An objective video quality assessment system based on human perception

Arthur A. Webster, Coleen T. Jones, Margaret H. Pinson,
Stephen D. Voran, Stephen Wolf

Institute for Telecommunication Sciences
National Telecommunications and Information Administration
325 Broadway, Boulder, CO 80303

ABSTRACT

The Institute for Telecommunication Sciences (ITS) has developed an objective video quality assessment system that emulates human perception. The system returns results that agree closely with quality judgements made by a large panel of viewers. Such a system is valuable because it provides broadcasters, video engineers and standards organizations with the capability for making meaningful video quality evaluations without convening viewer panels. The issue is timely because compressed digital video systems present new quality measurement questions that are largely unanswered.

The perception-based system was developed and tested for a broad range of scenes and video technologies. The 36 test scenes contained widely varying amounts of spatial and temporal information. The 27 impairments included digital video compression systems operating at line rates from 56 kbits/sec to 45 Mbits/sec with controlled error rates, NTSC encode/decode cycles, VHS and S-VHS record/play cycles, and VHF transmission. Subjective viewer ratings of the video quality were gathered in the ITS subjective viewing laboratory that conforms to CCIR Recommendation 500-3. Objective measures of video quality were extracted from the digitally sampled video. These objective measurements are designed to quantify the spatial and temporal distortions perceived by the viewer.

This paper presents the following: a detailed description of several of the best ITS objective measurements, a perception-based model that predicts subjective ratings from these objective measurements, and a demonstration of the correlation between the model's predictions and viewer panel ratings. A personal computer-based system is being developed that will implement these objective video quality measurements in real time. These video quality measures are being considered for inclusion in the Digital Video Teleconferencing Performance Standard by the American National Standards Institute (ANSI) Accredited Standards Committee T1, Working Group T1A1.5.

1. INTRODUCTION

The need to measure video quality arises in the development of video equipment and in the delivery and storage of video and image information. Although the work described in this paper is concerned specifically with NTSC video (the distribution television standard in the United States), the principles presented can be applied to other types of motion video and even still images. The methods of video quality assessment can be divided into two main categories: subjective assessment (which uses human viewers) and objective assessment (which is accomplished by use of electrical measurements). While we believe that assessment of video quality is best accomplished by the human visual system, it is useful to have objective methods available which are repeatable, can be standardized, and can be performed quickly and easily with portable equipment. These objective methods should give results that correlate closely with results obtained through human perception.

Objective measurement of video quality was accomplished in the past through the use of static video test scenes such as resolution charts, color bars, multi-burst patterns, etc., and by measuring the signal to noise ratio of the video signal.[1] These objective methods address the spatial and color aspects of the video imagery as well as overall signal distortions present in traditional analog systems. With the development of digital compression technology, a large number of new video services have become available. The savings in transmission and/or storage bandwidth made possible with digital compression technology depends upon the amount of information present in the original (uncompressed) video signal, as well as how much quality the user is willing to sacrifice. Impairments may result when the information present in the video signal is larger than the transmission channel capacity. However, users may be willing to sacrifice quality to achieve a substantial reduction in transmission and storage costs. But, how much quality is sacrificed for how much cost savings? We propose a set of measurements that offers a way to begin to answer this question. New impairments can be present in digitally compressed video and these impairments include both spatial and temporal artifacts.[2] The old objective measurement techniques are not adequate to assess the impact on quality of these new artifacts.[3]

After some investigation of compressed video, it becomes clear that the perceived quality of the video after passing through a given digital compression system is often a function of the input scene. This is particularly true for low bit-rate systems. A scene with little motion and limited spatial detail (such as a head and shoulders shot of a newscaster) may be compressed to 384 kbits/sec and decompressed with relatively little distortion. Another scene (such as a football game) which contains a large amount of motion as well as spatial detail will appear quite distorted at the same bit rate. Therefore, we directed our efforts toward developing perception-based objective measurements which are extracted from the actual sampled video. These objective measurements quantify the perceived spatial and temporal distortions in a way that correlates as closely as possible with the response of a human visual system. Each scene was digitized (at 4 times sub-carrier frequency) to produce a time sequence of images sampled at 30 frames per second (in time) and 756 x 486 pixels (in space).

2. DEVELOPMENT METHODOLOGY

Figure 1 presents a graphical depiction of the development process for the ITS quality assessment algorithm. A set of video scene pairs (each consisting of the original and a degraded version) was used in a subjective test. These scene pairs were also processed on a computer that extracted a large number of features. Statistical analysis was used to select an optimal set of quality parameters (obtained from features) that correlated well with the viewing panel results. This optimal set of parameters was then used to develop a quality assessment algorithm that gives results that agree closely with viewing panel results.

