"To suppose that the eye with all its inimitable contrivances for adjusting the focus to different distances, for admitting different amounts of light, and for the correction of spherical and chromatic aberration, could have been formed by natural selections, seems, I confess, absurd in the highest degree." ~ Charles Darwin (English Naturalist and Author of the theory of evolution by natural selection. 1809-1882)

Chapter 3

VIDEO QUALITY AND ITS ASSESSMENT

Work on picture quality metric goes back almost 50 years. Most of the quality metrics proposed over time are quality metrics for still images [**Wu 2006**]. First models of human vision were based on single-channel approach, in which the human visual system is represented by a single spatial filter characterized by the contrast sensitivity function. Mannos and Sakrison [**Mannos 1974**] developed the first image quality metric for encoded monochrome images. They considered integral square type of distortion measures, calculated the rate-distortion function and simulated the optimum encoding of a given image at a given bit-rate by varying different coding parameters. They also took into account some of the well-known characteristics of spatial vision and contrast sensitivity and developed a mathematical model for the human visual system, which was a closed-form expression for the contrast-sensitivity as a function of spatial frequency. The input images were filtered with this function after applying amplitude nonlinearity. Squared difference between the two images was the distortion measure. This was one of the first works in engineering that applied vision science in image processing.

The first color image quality metric was proposed by Faugeras [**Faugeras 1979**]. He presented a simple model of human color vision that quantitatively described different perceptual parameters such as brightness, hue, saturation etc. The idea was to structure the perceptual space as a vector space with spatial filtering properties, and introduced a norm on the vector space that would allow us to measure distances and to define a distortion measure in agreement with perceptual evaluation.

Soon after the introduction of a color image quality metric, Lukas and Budrikis [**Lukas 1982**] developed the first video quality metric based on a vision model. The first stage of the model constituted a nonlinear spatiotemporal model of a visual filter describing threshold characteristics on uniform background fields. The second stage incorporated a masking function to account for the non-uniform background fields. The model attempted to predict the subjective quality of moving monochrome television pictures containing arbitrary impairments.

In this chapter, various methods for determining and assessing the video quality are reviewed. These methods are broadly categorized into two groups, one that is based on engineering approach, relying on identifying and quantizing specific features or artifacts, and one that is based on psychophysical approach, involving the modeling of human visual system. Various video quality metrics based on color difference, image difference and image appearance modeling are also categorized under psychophysical approach, since they invariably take into account various mechanisms associated with the human visual system. Over the years, various international organizations have attempted to standardize video quality assessment methods and metrics. Some were more successful than the others. These standards are briefly reviewed in this chapter. The concluding section reviews several publications on subjective assessment of video quality.

3.1 Engineering Approach

In this approach, metric design is primarily based on extracting and analyzing some specific features or artifacts in the video. These are also called objective image quality metrics. The quality metrics developed using this approach can be further classified into three groups, full-reference, reduced-reference and no-reference [**Wu 2006**]. Full-reference metrics require the entire reference video to be available. Reduced-reference metrics require information on several features from the reference video. No-reference metrics do not use any information about the reference video.

As described in Chapter 2, most artifacts encountered in video are a direct result of compression. Miyahara et al [**Miyahara 1998**] presented a new methodology to obtain a picture quality scale (PQS) for coded achromatic still images. The authors used Weber-Fechner's law and the contrast sensitivity for achromatic images over the full range of image quality defined by the subjective mean opinion score. Some of the visual perception properties related to global image impairment were used while weighting the perceptually important structured and localized errors. The resulting PQS had high correlation with subjective mean opinion scores.

The work of Wu and Yuen [**Wu 1997**] focused on a specific compression artifact, the blocking artifact, which is quite common in compressed video. The authors presented a no reference, generalized block-edge impairment metric that enabled evaluation of reconstructed picture with blocking artifacts. The metric took into account luminance masking effects in extremely bright or extremely dark areas in a reconstructed image. The performance of this metric was found to be consistent with subjective evaluation.

Tan et al [**Tan 1998**] proposed a two-stage full-reference objective measurement model for MPEG-coded video. The first stage applies low-pass spatial filtering and spatial masking to each frame in the decoded picture and then computes perceptual impairment, taking into account the reference picture. The second stage simulates human visual system's processing of visual information, and thus acts as a cognitive emulator. The model was tested with several coded video sequences of 2-3 min duration. The model, when compared with PSNR and SSCQE, provided a closer approximation to the latter. This work put an emphasis on the need of a cognitive emulation stage in the objective measurement model.

