Are existing procedures enough? Image and video quality assessment: review of subjective and objective metrics

Ouni, Sonia, Chambah, Majed, Herbin, Michel, Zagrouba, Ezzeddine

Sonia Ouni, Majed Chambah, Michel Herbin, Ezzeddine Zagrouba, "Are existing procedures enough? Image and video quality assessment: review of subjective and objective metrics," Proc. SPIE 6808, Image Quality and System Performance V, 68080Q (28 January 2008); doi: 10.1117/12.760803

Event: Electronic Imaging, 2008, San Jose, California, United States

Are Existing Procedures Enough? Image and Video Quality Assessment: Review of Subjective and Objective Metrics

S. Ouni*#, M. Chambah*, M. Herbin*, E. Zagrouba#

* CReSTIC, Université de Reims Champagne Ardenne, France
E-mail : sonia.ouni@etudiant.univ-reims.fr
E-mail : {majed.chambah, michel.herbin}@univ-reims.fr

#URSIIVA, Institut Supérieur Informatique
E-mail: ezzeddine.zagrouba@fsm.rnu.tn

ABSTRACT

Images and videos are subject to a wide variety of distortions during acquisition, digitizing, processing, restoration, compression, storage, transmission and reproduction, any of which may result in degradation in visual quality.
That is why image quality assessment plays a major role in many image processing applications.
Image and video quality metrics can be classified by using a number of criteria such as the type of the application domain, the predicted distortion (noise, blur, etc.) and the type of information needed to assess the quality (original image, distorted image, etc.).
In the literature, the most reliable way of assessing the quality of an image or of a video is subjective evaluation [1], because human beings are the ultimate receivers in most applications. The subjective quality metric, obtained from a number of human observers, has been regarded for many years as the most reliable form of quality measurement. However, this approach is too cumbersome, slow and expensive for most applications [2].
So, in recent years a great effort has been made towards the development of quantitative measures. The objective quality evaluation is automated, done in real time and needs no user interaction. But ideally, such a quality assessment system would perceive and measure image or video impairments just like a human being [3].
The quality assessment is so important and is still an active and evolving research topic because it is a central issue in the design, implementation, and performance testing of all systems [4, 5].
Usually, the relevant literature and the related work present only a state of the art of metrics that are limited to a specific application domain. The major goal of this paper is to present a wider state of the art of the most used metrics in several application domains such as compression [6], restoration [7], etc.
In this paper, we review the basic concepts and methods in subjective and objective image/video quality assessment research and we discuss their performances and drawbacks in each application domain. We show that if in some domains a lot of work has been done and several metrics were developed, on the other hand, in some other domains a lot of work has to be done and specific metrics need to be developed.

Keywords: image quality evaluation, subjective and objective evaluation.

1. INTRODUCTION

The literature relating to image quality is vast and encompasses such different research areas in different domain applications. In this study, we present several image quality metrics, their advantages and disadvantages in different domains such as compression, communication, printing, displaying, analysis, registration, restoration, enhancement and watermarking.
Every domain application has specific degradations. For example, in the restoration domain we face the problem of color cast. The compression applications [8-11] present some artifacts like blur, noise, distortion and blocs. To evaluate the performance of a processing many metrics have been developed. There are two ways to evaluate the image quality. The most reliable way of assessing the quality of a video in different domain is the subjective

Image Quality and System Performance V, edited by Susan P. Farnand, Frans Gaykema,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6808, 68080Q, © 2008 SPIE-IS&T · 0277-786X/08/$18

approach because human beings are the ultimate receivers in most applications. But this approach presents some disadvantages.
That is why, in the last few years, new objective image and video quality metrics have been proposed in the literature. The objective quality metrics can be classified in three categories. An important criterion used in the classification of image quality measures is the type of information needed to evaluate the distortion in degraded images. Measures that require both the original image and the distorted image are called “full reference” or “non-blind” methods. Measures that do not require the original image are called “no-reference” or “blind” methods. Measures that require both the distorted image and partial information about the original image are called “reduced-reference” methods.
Although no-reference metrics are needed in some applications in which the original image is not available, they can be used to predict only a small number of distortion types. Reduced-reference measures are between full-reference and no-reference measures. The applicability of full-reference measures is much wider.
The performance of objective metrics depends on the degree of correlation with the subjective quality measures. A recent trend is to incorporate Human Vision System (HVS) features into the quality metrics to make the new measurements more consistent with human visual perception. This approach can be called hybrid evaluation wich is presented in figure 1.

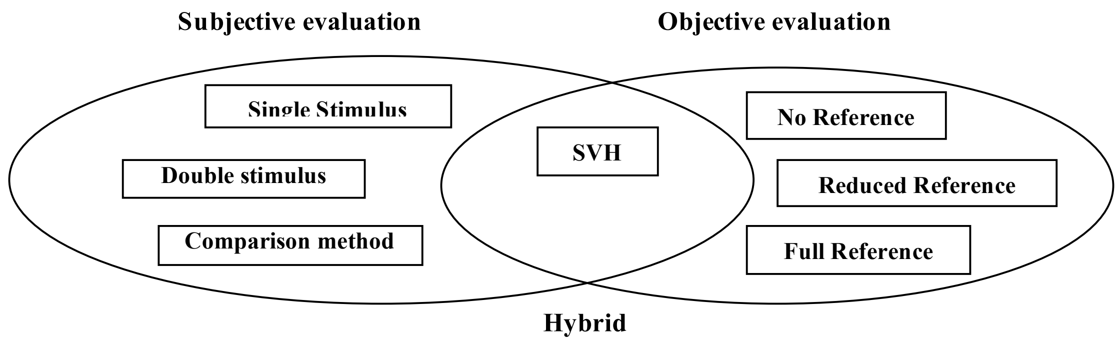

Figure 1. Classification of different quality metrics

Our state of art is different from common studies in the literature that focus on a specific application domain [12]. We present in this work a classification of different metrics used in several application domains. The organization of the paper is as follows: In the next section we will describe the subjective approaches and their different methods. In section 3, we describe the metrics and methods used in objective quality metrics in different application domains. In the last section, some conclusions are presented.

