Image Quality Assessment and Human Visual System

Xinbo Gao a, Senior Member, IEEE, Wen Lu a, Dacheng Tao b, Member, IEEE, Xuelong Li c, Senior Member, IEEE

a School of Electronic Engineering, Xidian University, Xi’an 710071, China
b School of Computer Engineering, Nanyang Technological University, 50 Nanyang Avenue, Blk N4, 639798, Singapore.
c State Key Laboratory of Transient Optics and Photonics, Xi’an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi’an 710119, Shaanxi, China.

ABSTRACT

This paper summaries the state-of-the-art of image quality assessment (IQA) and human visual system (HVS). IQA provides an objective index or real value to measure the quality of the specified image. Since human beings are the ultimate receivers of visual information in practical applications, the most reliable IQA is to build a computational model to mimic the HVS. According to the properties and cognitive mechanism of the HVS, the available HVS-based IQA methods can be divided into two categories, i.e., bionics methods and engineering methods. This paper briefly introduces the basic theories and development histories of the above two kinds of HVS-based IQA methods. Finally, some promising research issues are pointed out in the end of the paper.

Keywords: Image quality assessment, human visual system, visual physiology, visual psychophysics, bionics methods, engineering methods.

1. INTRODUCTION

With the rapid development of information technology, digital image as an important media for representing and communicating has witnessed a tremendous growth. A huge number of processing methods [1] have been proposed to treat image for different purposes. The performance of these methods [2] is usually evaluated by measuring the quality of the output image of the processing system. For instance, in the process of image compression, lossless compression techniques would introduce artificial block, blurring, and ringing effects, which lead to image degradation. In poor transmission channels, transmission errors or data dropping would happen, which lead to the imperfect quality and distortion of the received video data. Therefore, how to evaluate the image quality [3][4] has become a burning problem. The recent years has demonstrated and witnessed the tremendous and imminent demands for image quality assessment (IQA) [5][6] methods at least in the following three ways: 1) They can be exploited to monitor image quality for controlling quality of processing system [7]; 2) They can be employed to benchmark image processing systems and algorithms [8]; 3) They can also be embedded into image processing systems to optimize algorithms and parameter settings. Existing IQA metrics [9] can be categorized into subjective and objective methods. The former is based on the quality which is assessed by human observers, and the latter provides an objective index or real value which is obtained from an assessment model to measure the image quality.

Because human observers are the ultimate receivers of the visual information contained in an image, subjective method [10] whose results are directly given by human observers is probably a reliable way to assess the quality of an image. The subjective method is that the observers are asked to evaluate the picture quality of sequences using a continuous grading scale and to give one score for each sequence. A number of different subjective methods are represented by ITU-R Recommendation BT.500 [10]. The subjective quality measurement has been used for many years as the most reliable form of quality measurement [11]. However, subjective experiment requires human viewers working in a long period, and repeated experiments are needed for many image objects, it is expensive and time consuming, and cannot be easily and routinely performed for many scenarios, e.g., real time systems. Moreover, there has not been any precise mathematical model for subjective assessment currently. As a consequence, there arises the requirement of an objective quality metric that accurately matches the subjective quality and can be easily implemented in various image systems, leading to the emergence of the objective IQA.
The objective IQA [5][6][9] is proposed to provide a computational model to measure the perceptual quality of an image. It makes use of the variation of several original or distorted image characteristics which is caused by degradation to represent the variation of the image perceptual quality. Many objective quality metrics for predicting image distortions have been investigated. Traditional, the Peak Signal to Noise Ratio (PSNR) [12][13] has been a popular and commonly used metric to evaluate and quantify performance of the image processing algorithms. But it exhibits weak performance that has not been in agreement with perceived quality assessment based on pixel-wise error in image mathematically. Recently, a great deal of efforts has been made to the development of visual models that take advantage of the known characteristics of human visual system (HVS) [14]. The aim of the HVS-based IQA is to evaluate how strong the distorted information is perceived by HVS, according to the characteristics and cognitive mechanism of the HVS. A number of the IQA metrics based on HVS has been proposed to evaluate the perceptual quality. In the following section, the basic properties of human visual system are summarized first, and then the development of HVS-based IQA methods are reviewed and discussed.

2. HUMAN VISUAL SYSTEM

Human visual system [15][16] is a very complex system and not fully understood yet. Moreover, most visual properties of the HVS are not intuitive. So HVS is a crucial tool by which human beings understand and cognize nature world, and a breakthrough point that reveals the secrets of the brain. A large number of physiological [17] and psychophysical [18] experiments show physiological evidence and are the only way to understand the phenomenon. Nowadays, there is some research findings obtained in visual physiology and psychophysics.

