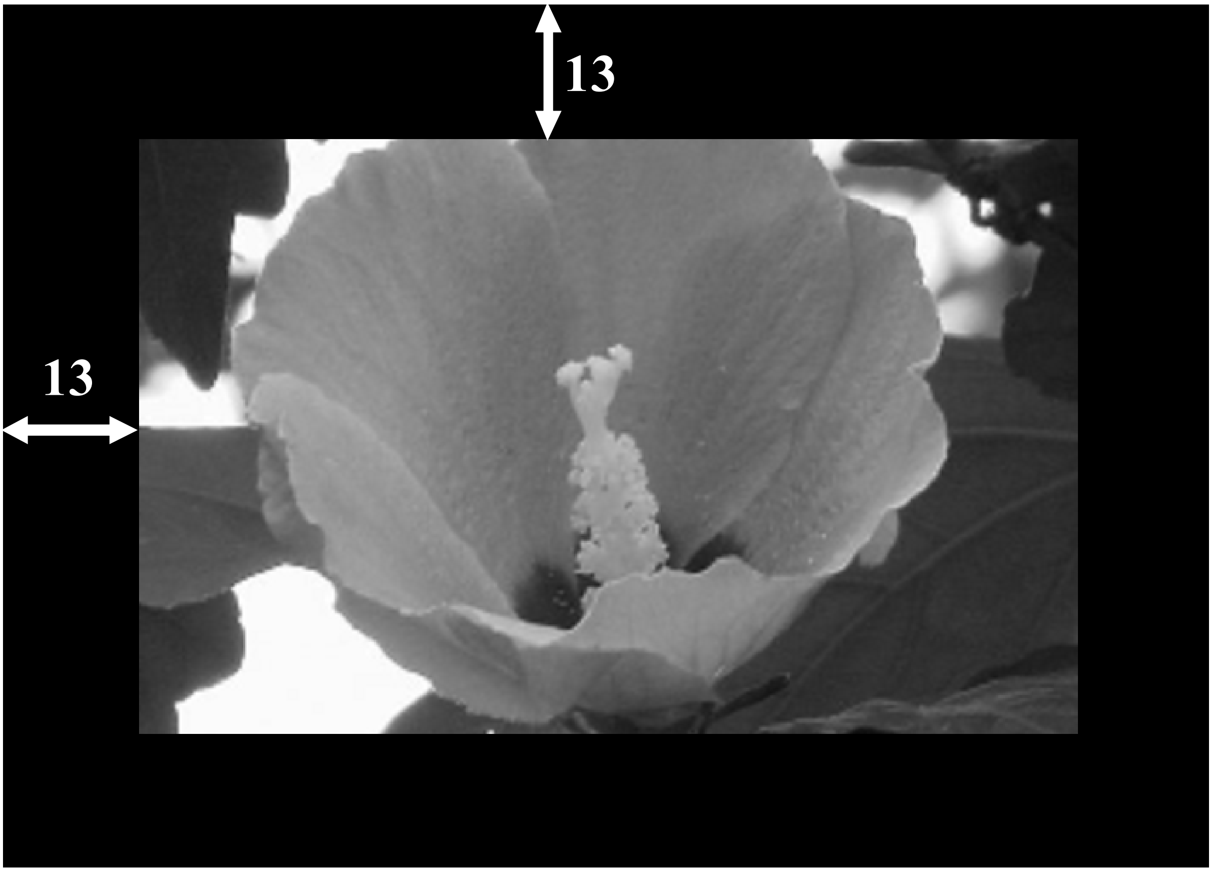
TABLE 6

Bits requirement per edge pixel

Video format	Size	Size after cropping	Bits for location	Bits for pixel value	Total bit per pixel
QCIF	176 × 144	168 × 136	15	8	23
CIF	352 × 288	338 × 274	17	8	25
VGA	640 × 480	614 × 454	19	8	27

FIGURE 10

An example of cropping (VGA) and the middle area



BT.1867-10

The model selects edge pixels from each frame in accordance with the allowed bandwidth (Table 5). Tables 7 to 8 show the number of edge pixels per frame which can be transmitted for the tested bandwidths.

