NTIA Technical Memorandum 13-497

Fast Low Bandwidth Model: A Reduced Reference Video Quality Metric

Margaret H. Pinson
Stephen Wolf

technical memorandum

U.S. DEPARTMENT OF COMMERCE • National Telecommunications and Information Administration

NTIA Technical Memorandum 13-497

Fast Low Bandwidth Model: A Reduced Reference Video Quality Metric

Margaret H. Pinson
Stephen Wolf

U.S. DEPARTMENT OF COMMERCE

June 2013

DISCLAIMER

Certain commercial equipment and materials are identified in this report to specify adequately the technical aspects of the reported results. In no case does such identification imply recommendation or endorsement by the National Telecommunications and Information Administration, nor does it imply that the material or equipment identified is the best available for this purpose.

This document describes software developed by NTIA. **NTIA does not make any warranty of any kind, express, implied or statutory, including, without limitation, the implied warranty of merchantability, fitness for a particular purpose, non-infringement and data accuracy. This software is provided "AS IS."** NTIA does not warrant or make any representations regarding the use of the software or the results thereof, including but not limited to the correctness, accuracy, reliability or usefulness of the software or the results. You can use, copy, modify, and redistribute the NTIA-developed software upon your acceptance of these terms and conditions, and upon your express agreement to provide appropriate acknowledgments of NTIA's ownership of and development of the software by keeping this exact text present in any copied or derivative works.

The user of this software ("Collaborator") agrees to hold the U.S. Government harmless and indemnifies the U.S. Government for all liabilities, demands, damages, expenses, and losses arising out of the use by the Collaborator, or any party acting on its behalf, of NTIA/ITS's software, or out of any use, sale, or other disposition by the Collaborator, or others acting on its behalf, of products made by the use of NTIA/ITS's software.

CONTENTS

FIGURES

FAST LOW BANDWIDTH MODEL: A REDUCED REFERENCE VIDEO QUALITY METRIC

Margaret H. Pinson[1] and Stephen Wolf[2]

This memorandum describes the Fast Low Bandwidth model. This summary is intended to help the reader understand the model from an algorithmic standpoint. Some knowledge of prior NTIA objective video quality models is necessary for the understanding of this document.

The Fast Low Bandwidth was designed to be operated in-service using the reduced reference (RR) methodology. The model requires access to the original video at one location, the processed video at another location, and low bandwidth data link between the two locations. That link is used to communicate RR features between the two locations. The Fast Low Bandwidth model is included in ITU-T Rec. J.249.

Key words: video; quality; models; metrics; features; parameters; objective; subjective; correlation; reduced-reference; spatial information (SI); temporal information (TI); impairments

1 INTRODUCTION

This report describes the NTIA Fast Low Bandwidth model, which appears in the normative section of ITU-T Rec. J.249 [1]. Reference code can be found in ITU-T Rec. J.249. This reduced reference (RR) model requires calibrated video.

The Fast Low Bandwidth model was designed to be run on video that has been calibrated using the NTIA RR calibration algorithms. Those calibration algorithms appear in the normative section of ITU-T Recommendation J.244 [2] and are described in [3].

Software implementing the Fast Low Bandwidth model and RR calibration algorithms is available for download [4]. This software can be used freely for any commercial or non-commercial purpose. Binary executable versions of these tools and their associated source code are available.

The NTIA Fast Low Bandwidth model is a direct successor to the NTIA General model, which appears in the normative sections of ITU-T Rec. J.144 [5] and ITU-R Rec BT.1683 [6]. The NTIA General model is summarized in [7], and as a result is known primarily by the acronym VQM. A full disclosure of VQM and background information can be found in [8].

---

[1] The author is with the Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, Boulder, CO 80305.
[2] The author was formerly with the Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, Boulder, CO 80305.

The RR calibration algorithms are successors to the full reference (FR) calibration algorithms. Those FR calibration algorithms, which also appear in [5]–[8],can be used with the Fast Low Bandwidth model, albeit not while running in-service.

For standard definition video (i.e., ITU-R Rec. BT.601 sampled video [9]), the Fast Low Bandwidth model uses RR features that require approximately 12 to 14 kbits/sec of transmission bandwidth. VQM is effectively an FR model, because the RR bandwidth requirements are too high to be practical. The Fast Low Bandwidth model uses RR features with much less bandwidth than VQM, yet the feature extraction and comparison process is similar. Thus, an understanding of the older model may be helpful.

2 BACKGROUND

VQM was finalized in 2000–2001. VQM uses RR features and parameters. These RR features require a high bandwidth, because the spatial-temporal (S-T) region size for each feature was optimized for model accuracy. This resulted in small S-T regions (e.g., 8 vertical lines by 8 horizontal pixels by 0.2 seconds) and consequently a high bandwidth requirement.