2.1 Visual Physiology

The research of visual physiology has shown that visual information processing involves the following four aspects [19]: optical processing [20], retina processing [21], lateral geniculate nucleus (LGN) processing [22] and visual cortex processing [23]. The optical processing [20] is completed by the eyes, and its main function is equal to a "camera". Correspondingly, sclera is similar to a spherical camera bellows, and cornea play an important role in concentrating light. The iris is regard as an aperture, and controls the size of a pupil, which is responsible for regulating the brightness on the retina. The lens is just like a lens, and retina as a film. Finally light focuses on the retina [21] and forms a clear object-image. Research shows that retina is mainly composed of three kinds of neurons, including photoreceptor cells, bipolar cells and ganglion cells, which is responsible for the photoelectric conversion and information transmission. Through the retina, optical signal is encoded into potential pulse and is passed to the LGN in the form of frequency modulation. As a transfer station for the signal from the retina to the visual cortex, LGN plays the role of threshold to control the amount of information. Ultimately, the visual cortex [17] realizes the process of object recognition, perception and understanding. Figure 1 shows the basic modules and process of the human visual system.

![Figure 1. The process of the human visual system](image)

2.2 Visual Psychophysics

Increasing physiological results [22] have greatly promoted the development of visual psychophysics [18]. Visual psychophysical method is an approach to use mathematical models and measurement techniques to study the relationship among visual psychological phenomena. Because of the great complexity of human visual system, many visual psychophysics phenomena are being in the hypothesis stage. Some experiments only have showed a number of low-level visual psychological characteristics, including luminance non-linearity [24], contrast sensitivity [25], masking effects [26], multi-channel parallel [27] and visual attention [28][29] and so on. **Luminance non-linearity** [30] refers to that human eyes have poor judgment to the absolute brightness of the observed objects, while strong to the relative differences in brightness discretion, i.e., the value of the contrast sensitivity is high. Specifically, the logarithm of the perceived brightness and the objective brightness are linear in a certain range. **Contrast sensitivity** describes frequency response characteristics of the human visual system, which reflects the ability of the human eyes to distinguish the
differences in intensity. Campbell and Robson [27] conducted a series of experiments by adjusting the amplitude and frequency of sinusoidal grating, given the well-known Campbell-Robson contrast sensitivity function map, and proposed that contrast sensitive function (CSF) can be seen as a band-pass filtering process. Masking effect [31] [32] is a general concept referring to the reduction of visibility of one image component due to the presence of another masker, which describes the phenomena that when there is a stimulus A, the perception of stimulus B would be to strengthen or weaken. Some studies indicate that vision is a multi-channel parallel [33], that is, the different visual information components are pre-processed through different neural channels, then as the input of the visual cortex. Next, they will be analyzed and processed by the different types of cortical cells. i.e., there are a series of different and independent channels which have selection to spatial frequency in human visual system memory. For example, in the primary visual cortex, most of the neurons are sensitive to the visual stimuli with specific spectrum location, frequency and orientation. Visual attention [29] is an appearance that we often pay attention to one or some scene due to the demanding for behavior purposes or local scene clues as the surrounding environment observed, so that some certain spots or area are selected as the representation of the scenery. James [34] proposed two strategies for developing the field of the "Principles of Psychology". The first one is bottom-up saliency to be influenced by the environment, such as Itti [35], GBVS [36], Gaffe [37], they are all based on the theory of the feature integration, with emphasis on the stimulation of the low-level features of the target. The second one is the top-down saliency, which depends on the visual processing tasks. For example, Searchlight [38] and Tsotsos [39] emphasized that the "attention" process can enhance the impact of the target features and affect the observer's experience and knowledge. Some visual psychophysics phenomena are present in Figure 2.

![Figure 2. The graphical representation of some visual psychophysics phenomena](image_url)

3. HVS-BASED IMAGE QUALITY ASSESSMENT

Since human beings are the ultimate receivers of the image in most case, it is necessary for the IQA research to deeply investigate the human visual system (HVS). According to the properties and cognitive mechanism of the human visual system, HVS-based image quality assessment can be categorized into bionics methods and engineering methods. In the following, the above two HSV-based IQA methods are introduced respectively.

