THE DEVELOPMENT OF OBJECTIVE VIDEO QUALITY MEASURES THAT EMULATE HUMAN PERCEPTION

Stephen Voran

U.S Department of Commerce
NTIA/ITS.N3
325 Broadway
Boulder, CO 80303
Phone 303-497-3839
Fax 303-497-5323

Abstract

The Institute for Telecommunication Sciences (ITS) is conducting research to derive objective measures of video quality that emulate human perception. These measures should agree closely with quality judgements made by a large panel of viewers. The measures are valuable because they provide video designers and standards organizations with means for making meaningful quality evaluations without convening viewer panels. The derivation of these measures involves the following steps. A set of test scenes is selected and distorted. Next, we extract a set of candidate objective measurements that quantify the video scene distortions that are thought to be important to the human visual and perceptual systems. A panel of viewers watches the same set of test scenes and their subjective judgements of the distortions are recorded. The final step in the derivation process is a simultaneous statistical analysis of the objective and subjective data sets. This analysis reveals which portion of the objective data set is meaningful, and how the objective data should be combined to create an objective metric that emulates human perception. After describing the need for these measures, this paper provides a detailed description of the derivation process and some preliminary results.

1. Introduction

The traditional performance measurements of video transport and storage systems use fixed test signals and assume that the system under test is time-invariant.[1] While these signals and the associated measurements are indispensable for the characterization of the electrical performance of conventional, time-invariant, analog video systems, the measurements often do not correlate well with video quality as perceived by the end users of the video system. For instance, weighted signal-to-noise ratio does not give an accurate indication of image quality when the noise is correlated with the image, as is the case with predictive video coders.[2] A video system with horizontal resolution limit of 200 television lines (TVL) may be adequate for head and shoulders video teleconferencing, but unacceptable when graphics are added to the scene. A chrominance phase error of 10 degrees might be insignificant while the weather map is being transmitted but it becomes objectionable when the meteorologist appears with false colored flesh. In each of these examples, it is the variability of video scenes that results in a range of perceived video quality levels for a fixed video system.

The primary motivation for characterizing the electrical performance of video equipment is to objectively and consistently maintain picture quality by controlling measurable electrical parameters. From the examples above, it is disturbing to note however, that perceived picture quality can change while measured electrical values remain constant. One solution to this problem is to set tight bounds on the electrical parameters to insure that all possible video scenes of interest will be reproduced with sufficient quality.

Over the past decade, the problem has become more complicated. Video signals are now commonly transmitted and stored in compressed digital form with a possible resulting loss of quality. The stringent bounds on electrical parameters adopted by the television broadcasting industry are not realistic limits for many of the new video services. On the other hand, some metric of system performance is essential. Ideally, this metric should mimic the metric of the human visual and perceptual system, so that measured video quality agrees with video quality as perceived by the end user who actually views the video signal.

Modification of the existing traditional test signals and measurements will not solve the problem because there is a fundamental incompatibility between traditional analog video signal testing and modern digital video systems. Effective compression algorithms are dynamic, with the input signal dictating the overall behavior of the algorithm through many sub-algorithms that perform motion prediction, adaptive transforms, adaptive pixel and/or frame prediction, and adaptive quantization, to name only a few. The resulting video systems are clearly time-varying systems. Due to the complex dynamic nature of these systems, the conventional family of static, deterministic test signals cannot provide an accurate characterization of their performance.

2. Overview

This paper describes a method for deriving objective measurements of video quality that can be used to predict the human perception of video quality. The intent is that these measures should work well over a wide range of analog and digital video transmission and storage technologies. Such measurements would be indispensable to persons who design, operate, and maintain video components, storage and delivery systems, as well as those involved in standards work. Our method for deriving these measurements is diagramed in Figure 1.