Once the VQM model was finalized, NTIA began development of an RR objective video quality model. The design goals were to:

- Match the accuracy of VQM.
- Use a very low RR bandwidth that could be reasonably achieved by most systems.
- Include improved performance for video resolutions both larger and smaller than standard definition video, such as QCIF, CIF, and HD.
- Have improved performance for the network impairments commonly appearing at that time. At the time, network impairments rarely changed the system delay, so the visual impact of such errors was expect to appear as corruption of individual frames.

In the 2003–2004 time frame, NTIA finalized two RR video quality models. These models were called the "Low Bandwidth model" and "Fast Low Bandwidth model." Based on training data, the Low Bandwidth model was expected to be slightly more accurate than the Fast Low Bandwidth model. These models often use an S-T region size of 30 vertical lines by 30 horizontal pixels by one second. A comparison of subjective ratings from video at resolutions from QCIF to HD indicated that this block size was appropriate for all resolutions.

The Fast Low Bandwidth model is a computationally efficient version of the Low Bandwidth model. The Fast Low Bandwidth model is about four times faster since it extracts spatial features from video frames that are time-averaged before the spatial features are extracted (as opposed to computing spatial features directly from the video frames). This is referred to as "pre-averaging", because it is a prefilter. The same algorithm appears in the Developer's model, which is a computationally efficient version of VQM (see [8]). The idea is that time-averaging mimics, in a simplistic way, the time averaging performed by the human visual system.

Additional computational savings for the Fast Low Bandwidth model resulted from computing temporal information (i.e., motion) features based on a random sub-sampling of pixels in the luma Y channel rather than using all pixels in all three video channels (Y, Cb, and Cr). Random algorithms originate in computer science theory:

> "A randomized strategy is typically useful when there are many ways in which an algorithm can proceed but it is difficult to determine a way that is guaranteed to be good….If the benefits of good choices outweigh the costs of bad choices, a random selection of good and bad choices can yield an efficient algorithm." [10]

NTIA wanted to submit both the Low Bandwidth and Fast Low Bandwidth models to the Reduced Reference TV (RRTV) tests for independent evaluation by the Video Quality Experts Group (VQEG). However, the test allowed only one model to be submitted for a given bandwidth. NTIA chose to submit these two models to different bit rate categories even though they have identical RR bit rate requirements. NTIA chose to submit the Low Bandwidth model to the 256k category and the Fast Low Bandwidth model to the 80k category since the performance of the Low Bandwidth model was expected to be superior to that of the Fast Low Bandwidth model. Both VQMs used the NTIA RR calibration algorithm which is included in ITU-T Recommendation J.244 [2]. This calibration algorithm requires approximately 22 to 24 kbits/sec of RR bandwidth to produce estimates for temporal delay, spatial shift, spatial scaling, gain, and level offset.

An interesting result from the VQEG RRTV evaluation tests [11] was that the Fast Low Bandwidth model outperformed the Low Bandwidth model for both the 525-line and 625-line test. This is an interesting result since it implies that extracting spatial features from averaged frames is superior to extracting them from non-averaged frames. Whether or not this result is true for other data sets is an area for further research. NTIA did not see a reason to standardize both models.

3 FAST LOW BANDWIDTH MODEL DESCRIPTION

3.1 Model Overview

The VQM description will encompass three primary areas: (1) the low bandwidth features that are extracted from the original and processed video streams, (2) the parameters that result from comparing like original and processed feature streams, and (3) the model calculation that combines the various parameters, each of which measures a different aspect of video quality. This description makes use of readily available references for the technical details.

3.2 Feature Description

3.2.1 Feature Overview

Three types of features are used by the Fast Low Bandwidth model: color, spatial, and temporal. Each of these feature types quantifies perceptual distortions in their respective domains. Reference code can be found in ITU-T Rec. J.249 [1].

3.2.2 Color Features

The color features are the same f_{COHER_COLOR} features that are used by VQM. These features are described in detail in Annex D.7.3 of ITU-T Recommendation J.144 [5]. The f_{COHER_COLOR} features provide a two-dimensional vector measurement of the amount of blue and red chrominance information (C_B, C_R) in each S-T region. Thus, the f_{COHER_COLOR} features are sensitive to color distortions. The f_{COHER_COLOR} features of the NTIA Fast Low Bandwidth model are extracted from S-T region sizes of 30 vertical lines by 30 horizontal pixels by one second of time (i.e., 30 × 30 × 1s) whereas VQM used S-T region sizes of 8 × 8 × 1 frame.

3.2.3 Spatial Features

The spatial features are the same f_{SI13} and f_{HV13} features that are used by VQM. These features are described in detail in Annex D.7.2.2 of ITU-T Rec. J.144 [5]. The f_{SI13} and f_{HV13} features measure the amount and angular distribution of spatial gradients in spatial-temporal (S-T) sub-regions of the luminance (Y) images. Thus, the f_{SI13} and f_{HV13} features are sensitive to spatial distortions such as blurring and blocking.

