NTIA Technical Memorandum TM-11-482

Video Quality Model for Variable Frame Delay (VQM_VFD)

Stephen Wolf
Margaret H. Pinson

technical memorandum

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

NTIA Technical Memorandum TM-11-482

Video Quality Model for Variable Frame Delay (VQM_VFD)

Stephen Wolf
Margaret H. Pinson

U.S. DEPARTMENT OF COMMERCE
September 2011

DISCLAIMER

Certain commercial software is 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 (NTIA), nor does it imply that the software identified is necessarily the best available for the particular application or use.

This document contains 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.** 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.

CONTENTS

Page

VIDEO QUALITY MODEL FOR VARIABLE FRAME DELAY (VQM_VFD)

Stephen Wolf and Margaret H. Pinson [1]

Time varying delays of the output (or processed) video frames with respect to the input (i.e., the original or reference) video frames present significant challenges for Full Reference (FR) video quality measurement systems. Time alignment errors between the output video sequence and the input video sequence can produce measurement errors that greatly exceed the perceptual impact of these time varying video delays. This document proposes a new video quality model (VQM) that properly accounts for the perceptual impact of variable frame delay (VFD). This new model, called VQM_VFD, also uses perceptual features extracted from spatial-temporal (ST) blocks of a fixed angular extent. This enables VQM_VFD to track subjective quality over a wide range of viewing distances and image sizes. VQM_VFD uses a neural network that achieves 0.9 correlation to subjective quality for subjective datasets at image sizes from Quarter Common Intermediate Format (QCIF) to High Definition TV (HDTV).

The model described in this memorandum uses algorithms from NTIA Technical Report 02-392, “Video quality measurement techniques,” NTIA Technical Memorandum TM-10-463, “A full reference (FR) method using causality processing for estimating variable video delays” and NTIA Technical Memorandum TM-11-475, “Variable frame delay (VFD) parameters for video quality measurements.”

Key words: alignment; angular; calibration; correlation; delay; distance; dropped; frames; Full Reference (FR); measurement; model; objective; parameters; pausing; quality; resolution; skipping; subjective; time; variable; video

1. INTRODUCTION

Digital video transmission systems normally consist of a video encoder, a digital transmission method (e.g., Internet Protocol—IP), and a video decoder. Video frames passing through these systems can be dropped and/or subject to variable time delays. As a result, the presentation of the video to the end user may contain unnatural or jerky motion, pauses or frame freezes, and fast forwards or missing segments. Reasons for these behaviors include:

---

[1] The authors are with the Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, 325 Broadway, Boulder, CO 80305.

- The video encoder may intelligently decide to reduce the video frame transmission rate in order to save bits.
- The video decoder may decide to freeze the last good video frame when:
    - Errors such as IP packet loss are detected.
    - Packets do not arrive on time, such as when the network drops packets or the delay through the network increases.
    - The video decoder cannot play the video at full frame rate.
- The video decoder may decide to skip forward, fast forward or even rewind when:
    - Errors are detected, such as IP packet loss.
    - The video decoder runs out of buffer space.

Whatever the reason, output video frames from modern video compression and transmission systems can contain significant time varying video delays and a Full Reference (FR) video quality model (VQM) must properly deal with these idiosyncrasies.

Time varying delays of the output (i.e., processed) video frames with respect to the input (i.e., original, or reference) video frames present a challenge for perception-based FR VQMs. This is because the time alignment errors between the output video sequence and the input video sequence can produce measurement errors that greatly exceed their impact on subjective video quality. For example, a one-frame video freeze without skipping will result in either the prior or later output segment being shifted by one video frame with respect to the original reference segment. While this is barely noticeable to viewers, the commonly used Peak-Signal-to-Noise-Ratio (PSNR) measurement [1] will detect a large impairment for the output video segment that is off by one video frame with respect to the original video segment.

Reference [2] describes an FR technique for estimating variable frame delay (VFD) information. Reference [3] goes further by proposing several perception-based quality parameters that can be extracted from VFD information. This document goes further yet and proposes a complete set of perception-based VFD quality parameters (Section 2) and a neural network for mapping these parameters to subjective quality (Section 3). The results showing the correlation between the subjective results and the objective results (obtained using the entire video quality model for variable frame delay, or VQM_VFD) are given in Section 4. Section 5 presents ideas for future work on improving VQM_VFD.

Figure 1 presents a high-level overview of the VQM_VFD system. The processed video is calibrated to remove gain and level offsets, spatial shifts and spatial scaling (if present). Next, the VFD information (i.e., the best matching original frame for every processed frame) is calculated and used to change the order of frames in the original video sequence so it matches the order of frames in the processed video sequence (i.e., VFD-matched original video). For instance, if the processed video sequence repeated every other frame, then the original sequence would match this behavior. The VFD information generated from this step, the calibrated processed video, and the VFD-matched original video are then used for perception-based VFD parameter extraction. Since all of the perceptual effects of variable frame delays (pauses, skips, etc.) have been removed from the VFD-matched original video (which will be used as the reference video for an FR VQM), quality parameters are also extracted from the VFD information itself. A neural network is used to map all of the perception-based VFD-quality parameters to a video-quality estimate.

