M.S. THESIS

Occlusion-Aware Motion Deblurring for Bilayer Scenes

가리어짐을 고려한 영상 디블러링

BY

BYEONGJOO AHN

FEBRUARY 2014

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Abstract

In this thesis, a novel blur model that can deal with occlusion in the blurred image from a scene with depth discontinuities is proposed. Existing deblurring methods usually ignore the occlusion that occurs near the depth variations but it causes severe artifacts near the object boundary, which is a critical factor in deblurring. Based on the analysis about the blur kernel near the depth discontinuities for a two-layer image model, a new occlusion-aware blur model which can make use of the information of occluded regions is proposed. Proposed model jointly recovers the depth map, foreground mask and restored image with accurate object boundary from two blurred observations. Also, a highly accurate optimization method is provided based on MCMC. Comparative experimental results on synthetic and real blurred images demonstrate convincingly that proposed model gives satisfactory results.

Key words: Image deblurring, Depth discontinuity, Occlusion, Blur model, Vision-based 3D reconstruction, Markov chain Monte Carlo

Student number: 2012-20803

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Motion blur problem has been studied extensively for a long time and can be roughly categorized into spatially invariant [2, 3] and spatially variant configurations [4, 5, 6]. In particular, spatially variant blur is usually motivated by the fact that blur is caused by camera motion. State-of-the-art approaches estimate the camera motion based on the assumption that the scene has no *depth variation* [4, 5, 6], but recently the depth variation is considered to deal with more general cases [1, 7, 8] because it is common in practice.

1.1 Background and Research Issues

However, the *occlusion* that occurs near the depth discontinuities during camera motion is still ignored in the conventional methods. To be specific, motion blur from camera motion is made by integration of all intermediate images that the camera sees along the trajectory of camera motion. When there are depth discontinuities in the static scene, in particular, the object closer to camera (foreground) moves

(a) Blurred image 1 (b) Blurred image 2 (c) Deblurring result (d) Recovered α

Figure 1.1: Blurred images and deblurring results

more than farther object (background) in the image during camera motion. Then, the background region which is exposed at the beginning of camera motion can be occluded by the foreground object (*i.e.* occlusion). Also, the background region occluded by the foreground object before can appear as the camera moves (*i.e.* disocclusion).

If the occlusion is ignored, exact boundary of the object cannot be restored which is a critical factor in deblurring. Moreover, when deblurring the image that contains large occlusions (*i.e.* large blur), not only boundary but the overall result image would have severe artifacts. Prior models can alleviate some artifacts, but they cannot solve the problem radically.

1.2 Outline of the Thesis

In this thesis, a novel blur model that can deal with the occlusion as well as depth variations is proposed for the first time, based on a two-layer image model (*i.e.* foreground and background). The camera motion composed of 2D translations and in-plane rotations is allowed, and two differently blurred images are required to handle the occlusion. Using this model, the foreground mask, depth map and the

restored image of each layer including the information of occluded region can be estimated simultaneously as shown in Fig. (1.1). Proposed model can also be applied to 3D reconstruction from blurred images [1] which inevitably contains some occlusions. Moreover, to deal with large blurs, which includes considerable occlusions, the proposed occlusion-aware blur model is necessary.

In Chapter 2, related works and the contributions are summarized. Conventional deblurring methods which handle depth variations or occlusions are introduced and the difference with this thesis is presented. In Chapter 3, the generation process of blur considering occlusion is analyzed and the new model is proposed with the comparison to the traditional model. In Chapter 4, the deblurring process which contains the problem statement, camera pose interpolation, objective function, and the optimization method is introduced. In Chapter 5, assumptions about camera poses are discussed. In Chapter 6, experimental results are shown. Finally, the thesis is concluded in Chapter 7.

Chapter 2

Related work

Motion blur from camera motion has been studied for a long time. State-of-the-arts are roughly categorized into spatially invariant and spatially variant configurations.

2.1 Uniform Blur

For spatially invariant blur, Fergus et al. [3] proposed a blind deconvolution method using natural image statistics. Shan et al. [9] used carefully chosen regularization term to reduce ringing artifacts. Cho and Lee [2] used an optimization technique that makes kernel estimation faster. Xu et al. [10] used the spatial prior and the iterative support detection (ISD) kernel refinement to restore pictures from significant motion blur.

2.2 Non-Uniform Blur

2.2.1 Non-Uniform Blur from Camera Motion

In practice, however, realistic camera motion includes in-plane rotation [11] and the blur is spatially variant because of camera rotation. Whyte et al. [6] proposed a model that can deal with rotational camera motion. Gupta et al. [4] proposed motion density function to represent camera motion, and also handled in-plane translation and rotation. Hirsch et al. [5] used efficient filter flow to reduce computational complexity. Xu et al. [12] proposed a robust kernel estimation method based on the sparse representation.

2.2.2 Non-Uniform Blur with Depth Variations

Although these methods solved spatially variant motion blur well, they were limited to the scenes without depth variations. To address the problem of depth variations, Xu and Jia [8] found depth dependent blur kernels using stereopsis for the case of translational motion. Lee and Lee [1] and Paramanand and Rajagopalan [7] utilized commutative property of convolution to estimate depth map. They restored the image with depth variations using blur kernels from the depth map, and handled in-plane rotations as well as in-plane translations. Paramanand and Rajagopalan handled bilayer scenes. This thesis is motivated by these works.

2.2.3 Non-Uniform Blur with Occlusions

While depth variations were effectively handled by these methods, these previous studies ignored the occlusion, which unavoidably occurs near the depth discontinuities. The occlusion caused by depth discontinuities during camera motion has not

been studied yet, but the occlusion caused by partial blur has been studied by Dai and Wu [13]. They made use of image matting method with user interaction and found the blur kernel from the mask. Several recent works utilize image matting to resolve the blur problem [14, 15, 16].

2.3 Contributions

Proposed method can deal with the occlusion as well as depth variations. Both in-plane translations and rotations of camera motion are allowed and user interaction is not needed at all. It has significant advantages over previous methods that can handle depth variations [1, 7] or occlusions [13]. First, the commutative property of convolution used in [1, 7] cannot deal with the occlusion. This is because commutativity does not hold near the depth discontinuities. Moreover, a blurred image with depth discontinuities cannot even be represented by a convolution of the single layer image and the blur kernel. Second, although occlusion is effectively handled by the method in [13], it is limited to partial blur and translational motion. In other words, neither foreground nor background should be blurred and rotational motion should not be allowed. Furthermore, user interaction is required to find the initial foreground mask.

Chapter 3

Analysis of Occlusion during Camera Motion

In this chapter, based on the two-layer image model [13], the generation process of motion blur from camera motion is analyzed for occlusion-aware blur model. The two-layer model of clear image is explained in Section 3.1 and how the two-layer image is projected when the camera moves is shown in Section 3.2. Finally, using the result of Section 3.1 and Section 3.2, a new blur model considering occlusion is proposed and then is compared with other models in Section 3.3.

