

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.340

(06/2010)

SERIES J: CABLE NETWORKS AND TRANSMISSION
OF TELEVISION, SOUND PROGRAMME AND OTHER
MULTIMEDIA SIGNALS

Measurement of the quality of service

**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**

Recommendation ITU-T J.340



International
Telecommunication
Union
Documents

Document Sharing Hub

Copyright International Telecommunication Union/ITU Telecommunication Sector
Provided by IHS Markit under license with ITU-T
No reproduction or networking permitted without license from IHS

Recommendation ITU-T J.340

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

Summary

Peak signal to noise ratio (PSNR) is a useful benchmark for evaluating performance improvements of new objective perceptual video quality metrics. This PSNR calculation method in Recommendation ITU-T J.340 has the advantage of automatically determining the highest possible PSNR value for a given video sequence over the range of spatial and temporal shifts. Only one temporal shift is allowed for all frames in the entire processed video sequence (i.e., constant delay).

This Recommendation defines a full reference (FR) algorithm for computing both the calibration and PSNR estimations for a processed video sequence: peak signal to noise ratio with compensation for constant spatial shifts, constant temporal shift, and constant luminance gain and offset (PSNR_{const}). Since the PSNR_{const} algorithm only examines the Y luminance channel (as defined by Recommendation ITU-R BT.601-6) distortions in the CB and CR chrominance channels will not be detected by the algorithm of this Recommendation. The intent of this Recommendation is to define and facilitate a standardized PSNR metric for use by industry and standard organizations. Reference code and test vectors have been included to assure accurate and consistent implementation of this PSNR_{const} metric.

History

Edition	Recommendation	Approval	Study Group
1.0	ITU-T J.340	2010-06-29	9

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2011

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

CONTENTS

	Page
1 Scope	1
2 References.....	2
3 Definitions	2
4 Abbreviations and acronyms	2
5 Conventions	2
6 PSNR _{const} Algorithm Description	2
6.1 Introduction	2
6.2 Algorithm	3
Appendix I – Reference Code to Calculate PSNR.....	5
Bibliography.....	20

100

Recommendation ITU-T J.340

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

1 Scope

Peak signal to noise ratio (PSNR) is a useful benchmark for evaluating performance improvements of new objective perceptual video quality metrics. For example, PSNR has been used as a benchmark for both the multimedia (MM) and reduced reference television (RRTV) test programs recently completed by the video quality experts group (VQEG). Since the calculation of PSNR is highly dependent upon proper estimation of spatial alignment, temporal alignment, gain, and level offset between the processed video sequence and the original video sequence, the method of measurement for PSNR should ideally include a method for performing these calibration procedures. This PSNR calculation method in this Recommendation has the advantage of automatically determining the highest possible PSNR value for a given video sequence over the range of spatial and temporal shifts. Only one temporal shift is allowed for all frames in the entire processed video sequence (i.e., constant delay).

This Recommendation defines a full reference (FR) algorithm for computing both the calibration and PSNR estimations for a processed video sequence: peak signal to noise ratio with compensation for constant spatial shifts, constant temporal shift, and constant luminance gain and offset ($\text{PSNR}_{\text{const}}$). Since the $\text{PSNR}_{\text{const}}$ algorithm only examines the Y luminance channel (as defined by [b-ITU-R BT.601-6]) distortions in the CB and CR chrominance channels will not be detected by the algorithm of this Recommendation. The intent of this Recommendation is to define and facilitate a standardized PSNR metric for use by industry and standards organizations. Reference code and test vectors have been included to assure accurate and consistent implementation of this $\text{PSNR}_{\text{const}}$ metric.

The intention of this $\text{PSNR}_{\text{const}}$ metric is to fully calibrate the video and then calculate PSNR on the luminance plane only. For these purposes, calibration consists of selecting the valid video region spatially (e.g., discarding the overscan region) and then removing from the entire video sequence a constant temporal shift (delay or advance), a constant spatial shift (vertically and horizontally), and a constant luminance gain and offset.

Common applications of the $\text{PSNR}_{\text{const}}$ algorithm include:

- A reference benchmark for evaluating the effectiveness of perceptual quality metrics.
- A FR quality of service metric (QoS) for video transmission systems.

It should be noted that the use of this calibration method will result in the best possible $\text{PSNR}_{\text{const}}$ value for a video sequence with constant delay. This value may differ from $\text{PSNR}_{\text{const}}$ obtained with perfectly calibrated video. Degradations due to gain are removed by this method. If higher $\text{PSNR}_{\text{const}}$ is obtained using an alternative calibration method, then this alternative calibration method can be used, although differences are expected to be small. It is noted that this Recommendation considers only integer-pixel shifts.

It should be noted that this $\text{PSNR}_{\text{const}}$ algorithm will not detect as an impairment a constant spatial shift, a constant luminance gain, a constant luminance offset, or a constant temporal shift.

2 References

None.

