Comparison-Based Image Quality Assessment for Selecting Image Restoration Parameters

Haoyi Liang, *Student Member, IEEE*, and Daniel S. Weller, *Member, IEEE*

***Abstract*— Image quality assessment (IQA) is traditionally classified into full-reference (FR) IQA, reduced-reference (RR) IQA, and no-reference (NR) IQA according to the amount of information required from the original image. Although NR-IQA and RR-IQA are widely used in practical applications, room for improvement still remains because of the lack of the reference image. Inspired by the fact that in many applications, such as parameter selection for image restoration algorithms, a series of distorted images are available, the authors propose a novel comparison-based IQA (C-IQA) framework. The new comparison-based framework parallels FR-IQA by requiring two input images and resembles NR-IQA by not using the original image. As a result, the new comparison-based approach has more application scenarios than FR-IQA does, and takes greater advantage of the accessible information than the traditional single-input NR-IQA does. Further, C-IQA is compared with other state-of-the-art NR-IQA methods and another RR-IQA method on two widely used IQA databases. Experimental results show that C-IQA outperforms the other methods for parameter selection, and the parameter trimming framework combined with C-IQA saves the computation of iterative image reconstruction up to 80%.**

***Index Terms*— Image quality assessment (IQA), human visual system (HVS), comparison-based image quality assessment (C-IQA), parameter selection.**

I. INTRODUCTION

OBTAINING an image with high perceptual quality is the ultimate goal of many image processing problems, such as image reconstruction, denoising and inpainting. However, measuring the perceptual image quality by subjective experimentation is time-consuming and expensive, so designing an image quality assessment (IQA) algorithm that agrees with the human visual system (HVS) [1]–[3] is a foundational image processing objective. Moreover, most image restoration algorithms require one or more parameters to regulate the restoration process, and no-reference IQA methods can be used to guide selecting the parameters. For instance, the regularization parameter of image reconstruction [4] is selected by a no-reference image quality index [5]. However, most existing no-reference IQA algorithms output the estimated image quality based on a single distorted image, ignoring that different degraded images can provide more information together to the quality estimation of each degraded image. This observation inspires us to develop a comparison-based IQA method to fill the gap between the increasing need of parameter selection for image processing algorithms and the lack of such an IQA algorithm that makes full use of the available information.

IQA algorithms are classified based on the amount of information from the reference image (the distortion-free image) that is required: full-reference (FR), reduced-reference (RR) and no-reference (NR). FR-IQA [6]–[10] is a relatively well-studied area. Traditional methods like mean squared error (MSE) and signal-to-noise ratio (SNR) are used as the standard signal fidelity indexes [11]. A more sophisticated FR-IQA algorithm, Structural Similarity Index Method (SSIM) [7], considers the structure information in images and performs well in different applications [11]–[14]. RR-IQA algorithms [15]–[18] require some statistical features of the reference image, such as the power spectrum, and measure the similarity of these features from the reference image and the distorted image. NR-IQA algorithms usually adopt two different approaches. The first kind of NR-IQA [19]–[23] algorithms have an approach similar to that of RR-IQA. The difference is that rather than extracting the features from the reference image, this kind of NR-IQA algorithm extracts statistical features from a training set. The second kind of NR-IQA algorithm [5], [14], [24] adopts a local approach to quantifying structure as a surrogate for quality. A common implementation of the second approach is calculating local scores by analyzing the coherence of image gradients. The overall score is synthesized by taking the average of the local scores.

Among these three kinds of IQA algorithms, speed and accuracy generally decrease from FR-IQA, RR-IQA to NR-IQA progressively. However, reference images do not exist in many cases. In applications like parameter selection for image restoration algorithms, a set of distorted images with the same image content are available. A novel comparison-based IQA (C-IQA) framework is proposed in this paper to make full use of these differently distorted images. A parallel two-step framework is adopted in C-IQA. First, a residual image is calculated by taking the difference between two input images, and the quality of the residual image is evaluated. Next, the contribution from two input images to the residual image is calculated. Finally, a simple procedure combines the first two parts: the input image that mainly contributes to high quality residual patches receives positive scores, while the

Manuscript received November 5, 2015; revised May 2, 2016 and June 21, 2016; accepted August 5, 2016. Date of publication August 19, 2016; date of current version September 13, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Weisi Lin.

The authors are with the Charles L. Brown Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904 USA (e-mail: hl2uc@virginia.edu; dweller@virginia.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2016.2601783

input image that is more responsible for the distorted residual patches receives negative scores. Depending on the type of the distortion, different quality indexes, such as the blockiness index [25], can be used in the first part and a multi-metric fusion scheme [26]–[28] can further improve the versatility of the proposed framework.

It is worth differentiating RR-IQA and C-IQA since they both use extra data beyond a single distorted image. The most significant distinction between RR-IQA and C-IQA is that the extra data required by most RR-IQA algorithms is distortion-free, while the extra data that C-IQA has is usually distorted. Therefore, RR-IQA algorithms specifically treat one image as the distorted image to be evaluated and the other information as truth or a reference. However, C-IQA treats the two input images equally without any prior knowledge of the quality of input images. In [18], Wu et al. take RR-IQA as a measurement of the fidelity of the distorted image to the original image. The RR-IQA method proposed in [15] depends on a distortion-free ancillary channel to transmit the features extracted from the original image. In [16], Soundararajan et al. point out that the output of their RR-IQA method is a single positive value that does not indicate if the input is better or worse than the reference image. On the contrary, the final output of our proposed C-IQA is a single real value that indicates which one of the input image is better and by how much. In brief, RR-IQA methods rely on the integrity of the extra data, while C-IQA is able to decide the relative quality only with two input images of the same scene.

The rest of the paper is organized as follows. Section II introduces and compares different NR-IQA methods. Section III elaborates on the details of C-IQA. The algorithm used for image reconstruction and the framework of parameter trimming are introduced in Section IV. In Section V experiments are conducted on two widely used IQA databases, LIVE [29] and CSIQ [6], to verify the performance of C-IQA on parameter selection. Section VI reviews the novelty and experimental results of the proposed C-IQA, and discusses further work on comparison-based IQA framework.

II. Existing NR-IQA Methods

In [14], Shnayderman et al. classify NR-IQA algorithms into two types: global approaches and local approaches. The underlying difference between these two methods are the features used by different NR-IQA algorithms. Statistical features, such as the distribution of wavelet coefficients [19], [20], are extracted for global approaches. Local approaches usually rely on the structure information, such as the edge prominence [5], [14]. Usually, a regression model is adopted to synthesize the statistical features into an overall image quality, while the structure features are able to reflect the image quality directly. The assumption of the statistical feature distribution is changed when considering the difference of two images, while the structure indexes that reflect the coherence of local gradients still work for the difference of two images. Therefore, the proposed C-IQA methods make use of a structure index, MetricQ. In the following part, we briefly review different NR-IQA methods and introduce one particular local structure index MetricQ [5].

A. Approaches With Statistical Features

The rationale of statistical feature-based NR-IQA methods [19]–[23] is that the distributions of natural scene statistics (NSS) share certain common characteristics among distortion-free images, and distortions will change these characteristics. For example, it is widely accepted that the wavelet coefficients of a natural image can be modeled by a generalized Gaussian distribution (GGD) [30], [31].

The main advantage of statistical features is that most of them are not dedicated to a specific distortion since the NSS features are a high-dimensional vector designed to be sensitive to various distortions. However, because of the high dimensionality of the statistical feature space, it is difficult to individually interpret and analyze these features quantitatively, and thus feature selection is largely an empirical work.

B. Approaches With Structure Indexes

Because human eyes are highly sensitive to the gradient in images, and the information in images can be well represented by their gradient [5], [7], [32], structure indexes usually reflect the spatial gradient information. Unlike the statistical indexes, most structure indexes represent the local quality directly without involving the learning process. However, the amount of the gradient, or total variation, itself is not a stable indicator of the quality [14]. Previous works [5], [14], [33] have shown that assessing the concentration of the gradient direction in an image is a promising way to evaluate the image quality. Among these works, MetricQ [5] shows encouraging results choosing denoising parameters. The underlying rationale of MetricQ is that the more concentrated the gradient direction is, the better the quality of the patch is. It is a reasonable assumption since both of the two most common distortions, noise and blurring, disperse the distributions of the gradient direction.

