# Visual-quality estimation using objective metrics

**Mylène C. Q. Farias**

***Abstract*** — In the past few years, a big effort in the scientific community has been devoted to the development of better image- and video-quality metrics that correlate well with the human perception of quality. In this paper, an overview of the main ideas used in the design of objective quality metrics is given. More specifically, we briefly describe the different types of objective metrics and present a representative set of the different approaches taken by these algorithms. Finally, the challenges and recent developments in the area of image and video quality are discussed.

***Keywords*** — *Video quality, image quality, artifact, quality measurement, quality metrics, compression, artifacts.*

DOI # 10.1889/JSID19.11.1

## 1 Introduction

The most accurate way to determine the quality of a video or an image is by measuring it using psychophysical experiments with human subjects.[1] Unfortunately, psychophysical experiments are very expensive, time-consuming, and hard to incorporate into a design process or an automatic quality of service control. Therefore, the ability to measure quality accurately and efficiently, without using human observers, is highly desirable in practical applications. Good visual-quality metrics can be employed to monitor video or image quality, compare the performance of image- and video- processing systems and algorithms, and to optimize the algorithms and parameter settings for a image- and video-processing system.

Objective visual-quality metrics can be classified as *data metrics,* which measure the fidelity of the signal without considering its content, or *picture metrics,* which estimate quality considering the visual information contained in the data. Customarily, quality measurements in the area of image processing have been largely limited to a few data metrics, such as the mean absolute error (MAE), the mean square error (MSE), and the peak signal-to-noise ratio (PSNR), supplemented by limited subjective evaluation. Although over the years data metrics have been widely criticized for not correlating well with perceived quality measurements, it has been shown that such metrics can predict subjective ratings with reasonable accuracy as long as the comparisons are made with the *same content*, the *same technique,* or the *same type of distortions*.[2–4] Because of their simplicity and physical significance, the use of data metrics is fairly standard in the published literature and it will probably not disappear in a near future, especially in areas such as broadcasting and coding.

One of the major reasons why these simple metrics do not generally perform as desired is because they do not incorporate any human-visual-system (HVS) features in their computation.[3,4] It has been discovered that, in the primary visual cortex of mammals, an image is not represented in the pixel domain, but in a rather different manner. Unfortunately, the measurements produced by metrics such as MSE or PSNR are simply based on a pixel-to-pixel comparison of the data, without considering what is the content and the relationships among pixels in an image (or frames). They also do not consider how the spatial and frequency content of the impairments are perceived by human observers. In the past few years, a big effort in the scientific community has been devoted to the development of better image- and video-quality metrics that incorporate HVS features (*i.e., picture metrics*) and, therefore, correlate better with the human perception of quality.[5–9]

Picture metrics can be divided into three different categories according to the availability of the original (reference) image or video signal: full reference (FR) metrics, reduced reference (RR) metrics, and no-reference (NR) metrics. Figures 1–3 show the block diagrams corresponding to the FR, RR, and NR quality metrics, respectively. On the FR approach, the entire reference is available at the measurement point. Quality metrics with best performances are generally FR picture metrics that try to incorporate aspects of the HVS considered relevant to quality, such as color perception, contrast sensitivity, and pattern masking. Unfortunately, the usage of FR metrics is restricted to studio applications since in many practical applications (*e.g.*, video conference, internet streaming, broadcasting, *etc.*) the reference is not available.

For the RR approach, only part of the reference is available through an auxiliary channel. In this case, the information available at the measurement point generally consists of a set of features extracted from the reference image or video. Metrics in this class may be less accurate than the FR metrics, but they are also less complex and make real-time implementations more affordable. Nevertheless, they still require some amount of information to be available at the receiver and synchronization. Requiring the

Based on a paper presented at the 30th International Display Research Conference (IDRC '10) held November 16–19, 2010, in São Paulo, Brazil, in conjunction with Latin Display 2010.
The author is with the University of Brasilia (UnB), Department of Computer Science, Campus Universitário Darcy Ribeiro, Asa Norte ICC Centro, Caixa Postal 4466, Brasilia, DF 70910-900 Brazil; telephone +55-61-8151-4957, e-mail: mylene@unb.br.

