DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-205

Objective Image Quality Assessment: Facing The Real-World Challenges

Zhou Wang
Dept. of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada
Email: zhou.wang@uwaterloo.ca

Abstract

There has been a growing interest in recent years in the development of objective image quality assessment (IQA) models, whose roles are not only to monitor image quality degradations and benchmark image processing systems, but also to optimize various image and video processing algorithms and systems. While the past achievement is worth celebrating, a number of major challenges remain when we apply existing IQA models in real-world applications. These include obvious ones such as the challenges to largely reduce the complexity of existing IQA algorithms and to make them easy-to-use and easy-to-understand. There are also challenges regarding the applicability of existing IQA models in many real-world problems where image quality needs to be evaluated and compared across dimensionality, across viewing environment, and across the form of representations – specific examples include quality assessment for image resizing, color-to-gray image conversion, multi-exposure image fusion, image retargeting, and high dynamic range image tone mapping. Here we will first elaborate these challenges, and then concentrate on a specific one, namely the generalization challenge, which we believe is a more fundamental issue in the development, validation and application of IQA models. Specifically, the challenge is about the generalization capability of existing IQA models, which achieve superior quality prediction performance in lab testing environment using a limited number of subject-rated test images, but the performance may not extend to the real-world where we are working with images of a much greater diversity in terms of content and complexity. We will discuss some principle ideas and related work that might help us meet the challenges in the future.

Introduction

Over the past decades, a growing number of researchers and engineers in the image processing community have started to realize the importance of image/video quality assessment (IQA/VQA) [40, 29, 4]. This is not surprising because no matter what image/video processing problems we are working on, the same issues repeatedly come up – How should we evaluate the images generated from our algorithms/systems? How do we know our algorithm/system is creating an improvement between the input and output images, and by how much? How can we know one algorithm/system performs better than another, and by how much? What should be the quality criterion for which the design of our algorithms/systems should be optimized? Since the human eyes are the ultimate receivers in most image processing applications, human subjective visual testing would be a reliable solution. However, with the exponential increase of the volume of image/video data being generated daily, it becomes impossible to address these quality issues in a timely manner by subjective visual testing, which is slow, cumbersome and expensive. Instead, only trusted objective IQA models may potentially meet these needs.

In academia, objective IQA has been a hot research topic, especially in the past 15 years [35, 4, 29]. First, the commonly used numerical disotrtion/quality measures in the past – the mean squared error (MSE) and the peak signal-to-noise ratio (PSNR) – have been shown to correlate poorly with perceived image quality [28, 30]. Second, a large number of perceptually more meaningful IQA models have been proposed, including full-reference (where a perfect quality reference image is available when evaluating a distorted image) [35, 4, 29], no-reference (where the reference image is not accessible) [34, 24, 31], and reduced-reference (where only partial information about the reference image is available) models [39, 36, 31, 29]. Third, several design principles have been discovered and repeatedly demonstrated to be useful in the design and improvement of IQA models. These include psychophysical and physiological visibility models [35, 4], the structural similarity (SSIM) approaches [28, 32, 33, 20, 49], the natural scene statistics (NSS) and information theoretic approaches [36, 39, 21, 31], the visual saliency based approaches [50], and the machine learning based approaches [6]. Fourth, a number of subject-rated image quality databases have been created and made publicly available [22, 7, 8, 17, 16, 47]. They provide a common benchmark platform for the evaluation and comparison of IQA models, among which several algorithms have achieved high correlations with the subjective mean opinion scores (MOSs) of the test images [23, 38, 33, 49]. In the video delivery industry, perceptual objective IQA methods such as the SSIM algorithm have been incorporated into many practical hardware and software systems to monitor image/video quality degradations and to test/compare image/video encoders and transcoders [27, 25, 26]. The wide use of SSIM has resulted in a Primetime Engineering Emmy Award given by the Academy of Television Arts and Sciences [1].

