White Paper: The Status of Objective Metrics

Bill Reckwerdt, CTO
Video Clarity, Inc.

This paper explores the challenges associated with assessing video quality. The challenge is modeling the customer's reaction on their new TV. Many factors affect the video before it gets to the TV: compression, image processing, scaling, decoding, transmission, etc. To isolate their video algorithm, companies perform verification using: subjective and objective video assessment techniques.

- Subjective assessment consists of bringing a group of experts into a room, and asking them which videos they like better. This is time consuming and costly.
- Objective testing requires an algorithm, which models the results of the video experts surveyed above.

While it would be far cheaper, to perform only objective test, nothing beats the human eye. Thus, Objective video quality measurements and Subjective video quality assessment are complementary rather than interchangeable. Subjective assessment is appropriate for research related purposes; objective measurements are required for equipment specifications, QA testing, and monitoring.

Problem Statement

When dealing with equipment to process TV & video transmissions, every design depends on accurate, repeatable measurements. A complex relationship between objective parameter measurements and subjective video quality exists. The goal is to achieve an objective metric; that is an automated measurement. All of the equipment must be tested from video processors, compression units, transmission gear, set-top boxes, and displays. The evaluation of the video quality and ultimately, the customer's reaction to the picture shown on their new HDTV drives the business.

Previous Options

For years, traditional techniques that looked at color, brightness, contrast, etc. were effective. However, the advent of compressed digital video transmission has complicated the process of evaluating video sequences, with respect to perceived picture quality. During compression, a certain amount of the original content is knowingly discarded. Visible impairments such as "blockiness" and Gaussian noise are by-products. Traditional measurement techniques that look at

color, brightness, contrast, etc. are no longer effective.

Objective Measurement Status

After analyzing the subjective results, considerable work has been done to come up with a quantifiable, repeatable measurement which is not dependent on the video sequence. To date, objective measurements have not proven to estimate the user's opinion. To introduce and qualify new algorithms, Video Quality Experts Group (VQEG) was formed in 1997, and generally acts in cooperation with ITU. VQEG has conducted two phases of testing; in the first phase ten algorithms were tested, and the conclusion reached was that all of them were statistically equivalent. A second phase of testing, conducted several years later, involved a smaller set of algorithms, more controlled video sequences, and a better defined test environment. The result of the second phases warranted the recent ITU-T Recommendation J.144.

Three basic types of objective video assessment metrics exist:

- Full Reference – A method that conducts a comparison of video source to resultant.
- Reduced Reference – A method that conducts a comparison of a reduced video source to a full result.
- No Reference – A method when there is no reference.

The three methods have different applications, and they provide different degrees of measurement accuracy, expressed in terms of correlation with subjective assessment results.

The work to date has centered on full reference algorithms. Full reference algorithms perform a detailed comparison of the input and output video sequence. This is a computationally intensive process, which involves per-pixel processing, and temporal, spatial alignment of the input and output streams. Full reference algorithms can achieve good levels of correlation with subjective test data. Having the reference video sequence available is only possible for certain applications: for example in lab testing, pre-deployment test or troubleshooting.

One of the earliest full reference algorithms is PSNR (Peak Signal to Noise Ratio), which is literally a measurement of the mean error between input and output as a ratio of the peak signal level, expressed in dB. A typical "good" PSNR is around 30dB and it is generally accepted that PSNR values of less than 18dB are unacceptable. PSNR is the most widely used technique for image and video quality measurement.

A wide range of full reference algorithms have been developed including: MPQM (Moving Pictures Quality Metric - 1996) from EPFL in Switzerland, the US Government NTIA ITS lab's VQM (Video Quality Metric – 1999), Sarnoff's JND (Just Noticeable Differences), and Wang's SSIM ('Structural SIMilarity). ITU-T J.144 does not actually specify a single algorithm but "provides guidelines on the selection of appropriate" techniques. J.144 does contain descriptions and test results for four full reference algorithms. The VQM algorithm from the US Government's NTIA ITS lab achieved slightly better performance than the other algorithms listed.

The Author

Bill Reckwerdt has been involved in digital video since the early 90's from digital compression, video on demand, to streaming servers. He received his MS specializing in Behavioral Modeling and Design Automation from the University of Illinois Urbana-Champaign.

He is currently the VP of Marketing and the CTO for Video Clarity, which makes quantitative, repeatable video quality testing tools. For more information about Video Clarity, please visit their website at http://www.videoclarity.com..