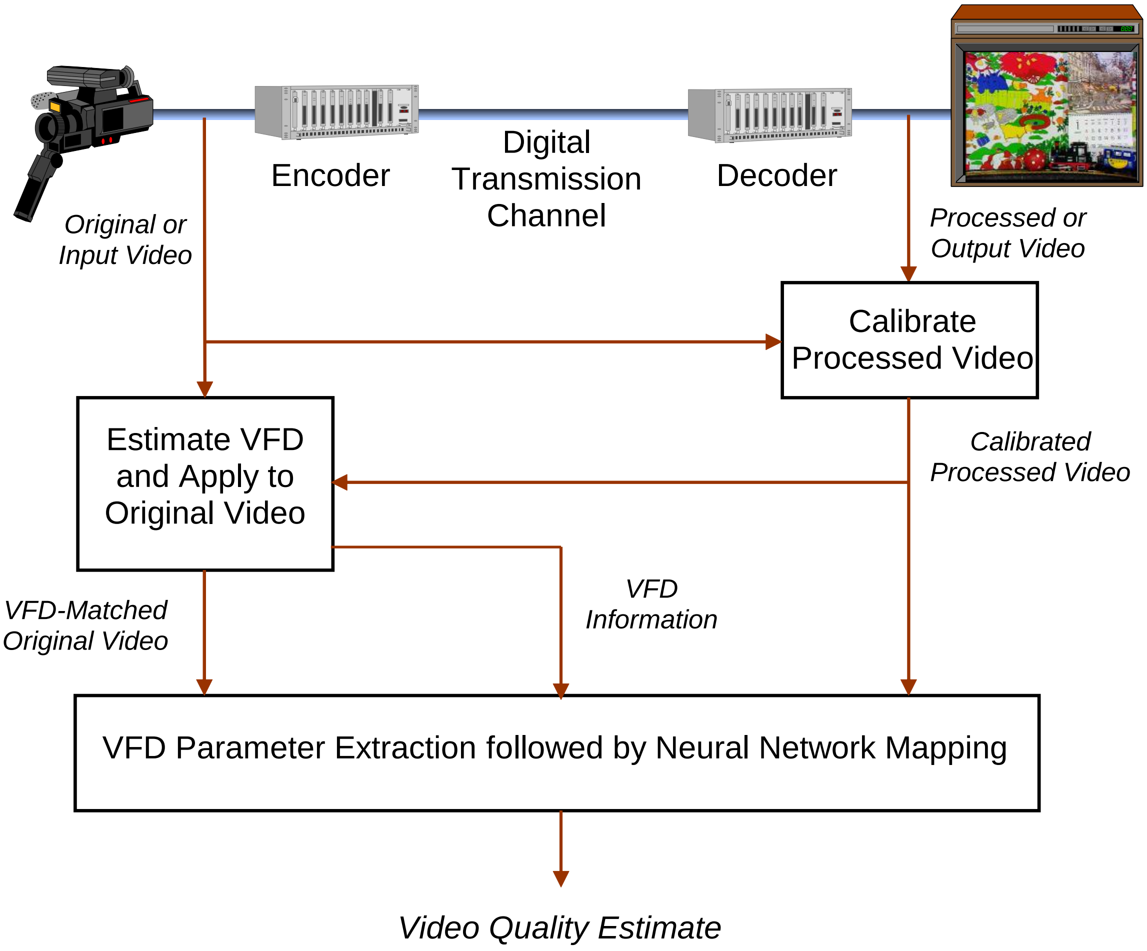


Figure 1. Schematic overview of VQM_VFD.

2. PERCEPTION-BASED VFD PARAMETERS

Features and parameters extracted from spatial-temporal (ST) blocks, like those shown in Figure 2, have been used to design VQMs with excellent correlation to subjective quality [4] [5]. A feature is a quantity of information associated with, or extracted from, an ST block. A parameter is a measure of video distortion that is the result of comparing two parallel streams of features, one stream from the original video and the corresponding stream from the processed video.

For standard definition video displayed at a viewing distance (*D*) of 4 to 6 times the picture height (*H*), studies have been performed to determine optimal ST block sizes for quality assessment purposes [6]. The picture height *H* is the total vertical extent of the video image. Controlled subjective tests normally fix the viewing distance as a multiple of the picture height. The optimal viewing distance *D* uses the full resolving power of the human eye but is not so close that individual pixels are discernable. Commonly selected values of *D* range from 3 (for HDTV) to 8 (for QCIF).

