J. Vis. Commun. Image R. 22 (2011) 297–312

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Perceptual visual quality metrics: A survey

Weisi Lin [a,*], C.-C. Jay Kuo [b]

[a] School of Computer Engineering, Nanyang Technological University, Singapore 639798, Singapore
[b] Ming Hsieh Department of Electrical Engineering and the Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089-2564, USA

ARTICLE INFO

Article history:
Received 17 November 2009
Accepted 21 January 2011
Available online 1 February 2011

Keywords:
Human visual system (HVS)
Vision-based model
Signal-driven model
Signal decomposition
Just-noticeable distortion
Visual attention
Common feature and artifact detection
Full reference
No reference
Reduced reference

ABSTRACT

Visual quality evaluation has numerous uses in practice, and also plays a central role in shaping many visual processing algorithms and systems, as well as their implementation, optimization and testing. In this paper, we give a systematic, comprehensive and up-to-date review of perceptual visual quality metrics (PVQMs) to predict picture quality according to human perception. Several frequently used computational modules (building blocks of PVQMs) are discussed. These include signal decomposition, just-noticeable distortion, visual attention, and common feature and artifact detection. Afterwards, different types of existing PVQMs are presented, and further discussion is given toward feature pooling, viewing condition, computer-generated signal and visual attention. Six often-used image metrics (namely SSIM, VSNR, IFC, VIF, MSVD and PSNR) are also compared with seven public image databases (totally 3832 test images). We highlight the most significant research work for each topic and provide the links to the extensive relevant literature.

1. Introduction

Quality evaluation for digital visual signals is one of the basic and challenging problems in the field of image and video processing as well as many practical situations, such as process evaluation, implementation, optimization (e.g., video encoding), testing and monitoring (e.g., in transmission and manufacturing sites). In addition, how to evaluate picture quality plays a central role in shaping most (if not all) visual processing algorithms and systems [50,114,124]. Examples of technological dependence upon visual quality evaluation include: signal acquisition, synthesis, enhancement, watermarking, compression, transmission, storage, retrieval, reconstruction, authentication, and presentation (e.g., display and printing).

Objective quality evaluation for images and video can be classified into two board types: signal fidelity measures, and perceptual visual quality metrics (PVQMs).

The signal fidelity measures refer to the traditional MAE (mean absolute error), MSE (mean square error), SNR (signal-to-noise ratio), PSNR (peak SNR), or one of their relatives [41]. Although they are simple, well defined, with clear physical meanings and widely accepted, they can be a poor predictor of perceived visual quality, especially when the noise is not additive [71,84]. Some metrics have been used to estimate delivered picture quality after transmission based on network parameters [108,138,183], such as throughput, jitter, delay, bit error and packet loss rates. However, the same network parameters may result in different degradation of visual content, and therefore different perceived quality. Quality determined by consumers' perception and satisfaction is much more complex than the statistics that a typical network management system can provide. It has been well acknowledged that a signal fidelity measure does not align well with human visual perception of natural images or computer generated graphics [41,52,97,149,161].

Since the human visual system (HVS) is the ultimate receiver and appreciator for the majority of processed images, video and graphics, it would be more logical, economical and user-oriented to develop a perceptual quality metric in system design and optimization. Naturally, perceptual visual quality (or distortion) can be evaluated by subjective viewing tests with appropriate standard procedures [65]. This is however time consuming, laborious and expensive, since the resultant mean opinion score (MOS) needs to be obtained by many observers through repeated viewing sessions. Moreover, incorporation of subjective viewing tests is not feasible for on-line visual signal manipulations (such as encoding, transmission, relaying, etc.). Even in situations where human examiners are allowed (e.g., visual inspection in a factory environment) and the manpower cost is not a problem, the assessment results still depend upon viewers' physical conditions, emotional states, personal experience, and the context of preceding display. Hence, it is necessary to build computational models to predict the evaluation of an average observer. In other words, objective means are sought to approximate human perception results (e.g.,

* Corresponding author.
E-mail addresses: wslin@ntu.edu.sg (W. Lin), cckuo@sipi.usc.edu (C.-C. Jay Kuo).

1047-3203/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jvcir.2011.01.005

MOS, when the number of subjects is sufficiently large). In comparison with the subjective viewing tests, objective metrics are advantageous in repeatability due to the nature of objective measurement.

Although physical variations in terms of MSE, SNR, PSNR, etc. reflect picture quality change, these traditional signal fidelity metrics fail to predict the HVS perception because of the following problems:

(1) Not every change in an image is noticeable;
(2) Not every pixel/region in an image receives the same attention level;
(3) Not every change leads to distortion (otherwise, many edge sharpening and post-processing algorithms would have not been developed);
(4) Not every change yields a same extent of perceptual effect with a same magnitude of change (due to spatial/temporal/chrominance masking).