3.1 The Two-Layer Model of Latent Image

When there are depth discontinuities in the scene, there occur occlusions and disocclusions during camera motion near pixels at the boundary of foreground as illustrated in Fig. (3.1). These pixels have background (foreground) information at the beginning of camera motion but after occlusion (disocclusion) occurs, they have the

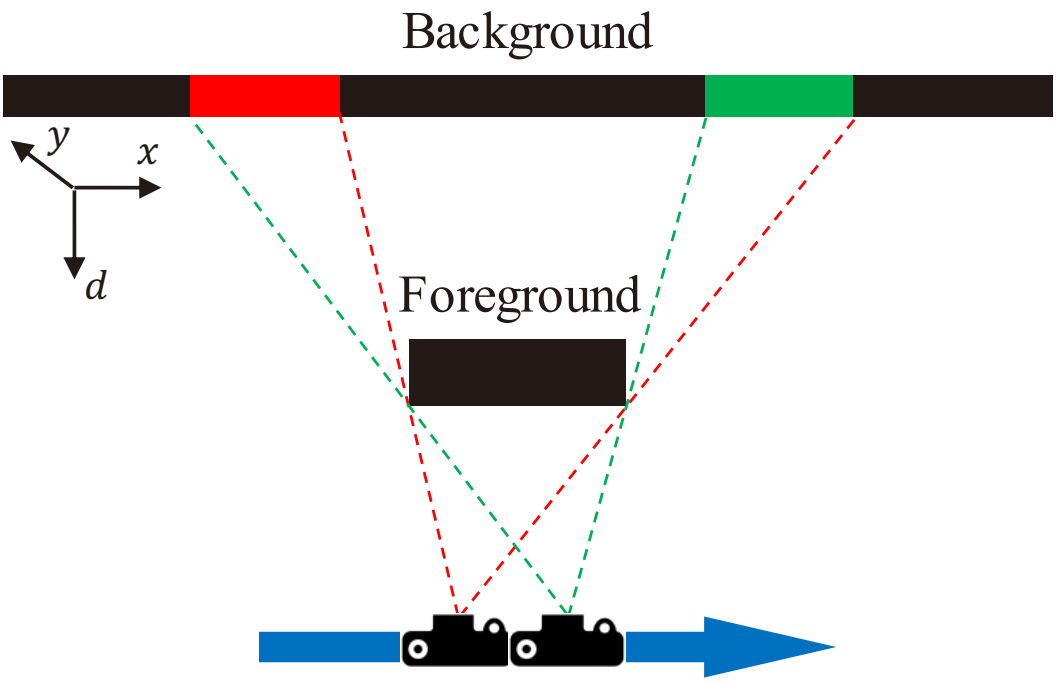


Figure 3.1: Illustration of the blur generation process from the scene with depth discontinuities. Red areas represents occlusion and green areas represents disocclusion.

foreground (background) information. Thus, these pixels have both foreground and background information at the same position. Such information cannot be captured by the traditional single layer image model which expresses the blur as the convolution of a single layer latent image and a spatially variant point spread function (PSF). Therefore, a layered image model is required to consider the occlusion during camera motion.

This thesis focuses on the case of two-layer, because it is common and representative in practice and can be generalized to the multiple layer model [13]. Then, the latent image L can be modeled as

$$L = \alpha F + (\mathbf{1} - \alpha)B, \tag{3.1}$$

where F is the layer closer to the camera (clear foreground layer), B is the farther layer (clear background layer), α is a clear foreground mask that represents where the foreground object is, and $\mathbf{1}$ is a matrix of ones. The size of F, B, α and $\mathbf{1}$ are same with that of L, and the pixel value of α is 1 at the pixels where the foreground object is and 0 elsewhere.

3.2 The Two-Layer Image Transformation

Given a two-layer image model, how the image is projected during the camera exposure time as the camera moves is analyzed. Let *the unblurred frame when camera shutter opens* be *the reference image*. Then, the projected image is represented as a transformation of the reference image, to model the blur as integration of them. Notably, the transformation should be able to represent the occlusion and disocclusion which is mentioned earlier.

Assume that the relative camera pose (*w.r.t the camera pose of the reference image*) at every time during the camera exposure period T is given, and camera motion is composed of 2D translation and in-plane rotation. Then, how each pixel of the reference image is projected to the image of a camera with pose $\mathbf{P}^\tau \in \mathbf{SE}(3)$ at exposure time $\tau \in [0, T]$ can be calculated as follows:

$$\begin{aligned} (x^\tau, y^\tau) &= h(\mathbf{K}((\mathbf{P}^\tau)^{-1}) \cdot \mathbf{X})), \\ \mathbf{X} &= \frac{1}{d}\mathbf{K}^{-1} \cdot (x, y, 1)^T, \end{aligned} \tag{3.2}$$

where (x^τ, y^τ) is projected pixel point, $h(\cdot)$ is the dehomogenization function, such that $h((x, y, z)^T) = (\frac{x}{z}, \frac{y}{z})$, $\mathbf{K}$ is the camera intrinsic matrix, $\mathbf{X}$ is a 3D scene point corresponding to pixel (x, y) at the reference image and d is the depth of the pixel.

In two-layer case, there are only two values that d can have. *i.e.* the depth

of foreground d_F and the depth of background d_B. Let $\mathcal{W}_F^\tau$ and $\mathcal{W}_B^\tau$ be the image transformation function that map the pixels in the foreground and background of the reference image to the corresponding pixels in the image of $\mathbf{P}^\tau$ via (2), respectively. Notably, given $\mathbf{P}^\tau$, $\mathcal{W}_F^\tau$ and $\mathcal{W}_B^\tau$ are functions of d_F and d_B, respectively. Then, the projected image L^τ at time τ can be expressed by a transformation of the reference image as follows:

$$L^\tau = \mathcal{W}_F^\tau(\alpha F) + (\mathbf{1} - \mathcal{W}_F^\tau(\alpha))\mathcal{W}_B^\tau(B). \tag{3.3}$$

This equation means the projected image is composed of transformed foreground, transformed background and transformed foreground mask. Foreground and its mask are transformed by $\mathcal{W}_F^\tau$ and background is transformed by $\mathcal{W}_B^\tau$. Then, L^τ has transformed foreground value at the pixel where $\mathcal{W}_F^\tau(\alpha)$ is 1 and transformed background value where $\mathcal{W}_F^\tau(\alpha)$ is 0. Since foreground mask is also transformed same with foreground, pixels at the background can be occluded or be disoccluded. Therefore, Eq. (3.3) can represent the projected image as a transformation of the reference image while considering occlusions.

3.3 Occlusion-Aware Blur Model

Now, the motion blur from camera motion can be modeled by integrating L^τ during the exposure time $\tau \in [0, T]$. Then the blurred image I considering occlusions can be represented as follows:

$$I = \frac{1}{T}\int_0^T \mathcal{W}_F^\tau(\alpha F) + (\mathbf{1} - \mathcal{W}_F^\tau(\alpha))\mathcal{W}_B^\tau(B)d\tau. \tag{3.4}$$

This is occlusion-aware blur model. It looks unfamiliar but by using matrix representation, it can be expressed in a similar form to traditional blur models. Let

us represent the image transformation function as the multiplication of a transformation matrix and a vector representation of the image. Then, Eq. (3.4) can be expressed as follows:

$$\mathbf{x}_I = \mathbf{K}_F\mathbf{x}_F + \mathbf{K}_B\mathbf{x}_B = \begin{pmatrix} \mathbf{K}_F & \mathbf{K}_B \end{pmatrix} \begin{pmatrix} \mathbf{x}_F \\ \mathbf{x}_B \end{pmatrix} = \mathbf{K}_{FB}\mathbf{x}_{FB}, \tag{3.5}$$

where $\mathbf{x}_I$, $\mathbf{x}_F$ and $\mathbf{x}_B$ are the vector representations of I, F and B. $\mathbf{K}_F$ is the integration of the matrix that achieves operation of element-wise multiplication of α, followed by the foreground transformation from camera motion, and $\mathbf{K}_B$ is the integration of the matrix that operates the background transformation from camera motion followed by element-wise multiplication of $(\mathbf{1} - \mathcal{W}_F^\tau(\alpha))$. In other words, $\mathbf{K}_F$ and $\mathbf{K}_B$ are represented as follows:

$$\begin{aligned} \mathbf{K}_F &= \frac{1}{T}\int_0^T \mathbf{W}_F^\tau diag(\mathbf{x}_\alpha) d\tau, \\ \mathbf{K}_B &= \frac{1}{T}\int_0^T (\mathbf{E} - diag(\mathbf{W}_F^\tau \mathbf{x}_\alpha))\mathbf{W}_B^\tau d\tau, \end{aligned} \tag{3.6}$$

where $\mathbf{W}_F^\tau$ and $\mathbf{W}_B^\tau$ are the matrix representations of $\mathcal{W}_F^\tau$ and $\mathcal{W}_F^\tau$, respectively. $\mathbf{x}_\alpha$ is the vector representation of α, and $\mathbf{E}$ is the identity matrix. The size of these matrices is p by p where p is the number of pixels in the image.