The remarkable development and successful deployment of modern IQA methods are definitely worth celebrating. Nevertheless, this does not necessarily mean that the existing IQA models have already met the real-world challenges. Otherwise, they should have made a much stronger impact and become a game-changing factor in the industry. Using the video delivery industry as an example, even now most practitioners are still equating bitrate with quality in the practical design of video delivery architectures. However, using the same bitrate to encode different video content could result in dramatically different visual quality. Clearly, the perceptual quality of the video itself, which is presumably the ultimate evaluation criterion of the whole video delivery system, has not been placed at the driver's seat. While it is understandable that quality degradation is inevitable at many stages in the video delivery chain due to practical constraints, the real concern here is that there is no existing protocol to monitor and control such quality degradation. As a result, various tricks have been used to manipulate the video content and network resources are allocated in suboptimal ways, leaving the creative intent of the content producers unprotected.

While it is certain that the industry needs to be better informed about the great potentials of making the best use of IQA/VQA models, we believe that an equally important aspect that slows down the process is that the existing IQA/VQA models still do not meet many real-world challenges. In the following sections, we will elaborate some of these challenges and then focus on a specific one, namely the generalization challenge. We wish our discussions on some fundamental ideas could provide some useful insights for the future development of IQA models that may meet these real-world challenges.

The Real-World Challenges

Here we make a list of real-world challenges, many of which are described in more details through examples of practical scenarios.

1. It is highly desirable to reduce the complexity of the IQA/VQA algorithms so that they can be computed in real-time or in an even faster speed. This is especially useful in time-sensitive applications such as live broadcasting and videoconferencing. Many existing models are far from meeting this challenge.
2. It is essential to make the IQA/VQA scores easy-to-use and easy-to-understand. For example, the raw SSIM score does not have an explicit perceptual meaning, making it difficult to determine what level of SSIM index can warrant an excellent video quality and how much improvement in the SSIM index is sufficient to create visible quality improvement. Mapping the raw scores into a perceptually linear domain that is easily linked to human expressions about image quality is desirable.
3. The same video stream shown on different display devices could result in very different perceptual quality. For example, a strongly compressed video that exhibits very annoying artifacts on a large TV could appear to have fine quality when viewed on the screen of a smartphone. The quality may also change significantly when the video is watched on the same TV but at two different viewing distances, one at the default distance and the other at a very close distance. However, existing IQA/VQA models give the same score based on the video stream only, completely ignorant of the viewing device and viewing condition.
4. In a video-on-demand application, a high-quality high-resolution (e.g., 4K) source video may be encoded into multiple video streams of different resolutions (e.g., 1080p, 720p, 360p, 240p, etc.) and different bit rates, aiming for satisfying a variety of user needs. In order to measure the quality of the encoded videos, most existing VQA models cannot be computed because the source (reference) and test videos have different spatial resolutions.
5. An image or video may need to be displayed on a screen that has a spatial resolution higher than that of the image resolution. As a result, spatial interpolation is performed. Again, most existing VQA models are not applicable because the reference and test images have different spatial resolutions.
6. An image or video of imperfect quality (e.g., being compressed at an earlier stage) is received and then transcoded to multiple images or videos with different bitrates and resolutions. Most existing IQA/VQA models are not applicable not only because they do not allow for cross-resolution quality assessment, but also because they assume the original reference image/video to have perfect quality, which is not the case here. How to carry out "degraded reference" IQA/VQA is a major challenge.
7. A high dynamic range (HDR) image (e.g., the pixels are in 10 or more bit depths) is tone mapped to a standard dynamic range (SDR) image (8 bits per pixel) in order to be visualized on an SDR display. There is certainly information loss that we would like to capture. However, most existing IQA models do not apply because they cannot compare images/videos with different dynamic ranges.
8. A high frame rate (HFR) video (e.g., 60fps, 120fps, or higher) is downsampled along temporal direction to a low frame rate (LFR) video (30fps, 24fps, 15fps, etc.). One would like to know what is the impact of the downsampling process on visual quality, or what is the quality gain/loss when switching between HFR and LFR videos, especially for the video content that contains significant amount of motion. Most VQA models do not apply because they cannot compare videos with different temporal resolutions.
9. A video with lower frame rate (LFR) needs to be played on a viewing device that allows for HFR video. Temporal interpolation methods may be applied beforehand to create HFR video content from the LFR videos. It is desirable to know whether the perceptual quality is improved through the process, but most existing VQA models are not useful because they cannot compare videos with different temporal resolutions.
10. A camera is used to capture a real natural scene multiple times, each with a different exposure level, so as to faithfully record all structural details in the scene, which includes both extremely dark and extremely bright regions. In order to visualize all details in a single image, multiple exposure fusion algorithms may be applied to combine a sequence of multi-exposure images into one image. Most existing IQA models cannot be used to evaluate the quality of the fused image, because the reference and test images are in different formats, one being a sequence of images and the other being a single image.
11. A color image is converted into a grayscale image for purposes in the subsequent processes (e.g., display, printing, etc.). In order to assess the quality of the resulting grayscale image (which inevitably loses information from the color image), we would need to use the original color image as the reference. However, most existing IQA model is not useful because the reference and test images are in different formats, one containing multiple color channels and the other being monotone.
12. An image or video of large size needs to be shown on a screen of a smaller size. Certain image retargeting algorithm

