Image Deconvolution with Multi-stage Convex Relaxation and Its Perceptual Evaluation

Tingbo Hou, Sen Wang, and Hong Qin

Abstract—This paper proposes a new image deconvolution method using multi-stage convex relaxation, and presents a metric for perceptual evaluation of deconvolution results. Recent work in image deconvolution addresses the deconvolution problem via minimization with non-convex regularization. Since all regularization terms in the objective function are non-convex, this problem can be well modeled and solved by multi-stage convex relaxation. This method, adopted from machine learning, iteratively refines the convex relaxation formulation using concave duality. The newly-proposed deconvolution method has outstanding performance in noise removal and artifact control. A new metric, transduced contrast-to-distortion ratio (TCDR), is proposed based on a human vision system (HVS) model that simulates human responses to visual contrasts. It is sensitive to ringing and boundary artifacts, and very efficient to compute. We conduct comprehensive perceptual evaluation of image deconvolution using visual signal-to-noise ratio (VSNR) and TCDR. Experimental results of both synthetic and real data demonstrate that our method indeed improves the visual quality of deconvolution results with low distortions and artifacts.

Index Terms—Image deconvolution, Non-convex regularization, Perceptual evaluation, Multi-stage convex relaxation.

I. INTRODUCTION

RECOVERING a sharp image from a single, motion-blurred photograph has been a long-standing, challenging research problem. If the blur kernel, i.e., point spread function (PSF), is shift-invariant, this problem reduces to that of image deconvolution. It is one of the commonly-encountered issues receiving considerable attention in a broad range of fields, such as image processing, astronomy, computational photography, medical imaging, and optimization. In recent work [1]–[7], notable progress has been made using the *natural image prior*, referring to the heavy-tailed distribution of image derivative magnitudes. The natural image prior can be represented by a hyper-Laplacian function, which is essentially a non-convex regularization in optimization. It forces large derivatives of the deblurred image to be sparse, resulting in nice performance on ringing control. Nonetheless, there are still challenges remaining in many aspects [8]. In this paper, we tackle the image deconvolution problem at several fronts: including a multi-resolution prior, non-convex regularization, and perceptual evaluation, with novel technical contributions for each aspect.

Image deconvolution has been intensively studied for decades. Despite the accomplishments, it is still hard to achieve both sharpness and cleanness in deblurred images, especially for some extreme cases such as strong edges and large noise. Strong edges are often accompanied by strong ringing artifacts that severely reduce the visual fidelity of the deblurred image. Noise is inevitable in deblurring, which comes from many sources including photon collection, quantization error, and blur kernel error. Artifacts may rapidly increase when the algorithm attempts to drive the blurred image sharper, resulting in some meaningless solution that may not be a natural image. For non-convex regularization that is frequently adopted by recent deblurring methods, it is difficult to find the global solution. Solutions obtained by existing methods are local minima, which depend not only on the relaxation, but also the numerical procedure including boundary padding, convolution computation, linear-system solving, etc. For example, it is well known that the frequency methods for computing convolutions have Gibbs phenomena and boundary artifacts. Therefore, the studies of robust image deconvolution, including regularization, relaxation, solving scheme, and numerical procedure, are important and necessary.

Quantitative evaluation is another important aspect of image deconvolution that has not been fully explored. In previous work, evaluations were conducted based on conventional metrics including the sum of squared differences (SSD) [8] and the peak signal-to-noise ratio (PSNR) [4], [6]. These metrics measure the difference between two signals, regardless of human response. However, for visual signals, absolute luminance values have different responses from human eyes, which means that the metrics based on luminance is neither accurate nor consistent from the visual perception's point of view. Visual perceptual metrics, on the other hand, have been studied with a rich literature in recent years [9]–[11] for measuring image fidelity, for example the visual signal-to-noise ratio (VSNR) [11]. Whereas these metrics are designed to output values highly correlated with perceived distortions, typical artifacts in image deblurring are not yet to be considered, such as ringing and boundary artifacts. Furthermore, these metrics have not been applied for perceptual evaluation of image deconvolution.

In this paper we focus on non-blind deconvolution and its perceptual evaluation, assuming that the blur kernel is shift-invariant. The contributions of this paper are summarized as follows:

- We present a multi-resolution sparse prior for image deconvolution, which is a generalization of the natural image prior. It naturally leads to a remarkable property

of natural images that derivatives in different resolutions are subject to similar heavy-tailed distributions.

- We develop a new non-blind deconvolution method using multi-stage convex relaxation. This scheme, adopted from machine learning with non-convex regularization, iteratively refines the convex relaxation using concave duality. The justification of convergence is automatically embedded in a joint minimization that induces a convex formulation.

- We propose a new perceptual metric, called *transduced contrast-to-distortion ratio* (TCDR), to evaluate visual quality of image deconvolution. This metric transforms luminance contrasts between two images to human visual responses based on a human visual system (HVS) model. It has great responses on ringing and boundary artifacts, since they are large contrasts and therefore large stimuli to human eyes.

- We conduct perceptual evaluation of image deconvolution results using VSNR (a state-of-the-art visual metric) and our new TCDR, which collectively make the evaluation more comprehensive and convincing. The evaluation is performed on both synthetic data and real data.

The remainder of this paper is organized as follows. We briefly review the previous work related to image deconvolution and image fidelity assessment in Section II. We then present the multi-resolution sparse prior and the deconvolution method using multi-stage convex relaxation in Section III. Our new perceptual metric for measuring image deconvolution based on a HVS model is detailed in Section IV. We finally demonstrate experimental results of perceptual evaluation on synthetic and real data in Section V, and conclude the paper with discussion and future work in Section VI.

II. PREVIOUS WORK

A. Image Deconvolution

Tremendous work has been dedicated to the problem of image deconvolution, which can be roughly classified into two categories: *non-blind* deconvolution and *blind* deconvolution, referring to deconvolution with known or unknown kernel, respectively. For non-blind deconvolution, a fundamental technique is the Richardson-Lucy (RL) deconvolution [12], which restores the latent image by a Bayesian-based model. Gaussian prior, inducing the Tikhonov regularization [13], was adopted in deconvolution at first. Later on, non-Gaussian prior that gives small penalties on large image edges became popular [14]–[16], such as total variation (TV) regularization [15], [17], half-quadratic regularization [18], and anisotropic regularization [16]. Recently, the natural image prior that is a non-convex regularization prevailed in image deblurring. Levin et al. [2] proposed the sparse prior as a concise representation of the natural image prior, which has an outstanding performance in ringing control. It was also referred to the hyper-Laplacian prior in [6], but with a fast solving strategy using variation and fast Fourier transform (FFT). According to a recent evaluation of deconvolution algorithms [8], the sparse prior has been shown to achieve the best performance in the non-blind deconvolution process. Yuan

et al. [19] proposed a progressive inter-scale and intra-scale approach based on an edge-preserving bilateral Richardson-Lucy (BRL) deconvolution. The deconvolution was performed on the residual image between the blurred image and the sharp image from the last iteration convolved with a blur kernel. It has a special feature in ringing control, but the residual deconvolution in the detailed layer is less stable in existence of noise. Joshi et al. [4] exploited color priors as well as the sparse prior in non-blind image deconvolution, and applied them to deblurring and denoising. Cho et al. [7] proposed a content-aware image prior that estimates spatially-varying gradient statistics, and applied it together with the sparse prior for non-blind deconvolution.

