Zhou Wang

Applications of Objective Image Quality Assessment Methods

The interest in objective image quality assessment (IQA) has been growing at an accelerated pace over the past decade. The latest progress on developing automatic IQA methods that can predict subjective quality of visual signals is exhilarating. For example, a handful of objective IQA measures have been shown to significantly and consistently outperform the widely adopted mean squared error (MSE) and peak signal-to-noise-ratio (PSNR) in terms of correlations with subjective quality evaluations [1]. It has been exciting to observe the new progress in both theoretical development and novel techniques on this multidisciplinary topic, which appears to be a converging point from a wide range of research directions and includes the following:

- signal and image processing
- computer vision
- visual psychophysics
- neural physiology
- information theory
- machine learning
- design of image acquisition, communication, and display systems.

While the field of objective IQA is still evolving quickly, and novel and better IQA methods will continue to emerge in the coming years, it is also interesting to discuss how we could make the best use of these tools in real-world applications. The purpose of this article is to provide an overview of the roles of IQA methods in these applications. We will start by a brief description of the current status of the IQA field, followed by discussions on benchmarking and monitoring applications of IQA measures. We will then discuss the applications of IQA measures in the design and optimization of advanced image processing algorithms and systems, where we perceive both great promises and major challenges. Finally, we will show how IQA measures could play important roles in an even more extended field of applications and provide a vision of the future.

Digital Object Identifier 10.1109/MSP.2011.942295
Date of publication: 1 November 2011

OBJECTIVE IMAGE QUALITY ASSESSMENT

Objective IQA measures aim to predict perceived image/video quality by human subjects, which are the ultimate receivers in most image processing applications. Depending on the availability of a pristine reference image that is presumed to have perfect quality, IQA measures may be classified into full-reference (FR), reduced-reference (RR), and no-reference (NR) methods. FR measures require full access to the reference image, while NR methods assume completely no access to the reference. RR methods provides a compromise in-between, where only partial information in the form of RR features extracted from the reference image is available in assessing the quality of the distorted image. IQA measures may also be categorized into application-specific or general-purpose methods. The former only applies to some specific applications where the types of distortions are often known and fixed, e.g., JPEG compression. The latter is employed in general applications, where one may encounter diverse types and levels of image distortions.

In the literature, a considerable number of IQA algorithms have been proposed, which exhibit substantial diversity in the methodologies being used. Meanwhile, they also share some common characteristics. In particular, all of them are rooted from certain knowledge in one or more of the following three categories (which, interestingly, constitute the basic building blocks in an information communication framework [1]):

1) knowledge about the image source, which can be either deterministic (when the reference image is fully available) or statistical (when certain statistical image models are employed)
2) knowledge about the distortion channel, which is often associated with some known facts about the specific distortion process that the images underwent, for example, blocking and blurring artifacts in JPEG compression, and blurring and ringing effects in wavelet-based image compression
3) knowledge about the receiver, i.e., the human visual system (HVS), where computational models originated from visual physiological and psychological studies play essential roles.

Until now, the area that has achieved the greatest success is FR IQA of grayscale still images. Several algorithms, including the structural similarity index (SSIM) [2] and its derivatives, and the visual information fidelity (VIF) [3], significantly outperformed PSNR and MSE in a series of tests based on large-scale subject-rated independent image databases. There have also been notable success in the areas of video quality assessment (VQA) as well as RR and NR IQA, especially application-specific methods [4].