2. SUBJECTIVE EVALUATION

The subjective measurement is a widely used method for the assessment of image or video quality, but it has several obvious disadvantages. It is very tedious, expensive and impossible to be executed automatically. The subjective evaluation methods are divided into three primary categories: the first is Single Stimulus methods; the second is the Double Stimulus methods and finally the stimulus-comparison methods. All of these methods were based on ITU-R Recommendation (International Telecommunications Union).
Both ITU-R and ITU-T have been developing recommendations on subjective quality assessment methodologies. Originally, CCIR (the former name of ITU-R) was basically addressing methodologies for evaluating audio and video quality for broadcasting and entertainment services, while CCITT (the former name of ITUT) was merely addressing methodologies for evaluating speech quality in telephony. However, the conventional ITU-R methodologies have not been sufficient to assess short extract of digitally coded video material because quality fluctuates widely depending on the scene content and impairments that may be short lived. Therefore, the conventional ITU-R methodologies have been changed and a new methodology suitable for picture quality evaluation in digital video systems was added. The ITU-R, besides the Rec.BT.500, that provides the fundamental description of the subjective video quality assessment methods, has produced recommendations describing the complete procedures, based on the methods illustrated in recommendation BT.500-7 [13], but adapted to specific systems as High Definition Television - HDTV [14], and enhanced PAL and SECAM [15].

Such tests have been standardized in the ITU-R Recommendation BT.500-11 [16]. This standardization consists on:

- The number of observers: At least 15 observers should be used and they should be non-expert;
- Selection of test materials;
- The viewing distance and the screen sizes are to be selected in order to satisfy the PVD (Preferred Viewing Distance).
- Viewing environment;
- Presentation: The observers should be asked to look at the picture for the whole of the duration of T1 and T3.Voting should be permitted only during T4.

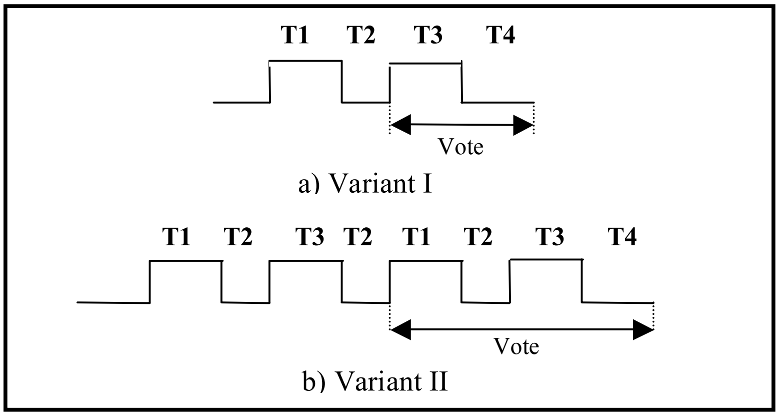

Figure 2. Presentation structure of Rec. ITU-R BT.500-11

Phases of presentation:
T1 = 10 s Reference picture
T2 = 3 s Mid-gray produced by a video level of around 200 mV
T3 = 10 s Test condition
T4 = 5-11s Mid-gray

- Session: The duration of the session is less than 30 minutes. There must be a reasonable time-gap (perhaps a half-day) between two sessions in the same group, unless the sessions last less than 20 minutes, in which case two sessions can be one after the other.
- Preliminary session: The first session for each group of observers. It has the same structure as any other session, but the results are not used.
- Presentation of results: The results should state the average and standard deviation of the scores for each test parameter. In addition, the following information should be given:
    1. Test configurations, reference and test cases used.
    2. Picture material used.
    3. Type and adjustment of the picture source(s).
    4. Type of display monitor(s).
    5. Type and number of observers.
    6. Number of groups and sessions.

We will present the different subjective measurement methods. Firstly, we present the single-stimulus method, then the Double Stimulus and finally the comparison methods.

2.1 Single Stimulus methods

In single-stimulus method, assessment is done on the image or sequence of images individually without any reference. The observers assign a value to each image or image sequence shown. There are two approaches:

- SS: With no repetition of test scenes.
- SSMR: The test scenes are repeated multiple times.

Two different methods are used:

- *Single stimulus continuous quality (SSCQ)*: used continuous scale with no numbers or a large range of 0 to 100 is used.
- *Single Stimulus Impairment Scale (SSIS):* In adjectival categorical judgments, observers assign an image or image sequence to one of a set of categories that, typically, are defined in semantic terms. The categories may reflect judgments of whether or not an attribute is detected. Categorical scales that assess image quality and image impairment have been used most often, and the ITU-R scales are given in figure 3.