Blind deconvolution is more challenging since both the latent image and the kernel are unknown. Some researchers attempted to address the problem using a single image with regularization terms including TV regularization [15], anisotropic regularization [16], non-convex regularization [1], etc. Fergus et al. [1] used a zero-mean mixture-of-Gaussians model learned from a natural image as image prior to recover the blur kernel. In [3], Shan et al. exploited a concatenation of two piece-wise continuous linear and quadratic functions to model image prior in their deblurring framework. Cai et al. [5] used the curvelet system for kernels and the framelet system for images to reduce the ill-posed problem to a joint optimization that maximizes the sparsity of the kernel and the sharp image. In [20], a fast deblurring method was proposed by introducing a prediction step. The graphics processing unit (GPU) implementation makes their algorithm very fast. In [8], Levin et al. analyzed maximum a posterior (MAP) failure in blind deconvolution, and evaluated single-image deconvolution algorithms using collected blur data with ground truth. Multiple blurred images can provide more information of the scene, which can reduce ambiguities of blind deblurring. Chen et al. [21] developed an algorithm to deblur two consecutively-captured blurred photos from camera shaking. In [22], Cai et al. relaxed this requirement by extremely sparse representation in the redundant curvelet system, but still required manual alignment of the two images. For non-uniform deblurring, Whyte et al. [23] proposed a parametrized geometric model based on the rotational velocity of the camera, and applied it to blind and non-blind deblurring.

Multiple images or specially-designed cameras were also adopted to aid deconvolution. In [24], a hybrid camera that simultaneously captures a high-resolution image together with a sequence of low-resolution images was employed in motion deblurring. Rasker et al. [25] proposed a *fluttered shutter* camera with coded exposure, which opens and closes the shutter during a normal exposure time. Yuan et al. [26] exploited blurred/noisy image pairs for deblurring, which contain a blurred image under a long exposure and a noisy image by fast shutter. Levin et al. [2] inserted a patterned occluder within the aperture of the camera lens, creating a *coded aperture* as known PSF. Zhou and Nayar [27] evaluated aperture patterns based on the quality of deblurring. In [28], Levin et al. built a prototype camera that translates within its exposure following a parabolic displacement rule. Hence blur can be removed by deconvolving the entire image with an identical, known PSF.

Hirsch et al. [29] proposed a class of linear transforms for space-varying filters in multiframe blind deconvolution. They used the efficient filter flow (EFF) framework for rapid computation of matrix-vector-multiplication. In [30], a prototype with gyroscopes and accelerometers to estimate the camera motion was presented, where deconvolution was performed based on the "measured" blur kernel by the sensors, coupled with a natural image prior.

In addition to deblurring, image deconvolution has been applied to other problems such as super-resolution and denoising. In [31], Shan et al. applied their deconvolution method in [3] to fast super-resolution by assuming that the convolution kernel is a Gaussian function. Joshi et. el. used deconvolution to perform denoising in [4]. In [32], three applications were performed: deblurring, super-resolution, and denoising.

B. Image Fidelity Assessment

Previous evaluations of image deconvolution were based on two conventional metrics: PSNR [4], [6] and SSD [8]. However, for visual signals such as images, it is well known that human eyes are more sensitive to luminance ratios rather than absolute luminance values. It prompts researchers to seek perceptual metric that responds to perceived distortions in the same or similar ways with human eyes. The HVS model has been studied for decades, with many efforts to design perceptual evaluation of image distortion. The widely-used property of HVS is contrast sensitivity, which induces metrics based on the contrast sensitivity function [33]–[37]. Another class of metric uses high-level properties of vision, based on overarching hypotheses of what the human visual system attempts to achieve. Wang et al. [9] proposed the structural similarity (SSIM) metric that operates based on the notion that the HVS has evolved to extract structural information from natural images. In [10], an informative-theoretic approach was proposed using an informative fidelity criterion (IFC) derived from natural-scene statistics. In the recent work [11], a wavelet-based VSNR was proposed to quantify the visual fidelity of natural images based on near-threshold and supra-threshold properties of human vision.

Recently, a concise HVS model was presented and applied to the design and assessment of High Dynamic Range (HDR) image display. Mantiuk et al. [38] proposed a framework for image processing operations that works in the visual response space, and applied it to contrast mapping, contrast equalization, and color-to-gray conversion. They studied the contrast discrimination model for HDR images, and introduced a transducer function that can represent the response of the HVS model. In [39], they defined visual distortion for tone reproduction and proposed a tone mapping operator to minimize the distortion. Aydin et al. [40] presented an image quality metric for HDR images, using the HVS model. Their metric is based on three types of distortions: loss of visible contrast, amplification of invisible contrast, and reversal of visible contrast.
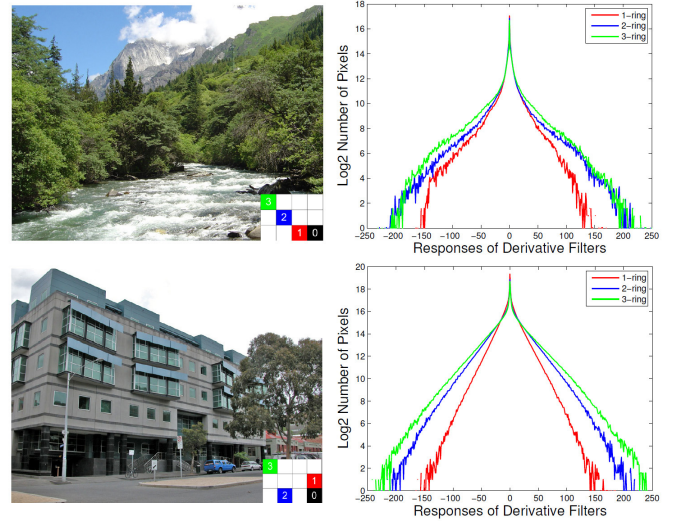

Fig. 1. Two natural images and their statistical responses of three derivative filters with different resolutions, which are selected from 1-ring, 2-ring, and 3-ring neighbors of the central pixel (0). The patterns of these three filters are shown at the lower-right corner of the left-hand images.

III. IMAGE DECONVOLUTION USING MULTI-STAGE CONVEX RELAXATION

In this section, we introduce our deconvolution method using a multi-resolution sparse prior and the multi-stage convex relaxation. Since we focus on non-blind deconvolution, a kernel is required as an input, which can be obtained by the blurred/noisy image pair [26] or Fergus's method in [1]).

A. Multi-resolution Sparse Prior

Natural images have an intrinsic property on the statistics of their gradient magnitudes: the heavy-tailed distribution. It generates sparse regularization in the objective function, and therefore, makes it non-convex. In our observation, magnitudes of multi-resolution gradients are also subject to this property. Fig. 1 shows two natural images and their statistical responses of three derivative filters with different resolutions, selected from 1-ring, 2-ring, and 3-ring neighbors of the central pixel. It illustrates that the responses of derivative filters within a certain size of local neighborhood (3-ring in this example) have similar distributions. Moreover, the tails of response distribution become lighter when the distance of the derivative filter increases, indicating the spatial affects to the distribution of filter outputs.