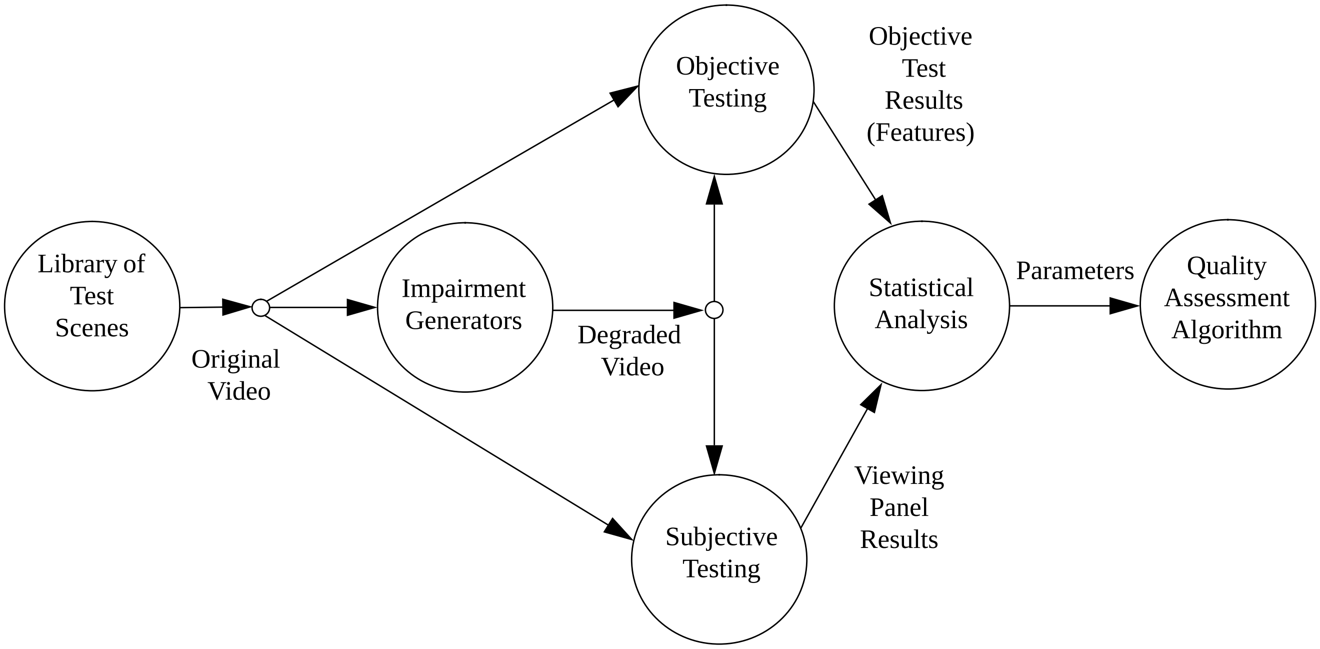

Figure 1. Development Process for Video Quality Assessment Algorithm

2.1 Library of test scenes

Several scenes, exhibiting various amounts of spatial and temporal information content, are needed to characterize the performance of a video system. Even more scenes are needed to guard against viewer boredom during the subjective testing. A set of 36 test scenes was chosen for the experiment. The test scenes spanned a wide range of user applications including still scenes, limited motion graphics, and full motion entertainment video.

2.2 Impairment generators

Twenty-seven video systems (plus the ‘no impairment’ system) were used to produce the degraded video that was used in the tests. The original video for this test was component analog video. The digital video systems included 11 video codecs (coder-decoders) from 7 manufacturers operating at bit rates from 56 kbits/sec to 45 Mbits/sec including bit error rates of 10^{-6} and 10^{-5}. Also included were analog video systems such as VHS and S-VHS recording and playback, and noisy RF transmission. All video systems except the ‘no impairment’ system included NTSC encoding and decoding.

2.3 Objective testing

Both the original video and the degraded video were digitized and processed to extract a large number of features. The processing included Sobel filtering, Laplace filtering, fast Fourier transforms, first-order differencing, color distortion measurements[4], and moment calculations. Typically, features were calculated from each original and degraded frame of the video sequence to produce time histories. Some features required the entire original and degraded video image (e.g., the variance of the error image calculated from the difference between the original and the degraded images). Other features required only the statistics of the original and degraded video images (e.g., the change in image energy obtained from the differences between the original and the degraded image variances). The time histories of the features were collapsed by various methods, e.g., maximum (MAX), root mean square (RMS), standard deviation (STD), etc., to produce a single scalar value (or parameter) for each test scene. These parameters defined the objective measurements and were used in the statistical analysis step shown in Figure 1.

2.4 Subjective testing

The subjective test was conducted in accordance with CCIR Recommendation 500-3.[5] A panel of 48 viewers were selected from the U.S. Department of Commerce Laboratories phone book in Boulder, Colorado. Each viewer completed four viewing sessions during a single week, attending one session per day. Each session lasted approximately 25 minutes and required viewing of 38 or 40, 30-second test clips. A clip is defined as a test scene pair consisting of the original video and the degraded video. The viewer was first shown the original video for 9 seconds followed by 3 seconds of grey and then 9 seconds of the degraded video. 9 seconds was allowed to rate the impairment on a 5 point scale before the next clip was presented. The viewer was asked to rate the difference between the original video and the degraded video as either (5) Imperceptible, (4) Perceptible but Not Annoying, (3) Slightly Annoying, (2) Annoying, or (1) Very Annoying. This scale covers a wide range of impairment levels and is specified as one of the standard scales in the CCIR Recommendation 500-3. Impairment testing was used since we were interested in measuring the change in video quality due to a video system. A mean opinion score was generated by averaging the viewer ratings.

The selection of 158 clips used in the test (out of 972 clips available) was made both deterministically and randomly. Random selections were made from a distribution table that paired video teleconferencing systems with more video teleconferencing scenes than entertainment scenes, and entertainment systems with more entertainment scenes than video teleconferencing scenes. The viewers rated 132 unique clips from the 158 actually viewed because some were used for training and consistency checks.