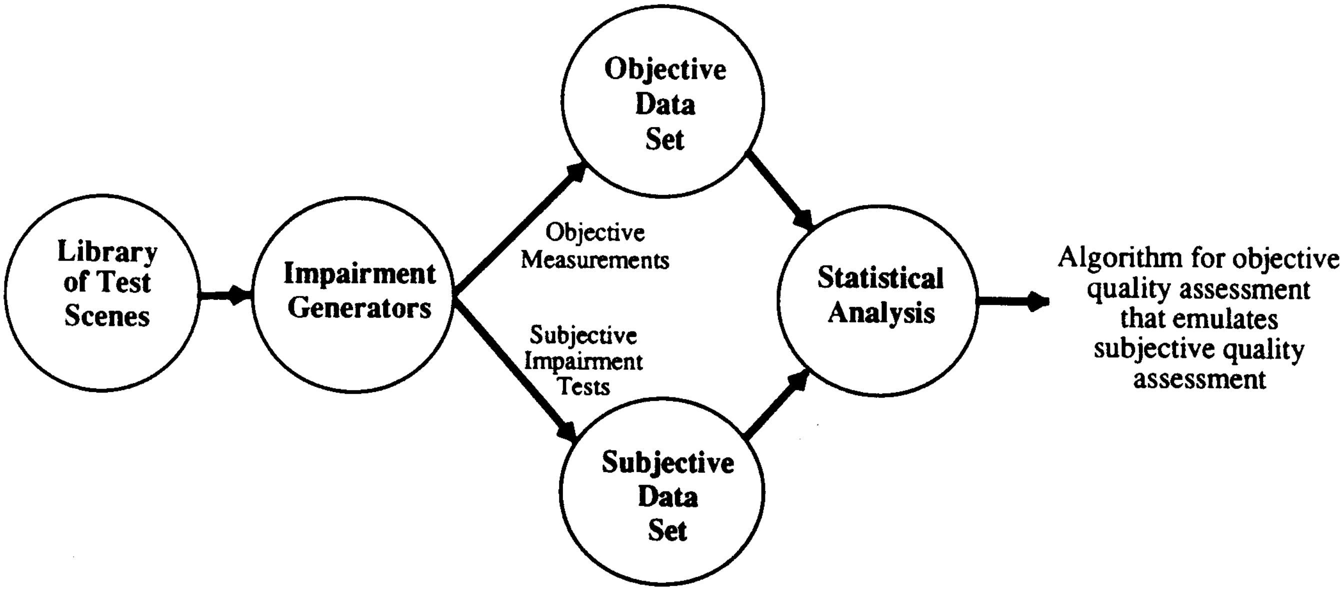

Figure 1. Measurement Derivation Process

First, a wide-ranging library of motion video test scenes is selected and recorded. Next, the scenes are dubbed to a second tape machine through representative video systems that distort the video signals and impair the resulting video images. The next step is a parallel objective and subjective measurement process. As the impaired video scenes are played back, they are digitized and a candidate set of objective measurements is computed from the digitized video data. These candidate measures are designed to extract and quantify the attributes of video scenes that are important to the human visual and perceptual systems. Data derived from these measures form the objective data set. The subjective measurement process involves playing the tapes for a panel of viewers. The viewers are asked to rate their perceptions of the impairments and these results form the subjective set. By utilizing the "null impairment" both the viewers and the objective measurement algorithm have access to the original, unimpaired version of each scene. This allows the use of differential measurement techniques: the viewers perform an "A versus B" comparison test, and the computer can calculate difference images of the form A-B.

The final step of the process is a simultaneous statistical analysis of the subjective and objective data sets. This analysis should indicate which of the candidate objective measurements provide unique information that is useful for predicting subjective assessments and which measures provide redundant or extraneous information, and hence can be discarded. The results of this analysis will be used to build a prediction algorithm that generates accurate predictions of subjective quality using the set of "good" objective measurements as inputs. This set of "good" measurements and the prediction algorithm are the final outputs of the derivation. Together, they provide an objective video quality metric that emulates the human perceptual video quality metric. By using a wide range of scenes and impairments as inputs to the derivation process, it is hoped that the resulting objective metric will provide an accurate, repeatable, reliable technique for predicting subjective video quality over a wide range of video applications without involving human viewers. The following sections describe the steps of the derivation process in detail and provide some preliminary results.

3. Test Scenes and Impairments

The selection of the test scenes used in the derivation must be done carefully. In particular, the spatial and temporal information content of the scenes are critical parameters. These parameters play a crucial role in determining the amount of video compression that is possible, and consequently, the level of impairment that is suffered when the scene is transmitted over a fixed-rate digital channel.