The f_{SI13} and f_{HV13} features of the NTIA Fast Low Bandwidth model are extracted from S-T region sizes of 30 vertical lines by 30 horizontal pixels by one second of time (i.e., 30 × 30 × 1s) whereas VQM used S-T region sizes of 8 × 8 × 0.2s. In addition, to save computations, the one second of luminance Y images are first pre-averaged across time before applying the two-dimensional 13 × 13 edge enhancement filters given in Annex D.7.2.1 of [5].

One additional spatial feature is extracted from the one second of pre-averaged luminance (Y) images. This feature is the *mean* luminance (Y) level of each 30 x 30 x 1s S-T region (denoted here as f_{MEAN}). The purpose of the f_{MEAN} feature is to provide a luminance-level perceptual weighting function for spatial information (SI) loss as measured by the f_{SI13} and f_{HV13} features. This will be described in Section 3.3.

3.2.4 Temporal Features

Powerful estimates of perceived video quality can be obtained from the color and spatial feature set described above. However, since the S-T regions from which these features are extracted span many video frames (i.e., one second of video frames), they tend to be insensitive to brief temporal disturbances in the picture. Such disturbances can result from noise or digital transmission errors; while brief in nature, they can have a significant impact on the perceived picture quality. Thus, a temporal-based RR feature is used to quantify the perceptual effects of temporal disturbances. This feature measures the absolute temporal information (ATI), or motion, in the luminance Y image plane and is computed as:

$$f_{ATI} = rms\ (rand5\%\{|Y(t) - Y(t - 0.2s)|\}) \quad (1)$$

where function *rms* computes root mean square error, and function *rand5%* retains 5% of pixels, chosen at random, and discards the other 95% of pixels. *Y* is randomly sub-sampled to contain only 5% of the image pixels for computational efficiency. The sub-sampled *Y* image at time *t*-0.2s is subtracted from the identically sub-sampled *Y* image at time *t* and the root mean square error of the result is used as a measure of ATI. Following the descriptive naming convention described in Annex D.8 of ITU-T Recommendation J.144, we denote f_{ATI} by the variable name:

$$f_{ATI} \cong Y\_rand5\%\ \_ati0.2s\_rms \quad (2)$$

Annex D.8 of ITU-T Recommendation J.144 defines the ITS video quality parameter naming convention. The goal of this naming convention is to identify the exact formula used to compute the parameter, when read from left to right. For example, “ati0.2s” is the ATI feature calculated using frames 0.2 seconds apart in time.

The feature f_{ATI} is sensitive to temporal disturbances. For 30 fps video, 0.2s is 6 video frames while for 25 fps video, 0.2s is 5 video frames. Subtracting images 0.2s apart makes the feature insensitive to real time 30 fps and 25 fps video systems that have frame update rates of at least 5 fps. The quality aspects of these low frame rate video systems, common in multimedia applications, are sufficiently captured by the f_{SI13}, f_{HV13}, and f_{COHER_COLOR} features. The 0.2s spacing is also more closely matched to the peak temporal response of humans’ visual perception than is taking the difference of two images that are one frame apart in time.

Figure 1 provides an example plot of the f_{ATI} feature for an original (solid blue) and processed (dashed red) video scene from a digital video system with transient burst errors in the digital transmission channel. Transient errors in the processed picture create spikes in the f_{ATI} feature. The bandwidth required to transmit the f_{ATI} feature is extremely low since it requires only 30 samples per second for 30 fps video. Other types of additive noise in the processed video, such as might be generated by an analog video system, will appear as a positive DC shift in the time history of the processed feature stream with respect to the original feature stream. Video coding systems that eliminate noise will cause a negative DC shift.

Before extracting a transient error parameter from the f_{ATI} feature streams shown in Figure 1, it is advantageous to increase the width of the motion spikes (red spikes in Figure 1). The reason is that short motion spikes from transient errors do not adequately represent the perceptual impact of these types of errors. One method for increasing the width of the motion spikes is to apply a maximum filter to both the original and processed feature streams before calculation of the error parameter function between the two waveforms. For the f_{ATI} based error parameter, a seven-point wide maximum filter (which will be denoted here as the *max7pt* function) was used that produces an output sample at each frame that is the maximum of itself and the three nearest neighbors on each side (i.e., earlier and later time samples).

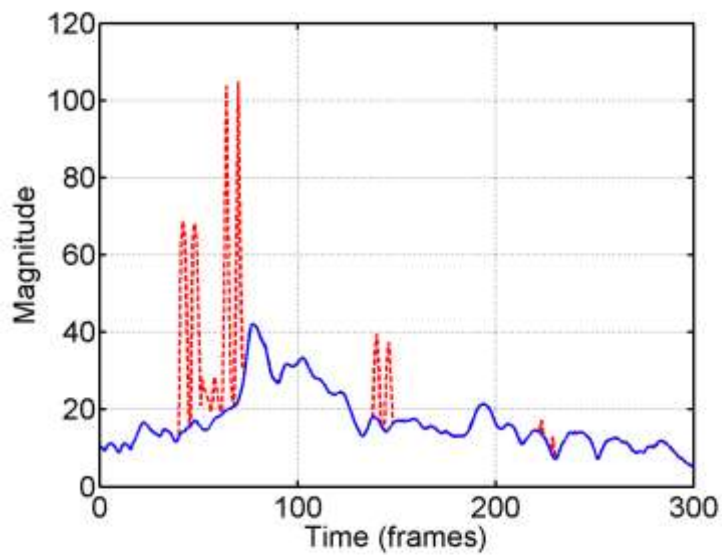

Figure 1. Example time history of f_{ATI} feature.