On the other hand, there is also an abundant menu of unresolved IQA problems left for future studies, including the following:

- General-purpose RR and NR IQA, where the types of image distortions are unavailable, is still at an immature stage.
- Methods for effective IQA of texture images are still lacking.
- There have not been good solutions for cross-dynamic range and cross-resolution IQA, where the reference image is available but has a different dynamic range of intensity levels or a different spatial or temporal resolution from the image being assessed.
- IQA for image signals with extended dimensions creates many challenging research problems, which include video, color, multispectrum, hyperspectrum, stereo, multiview, and three-dimensional (3-D) volume IQA.
- IQA algorithms that can be used for evaluating segmentation, halftoning, and fusion algorithms are lacking.
- In pattern recognition applications, effective IQA methods are missing that can assess how the recognition accuracy is affected by image distortions.
- In medical imaging applications, it is highly desirable to evaluate how image distortions affect the diagnostic values (rather than perceptual appeal) in images.
- In network visual communications, it is worth investigating how information regarding the communication channel conditions, such as channel fading characteristics and packet loss rate and delay could be utilized in the IQA process.
- In multimedia systems, visual quality may not be the only factor that affects the overall quality-of-experience (QoE) of users. Joint audio-visual quality assessment and joint quality assessment and visual discomfort evaluations are ongoing research topics. The complication of QoE assessment is raised to an even higher level in immersive multimedia environments such as panoramic 3-D displays.

In the past few years, there has been great effort in the research community to develop advanced IQA measures to solve the problems described above. For example, many recent projects carried out by the Video Quality Experts Group (VQEG) [12] are attempting to address these issues. Meanwhile, there are also many attempts to apply objective IQA measures for a wide variety of real-world applications, which will be our major focus in the next sections.

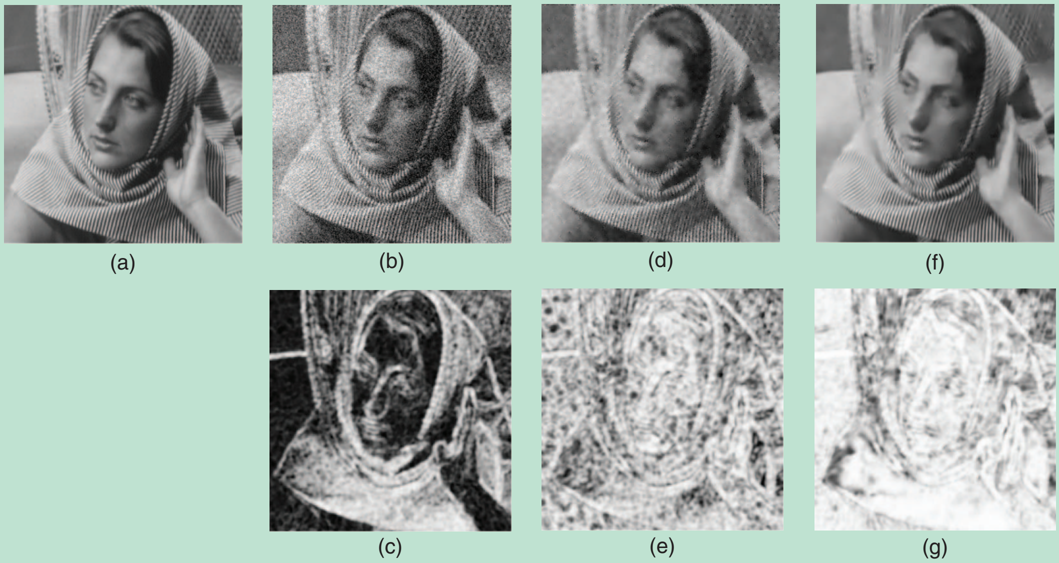
BENCHMARKING AND MONITORING APPLICATIONS

A direct application of IQA measures is to use them to benchmark image processing algorithms and systems. For instance, when multiple image denoising and restoration algorithms are available to recover images distorted by noise contamination and blur, a perceptual objective IQA measure can help pick the one that generates the best perceptual quality of the restored images. For another example, rate-distortion (RD) curves are often used to characterize the performance of image coding systems, where the RD function is defined as the distortion between the original and decoded images versus bit rate. A lower RD curve indicates a better image coder. Traditionally, MSE types of measures are employed to compute the distortion. If the role of MSE is replaced by a distortion function defined based on a perceptual IQA measure, then the RD curve could provide a perceptually more meaningful evaluation of the image coder.

