International Telecommunication Union

ITU-R

Radiocommunication Sector of ITU

Recommendation ITU-R BT.1866

(03/2010)

Objective perceptual video quality measurement techniques for broadcasting applications using low definition television in the presence of a full 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 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

RECOMMENDATION ITU-R BT.1866

Objective perceptual video quality measurement techniques for broadcasting applications using low definition television* in the presence of a full reference signal

(2010)

Scope

This Recommendation specifies methods for estimating the perceived video quality of broadcasting applications using low definition television (LDTV) when a full 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 full reference;

c) that Recommendation ITU-R BT.1833 describes multimedia systems for the 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.247[1] specifies objective measurement techniques of perceptual video quality applicable to LDTV applications in the presence of a full reference;

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

recognizing

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

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 models given in ITU-T Recommendation J.247 should be used for broadcasting applications using LDTV when a full reference signal is available.

NOTE 1 – Summaries of the measurement models are given in Annexes 2 to 5 for information. For more detail see ITU-T Recommendation J.247.

---

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

[1] ITU-T Recommendation J.247 is available at <http://www.itu.int/rec/T-REC-J.247-200808-P/en>.

Annex 1

1 Introduction

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

The estimation methods in this Recommendation are applicable to:

– codec evaluation, specification, and acceptance testing;

– potentially real-time, in-service quality monitoring at the source;

– remote destination quality monitoring when a copy of the source is available;

– quality measurement for monitoring of a storage or transmission system that uses either a single pass or concatenation of video compression and decompression techniques;

– lab testing of video systems.

The full reference measurement method can be used when the unimpaired reference video signal is readily available at the measurement point, as may be the case of measurements on individual equipment or a chain of processes in a laboratory or in a closed environment. The estimation methods are based on processing low definition 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 and dropped packets).

In a case where coding distortions are considered in the video signals, the encoder can use various compression methods (e.g. MPEG-2, H.264, etc.). The models described 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, and frame rate reduction). The models in this Recommendation can also be used for lab testing of video systems.

This Recommendation is deemed appropriate for services delivered at 4 Mbit/s or less presented on mobile receivers. The following conditions were used in the validation test for each resolution and found to be suitable:

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

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

– VGA (video graphics array (640 × 480 pixels)): 128 kbit/s to 4 Mbit/s.

TABLE 1

Factors used in evaluation of models

Test factors
Transmission errors with packet loss
Video resolution QCIF, CIF and VGA
Video bitrates QCIF: 16 to 320 kbit/s CIF: 64 kbit/s to 2 Mbit/s VGA: 128 kbit/s to 4 Mbit/s
Temporal errors (pausing with skipping) of maximum 2 s
Video frame rates from 5 to 30 fps

TABLE 1 (*end*)

Coding schemes
H.264/AVC (MPEG-4 Part 10), MPEG-4 Part 2, and three other proprietary coding schemes. (See Note 1 below.)

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. Note 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

Applications for the estimation models described in this Recommendation include but are not limited to:

1 codec evaluation, specification, and acceptance testing;

2 potentially real-time, in-service quality monitoring at the source;

3 remote destination quality monitoring when a copy of the source is available;

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

5 lab testing of video systems.

3 Model usage

This Recommendation includes the objective computational models shown in Table 2. An overview of the model performance is shown in Table 3. Further information is given in Appendix 1.

TABLE 2

Objective computational models

Model id	Proponent	Country	Annex
A	NTT	Japan	2
B	OPTICOM	Germany	3
C	Psytechnics	United Kingdom	4
D	Yonsei University	Korea (Rep. of)	5

All four models significantly outperform the peak signal to noise ratio (PSNR).

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

Models B and C tend to perform slightly better than Models A and D in some resolutions. Models B and C usually produce statistically equivalent results. For QCIF, Model A is often statistically equivalent to Models B and C. For VGA, Model D is typically statistically equivalent to Models B and C. The tables below provide an overview of the model's performances.

Although all four models can be used to adequately meet different industrie's needs, for VGA, it is highly advised that Models B, C, or D be used to obtain slightly better performance in most cases. For the same reason, it is highly advised that Models B or C be used for CIF and that Models A, B, or C be used for QCIF.