Caviedes et al [**Caviedes 2000**] used three impairment metrics for MPEG video quality assessment, namely blocking artifact level, ringing artifact level and corner outlier artifact level, and created a combined impairment metric. The metric was used for subjective and objective quality assessment. The authors concluded that before we could map such a deterministic metric onto a probabilistic perceptual space and thus use it in closed-loop quality control system, subjective test methods needed to be improved in order to increase resolution and certainty of quality prediction.

Caviedes and Oberti [**Caviedes 2003**] developed a no-reference objective quality metric for measuring improved and degraded video. The authors followed a heuristic, incremental approach to modeling quality and training the model using a variety of video sequences. The method involved dividing the training sequences into impaired, enhanced and impaired-enhanced sets in order to deal with individual impairments and enhancements.

3.2 Psychophysical Approach

In this approach, metric design is primarily based on modeling of human visual system. These metrics are essentially full-reference metrics. Ahumada [**Ahumada 1993**] presented a comprehensive overview of different vision model based quality metrics for monochrome still images. The review was however limited to image quality metrics relying on image difference between the original and a corrupted version of it. Further, several more vision model based metrics have been proposed since the time of this publication.

Lindh and van den Branden Lambrecht [**Lindh 1996**] presented a vision model for moving pictures. The model was a more advanced version of a simple spatiotemporal model proposed earlier [**van den Branden 1996**]. It accounted for the normalization of cortical receptive field responses and inter-channel masking. Two quality metrics for video were derived from the new and old model and were used to assess the coding quality of MEPEG-2 video streams. The new model yielded a better quality rating.

One of the earliest video quality metric was offered by Sarnoff Just-Noticeable Difference (JND) model. The model aims to provide accurate estimate of the visibility of differences between original and distorted image sequences without requiring direct measurement using human observers [**Lubin 1997**]. The model was based on physiological and psychophysical principles of human visual discrimination performance, and thus was applicable to a varied range of distortions. The paper discussed model performance in a range of video applications involving discrimination and fidelity assessment.

One of the most well known models for video quality was DVQ, an acronym for Digital Video Quality, described by Watson et al in 1999 [**Watson 1999**]. It was similar to Sarnoff JND model, but significantly different in implementation. DVQ metric attempted to incorporate several aspects of early visual processing into a simple image processing algorithm, including light adaptation, luminance and chrominance channels, spatial and temporal filtering, spatial frequency channels, contrast masking and probability summation. It accelerated the spatial filtering operation by using Discrete Cosine Transform (DCT), making use of already available efficient hardware and software.

In the subsequent publication following the proposal of DVQ metric, Watson and others [**Watson 2001**] reported new visual data on the visibility of dynamic DCT quantization noise. This was obtained by using an image composed of a square array of 8x8 pixel blocks. A DCT basis function of the same frequency was placed within each block, and over a sequence of frames, each basis function was temporally modulated by a Gabor function of a particular temporal frequency and phase. Human visual threshold for the DCT noise was measured through psychophysical experiments. The data were then fitted with a mathematical model and incorporated it into the DVQ metric. The DVQ metric was tested comparing its predictions to judgments of impairment for a video stream of 65 sequences to the quality estimates provided by 25 human subjects. DVQ model performed similarly to the Sarnoff model. However, a systematic failure of prediction was also observed. Like the Sarnoff model, the DVQ metric is aimed at predicting the probability of detection of threshold image differences. It does not include appearance modeling through spatial or temporal adaptation, or correlates of appearance attributes and thus, cannot be used for video rendering as well [**Fairchild 2005**].

Winkler presented a perceptual distortion metric for color video sequences [**Winkler-1 1999**]. It was based on a contrast gain control model of the human visual system incorporating spatial and temporal aspects of vision as well as color perception. After conversion of the input video sequence to the opponent color space, each of the resulting three components was subjected to a spatio-temporal perceptual decomposition, yielding a number of perceptual channels. They were weighted according to contrast sensitivity data and sent through a contrast gain control stage. Both the reference and the processed sequence were input to the model. Finally, all the sensor stages were combined into a distortion metric. The metric was used to assess the quality of MPEG coded sequence. The model achieved a close fit to contrast sensitivity and contrast masking data from several different psychophysical experiments for both luminance and color stimuli.