> IN MANY IMAGE PROCESSING ALGORITHMS, THERE ARE CERTAIN PARAMETERS THAT NEED TO BE DETERMINED BY USERS TO YIELD THE BEST RESULTS.

A useful feature of many IQA measures that is often overlooked by practitioners is that they not only create overall quality scores of distorted images, but also produce quality maps that indicate local quality variations over the image space. An example is given in Figure 1, where the original "Barbara" image (a) is contaminated by additive white Gaussian noise. Two denoising algorithms, spatially adaptive Wiener filtering (MATLAB Wiener2 function) and K singular-value-decomposition (KSVD) filtering [5], are employed to recover the original image from its noisy observation. The quality maps created by the SSIM index [2], a popular IQA measure, for the noisy image (b) and the denoised images (d) and (f) are given by (c), (e), and (g), respectively. These quality maps provide useful information in several aspects. First, despite the fact that noise is imposed uniformly over space, the perceptual quality varies significantly across the image. For example, the face region looks much noisier than the texture regions. These are clearly indicated by the quality map (c); Second, the quality maps help identify where in the image the denoisers yield the most improvement, and how one denoiser outperforms another. For instance, by comparing (e) and (g), we observe significant improvement of KSVD over Wiener filtering on the smooth regions as well as the stripe texture regions at the bottom part of the image. Third, the quality maps also indicate where the denoisers still need further improvement. For example, the textures in the upper-right region of the image are not well denoised by both algorithms.

In many image processing algorithms, there are certain parameters that need to be determined by users to yield the best results. This is often a difficult task for naive users as the best values may be image dependent. A good IQA measure could be a useful tool to help decide on these parameters automatically. For example, in [6], the Q-index, an NR sharpness and contrast measure, was used to automatically pick the parameters for image denoising algorithm. The idea may be extended further when multiple complementary algorithms (or multiple modes under the



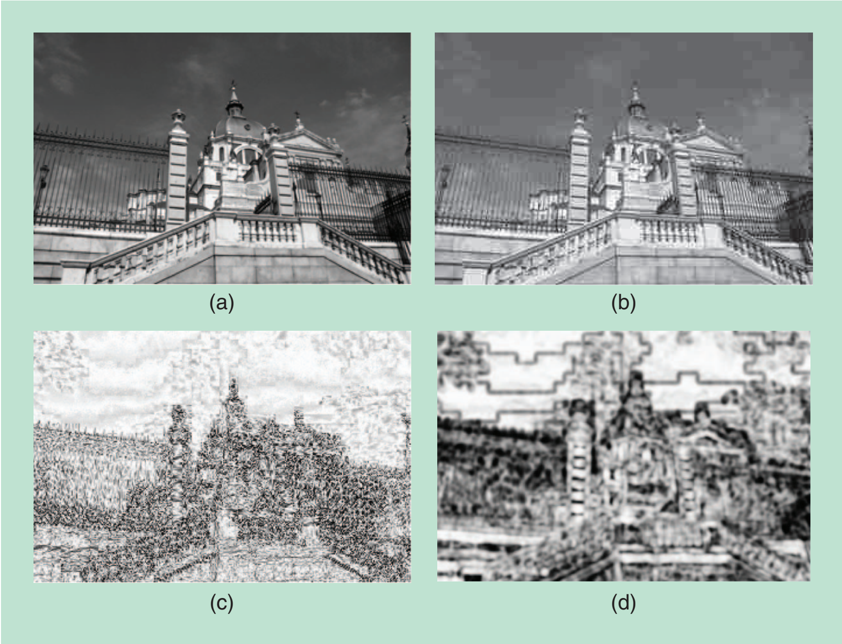
[FIG1] Example of performance analysis using IQA measures and quality maps. An original image (a) is contaminated by noise and (b) denoised by two denoising algorithms, resulting in (d) and (f), respectively. The SSIM-based quality maps [2] of the noisy and denoised images are shown in (c), (e), and (g), respectively, where brighter indicates better local quality.

same algorithm) are available for the same goal, for example, different coding modes in standard video compression systems. In such scenarios, an IQA measure can help select the right algorithm (mode), or to automatically switch between different algorithms (modes).