Inspired by the above observations, we propose a multi-resolution sparse prior

$$p(x) = \prod_{j=1}^{J} e^{-\phi(x, f_j)}, \tag{1}$$

where $\phi(x, f_j)$ is the potential function of image x and derivative filter $\{f_j\}$. For neighborhood size $m \times m$, $J = (m^2 - 1)/2$ forward filters are utilized. We adopt the hyper-Laplacian function to represent the natural image prior, and embed a spatial Gaussian function $G_\sigma(f_j)$ to reflect spatial affects, which yields

$$\phi(x, f_j) = G_\sigma(f_j)|x \otimes f_j|^p, \tag{2}$$

where $\otimes$ is the convolution operator, σ is the standard deviation of Gaussian, and p is a positive exponent value set in the range of [0.5, 0.8] as suggested by [2], [6]. The sparse function favors natural images, since their gradients are very non-Gaussian. Compared with $p=2$ (Gaussian prior) and $p=1$ (Laplacian prior), it applies less penalties on larger gradients. Hence, it is more likely to favor sharp explanations in image deconvolution than the Gaussian prior and Laplacian prior. For extensive justification and comparison, please refer to [8]. The spatial Gaussian $G_\sigma(f_j)$ is defined as the one in the bilateral filter [41]. This prior reflects the intrinsic characteristic of natural images with concise expression, and will serve as the regularization term in the subsequent minimization problem.

B. Deconvolution with Sparse Regularization

Shift-invariant motion blur is commonly modeled as a convolution process. The blurred image is interpreted as a latent image x convolved by a blur kernel k with additive noise n, written as

$$y = x \otimes k + n. \tag{3}$$

Derived from Bayes' rule, the estimation of the latent image x was modeled as a minimization problem with loss function $R_0(x)$ and a set of non-convex regularization terms $\{R_j(x)\}$,

$$\hat{x} = \arg\min_x \left[R_0(x) + \sum_{j=1}^J R_j(x) \right]. \tag{4}$$

The loss function here is set in a least-square fashion by assuming a Gaussian noise model,

$$R_0(x) = (x \otimes k - y)^2. \tag{5}$$

The non-convex regularization terms come from the adopted image prior, which are

$$R_j(x) = \lambda G_\sigma(f_j)|x \otimes f_j|^p, \tag{6}$$

in our work. Therefore, we can re-write Eq. (4) as,

$$\hat{x} = \arg\min_x \left[(x \otimes k - y)^2 + \lambda \sum_{j=1}^J G_\sigma(f_j)|x \otimes f_j|^p \right]. \tag{7}$$

C. Multi-stage Convex Relaxation

In previous work of image deblurring, the non-convex regularization was solved by FFT-based variational methods [3], [6] or iterative re-weighted least-square (IRLS) method [2]. FFT-based variational methods decompose the objective function into sub-problems. Their results are not solutions to the non-convex objective function, and also suffer from boundary artifacts due to FFT. The IRLS method can be interpreted as a one-stage convex relation approach, based on the majorization-minimization (MM) principle. We adopt the multi-stage convex relaxation [42] to solve the deconvolution with sparse regularization. This solving scheme, originally proposed in machine learning, was used for solving problems with non-convex objective functions. Specifically, it considers a non-convex component, e.g., $R_j(x)$ in Eq. (4). Let $h_j(x)$ be a

function with range Ω_j, and assume that there exists a function $\bar{R}_j(x)$ defined on Ω_j, which yields

$$R_j(x) = \bar{R}_j(h_j(x)). \tag{8}$$

Assume that there is a function u_j so that the $\bar{R}_j(u_j)$ is concave on $u_j \in \Omega_j$. Zhang [42] showed that the regularization function can be re-written, using the concave duality [43], as

$$R_j(x) = \inf_{v_j} \left[v_j^T h_j(x) + R_j^*(v_j) \right], \tag{9}$$

where $R_j^*(v_j)$ is the concave dual of $\bar{R}_j(u_j)$ given by

$$R_j^*(v_j) = \inf_{u_j} \left[-v_j^T u_j + \bar{R}_j(u_j) \right]. \tag{10}$$

According to [42], the minimization of the right-hand side of Eq. (10) is achieved at

$$\hat{v}_j = \nabla_{u_j} \bar{R}_j(u_j)|_{u_j = h_j(x)}. \tag{11}$$

Using concave duality, the minimization with non-convex regularization can be solved in a multi-stage convex relaxation method. Let $h_j(x)$ be a convex relaxation of $R_j(x)$. The minimization problem in Eq. (4) is relaxed to

$$\hat{x} = \arg\min_x \left[R_0(x) + \sum_{j=1}^J h_j(x)^T v_j \right], \tag{12}$$

where all of the components are convex. However, the solution obtained from this simple relaxation is different from the solution of Eq. (4). Using the representation of $R_j(x)$ in Eq. (9), the minimization in Eq. (4) can be re-written as

$$[\hat{x}, \hat{v}] = \arg\min_{x,\{v_j\}} \left[R_0(x) + \sum_{j=1}^J (h_j(x)^T v_j + R_j^*(v_j)) \right]. \tag{13}$$

This can be solved by an alternate optimization strategy, which optimizes x with fixed v and then optimizes v with fixed x. The first optimization is equivalent to Eq. (12), since $R_j^*(v_j)$ is independent of x. The second optimization is non-convex, but has a closed-form solution given in Eq. (11).

Above we outlined the general procedure of multi-stage convex relaxation proposed by [42]. It has some nice properties compared with the standard one-stage convex relaxation method. The justification of convergence is automatically embedded in the joint minimization in Eq. (13), which induces a convex formulation. Moreover, Zhang [42] showed that for sparse regularization, since the local solution found by this algorithm is the global solution of a refined convex relaxation, it should be closer to the desired solution of the non-convex regularization than that of the one-stage convex relation. Now, we consider the minimization problem in image deconvolution with our multi-resolution prior given in Eq. (7). We define the convex relaxation function $h_j(x)$ as