C. MetricQ

The local quality index used by MetricQ is based on singular values of the local gradient matrix, which have been widely used as low level features in different image processing problems, such as tracking feature selection [34], recognition [35] and image quality assessment [14]. For each $n \times n$ local patch (w), the gradient matrix is

$$G = \begin{bmatrix} \vdots & \vdots \\ p_x(k) & p_y(k) \\ \vdots & \vdots \end{bmatrix}, \tag{1}$$

in which $p_x(k)$ and $p_y(k)$ are the gradients of the k^{th} pixel in the patch w on x and y directions. The singular value decomposition (SVD) of the gradient matrix, G, is defined as

$$G = USV^T = U \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix} \begin{bmatrix} V_1 & V_2 \end{bmatrix}^T, \tag{2}$$

where U and V are both orthonormal matrices. Vector V_1 is of size 2×1 and corresponds to the dominant direction of the local gradient; V_2 is orthogonal to V_1 and thus represents

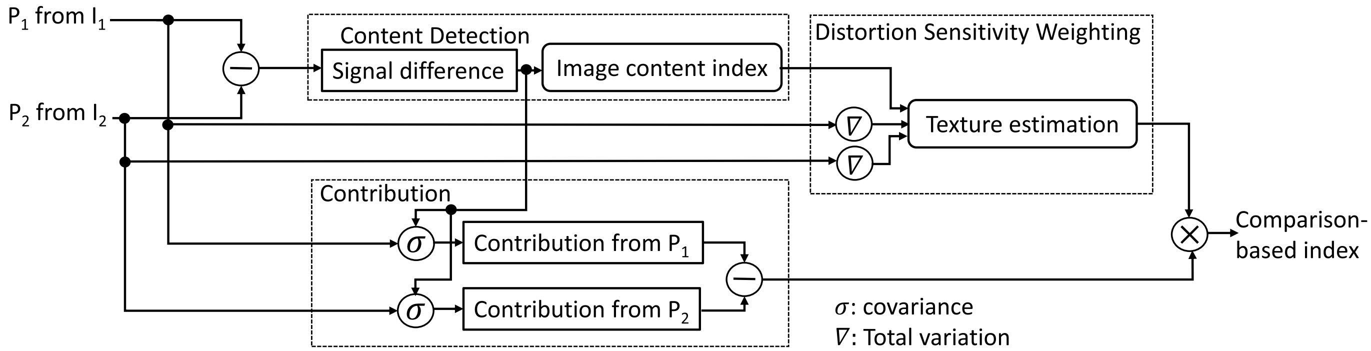


Fig. 1. Flow Chart of the Comparison-based IQA: P_1 and P_2 are local patches from input images, I_1 and I_2, at the same location respectively. The Content Detection module determines whether there is a meaningful structure in the difference patch; the Contribution module calculates which patch mainly contributes to the difference patch; the Distortion Sensitivity Weighting module compensates the distortion sensitivity difference of patches with various texture complexities. The output, comparison-based index, indicates the relative quality of P_1 based on P_2.

the edge direction. Singular values, s_1 and s_2, represent the luminance variances on V_1 and V_2 respectively. Intuitively, a large s_1 and a small s_2 indicate a prominent edge in the local patch.

In MetricQ [5], two indexes reflect the quality of a local patch: Image Content Index and Coherence Index. Image Content Index is defined as

$$Q = s_1 \frac{s_1 - s_2}{s_1 + s_2}, \tag{3}$$

and Coherence Index is defined as

$$R = \frac{s_1 - s_2}{s_1 + s_2}. \tag{4}$$

Q reflects the structure prominence in a local patch and R is used to determine whether a local patch is dominated by noise. The overall score of an image is calculated by

$$AQ = \frac{1}{MN} \sum_{i,j:R(i,j)>\tau} Q(i,j), \tag{5}$$

where $M \times N$ is the size of the image and τ is the threshold to decide whether a local patch is dominated by noise. $Q(i, j)$ and $R(i, j)$ are the Image Content Index and Coherence Index of the local patch centered at (i, j) in the image. A simplified interpretation of (5) is that AQ is the average structure index of local patches that have meaningful image content.

III. Comparison-Based Image Quality Assessment

Previous works on IQA [2], [3], [7], [21], [36] show that IQA performance can be significantly improved by taking advantage of the characteristics of HVS. For example, the structural information that human eyes are highly sensitive to is used by SSIM [7]. Traditional NR-IQA algorithms also try to exploit HVS features and make reasonable assumptions about natural scene images. However, one important aspect of HVS is ignored: comparison. In subjective IQA experiments [6], volunteers are required to evaluate the quality of an image by comparing it with a reference image, rather than giving an absolute score for the image. Although in most image processing applications, the reference image does not exist, a set of differently degraded images are available. In these cases, extending existing state-of-the-art FR-IQA and RR-IQA algorithms to comparison-based IQA algorithms is a natural thought. However, different from FR-IQA and RR-IQA algorithms, neither of the two input image qualities is known in the comparison-based IQA framework. As a result, in a comparison-based IQA algorithm, we not only measure the difference between two input images, but also assess the quality of the difference.

A. Framework of C-IQA

As shown in Fig. 1, C-IQA has two input images, I_1 and I_2, and the output indicates the relative quality of I_1 based on I_2. No prior knowledge about the quality of two input images is known to C-IQA, and the relative quality can be either positive or negative depending on whether I_1 is better than I_2. We refer to the second image in C-IQA as the base image to distinguish it from the reference image in FR-IQA and RR-IQA. The implemented C-IQA method consists of two basic modules: Content Detection and Contribution. The third module, Distortion Sensitivity Weighting, is optional and its description is deferred to Section III-C. In the rest of the paper, we refer to the comparison-based IQA variation composed by the two basic modules as CQ and the variation with three modules as CDQ. Content Detection determines whether the difference between two input images contains any meaningful structure, and Contribution decides which image mainly contributes to the difference. CQ composes these two modules by the criterion that the input image that contributes to a structured difference is better and the input image that contributes to a random difference is worse. The Distortion Sensitivity Weighting module added in CDQ adjusts the distortion sensitivity difference among patches with different texture complexity [7], [37].

1) Content Detection: The Content Detection module is based on the Coherence Index put forward in MetricQ [5]. Different from MetricQ, this index is calculated with the difference image between two input images in C-IQA. In MetricQ, limited by the information provided by a single input image, the algorithm does not know the texture complexity in the original image. Therefore, it is hard for an

Algorithm 1 Content Detection

$D_p = P_1 - P_2$
$G = [d_x(D_p)\ d_y(D_p)]$
$USV^T = SVD(G)$
$C_{ind} = \frac{s_1 - s_2}{s_1 + s_2}$ ▷ $s_1 > s_2$
if $C_{ind} > C_{thresh}$ **then**
  $is\_stru = 1$ ▷ structure
else
  $is\_stru = -1$ ▷ noise
end if

Algorithm 2 Contribution

$D_p = P_1 - P_2$
$M_p = max(\frac{mean(P_1)+mean(P_2)}{2}, \frac{1}{n \times n})$
$ctri1 = cov(P_1, D_p)$
$ctri2 = cov(P_2, -D_p)$
$ctri = \frac{ctri1 - ctri2}{M_p}$

algorithm to estimate how concentrated the gradient should be. However, by mimicking the comparative way HVS works, C-IQA removes the main image content in the images by taking the difference, and thus it is easy for the Content Detection module to differentiate the patches with noisy and structured content.

In Alg. 1, P_1 and P_2 are two patches of size $n \times n$ from I_1 and I_2 respectively, G is the same 2-column gradient matrix defined in (1), $SVD(G)$ represents taking the SVD operation on G, and s_1 and s_2 are the singular values of G. C_{thresh} is a constant threshold to binarize C_{ind}. The binary output is_stru indicates whether there is a meaningful structure in the difference of local patches.

2) Contribution: Once the difference is classified into noise or structure, the Contribution module is designed to find out which of the two input images mainly contributes to the difference image. In our implementation, the luminance-normalized covariance between the input image and the difference image is used to measure the contribution.

In Alg. 2, $mean(P_i)$ calculates the average of the local patch, and $cov(x_1, x_2)$ calculates the covariance between two input patches,

$$cov(x_1, x_2) = \frac{(x_1 - mean(x_1))^T (x_2 - mean(x_2))}{n^2 - 1},$$

x_1 and x_2 are vectorized patches of size $n^2 \times 1$. The output $ctri$ represents that P_1 contribute to D_p more than P_2 does by how much. A negative $ctri$ means that P_2 mainly contribute to D_p.

The comparative quality index for each local patch is calculated by

$$C_Q = is\_stru \cdot ctri.$$

The overall comparative quality of I_1 based on I_2 is

$$CQ(I_1, I_2) = \frac{1}{M \times N} \sum_{i,j=(n/2):(M-n/2)} C_Q(i, j),$$

where $C_Q(i, j)$ is the local comparative quality index centered at (i, j) in the image, $n \times n$ is the size of the local patch and $M \times N$ is the size of the image. Patches that are outside the boundary of the image are not included in the calculation. A positive $CQ(I_1, I_2)$ means I_1 is better than I_2, and the absolute value quantifies the quality difference. Due to the anti-symmetric design of the algorithm, $CQ(I_1, I_2) = -CQ(I_2, I_1)$.

