

# Joint Depth Estimation and Camera Shake Removal from Single Blurry Image

Zhe Hu<sup>1</sup>   Li Xu<sup>2</sup>   Ming-Hsuan Yang<sup>1</sup>

<sup>1</sup> University of California, Merced   <sup>2</sup> Image and Visual Computing Lab, Lenovo R & T

zhu@ucmerced.edu, xulihk@lenovo.com, mhyang@ucmerced.edu

## Abstract

Camera shake during exposure time often results in spatially variant blur effect of the image. The non-uniform blur effect is not only caused by the camera motion, but also the depth variation of the scene. The objects close to the camera sensors are likely to appear more blurry than those at a distance in such cases. However, recent non-uniform deblurring methods do not explicitly consider the depth factor or assume fronto-parallel scenes with constant depth for simplicity. While single image non-uniform deblurring is a challenging problem, the blurry results in fact contain depth information which can be exploited. We propose to jointly estimate scene depth and remove non-uniform blur caused by camera motion via exploiting their underlying geometric relationships, with only single blurry image as input. To this end, we present a unified layer-based model for depth-involved deblurring. We provide a novel layer-based solution using matting to partition the layers and an expectation-maximization scheme to solve this problem. This approach largely reduces the number of unknowns and makes the problem tractable. Experiments on challenging examples demonstrate that both depth and camera shake removal can be well addressed within the unified framework.

## 1. Introduction

Motion deblurring problem has drawn much attention in recent years with demonstrated success. Most approaches model a blurry image as the integration of intermediate frames captured by the camera along the motion trajectory, typically assuming in-plane translational camera motions. Recent endeavor in nonuniform deblurring further explores rotational camera motion model [6, 30, 28] and broadens the application range of deblurring approaches. These effective methods, however, do not consider large scene depth variation, which is ubiquitous in real-world scenes.

Figure 1 shows one outdoor image taken by a hand-held camera. Almost all existing deblurring methods do not perform well in this seemingly simple example due to the non-ignorable depth variation. While some algorithms on depth-

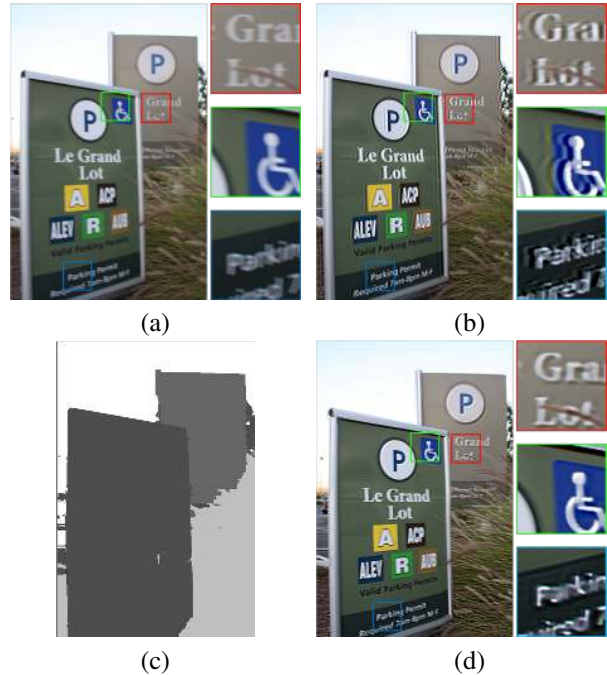


Figure 1. Depth-aware non-uniform deblurring. (a) blurry image; (b) restored image by the method [10] assuming constant scene depth; (c) estimated depth map by our method; (d) restored image by our method using the estimated depth map.

aware motion deblur have been recent proposed [31, 21], they nevertheless leverage additional observations for this challenging task. The inherent limitation of the existing deblurring methods to handle depth variation is readily explainable. On the one hand, even with known scene depth, the solution space of the deblurring problem with depth consideration is typically huge, making it not suitable for practical applications. On the other hand, depth information is not usually available and the depth inference from one single image is extremely difficult.

In this paper, we propose a unified framework for jointly restoration of scene depth and the latent clear image. While depth information is critical in blur removal, the blur image provides an additional cue for depth information. In addition, rather than inferring depth for isolated pixels [18],

we extract depth layers and model the layer-wise blur respectively. This makes the problem better constrained and is the major difference to the previous work. We develop an expectation-maximization (EM) scheme to solve scene depth and the latent image alternatively by exploiting their underlying geometric relationships, from one single blurry image. We demonstrate that the proposed model effectively removes non-uniform blur from images with depth variation.

## 2. Related Work

The problem of removing spatially invariant blur has been studied extensively and numerous algorithms have been proposed. Since blind deconvolution is an ill-posed problem, prior knowledge or additional information are often required for effective solutions. In the image deblurring literature, two types of additional information are often used: natural image priors [8, 17, 24, 19, 5, 7] and additional image observations [2, 22, 33, 27, 20].

Recently, the problem of removing spatially variant blur caused by camera motion has attracted much attention due to its wide range of applications [1, 6, 25, 27, 13, 30, 9, 10, 28, 29, 11, 12]. Early work splits an image into several regions where each is modeled with a constant PSF and recovered by a uniform deblur algorithm [1, 6]. However, the images are degraded from spatially variant blur even in the same region as shown in Figure 1. These methods are sensitive to region size which is required to be sufficiently small for the assumption of uniform PSFs to be hold. Hybrid camera systems have also been used to capture a local blur kernel at each pixel location based on optical flows [27, 13]. The methods with additional inputs are demonstrated to be effective on image deblurring, but such approaches are not applicable to single image non-uniform deblurring.

