Blind Image Quality Assessment is Not Impossible

Plenary Talk

Alan C. Bovik
Laboratory for Image and Video Engineering (LIVE)
The University of Texas at Austin

Abstract

The problem of blindly assessing the quality of visual signals – without reference, and without assuming a single distortion type – has long been regarded as, if not impossible, perhaps too difficult to bother with. After all, it requires sorting out the black box we call the visual apparatus of the human brain, which despite an increased level of transparency over the past 40 years, remains poorly understood. Moreover, it involves the very un-engineering-like concept of subjectivity, an uncomfortable thing altogether for the analytic types populating our profession. And lastly, it requires dispensing with older ideas of quality such as fidelity, similarity, and metric comparison. In this talk I will discuss our recent efforts on blind or "no reference" image quality assessment problems, including machine learning approaches, and the looming question of stereo (3D) image quality.

Prelude

I recall many years ago, when I was a graduate student laboring over my thesis on nonlinear digital image filters, I first wondered about image quality. As I slowly generated results at 2AM on the PDP-11 which was our computing engine – back then, running a median filter on a small image meant smoking a few cigarettes, sipping a beer, and taking a short nap – I'd look at the results and the inevitable printout of mean-squared errors (MSEs) for each filtered result. Then I'd compare these numbers with the images, and sit there puzzled: the relative image quality delivered by the various filters, at least as I saw it, didn't always connect with the MSE scores.

Since I didn't know what to do about this, I naturally turned to the supreme source of knowledge on any image-related topic, my advisor, the estimable Thomas Huang. So I showed my work to Tom, and inquired of him how I might best prove that our filter's results were better than "those other ones." After all, they looked better! Tom just looked at me, smiled, and quietly said (in effect) "There isn't any way; just put images in the paper for them to look at." So I did. When the results appeared later in the *Transactions,* as just a few, tiny, grainy, washed-out images (the process has since improved somewhat), I was naturally disappointed that one could tell very little, if any, difference between the results, and I couldn't prove to anyone that our way was better.

This mode of thinking continued for many years, with hardly anyone in the field thinking seriously about image quality until the early 1990's, when people started to acknowledge the deficiencies of the MSE [1], and to explore perceptual aspects of image quality [2]. Before long, the promise of consumer digital video led to the creation of such groups as the Video Quality Experts Group (VQEG), and the race to solve the problem was on.

But this talk isn't (yet another) historical summary or survey of image or video quality indices. To make a long story short, I am going to hazard the opinion that the so-called Full Reference (FR) still image quality assessment (IQA) problem is *largely solved.* There exist algorithms, such as Multi-scale SSIM, VSNR, and VIF [3]-[6] that correlate quite highly with human subjective judgment, on more than one large public database [7]-[9]. While I hope that I am wrong, I feel that so many avenues have been explored that any future significant gains in performance (relative to human judgment) will come as a result of deeper models of visual content and their perception that go well beyond current models of "saliency" (studied in the context of IQA [10], [11]) and "foveation" [12]. Better models of mid- to high-level processes of visual tasking, visual search, and visual recognition may drive this direction. While the conspicuity of distortions and low-level image features is important, the visual task being undertaken, be it hand-washing, reading, or navigation, is certainly more so [13]. I do believe that there is more room for improvement in objective *video* quality assessment (VQA), owing to the possibilities afforded by better models of motion perception [14], but I also think that a similar plateau in performance may soon be reached; progress on FR QA will thereafter be slow, absent an epiphany connecting higher-level and lower-level percepts of visual relevance and visual quality.

Yet, all of this work only goes a small part of the way towards resolving my 30-year difficulty. How to simply "eyeball" an image with an algorithm, and proclaim the level of visual quality it has? The problem has long been thought of as essentially unsolvable (hence the title of this talk). Although I have done a lot of work on the FR problem (after all it is doable), the fundamental underlying question of "What is Image Quality" has continued to distress me. There exists, after all, no "pristine" reference (I have never seen one) as image acquisition is an imperfect process. Moreover, in proliferating consumer applications, such as internet and mobile video, reference images are generally unavailable. What is needed, then, is a general theory of Image Quality that departs from the notions of "fidelity" or "similarity" that currently prevail [15].

