|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Question(s):** | 2 | | **Meeting, date:** | | July 1, 2010 | |
| **Study Group:** | 9 | **Working Party:** |  | **Intended type of document** (R-C-TD): | |  |
| **Source:** |  | | | | | |
| **Title:** |  | | | | | |
| **Contact:** | <Quan>  <Organization>  <Country> | | | | Tel:  Fax:  Email: | |
| **Contact:** | Margaret and Alina | | | | Tel:  Fax:  Email: | |
| Please don’t change the structure of this table, just insert the necessary information. | | | | | | |

<Recommendation No.>

Objective perceptual multimedia video quality measurement of HDTV    
for digital cable television in the presence of a full reference

Summary

This Recommendation provides guidelines on the selection of appropriate objective perceptual video quality measurement methods for HDTV when a full reference signal is available. The following are example applications that can use this Recommendation:

1. potentially real-time, in-service quality monitoring at the source;
2. remote destination quality monitoring when a copy of the source is available;
3. 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;
4. lab testing of video systems.

Keywords

<Optional>

Introduction

<Optional - This clause should appear only if it contains information different from Scope and Summary>

# 1 Scope

This Recommendation provides guidelines and recommendations on the selection of appropriate perceptual video quality measurement equipment for use in high definition television (HDTV) applications when the full reference (FR) measurement method can be used.

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 in the laboratory or in a closed environment such as a cable television head‑end. The estimation methods include both calibration and objective video quality estimations.

The validation test material contained both H.264 and MPEG-2 coding degradations and various transmission error conditions (e.g., bit errors, dropped packets). The models proposed 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 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 J.149]) to determine the models' accuracy for that particular context.

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

* 1080i 60 Hz (30 fps)
* 1080p (25 fps)
* 1080i 50 Hz (25 fps)
* 1080p (30 fps)
* 720p 60 Hz (60 fps)
* 720p 50 Hz (50 fps)

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

|  |
| --- |
| ***Test factors*** |
| Video resolution: HD, 720 x 1280 and 1080 x 1920, interlaced and progressive |
| Video frame rates 29.97 and 25fps |
| Video bitrates: |
| Temporal errors (pausing with skipping) of maximum 2 seconds |
| Transmission errors with packet loss |
| ***Coding technologies*** |
| H.264/AVC (MPEG-4 Part 10) |
| MPEG-2 |

## 1.1 Application

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

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

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

3) 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;

4) lab testing of video systems.

## 1.2 Model usage

## 1.3 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 (i.e. in two different laboratories) normally fall within the range 0.95 to 0.98. If this Recommendation is utilized to make video system comparisons (e.g., comparing two codecs), it is advisable to use a quantitative method (such as that in J.149) to determine the models’ accuracy for that particular context.

The models in this Recommendation were validated by measuring video that exhibits frame freezes up to 2 seconds. 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). The model was also not tested on frame-rates artificially and steadily reduced rather the common frame-rates for TV systems such as 29.97fps and 25fps in interlaced and 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 models may produce erroneous results. Here a subjective evaluation is required.

# 2 References

The following ITU-T 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-T 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.

Normative References

[ITU-T P.910] ITU-T Recommendation P.910 (2008), Subjective video quality assessment methods or multimedia applications.

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

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

Informative References

[ITU-T J.244] ITU-T Recommendation J.244 (2008), Calibration methods for constant misalignment of spatial and temporal domains with constant gain and offset

[ITU-R BT.500-11] ITU-R BT.500-11, Methodology for the subjective assessment of the quality of television pictures.

[ITU-T J.149] ITU-T Recommendation J.149 (1998), Subjective audiovisual quality assessment methods for multimedia applications.

[ITU-T Rec. J.247] ITU-T Recommendation J.247 (08/2008) – Prepublished version 6, Objective perceptual multimedia video quality measurement in the presence of a full reference

[ITU-T J.144] ITU-T Recommendation J.144 (2001), Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference.

[ITU-T P.931] ITU-T Recommendation P.931 (1998), Multimedia communications delay, synchronization and frame rate measurement.