In an important contribution to the volume of literature on perceptual video quality, Winkler summarized the issues in vision modeling for perceptual video quality assessment (PVQA) [**Winkler-2 1999**]. The author described quality factors in relation to a human observer, and how to measure them. Then the human visual system was described, along with the process to incorporate its component or phenomenon in a vision model for PVQA. The validation and evaluation of PVQA systems were also discussed by the author.

Traditional visual quality metrics measure image fidelity, i.e. the accuracy of the reproduction of the original image, instead of perceived quality. Winkler [**Winkler-3 2001**] investigated the addition of image appeal attributes to the metric in order to account for the perceived video quality. Sharpness and colorfulness were identified as important subjective attributes of quality and were integrated into the perceptual distortion metric [**Winkler-1 1999**]. These attributes were quantified by a sharpness rating based on local contrast and a colorfulness rating based on the distribution of chroma and saturation in the video sequence. Several subjective tests were performed to determine the relationship between these ratings and perceived visual quality. The results showed an improvement in the prediction of perceived quality by including sharpness and colorfulness ratings.

Yu et al published a review on human visual system (HVS) based digital video quality metrics [**Yu 2000**]. Three objective video quality metrics that represented the state-of-the-art of HVS based quality metric research were chosen for review, namely Sarnoff JND, Watson's DVQ and Perceptual Distortion Metric (PDM) proposed by Winkler. These models were also tested and verified by VEQG. Watson emphasized on easy implementation while formulating DVQ metric. As a result it was found to minimize the calculation and memory requirements, but at the cost of performance. Both Sarnoff JND and PDM incorporated recent results in vision research and showed similar overall performance. However, several critical aspects of HVS were missing from these models. These included i) temporal mechanism such as motion and light adaptation, spatial frequency adaptation, backward masking etc, ii) a thorough analysis of the opponent color space that models the perceptual pathways, iii) a working contrast gain control model which explains inter and intra-channel masking over temporal, spatial frequency and orientation bands and iv) supra-threshold quality metric.

Brill et al [**Brill 2004**] recognized the variability implicit in the psychophysical subjects and came up with techniques for determining the statistical resolving power of a video quality metric (VQM), defined as the minimum change in the value of the metric for which subjective test scores show a significant change. The primary data used in the analysis were subjective scores of various video-source materials subject to various kinds of digital-processing distortion. Original subjective mean opinion scores were converted to a common interval scale, and then the VQM scores were transformed to this common scale through statistical analysis. Fitting all VQMs to one scale provided a way for cross-calibration of those VQMs, in other words transformation of one VQM to another. Statistical probability was used to assess the resolving power of VQM. These new methods for assessing VQM accuracy and cross-calibrating VQMs were incorporated into the ATIS series of Technical Reports, which provided a comprehensive framework for characterizing and validating full-reference VQM.

3.2.1 Image Quality Metric Based on Image Difference

In an attempt to offer a comprehensive model for the complex process of image contrast judgment, Daly [**Daly 1993**] introduced an algorithm for the prediction of visual differences between two digital images based on a model of human visual system. The goal was to assess the image fidelity and develop an algorithm that could be used for the design and analysis of image processing algorithms, imaging systems etc. It consisted of three parts, amplitude nonlinearity, contrast sensitivity function and a hierarchy of detection mechanism. The algorithm was tested for a wide variety of image distortions including synthetic images designed for the purpose of psychophysical experiments and natural images with practical distortions. The sources of image distortion included blur, noise, data compression artifacts, banding, blocking, contouring, low frequency non-uniformities, hyper-acuity and tone-scale changes.

3.2.2 Image Quality Metric Based on Color Difference

Research on color difference equations had an important effect on the conceptualization of image quality metrics. Color difference research culminated with the introduction of CIEDE2000, but long before that, S-CIELAB presented the first incarnation of an image difference model based on CIELAB color space and color difference equations. Zhang et al [**Zhang 1996**] introduced this as a spatial extension to the CIELAB color metric for measuring color reproduction errors of digital images. The goal was to apply a spatial-filtering operation to the color-image data in order to simulate the spatial blurring by the human visual system. The model was essentially a spatial pre-processor to the standard CIE color difference equations to account for complex color stimuli such as halftone patterns. In case of large uniform areas, results had to be consistent with CIELAB calculations. To achieve this, the image data were first transformed into the opponent space, representing a luminance channel and two chrominance channels, red-green and yellow-blue. Then each opponent color images was convolved with a one-dimensional kernel to that color dimension. The shape of each kernel was determined by the visual spatial sensitivity and was such that the area under each of those kernels integrated to unity. Finally, the filtered data were transformed into a CIEXYZ representation, followed by a conversion to CIELAB. The calculation effectively segregated patterns and colors because spatial transformation was independent of color transformation, as also suggested by psychophysical experiments on human visual perception with regard to simple colored patterns. Because of this feature, S-CIELAB could be implemented as a pre-processor to existing CIELAB based systems. The results reported by Zhang et all showed better consistence with visual evaluation of color image difference than that predicted by general CIELAB equation.