2.5 Statistical analysis and quality assessment system

This stage of the development process utilized joint statistical analysis of the subjective and objective data sets. This step identifies a subset of the candidate objective measurements that provides useful and unique video quality information. The best measurement was selected by exhaustive search. Additional measurements were selected to reduce the remaining objective-subjective error by the largest amount. Selected measurements complement each other. For instance, a temporal distortion measure was selected to reduce the objective-subjective error remaining from a previous selection of a spatial distortion measure. When combined in a simple linear model, this subset of measurements provides predicted scores that correlate well with the true scores obtained in the subjective tests. In constructing the linear model we looked for p measurements $\{m_i\}$ and $p+1$ constants $\{c_i\}$, that allowed us to estimate the subjective mean opinion score. The estimated subjective mean opinion score is given by

$$s \approx \hat{s} = c_0 + \sum_{i=1}^{p} c_i m_i, \tag{1}$$

where s is the true subjective mean opinion score and $\hat{s}$ is the estimated score.

3. RESULTS

For the results presented here, three complementary video quality measurements (p=3) were selected. These three complementary measures (m_1, m_2, and m_3) have been used to explain most of the variance in subjective video quality that resulted from the impairments used in this experiment. The investigations and research that produced the m_1, m_2, and m_3 video quality metrics also provided insight into how the human perceives the spatial and temporal information of a video scene.

3.1 Spatial and temporal information features

The difficulty in compressing a given video sequence depends upon the perceived spatial and temporal information present in that video sequence. Perceived spatial information is the amount of spatial detail in the video scene that is perceived by the viewer. Likewise, perceived temporal information is the amount of perceived motion in the video scene. Thus, it would be useful to have approximate measures of perceived spatial and temporal information. These information measures could be used to select test scenes that appropriately stress the video compression system being designed or tested. Two different test scenes with the same spatial and temporal information should produce similar perceived quality at the output of the transmission channel. Measures of distortion could also be obtained by comparing the perceived information content of the video before and after passing through a video system. Although it is recognized that spatial and temporal aspects of vision perception cannot be completely separated from each other, we have found spatial and temporal features that correlate with human quality perception of spatial detail and motion. Both of these features require pixel differencing operations, which seem to be basic attributes of the human visual system. The spatial information (SI) feature differences pixels across space while the temporal information (TI) feature differences pixels across time. Here, both the SI and TI features have been applied to the luminance portion of the video.

3.1.1 Spatial information (SI)

The spatial information feature is based on the Sobel filter.[6] At time n, the video frame F_n is filtered with the Sobel operators. The standard deviation over the pixels in each Sobel-filtered frame is then computed. This operation is repeated for each frame in the video sequence and results in a time series of spatial information values. Thus, the spatial information feature, $SI\,[F_n]$, is given by

$$SI\,[F_n] = STD_{space}\,\{Sobel\,[F_n]\}\ , \tag{2}$$

where STD_{space} is the standard deviation operator over the horizontal and vertical spatial dimensions in a frame, and F_n is the n^{th} frame in the video sequence. Figure 2 shows a time sequence of 3 contiguous video frames for an original scene (top row) and degraded version of that scene (second row). These images were sampled at the NTSC frame rate of approximately 30 frames per second. The degraded version of the scene was obtained from a 56 kbits/sec codec. The third row of Figure 2 shows the Sobel filtered version of the original scene and the fourth row shows the Sobel filtered version of the degraded scene. The highly localized, clearly focussed edges in the third row produce a large STD_{space} since the standard deviation is a measure of the spread in pixel values. On the other hand, the non-localized, blurred edges shown in the fourth row produce a smaller STD_{space}, demonstrating that spatial detail has been lost. This is particularly evident for the images in the third column.