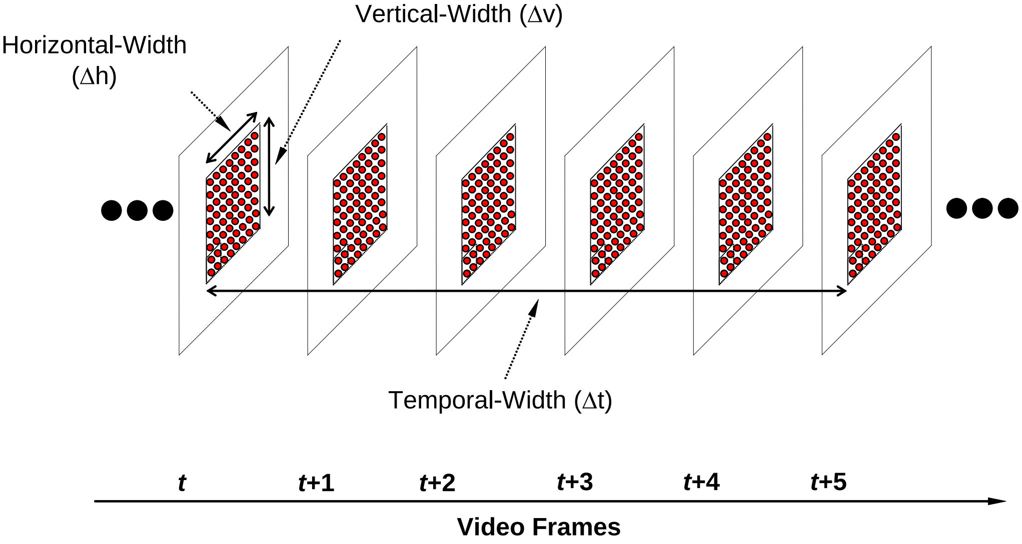


Figure 2. Image pixels contained within an ST block of size Δh x Δv x Δt.

To produce ST block parameters that are based on viewing physics, one must fix the angular extent θ subtended by the ST block at the eye. Using a fixed angular extent θ to determine ST block size allows one to design a VQM that works over a range of image sizes and viewing distances. The angular extent θ is chosen to realize the full potential of the extracted features and parameters (i.e., their correlation to subjective quality), while not being so small that the extracted statistics become unstable. In this respect, there seems to be little value to using an angular extent θ smaller than about 0.2 to 0.4 degrees [6]. Thus, VQM_VFD uses 0.4 degrees.

Assuming square blocks and the small angle approximation, the vertical and horizontal extent Δ of the ST blocks (in pixels) is given by

$$\Delta h = \Delta v = N_V \cdot D \cdot \theta \cdot \frac{\pi}{180} \text{ pixels,} \qquad (1)$$

where N_V is the total number of image pixels in the vertical direction, D is the viewing distance (in picture heights H), and θ is the angular extent of the ST blocks (in degrees).

Regarding the temporal extent of the ST block (Δt), [6] also demonstrates that there is little advantage to using less than about 0.2 seconds of video. Correlation to subjective quality tends to drop when smaller temporal extents are used, as the human visual system requires a sequence of video frames to form a quality opinion. VQM_VFD thus uses 0.2 seconds for the temporal extent of the ST blocks.

VQM_VFD uses eight quality parameters that are described in Sections 2.1 to 2.8. The parameter-naming conventions found in Section 5.6 of [4] have been adopted here, and the reader will have to reference the prior technical report to fully understand the parameters that are described herein.

2.1. HV_Loss

The *HV_Loss* parameter of the VQM_VFD model was derived from the one used by the General VQM described in detail in Sections 5.6.1 and 6.3 of [4], with four differences:

1. The VFD correction is performed on the luma Y-channel original video before feature extraction. Thus, there will be a *vfd* field added to the parameter name following *Y*.
2. Unique horizontal (H) and vertical (V) edge detection filters are used rather than the fixed-sized 13 x 13 filter described in Section 4.2.1 of [4]. The unique filter changes according to image size. This feature extraction will be denoted by a field called *hvA* in the parameter name (rather than the field *hv13* that is used for describing the general VQM *HV_Loss* parameter in Section 4.2.2 of [4]). The *hvA* filter will be described in Section 2.1.1.
3. ST block size is given by an angular extent (0.4 degrees) and a time duration (0.2 seconds) instead of rigidly specified in pixels and frames. This will be denoted by a field called *0.4deg_0.2s* in the parameter name rather than the field called *8x8_6F* that is used for describing the general VQM *HV_Loss* parameter.
4. The ST block parameters are weighted in a quadratic fashion by the product of the mean luma Y-channel level of the block and the root mean square (*rms*) of the absolute temporal information level of the block. This ST block weighting is performed after the block comparison function (e.g., *ratio_loss*) but before the spatial and temporal error pooling functions (e.g., *below5%_mean*). The exact form of this weighting will be described in Section 2.1.2.

The *HV_Loss* parameter of the VQM_VFD model is given by:

$$HV\_Loss = Y\_vfd\_hvA\_0.4\deg\_0.2s\_mean\_3\_ratio\_loss\_below5\%\_mean\_square\_clip\_0.06\,. \qquad (2)$$

A description of our parameter naming convention can be found in Section 5.6 of [4].