B. Justification of CQ

Inspired by Li's work [38] which claims that an IQA model should be based on three quantities: edge sharpness, random noise level and structure noise, we classify the distortions by residual images, the difference between a distorted image and the original image. In our classification, distortions can be categorized into two types: introducing a random residual image, or introducing a structured residual image. In most cases, random residual images correspond to noise-like distortions and structured residual images correspond to blurring-like distortions. In this part, we prove how C-IQA works under these two distortions.

Assume I_{true} is the original image, and I_1, I_2 are two distorted images. The residual images are calculated by,

$$e_i = I_i - I_{true}, \quad i = 1, 2.$$

Similarly, for each patch we have

$$e_{Pi} = P_i - P_{true}, \quad i = 1, 2.$$

1) Random Residual Image: Residual images behave like noise in this case. If we assume I_1 is more severely distorted than I_2, then we have $E[\|e_{P1}\|_2^2] > E[\|e_{P2}\|_2^2]$. The expectation of the local comparative quality index is

$$\begin{aligned}
E[C_Q] &= E[ctri \cdot is_stru] \\
&= E[(ctri1 - ctri2) \cdot is_stru] \\
&= E[cov(P_1, P_1 - P_2) - cov(P_2, P_2 - P_1)] \\
&\quad \cdot E[is_stru] \\
&= -E[cov(P_{true} + e_{P1}, e_{P1} - e_{P2}) \\
&\quad - cov(P_{true} + e_{P2}, e_{P2} - e_{P1})] \\
&= -E[2 \cdot cov(P_{true}, e_{P1} - e_{P2}) \\
&\quad + cov(e_{P1}, e_{P1}) - cov(e_{P2}, e_{P2})] \\
&= -E[cov(e_{P1}, e_{P1})] + E[cov(e_{P2}, e_{P2})] \\
&< 0.
\end{aligned}$$

The three most important properties in the derivation are the irrelevance between P_{true} and e_{Pi}, the randomness of e_{Pi}, and independence of is_stru and $ctri1$, $ctri2$. The result $E[C_Q] < 0$ agrees with our assumption that I_1 is more severely distorted than I_2. When I_2 is more severely distorted, the same proof shows $E[C_Q] > 0$.

2) Structured Residual Image: If the residual images show structured information, the most probable reason is that the image is distorted by a blurring-like distortion. Because the blurring filter acts as a low-pass filter, the residual images show a structure that is inversely related to the original image [39] to smoothen the high contrast on the edges.

Without loss of generality, we assume more blurring happens in I_1 than I_2, which means $E[|e_{P1}|] > E[|e_{P2}|]$.

The expectation of the local comparative quality index is

$$\begin{aligned} E[C_Q] &= E[ctri \cdot is\_stru] \\ &= E[(ctri1 - ctri2) \cdot is\_stru] \\ &= E[cov(P_1, P_1 - P_2) - cov(P_2, P_2 - P_1)] \\ &= E[cov(P_{true} + e_{P1}, e_{P1} - e_{P2}) \\ &\quad - cov(P_{true} + e_{P2}, e_{P2} - e_{P1})] \\ &= E[cov(2 \cdot P_{true}, e_{P1} - e_{P2}) \\ &\quad + cov(e_{P1} + e_{P2}, e_{P1} - e_{P2})] \\ &= E[cov(2 \cdot P_{true} + e_{P1} + e_{P2}, e_{P1} - e_{P2})] \\ &< 0. \end{aligned}$$

The most important step in this derivation is the last step. Since $E[|e_{P1}|] > E[|e_{P2}|]$, $e_{P1} - e_{P2}$ also demonstrates a structure that is inversely related to the original image as e_{Pi}. As long as the distortion is not severe enough to remove the structure in the original image, $2 \cdot P_{true} + e_{P1} + e_{P2} = P_1 + P_2$ is positively related to the original image. As a result, $E[cov(2 \cdot P_{true} + e_{P1} + e_{P2}, e_{P1} - e_{P2})] < 0$, which agrees with our assumption that I_1 is more severely distorted than I_2. Following the same steps, we can show $E[C_Q] > 0$ if I_2 is more severe distorted than I_1.

C. Distortion Sensitivity Weighting

We have proven that only with Content Detection and Contribution, the CQ can give correct results if both of the two input images are distorted by one distortion, either noise-like distortion or blurring-like distortion. However, another important property of HVS is missed in CQ: the response of HVS to the same distortion is texture-dependent. One example of this HVS property is that after being distorted by the same amount of Gaussian noise, the distortion in the image with simpler texture is more obvious. In this part, we first investigate such texture-based response of CQ and then design a weighting module to adjust the distortion sensitivity of CQ to different textures. We refer to the improved C-IQA method with Distortion Sensitivity Weighting module as CDQ.

In CQ, Content Detection is a qualitative module that detects the meaningful structure and the Contribution module quantifies the relative quality. Therefore, the Contribution module may implicitly include distortion sensitivity weighting. We design an experiment to explore the relation between the texture complexity and the output of Contribution. In this experiment, 140 patches of size 101×101 with homogeneous texture are selected from LIVE [29] and CSIQ [6], and eight samples of these patches are shown in Fig. 2. As the representatives of blurring-like and noise-like distortions, a bilateral filter and Gaussian noise with the same parameters are applied to each patch. According to the Weber-Fechner law [40], we use luminance-normalized total variation as the perceived texture complexity, $T\_ind = \frac{TV(P)}{mean(P)}$, where $TV(P)$ is the total variation in the original patch and $mean(P)$ is the average of the original patch. The relation between texture complexity, T_ind, and the output of Contribution module, $ctri$, are plotted in Fig. 3. Each circle in Fig. 3 represents a patch sample. It is clear that $ctri$ is almost linear

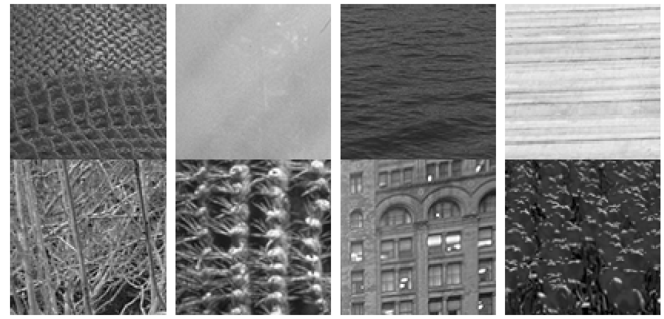
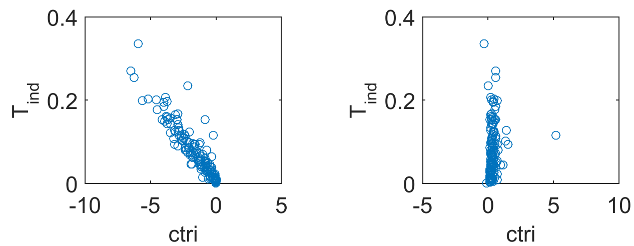


Fig. 2. Patch samples with various texture complexity are selected from LIVE [29] and CSIQ [6] to verify the difference of distortion sensitivity in CQ. The assumption is that patches with complex texture are more robust to noise, while patches with flat texture are more robust to blurring.



(a) (b)

Fig. 3. Relations between the output of Contribution module, $ctri$, and texture complexity, T_ind. Each circle in the figure represents a sample patch. All the sample patches are degraded by the same amount of distortion for blurring and noise.

Algorithm 3 Distortion Sensitivity Weighting

$T1\_ind = \frac{TV(P_1)}{mean(P_1)}$
$T2\_ind = \frac{TV(P_2)}{mean(P_2)}$
if is_stru = 1 **then**
    $T\_ind = max\{T1\_ind, T2\_ind\}$;
else
    $T\_ind = min\{T1\_ind, T2\_ind\}$;
end if
$S\_ind = log(1 + \frac{1}{C_1 \times T\_ind})$
if is_stru = 1 **then**
    $weight = 1$
else
    $weight = -S\_ind$
end if

related to texture complexity, T_ind, when blurring happens. On the contrary, T_ind shows no relation with $ctri$ when the distortion is noise. The reason for this is that blurring is a highly image-dependent distortion, and the residual image is more prominent at areas where total variation is high. After figuring out the blurring sensitivity compensation mechanism in CQ, we need to design an algorithm to compensate the sensitivity difference to noise.

Because noise-like distortion tends to increase the total variation while blurring-like distortion tends to decrease the total variation, Alg. 3 uses the output of Content Detection to synthesize $T1\_ind$ and $T2\_ind$ into T_ind. After texture complexity estimation, we transfer T_ind to the smoothness index, S_{ind}, and compensate the sensitivity to noise.

In CDQ, the comparative quality index for each local patch is

$$CD_Q = is\_stru \cdot ctri \cdot weight.$$

The overall comparative quality of I_1 based on I_2 is calculated by taking the average of local comparative quality index as CQ does.