3.1 Bionics Methods

The bionics method is also called bottom-up method, which is meant a metric to construct the whole procedure by modeling some characteristics of human visual system respectively, obtained on physiology and psychophysics experiments to do image quality assessment. Such as multi-channel decomposition, contrast sensitivity function, masking function and just-noticeable difference (JND) function, etc. The multi-channel decomposition is used to decompose an input image signal into multiple channels using localized, band-pass, and oriented filters. There are many multi-channel decomposition methods such as Cortex transform [40], multiscale geometric transform [41][42] and so on. Contrast
sensitivity function [43] is equal to a band-pass filter to describe a phenomenon that sensitivity varies with the spatial frequencies in HVS-modeling. Some CSF measurements are described in [44][45][46]. Masking function explains a fact when a stimulus that is visible by itself cannot be detected due to the presence of another, and how to disturb distortion is in certain regions of an image while they are hardly noticeable elsewhere. Several different types of masking have been described in [31][32]. JND function represents an appearance that there is a minimum visibility threshold [47] when visual contents are altered in the human visual system. [48] introduced JND in detail, and enriched and developed the JND. The process of the bionics methods is shown in Figure 3.

![Figure 3. The block-diagram of the process of bionics method](image)

Recently, a great deal of efforts has been made to the development of visual models that take advantage of the known characteristics of HVS. The famous model -- Daly visible differences predictor (VDP) is given by Daly in [49]. The Daly VDP uses a modified version of Watson’s cortex transform for the channel decomposition, receives distorted corresponding reference as input and produces difference map as output, which predicts the probability of detection for dissimilarities throughout the whole images. Another model that estimates the probability of detection of the difference between the reference image and distorted image is proposed by Lubin in [50]. The key step of this model is filtering the images using a low-pass PSF that simulates eye optics function. Safranek-Johnson et al. [51] used the generalized quadrature mirror filter transform for decomposing the image signals into sixteen subbands of the frequency space. At each subband, the coefficients are divided by corresponding overall normalization factor and pooled. Finally we obtain the quality of perceptual image coding. Teo-Heeger et al. introduced an image quality metric in [52], which adopted steerable pyramid decomposition with six orientations, whose parameters are optimized to fit psychophysical data. The watson metric [53] uses the DCT-based perceptual error measurement. The quantified errors for every coefficient in every block are scaled by the corresponding visual sensitivities of every DCT basis function in each block. Lu et al. presented a metric in [54] based on wavelet transform that took the variance of the visual sensitive coefficients into account to measure a distorted image. Watson et al. [55] proposed a method that can quantify noise by wavelet transform. This method was extended by Bradley [56]. Lukas and Budrikis [57] developed a spatial-temporal CSF model based on excitation channel and inhibition channel. Tong et al. [58] proposed a single-channel video quality evaluation method based on spatio-temporal CIELAB. Lambrecht and Verscheure proposed an evaluation method based on multi-channel visual properties [59] in 1996, which employ 16 Gabor filters for multi-channel decomposition. After the spatio-temporal CSF filtering and the inter-channel masking, the perceptual coefficients which take JND as unit are obtained. Then these perceptual coefficients are incorporated by Minkowski pooling, and the results are regarded as the overall video quality. In 1999, Winkler introduced a video quality assessment metric called perceptual distortion metric (PDM) in [60]. The first step is color space conversion, and then the temporal filtering which contains the transitory and sustained response is conducted. The spatial decomposition is conducted by the steerable pyramid transform. Afterwards, coefficients in each channel are weighted by contrast sensitivity function (CSF), and the final distortion measure can be obtained by integrating the difference between the two signals through Lp-norm summation. Gao and Lu et al. [61][62] proposed a novel framework for IQA to mimic the HVS by incorporating the merits from multiscale geometric analysis (MGA), CSF, and the Weber’s law of just noticeable difference (JND). Thorough empirical studies show that the novel framework with a suitable image decomposition method performs better than conventional standard reduced-reference image quality assessment method.

### 3.2 Engineering Methods

Engineering methods only take into account the relationship of the input and the output of the HVS that is modeled as a whole, i.e., it is equivalent to a “black box”. It is only concerned with the input images and the output value of the image quality regardless of its details. Figure 4 shows the diagram of the engineering method. The specific and application-oriented algorithms of the IQA are built according to the prior knowledge of distortion types which are one or some features of HVS. This kind of methods shows a lower computational complexity, better pertinence and performance, hence the method can be utilized in some particular applications that require higher real-time. Based on this, many
scholars have proposed a great amount of methods that can be divided into three classes, according to the types of feature: features of the original image, features in transform domain and distorted features of the image.