This formulation is similar to the conventional spatially variant blur model which represents the blur image as $\mathbf{K}\mathbf{x}_L$ ($\mathbf{K}$ is the spatially variant PSF matrix and $\mathbf{x}_L$ is the vector representation of a single layer latent image). Thus, $\begin{pmatrix} \mathbf{K}_F & \mathbf{K}_B \end{pmatrix}$ can be considered as a spatially variant PSF. The difference is, in the conventional model, PSF represents the pixels in the single layer reference image that influence a pixel in the blurred image. While, in the proposed model, the element of $\mathbf{K}_F(\mathbf{K}_B)$ represents the pixels in the foreground (background) layer of the reference image that influence a pixel in the blurred image.

For example, let us consider the (i, j) element of $\mathbf{K}_F(\mathbf{K}_B)$. This element shows how the jth pixel of $\mathbf{x}_F(\mathbf{x}_B)$ influence to the ith pixel of blurred image. In other words, the ith row of $\mathbf{K}_{FB}$ corresponds to the PSF of ith pixel of blurred image considering both foreground and background.

Fig. (3.2) shows the synthetic blurred images and their blur kernels from the conventional blur model in [1, 7] and the proposed blur model. In synthetic blurred images, conventional blur model (in (d)) does not represent the disocclusion of blue background and the occlusion of green background while the proposed model (in (e)) successfully represents it. It results from newly proposed PSF which considers two layers. In most of the regions, the proposed PSF is made from one layer which is same as the conventional model. On the other hand, in the region being exposed (in (f)(g)) or being occluded (in (h)(i)) during the exposure time, the PSF is from both the layers where whites are the pixels in F indicated by $\mathbf{K}_F$, blacks are the pixels in B indicated by $\mathbf{K}_B$, and grays are overlaps of whites and blacks. Notably, PSF at the pixel being exposed is overlapped and PSF at the pixel being exposed is separate.

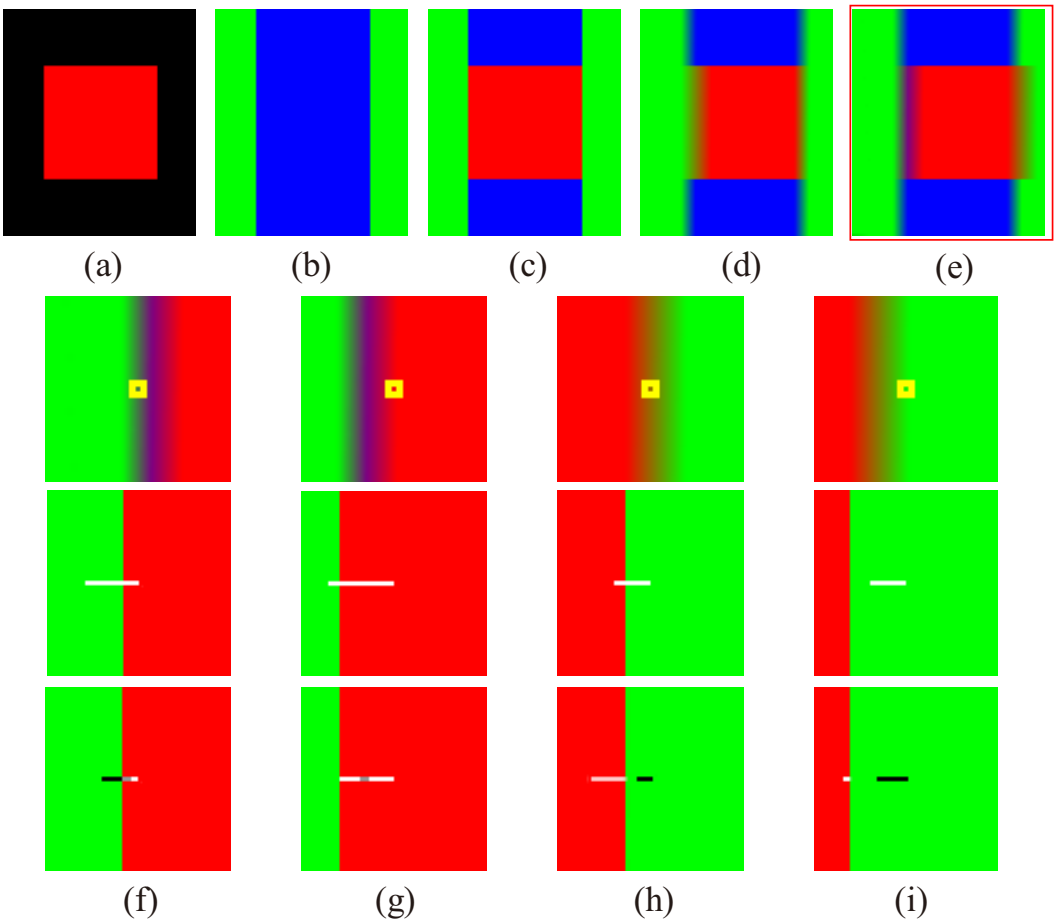


Figure 3.2: Illustration of the synthetic blurred images and their blur kernels from conventional blur model and the proposed blur model. Camera is moving to the left. (a) F. (b) B. (c) L (d) synthetic blurred image from conventional blur model. (e) synthetic blurred image from the proposed blur model. (f)(g)(h)(i) **1st row**: close-up views of (e) near boundary of red box. **2nd row**: conventional PSF of the yellow-marked pixel in 1st row. **3rd row**: proposed PSF of the yellow-marked pixel in 1st row, whites indicate PSF in F, blacks indicate PSF in B and grays indicate PSF in both. (f)&(g) are disoccluded region (left side of red box) and (h)&(i) are occluded region (right side of red box).

Chapter 4

Occlusion-Aware Motion Deblurring

In this chapter, how to deblur the image using the occlusion-aware blur model is explained. However, solving all unknowns (F, B, α, d_F, d_B) in Eq. (3.4) based on one blurred image observation I is severely under constrained. Moreover, even if α, d_F and d_B are given, F and B are not uniquely determined in the occluded or disoccluded region. For example, assume that the foreground object is white and background object is black. Then in the blurred image, the region near the boundary of the foreground object will be gray. In this case, same blurred image can be obtained if the boundary of the foreground object is black and the occluded/disoccluded region of the background is gray.

To solve this ill-posed problem, two blurred images which originate from the same scene (*i.e.* same F and B) are used as observations. Since two blurred images are blurred differently, the occluded region does not have the ill-posedness. Multiple observations can be easily obtained in practice and have been used frequently in