D. Comparison Between CDQ and SSIM

SSIM consists of three components: structure (loss of correlation), luminance (mean distortion) and contrast (variance distortion). In CDQ, the outputs of Content Detection and Distortion Sensitivity Weighting provide the quality of the difference image. The luminance and the contrast of an input image together determine the contribution of the input image to the difference image. Therefore, Content Detection and Distortion Sensitivity Weighting of CDQ together play the role of the structure part in SSIM. The difference is that without knowing which image has the better quality, CDQ has to analyze the quality of the structure in the difference image, rather than only measuring the structure distance as SSIM does. The Contribution module in CDQ is similar to the functions of luminance and contrast parts together in SSIM.

IV. Parameter Selection

As the motivation of C-IQA mentioned in the introduction, most image processing algorithms contain user-defined parameters (these image processing algorithms are referred as "target algorithms" in the following to differ from IQA algorithms). Parameter selection [4], [5], [41]–[47] is of importance to these target algorithms. By parameter selection, some of these target algorithms [45], [46] achieve a faster convergence rate; some [43], [44] obtain a better restored image.

In this section, we first introduce an image reconstruction algorithm and illustrate the importance of parameter selection with this reconstruction algorithm. Next, a boosted parameter selection framework for iterative image processing algorithm, parameter trimming [4], is introduced. In the following experimental section, we show target algorithms with the parameter trimming framework benefit from the parameters selected by CDQ.

A. Image Reconstruction

Total variation (TV) reconstruction [48] is aimed at minimizing the cost function,

$$E_\beta(x) = \beta\|Dx\|_1 + \frac{1}{2}\|Sx - y\|_2^2, \tag{6}$$

where x is the reconstructed image, y is the observed incomplete data set, S is the system matrix, D represents the difference matrix, and the TV regularizer $\|Dx\|_1$ combines gradients on two directions isotropically. In our implementation, $S = R\mathcal{F}$, where R represents the subsampling matrix and $\mathcal{F}$ represents the Fourier transform matrix. The regularization parameter β controls the sharpness of the reconstructed result. Large β oversmooths the reconstructed image, while small β leaves residual noise. A proper β is crucial to the performance of TV reconstruction. Split Bregman iteration [49] is used to solve (6). By making the replacement $d \leftarrow Dx$ and introducing the dual variable b, the split formulation of (6) becomes:

$$\min_{x,d} \beta\|d\|_1 + \frac{1}{2}\|Sy - y\|_2^2 + \frac{\mu}{2}\|d - Dx - b\|_2^2,$$
$$s.t.\ d = Dx. \tag{7}$$

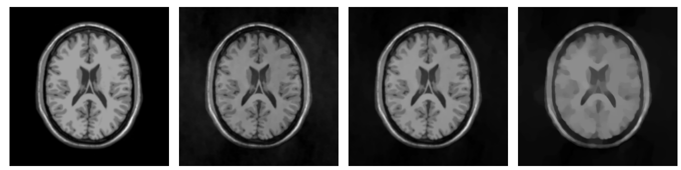
The Split Bregman iteration solution to (7) is Alg. 4. In Alg. 4 we use the notation $K = (R^T R - \mu\mathcal{F}D^T D\mathcal{F}^{-1})$, $L_k = (\mathcal{F}^T R^T y + \mu D^T(d^k - b^k))$ and $s^k = \sqrt{|Dx^k + b^k|^2}$. μ is set as 0.01β to ensure a fast convergence rate.

Algorithm 4 Split Bregman

Initialize: $x^0 = 0, d^0 = b^0 = 0$
while stop criterion is not satisfied **do**

$$x^{k+1} = \mathcal{F}^{-1}K^{-1}L_k$$
$$d_{k+1} = \max(s^k - \frac{1}{\mu}, 0)\frac{Dx^k + b^k}{s^k}$$
$$b^{k+1} = b^k + (Dx^k - d^{k+1})$$

end while



(a) (b) (c) (d)

Fig. 4. (a): original Brain image [50]; (b): reconstructed result with $\beta = 1.22 \times 10^{-6}$; (c): reconstruction result with $\beta = 4.46 \times 10^{-1}$; (d): reconstructed result with $\beta = 10$.

To illustrate the necessity of parameter selection of TV reconstruction, the Brain image [50] is reconstructed with 30 values of β. These candidate values of β are uniformly sampled from 1.22×10^{-6} to 10 in logarithmic scale and three of the reconstructed results are shown in Fig. 4. The image quality indexes during the convergence process are plotted in Fig. 5(a) where each line corresponds to one parameter candidate. The final reconstructed image qualities are plotted in Fig. 5(b). From Fig. 5, it is clear that if parameters that do not have the potential to achieve good results are terminated before convergence, considerable computation will be saved. This is the intuition of the parameter trimming in the next section.

B. Parameter Trimming

A traditional approach to parameter selection [41]–[44] is selecting the parameters after the convergence of all the target algorithm instances. However, since either the target

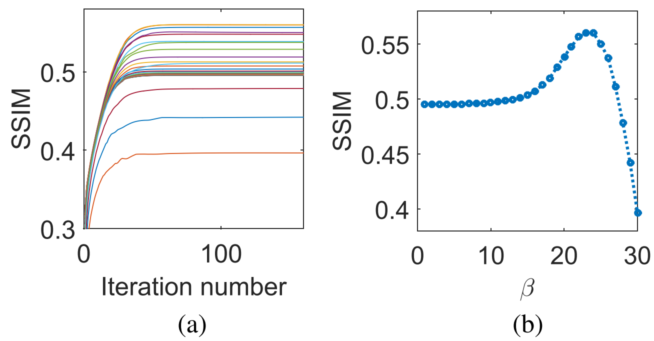


Fig. 5. (a): Each line corresponds to an algorithm instance with a different regularization parameter. (b): Qualities of reconstructed results with different regularization parameters after 160 iterations.

Fig. 6. Images in the gray scale [29] for the illustration of minimum resolution. (a) "caps" (b) "coinsinfountain".

algorithms converge quickly [5], [44] or the NR-IQA algorithm is time-consuming [43], computational efficiency is not considered in previous works.

In situations where target algorithms converge slowly or the set of parameter candidates is large, assessing image qualities and selecting the best parameter after all the algorithm instances converge would be too time-consuming to be practical. Instead of placing the quality monitor at the output end, we first proposed a parameter trimming framework [4] that integrates the quality monitor into the target algorithms. In this section, we use image reconstruction as the application to illustrate the parameter trimming framework.

Assume I_m^i is the reconstructed result of the m^{th} parameter candidate at the i^{th} iteration. The trimming decision is made based on three indexes, q_m^i, g_m^i and p_m^i, which are the reconstructed quality, the quality increasing gradient and the prediction of the quality of I_m^i respectively. Because the image quality index we use here is a comparison-based index, the definitions of the these three indexes are modified to fit CDQ into the parameter trimming framework in [4]. Denoting the best reconstructed result at the i^{th} iteration is $best_i$, it satisfies $CDQ(I^i_{best_i}, I^i_{best_i-1}) \geq 0$ and $CDQ(I^i_{best_i}, I^i_{best_i+1}) \geq 0$. The three indexes used for parameter trimming, q_m^i, g_m^i and p_m^i, are defined as,

$$\begin{aligned} q_m^i &= CDQ(I_m^i, I^i_{best_i}), \\ g_m^i &= CDQ(I_m^i, I^{i-1}_{best_{i-1}}) - CDQ(I_m^{i-1}, I^{i-1}_{best_{i-1}}), \\ p_m^i &= q_m^i + pre_{len} \cdot g_m^i. \end{aligned}$$

We set $pre_{len} = 4$ in all the experiments. More examples of the reconstruction and trimming process are shown in Section V-C.

V. Experiments

We first introduce a key property, minimum resolution, that is unique to C-IQA in Section V-A. In the next two parts, more comprehensive experiments are conducted to verify the effectiveness of C-IQA for parameter selection. The other NR-IQA algorithms that we use to compare CQ/CDQ with are DIIVINE (DII) [19], BRISQUE (BRI) [20], MetricQ (MQ) [5] and Anisotropy (Ani) [21]. Although RR-IQA methods are not suitable for parameter selection where the original image is not available, we include one RR-IQA [18] method to compare with C-IQA. The reason is that for some applications, such as delivering and decompressing visual data sent to networked devices [20], both C-IQA and RR-IQA are practical. In [18], Wu. et al proposed a RR-IQA method that uses two numbers containing the information of the order and disorder parts of a reference image to help evaluate the quality of the distorted image. A widely accepted FR-IQA algorithm, SSIM [7], is used as the ground truth to evaluate the performance of other IQA algorithms. Two IQA databases used in the experiments are LIVE [29] and CISQ [6]. Parameters in CQ/CDQ are set as $C_{thresh} = 0.12$, $C_1 = 4.6$ and the size of a local patch is 9×9 pixels.