TABLE 7

Number of edge pixels per frame (30 frames/s)

Video format	1 kbit/s	10 kbit/s	64 kbit/s	128 kbit/s
QCIF	1	14		
CIF		13	85	
VGA		12	79	158

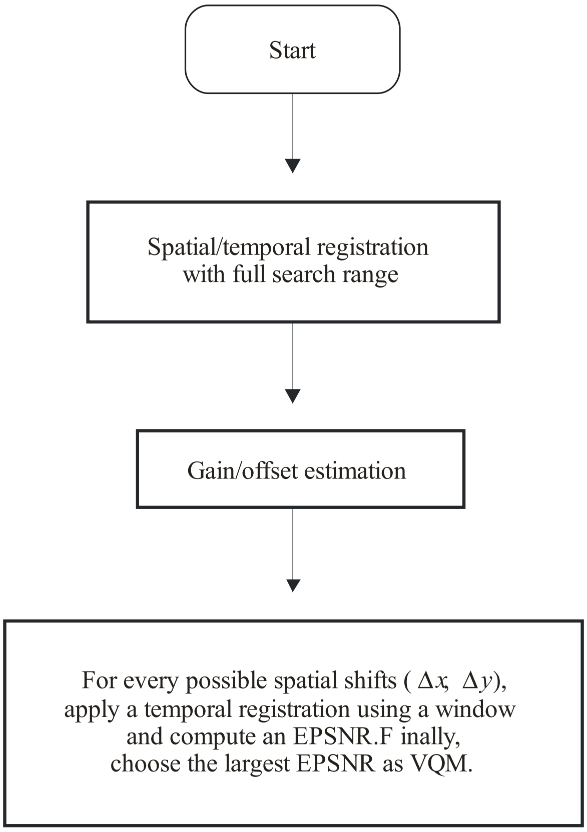
TABLE 8

Number of edge pixels per frame (25 frames/s)

Video format	1 kbit/s	10 kbit/s	64 kbit/s	128 kbit/s
QCIF	1	17		
CIF		16	102	
VGA		14	94	189

FIGURE 11

Flowchart of the model


Start
Spatial/temporal registration with full search range
Gain/offset estimation
For every possible spatial shifts (Δx, Δy), apply a temporal registration using a window and compute an EPSNR.F inally, choose the largest EPSNR as VQM.

BT.1867-11

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. First, a full search algorithm is applied to find global spatial and temporal shifts along with gain and offset values (Fig. 11). Then, for every possible spatial shifts (Δx,Δy), a temporal registration is performed and the EPSNR is computed. Finally the largest EPSNR is chosen as a video quality metric (VQM).

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:

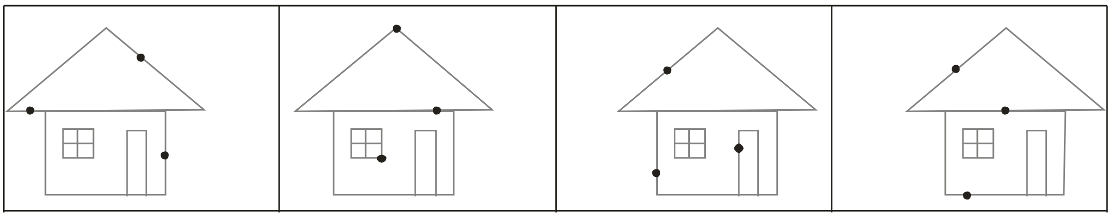
$$MSE_{window} = \frac{1}{N_{win}} \sum \left(E_{SRC}(i) - E_{PVS}(i) \right)^2$$

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 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.