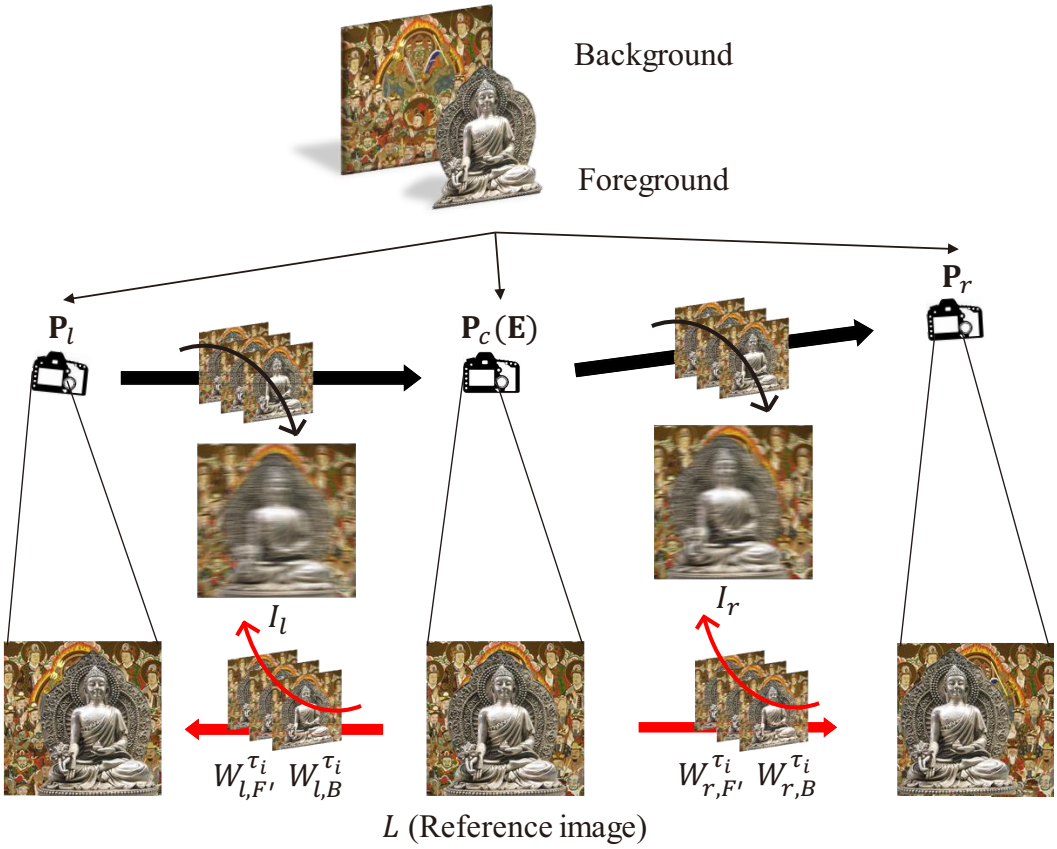


Figure 4.1: Illustration of the problem statement. Black arrows represent real blur generation process. Red arrows represent the proposed synthetic blur generation process for deblurring.

other deblurring approaches [17, 1, 7, 18, 19].

In section 4.1, the problem situation which uses two blurred images as observation is explained. In section 4.2, camera poses are interpolated at a uniformly sampled time during camera motion to make use of the occlusion-aware blur model in this problem situation. In section 4.3, the objective function using the proposed blur model with smoothness prior is proposed. Finally, the optimal values of (F, B, α, d_F, d_B) are obtained by using Markov chain Monte Carlo (MCMC) in section 4.4.

4.1 Problem Statement

For this thesis, the situation that we have considered is illustrated in Fig. (4.1). Two blurred observations are obtained by consecutively captured image frames. To be specific, blurred image I_l and I_r are made by integration of images that the camera sees from $\mathbf{P}_l$ to $\mathbf{P}_c$ and from $\mathbf{P}_c$ to $\mathbf{P}_r$, respectively. We are going to represent these blurred images by the image at $\mathbf{P}_c$. To use this image as the reference image L for both I_l and I_r, we consider the blurred image I_l is made by the camera motion from $\mathbf{P}_c$ to $\mathbf{P}_l$, rather than from $\mathbf{P}_l$ to $\mathbf{P}_c$. Since commutative property holds for addition, the direction of camera motion does not affect the blur model, which consists of the addition of intermediate images. Notably, by using same reference image for both blurred images, the information that two blurred images are from same scene is utilized.

4.2 Camera Pose Interpolation

The occlusion-aware model needs camera poses with respect to the camera pose of the reference image during the exposure time to parameterize $\mathcal{W}_F^\tau$ and $\mathcal{W}_B^\tau$ as the functions of d_F and d_B, respectively. In this framework, similar to [1], the camera poses $\mathbf{P}_l$, $\mathbf{P}_c$ and $\mathbf{P}_r$ are obtained from the registration-based camera localization algorithm. Let the c be the reference camera. *i.e.* $\mathbf{P}_c = \mathbf{E}$ (4 by 4 identity matrix). Then, $\mathbf{P}_l$ and $\mathbf{P}_r$ are the camera pose with respect to the camera pose of the reference image ($\mathbf{P}_c$). Now, the intermediate camera poses of $\mathbf{P}_l$ and $\mathbf{P}_r$ from $\mathbf{P}_c$ are obtained by interpolations assuming the camera motion is smooth.

To interpolate the camera pose, we divide the exposure time into M intervals uniformly, such that $\tau_i = \frac{T}{M} i$ where $i \in \{1, ..., M\}$. Then, an intermediate camera

pose $\mathbf{P}_s^{\tau_i}$ of $\mathbf{P}_s$ at time τ_i on the manifold of $\mathbf{SE}(3)$ is interpolated as follows:

$$\mathbf{P}_s^{\tau_i} = \exp\left\{\left(\frac{i}{M}\right)\log(\mathbf{P}_s)\right\}, \tag{4.1}$$

where $s \in \{l, r\}$. Note, the duration of the shutter period T is not needed in Eq. (4.1).

Using the intermediate camera poses with respect to the reference image (*i.e.* $\mathbf{P}_s^{\tau_i}$), the image transformation functions can be parameterized by d_F and d_B as Eq. (3.2). Let the image transformation functions using $\mathbf{P}_s^{\tau_i}$ be $\mathcal{W}_{s,F}^{\tau_i}$ and $\mathcal{W}_{s,B}^{\tau_i}$ for F and B, respectively. Then, we can represent a final proposal function using the image transformation functions and F, B, α of the reference image.

Notably, only *relative* camera pose with respect to the reference camera pose, rather than the *absolute* camera pose, is needed for the proposed model. The assumptions about the camera pose are not needed if we can obtain relative camera pose (*i.e.* camera transformation) from other method such as [7]. We will discuss this in detail in Chapter 5.

4.3 Objective Function

The final proposal function is

$$\begin{aligned} E(F, B, \alpha, d_F, d_B) = & \left\| \sum_{\substack{s\in \\ \{l,r\}}} \left[\frac{1}{M} \left(\sum_{i=1}^{M} \mathcal{W}_{s,F}^{\tau_i}(\alpha F) + (\mathbf{1} - \mathcal{W}_{s,F}^{\tau_i}(\alpha))\mathcal{W}_{s,B}^{\tau_i}(B) \right) - I_s \right] \right\|_2^2 \\ & + \sum_{X\in\{F,B,\alpha\}} \lambda_X R(X), \end{aligned} \tag{4.2}$$

where R is the regularization term for F, B and α, such that

$$R(X) = \sum_{x} \sum_{y \in \mathcal{N}(x)} |X(x) - X(y)|, \tag{4.3}$$

where $X \in \{F, B, \alpha\}$. We set $\lambda_F = 0.0005$, $\lambda_B = 0.0005$ and $\lambda_\alpha = 0.005$ in the experiments as the weights.

4.4 Optimization

For optimization, we use MCMC to efficiently explore the entire solution space including the domain of (F, B, α, d_F, d_B). This approach is of critical importance because all unknowns are highly correlated in the proposed method to deal with the occlusion while each unknown is independent of other unknowns in the methods which do not consider the occlusion [1, 7]. It is possible to heuristically optimize the objective function by iteratively solving each unknown which can be formulated as quadratic function [13]. However, this approach has a limitation that the user interaction is required for a good initial value. Proposed approach, on the other hand, can obtain optimal value without user interactions or other methods for good initial value by using MCMC.