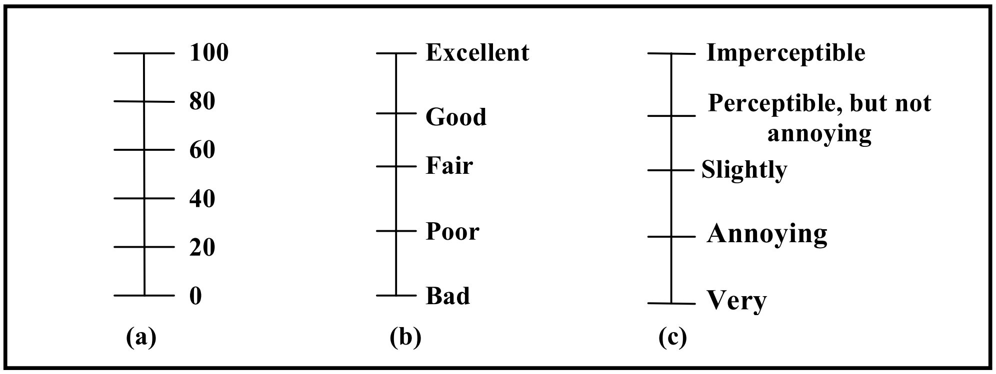

Figure 3. ITU-R Continuous scale (a), quality scale (b) impairment scale (c)

2.2 Double Stimulus methods

These methods can be applied in two variants:

- **Variant1:** Each assessor is let to switch between two conditions, A and B (two images), one of which is always the source and the other is the tested condition applied on the source. The identity of the images, whether it is the source or the test condition, should be known by the experimenter but not by the assessors. After evaluating the conditions the assessor moves to the next pair of images.
- **Variant2:** Multiple assessors are shown two conditions, A and B (two images), consecutively one of which is always the source and the other is the tested condition applied on the source. The identity of the images, whether it is the source or the test condition, should be known by the experimenter but not by the assessors. The next pair of conditions is shown after the assessors establish an opinion.

These methods can be divided in two methods: The first method, called Double Stimulus Impairment Scale – DSIS, the second is called Double Stimulus Continuous Quality Scale –DSCQS.

- **Double Stimulus Impairment Scale Method (DSIS or EBU)**: operates on the five level impairment grading. The reference image is always shown with the distorted one. Assessment of the images quality refers to the distortion level, not the absolute image quality.
- **Double-Stimulus Continuous-Quality Scale Method (DSCQS):** The picture quality is assessed on a continuous quality scale from excellent to bad. Experts are not informed which picture is the reference one, absolute image quality is assessed. In [16], DSCQS method is recommended for stereoscopic image pair assessment. However, in Video Quality Assessment [17] is the most used method.

2.3 Comparison method

In stimulus-comparison methods, two images or sequences of images are displayed and the viewer provides an index of the relation between the two presentations. In this method, the scale is based on a comparison within a set of images and the subject uses a comparison scale as shown in figure 4.

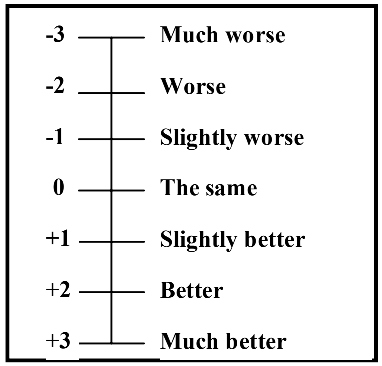

Figure 4. Comparison scale

These methods have generally different applications. The choice of a specific method depends on the context, the purpose and where, in the development process of the data the test is to be performed. DSCQS is the preferred method when the quality of the test and the reference sequence are similar, because it is quite sensitive to small differences in quality. The DSIS method is better suited for evaluating clearly visible impairments, such as artifacts caused by transmission errors. Single Stimulus Method is useful when the effect of one or more factors need to be assessed. The factors can either be tested separately or can be combined to test interactions. The Single Stimulus Continuous Quality Evaluation (SSCQ) method relates well to the time varying quality of today's digital video systems. The Stimulus Comparison Method is useful when two impaired images or sequences are required to be compared directly. This is the case for example when different image or video processing systems are compared on the basis of the visual quality of their results.
Subjective testing is currently the accepted method for establishing the quality of a particular processing algorithm. However, there are several difficulties associated with performing subjective quality tests [18, 19]. These intrinsic difficulties have driven the research on objective quality assessment. Some of these limitations are listed below:

- Most psychophysical experiments are conducted on simple patterns. But it is not known if a limited number of simple-stimulus experiments are sufficient to build a model that can predict the visual quality of complex structured natural images.
- Interactive visual processing (e.g. eye movement) influences the perceived quality. For example, subjects will give different quality scores if they are provided with different instructions. Prior information regarding the image content, or attention and fixation, may also affect the evaluation of the image or image sequence quality.
- Subjective tests are extremely time consuming and costly. Many groups do not possess the required equipment and have to conduct tests under non-standard conditions or in other laboratories. It is also difficult to obtain a large number of subjects. The process of subjective testing may take weeks or months, thus becoming a big limitation in the research of subjective quality assessment.
- A large number of subjects is required since there may be a large variation in individual viewing opinions, depending on the subject's age, sex, motivation and other personal factors.
- Subjective quality may vary depending on the length of the representation. Hamberg and de Ridder [20] found that subjects take around 1 second to react to a particular distortion in a scene and a further 2-3 seconds to stabilize their responses. Horita et al. [21] showed that distortions in the first or in the last part of the representation jeopardize the overall quality more than those appearing in the central part.
- The scale used by subjects can also introduce problems. For example, discrete scale with few levels asks for many subjects to reduce the variance. Subjects are usually reluctant to give very high or very low scores. For this reason open-ended scales may be used.