Model B shows the best overall minimum correlation. The minimum correlation coefficients of Models B, A, D and C are 0.68, 0.60, 0.59 and 0.57, respectively.

Model C obtained the highest number of occurrences of being in the top performing group. The total number of occurrences in the top group was 37 for Model C, 34 for Model B, 25 for Model A, and 24 for Model D.

TABLE 3

Model performance overview

VGA	Model A	Model B	Model C	Model D	PSNR
Avg. Correlation	0.786	0.825	0.822	0.805	0.713
Min. Correlation	0.598	0.685	0.565	0.612	0.499
Occurrences at Rank 1	8	10	11	10	3
Ranking analysis	Second	Best	Best	Best	–

CIF	Model A	Model B	Model C	Model D	PSNR
Avg. Correlation	0.777	0.808	0.836	0.785	0.656
Min. Correlation	0.675	0.695	0.769	0.712	0.440
Occurrences at Rank 1	8	13	14	10	0
Ranking analysis	Second	Best	Best	Second	–

QCIF	Model A	Model B	Model C	Model D	PSNR
Avg. Correlation	0.819	0.841	0.830	0.756	0.662
Min. Correlation	0.711	0.724	0.664	0.587	0.540
Occurrences at Rank 1	9	11	12	4	1
Ranking analysis	Best	Best	Best	Second	–

4 Limitations

The estimation models described in this Recommendation cannot be used to fully replace subjective testing. Correlation values between two carefully designed and executed subjective tests (in two different laboratories) normally fall between 0.95 to 0.98.

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

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

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

Appendix 1
to Annex 1

Findings of Video Quality Experts Group

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 and Radiocommunication Study Group 6. The recently completed Multimedia Phase I test of VQEG assessed the performance of proposed full reference perceptual video quality measurement algorithms for QCIF, CIF, and VGA formats.

Based on present evidence, four methods can be recommended by ITU-R at this time. These are:

Model A (Annex 2) − VQEG Proponent NTT, Japan

Model B (Annex 3) − VQEG Proponent OPTICOM, Germany

Model C (Annex 4) − VQEG Proponent Psytechnics, UK

Model D (Annex 5) − VQEG Proponent Yonsei University, Korea (Republic of).

The technical descriptions of these models can be found in Annexes 2 through 5 respectively. Note that the ordering of Annexes is purely arbitrary and provides no indication of quality prediction performance.

Table 4 provides details of the model's performances in the VQEG Multimedia Phase I test.

TABLE 4

a) VGA resolution: Model's performances in VQEG Multimedia Phase I test – Averages over 14 subjective tests

Metric	Model A	Model B	Model C	Model D	PSNR[(1)]
Annex	2	3	4	5	
Pearson correlation	0.786	0.825	0.822	0.805	0.713
RMS error	0.621	0.571	0.566	0.593	0.714
Outlier ration	0.523	0.502	0.524	0.542	0.615

TABLE 4 (*end*)

b) CIF resolution: Model's performances in VQEG Multimedia Phase I test – Averages over 14 subjective tests

Metric	Model A	Model B	Model C	Model D	PSNR[1]
Annex	2	3	4	5	
Pearson correlation	0.777	0.808	0.836	0.785	0.656
RMS error	0.604	0.562	0.526	0.594	0.720
Outlier ration	0.538	0.513	0.507	0.522	0.632

c) QCIF resolution: Model's performances in VQEG Multimedia Phase I test – Averages over 14 subjective tests

Metric	Model A	Model B	Model C	Model D	PSNR[1]
Annex	2	3	4	5	
Pearson correlation	0.819	0.841	0.830	0.756	0.662
RMS error	0.551	0.516	0.517	0.617	0.721
Outlier ration	0.497	0.461	0.458	0.523	0.596

(1) The PSNR values reported here are taken from the VQEG Multimedia Phase I final report (see: http://www.its.bldrdoc.gov/vqeg/projects/multimedia/). These values were calculated by NTIA/ITS.

Based on each metric, each FR VGA model was in the group of top performing models the following number of times:

Statistic	Model A	Model B	Model C	Model D	PSNR
Correlation	8	10	11	10	3
RMSE[1]	4	8	10	6	0
Outlier ratio	9	11	12	8	4

(1) RMSE: root mean square error.