Objective IQA measures are particularly desirable in network visual communication applications for the purpose of quality-of-service (QoS) monitoring. Image and video content delivered over various wired and wireless networks inevitably suffer from visual quality degradations during lossy compression and transmission over error prone networks. It is imperative for the network service providers to monitor such quality degradations in real time, so as to optimize network resource allocations and maximally satisfy user expectations within certain cost constraints. It was shown that typical error criteria used in network design and testing, such as bit error rate (BER), do not correlate well with the quality of experience of network consumers [4]. Therefore, accurate and high-speed perceptual IQA measures can play important roles. Apparently, FR IQA methods are less useful here because the original video signals (typically with extremely high data rate) would not be available at the mid or end nodes in the network. NR methods are desired but are difficult to develop. This is mainly due to the complication of the types of distortions that could occur during video transmission in modern communication networks, where the distortions could be caused by a combination of lossy compression, network delay and packet loss, scaling in temporal and spatial resolution, scaling in bandwidth, spatial and/or temporal interpolation at the receiver, and various types of pre- and post-processing filtering (e.g., error concealment, deblocking filtering, and sharpening). RR IQA provides a useful compromise between FR and NR solutions, where RR features extracted from the original images are transmitted to the receiver end to evaluate the quality of the received distorted images. It was shown that with only a fairly low RR data rate, one may achieve impressive quality prediction accuracy close to competitive FR methods [7].

OBJECTIVE IQA MEASURES ARE PARTICULARLY DESIRABLE IN NETWORK VISUAL COMMUNICATION APPLICATIONS FOR THE PURPOSE OF QoS MONITORING.

The difficulty with RR based methods is how to transmit the RR features to the receiver. This typically requires a guaranteed ancillary channel, which may be expensive or unavailable. An interesting method to trace network image quality degradations without using an ancillary channel is to incorporate modern image watermarking techniques [8]. The idea is to hide a watermark image or a pseudo-random bit sequence inside the image being transmitted. The degradation of the watermark image or the error rate of the embedded bit sequence gauged at the receiver side can then be employed as an indicator of the quality degradation of the host image. The idea of quality-aware image provides another means to incorporate watermarking techniques [7], where RR features extracted from the original image are embedded into the same image as invisible hidden messages. When a distorted



[FIG2] An original image (a) is compressed by JPEG (b). The absolute error map and the SSIM quality map are shown in (c) and (d), respectively. In both maps, brighter indicates better local quality (or lower distortion).

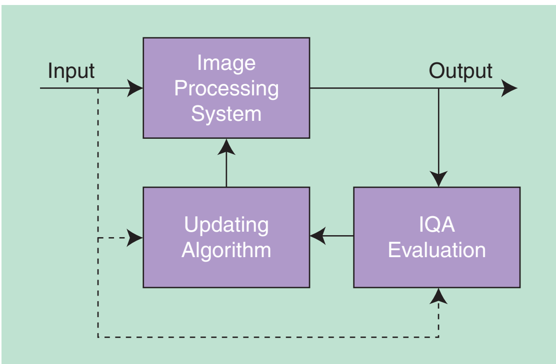
version of such a "quality-aware" image is received, users can decode the hidden RR features and use an RR IQA method to evaluate the quality of the distorted image. The advantages of watermarking-based methods are manifold. First, they do not require a separate data channel to transmit RR features or any other side information to the receiver. Second, they do not affect the conventional usage of the image content, because the data hiding process causes only invisible changes to the image. Third, compared with the approaches of including side information in image headers, they are more likely to survive image/video format conversion [7]. An additional and interesting benefit of quality-aware images is that they provide an opportunity for the end users to partially "repair" the received images using the decoded RR features. Such an idea of self-repairing image was demonstrated in [9] by matching certain statistical properties of the distorted image with those of the reference image (which are received as RR features). It was shown that this approach is quite effective for image deblurring [9].