The most reliable way of assessing the quality of an image or a video is subjective evaluation, because human beings are the ultimate receivers in most applications. The Mean Opinion Score (MOS), which is a subjective quality metric obtained from a number of human observers, has been regarded for many years as the most reliable form of quality measurement. However, the MOS method is too cumbersome, slow and expensive for most applications. Consequently, researchers have attempted to find objective measures to evaluate the quality of the processed images. The objective image or video quality metric can provide a quality value for a given image or video automatically in a relatively short time. This is very important for real world applications. These objective metrics are the subject of the following section.

3 OBJECTIVE IMAGE QUALITY METRICS

Objective metrics quantify specific physical attributes of the imaging process. Objective image or video quality metric can provide a quality value for a given image or video automatically in a relatively short time. This is very important for real world applications. Objective quality assessment methods of digital image or video can be classified into three categories. In the first category the quality is evaluated by comparing the image or video sequence to the original. It is called the Full reference FR metric. The second category called No Reference NR metric contains methods that compare features calculated from the original and the image or video sequences reference. The methods of the third category make observations only on image reference or video and estimate the quality using only that information.

3.1 Full-reference (FR) metrics

FR metrics have been intensively studied in literature. Most existing approaches are known as full-reference, meaning that a complete reference image is assumed to be known. This approach is based on measurement of differences between the original image or video and the distorted one. The metrics of this approach can be classified into two categories the first based on mathematical measures and the second on characteristics of the human visual system (HVS).

3.1.1 Model based on mathematical measures

The easiest objective quality measures are simple statistics features on the numerical error between the reference and the distorted image. The simplest used full-reference quality metric is the mean squared error (MSE), mean absolute error (MAE), root mean square error (RMSE), mean absolute error (MAE), Signal to Noise Ratio (SNR) and Peak signal-to-noise ratio (PSNR). The most widely and appealing in image and video compression and transmission are MSE and the PSNR because they are simple to calculate, have clear physical meanings, and are mathematically convenient in the context of optimization. But they are not very well matched to perceived visual quality [22-24], and do not correlate well with subjective quality measures because human perception of distortions and artifacts is unaccounted for. Indeed, these metrics neglect the properties of the HVS and thus cannot be a reliable predictor of the perceived visual quality. The HVS takes into account the neighborhood pixels and consequently, any loss of information in the pixel neighborhoods could be a good measure of image fidelity.

Hence these measures fail to take the HVS properties into consideration. Some improvement on these measures have been proposed by some researchers, such as the New weighted Mean Square Error (NwMSE) [25] which is easy to calculate and applicable to various image processing applications. It permits to evaluate noisy images with different types of distortions like Gaussian noise, White Uniform noise. The NwMSE takes into account the pixels neighborhood information in an effective way.

In the compression domain, the decompressed image and video exhibit various kinds of distortion artifacts such as blocking, blurring and ringing. The human visual sensitivity to different types of artifacts is very different. In recent years, many metrics have been developed. However, most of them simply use the mean squared error (MSE) as the distortion measure. Since the MSE is not good for perceptual image quality assessment [27], several improved distortion measures have been proposed [28-29]. Wang present in [30] a novel method for measuring the similarity between two images called the Structural SIMilarity (SSIM) index. The SSIM index can be viewed as a quality measure of one of the images being compared, provided the other image is regarded as of perfect quality. All of these techniques require access to the original images. The VQEG (Video Quality Expert Group) has defined some measures [31] in order to qualify performance of image quality criterion in terms of accuracy (RMSE weighted by the confidence on the MOS; RMSEw)

The growth of digital video has given rise to a need for computational methods for evaluating the visual quality of digital video. All video quality metrics are inherently models of human vision. For example, if root-mean-squared-error (RMSE) is used as a quality metric, this amounts to the assumption that the human observer is sensitive to the summed squared deviations between reference and test sequences, and is insensitive to aspects such as the spatial frequency of the deviations, their temporal frequency, or their color. In the domain of compression digital video quality metric, [26] present a new metric called DVQ (Digital Video Quality). It accepts a pair of digital video sequences, and computes a measure of the magnitude of the visible difference between them. The metric is based on the Discrete Cosine Transform. It incorporates aspects of early visual processing, including light adaptation, luminance and chromatic channels, spatial and temporal filtering, spatial frequency channels, contrast masking, and probability summation. It also includes primitive dynamics of light adaptation and contrast masking. We have applied the metric to digital video sequences corrupted by various typical compression artifacts, and compared the results to quality ratings made by human observers. Video Quality Metric (General Model) (VQM) [32] is a video FR-metric adopted as standard by the American National Standards Institute (ANSI) in 2003. The International

Telecommunication Union (ITU) has also included the NTIA General Model as a normative method in two Draft Recommendations.
In the context of transmission/ processing, video quality is a characteristic of a video passed through a video transmission/processing system, a formal or informal measure of perceived video degradation (typically, compared to the original video). Video processing systems may introduce some amounts of distortion or artifacts in the video signal, so video quality evaluation is an important problem.