3.2.5 Quantization of Features

Quantization to 9 bits of accuracy is sufficient for the Y_{MEAN}, f_{SI13}, f_{HV13}, and f_{COHER_COLOR} features, while the f_{ATI} feature should be quantized to 10 bits. To have minimal effect on the video quality parameter calculations, a non-linear quantizer design should be used where the quantizer error is proportional to the magnitude of the signal being quantized. Very low values are uniformly quantized to some cutoff value, below which there is no useful quality assessment information. Such a quantizer design minimizes the error in the corresponding parameter calculations because these calculations are normally based on an error ratio or log ratio of the processed and original feature streams (see Section 3.3 Parameter Description below).

Figure 2 provides a plot of the 9 bit non-linear quantizer used for the f_{SI13} original feature. The reference code subroutine “model_lowbw_compression” provides a complete mathematical description of the recommended quantizers used by the Fast Low Bandwidth model. If the features fall outside the range of the recommended quantizers on the low or high end (highly unlikely), then the S-T parameters derived from these features are zeroed so they do not influence the overall VQM.

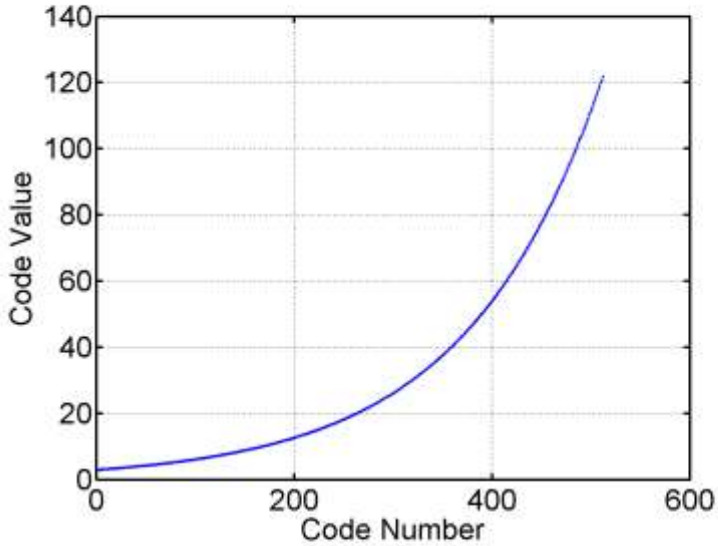

Figure 2. Non-linear 9 bit quantizer for the f_{SI13} feature.

3.3 Parameter Description

3.3.1 Parameter Overview

Several steps are involved in the calculation of parameters that track the various perceptual aspects of video quality. The steps may involve:

(1) Applying a perceptual threshold to the extracted features from each S-T region,
(2) Calculating an error function between processed features and corresponding original features, and
(3) Pooling the resultant error over space and time.

See Annex D.8 of ITU-T Recommendation J.144 for a detailed description of these techniques and their accompanying mathematical notation for parameter names, which will also be used here. The reference code subroutine "model_fastlowbw_parameters" provides a complete mathematical description of the parameters used by the Fast Low Bandwidth model. For simplicity, the description of the parameters in this section does not consider the effects of feature quantization (e.g., handling feature values that might lie outside of the recommended quantization ranges).

3.3.2 New Methods

This section will summarize new methods that have been found to improve the objective to subjective correlation of parameters based on RR features with very low transmission bandwidths such as those used for the NTIA Fast Low Bandwidth model (i.e., methods not found in the NTIA General Model from ITU-T Rec. J.144 [5]). It is worth noting that no improvements have been found for the basic form of the parameter error functions given in Annex D.8.2.1 of ITU-T Rec. J.144. The two error functions that consistently produce the best parameter results (for spatial and temporal parameters) are a logarithmic ratio {log10 [$f_p(s,t)$ / $f_o(s,t)$]} and an error ratio {[$f_p(s,t) - f_o(s,t)$] / $f_o(s,t)$}, where $f_p(s,t)$ and $f_o(s,t)$ are the processed feature and corresponding original feature extracted from the S-T region with spatial and temporal coordinates *s* and *t*, respectively. Errors must be separated into gains and losses, since humans respond differently to additive (e.g., blocking) and subtractive (e.g., blurring) impairments. Applying a lower perceptual threshold to the features before application of these two error functions prevents division by zero.