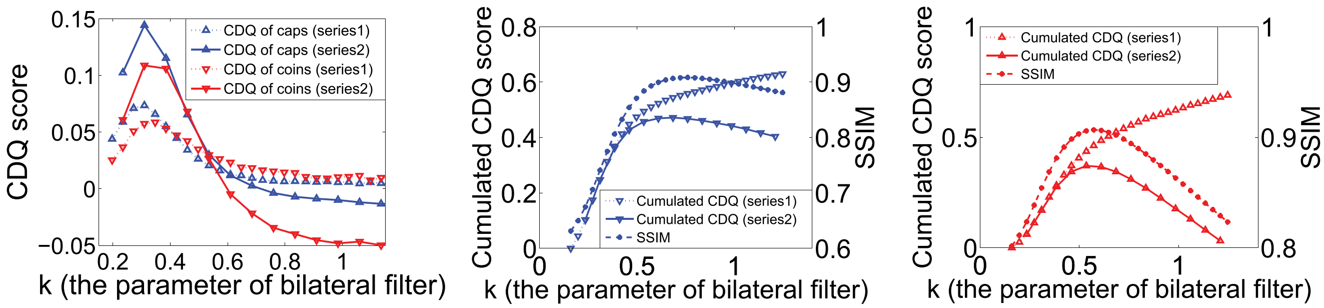
A. Minimum Resolution of C-IQA

Since the comparison-based IQA is a brand-new approach, new properties arise. In this section, we illustrate the minimum resolution of C-IQA and corresponding solutions based on two images from LIVE [29] as shown in Fig. 6.

Similar to HVS, IQA algorithms are not able to make a convincing quality comparison between images whose difference is sufficiently small. In this part, we define the minimum mean squared difference (MSD) between two images required to make a convincing quality comparison as the minimum resolution. It is worth noticing that minimum resolutions vary over different distortions and different IQA algorithms.

For the traditional single-image-input NR-IQA algorithms, minimum resolutions can be regarded as the minimum MSD required to ensure consistency on a series of increasingly distorted images. However, under the comparison-based framework, a distorted image has different scores compared with different base images. We cannot refer to the consistency to define the minimum resolution for a comparison-based IQA algorithm. The minimum resolution for comparison-based IQA is defined as the minimum MSD required to preserve transitivity among a series of increasingly distorted images. We conduct an experiment on the images in Fig. 6 to demonstrate the transitivity.

Assume I_{org} is the original image, and I_1 is created by adding Gaussian noise to I_{org}. A series of gradually filtered images, $(I_1, I_2, \cdots, I_N)$, are denoised by bilateral filters [51], $BF_{(r,d)}$, where r and d are the variances of Gaussian range kernel for smoothing differences in intensities and Gaussian



(a) (b) (c)

Fig. 7. Minimum resolution of Comparison-based IQA algorithm: (a) CDQ scores of denoised images compared with their previous images (series1) and the one before previous images (series2); (b) SSIM scores and cumulated CDQ scores of "caps" in (a); (c) SSIM scores and cumulated CDQ scores of "coinsinfountain" in (a).

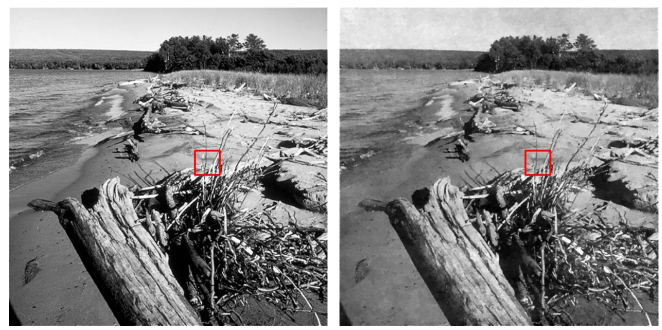
spatial kernel for smoothing differences in coordinates. For simplicity, we reduce the parameters of bilateral filters to one by fixing the ratio between r and d, $BF_k = BF_{(0.1k,3k)}$. In Fig. 7(a), we show the CDQ scores of each image compared with its previous one in the denoised sequence ($series1$) and the CDQ scores compared with the one before its previous one ($series2$). We can see that CDQ scores in $series1$ are always positive, but pass 0 in $series2$. Therefore, the denoised image qualities monotonically increases in $series1$, but reach a peak in $series2$. In Fig. 7(b) and Fig. 7(c), we plot the cumulated CDQ scores in $series1$ and $series2$. It is clear that the cumulated CDQ scores fail to characterize the trend of image quality in $series1$, but successfully reflect the peak in $series2$. In this example, the MSD between adjacent images in $series1$ is below the minimum resolution of the bilateral filter, but the MSD between adjacent images in $series2$ is above the minimum resolution of the bilateral filter.

There are two ways to avoid the unwanted result of operating below minimum resolution. First, increase the MSD between adjacent images by increasing the parameter steps. Second, avoid comparing the adjacent images in a series of increasingly distorted images. The Key Image algorithm introduced in the Section V-B2 is an implementation of the second approach.

B. Experiment Verification for Parameter Selection

Because the main motivation of Comparison-based IQA is parameter selection for image processing algorithms, two common image processing problems, image reconstruction and image denoising are used to demonstrate the parameter selection ability of the proposed C-IQA. The algorithm used for image reconstruction is the Split Bregman approach to total variation reconstruction [49]; the algorithm used for image denoising is the bilateral filter [51]. The optimal parameters of these two algorithms on different images are selected by SSIM, different NR-IQA algorithms, RR-IQA and C-IQA. The parameters selected by SSIM are compared with the ones selected by other IQA algorithms to evaluate the performance of other IQA algorithms.

1) Parameter Selection for TV Reconstruction: The algorithm used for image reconstruction is introduced in Section IV-A. In this experiment, 70% Fourier transform data

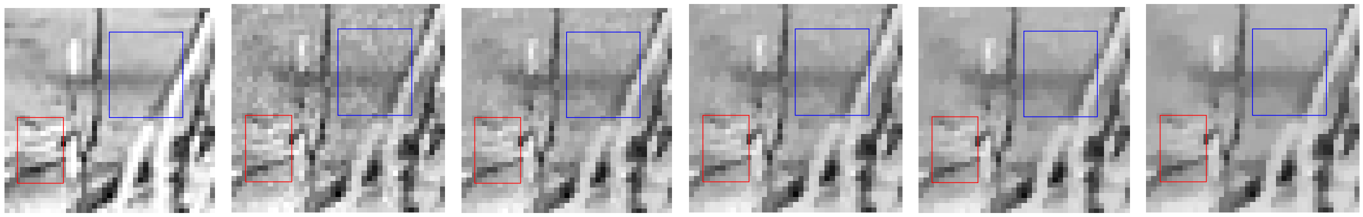


(a) (b)

Fig. 8. One example of image reconstruction parameter selection. The best regularization parameter for this image, $\beta^* = 2.81 \times 10^{-2}$. The area highlighted by the red box is enlarged in Fig. 9. (a) Original "log_seaside" for CSIQ. (b) Reconstructed "log_seaside" with β^*.

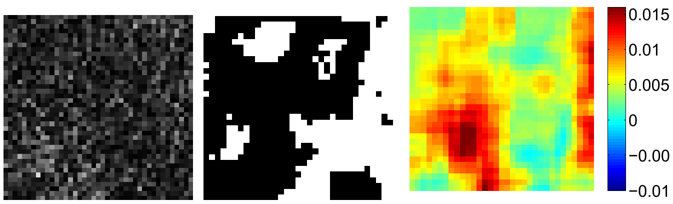
are used to reconstruct the image and in order to be more realistic, Fourier transform data are distorted by Gaussian noise. The SNR is kept at 20 dB in all reconstruction experiments. All 30 regularization parameter candidates are uniformly selected between $[10^{-5}, 10^{-1}]$ in logarithmic scale.