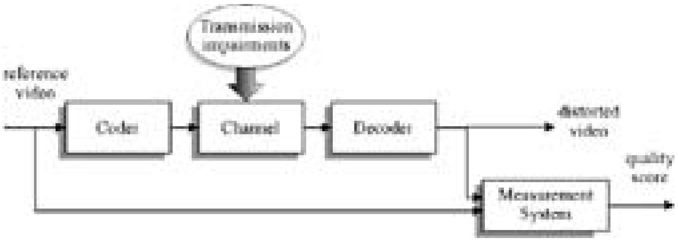


**FIGURE 1** — Block diagram of a full-reference (FR) objective quality metric.

reference or even a small portion of it becomes a serious impediment in many real-time transmission applications. For the NR approach, the reference image or video is not available and the quality estimation is obtained exclusively from the test signal. It turns out that, although human observers can usually assess the quality of an image or video without the reference, designing a NR metric is a difficult task. Considering the difficulties faced by the FR metrics, this is no surprise.

Picture metrics can be further classified according to the approach they use for estimating the quality of an image or video. There are basically three main approaches: error sensitivity, feature extraction, and statistical. The *error-sensitivity* approach analyzes visible differences (errors) between a test and a reference signal and gives an estimate of the quality of the test in comparison to the reference. This approach is mostly used by FR metrics since it is assumed that an undistorted (perfect quality) reference image or video is available at the measurement point. The error-sensitivity FR metrics are actually *similarity* or *fidelity* measures since they actually provide measures of how similar (or different) the two images or videos are. As a consequence, they are frequently incapable of identifying improvements in quality produced, for example, by an enhancement algorithm.

The *feature extraction* approach looks for higher-level features or attributes of the image or video (*e.g.*, sharpness/blur, contrast, fluidity, artifacts, *etc.*) that are considered relevant to quality.[10] These algorithms measure the magnitude of high-level features in order to estimate quality. A number of NR and RR metrics use a specific type of the feature extraction approach, which estimates the strength of the most perceptually relevant artifacts (perceptual defects) and combine these values in order to obtain a quality estimate. These specific metrics are referred as *application-specific* metrics and can only be used for the target application. In contrast, models that are designed

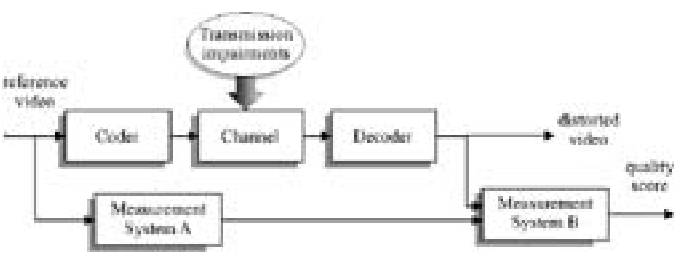


**FIGURE 2** — Block diagram of a reduced-reference (RR) objective quality metric.

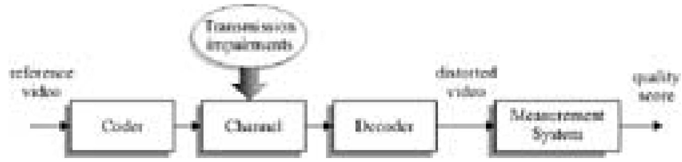


**FIGURE 3** — Block diagram of a no-reference (NR) objective quality metric.

using regular features or assumptions about the HVSs are known as *general-purpose* metrics. The feature extraction approach can also be designed to identify degradations or quality improvements through the analysis image attributes such as contrast, blurriness/sharpness, colorfulness, fluidity, *etc.*

The *statistical* approach consists of extracting statistical measures from the images or videos in order to obtain a quality estimation. The statistical measures may include mean, variance, covariance, kurtosis, *etc.* Since natural (undistorted) images and videos correspond to a small subset of all possible signals, the statistical algorithms try to identify if the test image or video present deviations from the statistics of natural images or videos.

The approaches discussed in this section are equally valid for image-and video-quality metrics. However, the video-quality metrics must take into account both spatial and temporal information. Also, video applications might contain artifacts that vary over the time and, therefore, a characterization of temporal errors and their impact on perceived quality is necessary. Unfortunately, this model is not usually implemented. Since a video can be seen as sequence of still images, most approaches in the literature simply extend the quality model for still images and apply it individually to every frame of the video. The scores obtained for each frame are combined using a pooling mechanism. Pooling techniques can range from a simple average of the frame scores to more complex methods that weigh frames differently according to their motion information.

