Dual-lens motion deblurring

A Project Report

submitted by

SHARATH GIRISH

in partial fulfilment of the requirements
for the award of the degree of

BACHELOR OF TECHNOLOGY

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY, MADRAS.
May 2019

THESIS CERTIFICATE

This is to certify that the thesis entitled **Dual-lens motion deblurring**, submitted by **Sharath Girish** (**EE15B058**), to the Indian Institute of Technology, Madras, for the award of the degree of **Bachelors of Technology** is a bona fide record of the research work carried out by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr.A.N. Rajagopalan
Research Guide
Professor
Dept. of Electrical Engineering
IIT-Madras, 600 036

Place: Chennai

Date:

ACKNOWLEDGEMENTS

I would like to use this opportunity to express my gratitude to everyone who supported and assisted me through the course of this project. Without their involvement in every step throughout the process, this thesis would have never been accomplished.

Firstly, I would like to thank my thesis advisor, Prof. A.N. Rajagopalan for his continuous support during my research and for his motivation, patience and immense knowledge. I am sincerely grateful to my friend and lab mate Mahesh Mohan, without whose supervision and constant help, this thesis would not have been possible. I would also like to thank all my course teachers for providing the knowledge needed for me to carry out this project.

I am extremely grateful to my family for their continual support and encouragement throughout my life. Finally, I would like to thank my friends and batchmates for helping me out and guiding me whenever I was going through difficult times.

ABSTRACT

Dual-lens (DL) mobile phones that can mimic the bokeh feature of bulky DSLRs are now in vogue. However, motion blur due to camera shake is a ubiquitous phenomenon in handheld photography. Because one of the cameras in DL phones typically comes with a narrow field-of-view, the effect of motion blur becomes magnified. The consideration of 3D scenes further aggravates the DL deblurring problem. While image aesthetics is the main objective in conventional single-lens deblurring, any method for DL deblurring is additionally tasked with ascertaining binocularity in the deblurred image-pair. Ours is the first work to systematically address the problem of motion deblurring in DL phones. We unveil the shortcomings of the mature single-lens blur model for the DL configuration, and propose a generalized model that elegantly explains the intrinsically coupled image formation process in DL. We proceed to reveal an intriguing challenge that stems from an inherent ambiguity unique to DL deblurring which naturally disrupts binocularity. We address this issue by devising a judicious prior that also accounts for the unknown center-of-rotation. Based on these findings, we propose a DL blind deblurring method which exhibits state-of-the-art performance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Motion blur is an ubiquitous phenomenon in hand-held photography. A plethora of works have been proposed to address the issue of blur in images due to camera shake. Unlike the problem of non-blind deblurring (NBD) which has the blur kernel as well as the blurry image as input, blind motion deblurring (BMD) methods estimate a clean image and/or the blur kernel by only utilizing the blurry image. The blurring process is typically represented as the convolution of a blur kernel with the sharp image. Each approach proposes a blur model based on some assumptions and invert this model to obtain the clean image estimate.

These single image deblurring approaches can broadly be classified under 2 classes: Models assuming uniform or spatially invariant (SIV) blur [3, 37, 38, 31, 39, 26] and models assuming non-uniform or spatially variant (SV) blur [33, 36, 18, 27, 35]. The methods assuming SIV blur estimate a single blur kernel and subsequently perform deconvolution to obtain the sharp image. The methods of [33] and [5] show that blur due to camera shake is, in general, non-uniform across the image. The former method assumes that this blur is mostly due to 3D rotational motion of the camera. This blur model is predominantly followed in subsequent works [36, 18, 27, 35]. In addition to model based solutions for single-lens (SL) BMD, several deep learning based solutions exist [17, 29] as well.

However, no methods exist for the case of dual-lens (DL) setup. Most modern cameras come with DL configuration which consists of two cameras separated by a baseline. These cameras have, in general, dissimilar configurations such as different focal lengths or field-of-views (FOV), image resolution and exposure times. Recently, there has been a renewed interest in leveraging these multiple camera systems with unconstrained settings for various applications such as HDR imaging [20, 2, 28], low-light photography [32], super-resolution [9] and visual odometry [16, 8]. Many smartphones today are equipped with DL cameras, one with a narrow FOV (high focal length) and the other with a wide FOV. Higher focal length of the narrow angle camera further amplifies the effect of handheld shake on motion blur. The problem of BMD for DL cameras is fraught with additional challenges over those present in normal cameras. This DL-setup warrants deblurring based on scene depth. Additionally, any method for DL-BMD must ensure scene-consistent disparities in the deblurred image pair. Finally, SL BMD methods assume that the centre of rotation (COR) of the camera is at the optical center, whereas in practice it may be located at a point far away. No methods exist which handle the COR issue, which is also exacerbated by higher focal length of narrow-angle camera.