$$h_j(x) = |x \otimes f_j|^q, \tag{14}$$

with exponent $q \geq 1$, and the concave function $\bar{R}_j(u_j)$ as

$$\bar{R}_j(u_j) = \lambda G_\sigma(f_j)|u_j|^{p/q}. \tag{15}$$
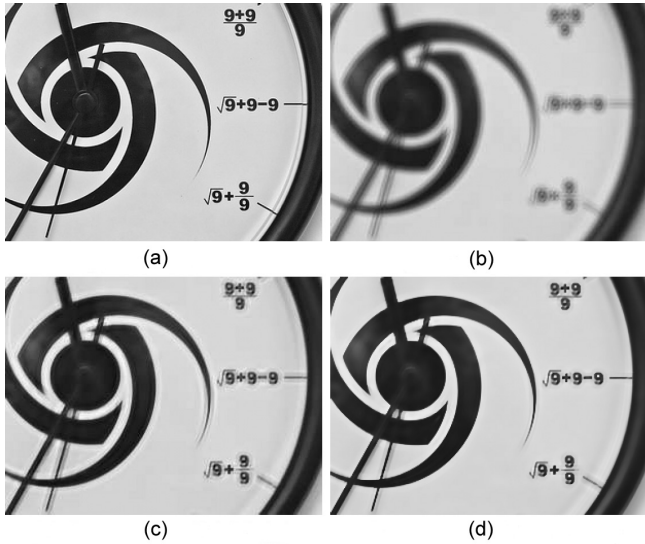
Fig. 2. Deconvolution with strong edges: (a) is the latent image with strong edges; (b) is a blurred image; (c) is the result by sparse prior [2] with PSNR 24.34dB; (d) is our result with PSNR 27.09dB.
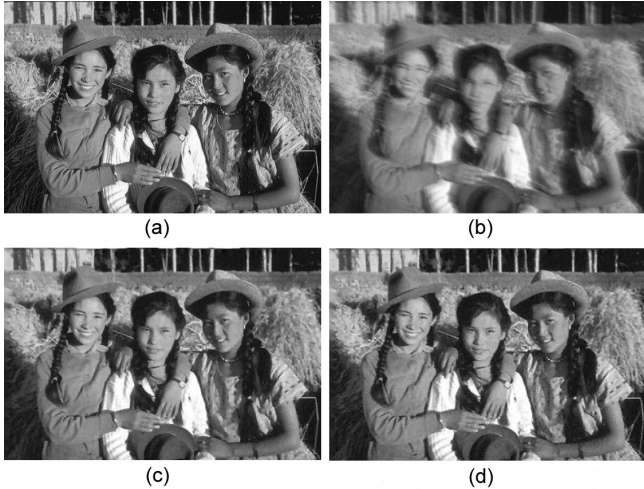

Fig. 3. Deconvolution with large noise: (a) is the latent image; (b) is a blurred image with 5% random noise; (c) is the result by sparse prior [2] with PSNR 25.67dB; (d) is our result with PSNR 26.18dB.

The first-step optimization can be expressed as

$$\hat{x} = \arg\min_{x} \left[(x \otimes k - y)^2 + \sum_{j=1}^{J} |x \otimes f_j|^q v_j \right], \quad (16)$$

which can be solved using the Euler equation and conjugate gradient method. According to Eq. (11), we take the derivative of Eq. (15) with respect to u_j, and obtain the second optimization

$$\hat{v}_j = \lambda(p/q)G_\sigma(f_j)|h_j|^{(p/q-1)}, \quad j = 1, 2, ..., J. \quad (17)$$

The proposed method with multi-resolution sparse prior and multi-stage convex relaxation has stronger ability for artifact and noise control than previous approaches using the sparse prior [2], [6], especially for the cases with strong edges and large noise, as shown in Fig. 2 and Fig. 3. The sparse prior, as a concise representation of natural image prior, gives small

penalties on large derivatives, which therefore restrain the noise while keeping edges. For comparison purpose, the multi-resolution prior is an extension of the sparse prior, but considers multi-resolution derivatives in a larger neighborhood. Moreover, our solving scheme using multi-stage relaxation ensures convergence to some local minimum, and has expected better solutions in the sparse relaxation. Extensive evaluations will be given in Section V.

IV. PERCEPTUAL METRIC: TCDR

We now introduce a concise HVS model, and propose a new perceptual metric TCDR for image deconvolution.

A. Human Visual System Model

A HVS model aims to transform input luminance to response that simulates human vision. It is essential for image quality assessment since human vision is sensitive to contrast not luminance values. We adopt a concise model of HVS in [38] based on perceived contrast. To simplify the formulation, we start with the contrast as logarithmic ratio of luminance:

$$G(L_1, L_2) = \log_{10}(L_1/L_2), \quad (18)$$

where L_1 and L_2 are two luminance values. Next, we need a transducer function that predicts the hypothetical response of the HVS for a given measured contrast. The transducer function in [38] is derived from [44], based on the assumption that the response value should change by one unit for each Just-Noticeable Difference (JND), beginning with

$$T(0) = 0, \quad T(G_\theta) = 1, \quad (19)$$

where the threshold of contrast detection (i.e., JND) G_θ is approximated with 1% contrast $\log_{10}(0.01 + 1)$. The rest of the transducer function is approximated by its first derivative:

$$\Delta T \approx \frac{dT(G)}{dG}\Delta G(G) = 1, \quad (20)$$

where $\Delta G(G)$ is the discrimination threshold obtained from Whittle's psychophysical experimental measurements [45]. We suggest that readers refer to [38] for more details. However, Mantiuk's model works in full range of luminance for HDR display. Thus we normalize the response by a scale

$$R(G) = \frac{1}{T(G_{max})}T(G), \quad (21)$$

where the maximum contrast for 8-bit low dynamic range images is $G_{max} = \log_{10}(255)$. Fig. 4 shows the normalized response of the transducer function, where the horizontal axis is the contrast defined by Eq. (18), and the vertical axis denotes the response of the transducer function defined by Eq. (21).

B. Perceptual Metric

Based on this HVS model, we define our pixel-wise measurement and perceptual metric for image quality assessment. This perceptual metric is directly computed from the contrast
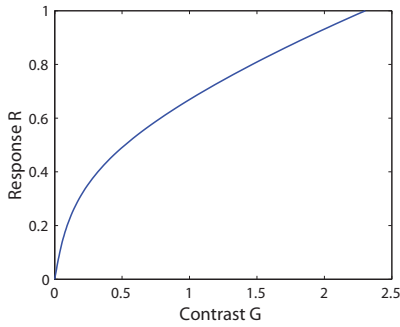
Fig. 4. Normalized response of transducer function in the HVS model. The x-axis is the contrast defined by Eq. (18), and the y-axis denotes the response of the transducer function defined by Eq. (21).
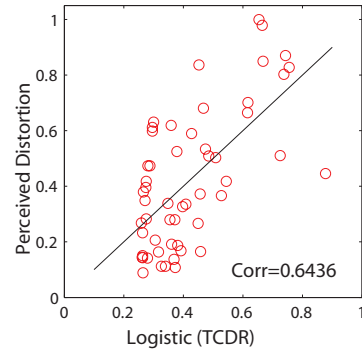

Fig. 5. Subjective ratings of perceived distortion plotted with respect to predicted values of metric TCDR. The y-axis denotes perceived distortion as reported by subjects, and the x-axis corresponds to transformed metric outputs.

between two images with luminance maps L_1 and L_2 at pixel i, given by

$$G(L_1(i), L_2(i)) = \log_{10}(L_1(i)/L_2(i)). \qquad (22)$$

According to Eq. (21), the visual distortion map of two images can be computed by

$$R_{L_1,L_2}(i) = R(G(L_1(i), L_2(i))). \qquad (23)$$

This is a pixel-wise measurement also referred as visual difference of two images, which can be interpreted as a visual distortion map. Then we compute the mean square distortion (MSD) on L_1 and L_2, given by

$$MSD(L_1, L_2) = \frac{1}{N} \sum_i^N |R_{L_1,L_2}(i)|^2, \qquad (24)$$

where N is the number of pixels. Ultimately, we reach the *transduced contrast-to-distortion ratio* (TCDR), defined as

$$TCDR = 10 \log_{10} \left[\frac{R(G_{max})^2}{MSD} \right], \qquad (25)$$

where G_{max} is the maximum contrast of L_1 and L_2.