DESIGN APPLICATIONS

The application scope of objective IQA measures is far beyond quality evaluation and algorithm comparison. In essence, any scientific design of image processing algorithms and systems would involve certain quality criterion, either explicitly or implicitly. If a good quality criterion is available, one may use it not only to assess the performance of these algorithms and systems, but also to optimize them so as to produce the best performance under this criterion.

Figure 2 demonstrates how a perceptual objective IQA measure could be useful in the context of image coding. An original image (a) is compressed using JPEG. Due to a limited bit budget, the resulting decompressed image (b) exhibits many highly visible distortions. In particular, the blocking artifacts in the sky can be clearly seen. The loss of details in the fence areas and the upper boundaries of the building is also obvious. Assume that some new bit budget is now available, and our goal is to decide on how to spend the new bits to enhance the image quality. Ideally, we would spend the bits at the locations that have the greatest potentials to improve the image quality. An IQA measure could assist us in identifying these locations. Figure 2(c) shows the absolute error map between (a) and (b), which is the first step in computing MSE and PSNR (as well as any l_p norm). Unfortunately, this error map provides us with wrong guidance, because it suggests that the inner parts of the building are where the largest distortions are located. By contrast, our visual observations are well consistent with the SSIM map (d) created by a perceptual IQA measure [2]. Realizing that most existing image coders are designed to optimize MSE/PSNR or similar criteria, the dramatic difference between the quality/error maps in (c) and (d) reveals the great potentials of perceptual image and video coding. Some recent work has shown great promises along this direction [4].

In the optimal design of image processing algorithms and systems, objective IQA measures may be employed in two different approaches. In the first approach, the core image processing module is kept unaltered, and the IQA measure is only used to create feedback control signals that help update the image processing module, likely in an iterative manner. This is illustrated in Figure 3, where depending on the application, either FR, RR, or NR IQA measures may be employed to create the feedback control signal. For example, in the case of image enhancement, an NR method may be employed and only the image created at the output end is needed for IQA



[FIG3] Diagram of IQA-based feedback optimization method.

computation. In image coding applications, an FR IQA measure could be used that requires both decoded image from the output end and the original reference image from the input (which is linked through the dashed line).

In the second approach of IQA-based optimal design, the objective IQA measure goes into the core of the image processing algorithm. To illustrate this, let us use the general image reconstruction problem as an example. Assume that there exists an original image X that is unknown to us. What is available is some distorted or partial information Y produced by applying an operator D on X: $Y = \mathrm{D}(X)$. Our goal is to design a reconstruction operator R, which, when applied to Y, yields a reconstructed image $\hat{X} = \mathrm{R}(Y)$, so that $\hat{X}$ is as close to X as possible. Depending on the operator D, this formulation could describe many practical problems. For example, when D denotes noise contamination, then this is a denoising problem. When D represents a downsampling operator, then it corresponds to an interpolation problem. Similarly, the same general framework could cover other problems such as image deblurring, decompression, inpainting, and reconstruction from compressed sensing data. Most of these problems are ill posed, in the sense that the solutions are not unique. To make the problem mathematically sound, one would need to define a cost function as the goal for minimization. For example, in a statistical approach, one treats X as a random variable associated with certain probability distribution and may define the optimization problem as

$$\hat{X}_{\mathrm{opt}} = \min_{\hat{X}} \mathbb{E}\{d(X, \hat{X})|Y\}, \qquad (1)$$

where $\hat{X}_{\mathrm{opt}}$ denotes the optimal solution, $\mathbb{E}$ represents the expectation operator, and d is an image distortion measure. The "standard" option for d is the MSE. To convert this to a perceptual optimization problem is straightforward—replacing d with a monotonically decreasing function with respect to a perceptual IQA measure.