This method addresses the BMD problem for an unconstrained DL setup accounting for COR. There exists an inherent ill-posedness present in DL-BMD which is addressed with a well devised prior. Finally, the method proposes a strategy which decomposes the high-dimensional BMD problem into subproblems, while enforcing the prior and preserving the convexity. The main contributions are summarized below:

- This is the first attempt to formally address blind motion deblurring for unconstrained camera configurations. To this end, a generalized DL blur model is introduced that allows for arbitrary COR.
- There exists an inherent ill-posedness present in DL-BMD, that disrupts scene consistent disparities. To address this, a prior is introduced that ensures biconvexity and allows for efficient optimization.
- Employing the introduced model and prior, a practical DL-BMD method is proposed that achieves state-of-the art performance for current DL set-up.

CHAPTER 2

Related works

A plethora of works exist for the single-lens BMD problem. While initial works restricted to space-invariant blur [3, 37, 38, 31, 39, 26], the methods of [33] and [5] show that motion blur is, in general, space variant. [5] propose a blur model assuming in-plane translations and rotations while [33] assume a 3D rotational model. Building on [33], other works propose models with different types of priors such as unnatural L_0 prior [35], dark channel prior [18] and extreme channel prior [36].

For the case of DL cameras, Xu *et al.*[34] restrict to a constrained set-up, *i.e.*, they use two identical cameras working in full synchronization, so the same camera shake applies to both cameras. Importantly, the method imposes strong assumptions on blur that it is primarily caused by inplane translations and that the scene is fronto-parallel with layered depth. For the case of light field (LF) cameras, all multi-view images share identical camera configurations and ego-motions and hence cannot apply, in general, to unconstrained camera setups. Except [15], none of the LF BMD methods apply to this problem because their objective warrants optimizing for a 4D light field consisting of images with identical FOVs, resolutions and exposure times. Arun *et al.*[1] propose a method for multi-shot BMD, but employ 4 images and restrict to layered depth scenes. Moreover, they require all the images to be registered within a few pixels, over the entire image coordinates. Though this requirement can be met for multi-shot capture, it is easily violated for stereo images because of strong depth-dependent disparities. Hu *et al.*[7] jointly estimate depth and the clean image from a single blurred image. However, they restrict the blur to be primarily due to inplane translations. Recently, DL video deblurring methods have been proposed [19, 22], but they address dynamic objects and necessitate as input multiple stereo image-pairs.

CHAPTER 3

Single Lens Blind Motion Deblurring

3.1 Blur Model for Single-Lens System

Following the model of [33], if a camera undergoes 3D rotation about its optical centre, the change in world coordinate system is given by

$$\boldsymbol{X}^{'} = \boldsymbol{R}\boldsymbol{X} \tag{3.1}$$

where $\boldsymbol{R}$ is the rotational matrix describing the motion of the camera, $\boldsymbol{X} = [X, Y, Z]^T$ and $\boldsymbol{X}^{'} = [X^{'}, Y^{'}, Z^{'}]^T$ are the 3D world coordinates with respect to initial and final camera positions respectively. Additionally, the relation between the 2D image coordinates and the 3D world coordinates is given as

$$\boldsymbol{x} = \frac{1}{Z}\boldsymbol{K}\boldsymbol{X} \tag{3.2}$$

where $\boldsymbol{x} = [x, y, 1]^T$ represents the 2D image coordinates $[x, y]^T$ in homogenous coordinates, $\boldsymbol{K}$ is the camera's internal calibration matrix and Z is the Z coordinate or the depth of the scene point (assuming the image plane to be the $x - y$ plane).