In the literature few frameworks have been interested to evaluate the image quality in the context of restoration. In [33] is presented a new metric called DAF for differential ACE filtering. This measure is a delta E distance in CIELab space under illuminant D65 computed between the original color image and its ACE filtered version.

The major advantage of the mathematical metrics is their simplicity. They can be very conveniently adapted by an image/video processing system. But, they don't take into consideration the HVS properties. In the last three to four decades, a great deal of effort has been made to develop objective image and video quality assessment methods (mostly for FR quality assessment), which incorporate perceptual quality measures by considering human visual system (HVS) characteristics. Some of the developed models are commercially available. However, image and video quality is far from being a mature research topic. In fact, only limited success has been reported from evaluations of sophisticated HVS-based FR quality assessment models under strict testing conditions and a broad range of distortion and image types. In the following sub-section certain of these HVS based assessment models will be presented.

3.1.2 Model based on characteristics of the human visual system

HVS-based models that incorporate the known visual phenomena, has shown promising results in the field of image and video quality assessment. Over the years many models have been proposed [34, 35]. HVS model combines and uses both the objective and subjective methods. The HVS-based models aim to simulate the processes of the human visual system from the eye to the visual cortex. These properties can be summarized as: luminance to contrast conversion (Weber's law), channel decomposition, frequency contrast sensitivity (Contrast Sensitivity Function) CSF, masking, and summation [34]. Most HVS-based models simulate the above processes, and try to sequence them in the way they occur in the HVS. A majority of these models requires a test image and its corresponding matching reference in order to determine the perceptual difference between them. Perceptual metrics are used in mechanistic modeling that incorporates HVS characteristics [36] such as luminance contrast sensitivity (Weber's law), frequency contrast sensitivity (Contrast Sensitivity Function) and masking effects [18, 37]. The Just-Noticeable-Difference (JND) is a very important concept in objective metrics using HVS features and ideally it provides each signal being represented with a threshold level of error visibility, below which errors are imperceptible. The JND is adopted in the Sarnoff Visual Discrimination Model proposed by Lubin [38].
HVS models can be divided into two groups. The first group comprises one-channel models [34, 39] that can be characterized by computing with the whole image. In the second one there are multi channel models [34, 35, 39] that simulate the neuron response of the brain cortex. The response is selective to spatial frequencies and orientations. These models decompose the image into the spatial frequency bands and/or orientations. Then, separate thresholds are set for each channel. At the end of the processing the channel are weighted and summed in order to get a number that represents the image quality.
In the last years, the development of novel video coding technologies has spurred the interest in developing digital video communications. The definition of evaluation mechanisms to assess the quality of video will play a major role in the overall design of video communication systems. A survey of video quality metrics based on HVS models can be found in [40]. A representative example of video quality metric based in HVS models is the Moving Picture Quality Metric (MPQM) developed by Van den Branden Lambrecht et al [41].
One important factor affecting the feasibility of HVS based video quality metrics is its computational complexity. While complex quality assessment methods may model the HVS more accurately, their computational complexity may be prohibitively large for many platforms, especially for real-time quality assessment of high-resolution video.

3.2 No reference metrics

Interestingly, human observers can easily assess the quality of distorted images without using any reference image. Consequently, most proposed no-reference quality metrics are designed for one or a set of predefined specific distortion types and are not generalized for evaluating images having other types of distortions. In image compression for example, NR quality metrics measure artifacts such as blockiness, blur, ringing, etc. [42, 43]. But,

the NR metrics has been predominantly used for quantifying the effects of block impairment artifacts [44-45]. This is because block impairment artifacts tend to be perceptually the most significant of all coding artifacts [44].
In the context of compressed video, *Venkatesh Babu* and al. [48] proposed two novel NR metrics. One metric to measure block edge impairments (or blockiness) and the other to measure the effectiveness of concealment strategies that try to mitigate the effects of packet loss on the overall video frame quality. The blockiness metric is based on measuring the activity around the block edges and on counting the number of blocks that might contribute to the overall perception of blockiness in the video frame while the effect of packet loss is measured by exploiting the structural pattern of this artifact. These metrics could be used for monitoring the quality of streaming video as well as being part of a feedback mechanism to assist in adapting the delivery of streaming video to varying network conditions. Yuukou Horita et al., [49] proposed a no-reference quality assessment model for JPEG/JPEG2000 coded image. This metric is defined in the spatial domain and based on the measurement of the blockiness for JPEG coded image, and the blur measure for JPEG2000 coded image. This assessment model is based on the blockiness around the block boundary, the average absolute difference between adjacent pixels within block, and the zero-crossing rate within block. The discrimination of JPEG coded image and JPEG2000 coded image is performed using the information of the blockiness and the average absolute difference between adjacent pixels. For image quality assessment of JPEG2000 coding, the blur measure is introduced instead of the blockiness of JPEG coding.
In today's world of multimedia communication over lossy networks like best-effort IP networks, it is crucial to be able to monitor the effects of compression and transmission related distortions in order to quantify the users Quality of Experience (QoE). The Venkatesh Babu et al. [50] present no-reference metrics to evaluate the following two major distortions in streaming of compressed video over packet-switched networks witch are the block-edge impairment and the effect of packet loss.
Objective No-Reference (NR) quality metrics is a very difficult task. This is mainly due to the limited understanding of the HVS, and it is believed that effective NR quality assessment is feasible only when the prior knowledge about the image distortion types is available. Although only several methods have been proposed for objective NR quality assessment [51-52], this topic has attracted a great deal of attention recently. For example, the video quality experts group (VQEG) [53] considers the standardization of NR video quality assessment methods as one of its future working directions.