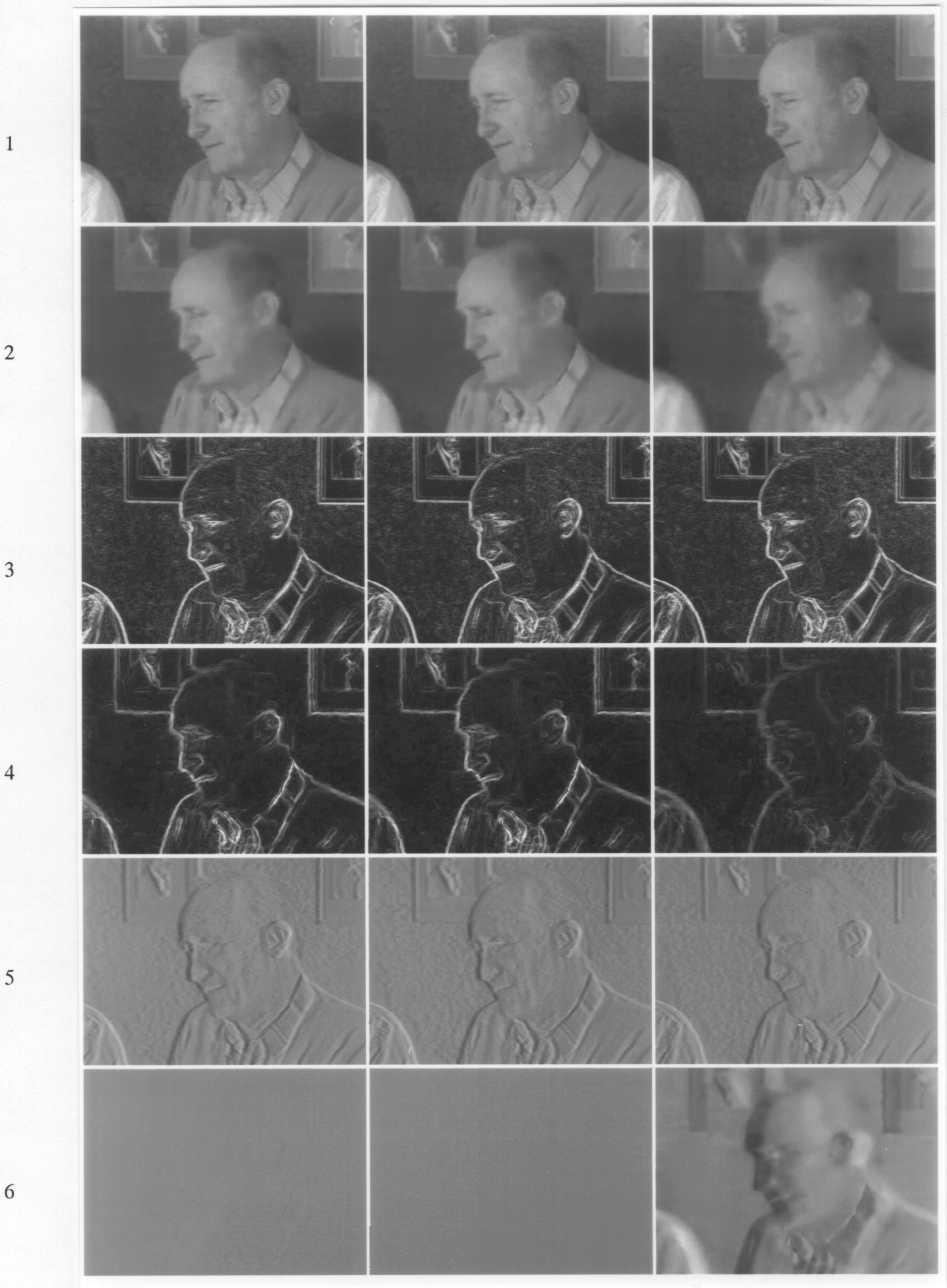

Figure 2. Video Processed to Demonstrate Perceived Spatial and Temporal Information

3.1.2 Temporal information (TI)

The temporal information feature is based upon the motion difference image, ΔF_n, which is composed of the differences between pixel values at the same location in space but at successive times or frames. ΔF_n , as a function of time *(n)*, is defined as

$$\Delta F_n = F_n - F_{n-1} . \tag{3}$$

The temporal information feature, $TI\,[F_n]$, is defined as the standard deviation of ΔF_n over the horizontal and vertical spatial dimensions, and is given by

$$TI\,[F_n] = STD_{space}\,[\Delta F_n] . \tag{4}$$

More motion in adjacent frames will result in higher values of $TI\,[F_n]$. The fifth row of Figure 2 shows the ΔF_n motion difference frames of the original video (top row) while the sixth row shows the motion difference frames of the degraded video (second row). The motion in the original scene was a smooth camera pan. Two of the motion difference frames for the degraded video shown in the sixth row contain very little motion energy. This resulted because the low bit-rate codec updated the information in the scene at less than 30 frames per second, the NTSC frame rate. In fact, the first degraded image in row 2 had already been repeated for several video frames prior to the time of column 1 (one can see this by comparing the first original image with the first degraded image in Figure 2 and noting that the first degraded image lags in time). When the codec does update the information (last column), the motion appears jerky. Humans perceive this as unnatural motion. Thus, the time history of TI quantifies the perception of this motion. In Figure 2, the flow of motion has been distorted from smooth and continuous in the original video to localized and discontinuous in the degraded video.

3.1.3 Spatial-temporal matrix

Figure 3 shows how the set of 36 test scenes used in the ITS video quality experiments can be placed on a spatial-temporal information plot. For clarity, only the maximum value over the 9 second time histories of SI and TI for each original scene was plotted. When entire time histories are plotted, most test scenes will produce trajectories in the spatial-temporal space. Along the TI=0 axis (x-axis) are found the still scenes and those with very limited motion. Along the SI=0 axis (y-axis) are found scenes with minimal spatial detail. These values of SI and TI can be compared to other test scenes measured using the above equations which have been spatially sampled at 4 times sub-carrier ($4f_{sc}$) or approximately 756 x 486 pixels, 30 frames per second, and digitized with white set to 235 and black set to 16. Note that no attempt has been made at this point to normalize SI and TI relative to their respective importance to the human visual system. As the distance from the origin increases, the total perceived information increases. This results in increasing coding difficulty and may result in increased distortion for a fixed bit-rate digital system.

3.2 Video quality measures

The three video quality measures presented here involve equational forms of the SI and TI features extracted from the original and degraded video. Results will be presented in section 3.3 that demonstrate the validity of the three video quality measures presented in this section. Since the range of video quality in the experiment was quite large, and since the linear predictor based on the three video quality measures closely tracked subjective mean opinion scores, we feel that the SI and TI features quantify basic perceptual attributes of the human visual system. The three video quality measurements have been normalized for unit variance so that we can interpret coefficient magnitudes as indications of the relative importance of m_1, m_2, and m_3. These normalization constants are included in the equations for m_1, m_2, and m_3.