Tong et al [**Tong 1999**] proposed a video quality metric based on a single-resolution spatial, temporal and chromatic model of human contrast sensitivity. This metric was an extension of the S-CIELAB and was termed spatio-temporal CIELAB (ST-CIELAB). The metric was designed such that it fitted published contrast sensitivity data and reduced to CIELAB value for uniform color field. ST-CIELAB was tested by conducting psychophysical experiments with MPEG video sequences. The metric ratings were consistent with subjective assessments, but overestimated the visibility of blocking artifacts in color MPEG video.

3.2.3 Image Quality Metric Based on Image Appearance Modeling

Formulation and application of **image appearance models** essentially began with the image measurement. As imaging systems became more complex, there was a need for device-independent images measurements, which started with CIE colorimetry and evolved into cross-media image reproduction, involving device-independent color imaging, gamut mapping and color-accurate computer graphics rendering with spectral imaging. However, CIE colorimetry did not provide a complete solution for image specification under widely disparate viewing conditions. Thus, color appearance models were developed to extend CIE colorimetry to the prediction of color appearance. CIECAM97s and CIECAM02 are the most widely studied color appearance models. While these models saw some successful applications in image reproduction, they did not adequately address spatially complex image appearance and image quality problems. Color appearance models do not directly incorporate any of the spatial and temporal properties of human vision and the perception of images. They treat each pixel as independent stimuli. To address the issues of device independent color imaging and modern color management systems with regard to spatial properties of vision, image perception and image quality, the concept of image appearance models was introduced [**Fairchild 2003**]. Apart from the attributes such as lightness, brightness, colorfulness, chroma and hue, which are adequately predicted by color appearance models, image appearance models also encompass different image attributes like sharpness, graininess, contrast and resolution. Image appearance models are essentially based on uniform color space. One of the most well-known image appearance model is iCAM [**Fairchild 2004**]. This model was also extended to predict the appearance of digital video sequences and high dynamic range scenes. Implementation of a temporal low-pass filter was proposed to model the time-course of chromatic and light adaptation of rendering applications. Conversion of spatial filter to spatio-temporal filters for image difference and quality applications was also proposed. Image appearance models like iCAM employ image-wise predictors of lightness, chroma and hue, from which an image quality metric can be derived.

Table 3.1 summarizes various video quality metrics discussed in this chapter. Some of these metrics were based on psychophysical approach relying on vision modeling, and others were based on engineering approach. Assessment of the developed metrics either was subjective, based on psychophysical experiments, or objective, based on various statistical metrics, or a combination of the two. In some cases, previously collected visual data were used to test the model, which has also been classified as a subjective assessment in this table.