3.3 Reduced-reference (RR) metrics

Reduced-reference (RR) image quality measures aim to predict the visual quality of distorted images with only partial information about the reference images. RR image quality metrics provide a solution that lies between FR and NR models. Recently, the video quality experts group has included RR image/video quality assessment as one of its directions for future development. RR methods are useful in a number of applications. For example, in real-time visual communication systems, these metrics can be used to track image quality degradations and control the streaming resources.
User-oriented image quality assessment has become a key factor in multimedia communications. However, existing image quality metrics such as peak signal-to-noise ratio (PSNR) are inappropriate for in-service quality monitoring, since they require the original image to be available at the receiver. In [54] is proposed a novel reduced-reference objective hybrid image quality metric (HIQM) that accounts for human visual perception and does not require a reference image at the receiver. This approach is based on a combination of various image artifact measures. The result is a single number, which represents an overall image quality. Experimental results indicate that HIQM outperforms PSNR.
In the domain of compression, [55] evaluates the quality of JPEG and JPEG2000 coded images whereas [56] provides assessment for JPEG and JPEG2000 compressed images, images distorted by white Gaussian noise, Gaussian blur, and the transmission errors in JPEG2000 bit streams. Reduced-Reference Image Quality Assessment (RRIQA) [57] the only RR-metric under study which is based on a Natural Image Statistic model in the wavelet transform domain and use the Kullback-Leibler distance between the marginal probability distributions of wavelet coefficients of the reference and distorted images as a measure of image distortion.
Wang [58] propose a new RR quality assessment method based on statistics computed for natural images in the wavelet transform domain.
Andrzej Glowacz et al. [59] proposed to evaluate an image transmitted especially with Fax over IP service. The "Fax over IP" service allows transmitting facsimile image over an IP network (like the Internet). But it can be potentially influenced by many mechanical and electronic distortions. He presents a novel approach to objective multimodal images comparing. The idea is to determine the visual image quality measure by combining partial measures from component algorithms operating on specified image features. He applies four such algorithms to successfully analyze image contrast, sharpness, granularity, and noise in relation to the original image.

In video transmission network, reference videos require tremendous amounts of storage space, and in many cases, are impossible to provide for most applications. Reduced-reference (RR) quality assessment does not assume the complete availability of the reference signal, only that of partial reference information that is available through an ancillary data channel. Perhaps the earliest RR quality assessment metric was proposed by Webster et al. [60] and is based on extracting localized spatial and temporal activity features. The work was extended in [61], where different edge enhancement filters are used, and two activity features are extracted from 3D windows. A successful RR quality assessment method must achieve a good balance between the data rate of RR features and the accuracy of image quality prediction.

4 CONCLUSION

Image and video quality assessment is still an active and evolving research topic. In this paper, many methods of image or video quality evaluation in different application domains were reviewed. The most reliable way of assessing the quality of an image or video is subjective evaluation, because human beings are the ultimate receivers in most applications. The mean opinion score (MOS), which is a subjective quality measurement obtained from a number of human observers, has been regarded for many years as the most reliable form of quality measurement. However, the MOS method is too inconvenient, slow and expensive for most applications.

The goal of objective image and video quality assessment research is to design quality metrics that can predict perceived image and video quality automatically. The performance of objective image quality assessment is to find an automatic metric that provides computed quality scores well-correlated with the ones given by human observers.

Because of the shortcomings of both, the subjective and the objective quality metrics mentioned above, many researchers have tried to overcome these shortcomings by developing a HVS-based quality metrics. Over the years many of these models were developed, because of the abundance of the HVS-based quality metrics in the literature. Even though all these models try to improve the accuracy of image quality prediction, no general-purpose metric has been agreed upon, to replace the subjective or the objective quality metrics. Hence there is still much left to be desired, and leave the door open to try and develop a new model that can improve the prediction of image quality measure accuracy.

REFERENCES

1. D.Semani, T.Bouwmans, C.Frélicot, and P. Courtellemont, «Automatic Fish Recognition in Interactive Live Video», International Workshop on IVRCIA, The 6th World Multi-Conference on SYSTEMICS, CYBERNETICS AND INFORMATICS, SCI 2002, Orlando, Florida (USA), July 14-18, 2002.
2. D.Semani, C. Saint-Jean, C. Frélicot, T. Bouwmans, and P. Courtellemont, «Alive fishes species characterization from video sequences», SSPR2002 & SPR2002 9th International workshop on Structural and Syntactic Pattern Recognition, 4th International workshop on Statistical Techniques in Pattern Recognition, Windsor, Canada, August 6-9 2002.
3. M. Chambah, B. Besserer, P. Courtellemont, "Recent Progress in Automatic Digital Restoration of Color Motion Pictures", SPIE Electronic Imaging 2002, San Jose, CA, USA, janvier 2002, vol. 4663, pp. 98-109.
4. A. Rizzi, C. Gatta, D. Marini, "From Retinex to ACE: Issues in Digital Images Unsupervised Color Equalization", *Journal of Electronic Imaging*, Vol. 13, No. 1, January 2004.
5. M. Chambah, A. Rizzi, C. Gatta, B. Besserer, D. Marini, "Perceptual approach for unsupervised digital color restoration of cinematographic archives", Electronic Imaging 2003, 21-24/01/03, S. Clara, California (USA).
6. M. Chambah, B. Besserer, P. Courtellemont, "Approach to Automate Digital Restoration of Faded Color Films", IS&T Conference on Graphics, Imaging, and Vision CGIV 2002, Poitiers, France, avril 2002, pp. 613-618.
7. D. Semani, C. Frélicot, and P. Courtellemont, "Feature selection using an ambiguity measure based on fuzzy OR-2 operators", Systems Association, IFSA'2003 - Istanbul - 2003.
8. Ismail Avcibas, Statistical evaluation of image quality measures Journal of electronic Imaging , Vol. 11(2) p 206,223