4.4.1 Markov chain Monte Carlo

Given a target probability distribution $\pi \propto \exp\{-E\}$, the aim is to find the state (F, B, α, d_F, d_B) where the probability is maximized. For $\{F, B, \alpha\}$, single-site Metropolis–Hastings is used. In this framework, each pixel of F, B and α are updated individually, rather than updating every pixel in the image all at once. To be specific, assume that lth state of each unknown is F^l, B^l and α^l and that the jth component is to be updated. Then the next state of the chain only differs from this state

on the jth component. *i.e.* $X_i^{l+1} = X_i^l, i \neq j$ where $X \in \{F, B, \alpha\}$. To generate the next state, we propose Y_j from a proposal density $q_X(X_j^l, \cdot)$ and let the next state $Y = (X_1^l, ..., X_{j-1}^l, Y_j, X_{j+1}^l, ..., X_p^l)$ where p is the number of pixels. Then this proposal is accepted according to the acceptance ratio a as follows:

$$\begin{aligned} a(X^l, Y) &= \min\left(1, \frac{\pi(Y, \cdot)^{\frac{1}{T}} q_X(Y_j, X_j^l)}{\pi(X^l, \cdot)^{\frac{1}{T}} q_X(X_j^l, Y_j)}\right) \\ &= \min\left(1, \exp\left\{\frac{E(X^l, \cdot) - E(Y, \cdot)}{T}\right\} \cdot \frac{q_X(Y_j, X_j^l)}{q_X(X_j^l, Y_j)}\right), \end{aligned} \tag{4.4}$$

where $\pi(X, \cdot)$ and $E(X, \cdot)$ means the functions using the same values as previous state except the X, and T is the temperature for simulated annealing. X is updated in order of F, B and α, and then $(j+1)$th component is updated. The order of j is determined to row-wise or column-wise randomly in each iteration.

After all pixels are updated, $d = \{d_F, d_B\}$ is updated. Assume that the lth state is $d^l = \{d_F^l, d_B^l\}$. Then the proposal y is generated from a proposal density $q_d(d^l, \cdot)$ and is accepted by following acceptance ratio:

$$\begin{aligned} a(d^l, y) &= \min\left(1, \frac{\pi(y, \cdot)^{\frac{1}{T}} q_d(y, d^l)}{\pi(d^l, \cdot)^{\frac{1}{T}} q_d(d^l, y)}\right) \\ &= \min\left(1, \exp\left\{\frac{E(d^l, \cdot) - E(y, \cdot)}{T}\right\} \frac{q_d(y, d^l)}{q_d(d^l, y)}\right). \end{aligned} \tag{4.5}$$

We iterate above process N times decreasing T by cooling ratio C. *i.e.* $T_{n+1} = CT_n$.

In this experiment, $q_F(q_B)$ is a mixture of gaussian of 5 neighboring pixels, q_α is a binary kernel such that $q_\alpha(x, y)$ is 1 for $y \neq x$ and 0 for $y = x$, and q_d is a gaussian kernel. σ of the q_F, q_B is 0.05×255, and σ of the q_d decreases linearly from 0.1 to 0.01 during the iteration and the correlation is 0. Initial F, B are given by blurred

observations, initial α is given by $\mathbf{0}$ and initial depth is given randomly such that $d_F > d_B$. Besides, $N = 1500$, $C = 0.993$ and initial $T = 2$.

Chapter 5

Discussion

In this chapter, more details about the assumptions of the proposed method are given. This thesis makes several assumptions on the camera pose; camera intrinsic matrix, $\mathbf{P}_l$, $\mathbf{P}_c$ and $\mathbf{P}_r$ are given, and the camera motion is smooth. These are required to estimate *camera motion* but many images available online do not provide a camera intrinsic matrix, and camera motion may not be smooth in practice. However, if we can obtain the camera motion from other method such as [7], it is possible to use the proposed approach without these assumptions. Since the main issue of this work is dealing with the occlusion, and finding camera motion from layered images is not perfect yet, these assumptions are used in this framework.

Chapter 6

Experiments

We demonstrate the effectiveness of the proposed method by experimental results on synthetic and real blurred images.

Figs. (6.1), (6.2) and (6.3) shows the results on synthetic blurred images and Fig. (6.3) shows results on real blurred images. Boxes in (a) represents the cropped region where red, green and blue boxes include the part of foreground, background, and occluded/disoccluded region, respectively. Each layer is successfully restored including the boundary of the foreground object (*i.e.* occluded/disoccluded region). Notably, α can contain some homogeneous regions in background because it does not affect the restored image. Also, B has smooth values where foreground object exists because this region does not affect the data term of the objective function and determined only by regularization term.

Fig. (6.5)-(6.9) compares the result of the proposed method to that of the conventional methods on the blurred images from the scene with depth variations. The method of Xu et al [12] fails to restore the image. The method of Lee and Lee [1] restores foreground (red boxes) and background (green boxes) well, but contains

(a) Original image (b) Blurred image 1 (c) Blurred image 2 (d) Deblurring result

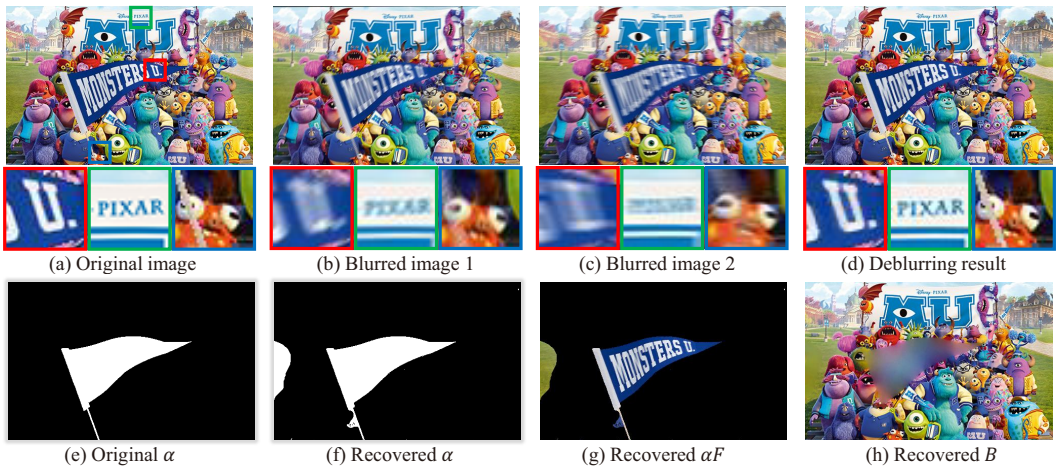
(e) Original α (f) Recovered α (g) Recovered αF (h) Recovered B

Figure 6.1: Person & background deblurring.

considerable artifacts at the boundary of the foreground object (blue boxes). On the other hand, the proposed method successfully restore the boundary of the foreground object as well as foreground and background. Peak signal-to-noise ratio(PSNR) for synthetic images also shows that proposed method achieves improved results.

Fig. (6.10) shows close-up view of restored background at the object boundary. Since the proposed method can make use of the information of occluded/disoccluded region, the background region exposed at least once during camera motion can be restored. Notably, the proposed method can restore the occluded/disoccluded region even if both foreground and background are blurred.

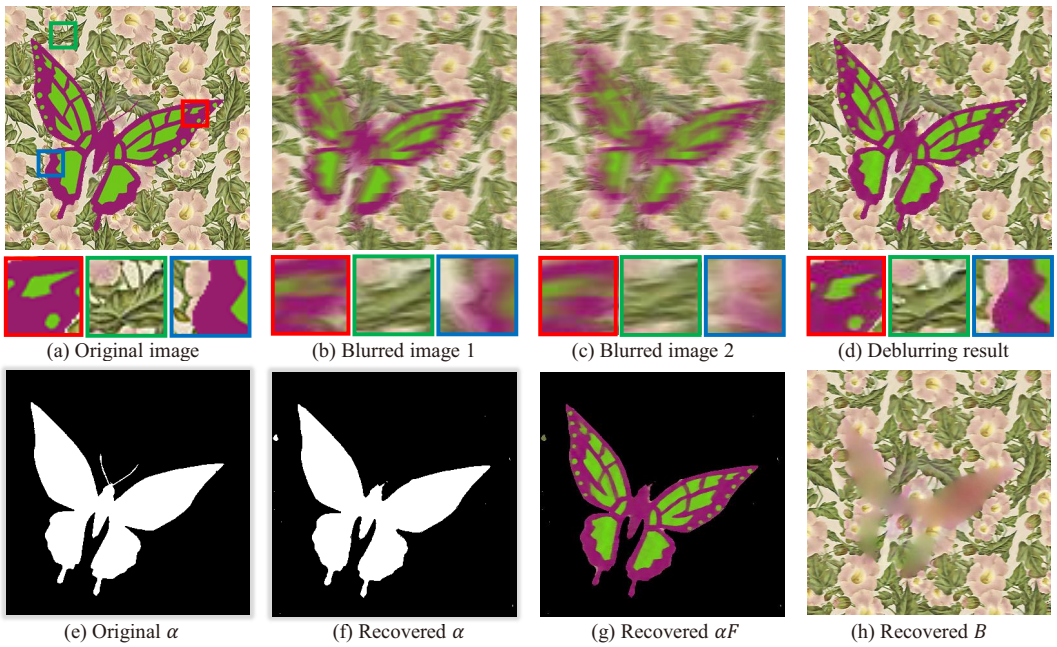
In the experiments, we obtained the camera intrinsic matrix and the camera poses by the method in [20].