3.2.1 Measurement m_1

Measurement m_1 was the first measurement selected by the statistical analysis. It is a measure of spatial distortion and is obtained from the SI features of the original and degraded video. The spatial distortions measured by m_1 include both blurring and false edges. Since the Sobel filter is an edge enhancement filter, edge energy gained or lost in an image after passing

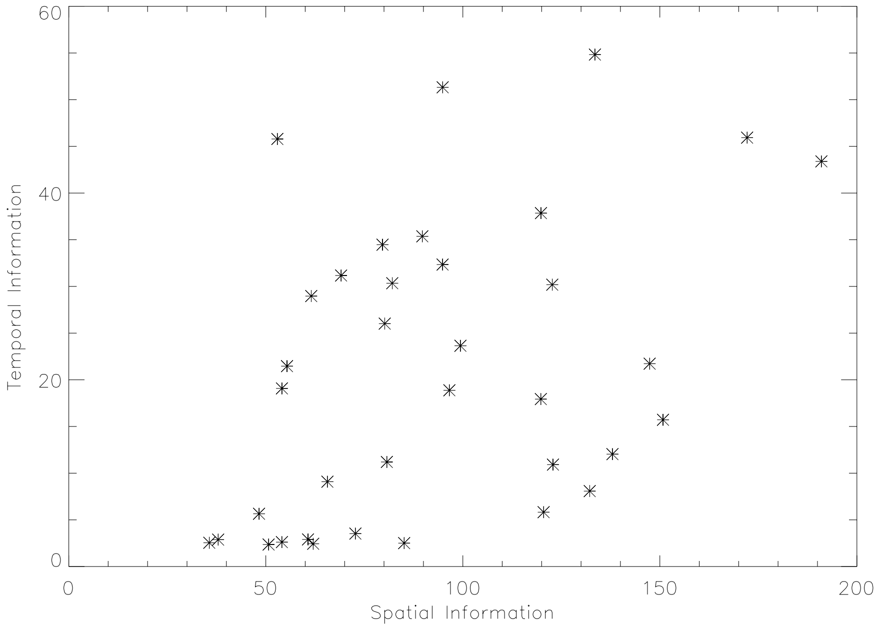

Figure 3. Spatial-Temporal Locations of ITS Test Scenes

through a video system will be measured by m_1. The equational form for m_1 is given by

$$m_1 = RMS_{time}\left(5.81\left|\frac{SI[O_n] - SI[D_n]}{SI[O_n]}\right|\right), \tag{5}$$

where O_n denotes the nth frame of the original video sequence, D_n is the nth frame of the degraded video sequence, *SI* [.] indicates the Spatial Information operator defined in Equation (2), and RMS_{time} denotes the root mean square time-collapsing function. Note that m_1 measures the relative change in SI between the original and degraded video. Figure 4 shows the time history of *SI* $[F_n]$ plotted for an original and a degraded version of a test scene. The degraded version was obtained from a 56 kbits/sec codec. Note that the SI of the degraded video was less than the SI of the original video, indicating spatial blurring.

3.2.2 Measurement m_2

The measurements m_2 and m_3 that were selected next by the statistical analysis are both measures of temporal distortion. These temporal distortion measures complement the spatial distortion measure m_1 and account for most of the remaining prediction error that resulted from using just m_1. Measurement m_2 is given by

$$m_2 = f_{time}[0.108 \cdot MAX\{(TI[O_n] - TI[D_n]), 0\}], \tag{6}$$

where

$$f_{time}(x_t) = STD_{time}\{CONV(x_t, [-1, 2, -1])\}, \tag{7}$$

and *TI* $[F_n]$ is the Temporal Information operator defined in Equation (4), *CONV* denotes the convolution operator, and STD_{time} denotes a standard deviation across time. Figure 5 shows *TI* $[F_n]$ for the same original and degraded video as shown