In this paper, we give an overview of the main ideas used in the design of picture metrics targeted at visual applications. More specifically, we briefly describe the different types of objective metrics and present a representative set of the different approaches taken by these algorithms. Finally, we discuss the challenges and recent developments in the area of image and video quality metrics.

## 2 Full-reference video-quality metrics

Full-reference (FR) video-quality metrics estimate the quality of a test image or video by comparing it with the reference (original). This approach assumes that the reference video has no degradations and that it is available at the measurement point. As previously mentioned, FR video-quality metrics have the best performance among the three types of metrics (FR, RR, and NR). This is mainly due to the availability of the reference video at the measurement point. Also, since FR metrics are intended for *off-line applications*, they can be more computationally complex and incorporate

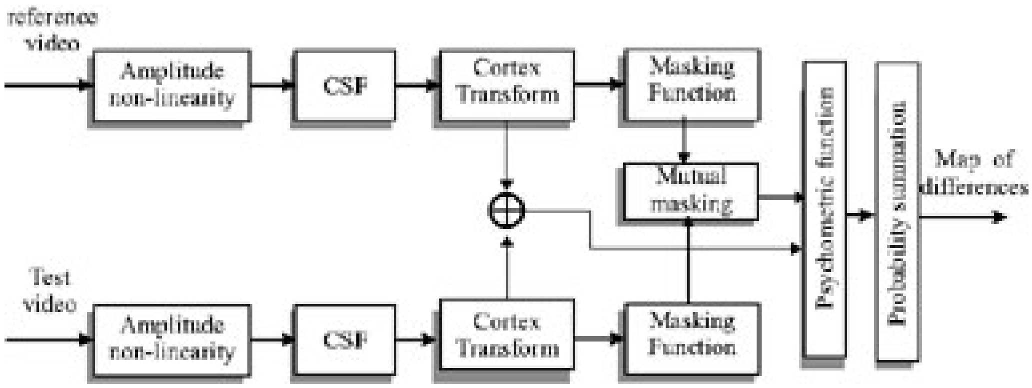


**FIGURE 4** — Block diagram of the visible-differences predictor (Ref. 5).

several aspects of the HVS. The major drawback of the FR approach is the fact that a large amount of reference information has to be provided at the final comparison point. Also, a very precise spatial and temporal alignment of reference and test videos is needed to guarantee the accuracy of the metric.

In this section, we give a brief description of the most common FR approaches, exemplified here by three different approaches of FR quality metrics: an *error-sensitivity approach*,[5] a *feature-extraction approach,*[9] and a *statistical approach*.[6]

## 2.1 Visible-differences predictor (VDP)

The FR model proposed by Daly is known as visible differences predictor (VDP).[5] This FR metric is a good example of a metric that uses an *error-sensitivity* approach, which makes an attempt to analyze and quantify the error signal in a way that simulates the human-quality judgment. Daly's algorithm finds the main parameters that limit the visual sensitivity and takes them into account when analyzing the differences between test and reference videos. The main sensitivity limitation (or variations) parameters considered by the model are *light level, spatial frequency,* and *signal content*. Each of these sensitivity variations corresponds to one of the stages of the model, as described below:

- **Amplitude non-linearity:** It is well known that sensitivity and perception of lightness are non-linear functions of luminance. The amplitude non-linearity stage of the VDP describes the sensitivity variations as a function of the gray scale. It is based on a model of the early retina network.
- **Contrast-sensitivity function (CSF):** The CSF describes the variations in the visual sensitivity as a function of spatial frequency. The CSF stage changes the input as a function of light adaptation, noise, color, accommodation, eccentricity,[a] and image size.
- **Multiple-detection mechanism:** It is modeled with four subcomponents:
  - **Spatial cortex transform** – Models the frequency selectivity of the visual system and creates the framework for multiple detection mechanisms. This is modeled by a hierarchy of filters modified from Watson's cortex transform that separates the image into spatial levels followed by six orientation levels.[11]
  - **Masking function** – Models the magnitude of the masking effect.
  - **Psychometric function** – Describes the details of the threshold.
  - **Probability summation** – Combines the responses of all detection mechanisms into a unified perceptual response.

A simplified block diagram of the VDP is depicted in Fig. 4. The output of Daly's metric is a probability-of-detection map, which indicates the areas where the reference and test images differ in a perceptual sense.