In [30, 28], geometric models are proposed to model the observed blurry image as the integration of all the intermediate images captured by the camera along the motion trajectory. The intermediate images are modeled as the homographic transformations of the latent image. Based on this model, the spatially-varying deblurring problem can be solved by estimating a general camera motion function and then the latent image. A similar model has been used to model three degrees of camera motion [9] (with in-plane translations and rotation) instead of camera rotations in [30]. To reduce the computational loads, fast non-uniform deblur methods have been developed to speed up the optimization step [10, 11]. In [14], a dataset for motion deblurring is constructed by recording and playing back the camera shake with a robot system. The blurry images are obtained by taking photos on sharp pictures which assume constant depth. Notwithstanding the demonstrated success, the above-mentioned non-uniform deblurring methods do not consider the effect of scene depth on camera motion

blur or assume constant scene depth for simplicity.

To exploit depth information from the blurry input, recent methods [6, 25, 3] partition an image into foreground/background and remove the blur caused by object motion. Coded aperture method has been proposed for deblurring and decoded images are used to infer pixel-wise depth in a Markov random field [18]. Depth effect on translational blur has also been discussed in [26] where the blur is estimated from the user specified region with approximately constant depth to infer depth. In [31], two blurry images of the same scene are used to infer depth with stereo techniques and further improve the blur removal. Similarly, a non-uniform motion deblurring method has been proposed to exploit bilayer-scene constraints from two blur observations [21]. In this work, we consider single-image non-uniform blur caused by camera motion and scene depth variation. Depth estimation from one single image itself is a difficult problem. A depth estimation method has been proposed based on the visual depth cues, texture and color, in [23]. However, the problem becomes more challenging when a blurry input is given as the depth cues are destroyed in the presence of motion blur. We address this issue by exploiting the underlying geometric properties of a blurry image.

## 3. Depth-Aware Deblurring

In this section, we present the depth-aware geometric model of non-uniform blur caused by camera motion, and propose an EM formulation for joint depth estimation and latent image recovery.

The commonly used geometric model of camera motion considers an observed blurry image  $y$  as the weighted sum of the transformed results based on the sharp image  $x$  [30, 9, 28],

$$y = \sum_{j=1}^t w_j (K(\theta_j)x) + n, \quad (1)$$

where  $\{\theta_j\}_{j=1}^t$  denotes the sampled camera poses, the weights  $W = (w_1, \dots, w_t)^\top$  indicate the exposure time the camera stays at each pose and  $\sum_j w_j = 1$ ,  $K(\theta_j)$  is the matrix that warps the latent image  $x$  to the transformed result at a sampled pose  $\theta_j$  and  $n$  as the observation noise. In this formulation,  $K(\theta_j)$  can also be viewed as a blur kernel basis of pose  $\theta_j$  where the  $r$ -th row is the weight vector representing how the  $r$ -th pixel of the latent image  $x$  contributes to the blurry image. In [9], the matrix  $K(\theta_j)$  is induced from the homography  $P_j$  of the latent image assuming constant depth  $d$ ,

$$P_j = C(R_j + \frac{1}{d}T_j[0, 0, 1])C^{-1}, \quad (2)$$

where  $R_j$  and  $T_j$  denote the rotational and translational components of pose  $\theta_j$ , and  $C$  is the intrinsic matrix of the camera. This equation also shows that depth variation

affects only translational blur. While this model assumes fronto-parallel scenes, we consider more generic cases with varying depth in this work. Considering depth variation, the  $r$ -th row of the warping matrix varies with the depth at the  $r$ -th pixel. Thus, we denote  $K(\theta_j, D)$  as the matrix induced by pose  $\theta_j$  and depth map  $D$ .

For practical applications, recent work simplifies the camera motion from the six-dimensional space to a three-dimensional subspace. In [30], only three rotations are used to approximate the camera motion, while in-plane translations and rotation are considered in [9, 10, 11]. It has been shown that these two models achieve similar performance under some assumptions [14]. In this work, we consider in-plane translations and rotation of camera motion, since out-of-plane rotations can be approximated with in-plane translations at certain depth in the proposed depth-aware model.

To make the problem tractable, we first split the image into disjoint regions,  $x = \sum_{i=1}^N x_i$  for latent image and  $y = \sum_{i=1}^N y_i$  for the blurry image, and assign a constant depth  $d_i$  to each region  $x_i$ . Here we use  $x_i$  to indicate the  $i$ -th region, with the intensity of a pixel same as the one in the original image  $x$  if it belongs to  $i$ -th region and 0 otherwise. We rewrite (1) as

$$y = \sum_{i=1}^N y_i = \sum_{i=1}^N \sum_{j=1}^t w_j (K(\theta_j, d_i) x_i) + n. \quad (3)$$

Given only blurry image, there are numerous unknown variables in (3), which makes it difficult to solve.