Attacking the Impossible Problem

My students and I have, in the past, dabbled in the so-called No Reference (NR) or blind IQA and VQA problems. Aside from innovations we have made in adapting natural scene statistic models to the IQA problem [16], our work on the NR IQA problem (as with the work of others) has really been "distortion assessment," whereby a specific distortion model, such as JPEG2K, is used to drive an objective algorithm that seeks to predict subjective quality. While distortion modeling is important, it does not necessarily embody perceptual relevance (distortion annoyance), since such factors as masking and contrast sensitivity need to be considered. Recently, we have been working hard on trying to make inroads on this formidable problem.

So, in this talk I will discuss some of our recent work on using machine learning principles to create effective NR IQA indices. Machine learning (ML), of course, implies that an algorithm is trained so that it can be successfully applied in application. Ideally, one would desire to be able to formulate and implement a quantitative theory of quality perception that would make it possible to avoid the "training and testing" solution. The reasoning is two-fold: firstly, we should, perhaps, be able to integrate quantitative visual models, distortion models, and natural scene statistics models in such a way as to make "training" unnecessary. This delivers the added advantage of understanding the algorithm, not simply creating it. The second reason is more important: the image database. Image quality databases are a touchy subject, especially when one creates both databases and algorithms, as we have done. It is even worse when one's algorithm does well on one's own database. I therefore applaud all efforts to create new IQA databases, as long as they are done properly - no simple task! More saliently, training on a database raises questions of the generalizability of a trained algorithm. Although precautions can be taken, one wonders if perhaps training can be avoided. Yet even as I try to convince myself of this, I realize that the perceptual process of distortion sensitivity in humans is the byproduct of neural adaptation to both images, and to some degree, distortions. Today's average consumer has been exposed to digital image and video distortions, is more savvy about why they are there, and may even know about such ideas as “compression” and “blocking.” This knowledge is likely coupled with a modified sensitivity to such distortions. In other words, our visual systems are trained, and perhaps algorithm training is likewise unavoidable.

So we have embarked upon ML-based methods for NR IQA towards learning whether the problem is, indeed, possible. To elaborate, whether it is possible to create NR IQA algorithms that, trained or otherwise, can either assess quality effectively in the absence of knowledge of the affecting distortion(s), or alternatively, can make an educated guess at the distortion(s) that are present, and then assess quality conditioned on that guess. Both approaches have merit, in my view.

The BLIINDS Index

As an example of the first of these, I will explain a new algorithm that we have developed that seeks to blindly assess image quality, by using measurements of natural image statistics as features to be trained on. The idea is simple: that natural images obey statistical laws that biological systems have evolved in respect to [17], and that moreover, man-made distortions modify these statistics, thus making the statistics of distorted images “unnatural” [6], [16]. Further, that appropriate features can be extracted that distinguish statistical unnaturalness and that can be trained and tested on. With an eye towards pragmatism, this first algorithm, which we have dubbed the **BL**ind **I**mage **I**ntegrity **N**otator using **D**CT **St**atistics (BLIINDS) index, operates by extracting statistics from the discrete cosine transforms (DCTs) of local image patches. I will describe the algorithm in more detail in the talk, but I will say that BLIINDS is an NR IQA algorithm that does not assume a specific image distortion type, uses no specific distortion model, extracts only a few simple features, and requires minimal training [18]. Nevertheless, the BLIINDS index correlates quite highly with human subjective judgment when trained and tested on a content-divided LIVE database: specifically, it outperforms the FR traditional metric, the Peak Signal-to-Noise Ratio (PSNR). While the PSNR is hardly an admirable image quality index (with apologies to traditionalists) [1], [19], it is *Full Reference.* Incidentally, the BLIINDS index algorithm is available for free download from the LIVE Image Quality website [20].