After computation of the S-T parameters using one of the error functions, the S-T parameters must be pooled over space and time to produce a parameter value for the video clip. This error pooling can occur in multiple stages (e.g., over space and then over time). One new error pooling method that is used by the Fast Low Bandwidth model is called Macro-Block (MB) error pooling. MB error pooling groups a contiguous number of S-T sub-regions and applies an error pooling function to this set. For instance, the function denoted as “MB(3,3,2)max” will perform a max function over parameter values from each group of 18 S-T sub-regions that are stacked three vertical by three horizontal by two temporal. For the 32 × 32 × 1s S-T sub-regions of the f_{SI13}, f_{HV13}, and f_{COHER_COLOR} features described above, each MB(3,3,2) region would encompass a portion of the video stream that spans 96 vertical lines by 96 horizontal pixels by two seconds of time. MB error pooling has been found to be useful in tracking the perceptual impact of impairments that are localized in space and time. Such localized impairments often dominate the quality decision process. Macro-Block error pooling can also be implemented as a filtering process so that instead of producing a single output value for each MB, each S-T sample is replaced with its MB filtered value, where the MB is centered on the S-T sample. This is called Overlapped MB (OMB) error pooling.

A second error pooling method is a generalized Minkowski(P,R) summation, defined as:

$$Minkowski(P,R) = \sqrt[R]{\frac{1}{N}\sum_{i=1}^{N}|v_i|^P} \quad (3)$$

Here v_i represents the parameter values that are included in the summation. This summation might, for instance, include all parameter values at a given instance in time (spatial pooling), or may be applied to the macro-blocks described above. The Minkowski summation where the power P is equal to the root R has been used by many developers of video quality metrics for error pooling. The generalized Minkowski summation, where $P \neq R$, provides additional flexibility for linearizing the response of individual parameters to changes in perceived quality. This is a necessary step before combining multiple parameters into a single estimate of perceived video quality, which is performed with a linear fit.

3.3.3 Color Parameters

Two parameters are extracted from the f_{COHER_COLOR} features. One of these parameters, called *color_extreme*, measures extreme color distortions that might be caused by colored blocks from transmission errors. The other parameter, called *color_spread*, provides an indication of the variance or spread in the color errors. Rather than using the Euclidean distance measure to quantify distortions (as in Annex D.8.2.2 of ITU-T Recommendation J.144), both of these parameters use the square root of the Manhattan distance. Following the mathematical notation in Annex D.8.2.2 of ITU-T Recommendation J.144 [5], where $\underline{f}_p(s,t)$ and $\underline{f}_o(s,t)$ represent the two-dimensional $\underline{f}_{COHER\_COLOR}$ feature extracted from a S-T region of the processed and original video streams, this feature comparison function is given by:

$$sqrtmanhat(s,t) = \sqrt{\sum_{C_B,C_R} \left|\underline{f}_p(s,t) - \underline{f}_o(s,t)\right|} \tag{4}$$

We have observed that the Manhattan distance measure seems to be more accurate than the Euclidean distance measure. The square root function is required to linearize the parameter's response to quality changes.

Annex D.8 of ITU-T Recommendation J.144 defines the ITS video quality parameter naming convention. The goal of this naming convention is to identify the exact formula used to compute the parameter, when read from left to right. For example, "30×30_1s" indicates a block size of 30 pixels horizontally, 30 lines vertically and 1 second in time. Following that naming convention, the color parameters are denoted by the parameter names:

$$color\_extreme = \text{color\_coher\_color\_30} \times \text{30\_1s\_mean\_sqrtmanhat\_OMB(3,3,2)above99\%\_Minkosski(0.5,1)}$$

$$color\_spread = \text{color\_coher\_color\_30} \times \text{30\_1s\_mean\_sqrtmanhat\_OMB(3,3,2)Minkosski(2,4)\_90\%.}$$

A combined color parameter (*color_comb*) that contains the optimal combination of both the *color_extreme* and *color_spread* parameters is then computed as:

$$color\_comb = 0.691686 * color\_extreme - 0.617958 * color\_spread$$

This positive valued *color_comb* parameter is then clipped at the low end, which is represented mathematically by (following the notation in Annex D.8.5 of ITU-T Recommendation J.144):

$$color\_comb = color\_comb\_clip\_0.114.$$

This *color_comb* parameter is included in the linear combination for the VQM calculation.