### 3.1. EM Formulation

Since there are numerous unknown variables in (3) and the equation cannot be solved directly, we propose an EM formulation to solve the problem. Given the observed image  $y$ , the goal is to estimate the latent image and the weights of camera motion  $\gamma = (x, W)$  in terms of the unknown depth map  $D$ . The maximum likelihood estimate (MLE) of this problem is determined by the marginal likelihood of the observed image,

$$L(\gamma; y) = p(y|\gamma) = \sum_D p(y, D|\gamma). \quad (4)$$

It can be solved with the EM algorithm by iteratively applying the following steps:

**E-step:** Compute the expected value of the log likelihood function under the current estimate of  $\gamma^{(l)}$ ,

$$Q(\gamma|\gamma^{(l)}) = E_{p(D|y, \gamma^{(l)})} [\log p(y, D|\gamma)], \quad (5)$$

**M-step:** Determine  $\gamma^{(l+1)} = \arg \max_{\gamma} Q(\gamma|\gamma^{(l)})$ .

We explain these steps in the following sections.

#### 3.1.1 E-step

To express  $Q(\gamma|\gamma^{(l)})$ , we first consider the likelihood

$$p(y|x, W, D) = \prod_{i=1}^N p(y_i|x_i, W, d_i), \quad (6)$$

in terms of disjoint regions. To simplify the problem, we assume a discrete sampled set of the potential depth values,  $\{v_k\}_{k=1}^s$ , and introduce an auxiliary variable  $\alpha_{i,k}$  indicating the value of depth  $d_i$  as

$$\alpha_{i,k} = \begin{cases} 1, & \text{if } d_i = v_k, \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

and  $\sum_{k=1}^s \alpha_{i,k} = 1$  since only one depth value is selected for each region. Instead of estimating the continuous depth values of  $\{d_i\}_i$ , we resort to finding the values of  $\{\alpha_{i,k}\}$ . The likelihood  $p(y_i|x_i, W, d_i)$  for each region is then expressed as the data fitting term and the auxiliary variable,

$$\begin{aligned} & p(y_i|x_i, W, d_i) \\ &= \exp\left(-\frac{1}{2\eta^2} \left\| y_i - \sum_{j=1}^t w_j K(\theta_j, d_i) x_i \right\|^2\right) \\ &= \exp\left(-\frac{1}{2\eta^2} \sum_{k=1}^s \alpha_{i,k} \left\| y_i - \sum_{j=1}^t w_j K(\theta_j, v_k) x_i \right\|^2\right), \end{aligned} \quad (8)$$

where we assume a zero mean Gaussian prior for signal noise  $n \sim N(0, \eta^2)$ . Here the warping matrix  $K(\theta_j, v_k)$  can be pre-computed once the sampling of camera poses and depth values is determined.

Assuming uniform distribution on  $p(D)$ , we write the log-likelihood  $\log p(y, D|\gamma)$  using (6) and (8),

$$\begin{aligned} & \log p(y, D|x, W) = \log p(y|x, W, D) p(D|x, W) \\ & \propto -\frac{1}{2\eta^2} \sum_{i=1}^N \sum_{k=1}^s \alpha_{i,k} \left\| y_i - \sum_{j=1}^t w_j K(\theta_j, v_k) x_i \right\|^2. \end{aligned} \quad (9)$$

and (5) can be written as

$$\begin{aligned} & E_{p(D|y, (x, W)^{(l)})} [\log p(y, D|x, W)] \\ & \propto -\frac{1}{2\eta^2} \sum_{i=1}^N \sum_{k=1}^s E[\alpha_{i,k}^{(l)}] \left( \left\| y_i - \sum_{j=1}^t w_j K(\theta_j, v_k) x_i \right\|^2 \right). \end{aligned} \quad (10)$$

Here  $E[\alpha_{i,k}^{(l)}]$  denotes the expectation of  $\alpha_{i,k}$  in the  $l$ -th iteration and it can be computed with the pseudo-likelihood

$$\begin{aligned} \bar{\alpha}_{i,k}^{(l)} &= E[\alpha_{i,k}^{(l)}] = p(\alpha_{i,k}^{(l)} = 1 | y_i, x_i^{(l)}, W^{(l)}) \\ &= \frac{\exp\left(-\frac{1}{2\eta^2} \left\| y_i - \sum_{j=1}^t w_j^{(l)} K(\theta_j, v_k) x_i^{(l)} \right\|^2\right)}{\sum_{k'=1}^s \exp\left(-\frac{1}{2\eta^2} \left\| y_i - \sum_{j=1}^t w_j^{(l)} K(\theta_j, v_{k'}) x_i^{(l)} \right\|^2\right)}, \end{aligned} \quad (11)$$

which also satisfies  $\sum_{k=1}^s \bar{\alpha}_{i,k}^{(l)} = 1$ .

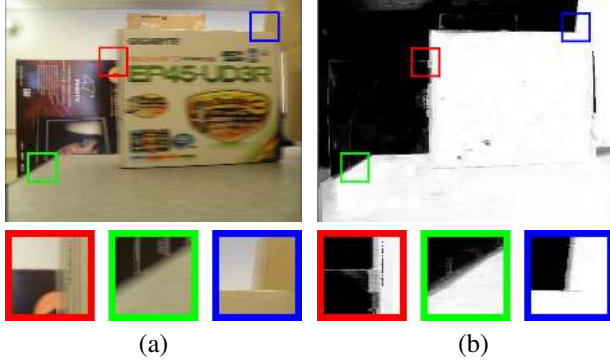


Figure 2. Image partition using matting. (a) blurry image and patches that are difficult for image partition, (b) one layer using matting and corresponding matting patches.