Based on each metric, each FR CIF model was in the group of top performing models the following number of times:

Statistic	Model A	Model B	Model C	Model D	PSNR
Correlation	8	13	14	10	0
RMSE[1]	6	10	13	9	0
Outlier ratio	10	13	12	11	1

(1) RMSE: root mean square error.

Based on each metric, each FR QCIF model was in the group of top performing models the following number of times:

Statistic	Model A	Model B	Model C	Model D	PSNR
Correlation	9	11	12	4	1
RMSE[1]	7	10	11	2	1
Outlier ratio	10	11	12	8	4

(1) RMSE: root mean square error.

NOTE 1 – As a general guideline, small differences in these totals do not indicate an overall difference in performance.

Secondary analysis

The secondary analysis averages all video sequences associated with each video system (or condition) and thus reflects how well the model tracks the average hypothetical reference circuit (HRC) performance. The following tables show the average correlations for each model and resolution in the secondary analysis.

VGA correlation

	Model A	Model B	Model C	Model D	PSNR
Average	0.891	0.914	0.903	0.864	0.809

CIF correlation

	Model A	Model B	Model C	Model D	PSNR
Average	0.915	0.919	0.913	0.892	0.817

QCIF correlation

	Model A	Model B	Model C	Model D	PSNR
Average	0.942	0.937	0.920	0.893	0.882

Annex 2

Model A

Model A is divided into three software modules: a video alignment module, temporal/spatial feature amount derivation module, and subjective video quality estimation module (Fig. 1).

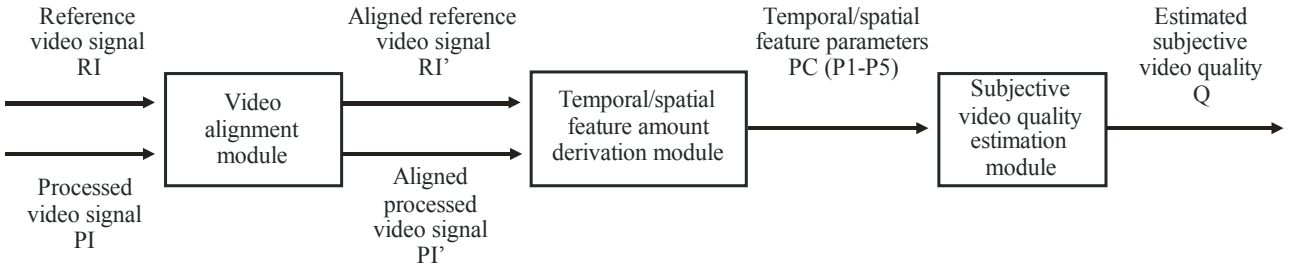
The video alignment module is divided into a macro alignment process and a micro alignment process. The macro alignment process matches pixels between reference video signals (RI) and processed video signals (PI) in spatial-temporal directions and filters the video sequences in consideration of the influence of video capturing and post-processing of the decoder. The micro alignment process matches frames between reference and processed video sequences in consideration of the influence of video frame skipping and freezing after the macro alignment process has finished.

The temporal/spatial feature amount derivation module calculates a spatial degradation parameter and a temporal degradation parameter (PC) by using an aligned reference video signal (RI') and an aligned processed video signal (PI'). The spatial degradation parameter is based on four parameters that reflect either the presence of overall noise, spurious edges, localized motion distortion, or localized spatial distortion. The temporal degradation parameter, calculated by weighted freeze-length summation, reflects frame freezing and frame-rate variation.

The subjective video quality estimation module calculates the objective video quality (Q) by using the previously mentioned parameters.

FIGURE 1

Block diagram of Model A



BT.1866-01

See Annex A of ITU-T Recommendation J.247 (08/2008) for the full description of Model A.

Annex 3

Model B

The basic idea of Model B is given in Fig. 2.

In the pre-processing step, a spatial region of interest (ROI) is extracted from the reference and test signal. All subsequent calculations are performed on this ROI only which is represented by the cropped signals S_P and P_P. The pre-processing is followed by a coarse alignment (registration) of the input sequences in the temporal and luminance domain. The "luminance and alignment information", obtained by these modules is used in the subsequent "detailed sequence alignment" process which performs the temporal frame by frame alignment of the two video sequences, a compensation for spatial shifts, and a compensation for differences in colour and brightness based on histogram evaluations. The results of the "detailed sequence alignment" are the "matching info", which is used to determine the perceptual impact of temporal degradations, as well as the cropped and aligned sequences S_A and P_A.