One reconstruction example, "log_seaside", from CSIQ is shown in Fig. 8. The highlighted area in Fig. 8 is shown in details in Fig. 9 for the original image and reconstructed results with different regularization parameters, β. The RR-IQA method [18] selects $\beta_1 = 8.53 \times 10^{-4}$ as the optimal parameters; DIIVINE and BRISQUE select $\beta_2 = 1.49 \times 10^{-2}$; Anisotropy selects $\beta_3 = 2.04 \times 10^{-2}$; SSIM, CQ and CDQ select $\beta^* = 2.81 \times 10^{-2}$; MetricQ selects $\beta_4 = 3.86 \times 10^{-2}$. It is clear from Fig. 9 that as β increases, noisy component disappears and blurring occurs. Fig. 10 and Fig. 11 show how CQ works by comparing different reconstructed results. In Fig. 10, the reconstructed result of β_1, Fig. 9 (b), is compared with the result with optimal parameter, β^*, Fig. 9 (e). From Fig. 10 (a), it is clear that the difference patch shows a noise pattern. The black areas in Fig. 10 (b) indicate the areas that are likely to be taken as noise. Fig. 10 (c) shows the contribution difference from reconstructed results with β_1 and β^* to Fig. 10 (a). It should be noticed that the contribution difference in white areas in Fig. 10 (b) tends to be assigned a much smaller absolute



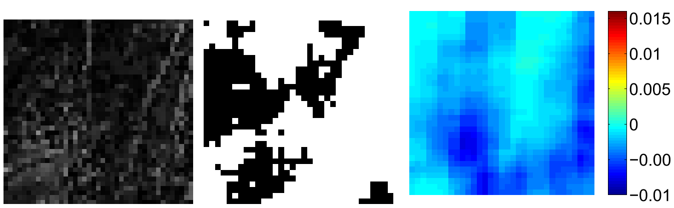
(a) (b) (c) (d) (e) (f)

Fig. 9. Patches from the highlighted areas in Fig.8. The regularization parameter of the total variation term is β. RR-IQA [18] selects β_1, DIIVINE and BRISQUE select β_2, Anisotropy selects β_3, CQ and CDQ select β^*, and MetricQ select β_4. As β increases, noise is suppressed as shown in blue rectangles, while subtle structures is blurred as shown in red rectangles. (a) Patches from the original image. (b) Reconstructed result with $\beta_1 = 8.53 \times 10^{-4}$. (c) Reconstructed result with $\beta_2 = 1.49 \times 10^{-2}$. (d) Reconstructed result with $\beta_3 = 2.04 \times 10^{-2}$. (e) Reconstructed result with $\beta^* = 2.81 \times 10^{-2}$. (f) Reconstructed result with $\beta_4 = 3.86 \times 10^{-2}$.



(a) (b) (c)

Fig. 10. CQ index of Fig. 9 (b) (β_1) based on Fig. 9 (e) (β^*). (a) is the difference between the image with β_1 and β^*. The white in (b) stands for structured areas in (a) and the black stands for noisy area. (c) shows the contribution difference between β_1 and β^*. On this local patch, $CQ(P_{\beta_1}, P_{\beta^*}) = -1.83 \times 10^{-3}$.

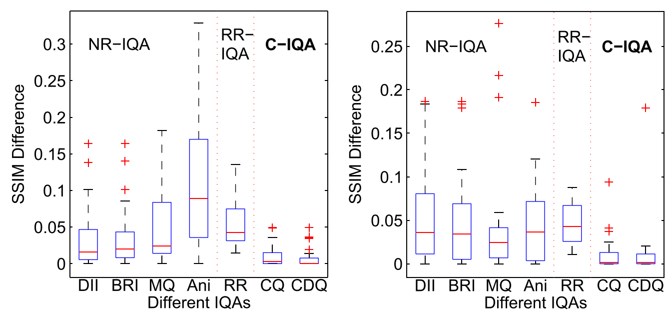


(a) (b) (c)

Fig. 11. CQ index of Fig. 9 (f) (β_4) based on Fig. 9 (e) (β^*). (a) is the patch difference between the image with β_4 and β^*. The white areas in (b) corresponds to structured area. In (c), the negative values means that the contribution comes from β^*. On this local patch, $CQ(P_{\beta_4}, P_{\beta^*}) = -2.50 \times 10^{-4}$.

value. Therefore, the CQ index of Fig. 9 (b) based on Fig. 9 (e) is a negative number that indicates Fig. 9 (b) is worse. Similarly, the comparison between reconstructed results with β_4 and β^* are shown in Fig. 11. A clear structured difference in Fig. 11 (a) is supported by the majority white area in Fig. 11 (b). The negative value in Fig. 11 (c) means that the contribution to the structured difference comes from β^*. It is worth to notice that although both the comparisons between results of β_1 and β^*, and β_4 and β^* lead to negative values that show β^* is better, the decision-making processes are different. When comparing β_1 and β^*, the difference is noisy and mainly comes from β_1; while when comparing β_4 and β^*, the difference is structured and mainly comes from β^*.

In the datasets of LIVE and CSIQ, there are 59 original images and each original image corresponds to 30 reconstructed results with different regularization parameters.



(a) (b)

Fig. 12. The SSIM differences of the best images chosen by different NR, RR, and Comparison-based IQA methods for image reconstruction parameter selection. (a) LIVE. (b) CSIQ.

The SSIM index difference between the best images chosen by SSIM and the one chosen by other IQA algorithms is used to evaluate other IQA methods in this experiment. The SSIM difference of each IQA algorithm is plotted in Fig. 12. More quantitative evaluation of different IQA algorithms are provided in Table I. Both the results in Fig. 12 and Table I show that the comparison-based methods, CQ and CDQ, have the best accuracy of selecting reconstruction parameters.

2) Parameter Selection for Bilateral Filter: A series of increasingly denoised images, $I_1, I_2, \cdots, I_{30}$, are created for each image the same as Section V-A. The SSIM index of the most oversmoothed image I_{30} is between 0.85 ± 0.01.

Because the MSD between the adjacent images are below minimum resolution of bilateral filtering, Alg. 5 is adopted to select the best result. Key images are a set of images among which the MSD is greater than the minimum resolution. Alg. 5 first separates the 30 increasingly denoised images into a few segments by key images and selected the best key image with CQ/CDQ. The MSD difference between key images are lower bounded by K_{thresh}. Next, images in the two segments that are adjacent to the best key image are evaluated based on the two key images on the ends. By doing so, we avoid comparing the adjacent images directly. K_{thresh} is set as 3.0 in this experiment.

The SSIM difference between the best denoised image chosen by SSIM and the one chosen by other IQA methods is plotted in Fig. 13. Table II shows more quantitative results

TABLE I

ACCURACY OF PARAMETER SELECTION FOR IMAGE RECONSTRUCTION

		DIIVINE	BRISQUE	MetricQ	Anisotropy	RR-IQA	**CQ**	**CDQ**
LIVE	median of all SSIM differences	1.59×10^{-2}	2.03×10^{-2}	2.42×10^{-2}	8.88×10^{-2}	4.28×10^{-2}	2.97×10^{-3}	$\mathbf{0}$
	average of all SSIM difference	3.45×10^{-2}	3.57×10^{-2}	5.07×10^{-2}	1.09×10^{-1}	5.45×10^{-2}	9.91×10^{-3}	$\mathbf{7.76\times10^{-3}}$
	average of non-outliers	2.59×10^{-2}	2.42×10^{-2}	5.07×10^{-2}	1.09×10^{-1}	5.45×10^{-2}	7.02×10^{-3}	$\mathbf{2.07\times10^{-3}}$
CSIQ	median of all SSIM difference	3.63×10^{-2}	3.43×10^{-2}	2.44×10^{-2}	3.66×10^{-2}	4.30×10^{-2}	$\mathbf{1.73\times10^{-3}}$	$\mathbf{1.73\times10^{-3}}$
	average of all SSIM difference	5.19×10^{-2}	4.97×10^{-2}	4.25×10^{-2}	4.30×10^{-2}	4.77×10^{-2}	$\mathbf{1.11\times10^{-2}}$	1.12×10^{-2}
	average of non-outliers	4.72×10^{-2}	3.49×10^{-2}	2.19×10^{-2}	3.81×10^{-2}	4.77×10^{-2}	6.02×10^{-3}	$\mathbf{5.38\times10^{-3}}$

TABLE II

ACCURACY OF PARAMETER SELECTION FOR BILATERAL FILTER

		DIIVINE	BRISQUE	MetricQ	Anisotropy	RR-IQA	**CQ**	**CDQ**
LIVE	median of all SSIM differences	1.65×10^{-2}	2.89×10^{-2}	$\mathbf{2.36\times10^{-3}}$	6.95×10^{-2}	7.98×10^{-3}	7.67×10^{-3}	3.80×10^{-3}
	average of all SSIM differences	2.53×10^{-2}	3.31×10^{-2}	6.73×10^{-3}	8.23×10^{-2}	1.07×10^{-2}	1.43×10^{-2}	$\mathbf{6.05\times10^{-3}}$
	average of non-outliers	2.53×10^{-2}	3.31×10^{-2}	4.63×10^{-3}	8.23×10^{-2}	1.07×10^{-2}	1.12×10^{-2}	$\mathbf{3.93\times10^{-3}}$
CSIQ	median of all SSIM differences	1.59×10^{-2}	3.84×10^{-2}	$\mathbf{2.83\times10^{-3}}$	2.18×10^{-2}	7.22×10^{-3}	5.28×10^{-3}	4.05×10^{-3}
	average of all SSIM differences	3.40×10^{-2}	3.86×10^{-2}	7.65×10^{-3}	4.30×10^{-2}	2.41×10^{-2}	1.09×10^{-2}	$\mathbf{6.24\times10^{-3}}$
	average of non-outliers	3.40×10^{-2}	3.86×10^{-2}	6.34×10^{-3}	3.17×10^{-2}	9.75×10^{-3}	6.36×10^{-3}	$\mathbf{5.32\times10^{-3}}$

Algorithm 5 Key Image

Key Images Selection;