In (2), the *HV* feature angle (0.225 radians) and minimum radius (20) shown in Figure 28 of [4] have been left out of the parameter name but they are identical to what was used for the *HV_Loss* parameter of the General VQM. For brevity, the luma and motion weighting functions described in Section 2.1.2 have also been left out of (2).

The following two subsections contain details on the *hva* filter and the *HV_Loss* weighting function.

2.1.1. *hva* Filter Description

The *hv13* feature uses two filter masks, each 13 x 13 pixels (see Section 4.2.1 of [4]). One is created by line replicating a bandpass filter 13 times (to detect vertical edges); the other is created by row replicating the same bandpass filter 13 times (to detect horizontal edges). The bandpass filter's coefficients are calculated using the equation:

$$w_x = k \cdot \left(\frac{x}{c}\right) \cdot \exp\left\{-\left(\frac{1}{2}\right)\left(\frac{x}{c}\right)^2\right\}, \qquad (3)$$

where x is the pixel displacement from the center of the filter (0, 1, 2, …, N), c is a constant that sets the width of the bandpass filter, and k is a normalization constant selected such that each filter would produce the same gain as a true Sobel filter.

Model VQM_VFD uses an a version of the *hv13* filters where the filter width adjusts, depending upon the image size. The filter widths were chosen empirically (see Table 1) to maximize correlation to subjective quality ratings.

Table 1. Bandpass Filter Weights for *hvA* Feature

Image Size	Filter Width	Filter Weights
High Definition (HD), Standard Definition (SD), VGA	13	[-0.0052625, -0.0173446, -0.0427401, -0.0768961, -0.0957739, -0.0696751, 0.0, 0.0696751, 0.0957739, 0.0768961, 0.0427401, 0.0173446, 0.0052625]
QVGA, CIF & SIF	9	[-0.0117050, -0.0628708, -0.1710348, -0.1988339, 0.0, 0.1988339, 0.1710348, 0.0628708, 0.0117050]
QCIF & QSIF	5	[-0.0512422, -0.7487578, 0.0, 0.7487578, 0.0512422]

2.1.2. HV_Loss Weighting Functions

The *HV_Loss* parameter of the General VQM was oversensitive to impairments for scenes with low and high luma levels and low and high motion levels. This suggested that a quadratic weighting function was required to de-weight those ST blocks that contained low and high luma levels and/or low and high motion levels. To compute the luma level of a processed ST block, the mean of the luma Y-channel was used. This feature is similar to the f_{CONT} feature in Section 4.4 of [4] except that the *mean* statistic is used rather than the standard deviation statistic (*std*). To compute the motion or temporal information (TI) of an ST block, the *rms* of the difference between successive frames was used. This feature is similar to the f_{ATI} feature in Section 4.5 of [4] except that the *rms* statistic is used rather than *std*.

The quadratic weighting shown in Figure 3 is first computed using the luma values of each ST block. These values range between 0 and 255. The weighting function is symmetric about a luma value of C2, where the corresponding weight is 1.0. At luma values of 0 and 2·C2, the weighting function falls to C1. The weighting function is limited so it can never fall below C3. For the quadratic luma weighting function, C1 = 0.64, C2 = 100, and C3 = 0.40 as shown in Figure 3. A similar weighting function is used for motion, except C1 = 0.75, C2 = 23, and C3 = 0.3. After the separate luma and motion weighting functions are computed, they are multiplied to obtain a composite weighting for the ST block. This composite weighting is applied after the ST block parameter calculation but before the spatial and temporal error pooling functions.