9. A. C. Bovik and S. Liu, “DCT-domain blind measurement of blocking artifacts in DCT-coded images,” Proceedings of International Conference on Acoustics, Speech, and Signal Processing, Salt Lake City, UT, May 7-11, 2001.
10. Z. Wang, A. C. Bovik and B. L. Evans, “Blind measurement of blocking artifacts in images,” Proceedings of IEEE 2000 International Conferencing on Image Processing, Vancouver, BC, Canada, September 10-13, 2000.
11. Z. Wang, H. R. Sheikh and A. C. Bovik, “No-reference perceptual quality assessment of JPEG compressed images,” Proceedings of IEEE 2002 International Conferencing on Image Processing, Rochester, NY, September 22-25, 2002.
12. L. Meesters and J.-B. Martens, “A single-ended blockiness measure for JPEG-coded images,” Signal Processing, Vol. 82, pp. 369-387, 2002.
13. ITU-R Rec. BT. 500:"Methods for the Subjective Assessment of the Quality of Television Pictures".
14. ITU-R Rec.BT.710:"Subjective Assessment Methods for Image Quality in HDTV".
15. ITU-R Rec. BT.811-1:"Subjective Assessment of Enhanced PAL and SECAM Systems".
16. Rec. ITU-R BT.500-11. Methodology for the subjective assessment of the quality of television pictures. ITU-R, 1974-2002.
17. Javůrek, R.: Different Approaches in Video Quality Assessment. In Proceedings of the 5th International Phd Workshop on Systems and Control. Budapest: Computer and Automation Research Institute, 2004, s. 60-64. ISBN 963-311-359-8.
18. W. Osberger (1999). Perceptual Vision Models for Picture Qaulity Assessment and Compression Applications. Ph.D. thesis, Queensland, University of Technology, Brisbane, Australia.
19. Z. Wang, A. Bovik, H. R. Sheikh, E. P. Simoncelli (2004). Image quality assessment: from error visibility to structural similarity. Image Processing 13(4):600–612.
20. R. Hamberg, H. de Ridder (1999). Time-varying image quality: Modeling the relation between instantaneous and overall quality. SMPTE Journal 108: 802–811.
21. Y. Inazumi, Y. Horita, K. Kotani, T. Murai (1999). Quality evaluation method considering time transition of coded quality. In International Conference on Image Processing, vol. 4, pp. 338–342.
22. Y. K. Lai and C.-C. J. Kuo, "A Haar wavelet approach to compressed image quality measurement," Journal of Visual Communication and Image Understanding, vol. 11, pp. 17-40, Mar. 2000.
23. Z. Wang and A. C. Bovik, “A universal image quality index,” IEEE Signal Processing Letters, vol. 9, no. 3, pp. 81-84, March 2002.
24. Z. Wang, A. C. Bovik and L. Lu, “Why is image quality assessment so difficult?” Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Proc., vol. 4, pp. 3313-3316, May 2002.
25. M.A. Ben Ayed, A. Samet, M. Loulou, and N. Masmoudi. New Perceptual Image Quality Assessment Metric Asian Journal of Information Technology 4(11): 996-1000, 2005.
26. A.B. Watson, J. Hu, and J.F. McGowan, DVQ: A digital video quality metric based on human vision, J. Electronic Imaging 10, 20-29, 2001.
27. B. Girod, “What’s wrong with mean-squared error,” Digital Images and Human Vision, A. B. Watson, ed., pp. 207-220, the MIT press, 1993.
28. S. A. Karunasekera and N. G. Kingsbury, “A distortion measure for blocking artifacts in images based on human visual sensitivity,” IEEE Trans. Image Processing, vol. 4, no. 6, pp. 71 3-724, June 1995.
29. M. Miyahara, “Quality assessments for visual service,” IEEE Communications Magazine, pp. 51-60, Oct. 1988.
30. Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004.
31. A. M. Rohaly, and al. Video quality experts group: Current results and future directions. vcip, 3:742-753, 2000.
32. M.H. Pinson, S. Wolf “A new standardized method for objectively measuring video quality” IEEE Transactions on broadcasting, Vol. 50, No. 3., pp. 312-322, Sep. 2004.
33. M. Chambah, A. Rizzi, C. Saint Jean, “Image quality and automatic color equalization”, SPIE/IS&T Electronic Imaging 2007, San Jose, California (USA).
34. Ch. J. van den B. Lambrecht, “Vision Models and Applications to Image and Video Processing”, Kluwer Academic Publisher, Netherlands 2001.
35. W. Osberger, Perceptual Vision Models for Picture Quality Assessment and Compression Applications ,Queensland Univ. of Technology, Brisbane, Australia, 1999.