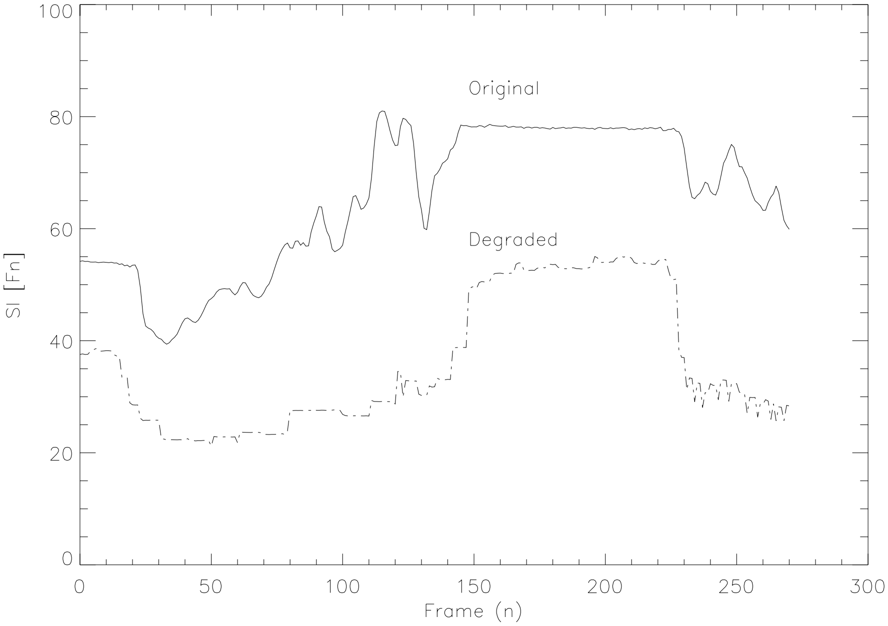

Figure 4. SI $[F_n]$ for the Original and Degraded Video Sequences

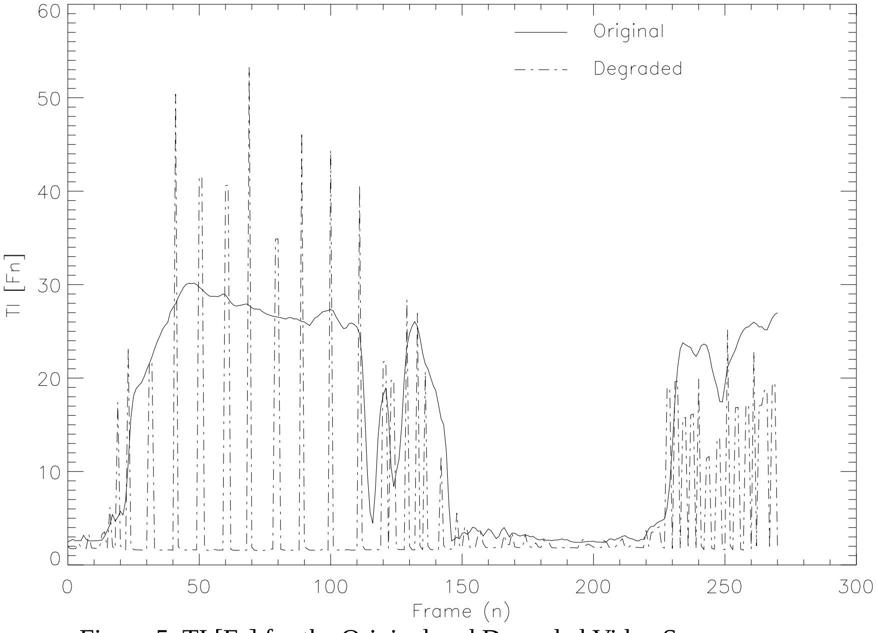

Figure 5. TI $[F_n]$ for the Original and Degraded Video Sequences

in Figure 4. The TI of the degraded video has areas of no motion (values that are approximately zero), and areas of large localized motion (spikes). By examining the frequency of the motion spikes, one can deduce the frame rate of the codec. For low bit rate systems, the frame rate is normally adaptive and changes depending upon the motion in the original scene. Areas of large motion in the original (such as the pan from frame 20 to frame 150) cause very large temporal distortions in the output. The longer the codec waits to update the video frame, the greater the perceived jerkiness. Note that m_2 measures the effect of temporally localized motion in the degraded video that were not in the original video. The convolutional kernel enhances these motion transitions since it is a high pass filter and the STD_{time} quantifies the spread in energy of the enhanced motion transitions. The m_2 measure is non-zero only when the degraded video has lost motion energy, such as during frame repetition as shown in Figure 2.

3.2.3 Measurement m_3

Measurement m_3 is another measure of temporal distortion and is given by

$$m_3 = MAX_{time}\{4.23 \cdot LOG_{10}(\frac{TI[D_n]}{TI[O_n]})\} \qquad (8)$$

where MAX_{time} returns the maximum value of the time history. Measurement m_3 selects the video frame that has the largest added motion. This may be the point of maximum jerky motion or the point in the video that has the worst uncorrected block errors (when errors in the digital transmission channel are present). Random or periodic noise are also detected by m_3. All of these impairments were included in the experimental data.

3.3 Assessment Algorithm Performance

Objective quality assessment is provided by estimating the level of perceived impairment present in a video sequence from the m_1, m_2, and m_3 quality measures. The $\{c_i\}$ constants for Equation (1) were found by applying least squares error criteria to a training set that was composed of 18 of the 36 test scenes (64 of the 132 clips). The remainder of the data was reserved for testing. The equation used to generate estimations of the subjective scores is

$$\hat{s} = 4.77 - 0.992m_1 - 0.272m_2 - 0.356m_3. \qquad (9)$$

The correlation coefficient between the subjective scores and the estimated (or objective) scores was 0.92 for the training set. With the testing set (68 of the 132 clips), the correlation coefficient was 0.94. This is a highly encouraging result. Figure 6 shows a plot of subjective versus estimated scores for both sets of data. The triangles represent the training data and the squares represent the testing data.[7] The testing data was clipped so that no scene was rated less than 1.

3.4 Extensions of the basic forms

The SI and TI features define a new class of quality metrics, not specifically limited to the above three equational forms. For some applications, collapsing SI and TI to single values per frame may be too limited. One could compute more localized measures by calculating SI and TI for image sub-regions. These subregions could be vertical strips, horizontal strips, rectangular regions, or even motion/still segmented regions. For instance, during recent analysis of contribution quality 45 Mbits/sec codecs from a different experiment, we have found that a more sensitive quality metric can be obtained by applying the TI distortion measures to horizontal strips of the degraded video that did not contain any motion in the original video. This measures the added noise in the (originally) still portion of the video scene.

Note that SI and TI are insensitive to shifts in the image mean (provided the dynamic range of the video system is not exceeded), but they are sensitive to image intensity scaling. In cases where one has reason to suspect that the video system under test has a constant, non-unity gain, the SI and TI of the degraded video can be compensated by dividing by the gain. This gain can be estimated from the original and degraded video images or found by other means.