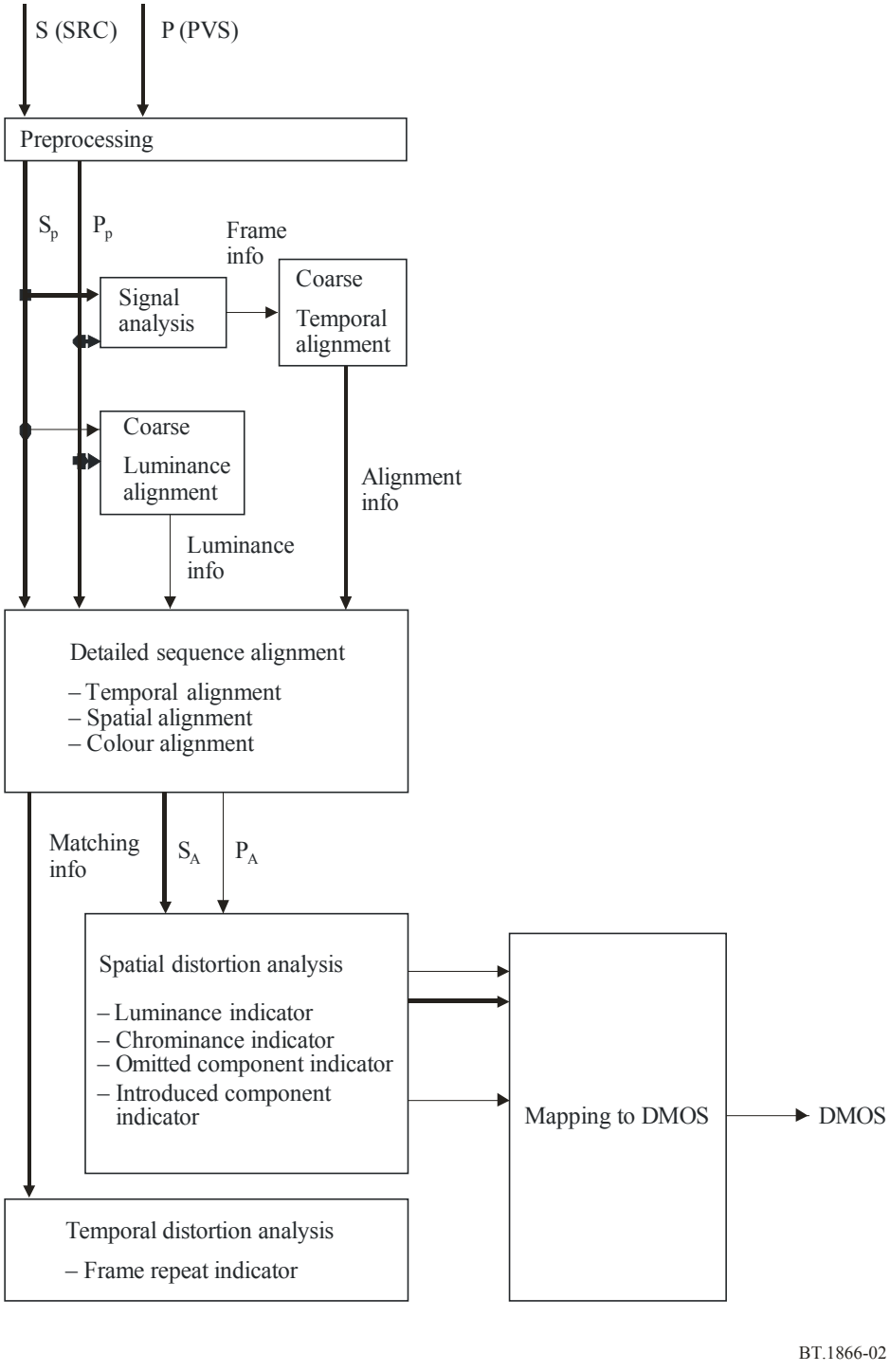
The spatial distortions are further analysed by the "spatial distortion analysis" block, which calculates the perceptual differences between the sequences in the spatial domain, resulting in four distortion indicators.