### 3.1.2 M-step

Once all the  $\bar{\alpha}_{i,k}^{(l)}$  are determined, we solve  $\gamma$  in the M-step with the optimization problem,

$$\gamma^{(l+1)} = \arg \max_{(x,W)} \sum_{i=1}^N \sum_{k=1}^s \bar{\alpha}_{i,k}^{(l)} (\|y_i - \sum_{j=1}^t w_j K(\theta_j, v_k) x_i\|^2). \quad (12)$$

This optimization problem is similar to the existing non-uniform methods based on the geometric model, except that the image is partitioned into multiple depth layers. We solve  $x$  and  $W$  alternatively in a way similar to the other non-uniform deblurring algorithms [30, 9, 10, 11] and explain the detailed steps in Section 3.4.

## 3.2. Image Partition

In this work, each blurry image is partitioned into multiple depth layers. There are two main issues for blurry-image partition as shown in Figure 2. One is the ringing artifacts along edges which makes it difficult to segment a blurry image. Another problem is the ambiguity of the region boundary, where is likely to be caused by the combination of objects appear in different layers. To address these issues, we first process a blurry image with a smoothing filter (e.g., Gaussian) to suppress the artifacts. We then apply a matting algorithm [4] on the filtered image to obtain multi-layer segmentation results, with the scribble for each layer from the user. The pixel with matting value less than one is likely to lie on the region boundary and we assign a weight based on the matting value for the data fitting term. We present an example in Figure 2 to show how a blurry image is partitioned into multiple layers.

## 3.3. Depth Estimation

In Section 3.1.1, we discuss how to compute the expectation of  $\alpha_{i,k}$  and use it in the M-step. Considering the adverse effect of the small edges, we apply the  $L_0$  smoothing

filter [32] to the estimated latent image  $x$  before computing  $\bar{\alpha}_{i,k}$ . To determine the depth value of each layer, we discard small values of  $\bar{\alpha}_{i,k}$  and only keep the highest one  $\hat{k}_i = \arg \max_{k=1}^s \bar{\alpha}_{i,k}$ . The depth value corresponding to the highest value  $\bar{\alpha}_{i,k}$  is assigned to the region  $x_i$  and  $y_i$ ,  $d_i = v_{\hat{k}_i}$ . One can also determine the depth by combining several several top  $\bar{\alpha}_{i,k}$  values.

## 3.4. Constant-Depth Non-Uniform Deblurring

In this section, we explain how to solve for  $x$  and  $W$  by fixing depth  $D$  (we abuse the notation to indicate  $\{\bar{\alpha}_{i,k}\}_{i,k}$  in (12)). We adopt the same formulation as existing non-uniform deblurring methods in an alternating way, i.e., solving one by fixing the others. At the first iteration, we initialize the weights  $W$  using backprojection as [11] to facilitate the convergence.

**Image Prediction:** Before the alternative optimization, we predict the approximate latent image using a bilateral filter and a shock filter similar to [5]. The bilateral filter is first applied to reduce noise and remove small edges, and then the shock filter is used to enhance strong edges. By doing so, the predicted image  $\tilde{x} = \sum_{i=1}^N \tilde{x}_i$  contains less artifacts and sharp edges, which leads to more accurate estimate of pose weights  $W$  as the predicted image resembles the latent image.

**Weight Estimation:** Image derivatives have been shown to be effective for suppressing ringing artifacts in deblurring problems [24, 5]. Thus, we replace the image intensity with derivatives in the data fitting term and express the weight estimation as

$$\begin{aligned} \bar{W} &= \arg \max_W \sum_{i=1}^N \sum_{k=1}^s \bar{\alpha}_{i,k} (\|\partial y_i - \sum_{j=1}^t w_j K(\theta_j, v_k) \partial \tilde{x}_i\|^2) \\ &= \arg \max_W \sum_{i=1}^N (\|\partial y_i - \sum_{j=1}^t w_j K(\theta_j, v_{\hat{k}_i}) \partial \tilde{x}_i\|^2), \end{aligned} \quad (13)$$

where  $\partial \in \{\partial_x, \partial_y\}$  represents the derivatives along horizontal and vertical directions. Since  $W$  represents the weights of camera motion trajectory, we add a smoothness term to enforce continuity and write (13) in a matrix form,

$$\bar{W} = \arg \max_W \|AW - b\|^2 + \lambda \|\Gamma W\|^2, \quad (14)$$

where  $A$  is the matrix with  $K(\theta_j, v_{\hat{k}_i}) \partial \tilde{x}_i$  at index  $(i, j)$  and  $b = (\partial y_1, \dots, \partial y_N)^\top$ . Here we denote  $\Gamma$  as the matrix form of the partial operator  $\partial$ . To address computational issues, we use a similar formulation as [11] by solving it with a subset of the camera pose space. The subset is determined by perturbing the one from the last iteration. We use the conjugate gradient method to minimize (14) and normalize  $W$  with unit sum.



**Latent Image Estimation:** To recover the latent image, we solve the following optimization problem,

$$\begin{aligned} \bar{x} &= \arg \max_x \sum_{i=1}^N \sum_{k=1}^s \bar{\alpha}_{i,k} (\|\partial y_i - \sum_{j=1}^t w_j K(\theta_j, v_k) \partial x_i\|^2) \\ &= \arg \max_x \sum_{i=1}^N (\|\partial y_i - \sum_{j=1}^t w_j K(\theta_j, v_{\hat{k}_i}) \partial x_i\|^2). \end{aligned} \quad (15)$$

Since  $\{x_i\}_{i=1}^N$  are disjoint regions, we solve  $x_i$  independently with constraints,

$$\bar{x}_i = \arg \max_{x_i} (\|\partial y_i - \sum_{j=1}^t w_j K(\theta_j, v_{\hat{k}_i}) \partial x_i\|^2 + \mu \|\partial x_i\|), \quad (16)$$

where we use a simple regularization to favor smooth gradients. For fast optimization, we adopt the strategy of splitting the image into grids and performing uniform deconvolution on each region, similar to [10]. It is also possible to use other constraints or deconvolution schemes [18, 24, 15, 16] to solve (15). For the final deconvolution, we use the method with sparse constraints [15] to recover the image.