The TCDR measures the visual difference of two images. Since the normalized transducer function $R(G)$ in Eq. (21) is monotonic with respect to the contrast G, greater contrasts will produce greater responses. Therefore, the metric TCDR is indeed a metric, which is sensitive to artifacts that have large contrasts.

C. Fast Implementation

One merit of the TCDR metric is that, it can be computed very fast. The domain transducer function in Eq. (21) can be transformed from the contrast G to the ratio of luminance $L_r = L_{max}/L_{min}$. Furthermore, this function can be implemented as a lookup table $R_{LUT}(L_r)$, which is subject to

$$R_{LUT}(L_{min}/L_{max}) = -R_{LUT}(L_{max}/L_{min}).$$

This feature results from the symmetry of the transducer function to contrast G. The relation between pixel intensity I and physical luminance L can be modeled by the gamma correction, given by

$$L(I) = I^\gamma, \qquad (26)$$

where γ is the gamma of display monitor, which is set to be 2.2 in our implementation. The range of intensity I is [0,1], and hence the range of luminance L is also [0,1]. The computation of two pixels is then reduced to a power operation, a division, and a read operation from a table stored in memory.

V. EXPERIMENTAL RESULTS

In this section, we first analyze the proposed perceptual metric TCDR, and then conduct perceptual evaluation of image deconvolution on both synthetic and real data. The deconvolution methods used for comparison include: the sparse prior method (sparse) [2], and the fast hyper-Laplacian method (fast) [6]. The sparse method with L_p regularization is solved by the conjugate gradient method and IRLS. The fast method has the same regularization with the sparse method, but is solved by FFT-based variation.

A. Metric Analysis

As a perceptual metric, The TCDR should be correlated to ratings of subjects on perceived distortion. Therefore, we examine the TCDR on the A57 database [1]. This database contains 18 distorted images with scaling ratings by subjects in psychophysical experiments. The perceived distortion includes "flat" allocation, baseline JPEG compression, baseline JPEG-2000 compression, JPEG-2000+DCQ compression, Gaussian blur, and Gaussian white noise. The plot of ratings with respect to predicted values of TCDR is shown in Fig. 5. The y-axis denotes perceived distortion as reported by subjects, and the x-axis corresponds to transformed metric outputs, given by a logistic function [11]

$$f(x) = \frac{\tau_1 - \tau_2}{1 + e^{\frac{x - \tau_3}{\tau_4}}} + \tau_2, \qquad (27)$$

where parameters τ_1, τ_2, τ_3, and τ_4 are obtained by minimizing the sum of squared errors between the transformed metric outputs $\{f(x)\}$ and the corresponding subjective ratings. The correlation between subjective ratings and transformed outputs of TCDR is $Corr$=0.6436. The plot of VSNR as well as other metrics is referred to [11].

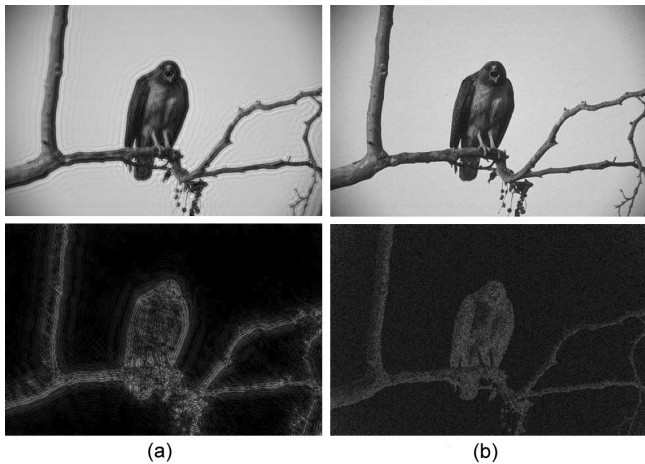[1] http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html

(a)　(b)

Fig. 6.　Two distorted images and their distortion maps with: (a) ringing and (b) noise. The evaluation values are shown in Table I.

TABLE I
EVALUATION VALUES ON IMAGES IN FIG. 6.

Metric	PSNR	VSNR	TCDR
Image (a) (dB)	31.09	10.37	18.36
Image (b) (dB)	30.89	6.90	18.45
Computation time (s)	0.0097	0.2650	0.0765

Perceptual metrics have been widely used to measure image fidelity in image compression. However, the psychophysical ratings are very subjective, and highly dependent on the dataset. We notice that familiar artifacts (such as ringing and boundary) in deblurring are not considered in previous subjective ratings of visual fidelity. It may be somehow confusing to evaluate image deconvolution using previous perceptual metrics such as VSNR. For instance, we obtain two distorted images with: ringing (a) and noise (b) in Fig. 6, and evaluate their qualities with respect to the original image. The evaluation values are given in Table I. Different conclusions are drawn using different metrics: by PSNR, (a) is slightly better than (b); by VSNR, (a) is much better than (b); and by TCDR, (b) is slightly better than (a). It is not odd to have these discrepancies when evaluating images with different metrics. In this particular case, the result by TCDR is more reasonable, since ringings are large stimuli to human eyes and strongly affect the visual quality. This experiment highlights the feature that the TCDR has strong responses to (ringing and boundary) artifacts in deblurring, which are large contrasts and therefore large stimuli to human eyes. It is interesting and necessary to have more than one perceptual metrics for comprehensive and convincing evaluation of image deconvolution, since artifacts in deblurring have not been included in the psychophysical ratings. Therefore, in the following experiments, we will use both VSNR and TCDR to conduct perceptual evaluations. The PSNR values are also given for comparison purpose.

Moreover, the computation of TCDR is very fast. It can be implemented by a lookup table storing the transducer function $R(L_1, L_2)$ in Eq. (21). In Fig. 6, the image size is 481×321, and the computation time of the three metrics is given in Table I, obtained on a laptop with Duo CPU 2.53GHz.
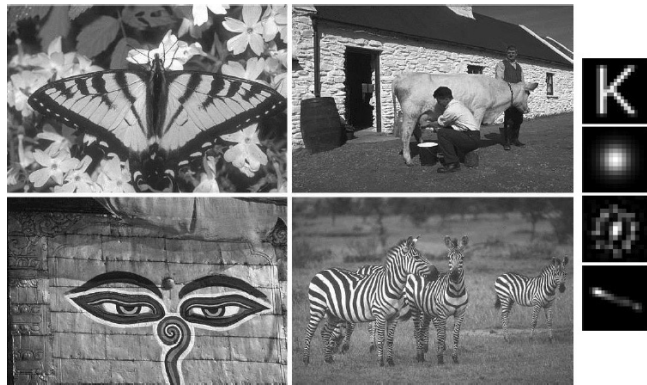