Table 3.1 A summary of various proposed video quality metrics

Authors	Approach (Psychophysical/ Engineering)	Assessment (Subjective/ Objective/ Both)	Description
Brill et al [**Brill 2004**]	Psychophysical	Subjective	Proposed techniques for determining the statistical resolving power and accuracy of a video quality metric (VQM) and thus for cross-calibrating various VQMs
Caviedes et al [**Caviedes 2000**]	Engineering	Both	Created a combined impairment metric for MPEG video quality assessment
Caviedes and Oberti [**Caviedes 2003**]	Engineering	Objective	Developed a no-reference objective quality metric for measuring improved and degraded video
Daly [**Daly 1993**]	Psychophysical	Objective	Introduced an algorithm for the prediction of visual differences between two digital images based on a vision model
Lindh et al [**Lindh 1996**]	Psychophysical	Subjective (visual data fitting)	Presented a vision model for moving pictures
Lubin [**Lubin 1997**]	Psychophysical	Subjective (visual data fitting)	Operation and general structure of Sarnoff JND model was described
Miyahara et al [**Miyahara 1998**]	Engineering	Both	Presented a new methodology to obtain a picture quality scale (PQS) for coded achromatic still images
Tan et al [**tan 1998**]	Engineering	Objective	Proposed a two-stage full-reference objective measurement model for MPEG-coded video
Tong et al [**Tong 1999**]	Psychophysical	Subjective (visual data fitting)	Proposed a video quality metric based on various models of human contrast sensitivity
Watson et al [**Watson 1999**]	Psychophysical	Subjective (visual data fitting)	Described Digital Video Quality (DVQ) metric
Winkler [**Winkler-1 1999**]	Psychophysical	Subjective (visual data fitting)	Presented a perceptual distortion metric for color video sequences
Winkler [**Winkler-3 2001**]	Psychophysical	Subjective (visual data fitting)	Integrated image appeal attributes like sharpness and colorfulness into the perceptual distortion metric proposed in [Winkler-1 1999]
Wu and Yuen [**Wu 1997**]	Engineering	Objective	Presented a no reference, generalized block-edge impairment metric

3.3 Standardization of Video Quality Assessment and Metrics

Since the early nineties, the rapid evolution of digital video technologies posed a significant challenge to the performance measurement task, which necessitated the development of a new measurement methodology for testing the performance of digital video systems. With this goal, American National Standard (ANSI) approved a standard in 1996, called ANSI T1.801.03. This standard provided much needed set of objective quality metrics that showed high correlation with subjective evaluations of digital video impairments. Wolf [**Wolf-1 1997**] presented an overview of different parameters that were a part of the American National Standard (ANSI) T1.801.03. These parameters were technology-independent and were relevant for a wide range of digital video compression, storage and transmission.

ANSI T1.801.03 did not include MPEG video systems and did not cover bit rates between 1.6 and 10 mbps. Wolf et al [**Wolf-2 1997**] presented the results from two MPEG studies, MPEG 1 and MPEG 2 codecs, whose bit-rates ranged from 1.5 to 8.3 mbps. An analysis of the results revealed that the objective video quality metrics primarily measured four features, addition of false edges, lost sharpness of edges, added motion and lost motion. A set of three or four of these measures achieved a high correlation with subjective responses. The authors concluded these could be used as effective predictors of subjective quality ratings for entertainment video systems.

Since digital video quality depends upon the dynamic characteristics of the input video and the digital transmission system, an accurate perception-based measurement must be performed in-service. The Institute of Telecommunication Sciences (ITS) developed spatial-temporal distortion metrics that were valid over a wide range of quality and that could be used for in-service quality monitoring [**Wolf-3 1999**]. It was essentially a reduced reference quality metric, for which reference information was extracted from the spatial-temporal region. Horizontal and vertical edge enhancement filters to estimate spatial gradient emphasized edges as long as 10 arc min, and suppressed a large amount of noise. Spatial-temporal feature compression factor of 384 was achieved. Two separate visual masking functions emulated human perception. Seven subjective data sets spanning a wide range of bit rates, test scenes and digital video systems were used to evaluate the metrics. The metrics accounted for a large percentage of variance of the mean opinion scores. The size of the spatial-temporal region could be adjusted to match the bandwidth of the in-service data channel.

IEEE Broadcast Technology Society Subcommittee on Video Compression Measurements initiated an approach to the issue of video quality assessment with the aim of developing a scale of video impairment and unit of measure to describe video distortion from both perceptual and engineering standpoint [**Libert 2000**]. The aim was to set a standard for specifying and testing new equipment used in television production and broadcasting. It was proposed that the IEEE study would attempt to define a scale of video impairment in terms of multiple measurements of the just-noticeable difference (JND) of compression-induced video impairments. The subcommittee agreed on using actual video clips with a single type of impairments for the subjective test. The results of this proposed study was never reported. However, in response to the perceived urgent need in the industry for the sanctioned guidance on video quality, the Telecommunications committee of Alliance for Telecommunications Industry Solutions (ATIS) released a series of technical reports. One of those reports provided full implementation details for Sarnoff's JND model [**ATIS 2001**].