A Framework for Design

We (my students and I) are also taking another approach to the design of general-purpose NR IQA algorithms. Actually, we are not designing algorithms for QA so much as we are creating an NSS- and ML-based *framework* for designing algorithms. We call this the Blind Image Quality Index (BIQI) framework. The framework is unique and proceeds in two steps: given an image to be quality-assessed, first decide what distortion(s) have impaired the image, then based on that decision, deploy distortion-specific QA index(s) appropriate to the distortion(s), combining them to yield an holistic assessment of the image’s quality. The method, which I will describe in detail in the talk, relies on the use of simple NSS features that we have found to be reliably modified by common image distortions [21]. Not only do distortions affect NSS, the effects are quite systematic and parameterizable, making them ideal as training features for ML-based IQA index design. Such distortion-specific signatures make it possible to classify an image into particular distortion category(s). Once such classification is achieved, it is as if the algorithm is aware of the distortion. The algorithm can then deploy a distortion specific IQA algorithm, from among the many interesting ones that can be devised. We have created and tested such algorithms using existing “off-the-shelf” blind, distortion-specific image quality indices, and again obtained levels of performance quite competitive with simple *Full Reference* IQA indices. Moving forward, we expect to design and deploy such algorithms but imbued with NSS-specific models, and using effective visual models to enhance the ML-based training and testing results.

DMA1.pdf

A Vision for Ubiquitous Learning Blind Quality Agents

Our development of general-purpose NR IQA indices, such as BLIINDS and BIQI, that can outperform the PSNR suggests that indeed, blind IQA is not impossible. We have just begun injecting specific perceptual attributes into these algorithms, and it is our hope that the result may be blind algorithms that compete with good FR IQA indices, such as multiscale SSIM (MS-SSIM) [4]. However, our vision extends far beyond blind IQA algorithms. What we would like to attain are blind, learning *Video Quality Agents* – software agents that we can blindly deploy in wide-area wireless and wireline data communications networks, going beyond simple network and packet checking algorithms. Eventually, as elements of every smart switch and router, in every set-top box, every smart phone, PC and laptop. If such agents can be created, then in today's increasing video-centric consumer data communications environment, such algorithms could represent a sea change in elevating the QoS of multimedia data delivery.

Just for a start, the presence of such autonomous Quality Agents in next generation video networking, computing and storage equipment as standard QoS tools could enable network adjustment, re-requests for videos, identification of faulty sensors (e.g., in visual sensor networks), and other corrective and control tasks. Deploying automatic quality agents in next-generation IP TV networks and HD video delivery systems could lead to significantly elevated levels of QoS control. VQA agents distributed over wireless networks or embedded in cell phones would allow cell operators to map video quality as functions of video source, cell and phone locations, power allocation and other conditions, enabling *visually optimized* dynamic bandwidth and carrier allocation, and perceptually optimized, adaptive source coding, channel coding and error protection.

The Real Blind Problem: Stereo Image Quality Assessment

Whatever one's opinion may be regarding the vapidity of the plotline of the blockbuster 3D adventure *Avatar*, it has created a suddenly heightened awareness of the incredible (and long-nascent) commercial possibilities of 3D video. No coincidentally, there has been a concurrent explosion in technical advances and commercial sales of 3D stereoscopic visualization products. Delivering the 3D visual experience has long been an extremely appealing goal of television, cinematic, and gaming industries, but only recently has the stereo experience become adequately satisfying to significantly drive products. Consumer interest is now high and growing rapidly, and temporally-interleaved stereoscopic presentation displays with synchronized polarized glasses (and also glasses-free auto-stereoscopic displays) are becoming affordable and provide very nice (and improving) stereoscopic experiences. Soon, desktop computers and laptops, digital cinema, HD TVs, and likely, mobile devices that offer 3D will be common. With this proliferation, being able to estimate the perceptual quality of the 3D experience is exceedingly important, as Quality-of-Service issues become paramount. Naturally, the goal is to maximize stereoscopic visual signal quality given constraints (e.g., bandwidth) and to deploy quality-optimized compression and transmission tools that ensure a certain minimum level of perceived quality by the end user.

Stereoscopic IQA, on the other hand, is a problem that is largely unsolved owing to a number of fundamental factors. First, it must be understood that the goal of Stereo IQA must ultimately be to capture or predict the *subjective quality of the 3-D stereoscopic experience.* This means that a Stereo IQA algorithm should, ideally, be able to assess the quality of the *cyclopean* 3D image, which consists of both a depth sensation in space, as well as a mapping of light patterns reflected from the surfaces in depth. It must be realized that *there is no possible ground truth* (reference stereo experience) for the stereo-perceived scene. I do believe that FR IQA of stereo images [22]-[32] is better than nothing, does have some important applications, and can serve as a useful bridge to the much more general NR approach.