Fig. 7.　Synthetic data: four selected images from the Berkeley Segmentation Dataset and four blur kernels.
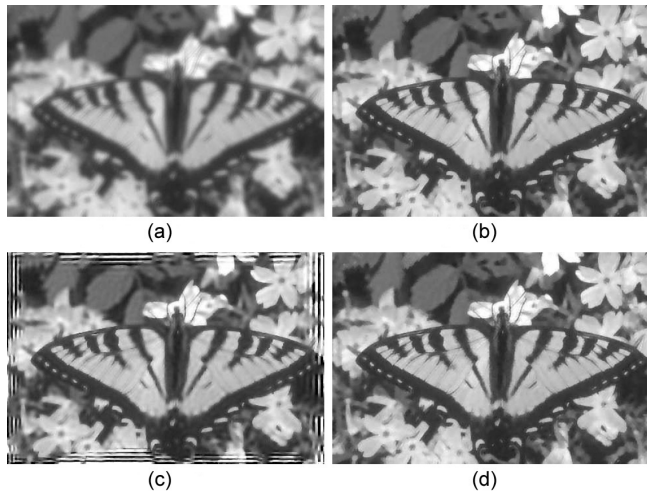

(a)　(b)

(c)　(d)

Fig. 8.　An example of the experiment on synthetic data: (a) a blurred image, (b) sparse method, (c) fast method, and (d) ours. Detailed evaluation values and the statistical analysis can be found in Fig. 9 and Table II.

B. Synthetic Data

In our synthetic data, we select four images from the Berkeley Segmentation Dataset [46] operated with four blur kernels, as shown in Fig. 7. The four kernels are: synthetic thin kernel, Gaussian kernel, coded aperture [2], and camera shake kernel. The blurred images form two groups of data by adding 1% and 3% random noise, respectively. This dataset therefore contains 32 blurred images with different levels of noise. The image resolution is 481×321, and the kernel size varies from 15×15 to 21×21.

We compare three methods on the synthetic data: sparse, fast, and our method. We use boundary value replication for boundary condition in our method. For all three methods, parameters are tuned on one image in each data group, and applied to the rest of the images in the group. Since images in one group have similar levels of blur and noise, they can share the same parameters. Specifically, the parameters of the three methods for the first data group (1% noise) are λ=0.0001, [λ=6000, α=0.4], and [σ=0.8, λ=0.0001]. The parameters for the second data group (3% noise) are λ=0.001, [λ=4000,α=0.1], and [σ=0.8, λ=0.0006]. Fig. 8 shows an example (with 3% noise) in this experiment. In this case, the
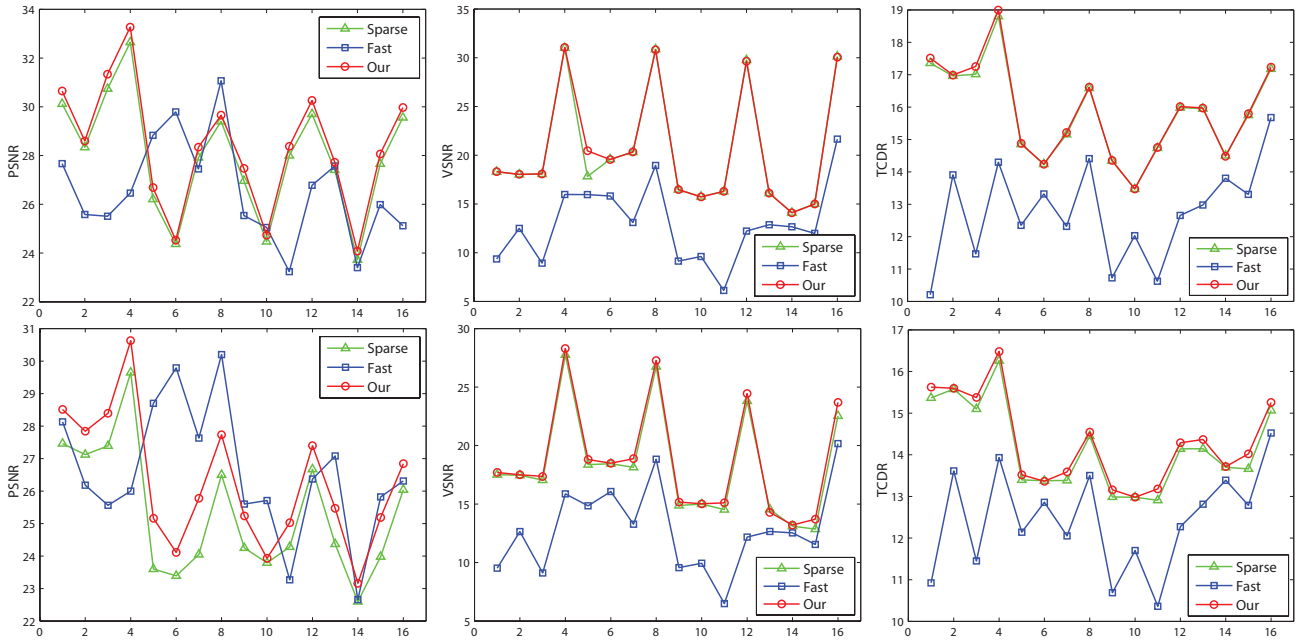
Fig. 9. Evaluation values on synthetic data: group 1 with 1% noise (first row) and group 2 with 3% noise (second row). The average improvements of our method over the other two methods are shown in Table II.

TABLE II
AVERAGE IMPROVEMENT OF OUR METHOD ON SYNTHETIC DATA.

Methods	PSNR (dB)	VSNR (dB)	TCDR (dB)
Our-Sparse	0.1156	0.1473	0.0546
	0.6512	0.3859	0.1602
Our-Fast	1.5019	7.6998	3.1063
	-0.5906	5.8564	1.8784

TABLE III
AVERAGE IMPROVEMENT OF OUR METHOD ON REAL DATA.

Methods	PSNR (dB)	VSNR (dB)	TCDR (dB)
Our-Sparse	0.1112	0.6211	0.2987
Our-Fast	-2.7471	2.3086	0.7483

fast method has severe boundary effects, because it uses FFT and IFFT to accelerate the computation. The sparse method and our method have high-quality results, while our method is slightly better. The complete evaluation values are shown in Fig. 9, and the average improvements of our method to the other two methods are shown in Table II. By PSNR only, it is not convincing to tell which method is better. By perceptual metrics (both VSNR and TCDR), our method is slightly better than the sparse method, and much better than the fast method. It demonstrates that our method has highest visual quality with artifacts and noise being considered. Moreover, the improvement of our method over the sparse method increases when the noise becomes larger.