36. Z. Wang . Objective Image/Video Quality Measurement - A Literature Survey. Tech. Rep. EE381K, Department of Electrical and Computer Engineering, University of Texas at Austin, USA, Department of Computer Science 9, University of Erlangen, 2003.
37. M. Sonka, V. Hlavic, R. Boyle. Image Processing, Analysis and Machine Vision. An International Thomson Publishing Company, 1999.
38. J. Lubin. A visual discrimination model for imaging system design and evaluation. In E. Peli (ed.), Vision Models for Target Detection and Recognition, pp. 245–283, World Scientific, New Jersey, 1995.
39. V. Slaninka, Objektivní hodnocení kvality obrazu, Diploma thesis, Prague, Czech technical University 2002.
40. Z. Wang, H. R. Sheikh, and A. C. Bovik, "Objective video quality assessment". The Handbook of Video Databases: Design and Applications (B.Furht and O. Marques, eds.), CRC Press, 2003.
41. C. J. van den Branden Lambrecht, D. M. Costantini, G. L. Sicuranza, and M. Kunt, "Quality assessment of motion rendition in video coding". IEEE Transactions Circuits and Systems for Video Technology, vol. 9, pp. 766-782. Aug. 1999.
42. P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi. Perceptual blur and ringing metrics: application to jpeg2000. Signal Processing: Image Communication, 19:163–172, 2004.
43. H.R. Sheikh, Z. Wang, L. Cormack, and A.C. Bovik. Blind quality assessment for JPEG2000 compressed images. IEEE Asilomar Conference on Signals Systems and Computers, Pacific Grove, California, Novembre 2002.
44. H. R.Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," IEEE Signal Processing Letters, vol. 70, no. 3, pp. 247–278, November 1998.
45. S. Winkler and A. Sharma and D. McNally, "Perceptual video quality and blockiness metrics for multimedia streaming applications," in Proc. 4th International Symposium on Wireless Personal Multimedia Communications, Aalborg, Denmark, pp. 553–556, September 2001.
46. Z. Wang and H. R. Sheikh and A. C. Bovik, "No reference perceptual quality assessment of JPEG compressed images," in Proc. ICIP'02, vol. 1, pp. 477–480, September 2002.
47. W. Gao, C. Mermer, and Y. Kim, "A de-blocking algorithm and a blockiness metric for highly compressed images," IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 12, pp. 1150–1159, Video Quality Experts Group (VQEG), December 2002.
48. R. Venkatesh Babu, Ajit Bopardikar, Andrew Perkis and Odd Inge Hillestad, "No-Reference Metrics for Video Streaming Applications," in the Proceedings of PV 2004, December 13-14, Irvine, CA, USA, 2004.
49. Yuukou Horita, Shinji Arata, and Tadakuni Murai, "No-reference image quality assessment for jpeg/jpeg2000 coding "
50. R. Venkatesh Babu, Ajit S. Bopardikar, Andrew Perkis, and Odd Inge Hillestad, "No-reference metrics for video streaming applications", in Proc. of the 14th International Packet Video Workshop (PV2004), December 13-14, 2004, University of California Irvine, USA.
51. H. R. Wu and M. Yuen, "generalized block-edge impairment metric for video coding," IEEE Signal Processing Letters, vol. 4, pp. 317–320, Nov. 1997.
52. P. Gastaldo, S. Rovetta, and R. Zunino, "objective assessment of MPEG-video quality: a neural-network approach," Proc. IJCNN, vol. 2, pp. 1432–1437, 2001.
53. VQEG: The Video Quality Experts Group, http://www.vqeg.org.
54. Kusuma, T.M.; Zepernick, H.-J. A reduced-reference perceptual quality metric for in-service image quality assessment. Mobile Future and Symposium on Trends in Communications, 2003. SympoTIC apos;03. Joint First Workshop on Volume , Issue , 26-28 Oct. 2003 Page(s): 71 – 74.
55. M. Carnec, P. Le Callet and D. Barba, "An image quality assessment method based on perception of structural information," *2003 International Conference on Image Processing*, Barcelona, Spain, September 14-17, 2003.
56. Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," *Human Vision and Electronic Imaging X Conference*, San Jose, CA, January 17-20, 2005, Proceedings of SPIE, Vol. 5666.
57. Z. Wang, E.P. Simoncelli "Reduced-reference image quality assessment using a wavelet-domain natural image statistics model" Human Vision and Electronic Imaging X, Proc. SPIE, vol. 5666. 2005.
58. 21 Zhou Wang, Guixing Wu, Hamid Rahim Sheikh, Quality-Aware Images, IEEE Transactions On Image Processing, vol. 15, no. 6, June 2006.
59. Głowacz, M. Grega, P. Gwiazda, L. Janowski, M. Leszczuk, P. Romaniak: Objective and subjective evaluation methods for selected aspects of fax image quality, XI symposium AES, Białystok Poland, September 2006.
60. A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran, and S. Wolf, "An objective video quality assessment system based on human perception," Proc. SPIE, vol. 1913, pp. 15-26, 1993.
61. S. Wolf and M. H. Pinson, "Spatio-temporal distortion metrics for in-service quality monitoring of any digital video system," Proc. SPIE, vol. 3845, pp. 266-277, 1999.