3.3.4 Spatial Parameters

Two spatial parameters are computed from the f_{SI13} feature, one that measures a loss of spatial information (*si_loss*) and one that measures a gain of spatial information (*si_gain*). Following the mathematical notation in Annex D.8 of ITU-T Recommendation J.144, these parameters are given by:

$$si\_loss = \text{avg1s\_Y\_si13\_30×30\_std\_3\_ratio\_loss\_OMB(3,3,2)Minkosski(1,2)\_Minkosski(1.5,2.5)\_clip\_0.12}$$

$$si\_gain = \text{avg1s\_Y\_si13\_30×30\_std\_3\_log\_gain\_clip\_0.1\_above95\%tail\_Minkosski(1.5,2)}.$$

As the mean luminance (Y) level of the S-T sub-region increases (i.e., as measured by the f_{MEAN} feature), the ability to perceive changes in spatial detail (e.g., such as blurring measured by *si_loss*) decreases. This can be accommodated by introducing a weighting function (*Y_weight*) as shown in Figure 3 to the *si_loss* values from each S-T sub-region (i.e., the *si_loss* values after the ratio loss comparison function is performed on each S-T sub-region but before the spatial and temporal collapsing functions). The weighting function *Y_weight* is equal to one (i.e., full weighting) until an average luminance level of 175 is reached and then it decreases linearly to zero as the luminance values increase from 175 to 255. This intermediate correction is applied only to the *si_loss* values, not the *si_gain* values.

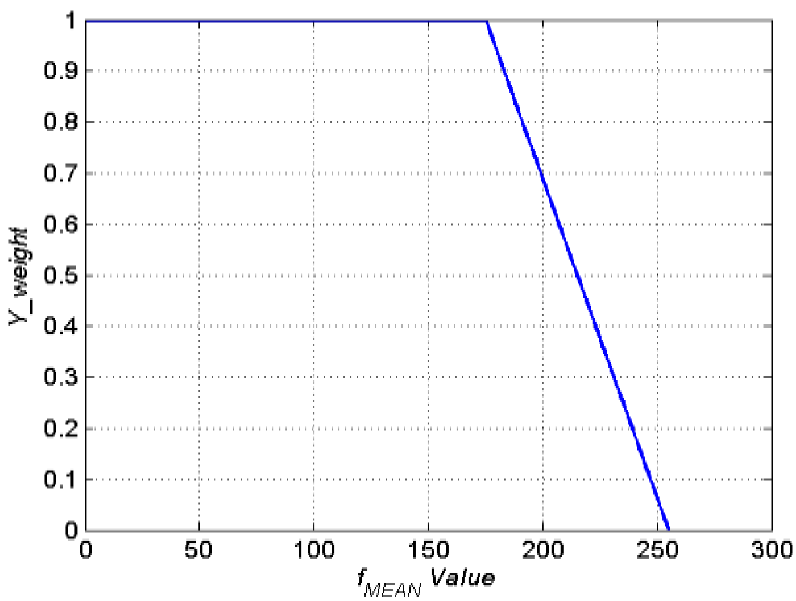

Figure 3. Weighting function *Y_weight* for modifying *si_loss* S-T parameters.

Two spatial parameters are computed from the f_{HV13} feature, one that measures a loss of relative Horizontal and Vertical (HV) spatial information (*hv_loss*) and one that measures a gain (*hv_gain*). Following the mathematical notation in Annex D.8 of ITU-T Recommendation J.144, these parameters are denoted by the parameter names:

$$hv\_loss = \text{avgls\_Y\_hv13\_angle0.225\_rmin20\_30×30\_mean\_4\_ratio\_loss\_...}$$
$$\text{OMB(3,3,2)below1\%\_Minkosski(1,1.5)\_clip\_0.08}$$

$$hv\_gain = \text{avgls\_Y\_hv13\_angle0.225\_rmin20\_30×30\_mean\_4\_log\_gain\_...}$$
$$\text{clip\_0.06\_OMB(3,3,2)above99\%tail\_Minkosski(1.5,3)\_}.$$

Not shown in the above equations is that the *Y_weight* function shown in Figure 3 is also applied to both the *hv_loss* and *hv_gain* values from each S-T sub-region before the spatial and temporal collapsing functions (after the ratio_loss and log_gain computations, respectively). An additional weighting function (*SI_weight* as shown in Figure 4) is applied to the *hv_loss* values from each S-T sub-region. This is necessary to reduce the sensitivity of *hv_loss* for S-T regions that have very little spatial information (i.e., low f_{SI13} original feature values).