[ITU-T J.148] ITU-T Recommendation J.148 (2003), Requirements for an objective perceptual multimedia quality model.

[ITU-T H.264] ITU-T Recommendation H.264 (2003), Advanced video coding for generic audiovisual services

[VQEG] Final report from the Video Quality Experts Group on the validation of objective models of HDTV quality-Phase I, 2010

# 3 Definitions

This recommendation uses the following terms defined elsewhere.

## 3.1 Terms defined elsewhere:

This Recommendation uses the following terms defined elsewhere:

**3.1.1 subjective assessment (picture)** [ITU-T J.144]: 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)** [ITU-T J.144]: 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** [ITU-T J.144]: 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:** is defined as the number of unique frames (i.e., total frames – repeated frames) per second

**3.2.2**  **Simulated transmission errors:** are defined as 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:** are defined as 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 P.910)

ACR-HR Absolute Category Rating with Hidden Reference (see 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 Color Space and file format

# 5 Conventions

*None.*

# 6 Description of full reference methodology

The double-ended measurement method with full reference, for objective measurement of perceptual video quality, evaluates the performance of systems by making a comparison between the undistorted input, or reference, video signal at the input of the system, and the degraded signal at the output of the system (Figure 1).

Figure 1 shows an example of application of the full reference method to test a codec in the laboratory.



Figure 1 − Application of the full reference perceptual quality measurement method   
to test a codec in the laboratory

The comparison between input and output signals may require a temporal alignment or a spatial alignment process, the latter to compensate for any vertical or horizontal picture shifts or cropping. It also may require correction for any offsets or gain differences in both the luminance and the chrominance channels. The objective picture quality rating is then calculated, typically by applying a perceptual model of human vision.

Alignment and gain adjustment is known as registration. This process is required because most full reference methods compare reference and processed pictures on what is effectively a pixel-by-pixel basis. The video quality metrics described in Annexes A through D include registration methods.

As the video quality metrics are typically based on approximations to human visual responses, rather than on the measurement of specific coding artefacts, they are in principle equally valid for analogue systems and for digital systems. They are also in principle valid for chains where analogue and digital systems are mixed, or where digital compression systems are concatenated.

Figure 2 shows an example of the application of the full reference method to test a transmission chain.



Figure 2 − Application of the full reference perceptual quality measurement method   
to test a transmission chain

In this case, a reference decoder is fed from various points in the transmission chain, e.g., the decoder can be located at a point in the network, as in Figure 2, or directly at the output of the encoder, as in Figure 1. If the digital transmission chain is transparent, the measurement of objective picture quality rating at the source is equal to the measurement at any subsequent point in the chain.

It is generally accepted that the full reference method provides the best accuracy for perceptual picture quality measurements. The method has been proven to have the potential for high correlation with subjective assessments made in conformity with the ACR-HR methods specified in [ITU‑T P.910].

# 7 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 and ITU-R Study Group 6. The recently completed high definition television phase I test of VQEG assessed the performance of proposed full reference perceptual video quality measurement algorithms.

**Primary Analysis of FR Models**

The performance of each FR model is summarized in the table below. “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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Metric** | **PSNR** | **NTT** | **Opticom** | **Swissqual** | **Tektronix** | **YonseiFR** |
| Superset RMSE | 0.71 | 0.74 | 0.88 | 0.56 | 0.65 | 0.74 |
| Top Performing Group Total | 1 | 0 | 0 | 5 | 3 | 1 |
| Better Than PSNR Total | 0 | 0 | 0 | 4 | 4 | 1 |
| Better Than Superset PSNR | No | No | No | Yes | No | No |
| Superset Correlation | 0.78 | 0.76 | 0.63 | 0.87 | 0.82 | 0.76 |

The body of this report includes other metrics including Pearson Correlation & 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., H.264 coding-only), scatter plots, and the fit coefficients for each model.

