A Comparative Study for Single Image Blind Deblurring

Wei-Sheng Lai[1] Jia-Bin Huang[2] Zhe Hu[1] Narendra Ahuja[2] Ming-Hsuan Yang[1]
[1]University of California, Merced [2]University of Illinois, Urbana-Champaign
{wslai.cv@gmail.com} {zhehu@ieee.org} {mhyang@ieee.org}
{ahuja@illinois.edu} {jbhuang1@illinois.edu}

Abstract

Numerous single image blind deblurring algorithms have proposed in various loss/regularized image/kernel/camera models. Instead, these algorithms are mainly evaluated on a few images with ground truth or real blurred images. It is unclear whether top three algorithms would perform on images acquired "in the wild" and how the performance is affected by the factors such as noise or image size, etc. We present in this comprehensive perceptual study and analysis of single image blind deblurring using real-world blurred images. First, we collect a dataset of real blurred images and a dataset of synthetic blurred images. Using these two datasets, we then evaluate large scale user studies to quantify the performance of several representative state-of-the-art blind deblurring algorithms. We also present a comprehensive analysis of these results, including the lack of agreement, significant role of noise differences, and robustness for top performing methods (see another: What we study the correlation between human subjective ratings and several full-reference and no-reference image quality metrics. Our analysis also indicates the performance gap between synthetically blurred images and real blurred images, and which top-performing methods perform well on real-world blurring.

Figure 1. Deblurring results of a blurred image from the dataset in Lai et al. [29]. The results of Cho and Lee [5], Xu and Jia [38], Levin et al. [22], Sun et al. [34], Michaeli and Irani [28], Zhong et al. [44], Pan et al. [31], and Yan et al. [41] are jointly ranked the 1st by the 11 blind image quality metrics in Table 3. The deblurring results of these top-ranked algorithms are qualitatively similar and are free of ringing artifacts. Which one do you prefer?

1. Introduction

The recent years have witnessed significant progress in single image blind deblurring (or motion deblurring). The seminal work of Fergus et al. [12] inspires a new line of efforts on various objectives/priors [5, 7, 17, 29, 34], various kernel/image priors [16, 27, 28, 38, 44], and more powerful optimization [4, 31, 41, 43]. To quantify and compare the performance of competing algorithms, existing work mainly relies on the ground-truth data of Levin et al. [21] (which uses synthetic uniform kernels) or a few images with known blur kernels, or (2) a non-uniform blurred benchmark with ground-truth latent images and blur kernels recorded in a controlled setting. However, existing datasets do not consider several critical factors, e.g., scene depth variation, camera orientation and motion, camera sensor noise, lens artifacts, and other unknown factors in the image formation process. Thus, the generalization of these top-performing algorithms to real-world blurred images and their relative performance are not well understood. It is unclear how the top-performing algorithms perform on images taken "in the wild" and how robust these algorithms are to various factors such as noise or image size, etc. We present in this comprehensive perceptual study and analysis of single image blind deblurring using real-world blurred images. First, we collect a dataset of real blurred images and a dataset of synthetic blurred images. Using these two datasets, we then evaluate large scale user studies to quantify the performance of several representative state-of-the-art blind deblurring algorithms. We also present a comprehensive analysis of these results, including the lack of agreements, significant role of noise differences, and robustness for top performing methods (see another). We also study the correlation between human subjective ratings and several full-reference and no-reference image quality metrics. Our analysis also indicates the performance gap between synthetically blurred images and real blurred images, and which top-performing methods perform well on real-world blurring.

Given the large number of blind deblurring methods, it is impossible to evaluate the relative merits by a follow-up work. To ease the study, we focus on recent state-of-the-art single image blind deblurring algorithms from the last blind deblurring benchmark [29] and the recent work [44]. We also include the seminal work of Fergus et al. [12] for a baseline. We hope that this study will provide a baseline for future research on single image blind deblurring.

The contributions of this work are threefold:

- We collect a dataset of real blurred images and a dataset of synthetically blurred images using top-performing algorithms from the last blind deblurring benchmark [29] and the recent work [44]. These datasets and the code for quantitative evaluation are available at: http://vllab.ucmerced.edu/wslai/blur.html.
- We conduct extensive user studies to quantify the performance of several representative state-of-the-art blind deblurring algorithms. We also present comprehensive analysis of these results, including the lack of agreement, significant role of noise differences, and robustness for top performing methods (see another).
- We study the correlation between human subjective ratings and several full-reference and no-reference image quality metrics. Our analysis also indicates the performance gap between synthetically blurred images and real blurred images, and which top-performing methods perform well on real-world blurring.