### 3.5. Initialization and Implementation Details

We summarize the main steps of the proposed method in Algorithm 1. In this work, we set the parameters  $\lambda = 0.5$  and  $\mu = 0.1$  in all experiments.

In the beginning, we initialize the depth of each layer to be the same value. Thus, the first iteration can be viewed as the non-uniform deblurring without depth variation. After applying  $L_0$  smoothing filter [32], the estimated latent image results in a good depth estimation. Figure 3 shows an synthetic example of the depth estimation in the first iteration. The proposed method provides a decent depth estimation except the sky region, which does not have much texture and therefore less important for image deblurring. With proper depth estimation in the first iteration, the proposed method converges to a solution efficiently.

---

#### Algorithm 1 Depth-aware non-uniform deblur

---

**Input:** blurry image  $y$   
**Output:** sharp image  $x$  and depth map  $D$   
**Algorithm:**  
 split the image into disjoint regions  
 initialize depth map  $D$   
**while**  $(x, W)$  do not converge (within a tolerance) **do**  
   **while**  $x$  do not converge given  $D$  **do**  
     latent image prediction  
     update  $W$  using (13)  
     update  $x_i, \forall i$  using (16)  
   **end while**  
   update depth map  $D$  using (11)  
**end while**

---

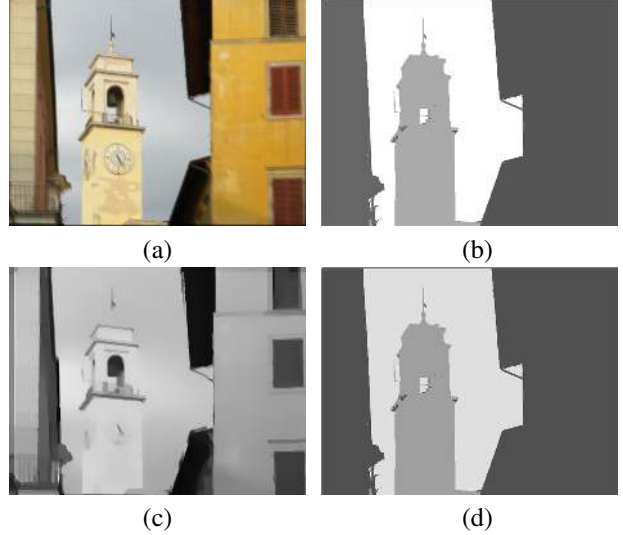


Figure 3. Synthetic example of the depth estimation in the first iteration. Dark region indicates small depth. (a) synthetic blurry image; (b) ground truth depth; (c) estimated latent image after smoothing filter; (d) our estimated depth using (c).

## 4. Experimental Results

In this section, we evaluate depth estimation results of the proposed algorithm and compare deblurring performance against several state-of-the-art methods for spatially variant blur. Implemented in MATLAB, it takes about 2.5 hours to process a blurry image of  $480 \times 640$  pixels on a 3.40 GHz machine with 16 GB RAM. More experimental results and the MATLAB code are available at [http://eng.ucmerced.edu/people/zhu/cvpr14\\_depthdeblur](http://eng.ucmerced.edu/people/zhu/cvpr14_depthdeblur).

### 4.1. Depth Estimation

Depth estimation from a single image is a difficult task. The recent approach [23] proposes a MRF-based method to infer scene depth from visual cues, e.g., texture and color. In Figure 4, we present an example of natural blurry image and the depth estimation by [23] and our approach. Although our method does not capture subtle depth variation, it estimates the rough depth well based on the geometry information contained in the blurry image. In contrast, the method [23] is not effective in estimating depth from blurry images where region texture and edge contrast are missing as a result of motion blur.

We evaluate our depth estimation algorithm on synthetic images with spatially variant blurs. First, we collect a set of sharp images and their depth maps (i.e., indoor images with Kinect sensors and outdoor scenes with laser scanners). We partition the image into different layers based on the depth map and assign the average depth to each layer. Synthetic camera motion trajectories and corresponding weights are

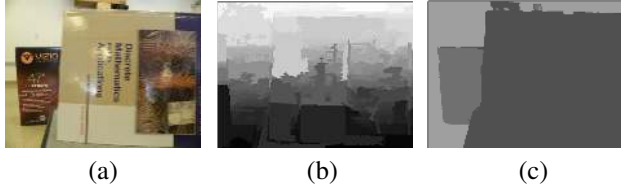


Figure 4. Comparison with depth-from-single-image method [23] on a real image. (a) blurry image; (b) estimated depth by [23]; (c) our estimated depth.