Video compression schemes attain various degrees of compression by removing the spatial and temporal statistical redundancies of the video signal. For a highly detailed scene, where the scale of the details is comparable to the scale of the spatial sampling grid, and the details are hard to predict, the video signal has very little spatial statistical redundancy. In this case, little compression can be gained by removing the small amount of spatial statistical redundancy. If the signal must be compressed, picture quality must be sacrificed. On the other hand, a scene with few details or with highly predictable details can often be greatly compressed. A parallel situation exists in the time domain. Scenes with large amounts of unpredictable motion have little redundancy in their temporal statistics and little temporal compression can be expected. Scenes with no motion or limited motion are highly redundant in their temporal statistics and can often be greatly compressed without loss of quality. Thus, there are direct links connecting information content of a scene, potential for compression, potential for transmission at a given rate, and received quality.

In light of these links, fair and relevant video quality measurements must use video scenes with spatial and temporal information content that is consistent with the video services that the device or system under test was intended to provide. As an example, video scenes of a soccer game contain too much spatial and temporal information to be useful in testing the performance of a codec designed to provide video teleconferencing services over 56 Kbps lines. In order to derive the most general possible measures, our library of test scenes contains scenes with widely varying amounts of spatial and temporal information. Figure 2 shows the relative amounts of spatial and temporal information for some possible test scenes.

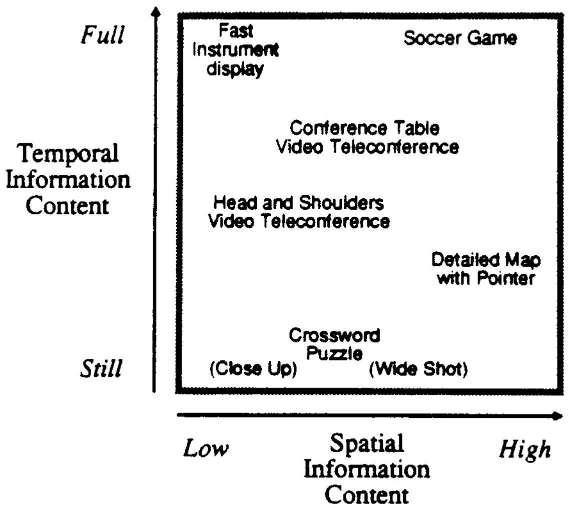

Figure 2. Information Content of Test Scenes

The subjective impairment tests require that viewers rate the impairment level of many versions of the test scenes. If this process fatigues and bores the viewers, their judgements may be less careful. Thus, it is important that the library of test scenes be fairly interesting. Our test scenes were recorded using a professional quality electronic news gathering camera and are stored on half-inch video tape in a component analog video (CAV) format. Our library also contains some test scenes from other labs which are NTSC encoded.

Since we are striving for a quality metric that works well across a wide range of video technologies, we have included a broad family of impairments in the derivation. The impairments include video codecs operating over simulated digital networks with controlled error rates and line rates that range from 45 Mbps down to 56 Kbps. Analog impairments include NTSC encode/decode cycles, VHS record/play cycles, and a noisy RF transmission link. All test scenes are subjected to all impairments to create a library of impaired test scenes.

4. Subjective Impairment Tests

The subjective data set is build from viewer judgements gathered in subjective impairment tests. These tests are conducted in a video viewing laboratory. The laboratory conforms to CCIR recommendation 500-3, which specifies the standard visual environment for conducting picture quality assessment in all respects except for the color temperature of room illumination.[3] The 14 x 11 x 8 foot room is finished with full length white drapes on three sides and gray drapes on the fourth (rear) wall. Gray carpet completes the subdued visual environment and provides noise reduction. An offset stud wall reduces noise transfer from an adjacent office area and a specially designed air handling system operates with a minimum of noise.