FIGURE 12

Edge pixel selection of the source video sequence



BT.1867-12

FIGURE 13

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

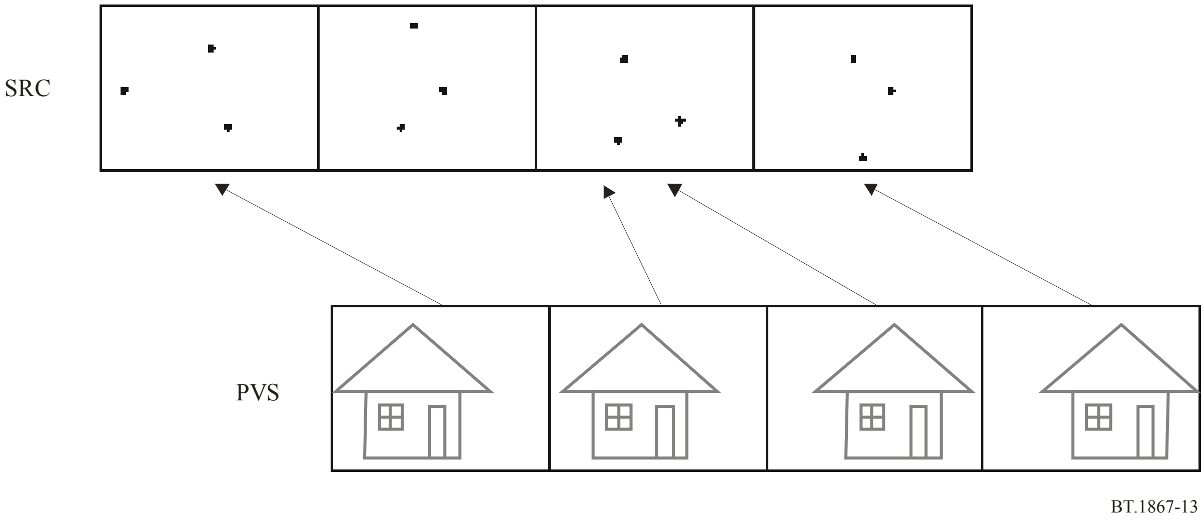
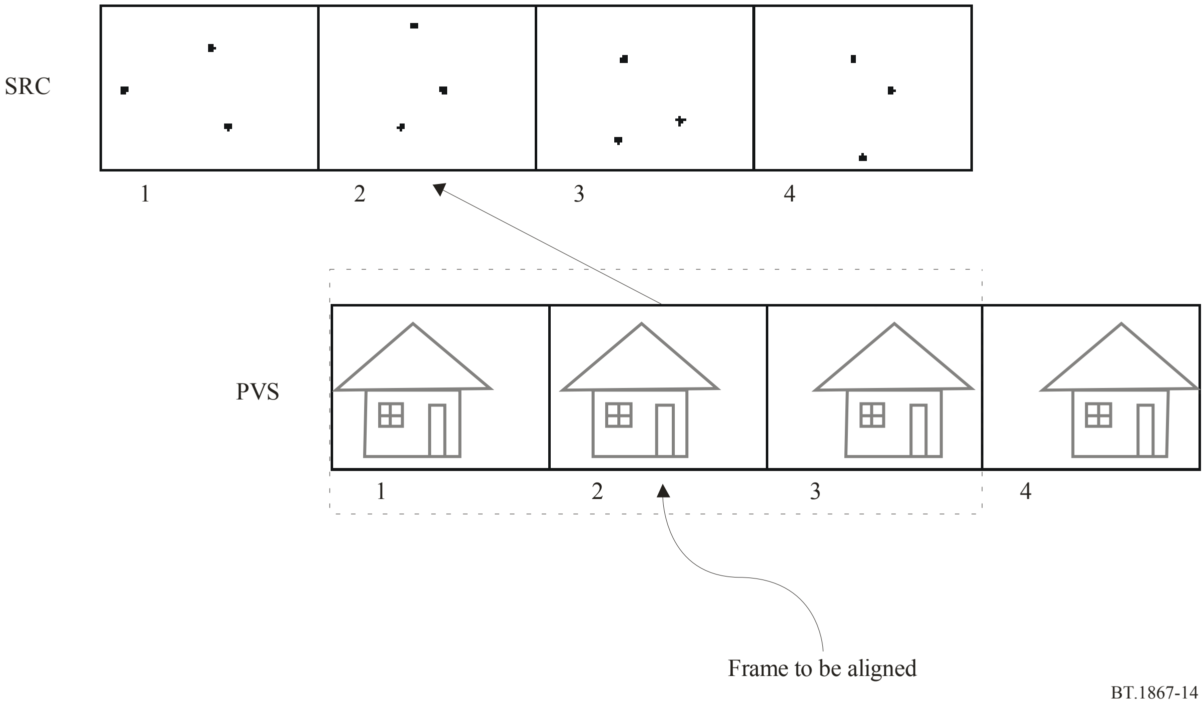