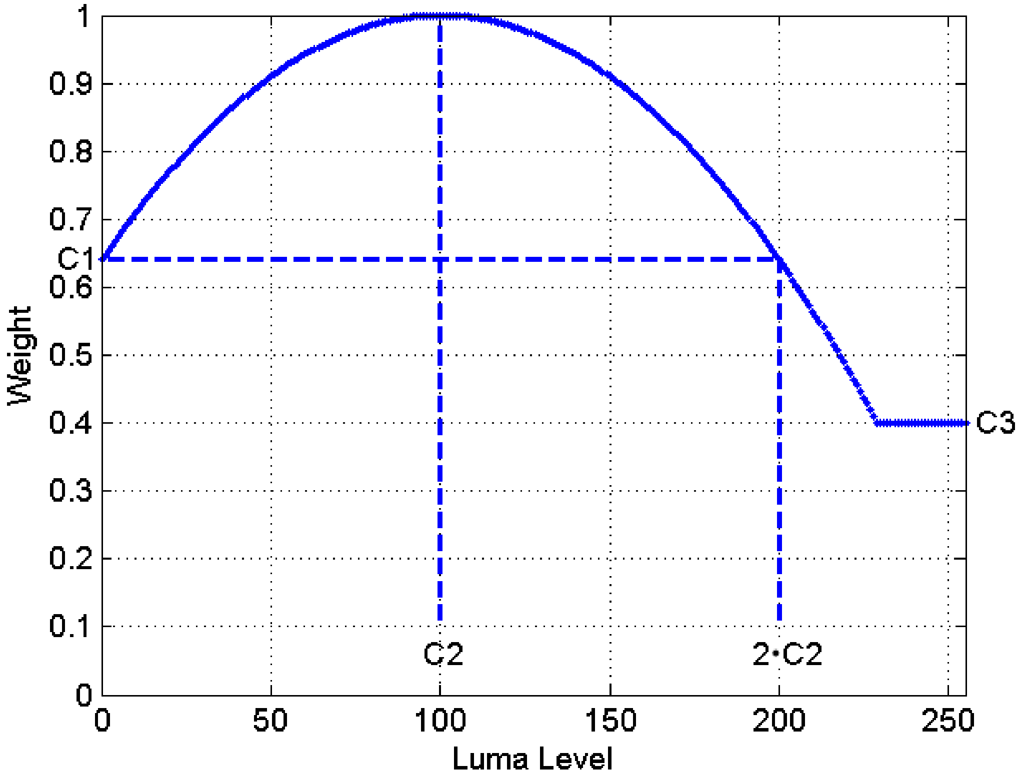


Figure 3. Quadratic weighting function for luma level.

2.2. HV_Gain

The *HV_Gain* parameter of the VQM_VFD model is similar to the one used by the General VQM described in detail in Section 6.3 of [4]. This *HV_Gain* parameter uses different spatial and temporal error pooling functions and has the differences listed in items 1–3, as described in Section 2.1. The *HV_Gain* parameter is given by:

$$HV\_Gain = Y\_vfd\_hvA\_0.4\deg\_0.2s\_mean\_3\_\log\_gain\_rms\_rms\,. \tag{4}$$

In (4), the *rms* spatial and temporal error pooling functions enhance the sensitivity of the *HV_Gain* parameter to HV edge coding noise.

2.3. SI_Loss

The *SI_Loss* parameter of the VQM_VFD model is similar to the one used by the General VQM described in detail in Section 6.3 of [4]. This *SI_Loss* parameter uses different spatial and temporal error pooling functions and has the differences listed in items 1–3, as described in Section 2.1. The *SI_Loss* parameter is given by:

$$SI\_Loss = Y\_vfd\_siA\_0.4\deg\_0.2s\_std\_12\_ratio\_loss\_mean\_above90\%\,. \tag{5}$$

In (5), the *siA* parameter field name denotes the use of the same edge filters as *hvA*, but the spatial information (SI) magnitude of the edge is used (i.e., the square root of the sum of the squares of the vertical and horizontal filter responses is the magnitude).

2.4. SI_Gain

The *SI_Gain* parameter of the VQM_VFD model is similar to the one used by the General VQM described in detail in Section 6.3 of [4]. This *SI_Gain* parameter uses different spatial and temporal error pooling functions and has the differences listed in items 1–3, as described in Section 2.1. The *SI_Gain* parameter is given by:

$$SI\_Gain = Y\_vfd\_siA\_0.4\deg\_0.2s\_std\_8\_\log\_gain\_above98\%tail\_rms\,. \tag{6}$$

In (6), the *above98%tail* spatial and *rms* temporal error pooling functions enhance the sensitivity of the *SI_Gain* parameter to transient added edges in the picture.

2.5. TI_Gain

The *TI_Gain* parameter uses the same feature described in Section 2.1 to compute the *rms* motion energy or temporal information of an ST block. Since the original video is VFD-matched to the processed clip, the *TI_Gain* parameter as calculated here does not have a large sensitivity to dropped or repeated frames, since these are compensated/corrected for by the VFD matching process. Rather, the *TI_Gain* parameter measures added transient distortions in the processed video (such as error blocks) that are not compensated for by the VFD correction. The *TI_Gain* parameter is given by:

$$TI\_Gain = Y\_vfd\_ti\_0.4\deg\_0.2s\_rms\_3\_\log\_gain\_STabove95\%tail\ . \qquad (7)$$

In (7), *STabove95%tail* is a three dimensional (3D) spatial-temporal error pooling function that is used to enhance the sensitivity of the *TI_Gain* parameter to transient-added errors in the picture.

2.6. RMSE_Gain

The *RMSE_Gain* parameter is a full reference parameter that is calculated as the *rms* error between ST blocks in the processed clip and the VFD-matched original clip. The *RMSE_Gain* parameter is given by:

$$RMSE\_Gain = Y\_vfd\_fr\_0.4\deg\_0.2s\_rms\_STmean\ . \qquad (8)$$

In (8), *fr* denotes a full reference calculation, which is the difference between the processed pixels and the original pixels. The *rms* calculation is performed over each ST block, and then a 3D mean is calculated over space and time (*STmean*). Thus, *RMSE_Gain* is the average *rms* error of the individual ST blocks.