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-205

may be applied to reduce the total number of pixels in the picture without sacrificing too much of the detailed content in the original picture. Again, existing IQA/VQA models do not apply because the reference and test images/videos have different spatial resolutions.

The list may be easily extended further, especially when we consider the variations of image/video content type in certain real-world application environments. For example, specific challenges may arise when we apply IQA/VQA methods in specific applications such as online gaming, videoconferencing and sports broadcasting.

It is worth mentioning that new methods have started to emerge in recent years to address some of these challenges. For example, the recently proposed SSIMplus algorithm is able to evaluate the quality of a test image/video in real-time using a reference image/video that has a different resolution, and meanwhile provide perceptually linear quality scores dependent on the viewing device and viewing condition [18]. The tone mapped image quality index (TMQI) is able to evaluate image quality across dynamic ranges [5]. The multi-exposure fusion SSIM algorithm (MEF-SSIM) is capable of assessing the quality of fused image created from a sequence of images captured at different exposure levels [10]. The C2G-SSIM algorithm is designed to assess the quality of a gray scale image that is converted from a color image [11]. The WIND algorithm can evaluate the quality of an interpolated high-resolution natural image using a low-resolution image (which was used to create the high-resolution image based on an interpolation algorithm) as the reference image [46]. There has been a clear trend that more effort will be dedicated to these research directions in the future.

The Generalization Challenge

As mentioned earlier, a large number of IQA models have been proposed in recent years. To validate and compare these models, the standard approach is to first build databases of images with various content and distortions, and then carry out subjective experiments to score all images for their quality. So far, several image databases with subjective ratings have been widely recognized and used by the research community. These include the LIVE [22], CSIQ [7], IVC [8], TID2008 [17], TID2013 [16], and VCL@FER [47] databases. Given these databases, several correlation metrics between subjective MOS scores and objective model predictions can then be calculated, and the models that obtain higher correlation numbers are believed to have better model performance. In fact, many existing IQA models have been reported to achieve very high correlations (mostly in the upper 80% to upper 90% range) with subjective scores when tested using these databases. The question is whether such high performance can generalize to real-world images outside these databases.

To validate the generalization capability of these IQA models is not an easy task. This is due to the conflict between the extremely large size of the space of images and the small scale of the affordable subjective experiment. It needs to be aware that subjective testing is time-consuming and expensive. A typical "large-scale" subjective experiment only allows for several hundreds (or at most a few thousands) test images to be rated by multiple human viewers. As a result, all of the widely known image quality databases can only accommodate at most a few dozens of source reference images, because in addition to these source images, a much larger number of distorted images need to be added based on the combination of distortion types and distortion levels. An additional problem is that many of the current well-known image quality databases employ similar or sometimes the same source images, which further limit the diversity of image content.