## 2.2 NTIA video-quality metric (VQM)

The video-quality metric (VQM) is a FR video-quality metric proposed by Wolf and Pinson from the National Telecommunications and Information Administration (NTIA).[9] The algorithm used by VQM uses a *feature-extraction approach* that includes measurements for the perceptual effects of several video impairments, such as blurring, jerky/unnatural motion, global noise, block distortion, and color distortion. These measurements are combined into a single metric that gives a prediction of the overall quality.

The VQM algorithm can be divided into the following stages:

- **Calibration:** This first stage has the goal of calibrating the video in preparation for the feature-extraction stage. With this purpose, it estimates and corrects the spatial and temporal shifts, as well as the contrast and brightness offsets of the processed video sequence with respect to the original video sequence.

[a]Eccentricity is the angle relative to the point of sharpest seeing, the fovea, is referred to as eccentricity. The more eccentric an object, the more difficult it is to see sharply. The fovea occupies the central 1° of vision.

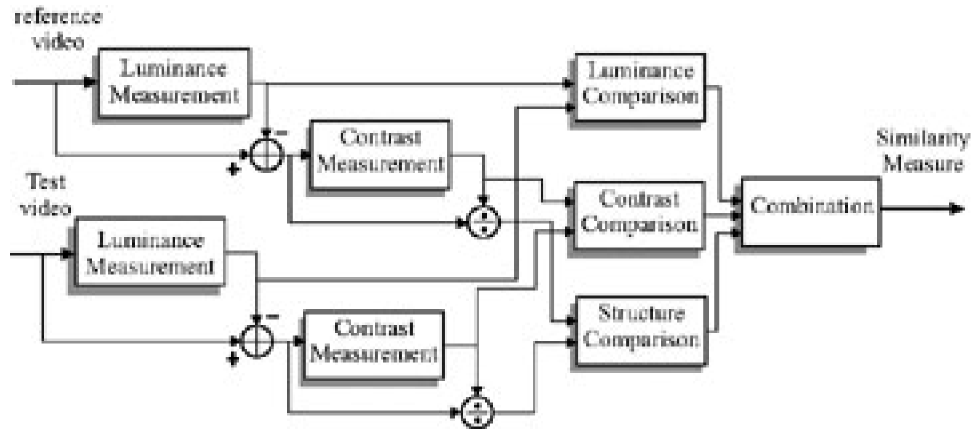


**FIGURE 5** — Block diagram of the SSIM algorithm (Ref. 6).

- **Extraction of quality features:** In this stage, the set of quality features that characterizes perceptual changes in the spatial, temporal, and chrominance domains are extracted from spatial-temporal subregions of the video sequence. For this, a perceptual filter is applied to the video to enhance a particular type of property, such as edge information. Features are extracted from spatio-temporal (ST) subregions using a mathematical function, and then, a visibility threshold is applied to these features.
- **Estimation of quality parameters:** In this stage, a set of quality parameters that describe the perceptual changes is calculated by comparing features extracted from the processed video with those extracted from the reference video.
- **Quality estimation:** The final step consists of calculating an overall quality metric using a linear combination of parameters calculated in previous stages.

The VQM was one of the best performance FR metrics tested by Video Quality Experts Group (VQEG) in Phase II.[7] It has recently been adopted by ANSI as a standard for objective video quality.

## 2.3 Structural similarity and image quality (SSIM)

The structural similarity and image quality (SSIM) is based on the idea that natural images are highly "structured".[6] In other words, image signals have strong relationships among themselves, which carry information about the structures of the objects in the scene.

To estimate the similarity between a test image and the corresponding reference, the SSIM algorithm measures the luminance $l(x, y)$, contrast $c(x, y)$, and structure $s(x, y)$ of the test image $y$ and the corresponding reference image $x$, using the following expressions:

$$l(x,y) = \frac{2\mu_y \mu_x + C_1}{\mu_y^2 + \mu_x^2 + C_1},$$

$$c(x,y) = \frac{2\sigma_y \sigma_x + C_2}{\sigma_y^2 + \sigma_x^2 + C_2},$$

and

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_y \sigma_x + C_3},$$

where $C_1$, $C_2$, and $C_3$ are small constants given by $C_1 = (K_1 \cdot L)^2$, $C_2 = (K_2 \cdot L)^2$, and $C_3 = C_2/2$. $L$ is the dynamic range of the pixel values (for 8-bits/pixel gray-scale images, $L = 255$), $K_1 << 1$, and $K_2 << 1$. The general formula of the SSIM metric is given