DMA1.pdf

However, the NR Stereo IQA problem is what is really needed to be solved – since there is no reference signal that adequately re-creates the stereoscopic experience: the perception of 3D surfaces in space that are textured by luma/chroma patterns – the fused *cyclopean image*. Simply comparing Left and Right images reconstructed by stereo algorithms (the most sophisticated approach to date) does not come close to this. Indeed, in our view such approaches do not really go beyond testing the quality of the monocular images and the stereo algorithm used, which are much easier problems [33].

Our own work on the general Stereo IQA problem is in very early stages. Yet since we think the problem is important and unsolved, we are putting significant effort into it. In my talk I will describe both our progress and our future plans, which are ambitious. As I will explain, we have developed a comprehensive and fundamental strategy for solving the NR Stereo IQA problem with algorithms that seek to access the direct stereo experience of human viewers; that will use the natural range and disparity statistics of stereo images which we are currently measuring [34], [35]; that will incorporate known neural properties of binocular vision; and that will be rigorously tested on a unique and large Stereo IQA Database complete with ground truth images. As 3D images and videos become increasingly important products of today's video-centric communications environment, and as the QoS of such services becomes more pressing, I believe that such algorithms will represent an important and timely component of 3D multimedia data delivery.

References

[1] B. Girod, "What's wrong with mean-squared error?" in *Visual Factors of Electronic Image Communications*. Cambridge, MA: MIT Press, 1993.

[2] P.C. Teo and D.J. Heeger, "Perceptual image distortion," *IEEE Int'l. Conf. Image Processing,* Austin, TX, Nov. 1994.

[3] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, pp. 600–612, Apr. 2004.

[4] Z. Wang, E.P. Simoncelli, and A.C. Bovik, "Multi-scale structural similarity for image quality assessment," *IEEE Asilomar Conf. Signals, Systems, Comput.*, Asilomar, CA, Nov. 2003.

[5] D.M. Chandler and S.S. Hemami, "VSNR: A wavelet-based visual signal to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sept. 2007.

[6] H.R. Sheikh and A.C. Bovik, "Image information and visual quality," *IEEE Trans. Image Processing*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[7] H.R. Sheikh and A.C. Bovik, "An evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Processing,* vol. 15, no. 11, pp. 3440-3451, Nov. 2006.

[8] H.R. Sheikh, Z. Wang, L.K. Cormack, and A.C. Bovik, "LIVE image quality assessment database," [Online]. Available: http://live.ece.utexas.edu/research/quality/

[9] N. Ponomarenko, M. Carli, V. Lukin, K. Egiazarian, J. Astola, F. Battisti "Color image database for evaluation of image quality metrics," *Proc. Int'l. Wkshp. Multimedia Signal Process.,* Australia, pp. 403-408, Oct. 2008.

[10] A. Ninassi, O. Le Meur, P. Le Callet and D. Barba, "Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric," *IEEE Int'l Conf Image Process,* San Antonio, TX, Sept 2007.

[11] A.K. Moorthy and A.C. Bovik, "Visual importance pooling for image quality assessment," *IEEE J. Special Topics Signal Process,* vol. 3, pp. 193-201, April 2009.

[12] S. Lee, M.S. Pattichis and A.C. Bovik, "Foveated video quality assessment," *IEEE Transactions on Multimedia,* vol. 4, no. 1, pp. 129-132, March 2002.

[13] M. Hayhoe and D. Ballard, "Eye movements in natural behavior," *Trends in Cognitive Sciences,* vol.9, no.4, pp. 188-194, April 2005.

[14] K. Seshadrinathan and A.C. Bovik, “Motion-tuned spatio-temporal quality assessment of natural videos,” *IEEE Trans Image Process,* vol. 19, no. 2, pp. 335-350, Feb 2010.

[15] A.C. Bovik, “Meditations on video quality,” invited Distinguished Position Paper, *IEEE Multimedia Communications E-Letter,* vol. 4, no. 4, pp. 4-10, May 2009.

[16] H.R. Sheikh, A.C. Bovik and L.K. Cormack, “No-reference quality assessment using natural scene statistics: JPEG2000,” *IEEE Trans Image Process,* vol. 14, no. 11, pp. 1918-1927, Nov. 2005.