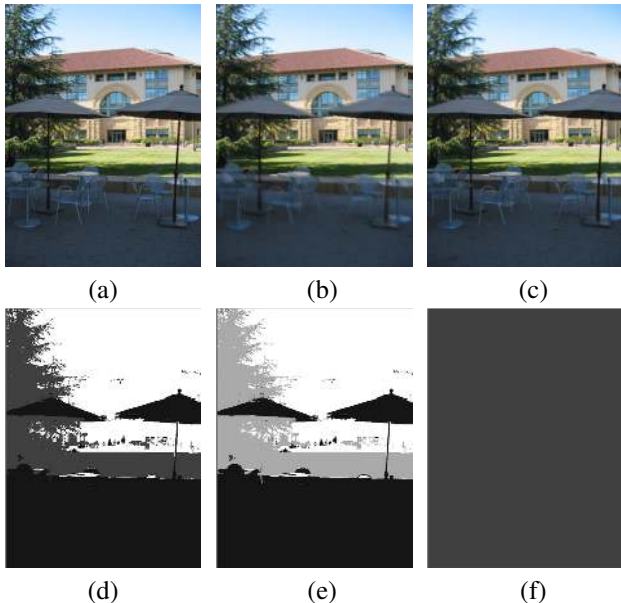


Figure 5. Depth estimation results on synthetic images. (a) sharp image; (b) synthesized blurry image caused by both translational and rotational blur; (c) synthesized blurry image caused by only rotational blur; (d) ground-truth depth; (e) our estimated depth from (b); (f) our estimated depth from (c).

then applied to generate blurry images. For experiments on depth estimation, 420 pairs of blurry and sharp images generated from 6 sharp images (3 indoor and 3 outdoor images from [23]), and 70 randomly synthesized camera motion are collected. Each sharp image has the depth variation from 2 meters to 20 meters with at least three depth layers. The average error per pixel is 2 meters and most errors are from the cases with small camera translations. As demonstrated in Table 1, our depth estimation becomes more accurate when there exists more translational camera motion. Figure 5 shows two depth estimation results by the proposed algorithm, from one blurry image caused by both translational and rotational camera motion and another caused only by rotational blur. With only rotational blur, our method estimates a uniform depth value for each layer as shown in Figure 5(f). The results, although not accurate with respect to depth estimation, satisfy our model that depth variation does not affect rotational motion blur as mentioned in Sec-

Table 1. Average error with respect to the minimal number of translated pixels in the camera motion.

# of pixels	1	3	5	7	9	11
error (meter)	2.02	1.67	1.47	1.35	1.10	1.03

tion 3. This further demonstrates that our model is able to infer the geometric information contained in motion blur.

## 4.2. Non-Uniform Deblurring

In this section, we compare our methods with several state-of-the-art single image deblurring methods [9, 10, 31]. The experiments on the synthetic blurry images and real blurry images are carried out to validate our model and EM algorithm. We also show that our method can be used to remove non-uniform blur caused by camera motion on the scene of small depth variation. Due to space limitations, only a few results are presented in this section and more results can be found in the supplementary material. We note that the dataset [14] is constructed by taking blurry photos on sharp pictures assuming constant depth, and thus it is not suitable for evaluating our depth-aware method.

**Synthetic Blurry Images:** We evaluate deblurring algorithms on the synthetic images obtained using the same process as described in Section 4.1. As shown in Figure 6(a), there are more than two objects at different depths in the synthetic image, and the objects in the back are less blurry than those in the front. The state-of-the-art non-uniform deblurring method based on constant depth [10] does not work well (Figure 6(b)) since the data-fitting terms of different objects penalize each other in weight estimation of camera motion. However, this problem can be alleviated by considering depth factor as formulated in this work. In Figure 6(c) and (d), we demonstrate that the proposed EM formulation generates reasonable depth estimation and improves the deblurred results. We select the text regions to better visualize the difference.

If the image contains small depth variation, our method naturally falls back to constant-depth non-uniform deblurring. We present an example in the supplementary material.

**Real Blurry Images:** In these experiments, we evaluate the proposed method on the real-world images blurred by hand-shake. Figure 7 shows some results with comparisons to the state-of-the-art methods for spatially variant blurs [10, 31, 11]. In the *Books* example, several books and a cookie box appear in the scene with small depth variation. In the *Boxes* example, there are a dozen objects which are appear in a cluttered scene with large depth variation. The uniform algorithm [5] and the method [10] assuming constant depth do not perform well in this case, while the proposed method recovers sharper images. Our method is not able to detect subtle depth change, and instead infers mainly

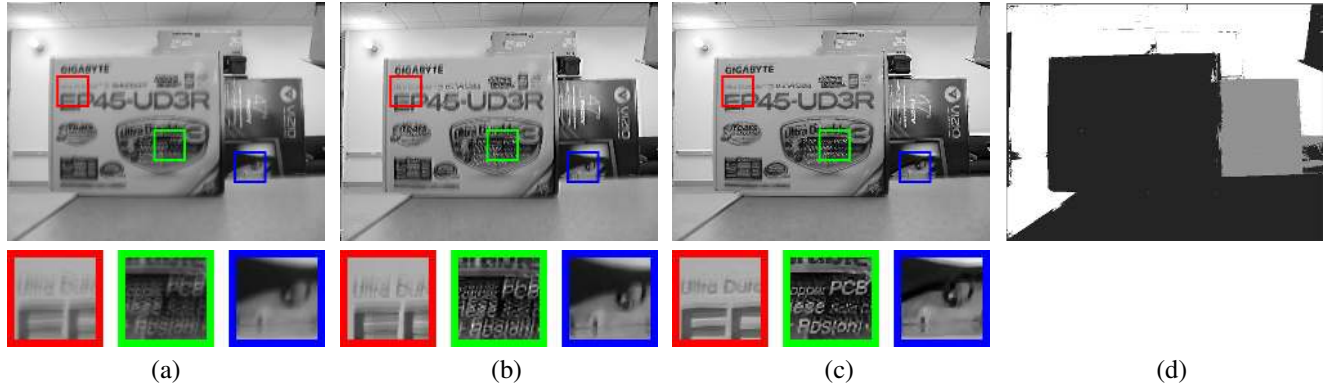


Figure 6. Depth-aware non-uniform deblurring on synthetic images. (a) synthetic blurry image; (b) deblurred result of [10]; (c) recovered image using our method; (d) our estimated depth (best viewed on a high-resolution display).