**FR Model Conclusions**

* VQEG believes that at least one FR model performed well enough to be included in normative sections of Recommendations.
* The scope of these Recommendations should be written carefully to ensure that the use of the models is defined appropriately.
* If the scope of these Recommendations includes video system comparisons (e.g., comparing two codecs), then the Recommendation should include instructions indicating how to perform an accurate comparison.
* None of the evaluated models reached the accuracy of the normative subjective testing.

Annex A  
  
Model description of the SwissQual HD model

(This annex forms an integral part of this Recommendation)

The model takes as input the full reference and the processed and possibly degraded video sequence. Score estimation is based on the following steps:

1. First, the video sequences are preprocessed. In particular, the frames are subsampled.
2. A temporal frame alignment between reference and processed video sequence is performed.
3. A spatial frame alignment between processed video frame and the corresponding reference video frame is performed and local similarity and a difference measure inspired by visual perception, a jerkiness measure, and a blockiness measure are computed.
4. The quality score is estimated based on a non-linear aggregation of the above features.
5. To avoid misprediction in case of relatively large spatial misalignment between reference and processed video sequence, the above steps are computed for 3 different horizontal and vertical spatial alignments of the video sequence, and the maximum predicted score among all spatial positions is the final estimated quality score.

Steps 1-3 are explained in more detail the following:

1. Preprocessing

Each frame of the reference and the processed video sequence is subsampled to 3 different resolutions, R1, R2, R3:

original frame R1 R2 R3

(height x width)

1080 x 1920 --> 540 x 960 --> 270 x 480 --> 96 x 128

2. Time Alignment

The time alignment is performed using the reference and processed video sequence at resolution R3. Time alignment is performed in a recursive manner as follows:

1. Find a ‘anchor’ frame in the reference sequence.
2. Match this frame to the best matching frame in the degraded sequence.
3. Take this best matched frame in the degraded sequence and match it to a frame close to the ‘anchor’ frame of the reference. Store this pair of matching frames.
4. If this matched frame-pair is a good match, according to a similarity criteria, split the reference and processed video sequence at the matching frame pair, each into two video sequences before and after the matching frames. Start at 1. with both pairs of subsequences.  
   Otherwise, if the matching frame pair is not a good match, start again at 1. with a different ‘anchor’ frame of the reference video.

The result of the time alignment is a sequence, assigning each frame of the processed video sequence a frame of the reference, or an indicator, indicating that no sufficiently good matching frame could be found.

3. Spatial Frame Alignment and Computation of Quality Related Features

Iterate over all frames of the processed video sequence and:

1. Perform a spatial alignment between the processed and corresponding  
   reference frame. As a prior, the spatial alignment of the previous matching frames is used. If a different spatial shift leads to a significantly smaller difference between the processed and corresponding reference frame, the spatial shift is adjusted.
2. A local similarity and difference measure is computed by iterating over equally distributed squared regions of size 13 x 13 of the processed and degraded frame, called s\_prc and s\_ref, and computing the similarity S and difference D,  
     
   S = (cor(s\_prc,s\_ref) + 25) / (var(s\_ref) + 25)  
     
   D = ( S \* (s\_prc-mean(s\_prc)) – (s\_ref-mean(s\_ref)) )^2  
     
   where cor is the correlation and var is the variance of the pixel values in the corresponding squared region.
3. A blockiness feature is computed using the frames at resolution R1, by computing average edge strength over the frame at different offsets and comparing to the resulting difference of the corresponding reference frame.
4. A jerkiness feature is computed by averaging the product of relative display time, a non-linear transform of display time, and a non-linear transform of motion intensity.

These four features, similarity, difference, blockiness, and jerkiness are the basis for score estimation.

Appendix I  
  
<Appendix Title>

(This appendix does not form an integral part of this Recommendation)

<Body of appendix I>

Bibliography

[b-VQEG Report] VQEG Final Report of HDTV Phase I Validation Test (2010), “*Video Quality Experts Group: report on the validation of video quality models for high definition video content*”, Video Quality Experts Group (VQEG), http://www.its.bldrdoc.gov/vqeg/projects/hdtv

[b-ITU-R BT.601-6] Recommendation ITU-R BT.601-6 (2007), “*Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios*”.