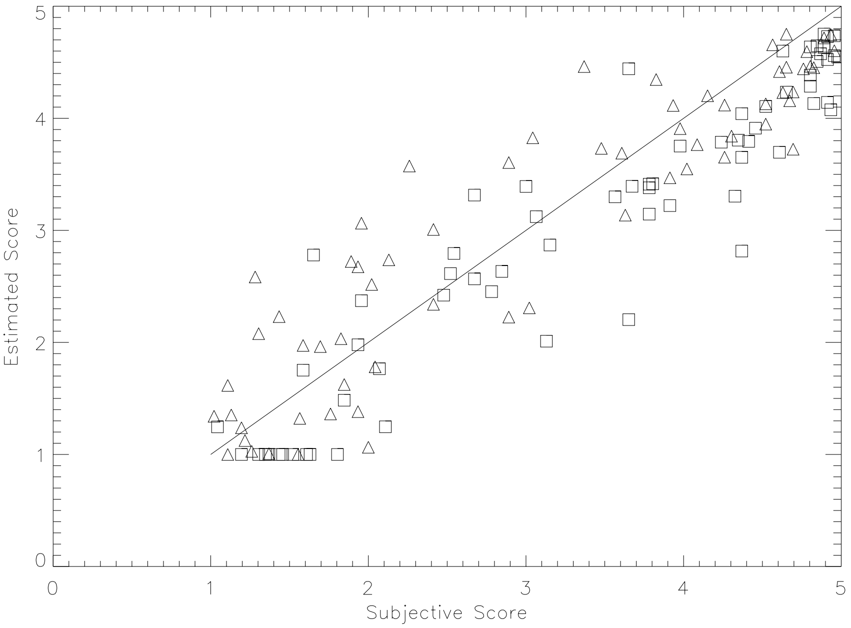

Figure 6. Estimated Scores Versus True Subjective Scores

4. REAL TIME IMPLEMENTATION

We are currently in the process of implementing our quality assessment algorithm in real time using a personal computer (PC). This has been made possible due to recent advances in frame sampling and image processing cards for PCs. The goal is the ability to estimate real-time video quality in the field, where the original and degraded video may be spatially separated by a great distance. Figure 7 is a block diagram of the real-time system. The final measurement system will consists of two PC's (one to process the source video and the other to process the destination video), two modems, and a phone line. A primary advantage of the video quality metrics is that they are based on the low-bit rate SI and TI quantities. This means that they can be transmitted over ordinary dial-up phone lines. This makes it possible to conduct economic in-service measurements of video quality when the source video and the destination video are separated by large distances. In-service measurements are important because they allow one to measure the actual video quality that is being delivered by the transmission channel. We have seen that this quality can change depending upon the source video being transmitted. Out-of-service testing requires only a PC at the destination end and a standard set of video scenes at the source end. In this case, the SI and TI features for the original source video are precomputed and stored in the PC at the destination end.

The video link begins with source video input to the system. This may be live video from a camera, video output from a VCR, a laserdisc player, or any other video source. The source video is input to the transmission channel shown in Figure 7. The example transmission service channel shown in Figure 7 is composed of an encoder, a digital communication circuit, and a decoder. The actual calculations of the video measurements and the predicted subjective score are performed in the parameter measurement equipment. This equipment consists of a PC with an image processing circuit board. The image processing board can process data at rates up to 30 MHz and is capable of digitizing video at 30 frames/sec. It performs both convolutions and image differencing. A PC, configured as described, is located at both the source video and the destination video sites.

For in-service measurements, the PC's will calculate the SI and TI features at or near real time for both the source and destination video. Although Figure 7 shows only the PC at the destination end calculating the m_1, m_2, and m_3 quality measures, the system is, in fact, symmetrical so that quality measurements are available at either the source or the destination end. The SI and TI features can be time tagged. This is particularly useful for calculating the video delay of the transmission channel, since the SI and TI features can be time-aligned with a simple correlation process. Once both the source and destination measurements are available at one PC, they can be combined in the linear predictor described previously to produce the estimated subjective rating of the live video.

The real-time system is currently in the development stage. It consists of one PC containing two image processing cards. One card processes the source video and the other processes the destination video. Our laboratory digital communication circuit consists of one codec with the coder output looped back to the decoder input. A camera supplies live source video.

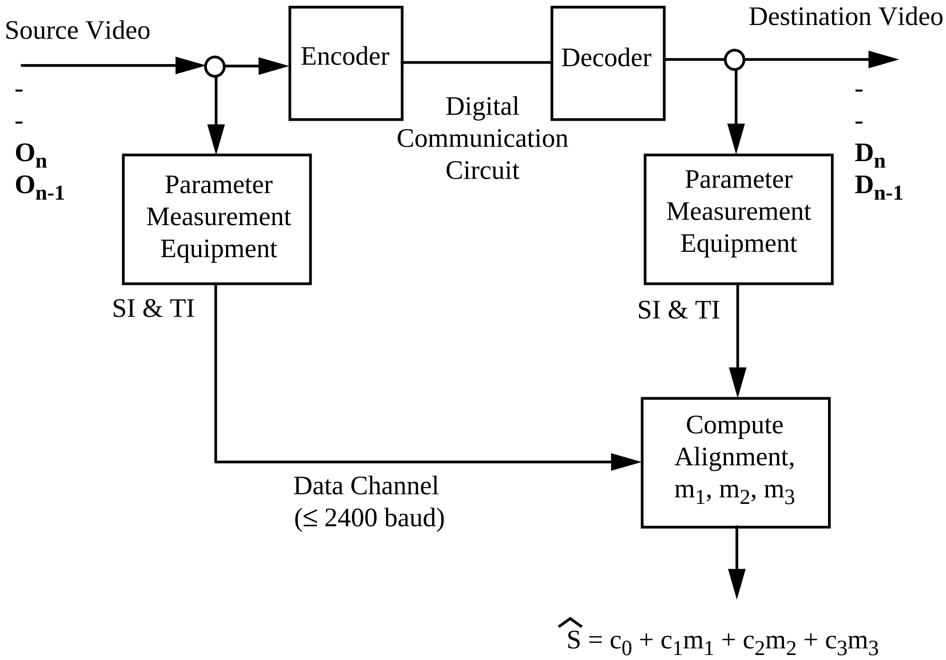
Source Video
-
-
O_n
O_{n-1}
Encoder
Digital
Communication
Circuit
Decoder
Destination Video
-
-
D_n
D_{n-1}
Parameter Measurement Equipment
SI & TI
Parameter Measurement Equipment
SI & TI
Compute Alignment, m_1, m_2, m_3
Data Channel
(≤ 2400 baud)
$\hat{S} = c_0 + c_1m_1 + c_2m_2 + c_3m_3$