two different depth layers in this image. However, by separating the image into two layers, the proposed method obtains more accurate weight estimation of camera motion compared to the one using constant depth, which also demonstrates the effectiveness of the depth-aware model. We note that the proposed method on real images does not estimate depth as accurate as in the cases of synthetic images. One reason is that there are many unknown variables to be estimated and the error surface is complex with alternating optimization in the M-step. Thus it is difficult to guarantee convergence to a good local minimum. Some other reasons include the simplified camera model assumption. Since our model is based on the pinhole model, lens error and defocus blur due to depth variation may affect the accuracy of depth estimation for real blurry images.

## 5. Concluding Remarks

In this paper, we discuss the depth influence on the geometric model of non-uniform blur caused by camera motion. We present a unified layer-based model for depth-involved deblurring and propose an EM formulation to recover depth and latent image jointly given only a blurry image. We demonstrate the effectiveness of our depth estimation scheme when there exists certain translations, and the estimated depth information helps improve the performance of latent image reconstruction over the state-of-the-art models that assume constant depth. When the depth variation of the scene is limited, the proposed algorithm performs similarly to the non-uniform deblur algorithms assuming constant depth.

Since our depth-aware blur model is based on multiple layers with constant depth on each, we mainly handle scenes with multiple objects in frontal views. It may fail for the scenes where contain elongated objects with large depth variation. Similar to the existing non-uniform deblurring methods, the computational load of the proposed algorithm is high. Our future work will focus on resolving more

complex scenes of depth variation and efficient solutions to reduce computational loads.

**Acknowledgment** The work is partially supported by NSF CAREER Grant #1149783 and NSF IIS Grant #1152576.

## References

- [1] J. Bardsley, S. Jefferies, J. Nagy, and R. Plemmons. A computational method for the restoration of images with an unknown, spatially-varying blur. *Optics Express*, 14(5):1767–1782, 2006. 2
- [2] M. Ben-Ezra and S. Nayar. Motion deblurring using hybrid imaging. In *CVPR*, pages 657–664, 2003. 2
- [3] A. Chakrabarti, T. Zickler, and W. T. Freeman. Analyzing spatially-varying blur. In *CVPR*, pages 2512–2519, 2010. 2
- [4] Q. Chen, D. Li, and C.-K. Tang. Knn matting. In *CVPR*, pages 869–876, 2012. 4
- [5] S. Cho and S. Lee. Fast motion deblurring. In *SIGGRAPH Asia*, 2009. 2, 4, 6, 8
- [6] S. Cho., Y. Matsushita, and S. Lee. Removing non-uniform motion blur from images. In *ICCV*, 2007. 1, 2
- [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*, pages 169–176, 2010. 2
- [8] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. In *SIGGRAPH*, pages 787–794, 2006. 2
- [9] A. Gupta, N. Joshi, L. Zitnick, M. Cohen, and B. Curless. Single image deblurring using motion density functions. In *ECCV*, pages 171–184, 2010. 2, 3, 4, 6
- [10] M. Hirsch, C. J. Schuler, S. Harmeling, and B. Scholkopf. Fast removal of non-uniform camera shake. In *ICCV*, pages 463–470, 2011. 1, 2, 3, 4, 5, 6, 7, 8
- [11] Z. Hu and M.-H. Yang. Fast non-uniform deblurring using constrained camera pose subspace. In *BMVC*, 2012. 2, 3, 4, 6, 8
- [12] H. Ji and K. Wang. A two-stage approach to blind spatially-varying motion deblurring. In *CVPR*, pages 73–80, 2012. 2