FIGURE 14

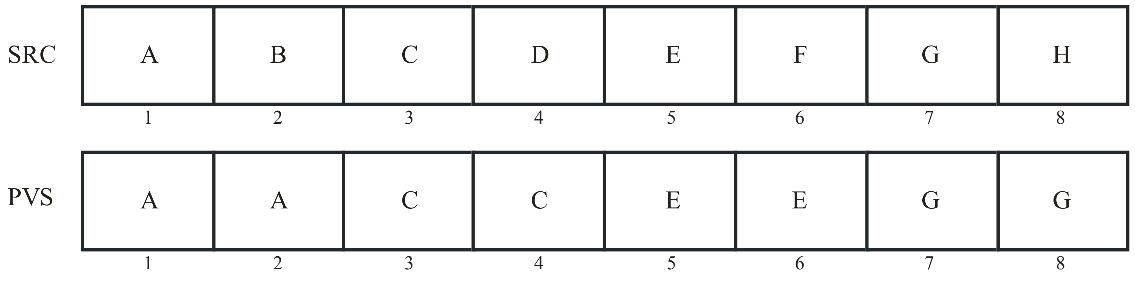
Aligning the processed video sequence to the edge pixels using a window



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 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

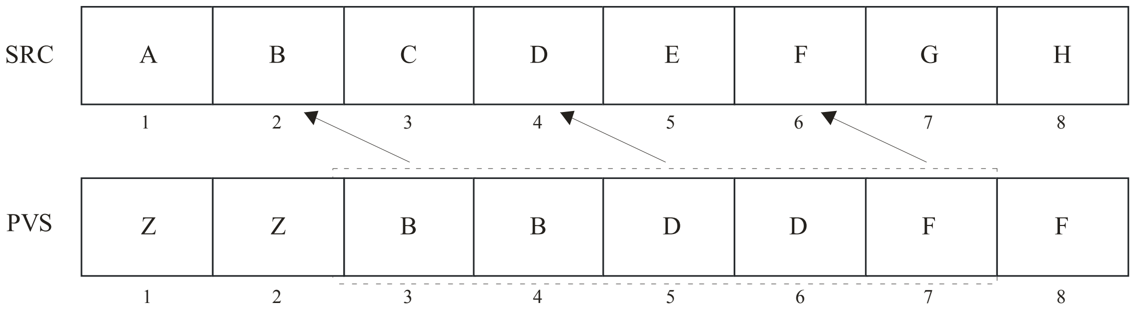
Example of repeated frames



BT.1867-15

FIGURE 16

Handling repeated frames

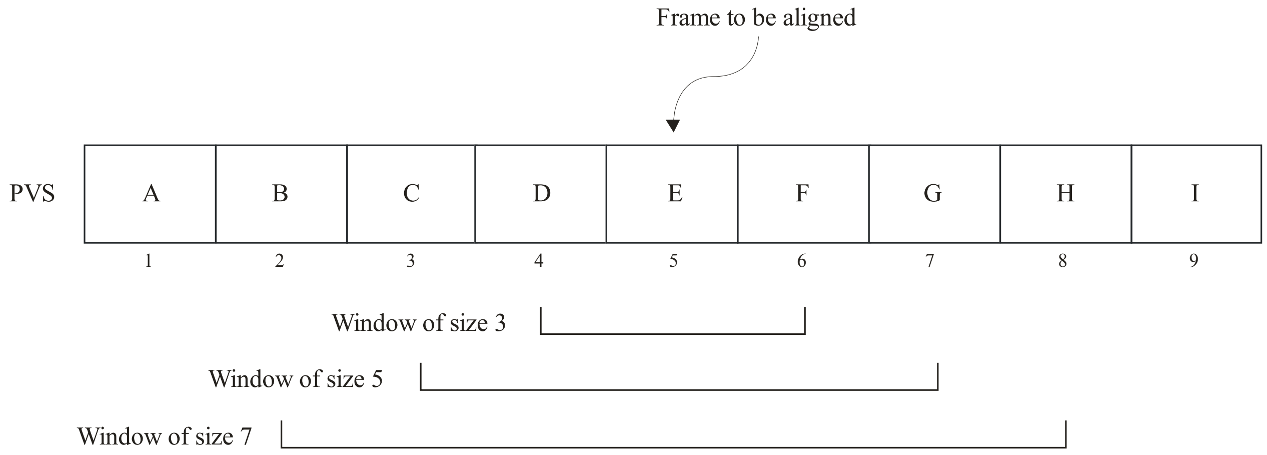


BT.1867-16

It is possible to have a processed video sequence with irregular frame repetition, which may cause the temporal registration method using a window to produce inaccurate results. To address this problem, it is possible to locally adjust each frame of the window within a given value (e.g. ±1), as shown in Fig. 18 after the temporal registration using a window. Then, the local adjustment which provides the minimum MSE is used to compute the EPSNR.