Although the second approach for IQA-based optimal design looks appealing, when it comes to solving the optimization problem in (1), one often faces major difficulties. This is largely due to the lack of desirable mathematical properties in most perceptual IQA measures. To understand this, let us consider why the MSE is still the prevailing optimization criterion, regardless of the wide criticism on its poor correlation with perceptual image quality (as demonstrated in Figure 2). Indeed, the MSE is an ideal target for optimization [1]. It is based on a valid distance metric (l_2) that satisfies positive definiteness, symmetry, and triangular inequality properties. It is convex, differentiable, memoryless, and additive for independent sources of distortions. It is also energy preserving under orthogonal or unitary transformations [1]. Hardly any perceptual IQA measures with good quality prediction performance satisfies any of these properties. In [10], some initial attempts has been made to develop novel image distortion metrics that approximate the SSIM index while maintaining some of the desirable mathematical properties. It was shown that a valid distance metric exists that can very well approximate the SSIM index. In addition, the metric also possesses some useful convexity properties.

WITH THE FAST ADVANCES OF MEDICAL IMAGING TECHNOLOGIES, THE AMOUNT OF MEDICAL IMAGE DATA BEING ACQUIRED EVERY DAY HAS BEEN INCREASING DRAMATICALLY.

EXTENDED APPLICATIONS

In most of the IQA applications we discussed so far, the final outputs are images. Besides these, IQA measures may also be extended to an even broader range of applications where the outputs are interpretations or classification labels of images. Image-based pattern recognition is one such example, where the quality of images is often a critical factor that affects the accuracy of the recognition algorithms. For example, in biometrics, the purpose is to recognize humans or verify human identities based on one or more unique physiological characteristics of humans. Many biometric methods are based on images, including images of faces, fingerprints, palmprints, hand shapes, and handwritings. In practice, the acquisition process of these images may not be perfect, and thus the biometric systems may have to work under the conditions of noisy, distorted, or partially impaired images. In these application scenarios, it would be useful to know the level of quality degradations of these images and what recognition accuracy can be expected under such quality degradations. Different from traditional performance evaluation of IQA measures, here the IQA measures should be assessed and compared based on how they can predict the impact of image quality degradations on the final recognition performance, rather than the perceptual appealingness of the images. Once the image quality is estimated, some preprocessing procedure may be performed to enhance the quality of the image before the pattern recognition algorithm is applied. Another way of using the IQA results is to use them to help select between multiple recognition algorithms or to fuse the results from multiple algorithms, so as to improve the overall recognition performance. Such an approach has been successfully used in fingerprint verification systems [11].

With the fast advances of medical imaging technologies, the amount of medical image data being acquired every day has been increasing dramatically, largely surpassing the increase of available storage space. Efficiently storing, transmitting, and retrieving medical image information in large-scale databases has become a major challenge in hospitals and medical organizations. Lossy image compression provides a powerful means to reduce the data rate, but runs the risk of losing or altering important diagnostic information contained in medical images. It is therefore important to provide specific objective IQA measures that can help maximize the level of compression, but without affecting the diagnostic value of medical images. Moreover, many modern medical imaging devices acquire images with much higher dynamic range of intensity levels than what can be appropriately shown on standard dynamic range displays. Therefore, it is desirable to employ those IQA measures that can provide meaningful quality evaluations of the images after dynamic range compression. Furthermore, both data rate and dynamic range compression of medical images should be optimized for the IQA measures specifically designed for medical applications.

OUTLOOK

We have discussed the application aspects of modern objective IQA methods. Rather than providing an exhaustive survey of all applications, we have emphasized on the great potentials of IQA applications, provided instructive examples, and also discussed the main challenges. In the future, it is expected that the development and application sides of objective IQA measures will mutually benefit each other. On one hand, more accurate and more efficient IQA measures will certainly enhance their applicability in real-world applications. On the other hand, new challenges arising from real applications (e.g., desired mathematical properties for optimization purposes) will impact the new development of future IQA measures.