[17] B.A. Olshausen and D.J. Field, “Natural image statistics and efficient coding,” *Network: Computation in Neural Syst,* vol. 7, pp. 333–339, 1996.

[18] M. Saad, A.C. Bovik and C. Charrier, “A DCT statistics based blind image quality index,” *Signal Processing Letters.*

[19] Z. Wang and A.C. Bovik, “Mean squared error: Love it or leave it? — A new look at signal fidelity measures,” *IEEE Signal Processing Magazine,* vol. 26, no. 1, pp. 98-117, Jan. 2009.

[20] “Image Quality Research at LIVE” [Online]. Available: http://live.ece.utexas.edu/research/quality/.

[21] A.K. Moorthy and A.C. Bovik, “Statistics of natural image distortions,” *IEEE International Conference on Acoustics, Speech and Signal Processing,* Dallas, Texas, March 14-19, 2010.

[22] P. Campisi, P. Le Callet, and E. Marini, “Stereoscopic images quality assessment,” *Euro Signal Process Conf.*, Poznan, Poland, September 2007.

[23] A. Tikanmäki, A. Gotchev, A. Smolic and K. Müller, “Quality assessment of 3D video in rate allocation experiments,” *IEEE Int’l Symp Consum Electr,* Algarve, Portugal, April 2008.

[24] H. Shao, X. Cao and G.H. Er, “Objective quality assessment of depth image based rendering in 3DTV system,” *IEEE 3DTV Conference,* Potsdam, Germany, May 4-6, 2009.

[25] J. You, L. Xing, A. Perkis and X, Wang, “Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis,” *Int’l Wkshp Video Process Quality Metrics for Consumer Electronics,* Scottsdale, AZ, January 2010.

[26] A. Benoit, P. Le Callet, P. Campisi and R. Cousseau, “Quality assessment of stereoscopic images,” *EURASIP Journal on Image and Video Processing,* vol. 2008, Article ID 659024, 13 pages, 2008. doi:10.1155/2008/659024.

[27] M. Subedar and L. Karam, “A study of relation between blur and depth in stereoscopic images,” *Int’l Wkshp Video Process Quality Metrics for Consumer Electronics,* Scottsdale, AZ, January 2010.

[28] L. Goldmann, F. De Simone and T. Ebrahimi, “Impact of acquisition distortions on the quality of stereoscopic images ,” *Int’l Wkshp Video Process Quality Metrics for Consumer Electronics,* Scottsdale, AZ, January 2010.

[29] A. Boev, A. Gotchev, K. Egiazarian, A. Aksay, G. Bozdagi Akar, "Towards compound stereo-video quality metric: a specific encoder-based framework", *Southwest Symp. Image Anal Interp,* Denver, Colorado, USA, March 2006.

[30] C.T.E.R. Hewage, S. Worrall, S. Dogan and A.M. Kondoz, “Prediction of stereoscopic video quality using objective quality models of 2D video,” *Elec. Lett.*, vol. 44, pp. 963-965, July 2008.

[31] P. Gorley and N. Holliman, "Stereoscopic image quality metrics and compression,” *SPIE Conf Stereoscopic Displays Appl.*, vol. 6803, Jan 2008.

[32] J. Yang, Ch. Hou, “Objective quality assessment method of stereo images,” *IEEE 3DTV Conference,* Potsdam, Germany, May 4-6, 2009.

[33] W. Malpica and A.C. Bovik, “Range image quality assessment by structural similarity,” *IEEE Int’l Conf Acoust, Speech Signal Processing,* Taipei, Taiwan, Apr 2009.

[34] Y. Liu, L.K. Cormack and A.C. Bovik, “Disparity statistics in natural scenes,” *Journal of Vision,* vol. 8, no. 11, art. 19, pp. 1-14, Aug 2008.

[35] Y. Liu, L.K. Cormack and A.C. Bovik, “Dichotomy between luminance and disparity features at binocular fixations,” *Journal of Vision,* in revision, 2010.

[36] Y. Liu, A.C. Bovik and L.K. Cormack, “Statistical modeling of luminance, range, and disparity in natural scenes,” *IEEE Transactions on Image Processing,* submitted.