FIGURE 17

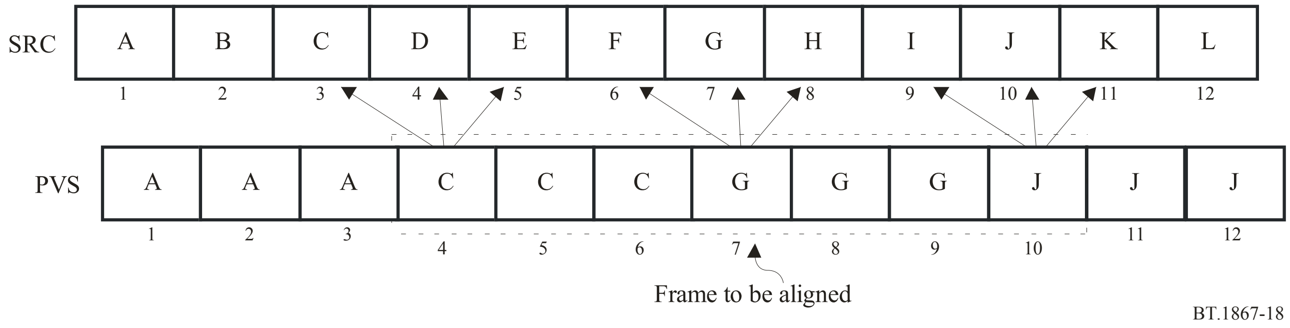
Windows of various sizes



BT.1867-17

FIGURE 18

Local adjustment for temporal registration using a window



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:

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

where:

p is the peak value of the image.

In multimedia video encoding, there can be frame repeating due to reduced frame rates and frame freezing due to transmission error, which will degrade perceptual video quality. In order to address this effect, the model applies the following adjustment before computing the EPSNR:

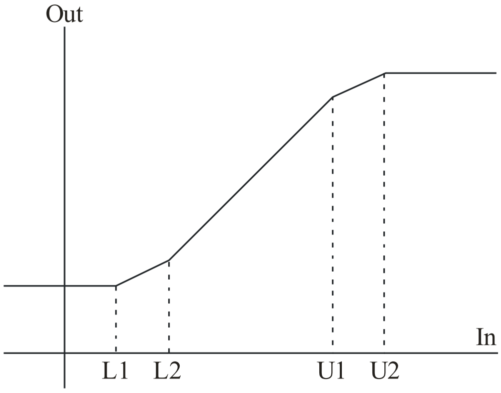
$$MSE_{freezed\_frame\_considered} = MSE_{edge} \times \frac{\mathrm{K} \times N_{total\_frame}}{N_{total\_frame} - N_{total\_freezed\_frame}}$$

where $MSE_{freezed\_frame\_considered}$ is the mean squared error which takes into account repeated and freezed frames, N_{total_frame} is the total number of frames, $N_{total\_freezed\_frame}$, K is a constant. In the model tested in the VQEG multimedia test, K was set to 1.

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. 19. In the model tested in the VQEG multimedia test, only the upper bound is set to 50 since polynomial curve fitting was used.

FIGURE 19

Piecewise linear function for linear relationship between the EPSNR and DMOS



BT.1867-19

2.5 Optimal bandwidth of side channel

The Appendix shows the performance comparison as the bandwidth of the side-channel increases. For the QCIF format, it is observed that the correlation coefficients are almost saturated at about 10 kbit/s. After that, increasing the bandwidth produces about 1% improvement. For the CIF format, it is observed that the correlation coefficients are almost saturated at about 15 kbit/s. After that, increasing the bandwidth produces about 0.5% improvement. For the VGA format, it is observed that the correlation coefficients are almost saturated at about 30 kbit/s. After that, increasing the bandwidth produces about 0.5% improvement.