In the real-world, however, digital images live in a high dimensional space. Note that the dimension equals the number of total pixels in the image. For example, an image that contains one million pixels lives in a one million dimensional space. By contrast, a manageable subjective experiment can only evaluate several hundreds or at most a few thousands images. This corresponds to sampling the one million dimensional space with a few thousands samples, which is deemed to be extremely sparsely distributed in the space. Furthermore, in terms of the diversity of image content, no matter how the source images are selected, it is difficult to justify how a few dozens of source images can sufficiently represent the diversity of the content types in the real-world. Considering all the facts above, it is thus natural to question if the reported highly competitive performance of some of the existing IQA models can be generalized to the real-world, where images have much richer content types and are undergoing a much wider variety of distortions. Indeed, recently there has been direct evidence showing that the performance of existing IQA models degrade largely in a relatively larger database composed of real-world Internet images [2].

In the future development and validation of novel IQA models, in order to properly address the generalization challenge, we would need to first of all work with a much larger image database, which is ideally in the scale of at least thousands of source images and hundreds of thousands of distorted images. Of course, carrying out a complete subjective test that could allow us to collect the MOS scores for all the images in such a database becomes impossible. Therefore, innovative ways on how to explore such an image database with limited resource for subjective testing is the key to success. Such innovations may need us to change the fundamental principles in model selection and subjective testing.

One of such novel design principles is the MAximum Differentiation (MAD) competition methodology [37] and its extensions [9]. The most fundamental idea is to switch our goal from attempting to "prove" a model to "disapprove" a model. Note that the application domain of an IQA model is the space of all possible images. To prove a model in such a domain, as in the traditional approach to validating IQA models, we would need to use samples that provide sufficient coverage or representation of all images live in the domain. Given the complexity of the IQA problem and the size of the domain, this requires an extremely large number of samples, which is a task impossible to achieve with the current capacity of subjective testing. By contrast, to disapprove a model is much "easier", because ideally even one "counter-example" is sufficient to achieve the goal. Apparently, a significant advantage of the new principle is the great potential to reduce the total number of necessary samples subject to subjective testing. These test samples need to be carefully selected, which is the focus of the second essential component of the MAD methodology, where rather than hand designing or manually searching for the best counter-examples, we use an efficient and automatic way to find potential "counter-examples" that are most efficient in falsifying one model using another. We refer the model we at-

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-205

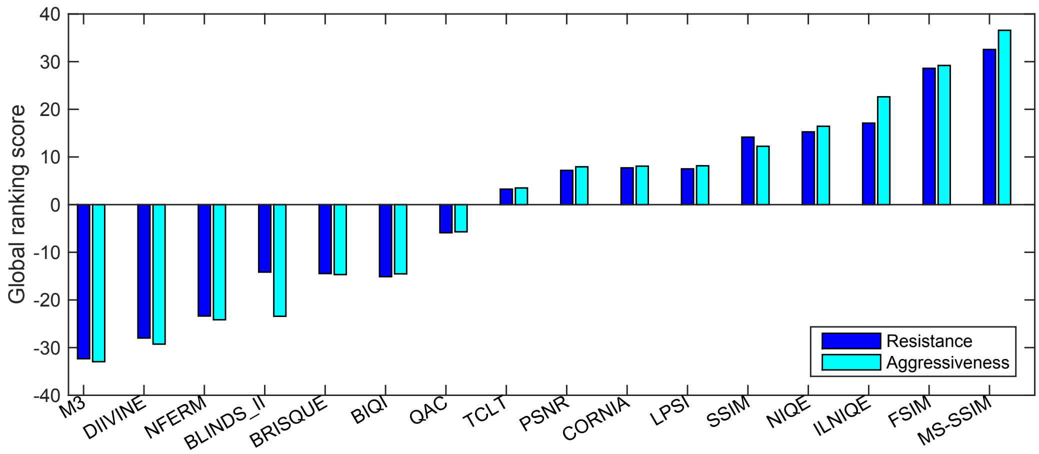

Figure 1: IQA models ordered based on Resistance and Aggressiveness in MAD competition.