The six light fixtures in the laboratory are independently steerable and dimmable. This flexible lighting feature allows one to obtain the specified screen-to-background luminance ratio while simultaneously providing comfortable lighting conditions for writing. Recommendation 500-3 indicates that a light source with a color temperature of 6500 degrees Kelvin is preferred, presumably to match NTSC white. Several problems (power requirements, heat dissipation, bulb life, fixed light output) seem to be intrinsically linked to such sources. As a practical matter, we chose to use color-corrected incandescent bulbs. The tinted bulbs presently in use have a maximum color temperature of 4000 degrees Kelvin. This means that the "white" room illumination does not exactly match NTSC white. Given the adaptive nature of human color perception, effects of this fixed background color error are expected to be minimal. If necessary, color correction filters may be added in order to more closely approximate 6500 degree sources.

Video scenes are displayed on a broadcast quality monitor. The monitor displays CAV video signals with a maximum of 900 lines of horizontal resolution on a 19-inch screen. The monitor includes an option that allows repeatable monitor setup which is referenced to digitally stored values. This feature eliminates monitor setup as a potential source of variation in the subjective tests. Comfortable seating for three viewers is provided at a distance of six picture heights from the monitor screen. By restricting the number of viewing locations to three, the viewers in the end locations are only 20 degrees off-center and potential test variations due to viewer location are minimized. The laboratory is equipped with hidden studio quality speakers to allow for recorded voice instructions as well as future subjective video-with-audio quality assessment tests.

The following paragraphs describe the test methodology used in the subjective impairment tests. The methods are based on standard picture quality assessment procedures augmented by the results of our preliminary viewing sessions.[3][4] Those sessions used a viewer questionnaire to determine the testing rates and scales most likely to provide accurate responses.

Subjects for the impairment tests are selected at random from the site telephone directory. This gives a pool of 1,750 prospective viewers. The pool of viewers includes maintenance workers, office workers, administrators, scientists, and engineers. The use of a large, random sample of viewers from this diverse pool should substantially reduce any occupational bias that might be present. By recording the occupation of each viewer as a potential correlate, subjective impairment ratings within occupation can be calculated and results can be weighted to approximate those of other populations. In addition, each subject's exposures to several types of video, including broadcast and cable television, video teleconferencing and computer video displays are recorded as potential correlates. In accordance with CCIR Recommendation 500-3, each viewer is given a visual acuity test and a color vision test. The results of these tests along with the age and sex of the viewer become part of the viewer's data record.

Each viewer participates in four twenty-minute sessions. These sessions are conducted on four consecutive days. The first third of the initial session is a training period. The training period exposes the subjects to a wide range of video impairments, many of which may be new to the subjects, as well as a wide range of scene types that can also affect the perceived severity of the impairments. The final portion of the training period allows the subjects to practice marking the response form between scenes. The remainder of the initial session and the subsequent three sessions contain 30-second impairment tests. Here the subject is presented with nine seconds of a scene, three seconds of grey screen, nine seconds of the impaired same scene, and finally a nine second period in which to mark the response form. The subjects are instructed to decide on and mark the level of impairment in the second scene, using the first scene as a reference. The five possible responses offered are: imperceptible, perceptible but not annoying, slightly annoying, annoying, and very annoying. This scale intentionally covers a very wide range of impairment levels in a nonlinear fashion. By including reference scenes, impairment tests take advantage of the fact that the human eye excels at making comparisons. Impairment tests also tend to reduce inter-laboratory testing variances. To reduce unwanted comparison effects, the order of scene presentation is randomized.

After allowance for training periods, rest intervals, and some redundancy (to provide consistency checks), the cumulative 80 minutes of testing allow for the viewing and rating of 127 test scenes. In order to hold the subjects' interest, this body of 127 scenes is composed of 36 distinct scenes, with each scene appearing 3 or 4 times. Thus, the impact of 3 or 4 different impairments can be measured for each scene. The scene-impairment combinations are selected from the library of impaired test scenes described in the previous section.

5. Objective Impairment Measurements

The objective data set is built up from objective measurements or computations performed on the digitized (756 x 486 x 24 bits) video signal. The measurements are performed on every frame of each test scene. This intensive measurement approach, along with the need for large data sets, dictates that the measurements be automated. A controlling program with a windowed user interface passes instructions to device drivers that in turn control the tape machines, frame digitizer, video routing switcher, and video codecs. Additional software distributes the computation of measurements across several workstations to reduce the total measurement time.