The EPSNR reduced reference models for objective measurement of video quality are based on edge degradation. The models can be implemented in real time with moderate use of computing power. The models are well suited to applications which require real-time video quality monitoring where side channels are available.

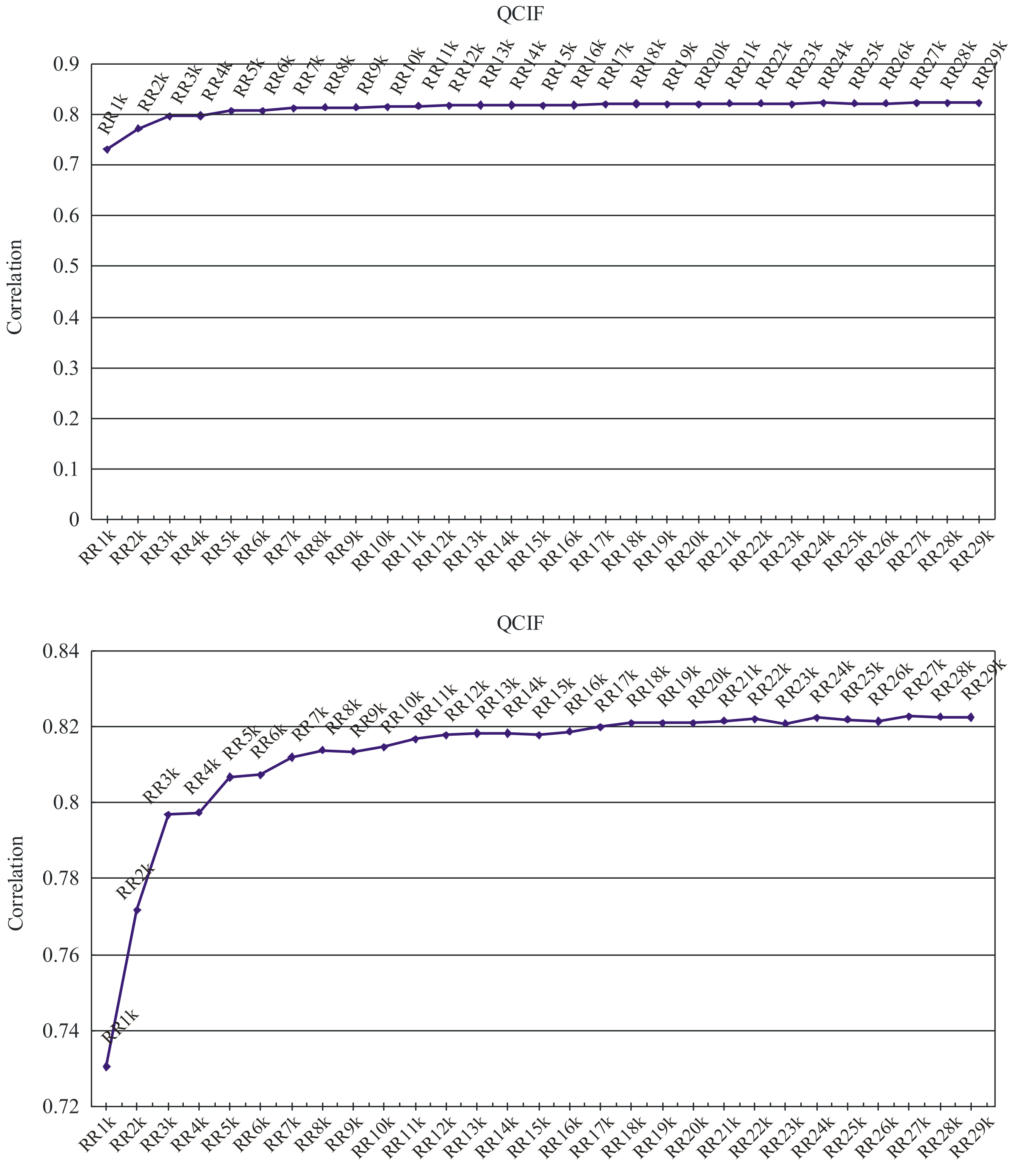
Appendix 1
to Annex 2

1 Optimal side-channel bandwidths

Figure 20 shows correlation coefficients for different side-channel bandwidths for the QCIF video sets. It can be seen that the correlation coefficients are almost saturated at about 10 kbit/s. After that, increasing the bandwidth produces about 1% improvement.