2.7. VFD_Par1

All of the parameters in Sections 2.1 to 2.6 use an original video clip that has been VFD-matched to the processed video clip. None of these parameters can quantify the perceptual impact of repeated or dropped frames. *VFD_Par1* captures the perceptual impact of repeated or dropped frames and variable video delays using the VFD information shown in Figure 1. *VFD_Par1* is fully documented in [3]. Essentially, *VFD_Par1* captures abnormal frame jumps that differ from the normal progression of video frames over time. Longer frame freezes produce a larger abnormal frame jump when the video is finally updated.

2.8. VFD_Par1·PSNR_VFD

When VFD is present in the processed video clip, PSNR has issues that limit its usefulness for tracking subjective ratings (see Section 1). PSNR calculated after the original clip is VFD-matched to the processed clip (*PSNR_VFD*) overcomes these issues [3]. But *PSNR_VFD* does not impose any penalties for dropped or repeated frames and variable video delays. A parameter that is computed as the product of *VFD_Par1* and *PSNR_VFD* (i.e., *VFD_Par1·PSNR_VFD*) captures the perceptual attributes of both PSNR and VFD.

3. NEURAL NETWORK MAPPING

The eight objective video quality parameters described in Sections 2.1 to 2.8 must be mapped to subjective quality estimates. The data available to determine this output mapping included 83 subjectively rated datasets with a total of 11255 processed video clips at five image sizes: Quarter Common Intermediate Format (QCIF), Common Intermediate Format (CIF), Video Graphics Array (VGA), Standard Definition (SD), and High Definition (HD). The Iterative Nested Least Squares Algorithm (INLSA) was used to map the subjective scores onto the nominal (0, 1) common scale [7]. This enabled the combined dataset to be used for developing and testing the output mapping.

Due to the abundance of data, a neural network (NN) was chosen to perform the output mapping. The MATLAB® NN training tool (*nntraintool*) was used to train and test the NN. Data was randomly divided into 70% training and 30% testing (the default values used by *nntraintool*) . Figure 4 shows a screen snapshot of the upper portion of the *nntraintool* window, which gives the configuration of the NN that was used to perform the output mapping.

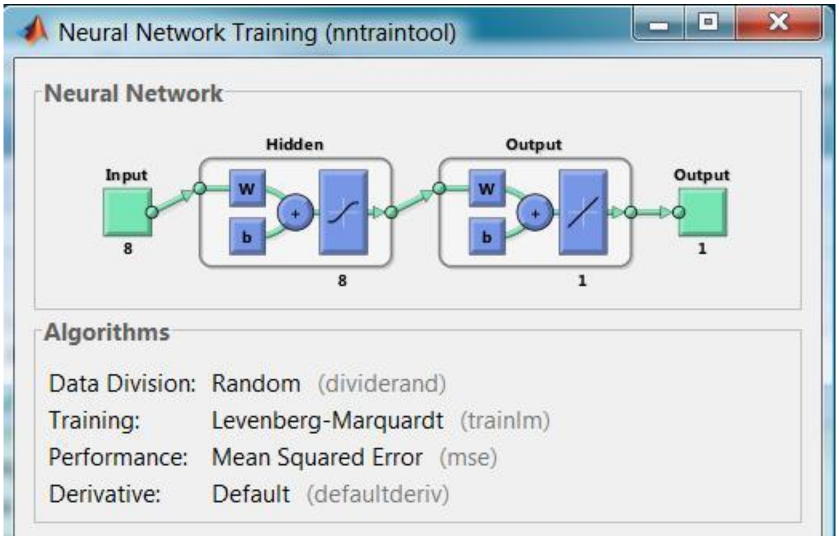
The eight-parameter input vector is multiplied by an 8 x 8 weighting matrix, which is added to a bias vector, and sent to a hidden layer consisting of 8 tan-sigmoid (*tansig*) neurons.[2] The outputs of these 8 *tansig* neurons are in turn weighted, summed together with a bias, and sent to a pure-linear (*purelin*) output neuron. There are thus 72 weights and 9 biases in the NN, for a total of 81 free parameters, which are determined in the training phase. A *tansig*/*purelin* NN was chosen because for its ability to act as a generalized function approximator.

Since there were a total of 11255 clips and 70% of the data was used for training, the training ratio (i.e., number of training samples per free parameter) was 0.7 times 11255 divided by 81, which is 97. This high training ratio produced a robust NN with testing performance that was essentially identical to the training performance (see Section 4).

The output of the NN in Figure 4 did not include a clipper or soft limiter function (e.g., see Sections 6.1 to 6.4 of [4]) to limit excursions beyond the nominal (0, 1) output range. The addition of such a function would be an extra safety precaution should the NN be presented with an input vector that falls outside of the quality space spanned by the training data.

---

[2] The *tansig* transfer function is given by: $tansig(n) = -1 + 2 / (1+\exp(-2n))$.