In an exact parallel to the subjective tests, these measurements are all differential. That is, they involve both the impaired and the unimpaired versions of each scene. In order to make meaningful differential measurements, the two frame sequences must be aligned as closely as possible. Spatial alignment is determined by the video frame acquisition hardware and is accurate to a fraction of a pixel. Because there are unknown delays in the video systems used to create the impaired sequences, the temporal alignment of the two, 30 frame-per-second sequences is a non-trivial task. To further complicate the matter, many video codecs output the same frame two or more times. This frame repetition technique allows significant data reduction, but it also means that any possible one-to-one correspondence between the input and output sequences is lost. We have adopted the following technique for temporal alignment. For each output frame, match it with the input frame that is the closest (in terms of squared pixel error) and consistent with a causal system under test:

$$\underset{k:\ k \geq 0}{\mathrm{MIN}} \sum_{x=1}^{m} \sum_{y=1}^{n} [S(x,y,t-k) - D(x,y,t)]^2.$$

After the minimizing value of $k=k_0$ is found, the output frame D(x,y,t) is paired with the input frame $S(x,y,t-k_0)$. The value of k_0 is an estimate of the time delay of the system-under-test at time t.

Since we are seeking measurements that emulate those made by the human visual system, we are well-advised to mimic its properties whenever possible. The human visual system has greater resolving power on still scenes than on moving objects. Further, since the time-averaged information content of a still video scene is much less than the time-averaged information content of a moving scene, most compressed digital video systems have very different static and dynamic responses. In light of these observations, it is clear that measurements could be enhanced by partitioning each video frame into still and motion parts, performing separate measurements on each part and then recombining these measurements in an appropriate way. We have adopted this technique.

The partitioning of each frame is accomplished according to the following algorithm. To partition the k^{th} frame, first compute the absolute difference image, $|Frame_{k+1} - Frame_{k-1}|$. Then compare each pixel of the absolute difference image with a threshold value of 15. (The 8 bit luminance pixel values range from 0 to 255.) Those pixels exceeding the threshold value are declared to be motion pixels, and those below are considered to be still. A three by three dilation operation serves to smooth and fill the thresholded image, and the resulting binary motion mask indicates which regions of the k^{th} frame are still and which are moving. The threshold value of 15 was selected following a statistical analysis of 3 million motion pixels and 3 million still pixels.[5][6]

Once the video sequences have been digitized, time-aligned, and partitioned into motion and still regions, a family of over 90 differential measurements is computed and stored in the objective data set. These candidate measures were selected with an eye toward the following desirable properties: correlation with subjective quality, applicability to many types of scenes, value as a local estimate of quality in space and time, computational efficiency, stability, functional independence (each measure should provide different information), and technology independence (each measure should be useful for a wide range of video storage and transmission technologies). All video impairments can be described as distortions of the amplitude or the timing of the video waveform. On the other hand, when displayed on a monitor for human use, this one-dimensional voltage waveform is interpreted as a continuously evolving, multi-colored, two-dimensional signal. Useful measures must take note of this human interpretation and mimic it to the extent possible. Thus, our candidate set of measures includes those designed to measure temporal, luminance, chrominance, and spatial distortions. A detailed description of the set of candidate measurements is available.[7] These techniques are currently being considered for inclusion into the draft standard of "Analog Interface Performance Specifications for Digital Video Teleconferencing/Video Telephony Service" by the American National Standards Institute (ANSI) Accredited Standards

Committee T1, Working Group T1Q1.5, Video Teleconferencing/Video Telephony Sub-working Group.

Temporal distortions are perceived by the viewer as unnatural motion. Both the residual squared pixel error sequence given in the above equation, and the sequence of minimizing values of k_0 contain information which aids in quantifying unnatural motion. Color distortions are measured after transforming luminance and chrominance values to the International Commission on Illumination (CIE) LUV color space. In that 3-dimensional space, each color axis is perceptually independent and psychometrically uniform.[8] The remainder of this section provides some discussion and examples of spatial distortion measurements.