(a) Original image (b) Blurred image 1 (c) Blurred image 2 (d) Deblurring result

(e) Original α (f) Recovered α (g) Recovered αF (h) Recovered B

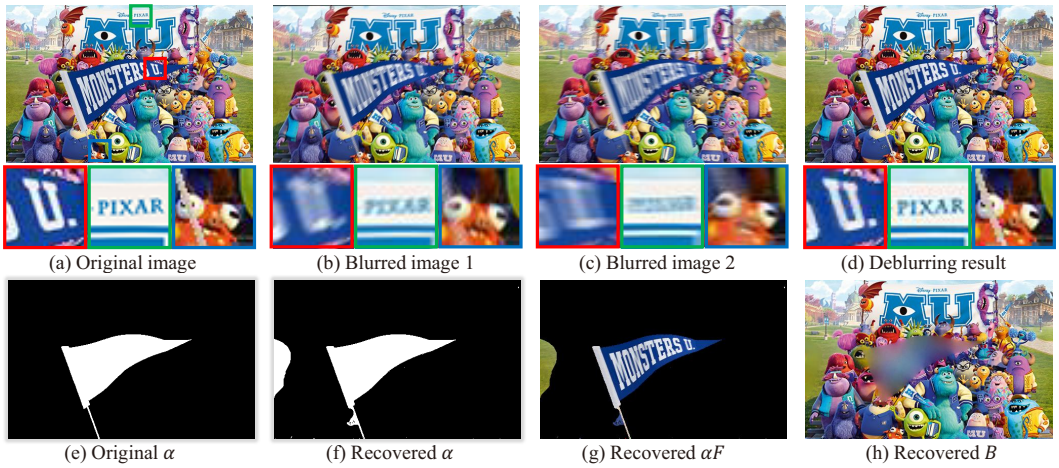
Figure 6.2: Butterfly & background deblurring.



(a) Original image (b) Blurred image 1 (c) Blurred image 2 (d) Deblurring result

(e) Original α (f) Recovered α (g) Recovered αF (h) Recovered B

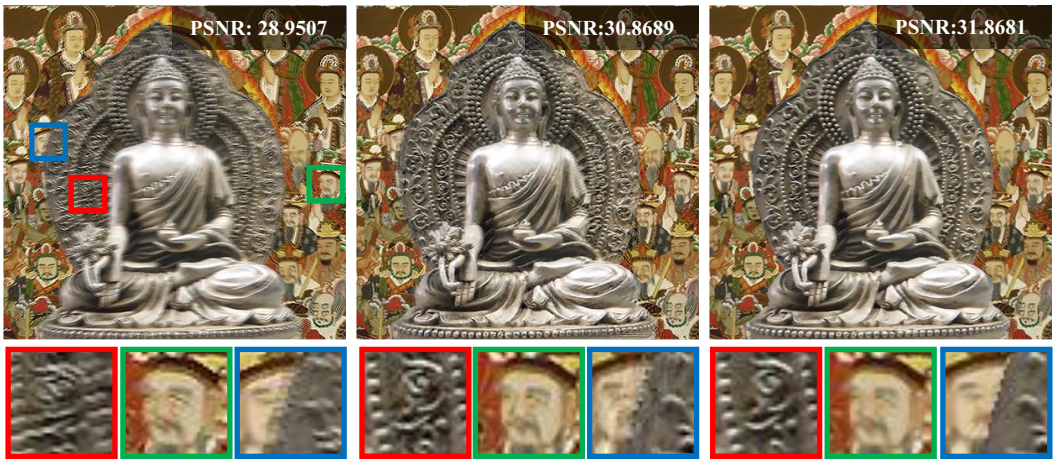
Figure 6.3: Flag & background deblurring.



(a) Original image (b) Blurred image 1 (c) Blurred image 2 (d) Deblurring result

(e) Original α (f) Recovered α (g) Recovered αF (h) Recovered B

Figure 6.4: Book & background deblurring.

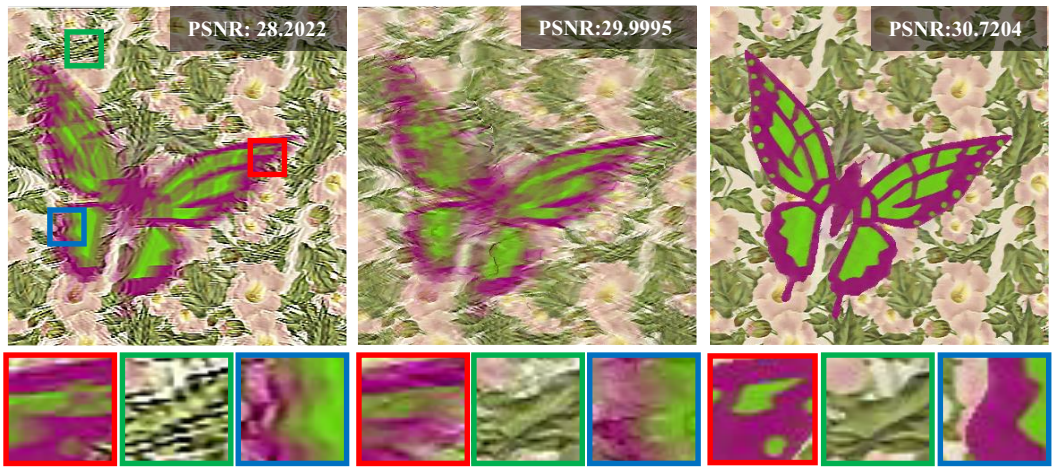


(a) Xu et al [21] (b) Lee & Lee [10] (c) Proposed method

Figure 6.5: Comparison of buddha & background deblurring.

(a) Xu et al [21] (b) Lee & Lee [10] (c) Proposed method

Figure 6.6: Comparison of person & background deblurring.



(a) Xu et al [21] (b) Lee & Lee [10] (c) Proposed method

Figure 6.7: Comparison of butterfly & background deblurring.

(a) Xu et al [21] (b) Lee & Lee [10] (c) Proposed method

Figure 6.8: Comparison of flag & background deblurring.

(a) Xu et al [21] (b) Lee & Lee [10] (c) Proposed method

Figure 6.9: Comparison of book & background deblurring.