Data Division: Random (dividerand)
Training: Levenberg-Marquardt (trainlm)
Performance: Mean Squared Error (mse)
Derivative: Default (defaultderiv)

Figure 4. MATLAB® NN training tool configuration.

4. CORRELATION RESULTS OF VQM_VFD

This section presents the performance of the VQM_VFD model. There are many different ways to measure the performance of a VQM model. Here, we will simply present subjective vs. objective scatter plots and Pearson correlation coefficients. The interpretation of the Pearson correlation coefficient (denoted here as ρ) is that its square (i.e., ρ^2) provides an estimate of the subjective mean opinion score (MOS) variance that is explained by the objective model. The scatter plot provides a graphical means to display the tightness of the relationship and the location of outliers.

Performance results of the VQM_VFD model are presented separately for each dataset (QCIF, CIF, VGA, SD, and HD) in Figure 5. The Pearson correlation coefficients for QCIF, CIF, VGA, SD, and HD are 0.91, 0.91, 0.90, 0.91, and 0.90, respectively. It is quite remarkable to achieve $\rho \geq 0.9$ with one video quality model over a range of subjective data that spans from QCIF (176 x 144 pixels) to HD (1920 x 1080 pixels). In addition, the response of the VQM_VFD model is well behaved, as there are very few outliers in the plots.

In Figure 5, subjective scores vary between nominal values of 0 (best quality) and 1 (worst quality). Scores below 0 are possible and may result from quality improvements in the processed video with respect to the original video (e.g., noise removal, color enhancement). In addition, the INLSA scaling process used to map the subjective scores to the nominal (0, 1) range may cause a small number of subjective scores to be scaled to slightly negative values. Note that the HD plot demonstrates a lack of very low quality samples in comparison to the other image sizes.

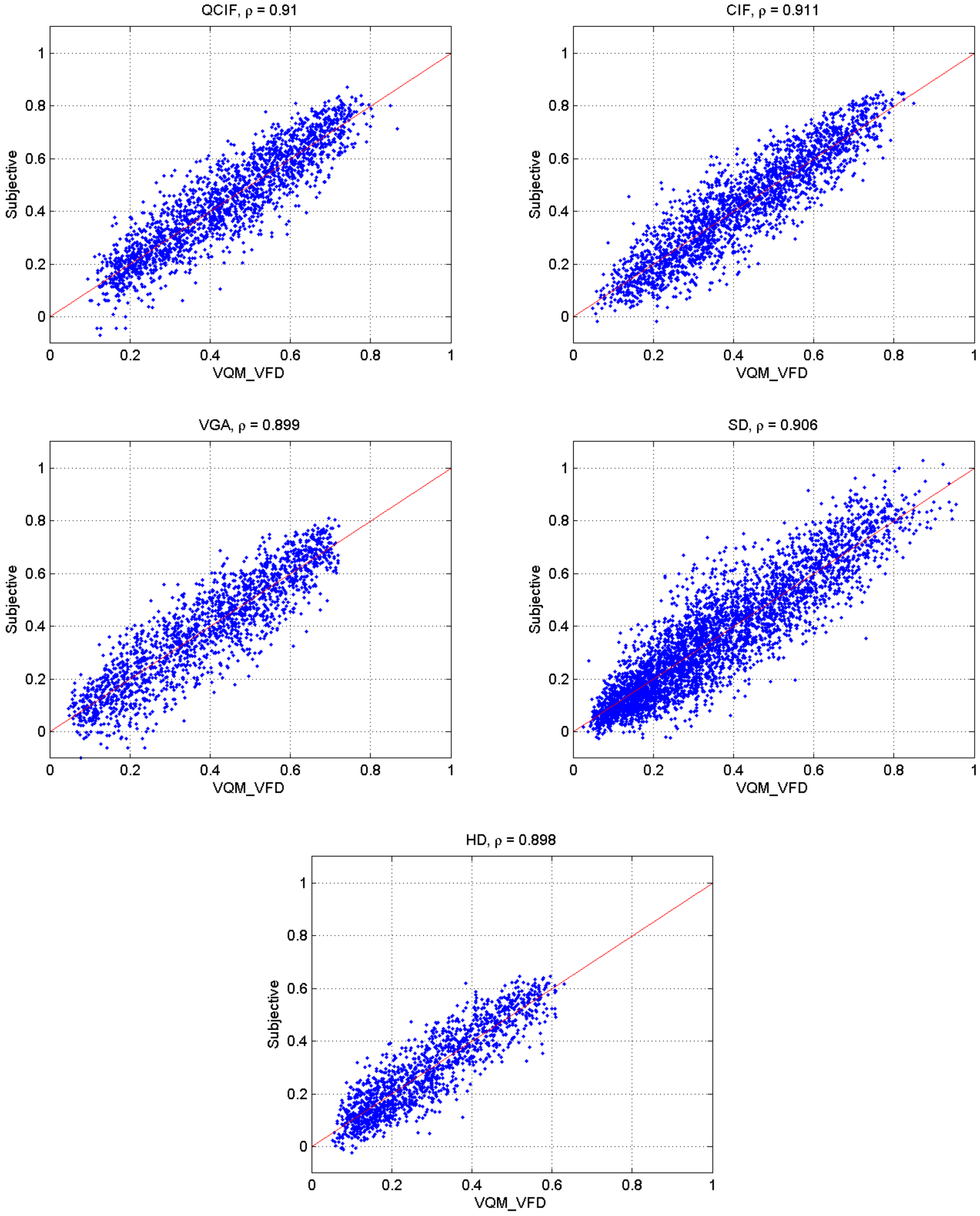


Figure 5. Performance of the VQM_VFD model.