Figure 7. Real-time System Implementation

To make the measurements as efficient as possible, we are examining several alternate computations of SI and TI. These alternate computations are:

1) The true Sobel filter is replaced with the pseudo-Sobel filter for the computation of the SI feature in Equation (2). The true Sobel filter convolves the image with two convolution kernels. One detects horizontal edges, and the other detects vertical edges. The final result is the square root of the sum of the squares of the individual convolutions. This calculation requires two passes of the image on the image processing boards we are using. The pseudo-Sobel uses the same two kernels as the true Sobel, but the final result is simply the sum of the absolute values of the individual convolutions. This calculation is twice as fast as the true Sobel calculation because it can be done in a single pass over the image. In addition, we have conducted investigations that show that the SI feature can be temporally sub-sampled at a rate as low as 3 times per second (every 10th NTSC video frame) without appreciably affecting the results.

2) The mean of the absolute value of the difference image (see Figure 7) is substituted for the TI feature in Equation (4). This modified TI measure is simpler to compute in real time since it is easier to accumulate pixel values than it is to accumulate the squares of pixel values. Investigations have shown that this new TI measure contains similar quality information as the original TI measure in Equation (4). The new TI measure is also spatially sub-sampled by a factor of 2 in the horizontal direction and 2 in the vertical direction (sampled at 2 times sub-carrier rather than 4 times sub-carrier and applied to only one field of the NTSC frame).

The affects of the above changes to SI and TI on the video quality metrics m_1, m_2, and m_3 are currently being investigated. The real-time video quality system provides a means for rapidly testing the performance of these alternate video quality metrics. Future uses for the real-time system include developing and testing other video quality metrics in the laboratory, and measuring end-to-end video quality in the field for applications such as video teleconferencing and the delivery of digital video to the home.

5. CONCLUSIONS

Based on the results of this experiment, it appears that the human visual system may be perceiving spatial and temporal distortions in video by performing differences in space and time, and by comparing these quantities to a reference. The ITS video quality algorithm used the original video scene as a source of reference features. These video quality reference features extracted from the original video scene can be communicated at a low bit rate. This seems to indicate that the perceptual bandwidth of the human visual system, for quality considerations, is much less than the bandwidth of the original video scene. Therefore, one does not require the original video scene in its entirety to perform good quality estimation.

We have presented a family of video quality measures that quantify the perceived spatial and temporal distortions in video systems. These measures have proven useful in predicting subjective quality of both compressed digital video systems and traditional analog video systems. The development methodology that was used and the resulting perceptual measures of video quality are general enough to be applied to the wide range of new video systems and services being introduced today. These video applications span a wide range of bit rates and quality levels from videophone to HDTV. Furthermore, the video quality measures can be performed in real-time on a PC-based field unit.

The metrics presented here were applied to the luminance portion of the video signal. Although we investigated a very large number of color distortion measures, none of these were selected by our optimization algorithms because color distortions were insignificant relative to the spatial and temporal distortions. We feel that an investigation of color distortion metrics will require a data set where the spatial and temporal distortions are minimal.

6. REFERENCES

1. EIA-250-B, “Electrical Performance Standard for Television Relay Facilities,” Electronic Industries Association, Washington, D.C., US, 1976.

2. S. Wolf, “Features for Automated Quality Assessment of Digitally Transmitted Video,” US Department of Commerce, National Telecommunications and Information Administration Report 90-264, June, 1990.

3. A. N. Netravali, and B. G. Haskell, “Digital Pictures: Representation and Compression,” Plenum Publishing Corporation, 1988.

4. “Recommendations on Uniform Color Spaces - Color Difference Equations, Psychometric Color Terms,” CIE Supplement No. 2 to CIE Publication No. 15 (E-1.3.1) 1971/(TC-1.3), 1978.

5. CCIR Recommendation 500-3, “Method for the Subjective Assessment of the Quality of Television Pictures,” Recommendations and Reports of the CCIR, 1986, XVIth Plenary Assembly, Volume XI, Part 1.

6. A. K. Jain, “Fundamentals of Digital Image Processing,” Prentice-Hall Inc., Englewood Cliffs, New Jersey, US, 1989.

7. S. D. Voran and S. Wolf, “The development and evaluation of an objective video quality assessment system that emulates human viewing panels,” International Broadcasting Convention, Amsterdam, The Netherlands, 1992.