COMMITTEE T1

CONTRIBUTION

Document Number: T1A1.5/92-135

****

STANDARDS PROJECT: Analog Interface Performance Specifications for Digital Video Teleconferencing/Video Telephony Service

****

TITLE: An Update on ITS Video Quality Parameters

****

ISSUE ADDRESSED: Objective Video Performance Testing

****

SOURCE: NTIA/ITS, Stephen Wolf

****

DATE: July 13, 1992

****

DISTRIBUTION TO: T1A1.5

****

KEYWORDS: Video Performance Testing, Objective Quality, Subjective Quality

****

DISCLAIMER:

****

Introduction

At the last T1A1 meeting, the Institute for Telecommunication Sciences (ITS) presented contribution T1A1.5/92-112 entitled "Objective Measures of Video Impairment: Analysis of 128 Scenes." That contribution contained an International Broadcasting Convention (IBC) paper which discussed 3 objective parameters, (m_1, m_2, m_3), that have been found to be very useful for predicting subjective quality. This contribution contains:

1. A minor correction to one of the prediction coefficients, c_3.
2. An alternate method of computing the spatial distortion parameter m_1. This newer method has several advantages, including more efficient computation, and a relaxation of the requirement to align the original and degraded video frame samples.
3. A brief summary of other ongoing efforts aimed at reducing the computational complexity of the m_1, m_2, and m_3 quality parameters.
4. A brief summary of recent efforts to construct a real-time, in-service, PC-based video quality measurement system.

Coefficient Correction

Prediction coefficient c_3 (on page 4 of the IBC paper) should have been listed as -.3341 instead of +.3341. Prediction coefficients c_1, c_2, and c_3 are all negative. Since m_1, m_2, and m_3 are all measures of impairments (m_1 measures spatial impairments while m_2 and m_3 measure temporal impairments), quality is reduced for positive values of m_1, m_2, and m_3 according to equation (4) on page 3 of the IBC paper, reproduced here as:

$$\hat{s} = c_0 + c_1 m_1 + c_2 m_2 + c_3 m_3$$

Prediction coefficient c_0, which is slightly less than 5 (a perfect score), gives the predicted quality when no spatial or temporal impairments are present.

Alternate Method for Computing m_1

ITS has been investigating methods of making the computation of m_1, m_2, and m_3 as efficient as possible. This section discusses the results from one such investigation of the spatial distortion parameter m_1.

The original equation for computing the spatial distortion parameter m_1 was given on page 3 of the IBC paper as:

$$m_1 = rms_{time}\left(5.78\left|\frac{std_{space}(Sobel(O_n)) - std_{space}(Sobel(D_n))}{std_{space}(Sobel(O_n))}\right|\right)$$

$$n = 1, 2, 3, \ldots, N$$

Where O_n denotes the nth frame of the original video sequence, D_n denotes the

n^{th} frame of the degraded video sequence, N denotes the total number of frames measured (in our case N=271 frames, or 9 seconds of video), Sobel(•) indicates the Sobel filtering operation, std_{space} indicates that a spatial standard deviation of the image pixel values is computed, and rms_{time} indicates a root mean square operation of the frame samples is computed.

The following permutation of the above equation for m_1 was found to give virtually identical results (we shall denote this new measurement as m_1'):

$$m_1' = 5.78\left|\frac{rms_{time}\{std_{space}(Sobel(O_n))\} - rms_{time}\{std_{space}(Sobel(D_n))\}}{rms_{time}\{std_{space}(Sobel(O_n))\}}\right|$$

$$n = 1, 11, 21, \ldots, N$$

This equation has several advantages over the original method for computing m_1. Among these are (1) the Sobel and standard deviation operations are only computed for every 10th video frame, and (2) the root mean square operation is computed independently on the input and output frame samples. The ramification of (1) is that the measurement only requires 1/10 th of the computational processing time relative to the original measurement. The ramification of (2) is that only coarse alignment of the original and degraded frame samples is required. The second condition is not really an issue since an accurate, time alignment algorithm has been developed (see contribution T1A1.5/92-139).

Results using the revised m_1' equation demonstrate that the spatial information varies slowly relative to the video frame rate, and hence may be sampled at rates lower than the video frame rate. The revised least-squares solution for predicting the subjective quality is:

$$\hat{s}' = c_0' + c_1' m_1' + c_2' m_2 + c_3' m_3$$

where $c_0' = 4.8118$, $c_1' = -.9360$, $c_2' = -.3828$, and $c_3' = -.3675$

Figure 1 shows the prediction performance of the revised equation for the 64 testing scenes (+ symbols) plotted on the same graph as the prediction performance of the original equation in the IBC paper (❑ symbols). As one can see from Figure 1, substitution of the new m_1' produced only very minor changes in the prediction performance. In fact, most of the + symbols fall into the ❑ symbols. The coefficient of correlation for the revised equation is .949, which is slightly better than the coefficient of correlation for the original equation (=.947).