2. Related Work

Datasets and benchmarks for image deblurring. Several datasets have been used to evaluate and compare the performance of image deblurring algorithms. Levin et al. [21] collect 32 images and 8 blur kernels, generating 32×8 blurred images. Sun et al. [34] use the dataset of Levin et al. [21] and a uniform motion blur model to generate 4 images with 112×112 pixels. However, the five grayscale images coupled with spatial invariance at 256×256 pixels are not sufficient to cover a wide variety of real-world scenarios. Yan et al. [41] propose a dataset of 48 images (12 scenes × 4 blur kernels) of indoor scenes, and synthetically blurring each image with 32 depth-dependent blur kernels. Recently, a benchmark dataset of 4 images (2 indoor and 2 outdoor scenes) with ground-truth sharp images has been proposed for evaluating deblurring or image kernel estimation [29]. Recently, Zhong et al. [44] propose a dataset of 10 images (5 indoor and 5 outdoor scenes) with 6 blur kernels. Although the dataset of Zhong et al. [44] is the largest, it is still limited in the number of scenes and lacks ground-truth latent images. Unlike the above-mentioned datasets, Kim et al. [16] propose a dynamic scene deblurring dataset with 6 sequences of 720×480 pixel images. However, the ground-truth latent images are not available, and the evaluation is based on the deep learning model. Recently, Hu et al. [14] propose a dataset of 25 images with ground-truth latent images. However, the blur kernels are not spatially invariant, and the ground-truth latent images are not sharp. In this work, we propose two datasets with ground-truth latent images for real and synthetic blurred images, and we study the performance of several state-of-the-art algorithms using these datasets.

Deblurring algorithms. Early work on single image blind deblurring relies on strong assumptions on the blur kernel and/or the latent image. Fergus et al. [12] propose a variational Bayes (VB) method with a mixture of gaussians model as the prior on the gradients of the latent image. Shan et al. [32] propose a unified probabilistic model for the blur kernel, the latent image, and the noise. Cho and Lee [5] propose a fast deblurring method based on a series of well-motivated steps involving prediction of sharp edges, kernel estimation, and non-blind deconvolution. Xu and Jia [38] analyze the effectiveness of some of these steps and propose a simple and effective formulation. Levin et al. [22] show that the MAP formulation can be favored for blind deblurring if a sparse gradient prior is used to model the latent image. Perrone and Favaro [33] show that the success of the MAP formulation proposed by Levin et al. [22] depends on the choice of the initialization and how the kernel is updated. Recently, Zhong et al. [44] show that the motion blur can be accurately estimated from a single image using a linear classifier with motion-invariant features. Chakrabarti [4] proposes a deep neural network to predict the Fourier coefficients of a deblurring filter for a fixed window of the input image. Sun et al. [34] use a deep network to estimate a spatially invariant motion blur kernel. Nah et al. [30] propose a multi-scale deep network to jointly estimate the latent image and the blur kernel. Yan et al. [41] propose to jointly estimate the blur kernel and the salient edges for deblurring. Pan et al. [31] show that the dark channel prior can be used to deblur images with various scenes and blur kernels. Zhong et al. [44] propose a fast method based on the conjugate gradient algorithm. Ren et al. [43] propose a deep neural network to jointly estimate the segmentation, the motion blur field, and the latent image. Kim et al. [16] propose a dynamic scene deblurring method based on a recurrent neural network. Recently, Hu et al. [14] propose a method based on the statistics of local image patches. Although there are many deblurring algorithms, most of them are evaluated on a few images with ground-truth or real blurred images. It is unclear how these top-performing algorithms perform on other real-world blurred images and how robust these algorithms are to various factors such as noise, image size, etc.

Synthetic datasets. To examine the performance, some work utilizes real and synthetic images. We extract 20 sharp images from the Berkeley segmentation dataset (BSDS) [26] and use the 8 blur kernels from [21] to generate 20×8×3 blurred images for each method. We then follow the same protocol as [29] to select 25 patches from the 160 blurred images for each method. We also use the 4 blur kernels from [29] to generate 20×4×3 blurred images. We follow the same protocol to select 25 patches from the 120 blurred images for each method. We use the 8 blur kernels from [21] to generate 20×8×3 blurred images. We follow the same protocol to select 25 patches from the 160 blurred images for each method. We