FIGURE 20

Performance improvement as the side-channel bandwidth increases (QCIF)

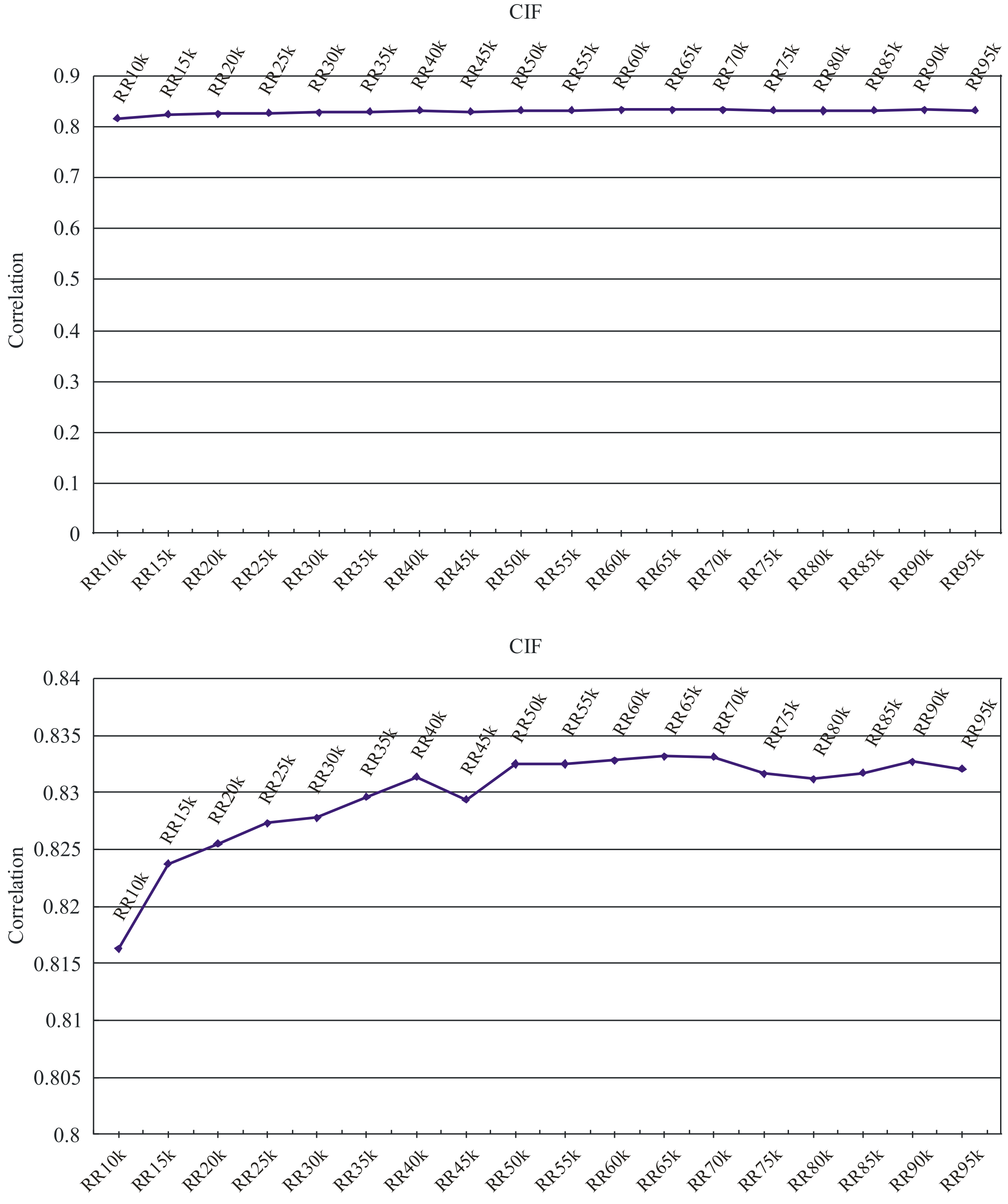


BT.1867-20

Figure 21 shows the correlation coefficients for different side-channel bandwidths for the CIF video sets. It can be seen that the correlation coefficients are almost saturated at about 15 kbit/s. After that, increasing the bandwidth produces about 0.5% improvement.