![Figure 4. The framework of engineering methods](image)

Features of the original image based methods: Miyahara et al. [63] proposed a picture quality scale (PQS), which exploited the luminance encoding error, spatial frequency error weighted, random error and others features to evaluate the image quality. Wekrn et al. [64] utilized the fuzzy theory to compare the similarity and consistency between two images proposed. Wang et al. [65] proposed a structural similarity (SSIM) image quality assessment corresponding to sensitivity to structural information in HVS. Gao et al. presented the content-based metric (CBM) in [66]. The method based on SSIM, to extract the structural information, then divides them to edges, textures and flat regions according to content. Hekstra et al. [67] gave a perceptual measurement that performs the evaluation by exploiting the contour feature of the luminance image, normalized color error and temporal variance. Wolf and Pinson’s video quality metric (VQM) [68] utilizes spatial distortion pooling and temporal distortion pooling. Pahalawatta et al. [69] proposed a motion estimated temporal consistency metrics for objective video quality assessment. Gunawan et al. [70] utilized the discriminative local harmonic strength, which considered the motion, to evaluate the video quality. Ninassi et al. [71] take into account temporal variations of spatial distortions for the proposed method. Zhai et al. proposed a cross-dimensional perceptual quality assessment for low bit-rate videos [72]. Hewage et al. introduced a quality evaluation of color plus depth map-based stereoscopic video in [73]. Barkowsky proposed a temporal trajectory aware video quality measure [74]. Winkler et al. reviewed the evolution of video quality measurement techniques and the state of the art methods [75].

Features in transform domain of the image based methods: Shnayderman et al. presented singular value decomposition (SVD) [76]. Chandler et al. proposed a visual signal-to-noise ratio (VSNR) [77]. This method is based on that the adaptability of visual luminance and angle to various visual conditions. Sheikh et al. introduced the visual information fidelity (VIF) in [78][79], which implemented image quality assessment in the point of view of information communication and sharing. The method not only models the input image statistically with Gaussian mixture model in wavelet domain and extracts the features, but also models the image distortion channel and human visual distortion channel, evaluate the images. Li et al. [80] proposed a metric based on a divisive normalization image representation. Rajashekar et al. described a framework for quantifying color image distortion based on adaptive signal decomposition in [81]. Watson gives a digital video quality assessment metric in [82], it exploit the model of the JND in DCT domain to simulate the HVS. Callet et al. [83] proposed a video quality metric based on neural network convolution. Baroncini and Pierotti proposed a video quality assessment model based on DCT encoding video [84]. The advantages of this model are to detect block artifacts by checking the error visibility caused by the temporal fluctuation of visual quality. Gunawan et al. presented a reduced reference video quality assessment based on harmonic analysis [85], through analyzing the visibility of harmonic amplitude, the score will be worked out by pooling the harmonic gain and lost. Li [86] et al. showed a natural image quality assessment in [86], which performed the task by means of hybrid wavelets and directional filter banks (HWD). Farias et al. introduced a water mark based quality metric [87]. The model estimates the video quality by checking the embedded watermarking sequent. Gastaldo presented a metric which is based on neural network in [86]. Tao et al. presented a no reference image quality metric in contourlet domain [89].

Distorted features of the image based methods: Damera-Venkata et al. [90] exploited an objective quality assessment method based on image degradation model. Engelke et al. [91] designed the normalized hybrid image quality metric, which account for different structural artifacts that have been observed in the distortion model of a wireless link. Angelis et al. [92] proposed a metric that consists in the evaluation of a vector error whose components include noise cancellation and detail preservation affecting the luminance and chroma. Tan et al. [93] proposed an objective MPEG-specific video quality assessment that commits itself to convert each frame distortion metric to continuous distortion metric. In 2000, Tan et al. accomplished the evaluation of blurring and block artifacts using the distorted MPEG video.
sequences [94]. Caviedes and Oberti [95] showed a model based on distorted video features. The model contains four layers of video features and estimates the resulting score by pooling these features. Sheikh et al. [96] proposed an no reference IQA metric that uses natural scene statistics (NSS) for JPEG2000 compressed images. In [97] the NSS model is improved by Lu et al., in this metric, the image is decomposed by contourlet transform, and the relationship of the contourlet coefficients is represented by the joint distribution. The final quality obtained from the variation of statistics law of the contourlet coefficients.