tempt to disapprove as the "defender" model, and the one used to disapprove the defender model as the "attacker" model. Specifically, in MAD competition, for any given quality level defined by the defender model, we automatically search in the space of all images (or practically an image database that is as large as possible) to find a pair of images with the maximum/minimum quality in terms of the defender model. This pair of images are subject to a subjective discriminative test, where human subjects are asked to choose the image that has better quality than the other. If one image is clearly better than the other (with a significantly higher number of subject votes), than the attack is successful and the defender model is disapproved. Otherwise, the attack "fails" and the defender model "survives" in this attack, which provides useful evidence on the robustness or reliability of the model, and in turn reveals certain drawbacks of the attacker model.

In [9], 16 classic and state-of-the-art IQA models are included in a group MAD competition experiment. These include These include full-reference models PSNR, SSIM [32], MS-SSIM [38] and FSIM [49], and no-reference models BIQI [14], BLINDS_II [19], BRISQUE [12], CORNIA [45], DIIVINE [15], IL-NIQE [48], LPSI [42], M3 [43], NFERM [3], NIQE [13], QAC [44] and TCLT [41]. The experiment exploits a database that contains more than 4,000 source natural images together with more than 90,000 distorted images generated from them. This database is much larger than all exiting image quality databases combined. The MAD competition is performed for all possible combinations of attacker-defender roles of the competing IQA models on 6 quality levels. This results in 1440 pairs of images selected. Note that the number of the selected image pairs depends on the number of competing models and the number of test quality levels, but does not depend on the size of the image database. As a result, exploiting an even larger image database does not lead to increased cost of the subjective testing. These selected image pairs are employed in a subjective discriminative test with 31 participating subjects. Statistical analysis is performed after the subjective testing, and two evaluation criteria are used to summarize the overall behavior of each IQA model – a "resistance" index that measures how robust of an IQA model as a defender under the attacks of all other models, and a "aggressive" index that evaluates how successful of an IQA model as an attacker that fails all other models as the defender. Figure 1 shows the competition results of the IQA models ordered based on their resistance and aggressiveness. The test results as well as the pairs of images selected through the competition not only provide a novel measure about the relative performance and reliability of an IQA model, but also reveals the relative strength and weakness of the models, which further provide useful insights on potential ways to improve the models.

Conclusion

Remarkable progress has been made in the past decades in the field of IQA/VQA, evidenced by a number of state-of-the-art IQA/VQA models achieving high correlations with subjective quality opinions on images/videos when tested using publicly available image/video quality databases. Nevertheless, we show that these achievements are still not enough to facilitate the wide usage of IQA/VQA models due to a number of real-world challenges. It is very important to be aware of such challenges so as to develop future IQA/VQA models or improve existing ones such that they can be converted to industrial products that are fast, accurate, easy-to-use, easy-to-understand, and applicable across dimensionality, viewing environment, and the form of representations. We have also discussed in more detail on the generalization challenge, which is believed to be a more fundamental issue related to the way how IQA models should be developed and validated. We discussed the idea behind the MAD competition methodology, which has shown some promising characteristics that may lead to significant changes in the future development, validation and application of IQA models.

Acknowledgments

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada, under the Discovery, Strategic, and Steacie Memorial Fellowship programs.