FIGURE 21

Performance improvement as the side-channel bandwidth increases (CIF)

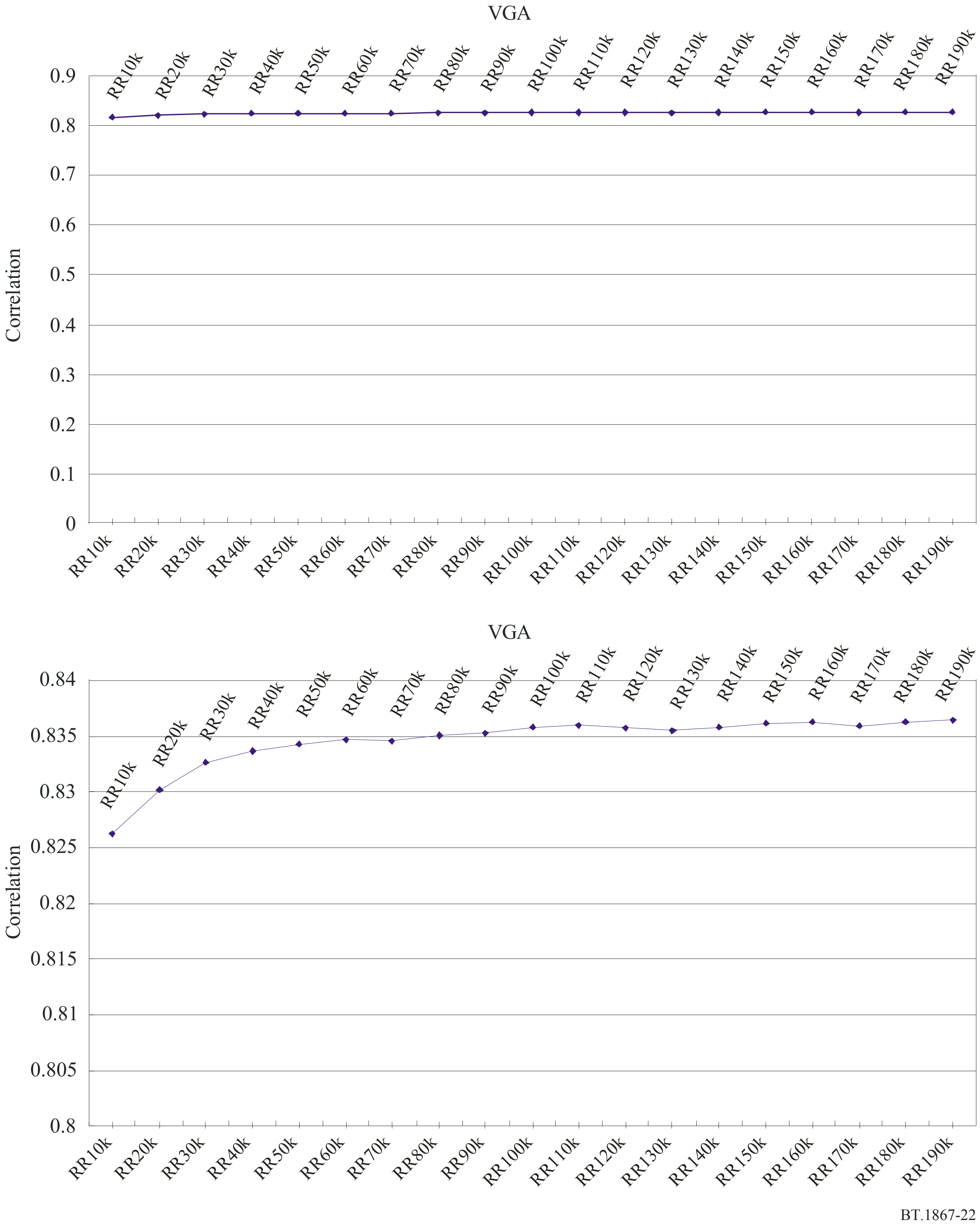


BT.1867-21

Figure 22 shows correlation coefficients for different side-channel bandwidths for the VGA video sets. It can be seen that the correlation coefficients are almost saturated at about 30 kbit/s. After that, increasing the bandwidth produces about 0.5% improvement.

FIGURE 22

Performance improvement as the side-channel bandwidth increases (VGA)



BT.1867-22