C. Real Data

For real data, we utilize the dataset in [8]. This dataset contains 32 images obtained from four photos and 8 kernels. The photos are nailed to a large board, and the ground truth is obtained by mounting the camera on a tripod. The kernels estimated by minimizing $\|k \otimes x - y\|^2$ have some noise, which challenges the deconvolution. The image resolution is 255×255, and the kernel size varies from 10 to 25 pixels. Therefore, the blurries are very large in this dataset. Similar to the experiment of synthetic data, parameters are tuned on

one image in each group with the same kernel, and applied to the rest of the images in the group. Fig. 10 shows two experimental examples on real data. The complete evaluation values are shown in Fig. 11, and the average improvements of our method over the other two methods are documented in Table III. Again, the result by PSNR does not clearly indicate which method is indeed better. For most cases, the fast method reaches higher PSNRs than the other two, which is consistent with the evaluation results in [6]. However, by perceptual metrics, the fast method has lower visual qualities due to severe artifacts. Our method achieves highest VSNRs and TCDRs in all cases, which indicates highest visual fidelity. This perceptual experiment further demonstrates that our method has high visual quality.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have developed a non-blind deconvolution method using a multi-resolution sparse prior and the multistage convex relaxation. Our method enhances the robustness of deconvolution, especially when strong edges and large errors are presented. We have also proposed a new perceptual metric for measuring image deconvolution in a quantitative way. This metric, equipped with fast computation, has strong responses on artifacts in deblurring. From a new perspective of quantifying image deconvolution, we have conducted comprehensive perceptual evaluations using VSNR and TCDR. The
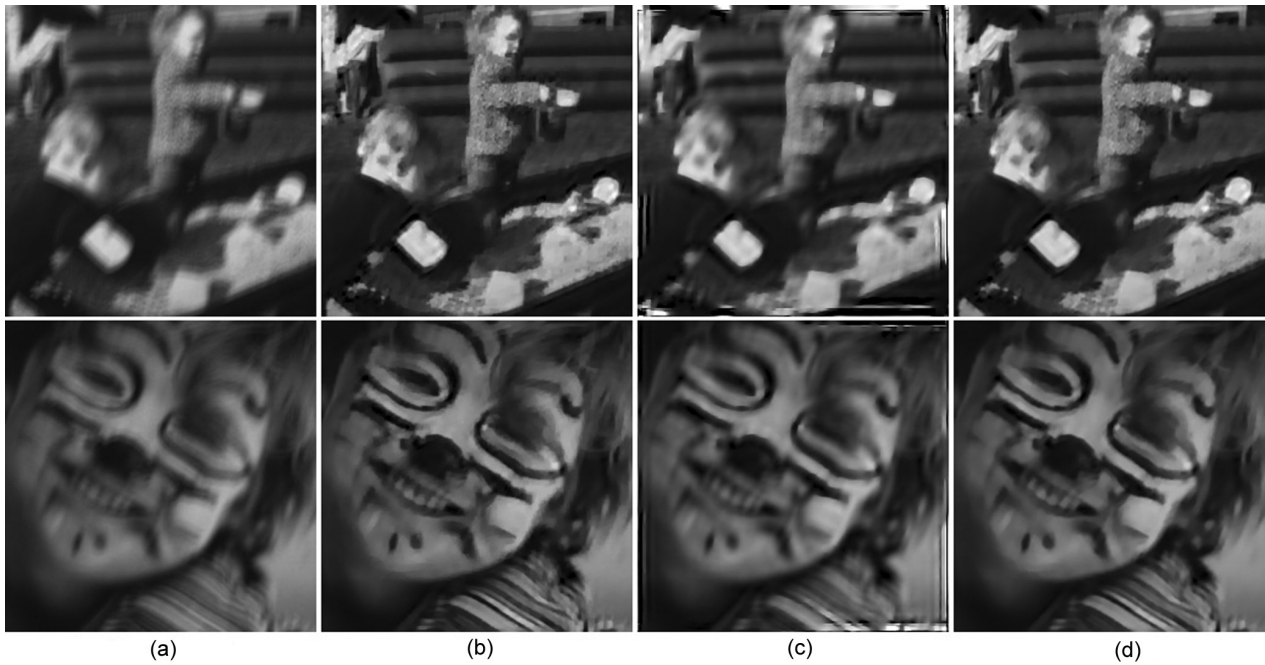
Fig. 10.   Two examples of the experiment on real data: (a) blurred images, (b) sparse method, (c) fast method, and (d) ours. Detailed evaluation values and the statistical analysis can be found in Fig. 11 and Tab. III.
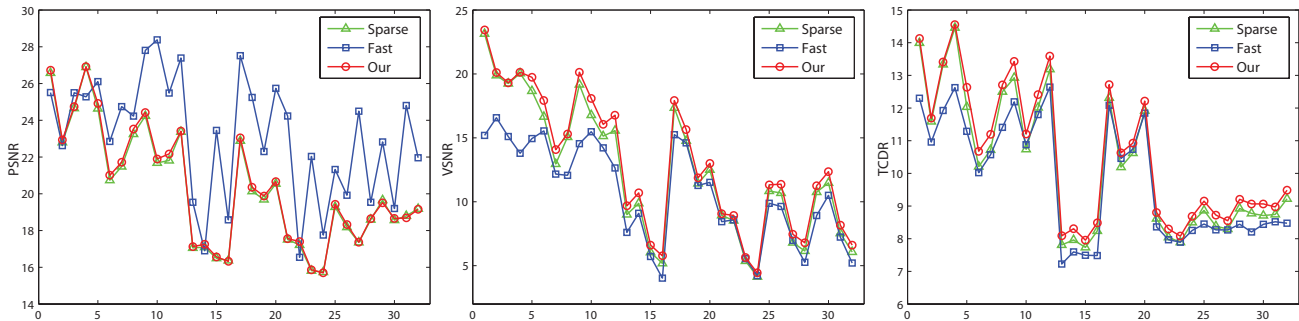

Fig. 11.   Evaluation values on real data. The average improvements of our method over the other two methods are shown in Table III.

evaluation results on both synthetic and real data demonstrate that our new method indeed improves the visual quality with low distortion and small artifacts.

In the near future, we plan to further improve our deconvolution algorithm via GPU acceleration to achieve real-time computation. We will also continue to investigate the stability of deconvolution methods from other aspects, such as shift-variant kernels.

ACKNOWLEDGMENTS

REFERENCES

[1] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, no. 3, 2006.

[2] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, 2007.

[3] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Trans. Graph.*, vol. 27, no. 3, 2008.

[4] N. Joshi, C. L. Zitnick, R. Szeliski, and D. J. Kriegman, "Image deblurring and denoising using color priors," in *CVPR*, 2009, pp. 1550–1557.

[5] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "Blind motion deblurring from a single image using sparse approximation," in *CVPR*, 2009, pp. 104–111.

[6] D. Krishnan and R. Fergus, "Fast image deconvolution using hyperlaplacian priors," in *NIPS*, 2009.

[7] T. S. Cho, N. Joshi, C. L. Zitnick, S. B. Kang, R. Szeliski, and W. T. Freeman, "A content-aware image prior," in *CVPR*, 2010, pp. 169–176.

[8] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *CVPR*, 2009, pp. 1964–1971.

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

[10] 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 Processing*, vol. 14, no. 12, pp. 2117–2128, 2005.

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

[12] L. Lucy, "Bayesian-based iterative method of image restoration," *Journal of Astronomy*, vol. 79, pp. 745–754, 1974.

[13] A. Tikhonov, "On the stability of inverse problems," *Doklady Akademii Nauk SSSR*, vol. 39, no. 5, pp. 195–198, 1943.

[14] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 367-383, p. 3, 1992.

[15] T. F. Chan and C.-K. Wong, "Total variation on blind deconvolution," *IEEE Trans. Image Processing*, vol. 7, no. 3, pp. 370–375, 1998.