$$SSIM(x,y) = l^{\alpha}(x,y) \cdot c^{\beta}(x,y) \cdot s^{\gamma}(x,y),$$

where $\alpha$, $\beta$, and $\gamma$ are parameters that define the relative importance of the luminance, contrast, and structure components. If $\alpha = \beta = \gamma = 1$, these parameters are given equal importance. The SSIM has a range of values varying between "0" and "1", with "1" being the best value possible. A study on the performance of SSIM has shown that this simple metric presents good results.[6] A block diagram of the SSIM algorithm is depicted in Fig. 5.

## 3 Reduced-reference video-quality metrics

Reduced-reference (RR) video-quality metrics require only partial information about the reference video. To help evaluate the quality of the video, certain features or physical measures are extracted from the reference and transmitted to the receiver as side information. One of the interesting characteristics of RR metrics is the possibility of choosing the amount of side information. In practice, the exact amount of information will be dictated by the characteristics of the side channel that is used to transmit the auxiliary data or, similarly, by the available storage to cache them. Bit rates of the RR channel can go from zero (for no-reference metrics) to 15, 80, or 256 kbps, depending on the application in target.

Metrics in this class may be less accurate than the FR metrics, but they are also less complex and make real-time implementations more affordable. Nevertheless, synchronization between the original and impaired data is still necessary. Frequently, RR metrics are designed from FR metrics that use a *feature-extraction* approach. This is the case for the RR metric described in this section.

### 3.1 NTIA reduced-reference video-quality metric

One of the earliest RR metrics was proposed by Webster *et al.* from NTIA.[12] Their metric is a feature extraction metric that estimates the amount of impairment in a video by extracting localized spatial and temporal activity features using especially designed filters. The block diagram of this metric is depicted in Fig. 6. As expected, this metric is very similar to the VQM metric described in the previous section.

The authors define two types of features: spatial information (SI) and temporal information (TI). The SI feature corresponds to the standard deviation of edge-enhanced frames, which assumes that the inserted degradation will modify the edge statistics in the frames. The TI feature corresponds to the standard deviation of difference frames, *i.e.*, the amount of perceived motion in the video scene. As shown in Fig. 3, at the transmission side or the encoding point, these features are extracted from the reference video and sent over an auxiliary data channel. The size of the RR data depends upon the size of the window over which SI and TI features are calculated. At the receiver side or comparison point, the features from the reference video are received. Then, the exact same features are extracted from the test video. Afterwards, three comparison metrics ($m_1$, $m_2$, and $m_3$) are derived from the SI and TI features of the reference and test videos and, finally, used to estimate the quality of the test video.

## 4 No-reference video-quality metrics

Requiring the reference video or even a small portion of it becomes a serious impediment in many real-time transmission applications. In this case, it becomes essential to develop ways of blindly estimating the quality of a video using a no-reference video-quality metric. It turns out that, although human observers can usually assess the quality of a video without using the reference, designing a no-reference metric is a very difficult task. Considering the difficulties faced by the FR video-quality metrics, this is no surprise. For this reason, most of the proposed NR metrics available in the literature use the *engineering approach* to estimate quality. One of the exceptions to this is the metric by Gastaldo *et al.* that uses a neural network to estimate the quality of a given test video.[13]

The *engineering approach* is similar to the feature extraction approach described earlier. But, in this case, a greater emphasis is given to the distortion analysis and, consequently, specific features of the video are extracted. These extracted features can be either *structural elements* (contours, lines, activity, *etc.*) or *artifacts* (specific distortions) introduced by a particular technology. In this approach, the algorithm estimates how pronounced these features are and uses this information to estimate the overall quality or degradation of the video.

Most NR video-quality metrics consider *artifacts* as features, *i.e.*, they estimate the strength of the artifact signals in the video and combine these measures to obtain an estimation of the amount of degradation in the video. The most popular artifacts considered by NR metrics are blockiness, blurriness, noisiness, and ringing. In this section, we describe the NR metric by Farias and Mitra,[14] which uses a multi-dimensional approach to combine the outputs from a set of artifact submetrics.