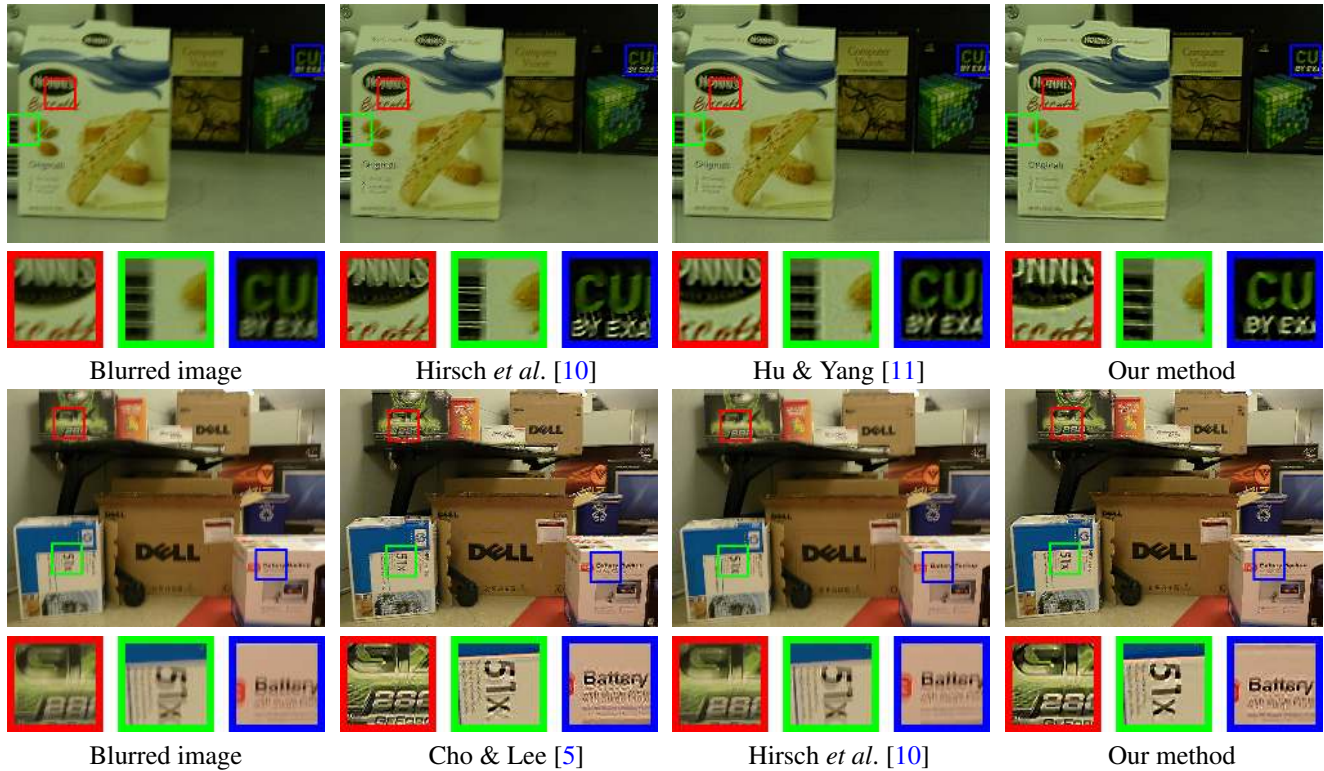


Figure 7. Depth-aware non-uniform deblurring on real images. Top to bottom: *Books*, and *Boxes* examples (best viewed on high-resolution displays).

[13] N. Joshi, S. B. Kang, C. L. Zitnick, and R. Szeliski. Image deblurring using inertial measurement sensors. In *SIGGRAPH*, 2010. 2

[14] R. Köhler, M. Hirsch, B. Mohler, B. Schölkopf, and S. Harmeling. Recording and playback of camera shake: benchmarking blind deconvolution with a real-world database. In *ECCV*, pages 27–40, 2012. 2, 3, 6

[15] D. Krishnan and R. Fergus. Fast image deconvolution using hyper-laplacian priors. In *NIPS*, 2009. 5

[16] D. Krishnan, T. Tay, and R. Fergus. Blind deconvolution using a normalized sparsity measure. In *CVPR*, pages 233–240, 2011. 5

[17] A. Levin. Blind motion deblurring using image statistics. In *NIPS*, pages 841–848, 2006. 2

[18] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. In *SIGGRAPH*, 2007. 1, 2, 5

[19] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Understanding and evaluating blind deconvolution algorithms. In *CVPR*, pages 1964–1971, 2009. 2

[20] F. Li, J. Yu, and J. Chai. A hybrid camera for motion deblurring and depth map super-resolution. In *CVPR*, 2008. 2

[21] C. Paramanand and A. N. Rajagopalan. Non-uniform motion deblurring for bilayer scenes. In *CVPR*, 2013. 1, 2

[22] R. Raskar, A. Agrawal, and J. Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. In *SIGGRAPH*, pages 795–804, 2006. 2

[23] A. Saxena, M. Sun, and A. Y. Ng. Make3d: Learning 3d scene structure from a single still image. *PAMI*, 31(5):824–840, 2009. 2, 5, 6

[24] Q. Shan, J. Jia, and A. Agarwala. High-quality motion deblurring from a single image. In *SIGGRAPH*, 2008. 2, 4, 5

[25] Q. Shan, W. Xiong, and J. Jia. Rotational motion deblurring of a rigid object from a single image. In *ICCV*, 2007. 2

[26] M. Sorel and J. Flusser. Space-variant restoration of images degraded by camera motion blur. *TIP*, 17(2):105–116, 2008. 2

[27] Y. W. Tai, H. Du, M. S. Brown, and S. Lin. Image/video deblurring using a hybrid camera. In *CVPR*, 2008. 2

[28] Y. W. Tai, P. Tan, and M. S. Brown. Richardson-Lucy deblurring for scenes under projective motion path. *PAMI*, 33(8):1603–1618, 2011. 1, 2

[29] O. Whyte, J. Sivic, and A. Zisserman. Deblurring shaken and partially saturated images. In *ICCVWorkshops*, pages 745–752, 2011. 2

[30] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce. Non-uniform deblurring for shaken images. In *CVPR*, pages 491–498, 2010. 1, 2, 3, 4

[31] L. Xu and J. Jia. Depth-aware motion deblurring. In *ICCP*, 2012. 1, 2, 6

[32] L. Xu, C. Lu, Y. Xu, and J. Jia. Image smoothing via l0 gradient minimization. *SIGGRAPH*, 30(6):174, 2011. 4, 5

[33] L. Yuan, J. Sun, L. Quan, and H. Shum. Image deblurring with blurred/noisy image pairs. In *SIGGRAPH*, 2007. 2