Investigators in the fields of human vision and object recognition have noted the importance of sharp edges in the visual perception and recognition of objects. Thus, an important class of spatial distortions are those that destroy, soften, blur, displace, or create edges in the video image. Our measurements of these effects utilize the Sobel and Laplace edge extraction or edge enhancement filters.[9][10] These filters are followed by differencing operators, energy accumulators, averagers, et-cetera, to create a family of measures that quantify edge distortions. As an example, the blurring of the sharpest edges can be measured by examining the decrease in the number of pixels in the filtered output that exceed a fixed threshold. More detailed examples of two spatial distortion measurements follow.

The measurements p_{77} and p_{60} are described mathematically at the end of this section. The measurement p_{77} has been named "Edge Fraction Gained, Still Portion", because it quantifies edges in the distorted frame that are not present in the original frame. The measurement is restricted to the still portion of each frame. The second measure, p_{60} is called "Absolute Edge Energy Difference, Motion-Still Weighted". Here a logarithmic energy difference measure is computed for both the motion and the still portions of each frame. The measures are passed through the absolute value operator and then combined using the weighting factors α and $(1-\alpha)$, which indicate the relative amounts of stillness and motion in the frame. The merit of p_{77} and p_{60} as predictors of subjective picture impairment is discussed in the following section.

$$p_{77} = -\text{mean\_}\{\text{still}(\text{Sobel}(S)) - \text{still}(\text{Sobel}(D))\},$$

$$p_{60} = \alpha \cdot 20 \cdot \left| \log_{10}\left\{\frac{\text{std}(\text{still}(\text{Sobel}(S)))}{\text{std}(\text{still}(\text{Sobel}(D)))}\right\}\right| + (1-\alpha)\cdot 20 \cdot \left| \log_{10}\left\{\frac{\text{std}(\text{motion}(\text{Sobel}(S)))}{\text{std}(\text{motion}(\text{Sobel}(D)))}\right\}\right|,$$

where:

S is the original video frame,
D is the distorted frame,
mean_ is the mean of the negative pixels,
still takes only the still parts of frame,
motion takes only the motion parts of frame,
std is the standard deviation of frame pixel values,
α = (number of still pixels)/(total number of pixels).

6. Analysis of Measurements

The final stage of the derivation process involves joint statistical analysis of the subjective and objective data sets. As of this writing, only preliminary work has been done in this stage. The step is intended to single out a subset of the candidate set of objective measures and provide information that aids in the design of the prediction algorithm. The members of this set of "good" measures should provide enough information to enable an appropriate algorithm to make accurate predictions of subjective impairment scores. Rather than predicting only the mean subjective impairment score, we hope to predict the fraction of persons that vote in each of the five impairment categories. From this predicted histogram we can compute predicted mean subjective impairment scores as well as predicted dispersions about that mean. Potential prediction, estimation and classification techniques include linear and quadratic predictors, hybrid linear-nonlinear decision structures (possibly adaptive), and Bayesian inferencing.

The remainder of this section gives the results of preliminary statistical analysis that was performed on relatively small data sets. For this preliminary study, we are utilizing the results of an experiment conducted by Fish and Judd.[11] Their team selected 5 NTSC encoded test scenes. Each scene consists of a three second still followed by five seconds of full motion video. They created two impaired versions of each scene: a VHS record-play cycle and a simulated codec operating at the DS1 signaling rate. The resulting 15 scenes were shown to 45 viewers who rated each scene in terms of its "distance from ideal". The researchers provided our lab with a copy of their test scenes and their subjective data set. We applied our family of objective measures to the test scenes to create a companion objective data set. This involved the processing of roughly four seconds from the motion part of each of the 15 scenes, resulting in approximately 120 values for each of 92 candidate measures. (One value for each frame of the four second sequence.) Since the subjective human assessments that we seek to emulate consist of a single value for the entire 4 seconds, we must reduce each temporal sequence of 120 objective measurements to a single value. This data reduction step should be done the same way human viewers do it. It seems reasonable that the median value of the sequence, augmented in some way by the minimum and maximum values would provide a good approximation. We are currently experimenting with a family of 10 data reduction functions. Included in this family of functions are mean value, maximum value, minimum value and median value. The final contents of the objective data set is a collection of 92 measures on 15 scenes for each of the 10 data reduction functions used.