The Video Quality Experts Group or VQEG was formed in October 1997 in Turin, Italy to address video quality issues. The first task undertaken by VQEG was to provide a validation of objective video quality measurements methods leading to recommendations in both the Telecommunication (ITU-T) and Radiocommunication (ITU-R) sectors of the International Telecommunications Union (ITU). VQEG outlined, designed and executed a test program to compare subjective video quality evaluations to the predictions of a number of proposed objective measurement methods for video quality in the bit rate range of 768 kbps to 50 mbps [**VQEG 2000**]. VQEG solicited submission of objective models to be included in an ITU verification process leading to one or more ITU recommendations. It required all models to receive as input a processed sequence and its corresponding source sequence. Based on this input, the model was supposed to provide one unique figure of merit that correlated with the value obtained from subjective assessment of the proposed material. A set of test sequence was selected and two test sequences for subjective and objective evaluations were executed in parallel. Psychophysical experiment was performed on a total of 287 viewers to collect the subjective data, while ten objective models were evaluated through statistical analysis with respect to three aspects of their ability to estimate subjective assessment of video quality, namely prediction accuracy, prediction monotonicity and prediction consistency. The result of the test did not find an objective measurement system that was able to replace subjective testing. Depending on the metric used for the evaluation, the performance of eight or nine out of ten models was found to be statistically equivalent, leading to the conclusion that no single model outperformed the others in all cases. One of the major achievements of the first validation effort by VQEG was the unique data set assembled to help future development of objective models.

3.4 Subjective Assessment of Video Quality

Subjective assessment involves human observers who evaluate, compare or assess the quality of a given video. Although more time and resource intensive than an objective assessment, it is the most reliable way to determine perceived picture quality. Various aspects of video quality testing have been dealt in detail in [**Wu 2006**].

Corriveau et al [**Corriveau 1999**] used different test methods to evaluate the same video materials, and compared the stability of measurements in presence of contextual effect (due to varied impairments). Double Stimulus Continuous Quality Scale (DSCQS), Double Stimulus Impairment Scale method variant II (DSIS II) and comparison method were used in this study. DSCQS was found to be free from contextual effect, while comparison method had moderate and DSIS II had large contextual effect.

Pinson and Wolf [**Pinson 2003**] used data sets from six different subjective video quality experiments performed with single stimulus continuous scale evaluation (SSCQE), double stimulus continuous quality scale (DSCQS) and double stimulus comparison scale (DSCS) methods. A subset of video clips from each of these six experiments were combined and rated in a secondary SSCQE subjective video quality test. It was found that SSCQE with hidden surface removal and multiple randomized viewer orderings produced quality estimates comparable to DSCQS or DSCS value, which indicated that viewers performed the same error pulling function in all methods. The study also showed that the SSCQE test subjects utilized at most 9 to 15 seconds of video. Based on their findings, the authors concluded that properly designed SSCQE testing might be an effective substitute for more complicated DSCQS method.

Pearson [**Pearson 1998**] reported a three-stage method of measuring time-varying video quality. The first stage was an SSCQE method of instantaneous quality, the second stage was a calibration process to convert SSCQE to DSCQS metric, and the last stage was a numerical procedure for relating continuous and overall quality. It was found in this study that three effects played a role in subjective tests. These are: i) a forgiveness effect (momentary occurrence of poor quality ignored after a period of time, ii) a recency effect (quality in last 10-20 seconds of the presentation has a significant influence on overall perceived quality), and iii) a negative-peak effect (more influence of depth rather than the duration of the negative peaks in quality).

Zhao and de Haan presented a subjective assessment of various de-interlacing techniques [**Zhao 2005**]. Five typical algorithms were used in split screen paired comparison experiments. The test sequences included stationary, horizontal and vertical moving sequences, and object with complex motion. Rankings of different methods were derived from the experiment and were compared with objective scores using peak signal-to-noise ratio (PSNR). Since the subjective and objective scores were highly correlated, the authors concluded that objective performance criteria like PSNR are good predictors of quality of reconstructed video resulting from various de-interlacing techniques.

3.5 Conclusions

As this chapter illustrates, a vast body of literature is available on objective video quality metrics. However, the development of reliable metrics is still in a nascent stage, with many challenging issues remaining to be resolved [**Wu 2006**]. More comparative analysis is needed to evaluate the prediction performance of metrics. More experiments with natural images need to be conducted so that more visual data are available for vision modeling purposes. The focus of such experiments should be on supra-threshold conditions, rather than on the threshold of visibility. A reliable perceptual video quality metric will eventually help in benchmarking various video processing techniques. This will require coordinated research efforts in the areas of human vision, color science and video processing.