In this field of image quality assessment, it is crucial to deep research the physiology and psychology of human visual system. If the procedure of the human visual system could be simulated accurately, the precise evaluation would be designed accordingly. However, there are some limitations because of the complication of the HVS at present. Bionics method is one way to simulate human visual system. Comparing engineering method, it is more reasonable and more problematic consequently. Engineering method resolves some problems such as complexity and de-correlation of natural image to some extent, explains some high-level cognitive procedures as well, however, since the engineering method depends on mathematical models completely, the issues about modeling and parameter selection are more difficult than bionics method. Although the totally different designing philosophy between them, they are complementary to a larger extent, not contradictory. Hereby, there is not evident boundary between engineering and bionics method. In the future, there is an extremely significant means that integrates two kinds of methods to build a new metric for the IQA.

4. CONCLUSION AND FUTURE DIRECTIONS

This paper mainly reviews image quality assessment [98] based on human visual system. Since the ultimate aim of IQA is to employ a computational model to replace image quality scores observed by human beings. HVS is a vital position in image quality assessment. A metric based on HVS is a most significance way for studying IQA. By simulating the properties and cognitive mechanism of the HVS, IQA metrics can be divided into bionics methods and engineering methods. The paper briefly reviews two kinds of metrics respectively. Although these metrics have good consistency with subjective perception values and the objective assessment results can well reflect the visual quality of images, there are still some issues deserve to be further investigated in the future.

1) The problems of the human visual system: Since image is ultimately consumed by human beings, from the points of simulating the HVS, it is significant to build a computationally physiological and psychological model of the HVS. The HVS, however, is a very sophisticated system. The implementation has always given rise to a complex algorithm and high computational cost accordingly. In addition, it is limited to build an exact and unified model owning to the lack of a deep insight into human visual system. Furthermore, the assessment of image quality is to be done by lots of the high-level neural activities in human brain, e.g., the cognitive stage, it is far from adequate to model the early visual stages merely. Hence there is tremendous room for development of HVS-based image quality assessment.

2) The cross problems of two kinds of the metrics based on HVS: The assessment methods based on the HVS are divided into two major categories: bionics methods and engineering methods. Both of them have advantages and disadvantages respectively, but to a great extent they are complementary rather than contradictory. Because of no clear boundaries between them, bionics methods and engineering methods are desired to find the common ground, so that a comprehensive solution can be offered including these methods. The characteristics of the HVS are described more reasonable and more effective in order to obtain a much more outstanding image quality assessment.

3) The problems of the statistical features of the image: Due to the natural image with the certain higher level statistics law, the features of image can be modeled effectively. So, how to find an effective model to measure the image quality is the research highlight. Recently, some lower level statistics models of images in existence cannot describe the probability distribution of natural images adequately. The models which were only used to find the models to depict the general rule of image have serious limitations. Next, the accurate description of the higher order statistics feature of image is of great significance for the image quality assessment.

4) The problems of the visual attention mechanism: Despite of a large field of view, the human eyes process only a tiny central region on the retina with great detail, that is to say, we always concentrate on the region that is gazed. Normally when observing an image, we spend much more time paying attention to attentive points, and also obtain more detailed information. Therefore whether an image degrades on the attentive region greatly impacts the results of visual perception. Visual studies have shown that regions of large contrast, slender shape, distinct color, and with
the trend of movement can easily cause visual attention. Consequently, image quality evaluation should focuses on such factors that affect visual attention mechanism.

5) **The problems of the color model of the image:** So far, the existing image quality assessment is to be done by evaluating the pixel-wise differences, but the images existed in nature scene are of various colors. In human brains the perception to chromatic image is far more complicated than the gray-scale one due to the different processes that human beings perform. In order to accomplish the chromatic image quality assessment, the color space, colorimetry and other relevant aspects should be considered because of the likely emerging of distortion in different color gamut. Therefore, from the point of view of color models, how to build a model capable of quantifying the color degeneration and to evaluate perceived color image quality is worth deeply studying.

6) **The extensive applications of the image processing:** Image quality evaluation is widely used in many fields including image acquisition, compression, communication, printing, display, reconstruction, enhancement, noise reduction, segmentation, detection and medical, geographical, satellite, and astronomic image classification. Since quality evaluation is in close touch with practical application, it is sensible to start from concrete conditions to design evaluation methods which target at a certain application environment. When the combination of image quality evaluation and application is becoming more and more tight and widespread, amounts of professional problems will arise. The research on image quality assessment will be further promoted by the various applications.

**REFERENCES**

34. James, W., The principles of psychology, New York: H. Holt, (1890).