5. SUGGESTIONS FOR FUTURE WORK

While the VQM_VFD model achieves good performance in predicting subjective ratings, there is always room for improvement. One obvious place for improvement is the addition of color distortion parameters to the VQM_VFD model. Many investigations have been performed looking for good color distortion metrics that might help to explain more of the error variance in the VQM_VFD model. The best improvements (although very minor) have resulted from using the two-dimensional (2D) coherent color feature (Section 4.3 of [4]) together with the Euclidean distance feature comparison function (Section 5.2.2 of [4]).

One possible reason for the difficulty in obtaining a robust color distortion metric that brings added information to the VQM_VFD model might be the lack of independent color distortions in the subject datasets. While distortions might appear in the chroma channels (C_B, C_R), these distortions nearly always also appear in the luma channel (Y). Another reason might be that some of the color distortions are actually pleasing to the eye (e.g., colors are made more vibrant). Thus, a color distortion metric probably needs to be bipolar, where some distortions produce increases in subjective quality while others produce decreases in subjective quality.

6. REFERENCES

[1] ITU-T Recommendation 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," Recommendations of the ITU, Telecommunication Standardization Sector.

[2] S. Wolf, "A full reference (FR) method using causality processing for estimating variable video delays," NTIA Technical Memorandum TM-10-463, Oct. 2009.

[3] S. Wolf, "Variable frame delay (VFD) parameters for video quality measurements," NTIA Technical Memorandum TM-11-475, Apr. 2011.

[4] S. Wolf and M. Pinson, "Video quality measurement techniques," NTIA Technical Report 02-392, Jun. 2002.

[5] M. Pinson and S. Wolf, "A New Standardized Method for Objectively Measuring Video Quality," *IEEE Transactions on Broadcasting, v. 50*, pp. 312–322, Sep. 2004.

[6] S. Wolf and M. H. Pinson, "The relationship between performance and spatial-temporal region size for reduced-reference, in-service video quality monitoring system," in *Proc. SCI / ISAS 2001 (Systematics, Cybernetics, and Informatics / Information Systems Analysis and Synthesis)*, Jul. 2001.

[7] M. Pinson, and S. Wolf, "An objective method for combining multiple subjective data sets," *SPIE Video Communications and Image Processing Conference*, Lugano, Switzerland, Jul. 2003.

FORM **NTIA-29** U.S. DEPARTMENT OF COMMERCE
(4-80) NAT'L. TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION

BIBLIOGRAPHIC DATA SHEET

1. PUBLICATION NO. TM-11-482	2. Government Accession No.	3. Recipient's Accession No.
4. TITLE AND SUBTITLE Video Quality Model for Variable Frame Delay (VQM_VFD)		5. Publication Date September 2011 6. Performing Organization NTIA/ITS.T
7. AUTHOR(S) Stephen Wolf and Margaret H. Pinson		9. Project/Task/Work Unit No. 3141011
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 No.
11. Sponsoring Organization Name and Address National Telecommunications & Information Administration Herbert C. Hoover Building 14th & Constitution Ave., NW Washington, DC 20230		12. Type of Report and Period Covered

14. SUPPLEMENTARY NOTES

15. ABSTRACT

Time varying delays of the output (or processed) video frames with respect to the input (i.e., the original or reference) video frames present significant challenges for Full Reference (FR) video quality measurement systems. Time alignment errors between the output video sequence and the input video sequence can produce measurement errors that greatly exceed the perceptual impact of these time varying video delays. This document proposes a new video quality model (VQM) that properly accounts for the perceptual impact of variable frame delay (VFD). This new model, called VQM_VFD, also uses perceptual features extracted from spatial-temporal (ST) blocks of a fixed angular extent. This enables VQM_VFD to track subjective quality over a wide range of viewing distances and image sizes. VQM_VFD uses a neural network that achieves 0.9 correlation to subjective quality for subjective datasets at image sizes from Quarter Common Intermediate Format (QCIF) to High Definition TV (HDTV).

16. Key Words

alignment; angular; calibration; correlation; delay; distance; dropped; frames; Full Reference (FR); measurement; model; objective; parameters; pausing; quality; resolution; skipping; subjective; time; variable; video

17. AVAILABILITY STATEMENT ☒ UNLIMITED.	18. Security Class. (This report) Unclassified	20. Number of pages 26
	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.