References

[1] Academy of Television Arts and Sciences. 67th Engineering Emmy Award. 2015. www.emmys.com/news/press-releases/honorees-announced-67th-engineering-emmy-awards. 1

[2] D. Ghadiyaram and A. C. Bovik. LIVE in the wild image quality challenge database. 2015. live.ece.utexas.edu/research/ChallengeDB/. 3

[3] K. Gu, G. Zhai, X. Yang, and W. Zhang. Using free energy principle for blind image quality assessment. *IEEE TMM*, 2015. 4
[4] H. R. Wu and K. R. Rao, Ed. *Digital Video Image Quality and Perceptual Coding*. CRC Press, Taylor & Francis Group, Boca Raton, 2005. 1
[5] Y. Hojatollah and W. Zhou. Objective quality assessment of tone-mapped images. *IEEE Transactions on Image Processing*, 22(2):657–667, 2013. 3
[6] L. Kang, P. Ye, Y. Li, and D. Doermann. Convolutional neural networks for no-reference image quality assessment. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1733–1740, 2014. 1
[7] E. C. Larson and D. M. Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *SPIE JEI*, 2010. 1, 3
[8] P. Le Callet and F. Autrusseau. Subjective quality assessment IRCCyN/IVC database 2005 [online]. Available: http://www.irccyn.ec-nantes.fr/ivcdb/. 1, 3
[9] K. Ma, Q. Wu, Z. Wang, Z. Duanmu, H. Yong, L. Zhang, and H. Li. A new methodology to compare objective image quality models. 2015. submitted. 3, 4
[10] K. Ma, K. Zeng, and Z. Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24:3345–3356, 2015. 3
[11] K. Ma, T. Zhao, K. Zeng, and Z. Wang. Objective quality assessment for color-to-gray image conversion. *IEEE Transactions on Image Processing*, 24:4673–4685, 2015. 3
[12] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. *IEEE TIP*, 2012. 4
[13] A. Mittal, R. Soundararajan, and A. C. Bovik. Making a "completely blind" image quality analyzer. *IEEE SPL*, 2013. 4
[14] A. K. Moorthy and A. C. Bovik. A two-step framework for constructing blind image quality indices. *IEEE SPL*, 2010. 4
[15] A. K. Moorthy and A. C. Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE TIP*, 2011. 4
[16] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 2015. 1, 3
[17] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti. TID2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, 2009. 1, 3
[18] A. Rehman, K. Zeng, and Z. Wang. Display device-adapted video quality-of-experience assessment. In *IS&T/SPIE Electronic Imaging*, pages 939406–939406–11, 2015. 3
[19] M. A. Saad, A. C. Bovik, and C. Charrier. Blind image quality assessment: A natural scene statistics approach in the DCT domain. *IEEE TIP*, 2012. 4
[20] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey. Complex wavelet structural similarity: A new image similarity index. *IEEE Trans. Image Processing*, 18(11):2385–2401, Nov. 2009. 1
[21] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Trans. Image Processing*, 15(2):430–444, Feb. 2006. 1
[22] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE TIP*, 2006. 1, 3
[23] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Processing*, 15(11):3440–3451, Nov. 2006. 1
[24] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. Blind quality assessment for JPEG2000 compressed images. In *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, Nov. 2002. 1
[25] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao. SSIM-motivated rate-distortion optimization for video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(4):516 – 529, 2012. 1
[26] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao. Perceptual video coding based on SSIM-inspired divisive normalization. *IEEE Transactions on Image Processing*, 22(4):1418–1429, 2013. 1
[27] Z. Wang. Applications of objective image quality assessment methods. *IEEE Signal Processing Magazine*, 28:137–142, Nov. 2011. 1
[28] Z. Wang and A. C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, Mar. 2002. 1
[29] Z. Wang and A. C. Bovik. Modern image quality assessment. *Synthesis Lectures on Image, Video, and Multimedia Processing*, 2006. 1
[30] Z. Wang and A. C. Bovik. Mean squared error: love it or leave it? a new look at signal fidelity measures. *IEEE SPM*, 2009. 1
[31] Z. Wang and A. C. Bovik. Reduced- and no-reference visual quality assessment - the natural scene statistic model approach. *IEEE Signal Processing Magazine*, 28:29–40, Nov. 2011. 1
[32] 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*, 13(4):600–612, Apr. 2004. 1, 4
[33] Z. Wang and Q. Li. Information content weighting for perceptual image quality assessment. *IEEE Trans. Image Processing*, 20(5):1185–1198, May 2011. 1
[34] Z. Wang, H. R. Sheikh, and A. C. Bovik. No-reference perceptual quality assessment of JPEG compressed images. In *Proc. IEEE Int. Conf. Image Proc.*, Rochester, Sept. 2002. 1
[35] Z. Wang, H. R. Sheikh, and A. C. Bovik. Objective video quality assessment. In B. Furht and O. Marques, editors, *The Handbook of Video Databases: Design and Applications*, pages 1041–1078. CRC Press, Sept. 2003. 1
[36] Z. Wang and E. P. Simoncelli. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In *Human Vision and Electronic Imaging X, Proc. SPIE*, volume 5666, San Jose, CA, Jan. 2005. 1
[37] Z. Wang and E. P. Simoncelli. Maximum differentiation (MAD) competition: A methodology for comparing computational models of perceptual quantities. *Journal of Vision*, 8(12):1–13, Sept. 2008. 3
[38] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multi-scale structural similarity for image quality assessment. In *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, pages 1398–1402, Pacific Grove, CA, Nov. 2003. 1, 4
[39] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E.-H. Yang, and A. C. Bovik. Quality-aware images. *IEEE Trans. Image Processing*, 15(6):1680–1689, June 2006. 1
[40] S. Winkler. *Digital Video Quality: Vision Models and Metrics*. Wiley Publishing, 2005. 1
[41] Q. Wu, H. Li, F. Meng, K. N. Ngan, B. Luo, C. Huang, and B. Zeng. Blind image quality assessment based on multi-channel features fusion and label transfer. *IEEE TCSVT*, 2015. 4
[42] Q. Wu, Z. Wang, and H. Li. A highly efficient method for blind image quality assessment. In *IEEE ICIP*, 2015. 4
[43] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng. Blind image quality assessment using joint statistics of gradient magnitude and