The "matching info" is further processed by the perceptual "temporal distortion analysis", which results in one indicator representative for frame repeats and other temporal distortions.

In the last step of Model C the five indicators that were derived above are weighted by logistic functions and combined to form the final PEVQ (perceptual evaluation of video quality) score, which correlates highly with a MOS (mean opinion score) obtained from subjective tests.

FIGURE 2

Overview of Model B



See Annex B of ITU-T Recommendation J.247 (08/2008) for the full description of Model B.

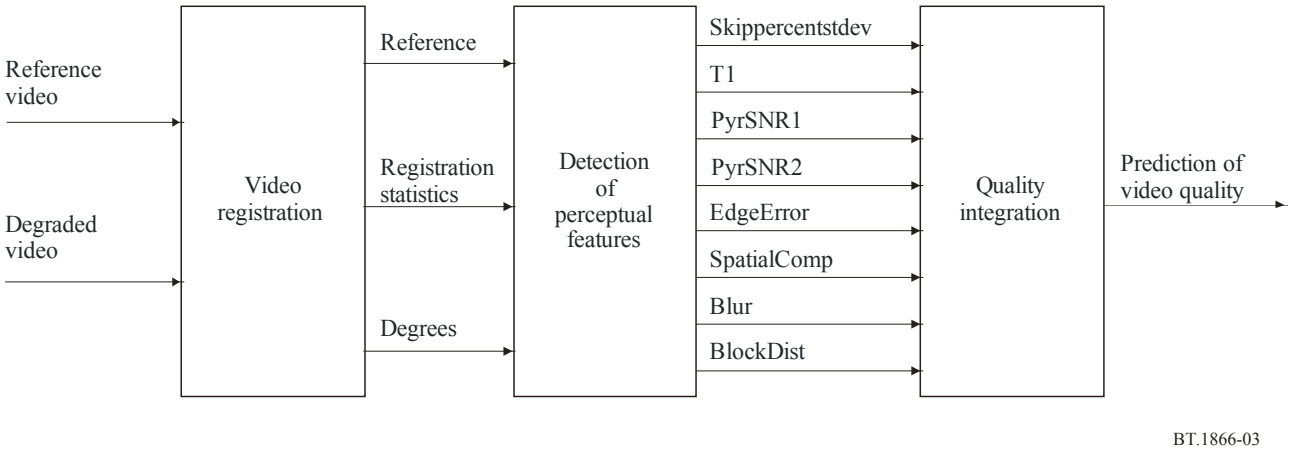
Annex 4

Model C

An overview of Model C is provided in Fig. 3. The model includes three main stages: (1) video registration, (2) detection of perceptual features and (3) integration of these features into an overall quality prediction score.

FIGURE 3

Overview of Model C



The first stage consists of a video registration which temporally pairs each frame in the degraded video to the best matching frame in the reference video. Video registration is necessary to provide accurate quality measurement. From the original reference and degraded videos, matched reference and matched degraded video sequences are obtained, together with registration information. The matched reference video contains all the reference frames that have been matched to frames of the degraded video. The matched degraded video contains all degraded frames for which a reference frame has been identified. The video registration algorithm is able to cope with time-varying temporal and spatial offsets between reference and degraded videos.

The second stage consists in the analysis of a set of perceptually meaningful features extracted either directly from the degraded video or from a comparison between the reference and degraded videos. These perceptual features are not based on any assumption on the type of content in the video or how the video has been encoded and transmitted to the end user. The perceptual features act as a sensory layer that filters out the image components to which the human viewer is insensitive. This is because video codecs reduce the amount of data to encode by applying image processing techniques to remove components of the video signal to which viewers would be least likely to notice as missing. By incorporating a model of the human visual system, this stage of the algorithm is able to determine how effective these processes are and identify visible errors. Where a coding scheme does not successfully achieve this goal, the degradations will be visible to the end user and will therefore form a part of the output from the sensory layer.

In the final stage, the perceptual parameters are combined to produce a single overall prediction of video quality. The optimum form of the integration function was obtained by exercising the model over an extensive set of subjective experiments (training set) and verifying its performance on a set of unknown experiments (validation set).