### 4.1 NR metric based on artifact measurements

The algorithm proposed by Farias and Mitra is based on the assumption that the perceived quality of a video can be affected by a variety of artifacts and that the strengths of these artifacts contribute to the overall annoyance.[14] This *multi-dimensional* approach requires a good knowledge of the types of artifacts present in digital videos and extensive studies of the most relevant artifacts. The authors performed a series of psychophysical experiments to understand how artifacts depend on the physical properties of the video and how they combine to produce the overall annoyance.

The block diagram of the proposed metric is depicted in Fig. 7. The algorithm is composed of a set of artifact submetrics (artifact physical-strength measurements) to estimate the strengths of blockiness, blurriness, and ringing/noisiness artifacts. The metrics are simple enough to be used in real-time applications, as briefly described below.

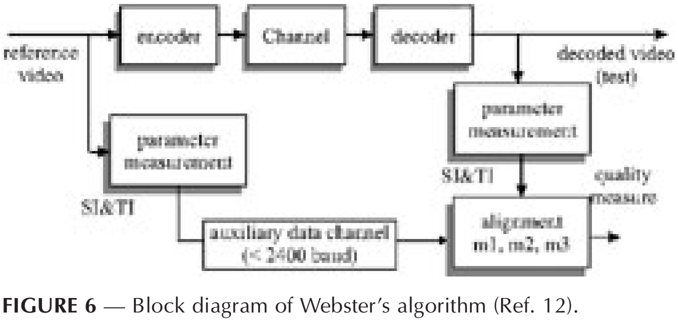


**FIGURE 6** — Block diagram of Webster's algorithm (Ref. 12).

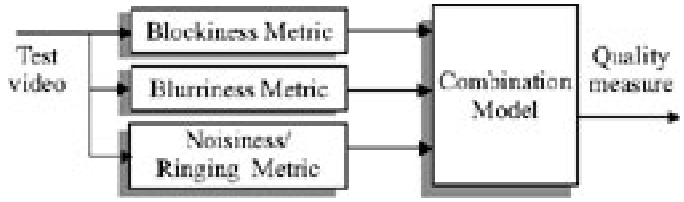


**FIGURE 7** — Block diagram of the no-reference metric proposed by Farias and Mitra (Ref. 14).

- The blockiness metric is a modification of the metric by Vlachos.[15] It estimates the blockiness signal strength by comparing the cross-correlation of pixels inside (intra) and outside (inter) the borders of the coding blocking structure of a frame.
- The blurriness metric is based on the idea that blur makes the edges larger or less sharp.[16] The algorithm measures blurriness by estimating the width of the edges in the frame.
- The noisiness/ringing metric is based on the work by Lee,[17] which uses the well-known fact that the noise variance of an image can be estimated by the local variance of a flat area. A cascade of 1-D filters was used as a pre-processing stage to reduce the content effect.

The performance of each artifact submetric is evaluated according to their ability to detect and estimate the artifact signal strength. Hence, test sequences containing (1) only the artifact being measured, (2) artifacts other than the artifact being measured, and (3) a combination of all artifacts are used in the design. The outputs of the individual metrics are also compared to artifact perceptual strengths gathered from psychophysical experiments. A model for overall annoyance is obtained based on a combination of the artifact metrics using a Minkowski metric.

## 5 Alternative approaches

In this section, we discuss three alternative approaches to quality metrics that have received a lot of attention in the last few years: hybrid metrics, saliency-based metrics, and data hiding metrics.

### 5.1 Hybrid video-quality metrics

A new trend in video-quality design is the development of hybrid metrics, which are metrics that estimate the quality of the video by analyzing the network information, the bitstream headers, and the decoded video.[8] The idea here is to improve the precision of regular *picture-quality* metrics by also considering parameters extracted from the transport stream and the bitstream headers (without decoding) in the computation. The main advantages of this approach are the lower bandwidth and processing requirements, when compared to metrics that only consider the fully decoded video. The disadvantage of hybrid metrics is that they are technology dependent, *i.e.,* their solution is only valid to the technology (compression, transmission, *etc.*) they were designed for.

One interesting example of this type of metric is the work by Winkler and Mohandas.[8] Their algorithm does not require the reference video and is targeted at MPEG-2 and H.264 video streaming over IP networks. To estimate the quality of a compressed and transmitted video, the authors define a V-factor that combines the information gathered by analyzing the transport stream (TS) headers, the packetized elementary stream (PES) headers, the video coding layer (VCL), and the decoded video signal. It is worth pointing out that the analysis of decoded video signal is done by using a picture-quality metric, similar to the ones described in the previous sections of this paper.