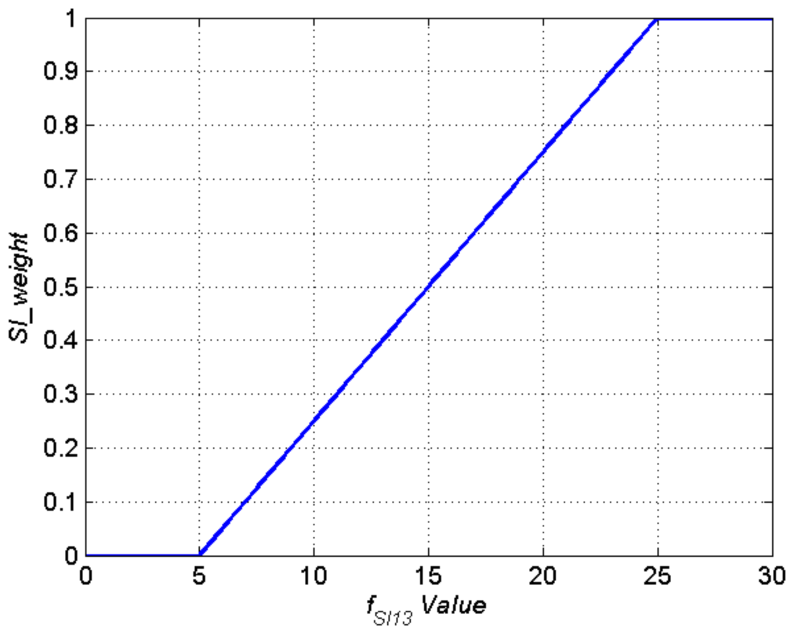

Figure 4. Weighting function *SI_weight* for modifying *hv_loss* S-T parameters.

The spatial distortion parameters can be crushed (i.e., excessive excursions beyond those found in the training data are limited or compressed) using functions like the VQM crushing function found in the VQM Calculation Section.

3.3.5 Temporal Parameters

Two temporal parameters are computed from the f_{ATI} feature, one that measures added random noise (*ati_noise*) and one that measures motion disturbances caused by transmission errors (*ati_error*). Following the mathematical notation in Annex D.8 of ITU-T Recommendation J.144, these parameters are denoted by the parameter names:

$$ati\_noise = Y\text{\_rand5\%\_ati0.2s\_rms\_5\_ratio\_gain\_between25\%50\%}$$

$$ati\_error = Y\text{\_rand5\%\_ati0.2s\_rms\_max7pt\_12\_ratio\_gain\_above90\%.}$$

To make the *ati_noise* and *ati_error* parameters more robust against temporal misalignments, the parameters are computed for all temporal alignments of the processed video that are within plus or minus 0.4 seconds of the best estimated temporal alignment to the original video, and then the minimum parameter value is selected.

3.4 Model Calculation

Similar to VQM in Annex D of ITU-T Recommendation J.144, the Fast Low Bandwidth model calculation linearly combines two parameters from the f_{HV13} feature (*hv_loss* and *hv_gain*), two parameters from the f_{SI13} feature (*si_loss* and *si_gain*), and two parameters from the f_{COHER_COLOR} feature (except that the two parameters have been combined into a single color distortion parameter called *color_comb*). The one noise parameter in VQM has been replaced with two parameters based on the low bandwidth f_{ATI} feature described here (*ati_noise* and *ati_error*).

Thus, FLB (Fast Low Bandwidth model) consists of a linear combination of eight parameters. FLB is given by:

FLB = { 0.38317338378290 * *hv_loss* + 0.37313218013131 * *hv_gain* +

0.58033514546526 * *si_loss* + 0.95845512360511 * *si_gain* +

1.07581708014998 * *color_comb* +

0.17693274495002 * *ati_noise* + 0.02535903906351 * *ati_error* }

The total (after the contributions of all the parameters are added up) is clipped at a lower threshold of 0.0 to prevent negative numbers. Finally, a crushing function that allows a maximum of 50% overshoot is applied to values over 1.0 to limit values for excessively distorted video that falls outside the range of the training data.

If FLB > 1.0, then FLB = (1 + c)*FLB / (c + FLB), where c = 0.5.

FLB computed in the above manner will have values greater than or equal to zero and a nominal maximum value of one. FLB may occasionally exceed one for video scenes that are extremely distorted.

To make FLB more robust against spatial misalignments, FLB is computed for all spatial alignments of the processed video that are within plus or minus one pixel of the best estimated spatial alignment to the original video, and then the minimum FLB is selected.

4 CONCLUSION

The NTIA Fast Low Bandwidth model was optimized for in-service operation over a reduced reference communications link. The RR link is required for access to the original video or a very high quality version of the video. The original video acts as "truth data" (i.e., what the processed video ought to look like). The parameter selection includes compromises between accuracy and the reduced bandwidth requirements for the communications link between the original and processed video. In practical terms, this means that the Fast Low Bandwidth model is optimal only when the user has access to both the original and processed video as part of a live system.

5 REFERENCES

[1] ITU-T Rec. J.249 (01/10), "Perceptual video quality measurement techniques for digital cable television in the presence of a reduced reference."

[2] ITU-T Recommendation J.244 (04/08), "Full reference and reduced reference calibration methods for video transmission systems with constant misalignment of spatial and temporal domains with constant gain and offset."

[3] M. H. Pinson and S. Wolf, "Reduced reference video calibration algorithms," NTIA Technical Report TR-06-433b, Nov. 2007. Available: www.its.bldrdoc.gov.

[4] Video Quality Model (VQM) Software Tools - Binary Executables and Source Code, available from the National Telecommunications and Information Administration (NTIA) at http://www.its.bldrdoc.gov/n3/video/VQM_software.php.

[5] ITU-T Recommendation J.144 (03/04), "Objective Perceptual Video Quality Measurement Techniques for Digital Cable Television in the Presence of a Full Reference."