The format of the input reference and degraded videos supported by the software implementation of the model that was submitted to the VQEG Multimedia Phase I evaluation was uncompressed AVI with UYVY (YUV 4:2:2) colour space format, as specified by the VQEG Multimedia test plan. However, the quality assessment model is independent from this format and is therefore equally applicable with other formats (e.g. uncompressed AVI RGB24) provided that a proper input filter (e.g. colour space conversion filter or file reader) is applied first.

Both reference and degraded videos must be in progressive format. The absolute frame rate of the degraded video must be identical to the one of the reference video (e.g. 25 fps), where absolute frame rate is the number of (progressive) frames per second. However, the effective frame rate of the degraded video can be different from the frame rate of the reference video, where effective frame rate is the (average) number of unique frames per second. The effective frame rate of the degraded video can also be variable in time. The following example is provided as illustration. Reference video (A) has an absolute frame rate of 25 fps and is encoded with an effective (target) frame rate of 12.5 fps (B). The encoded video is played back (i.e. decoded) and captured at 25 fps (C). Consequently, every other frame in C is identical to the previous one. The input reference and degraded videos to the quality assessment model are A and C, respectively.

See Annex C of ITU-T Recommendation J.247 (08/2008) for the full description of Model C.

Annex 5

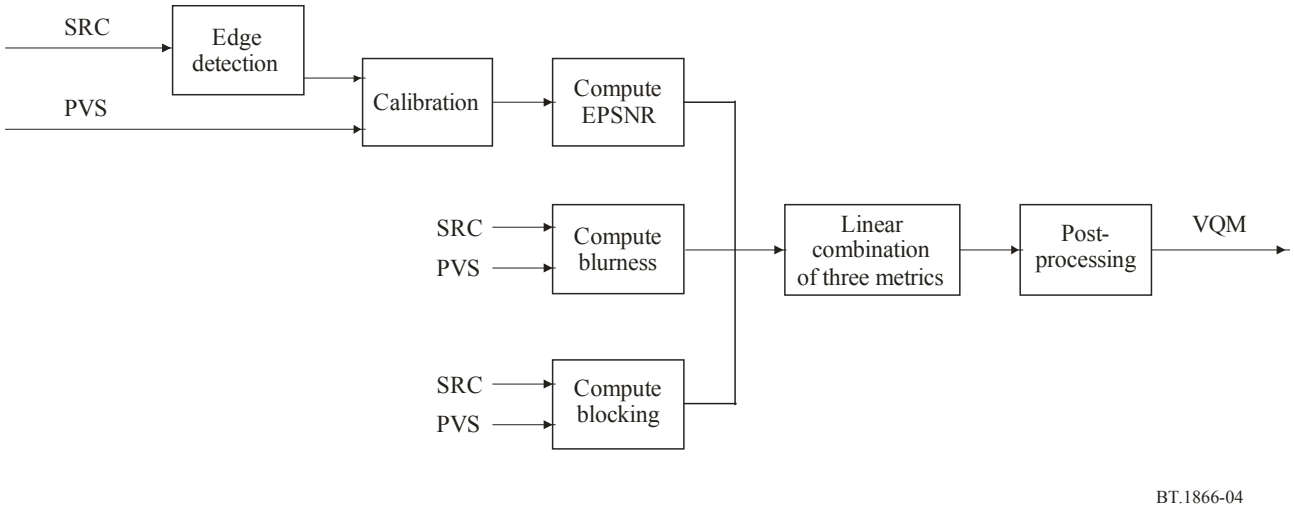
Model D

It is observed that the human visual system is sensitive to degradation around the edges. It is further observed that video compression algorithms tend to produce more artefacts around edge areas. Based on this observation, the model provides an objective video quality measurement method that measures degradation around the edges. In the model, an edge detection algorithm is first applied to the source video sequence to locate the edge areas. Then, the degradation of those edge areas is measured by computing the mean squared error. From this mean squared error, the edge PSNR (EPSNR) is computed. Furthermore, the model computes two additional features which are combined with the EPSNR to produce the final video quality metric (VQM).

Figure 4 shows the block diagram of Model D based on edge degradation of a full reference (FR) model, which takes two inputs: source video sequence (SRC) and processed video sequence (PVS).

FIGURE 4

Block diagram of Model D



See Annex D of ITU-T Recommendation J.247 (08/2008) for the full description of Model D.