### 5.2 Video-quality metrics based on salient region detection

A recent development in the area of quality metrics consists of trying to incorporate the perceptual importance of the different areas of the video in the design of the metric. The assumption here is that the quality will suffer more if an impairment affects an "important" area of the video, rather than "non-important" areas.

The work by Oprea *et al.* uses this approach to weigh distortion measurements.[18] The first step of Oprea *et al.*'s algorithm is to find the perceptually important areas of the video frame. For this, the model estimates key features that attract attention: color contrast, object size, orientation, and eccentricity. The measurement of these features will determine the important (or salient) areas, producing a saliency map. Extracting saliency from video sequences is a complex task because both the spatial extent and the dynamic evolution of the regions should be considered. For the detected salient areas, a distortion measure is computed using a specialized no-reference metric. The metric considered by this algorithm is a blurriness metric.

### 5.3 Video-quality metrics using data hiding

Yet another alternative way of implementing a video-quality metric is to use data-hiding techniques. The idea consists of embedding into the original video the necessary information to estimate its quality at the time of display. One example of a metric that uses this approach is the work by Farias *et al.*[19] Figure 8 depicts the block diagram of this video-quality-assessment system.

At the transmitter, the algorithm embeds a mark in each frame of the video using a spread-spectrum technique.[20] In the embedding stage, a pseudo-random algorithm is first used to generate a mark, which is then multiplied by a scaling factor and added to the (mid-frequencies) DCT coefficients. At the receiver, the mid-frequency DCT coefficients (where the mark was inserted) are multiplied by the corresponding pseudo-random image. The result is then averaged over a number of frames and the mark is extracted. A measure of the degradation of the mark

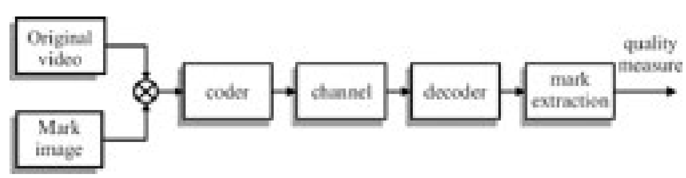


**FIGURE 8** — Block diagram of the video-quality-assessment system based on data hiding.

is given by the total square error (TSE) between the extracted mark and the original mark. The less the amount of errors caused by processing, compression or transmission, the smaller the error between the two marks.

## 6 Conclusions

In this paper, we gave an overview of the main ideas used in the design of objective visual-quality metrics. More specifically, we briefly described the different types of objective metrics and presented a representative set of the different approaches taken by these algorithms.

To date, most of the achievements have been in the development of FR video-quality metrics. Much remains to be done in the area of NR and RR quality metrics, which would certainly benefit from the incorporation of better perception models. Concerning the development of artifact metrics for NR and RR metrics, we believe there is still a great need to characterize the different types of artifacts that affect visual quality. In particular, there are very few works that investigate the interactions among the different types of artifacts or even attributes.

Concerning the area of video quality, there is a good number of algorithms that incorporate some sort of temporal information to estimate quality. Nevertheless, we believe improvements can be achieved by incorporating better models for motion perception, pooling, and visual attention. A new trend in video-quality design is the development of hybrid metrics, which are metrics that use a combination of the packet information, the bitstream header (without decoding), and the decoded video as inputs to the video-quality estimation algorithm.

With respect to applications, given the growing popularity of video delivery services over IP networks (*e.g.*, internet streaming and IPTV) or wireless channels (*e.g.*, mobile TV), there is a great need for metrics that estimate the quality of the video in these applications. Another area that has attracted attention is multimedia quality. So far, very few metrics have addressed the issue of simultaneously measuring the quality of all media involved (*e.g.*, video, audio, text). Since video and audio quality can influence each other,[21] quality metrics for the multimedia scenario cannot be a simple combination of an audio and a video metric.

Another area that has attracted a lot of attention is the quality assessment of 3-D videos.[22] Assessing the quality of 3-D videos is a big challenge. So far, it is not clear if the same approaches used for the 2-D case can be extended to the 3-D.

## References


1 International Telecommunication Union, "ITU-T Recommendation BT.500-8: Methodology for the subjective assessment of the quality of television pictures" (1998).