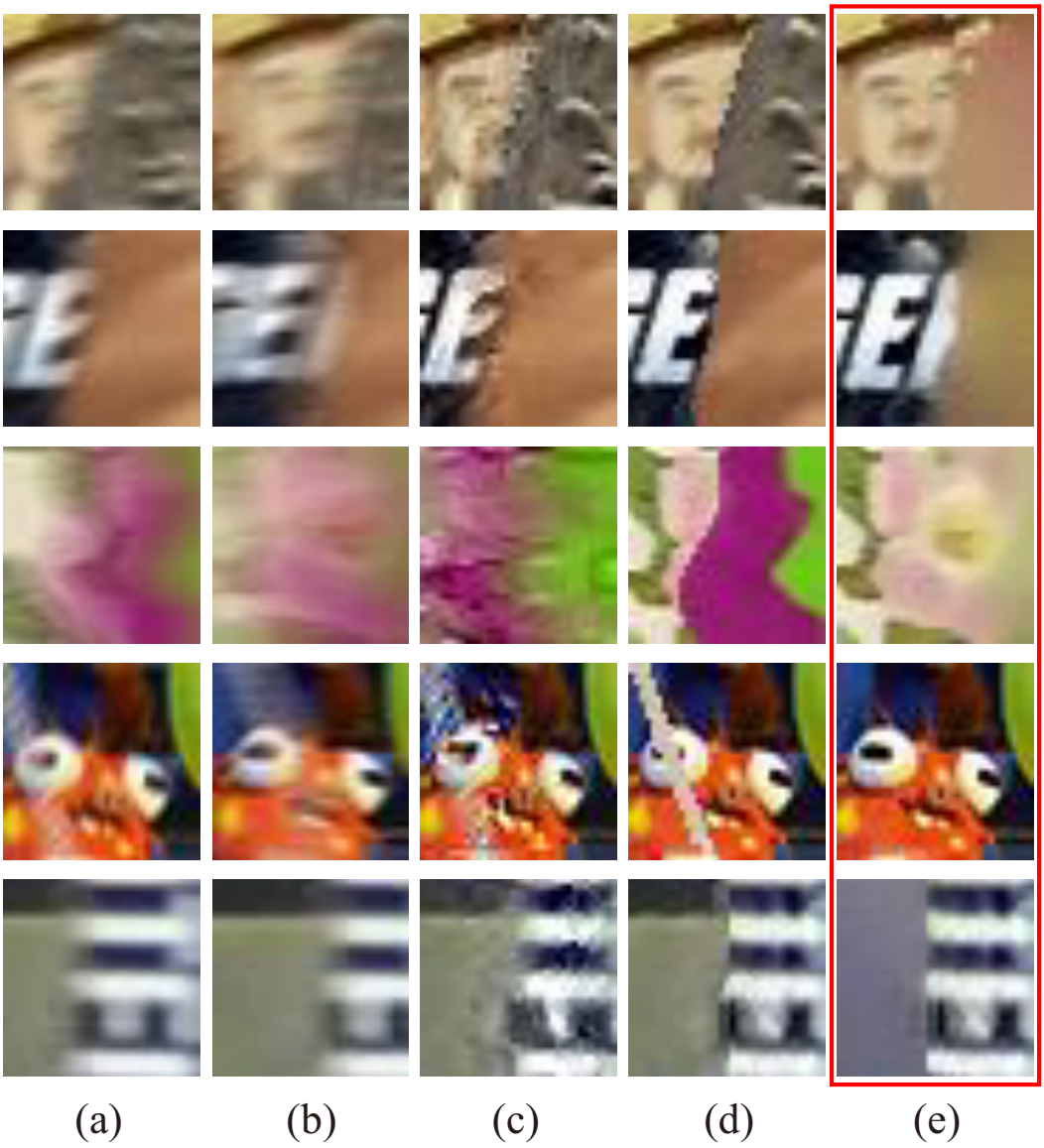


Figure 6.10: Recovered background at the object boundary. (a) blurred image 1. (b) blurred image 2. (c) Lee and Lee [1]. (d) L from proposed method. (e) B from proposed method.

Chapter 7

Conclusion

7.1 Summary of the Thesis

In this thesis, a novel blur model is proposed to deal with the occlusion from the blurred image with depth variations for the first time. This work is based on the analysis that the blur kernels near the depth discontinuities are peculiar (separate or overlapped PSF) which cannot be represented by the traditional blur model. Foreground layer, background layer, foreground mask and depth map are estimated from two blurred observations. By utilizing the information of occluded regions, exact boundary of the object is restored. Since human evaluate the deblurring results on clearness of the edge in the scene, the proposed method that gives exact and clear boundary of the object is of critical importance. Also, the background region exposed at least once during camera motion can be restored, which is impossible in conventional methods. Experimental results on synthetic and real blurred images demonstrate outstanding performance of our method. This method can contribute not only to deblurring, but also to 3D reconstruction.

7.2 Future Directions

7.2.1 Multi Layer Scenes

Proposed occlusion-aware motion deblurring method is focused on two-layer scenes. However, the scene may not be two-layer in practical situations. Although proposed method has contribution on finding exact edge in the blurred scene with depth discontinuities, it should handle multi layer scenes to be a fully general deblurring method. The key idea to deal with multi layer scenes is same with the case of two-layer which represents every projected image during camera motion as the transformation of the reference image. Remaining issue is how to make the formulation of blur model easy to optimize or finding new optimize scheme that can obtain the solution quickly and accurately. To detect depth discontinuities and using [1, 7] except depth discontinuities can also be a effective solution. By dealing with multi layer scenes, the proposed approach will be not only general deblurring method, but also general 3D reconstruction method by finding exact depth map of the scene.

7.2.2 Projective Motion

The camera motion is restricted to in-plane translation and rotation in this method. Although camera pose interpolation used in the proposed method can represent every 3D camera motion, it requires more consideration about the occlusion in a layer to deal with projective motion. If we represent the projected image during camera motion by homography, not by warping each pixel, projective motion could be handled. However, since homography needs more computational costs than current image transformation function, optimization speed should be improved to deal with projective motion.

7.2.3 Dynamic Scenes

Proposed method does not deal with moving or deformable objects because it parameterize the blur kernel by the camera motion and the depth of the scene. Combining a proposed framework to the case of moving object [21, 22] is one of the future aspirations. If we combine the proposed method with [1, 7] by detecting depth discontinuities, it can be applied to detecting occlusion in the case of dynamic scenes.

7.2.4 Real-Time Deblurring and 3D Reconstruction

Real-time is one of the important issues in the filed of 3D reconstruction. To contribute to 3D reconstruction, the proposed method should be faster. Current approach is slow because of sampling for optimization, but the sampling is needed only near the depth discontinuities because layers and foreground mask is highly correlated only in this region. In other regions, depth is independent of other unknowns and can be estimated quickly. Therefore, combining the proposed method with [1, 7] can also be a solution for this problem.

In conclusion, detecting depth discontinuities and using [1, 7] except depth discontinuities can solve a lot of current problems including multi layer scenes, projective motion, dynamic scenes and the speed of the method. Thus, detecting depth discontinuities and combining this with the proposed method is our main future direction.

Bibliography

[1] H. S. Lee and K. M. Lee, "Dense 3d reconstruction from severely blurred images using a single moving camera," in *CVPR*, 2013.

[2] S. Cho and S. Lee, "Fast motion deblurring," in *ACM Transactions on Graphics (TOG)*, vol. 28, no. 5, p. 145, 2009.

[3] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," in *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3, pp. 787–794, 2006.

[4] A. Gupta, N. Joshi, C. L. Zitnick, M. Cohen, and B. Curless, "Single image deblurring using motion density functions," in *ECCV*, 2010, pp. 171–184.

[5] M. Hirsch, C. J. Schuler, S. Harmeling, and B. Scholkopf, "Fast removal of non-uniform camera shake," in *ICCV*, pp. 463–470, 2011.

[6] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce, "Non-uniform deblurring for shaken images," *International journal of computer vision*, vol. 98, no. 2, pp. 168–186, 2012.

[7] C. Paramanand and A. N. Rajagopalan, "Non-uniform motion deblurring for bilayer scenes," in *CVPR*, pp. 1115–1122, 2013.

[8] L. Xu and J. Jia, "Depth-aware motion deblurring," in *ICCP*, pp. 1–8, 2012.

[9] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," in *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, p. 73, 2008.

[10] L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," in *ECCV*. Springer, 2010, pp. 157–170.

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

[12] L. Xu, S. Zheng, and J. Jia, "Unnatural l0 sparse representation for natural image deblurring," in *CVPR*, 2013.

[13] S. Dai and Y. Wu, "Removing partial blur in a single image," in *CVPR*, pp. 2544–2551, 2009.