AUTHOR

Zhou Wang (zhouwang@ieee.org) is an associate professor in the Department of Electrical and Computer Engineering, University of Waterloo, Canada.

REFERENCES

[1] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Processing Mag.,* vol. 26, pp. 98–117, Jan. 2009.

[2] 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.

[3] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Processing,* vol. 15, pp. 430–444, Feb. 2006.

[4] H. R. Wu and K. R. Rao, Ed., *Digital Video Image Quality and Perceptual Coding.* Boca Raton: CRC, 2005.

[5] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Processing,* vol. 15, pp. 3736–3745, Dec. 2006.

[6] X. Zhu and P. Milanfar, "Automatic parameter selection for denoising algorithms using a no-reference measure of image content," *IEEE Trans. Image Processing,* vol. 19, pp. 3116–3132, Dec. 2010.

[7] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E.-H. Yang, and A. C. Bovik, "Quality-aware images," *IEEE Trans. Image Processing,* vol. 15, pp. 1680–1689, June 2006.

[8] M. C. Q. Farias, S. K. Mitra, M. Carli, and A. Neri, "A comparison between an objective quality measure and the mean annoyance values of watermarked videos," in *Proc. IEEE Int. Conf. Image Processing,* Rochester, MN, Sept. 2002, pp. 469–472.

[9] A. Rehman and Z. Wang, "Reduced-reference SSIM estimation," in *Proc. IEEE Int. Conf. Image Processing,* Hong Kong, China, Sept. 2010, pp. 289–292.

[10] D. Brunet, E. R. Vrscay, and Z. Wang, "A class of image metrics based on the structural similarity quality index," in *Proc. Int. Conf. Image Analysis and Recognition (Lect. Notes Comput. Sci.),* vol. 6753, Burnaby, BC, Canada, June 2011, pp. 100–110.

[11] J. Fierrez-Aguilar, Y. Chen, J. Ortega-Garcia, and A. K. Jain, "Incorporating image quality in multi-algorithm fingerprint verification," *Lect. Notes Comput. Sci.,* vol. 3832, pp. 213–220, 2005.

[12] Video Quality Experts Group Web site [Online]. Available: www.vqeg.org

[SP]

 (continued from page 12)

fed into a spatial light modulator, a device that can modulate light spatially in amplitude and phase. "You could get all that information and display it in 3-D and it can actually be in real time," he says.

While most existing telepresence systems are geared toward conferencing applications, Peyghambarian feels that real-time holography has the potential to drive the technology into a wider range of fields. "Benefits include 3-D social networking, 3-D remote surgery, and 3-D collaborative research," he says. "The advantage of our technology is that it can continuously read and replace data, so you can use it in an magnetic resonance imaging or computer assisted tomography scan system that would provide the information it gathered in 3-D to doctors." The technology could, for example, help surgeons performing brain surgery or other types of delicate operations. "They could use that [technology] to see the information as they do the operation," Peyghambarian says.

John Apostolopoulos, director of the Mobile and Immersive Experience Lab at Hewlett-Packard (HP) Laboratories in Palo Alto, California, believes that signal processing will be vital to overcoming many of the challenges telepresence researchers currently face. "This includes video and audio capture, noise reduction, compression, transmission over a packet network, packet-loss concealment, multichannel echo cancellation, efficient signal-processing algorithms for multicore and GPU systems and so on," he says. "I believe that advances in signal processing will continue to be central to improving telepresence in the future."

None of these improvements will come too soon for Microsoft's Zhang, who admits that he has a personal interest in seeing sophisticated telepresence systems becoming commonplace. "I have frequent phone calls with my parents and family members in China as well as my research collaborators at Microsoft Research Asia in Beijing," he says. "Telephony is a great invention, but leaves much more to be desired compared with a face-to-face meeting."

[SP]