2 E. G. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Image Processing* **43**(12), 2959–2965 (1995).

3 Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it? A new look at signal fidelity measures," *IEEE Signal Processing Mag.* **26**(1), 98–117 (2009).

4 B. Girod, "What's wrong with mean-squared error?," *Digital Images and Human Vision,* ed. A. B. Watson (MIT Press, Cambridge, Massachusetts, 1993), pp. 207–220.

5 S. Daly, "The visible differences predictor: an algorithm for the assessment of image fidelity," *Digital Images and Human Vision,* ed. A. B. Watson (MIT Press, Cambridge, Massachusetts, 1993), pp. 179–206.

6 Z. Wang *et al.,* "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing* **13**, 600–612 (2004).

7 VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment – Phase II," *Tech. Report* (2003).

8 S. Winkler and P. Mohandas, "The evolution of video quality measurement: From PSNR to hybrid metrics," *IEEE Trans. Broadcasting* **54**(3), 660–668 (2008).

9 M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcasting* **50**(3), 312–322 (2004).

10 B. W. Keelan, *Handbook of Image Quality: Characterization and Prediction* (CRC Press, 2002).

11 A. Watson, "The cortex transform: rapid computation of simulated neural images," *Computer Vision, Graphics, and Image Processing* **39**(3), 311–327 (1987).

12 A. A. Webster *et al.,* "An objective video quality assessment system based on human perception," *Proc. SPIE Conference on Human Vision, Visual Processing, and Digital Display IV* **1913**, 15–26 (1993).

13 P. Gastaldo *et al.,* "Objective assessment of MPEG video quality: a neural-network approach," *Proc. Intl. Joint Conf. on Neural Networks* **2**, 1432–1437 (2001).

14 M. C. Q. Farias and S. K. Mitra, "No-reference video quality metric based on artifact measurements," *Proc. IEEE Intl. Conf. on Image Processing* **III**, 141–144 (2005).

15 T. Vlachos, "Detection of blocking artifacts in compressed video," *Electron. Lett.* **36**(13), 1106–1108.

16 J. Lu, "Image analysis for video artifact estimation and measurement," *Proc. SPIE Conf. on Machine Vision Applications in Industrial Inspection IX* **4301**, 166–174 (2001).

17 J. Lee and K. Hoppel, "Noise modeling and estimation of remotely-sensed images," *Proc. Intl. Geoscience and Remote Sensing* **2**, 1005–1008 (1989).

18 C. Oprea *et al.,* "Perceptual video quality assessment based on salient region detection," *Fifth Advanced Intl. Conf. on Telecommunications* (AICT '09), 232–236 (May 2009).

19 M. C. Q. Farias *et al.,* "A comparison between an objective quality measure and the mean annoyance values of watermarked videos," *Proc. IEEE Intl. Conf. on Image Processing* **III**, 469–472 (2002).

20 I. Cox *et al.,* "Secure spread spectrum watermarking for multimedia," *IEEE Trans. on Image Processing* **6**(12), 1673–1687 (1997).

21 J. G. Beerends and F. E. De Caluwe. "The influence of video quality on perceived audio quality and vice versa," *J. Aud. Eng. Soc.* **47**(5), 355–362 (1999).

22 P. Gorley and N. Holiman. "Stereoscopic image quality metrics and compression," *SPIE Stereoscopic Displays and Applications* **XIX**, 6803 (2008).



**Mylene C. Q. Farias** received her B.S. degree in electrical engineering from the Universidade Federal de Pernambuco (UFPE), Brazil, in 1995, and her M.Sc. degree in electrical engineering from the Universidade Estadual de Campinas (UNICAMP), Brazil, in 1998. She received her Ph.D. degree in electrical and computer engineering from the University of California Santa Barbara (UCSB), USA, in 2004, for work in no-reference video-quality metrics. She has worked as a research engineer at CPqD (Brazil) in video-quality assessment and validation of video-quality metrics. She has also worked as an intern for Philips Research Laboratories (The Netherlands) in video-quality assessment of sharpness algorithms and as a consultant for Intel Corp. (Phoenix, USA) developing no-reference video-quality metrics. Currently, she is an Assistant Professor of the Department of Computer Science at the University of Brasilia (UnB), Brazil. Her current interests include video-quality metrics, video processing, multimedia, watermarking, and information theory.