A significant amount of research efforts has been made toward HVS-based picture quality evaluation during the past decade [26,27,51,70,106,118,123,156,157,160,183,188] so as to tackle the abovementioned four problems of traditional measures.

2. The problem

2.1. Nature of the problem

Visual quality assessment can be of the first party (the photographer or image maker), the second party (the subject of an image) and the third party (neither the photographer nor the subject) [72]. The concern in this survey is the perception of third-party observers, since this represents the most general and meaningful situation in modeling and applications.

PVQMs refer to the objective models for predicting subjective visual quality scores (i.e., the MOS). In this paper, we will focus on surveying the PVQMs developed so far that carry out direct evaluation of the actual picture under consideration, rather than some predefined signal patterns that go through the same processing [66]. This is because picture quality is a function of visual contents, so the change of predefined test signals through a system is not necessarily a reliable source of visual quality measurement for actual signals; and in addition, the predefined visual signal adds to the overheads of transmission/storage.

In spite of the recent progress in related fields, objective evaluation of picture quality in line with human perception is still a long and difficult odyssey [38,123,156,157,163,183] due to the complex, multi-disciplinary nature of the problem (related to physiology, psychology, vision research and computer science), the limited understanding of the HVS mechanism, and the diversified scope of applications and requirements.

Despite the difficulties, perceptual visual quality evaluation should be less demanding than computer vision in general, since it can be performed without the need of emulating "*the process of discovering from images what is present in the world, and where it is*" (Marr's words on vision [99]), in most meaningful and practical situations for visual quality evaluation. With proper modeling of major underlying physiological and psychological phenomena, it is possible to develop better visual quality metrics to replace non-perceptual criteria widely used nowadays, in various specific practical situations.

2.2. Organization of this paper

Due to the vast scope of this survey, we divide the main body of the survey that follows into two parts for clearer presentation: in Section 3 below, a review will be given on basic computational modules in building various PVQMs; in Section 4, two major categories of PVQMs will be then discussed. The further rationale for such a 2-step organization strategy is as follows.

The basic computational modules include signal decomposition (decomposing an image or video into different color, spatial and temporal channels), detection of common features (like contrast and motion) and artifacts (like blockiness and blurring), just-noticeable distortion (JND) (i.e., the maximum change in visual content that cannot be detected by the majority of viewers), and visual attention (VA) (i.e., the HVS's selectivity to respond to the most attractive activities in the visual field). First, many of these are based upon the related physiological and psychological knowledge. Second, most of them are independent research topics themselves, like JND and VA modelling, and have other applications (image/video coding [10,194], watermarking [187], error resilience [48], computer graphics [136], just to name a few) in addition to PVQMs. Third, these modules can be simple PVQMs themselves in specific situations (e.g., blockness and burring). After the discussion of these basic building modules, we will be able to focus on system-level issues related to the major PVQMs in Section 4.

In Section 5, we will compare six existing image quality metrics (SSIM [167], VSNR [17], IFC [139], VIF [140], MSVD [42], and PSNR) against the subjective viewing data, from seven publicly available databases. These databases are with a wide variety of visual contents and distortion types to enable a meaningful and convincing benchmarking.

Before going to the main body of this paper, let us briefly explain several psychophysical phenomena that have been commonly used in PVQM development. The contrast sensitivity function (CSF) denotes the HVS's sensitivity toward signal contrast with spatial frequencies and temporal motion velocities [73,155], and exhibits a parabola-like curve with the increase of spatial and temporal frequencies, respectively. Luminance adaptation refers to the noticeable luminance contrast as a function of background luminance; for digital images, luminance adaptation takes a parabola-like curve [23,67]. Visual masking is usually the increase of the HVS's contrast threshold for a signal in the presence of another one; it can be divided into intra-channel masking [7] by the signal itself, and inter-channel masking [13,82] by signals with different frequencies and orientations.

For the convenience of the reader, the major abbreviations and notations used in this paper are listed in Tables 1 and 2.

3. Basic computational modules

There have been basically two categories of PVQMs [183]: the vision-based modeling and signal-driven approach. For the first category [30,93,174,178], PVQMs are developed based upon systematical modeling of relevant psychophysical properties and physiological knowledge, including temporal/spatial/color decomposition, CSF, luminance adaptation, and various masking effects. The second category attempts to tackle the problem from the viewpoint of signal extraction and analysis, such as statistical features [185], structural similarity [162], luminance/color distortion [107], and the common visual artifacts (e.g., blockiness and blurring) [100,189]. These metrics look at how pronounced the related features are in image/video to estimate overall quality. This does not necessarily mean that such metrics disregard human vision knowledge, as they often consider psychophysical effects as well (e.g., a JND model), but image content and distortion analysis rather than fundamental vision modeling is the basis for design.

There are metrics making use of both classes. For example, a scheme was proposed in [148] to switch between a model-based scheme and a signal-driven one according to the extent of blocki-