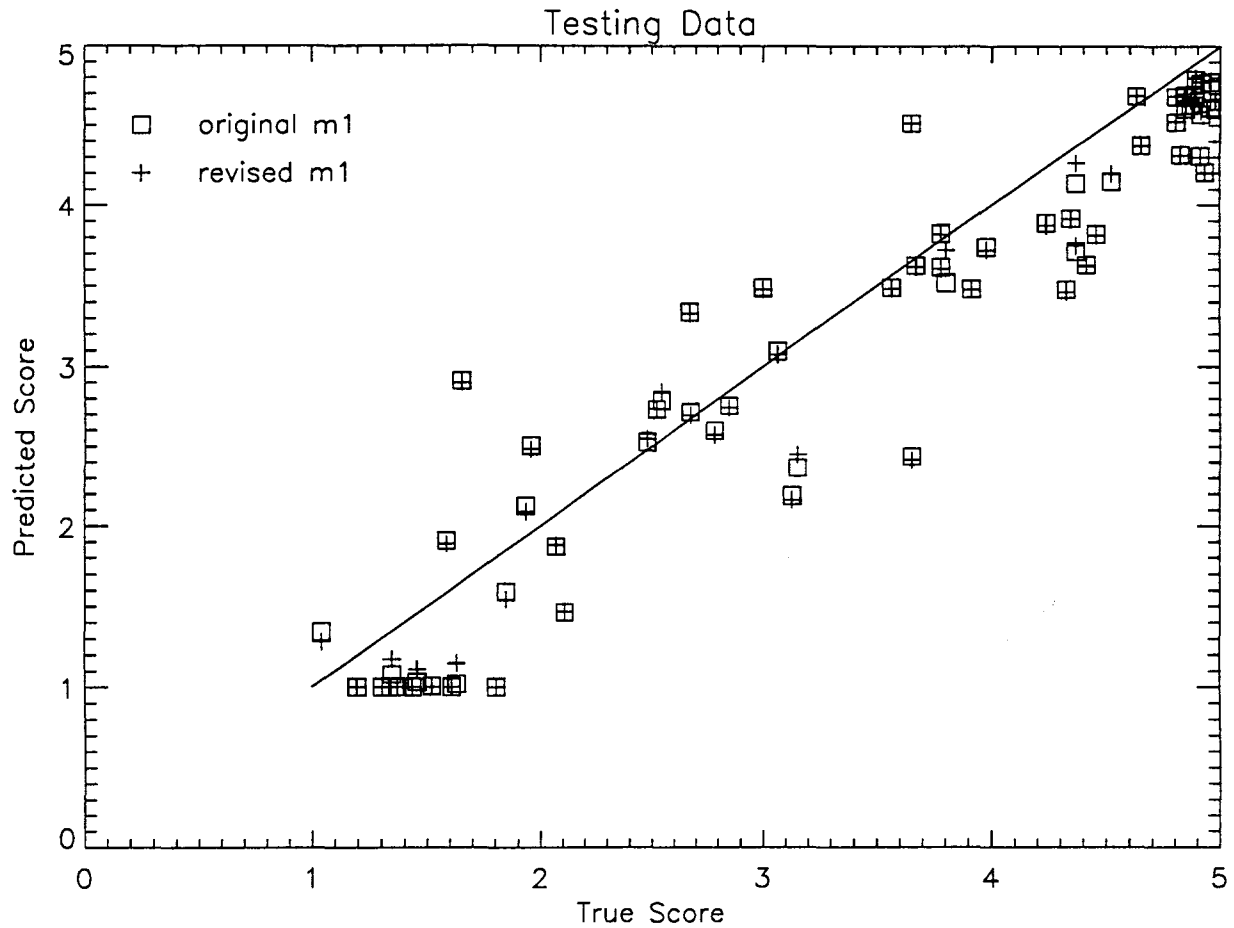

Figure 1 Prediction Performance, 64 Test Scenes

Other Efforts to Reduce Computational Complexity

ITS is investigating several other quality parameter modifications, the results of which are not available at this time, that are aimed at reducing the computational complexity of m_1, m_2, and m_3 parameters that were specified in the IBC paper. These other modifications include:

1. Using the pseudo-Sobel filter for computation of m_1 instead of the Sobel filter. This substitution would save two squares and one square root for every image pixel processed.

2. Spatial sub-sampling of the images for computation of m_2 and m_3. Since m_2 and m_3 are motion parameters (they detect frame to frame changes), sub-sampling of the images before computation of m_2 and m_3 should not destroy relevant information. Note that this applies only to the motion parameters m_2 and m_3 and that we are still proposing full $4xf_{sc}$ sampling for parameter m_1. The computational savings would depend upon the sub-sampling rate. ITS is investigating sub-sampling by a factor of 6: 3 in the horizontal direction (using approximately 256 samples per line), and 2 in the vertical direction (using one field of the NTSC frame).

Real-time, PC-based Video Quality Measurement System

ITS is currently in the process of constructing a real-time, in-service, PC-based video quality measurement system. In addition to real-time measurements of the m_1', m_2, and m_3 quality parameters, the completed system will also provide real-time measurements of the one-way video delay discussed in T1A1.5/92-139, and the transmitted frame rate/average frame rate discussed in T1A1.5/92-138. The primary component for the PC-based system is a low cost frame grabber/image processing card for the AT bus. The components have been ordered and the system will be prototyped on a 486 PC. The real-time video quality measurement system is expected to be integrated with similar voice and data measurement systems to produce a fully integrated voice/video/data measurement system for multimedia applications. In the future, this integrated performance measurement system is expected to be used for evaluating multimedia systems. Results of this effort will be made available to the T1A1.5 committee to assist in the development of related performance standards.

Conclusions

The m_1, m_2, and m_3 parameters have been shown to give accurate in-service measurements of digital video quality (see T1A1.5/92-136). This contribution has presented methods of making these parameters as computationally efficient as possible. ITS intends to demonstrate in the near future (4-6 months) a low cost, real-time PC based system for making these measurements. We will continue to investigate methods of making these measurements as efficient as possible before making our final recommendation to the T1A1.5 committee. The objective quality measurements that have been developed can be used for in-service (section 6 of the VTC/VT draft standard, T1A1.5/92-107) as well as out-of-service (section 5 of the VTC/VT draft standard) testing of digital video systems.