Next we performed a correlation analysis between the objective and subjective data sets, and an analysis within the objective data set. Correlation analysis detects monotonic relationships between data sets. As relationships become more monotonic and closer to linear, the coefficient of correlation tends towards ± 1. A correlation coefficient (across the 15 scenes) was computed between the mean subjective impairment score and each of 92 candidate objective measures. We found absolute correlation coefficients larger than .8 for a large group of measures, but many of the objective measures are highly correlated with each other, indicating that all of them cannot contribute unique information to the prediction problem. If we select a subset of these measures by requiring that the absolute correlation coefficient between every possible pair of members of the subset be less than .9, we find 14 measurements in the subset. The correlation thresholds (.8 and .9) are somewhat arbitrary, but were chosen because they provide the "best family" of 14 objective measurements. If one were interested in larger or smaller families, one could lower or raise the appropriate thresholds. Of the 14 objective measurements mentioned above, two are those described in section 4. The coefficient of correlation between subjective score and the median value of the objective measurement called "Edge Fraction Gained, Still Portion" (p_{77}) is .96. For "Absolute Edge Energy Difference, Motion-Still Weighted" (p_{60}) the correlation value is .94. In general, the median function seems to provide the best data reduction across time. The majority of the remaining 12 top measures provide additional information about edges lost and/or gained in the motion and/or still portions of the video scene. Some of them are linear measurements, some are quadratic, and some are logarithmic.

The correlation values attained indicate that for this set of scenes, a trivial, first order linear predictor would do a respectable job of predicting mean subjective impairment values from either measurement. While these preliminary results are encouraging, the data sets are much too small to draw any firm conclusions from the results. Larger data sets might yield lower correlation values and present a greater challenge in terms of designing an objective video quality prediction algorithm. We are confident that a sufficient set of measures can be found and that an accurate prediction algorithm can be designed. As of this writing, we are building up the data sets to enable a much more thorough analysis, and more sophisticated prediction algorithm design.

7. Conclusion

We have described a method for deriving an objective video quality metric that emulates the human video quality metric. The objective metric comprises a family of measurements that quantify spatial, temporal, luminance and chrominance distortions, followed by a prediction algorithm. The intent is that the derived metric will work well over a wide range of digital and analog video transmission and storage technologies. The metric promises to yield reliable predictions of perceived video quality without the effort and expense of polling a large group of human subjects in a controlled environment. This provides a valuable tool for persons who design video components, storage and delivery systems, and those involved in standards work.

Acknowledgements

Many persons have contributed directly and indirectly to the work described here. Major contributors to this research effort include Stephen Wolf, Margaret Pinson, and Arthur Webster. We express thanks to Robert Fish and Thomas Judd at Bellcore for providing us with video scenes and data from their experiment.

References

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

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

3. CCIR Recommendation 500-3, Method for the Subjective Assessment of the Quality of Television Pictures.

4. CCIR Report 405-5, Subjective Assessment of the Quality of Television Pictures.

5. S. Wolf, M. Pinson, S. Voran, and A. Webster, "Objective Quality Assessment of Digitally Transmitted Video," in Proceedings of IEEE Pacific Rim Conference on Communications, Computers, and Signal Processing, 1991.

6. S. Voran and S. Wolf, Motion-Still Segmentation Algorithm for VTC/VT Objective Quality Assessment, a contribution to the ANSI Accredited Standards Committee T1, Working Group T1Q1.5, Video Teleconferencing/Video Telephony Sub-Working Group, document number T1Q1.5/91-110, January 22, 1991.

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

8. 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.

9. R. C. Gonzalez and P. Wintz, Digital Image Processing, second edition, Addison-Wesley Publishing Company, 1987.

10. A. K. Jain, Fundamentals of Digital Image Processing, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1989.

11. R. Fish and T. Judd, A Subjective Visual Quality Comparison of NTSC, VHS, and Compressed DS1-Compatible Video, Proceedings of the Society for Information Display, Volume 32, Number 2, (1991).