DOI: 10.2352/ISSN.2470-1173.2016.13.IQSP-205

Laplacian features. *IEEE TIP*, 2014. 4

[44] W. Xue, L. Zhang, and X. Mou. Learning without human scores for blind image quality assessment. In *CVPR*, 2013. 4

[45] P. Ye, J. Kumar, L. Kang, and D. Doermann. Unsupervised feature learning framework for no-reference image quality assessment. In *CVPR*, 2012. 4

[46] H. Yeganeh, M. Rostami, and Z. Wang. Objective quality assessment of interpolated natural images. *Image Processing IEEE Transactions on*, 24:4651–4663, 2015. 3

[47] A. Zarić, N. Tatalović, N. Brajković, H. Hlevnjak, M. Lončarić, E. Dumić, and S. Grgić. VCL@FER image quality assessment database. *AUTOMATIKA*, 2012. 1, 3

[48] L. Zhang, L. Zhang, and A. Bovik. A feature-enriched completely blind image quality evaluator. *IEEE TIP*, 2015. 4

[49] L. Zhang, L. Zhang, X. Mou, and D. Zhang. FSIM: a feature similarity index for image quality assessment. *IEEE TIP*, 2011. 1, 4

[50] W. Zhang, A. Borji, Z. Wang, P. Le Callet, and H. Liu. The application of visual saliency models in objective image quality assessment: A statistical evaluationapplications of objective image quality assessment methods. *IEEE Transactions on Neural Networks and Learning Systems*, 2015. 1

Author Biography

Zhou Wang received his PhD degree from The University of Texas at Austin (2001). He is currently a Professor in the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include perceptual image processing and quality assessment, computational vision, multimedia communications, and biomedical signal processing. He has over 100 publications with over 25,000 citations in these fields. He is a Fellow of IEEE, and a recipient of an Engineering Emmy Award, a Steacie Memorial Fellowship, an IEEE Signal Processing Society Best Paper Award, and an IEEE Signal Processing Society Best Magazine Paper Award. He has served on IEEE Multimedia Signal Processing Technical Committee, and the editorial boards of several leading IEEE journals.