[14] ——, "Motion from blur," in *CVPR*, pp. 1–8, 2008.

[15] J. Jia, "Single image motion deblurring using transparency," in *CVPR*, pp. 1–8, 2007.

[16] H. T. Lin, Y.-W. Tai, and M. S. Brown, "Motion regularization for matting motion blurred objects," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 11, pp. 2329–2336, 2011.

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

[18] T. Portz, L. Zhang, and H. Jiang, "Optical flow in the presence of spatially-varying motion blur," in *CVPR*, pp. 1752–1759, 2012.

[19] F. Sroubek and J. Flusser, "Multichannel blind deconvolution of spatially misaligned images," *Image Processing, IEEE Transactions on*, vol. 14, no. 7, pp. 874–883, 2005.

[20] G. Zhang, J. Jia, T.-T. Wong, and H. Bao, "Consistent depth maps recovery from a video sequence," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 6, pp. 974–988, 2009.

[21] T. Kim, B. Ahn, and K. M. Lee, "Dynamic scene deblurring," in *ICCV*, 2013.

[22] A. Levin, "Blind motion deblurring using image statistics," in *Advances in Neural Information Processing Systems*, pp. 841–848, 2006.

국문초록

영상 디블러링 연구는 영상 전체가 일정하게 블러된 경우부터 시작하여, 위치에 따라 다르게 블러된 경우, 그리고 영상에 깊이 차이가 있는 경우의 순서로 발전되어 왔다. 그런데 영상의 깊이가 불연속적으로 변하는 경우, 블러 과정동안에 영상에 가리어짐이 발생한다. 기존의 연구는 가리어짐 효과를 무시해왔는데, 이는 결과 영상에서 물체의 가장자리 부분에 상당한 오류를 일으킨다. 가장자리의 복원이 중요한 디블러링에서 이러한 오류는 치명적인 문제이다. 본 논문에서는 두 개의 층으로 이루어진 영상 모델을 이용하여 깊이가 불연속적으로 변하는 영상의 블러 형성 과정을 분석함으로써, 가리어짐 효과를 고려한 블러 모델을 제안한다. 그리고 이를 통해 두 개의 블러 영상으로부터 깊이 지도, 전경 마스크, 복원된 이미지를 얻어낸다. 합성된 영상과 실제 영상을 이용한 실험을 통하여 제안한 방법이 가장자리에서의 문제를 효과적으로 해결하였으며, 기존의 방법에 비해 우수한 성능을 내는 것을 확인하였다. 본 연구는 영상 디블러링뿐만 아니라 블러된 영상에서 3차원 복원을 해내는 데도 응용될 수 있다.

주요어: 영상 디블러링, 깊이 변화, 가리어짐, 블러 모델, 영상 기반 3차원 복원

학번: 2012-20803

감사의 글

2년간의 짧은 기간이었지만 연구자로서의 첫 걸음을 훌륭하신 분들과 함께 보낼 수 있어 정말 행복하고 감사합니다. 늘 좋은 말씀과 함께 열성적으로 지도해주신 교수님께서 계셨기에, 전공지식만이 아닌 연구에 임하는 자세와 연구자로서 갖추어야 할 태도에 대해 어렴풋이나마 깨달을 수 있었습니다. 교수님의 가르침을 가슴 속에 새기고 논문을 위한 연구가 아닌 세상에 기여할 수 있는 연구, 스스로 즐길 수 있는 연구를 할 수 있도록 노력을 게을리하지 하겠습니다. 교수님 정말 감사드립니다. 앞으로도 많은 가르침 부탁드립니다. 논문 심사를 통해서 부족한 제 연구를 보충할 수 있도록 도움을 주신 이상욱 교수님, 조남익 교수님께도 깊은 감사를 드립니다.

때로는 존경스러운 연구자로서, 때로는 친근한 형 누나 친구 동생으로서 매일매일을 함께 해온 연구실 선후배들이 있어 지난 2년은 정말 행복했습니다. 저에게 연구실을 다시 선택할 수 있는 기회가 주어진다 해도, 저는 다시 여기 컴퓨터 비전 연구실을 선택할 것입니다. 연구실 첫 면담 때부터 뵙고 연구자로서의 자세와 마음가짐에 대해 말씀해주신 민수형, 자상한 가르침과 함께 연구에 대한 자신감을 북돋아주신 준석이형, 부드럽고 온화한 미소로 입학부터 졸업까지 연구실 생활을 챙겨주신 영민이형, 묵직하면서도 부드럽게 늘 든든한 모습으로 연구실 선배들의 끈끈한 정을 느끼게 해주신 동우형, 연구주제도 함께 취미생활도 함께 하며 연구실에 나오는 즐거움을 느끼게 해주신 희석이형, 석사과정 내내 바로 뒤에서 무수한 연구 지도와 동시에 생활 전반을 세심하게 챙겨주시고 재미있게 지낼 수 있도록 해주신 태현이형, 늘 유쾌하면서도 자상하고 피곤한 모습으로 연구실 생활에서 많

은 가르침과 즐거운 웃음을 주신 원식이형, 큐브로 연구실 생활을 풍성하게 만들어주시고 날카로운 질문으로 다양한 깨달음을 주신 비상한 두뇌의 정민이형, 폭넓은 지식을 바탕으로 연구실 생활에 꼭 필요한 흥미로운 이야기를 들려주시고 늘 다정하게 챙겨주신 희수형, 정겨운 말투와 매력적인 미소로 연구실 생활의 시작을 즐겁게 만들어주신 배려의 아이콘 상돈이형, 친근하면서도 꼼꼼하게 생활과 진로를 챙겨주시고 자동화에서의 생활을 아기자기하게 만들어주신 앞으로도 매일 보게 될 준하형, 차분한 모습으로 고민을 들어주시고 마음이 즐거워지는 이야기를 나눌 수 있도록 해주신 유민이누나, 부드러운 모습 속에 마성의 매력을 숨기고 있는 좋은 형이자 동기인 탁구왕 광모형, 다양한 IT 지식과 따뜻한 마음씨로 재미있는 이야기뿐만 아니라 필요할 때마다 도움을 준 장훈이, 퀄컴 투어의 소중한 경험을 함께 했던 연구실에 활기를 불어넣는 다재다능한 명섭이, 늘 반가운 얼굴로 인사해주는 바른생활 사나이 동갑내기 세르게이, 회식 자리에서 조곤조곤 이야기를 나누었던 폴란드 친구 카밀, 앞으로 연구실의 미래를 이끌어나갈 승준이와 희제. 한분한분 모두가 저에게 잊지 못할 소중한 추억과 가르침을 주셨습니다. 정말로 감사드립니다. 졸업 후에도 CVLAB 출신이라는 자부심과 함께 한분한분이 주신 가르침과 고마움을 오래오래 간직할 것입니다. 비전세미나에서, 학회에서 동고동락하며 따뜻한 추억을 만들어주신 신호처리연구실의 상현이형, 해솔이형, 성민이형, 보준이형, 승연이형, 세라누나께도 깊은 감사를 전합니다.

마지막으로 사랑하는 가족들에게 무한한 감사를 전합니다. 늘 저를 믿고 응원해주시는 부모님의 사랑이 가득 담긴 목소리가 있었기에 힘들고 어려운 일이 있더라도 자신감을 가지고 용기를 얻을 수 있었습니다. 동생 정수와 함께 했던 공부와 오락도 저에게 큰 활력소였습니다. 가족이 있었기에 지금의 제가 있을 수 있었습니다. 항상 감사하고 사랑한다는 말을 전하고 싶습니다.

다시 한 번, 도움을 주신 모든 분들께 감사드립니다. 여러모로 부족한 제가 여기까지 무사히 올 수 있었던 것은 모두의 관심과 도움 덕분입니다. 늘 감사하고

사랑하는 마음으로, 받은 것을 세상에 베풀 수 있는 사람이 되겠습니다. 감사합니다.