[6] ITU-R Recommendation BT.1683 (06/04), "Objective Perceptual Video Quality Measurement Techniques for Standard Definition Digital Broadcast Television in the Presence of a Full Reference."

[7] M Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Transactions on Broadcasting*, vol.50, no.3, pp. 312-322, Sept. 2004.

[8] S. Wolf and M. Pinson, "Video Quality Measurement Techniques," NTIA Technical Report TR-02-392, Jun. 2002.

[9] ITU-R Recommendation BT.601-6 (01/07), "Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios."

[10] T. H. Cormen, C. E Leiserson, and R.L. Rivest, *Introduction to Algorithms*, MIT Press and McGraw-Hill, 1997, pp. 161-163.

[11] VQEG Final Report of RRNR-TV Validation Test (2009), "Validation of reduced reference and no-reference objective models for standard definition television, phase 1", Video Quality Experts Group (VQEG), http://www.its.bldrdoc.gov/vqeg/projects/multimedia, ITU-T Study Group 9 TD923, Study Period 2005-2008.

FORM **NTIA-29**
(4-80)

U.S. DEPARTMENT OF COMMERCE
NAT'L. TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION

BIBLIOGRAPHIC DATA SHEET

1. PUBLICATION NO. TM-13-497	2. Government Accession No.	3. Recipient's Accession No.
4. TITLE AND SUBTITLE Fast Low Bandwidth Model: A Reduced Reference Video Quality Metric		5. Publication Date June 2013 6. Performing Organization Code NTIA/ITS.D
7. AUTHOR(S) Margaret H. Pinson and Stephen Wolf		9. Project/Task/Work Unit No. 3141011.300
8. PERFORMING ORGANIZATION NAME AND ADDRESS Institute for Telecommunication Sciences National Telecommunications & Information Administration U.S. Department of Commerce 325 Broadway Boulder, CO 80305		10. Contract/Grant Number.
11. Sponsoring Organization Name and Address National Telecommunications & Information Administration Herbert C. Hoover Building 14^{th} & Constitution Ave., NW Washington, DC 20230		12. Type of Report and Period Covered
14. SUPPLEMENTARY NOTES		

15. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)

This memorandum describes the Fast Low Bandwidth model. This summary is intended to help the reader understand the model from an algorithmic standpoint. Some knowledge of prior NTIA objective video quality models is necessary for the understanding of this document.

The Fast Low Bandwidth was designed to be operated in-service using the reduced reference (RR) methodology. The model requires access to the original video at one location, the processed video at another location, and low bandwidth data link between the two locations. That link is used to communicate RR features between the two locations. The Fast Low Bandwidth model is included in ITU-T Rec. J.249.

16. Key Words (Alphabetical order, separated by semicolons)

video; quality; models; metrics; features; parameters; objective; subjective; correlation; reduced-reference; spatial information (SI); temporal information (TI); impairments

17. AVAILABILITY STATEMENT	18. Security Class. (This report)	20. Number of pages
☒ UNLIMITED.	Unclassified	26
☐ FOR OFFICIAL DISTRIBUTION.	19. Security Class. (This page) Unclassified	21. Price:

NTIA FORMAL PUBLICATION SERIES

NTIA MONOGRAPH (MG)

A scholarly, professionally oriented publication dealing with state-of-the-art research or an authoritative treatment of a broad area. Expected to have long-lasting value.

NTIA SPECIAL PUBLICATION (SP)

Conference proceedings, bibliographies, selected speeches, course and instructional materials, directories, and major studies mandated by Congress.

NTIA REPORT (TR)

Important contributions to existing knowledge of less breadth than a monograph, such as results of completed projects and major activities. Subsets of this series include:

NTIA RESTRICTED REPORT (RR)

Contributions that are limited in distribution because of national security classification or Departmental constraints.

NTIA CONTRACTOR REPORT (CR)

Information generated under an NTIA contract or grant, written by the contractor, and considered an important contribution to existing knowledge.

JOINT NTIA/OTHER-AGENCY REPORT (JR)

This report receives both local NTIA and other agency review. Both agencies' logos and report series numbering appear on the cover.

NTIA SOFTWARE & DATA PRODUCTS (SD)

Software such as programs, test data, and sound/video files. This series can be used to transfer technology to U.S. industry.

NTIA HANDBOOK (HB)

Information pertaining to technical procedures, reference and data guides, and formal user's manuals that are expected to be pertinent for a long time.

NTIA TECHNICAL MEMORANDUM (TM)

Technical information typically of less breadth than an NTIA Report. The series includes data, preliminary project results, and information for a specific, limited audience.

For information about NTIA publications, contact the NTIA/ITS Technical Publications Office at 325 Broadway, Boulder, CO, 80305 Tel. (303) 497-3572 or e-mail info@its.bldrdoc.gov.

This report is for sale by the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161,Tel. (800) 553-6847.