3 Definitions

None.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

FR	Full Reference
MSE	Mean Squared Error
PSNR	Peak Signal to Noise Ratio
PSNR _{const}	Peak Signal to Noise Ratio with Compensation for Constant Spatial Shifts, Constant Temporal Shift, and Constant Luminance Gain and Offset
QoS	Quality of Service
SROI	Spatial Region of Interest
ST	Spatial-Temporal
TROI	Temporal Region of Interest

5 Conventions

None.

6 PSNR_{const} algorithm description

6.1 Introduction

The PSNR_{const} algorithm performs an exhaustive search for the maximum Y-channel PSNR_{const} over plus or minus the horizontal and vertical spatial uncertainties (in pixels) and plus or minus the temporal uncertainty (in frames). The processed video segment is fixed and the original video segment is shifted over the search range. For each spatial-temporal (ST) shift, a linear fit between the Y channel processed and the original pixels is performed such that the mean squared error (MSE) of original – (gain*processed + offset) is minimized, hence maximizing PSNR. Thus, this calculation of PSNR_{const} yields values that are greater than or equal to commonly used PSNR implementations if the exhaustive search covered enough ST shifts.

The ST search range, spatial region of interest (SROI), and temporal region of interest (TROI) are input parameters to the algorithm and the user of this Recommendation is advised to carefully consider appropriate input arguments for these quantities. For instance, some video systems truncate the border picture elements replacing these pixels with black. In these cases, the SROI should be reduced so as to not include these pixels in the PSNR_{const} calculation.

Since the ST search performed by the algorithm only considers integer shifts in space and time, this method is not appropriate for video systems that contain sub-pixel shifts. Caution should also be observed when analysing video frames that contain interlaced video frames to assure that the interlaced framing of the processed video sequence is identical to the original video sequence before applying the algorithm (i.e., the algorithm does not account for one-field or half frame timing shifts between the processed and original).

Since the video delay is assumed to be constant for all video frames in the processed video segment, this algorithm may not be appropriate for video systems that contain variable video delays (i.e., where the video delay of individual frames may vary). The algorithm may be used in these cases if one desires the PSNR_{const} calculation to include distortions due to variable video delays.

6.2 Algorithm

PSNR is defined as $10 \cdot \log_{10}$ of the ratio of the peak signal energy to the MSE observed between the processed video signal and the original video signal. For the algorithm presented here, the peak signal energy is assumed to be $(2^R - 1)^2$ where R is the pixel depth, and the MSE summation is performed over the selected SROI and TROI of the processed video sequence. It is noted that the peak signal energy is 255^2 (i.e., $255 \cdot 255$) for 8-bit video and 1023^2 for 10-bit video. The algorithm performs a linear fit of the processed image pixels to the corresponding original image pixels for each ST shift that is examined before computing the MSE. This is equivalent to removing gain (contrast) and level offset (brightness) calibration errors in the processed video before performing the $\text{PSNR}_{\text{const}}$ calculation.

Computation of $\text{PSNR}_{\text{const}}$ for an (original, processed) video clip pair involves the following steps:

- 1) Determine the appropriate ST search range for the processed video clip. This involves estimating the x and y spatial uncertainty (in pixels, denoted here as x_{uncert} and y_{uncert}) of spatial registration errors that might be present, as well as estimating the t temporal uncertainty (in frames, denoted as t_{uncert}) of any temporal registration errors that might be present. Since the algorithm will perform an exhaustive search over plus or minus x_{uncert} , y_{uncert} , and t_{uncert} shifts of the original video sequence with respect to the processed video sequence, these estimates should be as tight as possible while still including the optimal ST registration.
- 2) Determine the maximum SROI and TROI that can be used for the processed video clip. If the processed video clip contains truncation of border pixels (i.e., black border), the SROI of the processed video clip should be reduced to eliminate these pixels. If the processed video clip contains transition video frames at the beginning or end of the video clip (perhaps due to prior or following scene content), these frames should be eliminated from the TROI. If necessary, reduce the maximum SROI and TROI to allow for the x_{uncert} , y_{uncert} , and t_{uncert} shifts found in step 1. The final SROI and TROI of the processed video clip that is determined by this step will remain fixed for all PSNR calculations. Since the original video clip will be shifted by a maximum of plus or minus x_{uncert} and y_{uncert} pixels and plus or minus t_{uncert} frames with respect to the processed video clip, one must assure that there are valid original video pixels that align to every processed video pixel within the final SROI and TROI.
- 3) For each ST shift of the original sequence in step 1 (i.e., shifts in the x , y and t directions will be denoted here as x_s , y_s and t_s , respectively), perform a linear fit of the processed pixels to the shifted original pixels. This linear fit is performed for all pixels in the entire ST region encompassed by the processed video SROI and TROI selected in step 2. For a given ST shift, this can be expressed as finding the $\text{Gain}(x_s, y_s, t_s)$ and $\text{Offset}(x_s, y_s, t_s)$ that minimizes the MSE given by:

$$\text{MSE} = \frac{1}{N} \sum_x \sum_y \sum_t \{O(x+x_s, y+y_s, t+t_s) - [\text{Gain}(x_s, y_s, t_s) * P(x, y, t) + \text{Offset}(x_s, y_s, t_s)]\}^2;$$

$(x, y) \in \text{SROI}, t \in \text{TROI}$