$key\_img = \{1\}$
$key_{num} = 1$
for $i = 1 : N$ **do**
  if $MSD(I_i, I_{pre_{key}(key_{num})}) > K_{thresh}$ **then**
    $key\_img = \{key\_img, i\}$
    $key_{num} = key_{num} + 1$
  end if
end for

Key Images Comparision;

for $i = 2 : (key_{num} - 1)$ **do**
  if $C\text{-}IQA(I_{key\_img(i)}, I_{key\_img(i-1)}) > 0$ and $C\text{-}IQA(I_{key\_img(i)}, I_{key\_img(i+1)}) > 0$ **then**
    $best_{key} = i$
    break;
  end if
end for

Best Image Selection;

$start_{num} = key\_img(best_{key} - 1)$
$end_{num} = key\_img(best_{key} + 1)$
for $i = start_{num} : end_{num}$ **do**
  $score_{start}(i) = C\text{-}IQA(I(i), I(start_{num}))$
  $score_{end}(i) = C\text{-}IQA(I(i), I(end_{num}))$
end for
$best_{img} = max(score_{start} + score_{end})$

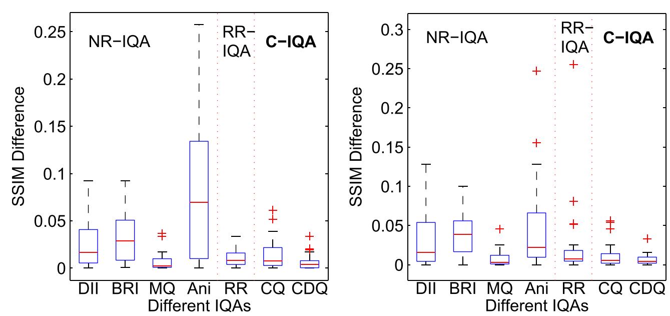
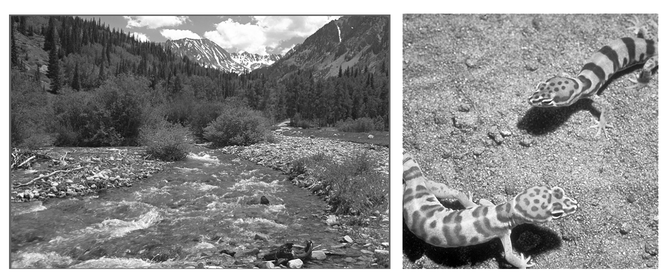


Fig. 13. The SSIM differences of the best images chosen by different NR, RR, and Comparison-based IQA methods for the bilateral filter parameter selection. (a) LIVE. (b) CSIQ.



(a) (b)

Fig. 14. Outliers of parameter selection. (a) “stream” (LIVE) (b) “geckos” (CSIQ).

of different IQA algorithms. Both CDQ and MetricQ give satisfying results for bilateral denoising parameter selection.

In order to better analyze the performance of comparison-based methods, two of the outliers of denoising parameter selection by C-IQA are shown in Fig. 14. Comparison-based methods tend to choose the parameters that lead to over-smoothed denoised results. This is a common challenge for all the NR-IQA algorithms in our experiments. For the lack of texture complexity information from the original image, NR-IQA algorithms are easy to confuse the fine texture with noise component. On the contrary, the RR-IQA in [18] is good at handling images with different global texture complexity because an index that indicates the energy in the disorder part in the original image is available for the image quality assessment.

TABLE III
ACCURACY OF PARAMETER SELECTION FOR BM3D

		DIIVINE	BRISQUE	MetricQ	Anisotropy	RR-IQA	**CQ**	**CDQ**
LIVE	median of all SSIM differences	1.25×10^{-2}	$\mathbf{5.45 \times 10^{-3}}$	7.15×10^{-2}	7.43×10^{-3}	7.43×10^{-3}	1.67×10^{-2}	6.23×10^{-3}
	average of all SSIM differences	1.94×10^{-2}	9.50×10^{-3}	6.43×10^{-2}	1.74×10^{-2}	1.02×10^{-2}	2.24×10^{-2}	$\mathbf{7.52 \times 10^{-3}}$
	average of non-outliers	1.11×10^{-2}	7.25×10^{-3}	6.08×10^{-2}	1.23×10^{-2}	9.45×10^{-3}	1.70×10^{-2}	$\mathbf{6.16 \times 10^{-3}}$
CSIQ	median of all SSIM differences	2.33×10^{-2}	7.25×10^{-3}	6.05×10^{-2}	4.34×10^{-3}	4.57×10^{-3}	8.00×10^{-3}	$\mathbf{8.41 \times 10^{-4}}$
	average of all SSIM differences	2.79×10^{-2}	1.72×10^{-2}	5.49×10^{-2}	9.16×10^{-3}	5.78×10^{-3}	2.05×10^{-2}	$\mathbf{3.05 \times 10^{-3}}$
	average of non-outliers	2.43×10^{-2}	9.48×10^{-3}	5.49×10^{-2}	6.67×10^{-3}	4.88×10^{-3}	1.62×10^{-2}	$\mathbf{1.94 \times 10^{-3}}$

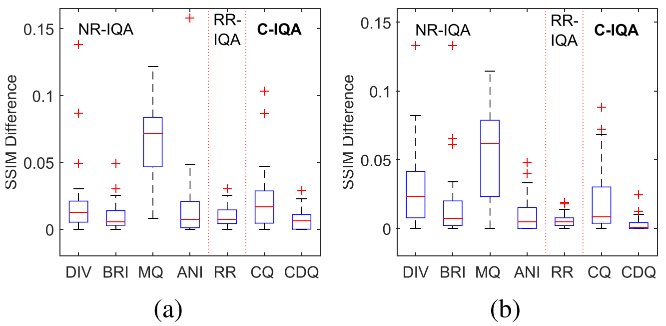


Fig. 15. The SSIM differences of the best images chosen by different NR, RR, and Comparison-based IQA methods for the BM3D parameter selection. (a) LIVE. (b) CSIQ.

3) Parameter Selection for BM3D: Similar to the settings of parameter selection for bilateral filter, a series of increasingly denoised images, $I_1, I_2, \cdots, I_{30}$, are created with BM3D [52]. The SSIM index of the most oversmoothed image I_{30} is between 0.85 ± 0.01. Alg. 5 is adopted to select the best result.

The SSIM difference between the best denoised image chosen by SSIM and the one chosen by other IQA methods is plotted in Fig. 15. Table III shows more quantitative results of different IQA algorithms. It should be noticed that the performance of MetricQ decreases significantly compared with the results of bilateral filters. The reason is that BM3D mainly removes the subtle details as the parameter increases, while MetricQ takes these details as noise. On the contrary, bilateral filters blur major structures as well when the parameter increases. The results of MetricQ and C-IQA on BM3D also reveal that although the implemented C-IQA makes use of MetricQ, the performance of C-IQA is significantly improved based on MetricQ due to the comparison framework.

C. Application in Parameter Trimming

In this section, we combine CDQ with the parameter trimming framework and show that considerable computation can be saved while preserving the accuracy of parameter selection. In this part, all the parameter settings for image reconstruction are the same as the settings in Section V-B1. Fig. 16(a) shows one example image in parameter trimming. The SSIM indexes in Fig. 16(b) and (c) are only used to demonstrated the convergence process. Fig. 16(b) shows the parameter selection after all the algorithm instances with different parameters converge. Fig. 16(c) illustrates the parameter trimming process with CDQ. From Fig. 16, we can see that the trimming decision based on CDQ achieves the goal of terminating the iteration of parameters that are far from the best choice. On LIVE [29], all the parameters selected with parameter trimming are the same as the parameters selected after convergence; on CSIQ [6], only one of the best parameters selected by parameter trimming is different from the one selected after convergence. From Table IV, it is clear that considerable computation is saved with parameter trimming.

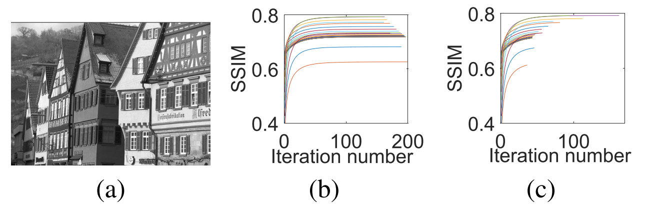


Fig. 16. Comparison between convergence with and without parameter trimming on "buildings". (a) "buildings" (from LIVE). (b) Convergence without parameter trimming. (c) Convergence with parameter trimming.

TABLE IV
COMPUTATION SAVED BY PARAMETER TRIMMING

	Ave # of iteration without parameter trimming	Ave # of iteration with parameter trimming	saved computation (%)
LIVE	4651.9	847.7	81.78
CSIQ	4565.6	941.1	79.39

VI. CONCLUSION

Motivated by the parameter selection for image restoration algorithms, for the first time we proposed a comparison-based IQA framework. The comparison-based method implemented in this paper includes three primary modules, Content Detection, Contribution and Distortion Sensitivity Compensation. One important property that is unique to comparison-based IQA, minimum resolution, is analyzed. At last, the comparison-based IQA compares favorably with other NR-IQA and RR-IQA algorithms on two widely used databases for image reconstruction and bilateral filter parameter selection.