[16] Y.-L. You and M. Kaveh, "Blind image restoration by anisotropic regularization," *IEEE Trans. Image Processing*, vol. 8, no. 3, pp. 396–407, 1998.

[17] J. M. Bioucas-Dias, M. A. T. Figueiredo, and J. P. Oliveira, "Total variation-based image deconvolution: a majorization-minimization approach," in *ICASSP*, vol. 2, 2006, pp. 861–864.

[18] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Processing*, vol. 4, no. 7, pp. 932–946, 1995.

[19] L. Yuan, J. Sun, L. Quan, and H. Y. Shum, "Progressive inter-scale and intra-scale non-blind image deconvolution," *ACM Trans. Graph.*, vol. 27, no. 3, 2008.

[20] S. Cho and S. Lee, "Fast motion deblurring," *ACM Trans. Graph.*, vol. 28, no. 5, 2009.

[21] J. Chen, L. Yuan, C.-K. Tang, and L. Quan, "Robust dual motion deblurring," in *CVPR*, 2008, pp. 1–8.

[22] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "High-quality curvelet-based motion deblurring from an image pair," in *CVPR*, 2009, pp. 1566–1573.

[23] O. Whyte, J. Sivc, A. Zisserman, and J. Ponce, "Non-uniform deblurring for shaken images," in *CVPR*, 2010, pp. 491–498.

[24] M. Ben-Ezra and S. Nayar, "Motion deblurring using hybrid imaging," in *CVPR*, 2003, pp. 657–664.

[25] R. Raskar, A. Agrawal, and J. Tumblin, "Coded exposure photography: Motion deblurring using fluttered shutter," *ACM Trans. Graph.*, vol. 25, no. 3, 2006.

[26] L. Yuan, J. Sun, L. Quan, and H. Shum, "Image deblurring with blurred/noisy image pairs," *ACM Trans. Graph.*, vol. 26, no. 3, 2007.

[27] C. Zhou and S. Nayar, "What are good apertures for defocus deblurring?" in *ICCP*, 2009, pp. 1–8.

[28] A. Levin, P. Sand, T. S. Cho, F. Durand, and W. T. Freeman, "Motion-invariant photography," *ACM Trans. Graph.*, vol. 27, no. 3, 2008.

[29] M. Hirsch, S. Sra, B. Schölkopf, and S. Harmeling, "Efficient filter flow for space-variant multiframe blind deconvolution," in *CVPR*, 2010, pp. 607–614.

[30] N. Joshi, S. B. Kang, C. L. Zitnick, and R. Szeliski, "Image deblurring using inertial measurement sensors," *ACM Trans. Graph.*, vol. 29, no. 3, 2010.

[31] Q. Shan, Z. Li, J. Jia, and C.-K. Tang, "Fast image/video upsampling," *ACM Trans. Graph.*, vol. 27, no. 5, 2008.

[32] T. Hou, S. Wang, H. Qin, and R. L. Miller, "Image deconvolution using multigrid natural image prior and its applications," in *ICIP*, 2010, pp. 3569–3572.

[33] F. Lukas and Z. Budrikis, "Picture quality prediction based on a visual model," *IEEE Trans. Communications*, vol. 30, no. 7, pp. 1679–1692, 1982.

[34] S. J. P. Westen, R. L. Lagendijk, and J. Biemond, "Perceptual image quality based on a multiple channel hvs model," in *ICASSP*, vol. 4, 1995, pp. 2351–2354.

[35] S. Winkler, "A perceptual distortion metric for digital color images," in *ICIP*, vol. 3, 1998, pp. 399–403.

[36] A. Bradley, "A wavelet visible difference predictor," *IEEE Trans. Image Processing*, vol. 8, no. 5, pp. 717–730, 1999.

[37] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *ICIP*, vol. 3, 2002, pp. 57–60.

[38] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "A perceptual framework for contrast processing of high dynamic range images," *ACM Trans. Applied Perception*, vol. 3, no. 3, pp. 286–308, 2006.

[39] R. Mantiuk, S. Daly, and L. Kerofsky, "Display adaptive tone mapping," *ACM Trans. Graph.*, vol. 27, no. 3, 2008.

[40] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Dynamic range independent image quality assessment," *ACM Trans. Graph.*, vol. 27, no. 3, 2008.

[41] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *ICCV*, 1998, pp. 839–846.

[42] T. Zhang, "Analysis of multi-stage convex relaxation for sparse regularization," *Journal of Machine Learning Research*, vol. 11, pp. 1081–1107, 2010.

[43] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.

[44] H. Wilson, "A transducer function for threshold and suprathreshold human vision," *Biological Cybernetics*, vol. 38, pp. 171–178, 1980.

[45] P. Whittle, "Increments and decrements: Luminance discrimination," *Vision Research*, vol. 26, no. 10, pp. 1677–1691, 1986.

[46] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *ICCV*, vol. 2, 2001, pp. 416–423.
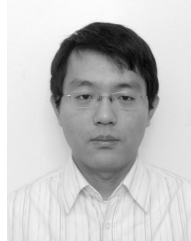
Tingbo Hou is a Ph.D. candidate in Department of Computer Science at Stony Brook University (SUNY Stony Brook), USA. He received his B.S. degree from University of Science and Technology of China in 2004, and his M.E. degree from Chinese Academy of Sciences in 2007. In the summer of 2009, he worked as a research intern in Kodak Research Laboratories, Rochester. His research interests span across visual computing, shape analysis, shape matching/registration, image processing, and geometric modeling.

Sen Wang is a senior research scientist in Kodak Research Laboratories, Eastman Kodak Company. He received his Ph.D. degree in Computer Science from Stony Brook University in 2008 and a M.S. degree in 2006. He also received a M.E. degree from Chinese Academy of Science in 2003, and a B.S. degree from Shandong University in 2000. Currently, Dr. Wang serves as an associate editor for the Journal of Mathematical Imaging and Vision. He is also on the editorial board of Journal of Real-time Image Processing and Journal of Personal and Ubiquitous Computing. His main interests include Computer Vision, Computer Graphics, Stereo/3D Imaging, Biometrics, Image/Video Processing, and Human Computer Interaction. Dr. Wang has more than 20 research papers and 30 patent applications in the above areas.

Hong Qin is a professor of Computer Science in Department of Computer Science at State University of New York at Stony Brook (Stony Brook University). He received his B.S. degree and his M.S. degree in Computer Science from Peking University, China. He received his Ph.D. (1995) degree in Computer Science from the University of Toronto. During his years at the University of Toronto (UofT), he received UofT Open Doctoral Fellowship. He was also a recipient of NSF CAREER Award from the National Science Foundation (NSF), Honda Initiation Award, and Alfred P. Sloan Research Fellow by the Sloan Foundation. Currently, he serves as an associate editor for The Visual Computer, Graphical Models, and Journal of Computer Science and Technology. His research interests include geometric and solid modeling, graphics, physics-based modeling and simulation, computer aided geometric design, human-computer interaction, visualization, and scientific computing. Detailed information about Dr. Hong Qin can be found from his website: http://www.cs.sunysb.edu/~qin. He can be reached at qin@cs.sunysb.edu.