We take CQ and CDQ in this paper as two specific implementations of the comparison-based IQA method. By employing and fusing multiple image quality metrics [26], [27], other comparison-based IQA methods can be designed for different application scenarios in the future.

ACKNOWLEDGMENT

The authors would like to thank Xiang Zhu for providing the MetricQ [5] code and Anush Krishna Moorthy for explaining the usage of DIIVINE [19] code.

REFERENCES

[1] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, no. 1, pp. 1193–1216, Mar. 2001.

[2] J. Lubin, "A human vision system model for objective picture quality measurements," in *Proc. Broadcast. Conv.*, Sep. 1997, pp. 498–503.

[3] S. J. P. Westen, R. L. Lagendijk, and J. Biemond, "Perceptual image quality based on a multiple channel HVS model," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 4. May 1995, pp. 2351–2354.

[4] H. Liang and D. S. Weller, "Regularization parameter trimming for iterative image reconstruction," in *Proc. IEEE Asilomar Conf. Signals, Syst. Comput.*, Nov. 2015, pp. 755–759.

[5] X. Zhu and P. Milanfar, "Automatic parameter selection for denoising algorithms using a no-reference measure of image content," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3116–3132, Dec. 2010.

[6] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, p. 011006, Jan. 2010.

[7] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[8] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3441–3451, Nov. 2006.

[9] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2. Nov. 2003, pp. 1398–1402.

[10] S. J. Daly, "Visible differences predictor: An algorithm for the assessment of image fidelity," in *Proc. Symp. Electron. Imag. Sci. Technol.*, Aug. 1992, pp. 179–206.

[11] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. San Rafael, CA, USA: Morgan & Claypool, 2006.

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

[13] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.

[14] A. Shnayderman, A. Gusev, and A. M. Eskicioglu, "An SVD-based grayscale image quality measure for local and global assessment," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 422–429, Feb. 2006.

[15] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 202–211, Apr. 2009.

[16] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517–526, Feb. 2011.

[17] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E.-H. Yang, and A. C. Bovik, "Quality-aware images," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1680–1689, Jun. 2006.

[18] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1700–1705, Nov. 2013.

[19] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[20] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[21] S. Gabarda and G. Cristóbal, "Blind image quality assessment through anisotropy," *J. Opt. Soc. Amer. A*, vol. 24, no. 12, pp. B42–B51, 2007.

[22] M. A. Saad, A. C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 583–586, Jun. 2010.

[23] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. Int. Conf. Image Process.*, vol. 1. 2002, pp. 477–480.

[24] Q. Wu, Z. Wang, and H. Li, "A highly efficient method for blind image quality assessment," in *Proc. Int. Conf. Image Process.*, Sep. 2015, pp. 339–343.

[25] L. Li, Y. Zhou, W. Lin, J. Wu, X. Zhang, and B. Chen, "No-reference quality assessment of deblocked images," *Neurocomputing*, vol. 177, pp. 572–584, Feb. 2016.

[26] T.-J. Liu, W. Lin, and C.-C. J. Kuo, "A multi-metric fusion approach to visual quality assessment," in *Proc. 3rd Int. Workshop Quality Multimedia Exper.*, Sep. 2011, pp. 72–77.

[27] T.-J. Liu, K.-H. Liu, J. Y. Lin, W. Lin, and C.-C. J. Kuo, "A paraboost method to image quality assessment," *IEEE Trans. Neural Netw. Learn. Syst.*, doi: 10.1109/TNNLS.2015.2500268.

[28] Q. Wu and *et al.*, "Blind image quality assessment based on multi-channel features fusion and label transfer," *Pattern Recognit.*, vol. 26, no. 3, pp. 425–440, Mar. 2016.

[29] H. R. Sheikh, A. C. Bovik, L. Cormack, and Z. Wang. (2005). *LIVE Image Quality Assessment Database Release 2*. [Online]. Available: http://live.ece.utexas.edu/research/quality

[30] R. L. Joshi, V. J. Crump, and T. R. Fischer, "Image subband coding using arithmetic coded trellis coded quantization," *IEEE Trans. Circuits Syst. Video Technol*, vol. 5, no. 6, pp. 515–523, Dec. 1995.

[31] S. M. LoPresto, K. Ramchandran, and M. T. Orchard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework," in *Proc. Data Compress. Conf.*, Snowbird, UT, USA, Mar. 1997, pp. 221–230.

[32] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500–1512, Apr. 2012.

[33] Y. Wang, Q. Zhang, and B. Li, "Structure-preserving image quality assessment," in *Proc. Int. Conf. Multimedia Expo*, Jun./Jul. 2015, pp. 1–6.

[34] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1994, pp. 593–600.

[35] Z.-Q. Hong, "Algebraic feature extraction of image for recognition," *Pattern Recognit.*, vol. 24, no. 3, pp. 211–219, 1991.

[36] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

[37] S. A. Golestaneh and L. J. Karam, "Reduced-reference quality assessment based on the entropy of DNT coefficients of locally weighted gradients," in *Proc. Int. Conf. Image Process.*, Sep. 2015, pp. 4117–4120.

[38] X. Li, "Blind image quality assessment," in *Proc. Int. Conf. Image Process.*, vol. 1. 2002, pp. 449–452.

[39] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2005, pp. 60–65.

[40] L. L. Thurstone, "A law of comparative judgment," *Psychol. Rev.*, vol. 34, no. 4, pp. 273–286, 1927.

[41] B. Peng and O. Veksler, "Parameter selection for graph cut based image segmentation," in *Proc. Brit. Mach. Vis. Conf.*, 2008, pp. 1–10.

[42] Y. Yitzhaky and E. Peli, "A method for objective edge detection evaluation and detector parameter selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 1027–1033, Aug. 2003.

[43] M. S. C. Almeida and M. A. T. Figueiredo, "Parameter estimation for blind and non-blind deblurring using residual whiteness measures," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2751–2763, Jul. 2013.

[44] J. L. Herraiz, S. Gabarda, and G. Cristobal, "Automatic parameter selection in PET image reconstruction based on no-reference image quality assessment," in *Proc. Nucl. Sci. Symp. Med. Imag. Conf. Rec.*, Oct./Nov. 2012, pp. 3371–3374.

[45] D. S. Weller, S. Ramani, and J. A. Fessler, "Augmented Lagrangian with variable splitting for faster non-Cartesian L_1-SPIRiT MR image reconstruction," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 351–361, Feb. 2014.

[46] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 37, no. 12, pp. 2024–2036, Dec. 1989.

[47] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," in *Proc. Int. Conf. Image Process.*, vol. 3. Oct. 2001, pp. 542–545.

[48] L. I. Rudin and S. Osher, "Total variation based image restoration with free local constraints," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, vol. 1. Nov. 1994, pp. 31–35.

[49] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, Jan. 2009.

[50] R. K. S. Kwan, A. C. Evans, and G. B. Pike, "MRI simulation-based evaluation of image-processing and classification methods," *IEEE Trans. Med. Imag.*, vol. 18, no. 11, pp. 1085–1097, Nov. 1999.

[51] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. Int. Conf. Comput. Vis.*, Jan. 1998, pp. 839–846.

[52] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2096, Aug. 2007.

Haoyi Liang (S'15) received the B.S. degree in information engineering from Shanghai Jiao Tong University, China, in 2014. He is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of Virginia. His research interests are in the domain of image processing and computer vision, MRI reconstruction, deblurring, super-resolution, and image quality assessment.

Daniel S. Weller (S'05–M'12) received the B.S. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2006, and the M.S. and Ph.D. degrees in electrical engineering from MIT, Cambridge, MA, USA, in 2008 and 2012, respectively. He was a Post-Doctoral Research Fellow with the University of Michigan, Ann Arbor, MI, USA, where he was involved in imaging research supported by a U.S. National Institutes of Health Ruth L. Kirschstein National Research Service Award Post-Doctoral Fellowship. He is currently an Assistant Professor with the Charles L. Brown Department of Electrical and Computer Engineering, University of Virginia (UVA), Charlottesville, VA, USA, where he leads the Virginia Imaging Theory and Algorithms Laboratory. He is also a Team Leader for UVA CHARGE, an NSF ADVANCE program aimed at recruiting and retaining women faculty in STEM and SBE fields. He is a member of the ISMRM, AHA, Eta Kappa Nu, and Tau Beta Pi. He is a member of the special interest group on Computational Imaging in the IEEE Signal Processing Society. He serves as an Associate Editor of the IEEE TRANSACTIONS ON MEDICAL IMAGING.