Therefore, we get,

$$\begin{aligned}\boldsymbol{x}' &= \frac{1}{Z'}\boldsymbol{K}\boldsymbol{X}' \\ &= \frac{1}{Z'}\boldsymbol{K}\boldsymbol{R}\boldsymbol{X} \\ &= \frac{1}{Z'}\boldsymbol{K}\boldsymbol{R}(Z\boldsymbol{K}^{-1}\boldsymbol{x}) \\ &= \lambda\boldsymbol{H}\boldsymbol{x}\end{aligned} \tag{3.3}$$

where $\boldsymbol{H} = \boldsymbol{K}\boldsymbol{R}\boldsymbol{K}^{-1}$ represents the 2D homographic transformation between the 2 image coordinates $\boldsymbol{x}$ and $\boldsymbol{x}'$ and $\lambda = \frac{Z}{Z'}$ is the normalization constant.

The matrix $\boldsymbol{R}$ is described by an angle θ moved about some axis, consisting of 3 parameters. The matrix $\boldsymbol{K}$ is taken to be of the form

$$\boldsymbol{K} = \begin{bmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.4}$$

where f is the focal length of the camera and $[x_0, y_0]^T$ represents the coordinate on the image plane which is intersected by the camera optical axis.

Now, as the camera undergoes a sequence of rotations while the camera shutter is open for the exposure time t_e, the blurred image $\boldsymbol{I}_B$ is an integration of the sharp images corresponding to each rotation. Thus the equation is given as

$$\boldsymbol{I}_B = \frac{1}{t_e}\int_{t \in t_e} P(\boldsymbol{R}_t\boldsymbol{X})dt + \epsilon \tag{3.5}$$

where ϵ represents noise in the observed image, $\boldsymbol{R}_t$ is the matrix corresponding to the rotational pose at time t, and $P(\bar{\boldsymbol{X}})$ represents the image formed by projection of the world coordinates $\bar{\boldsymbol{X}}$ onto the image sensor. Note that this model assumes that the COR is at the optical centre.

As the blurry image is independent of the chronological order of the poses it undergoes, Eq. 3.5 can equivalently represented as a weighted integral over a set of camera rotations:

$$\boldsymbol{I}_B = \int_{\boldsymbol{\theta}\in\mathbb{P}^3} P(\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{X})\omega_{\boldsymbol{\theta}} d\boldsymbol{\theta} + \epsilon \tag{3.6}$$

Discretizing Eq. 3.6, we get

$$\boldsymbol{I}_B = \sum_{\boldsymbol{\theta}\in\mathbb{P}^3} P(\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{X})\omega_{\boldsymbol{\theta}} + \epsilon \tag{3.7}$$

where $\mathbb{P}^3$ represents the pose space containing all the possible rotations of the camera and $\omega_{\boldsymbol{\theta}}$ is the weight representing the fraction of exposure time the camera stayed in the pose $\boldsymbol{\theta}$.

Thus, the blurry image $\boldsymbol{I}_B \in \mathbb{R}^N$ (N is total number of pixels in the image) can be equivalently represented by a weighted summation of all the projectively warped versions of the clean image $\boldsymbol{I}_c \in \mathbb{R}^N$, where $\boldsymbol{I}_c = P(\boldsymbol{X})$. If the projective warping operation of $\boldsymbol{I}_c$ does bilinear interpolation then it can be represented as a linear transformation, *i.e.*,

$$P(\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{X}) = \boldsymbol{M(\theta)}P(\boldsymbol{X}) \tag{3.8}$$

where the $N \times N$ matrix $\boldsymbol{M}(\boldsymbol{\theta})$ does the projection corresponding to rotation $\boldsymbol{\theta}$. Thus Eq.3.7 can be represented as

$$\boldsymbol{I}_B = \sum_{k=1}^{K} \omega_k \boldsymbol{M}_k \boldsymbol{I}_c + \epsilon \tag{3.9}$$

where K is the total number of 3D poses in the pose space $\mathbb{P}_{\boldsymbol{\theta}}$, $\boldsymbol{M}_k$ and ω_k correspond to some camera orientation $\boldsymbol{\theta}_k$. The set of all weights can be represented by a single vector $\boldsymbol{\omega} \in \mathbb{R}^K$, called the Motion Density Function (MDF).

By closely observing Eq. 3.9, it can be seen that the blurry image $\boldsymbol{I}_B$ is linear in either $\boldsymbol{\omega}$ or $\boldsymbol{I}_c$ when the other quantity is known. Thus, it can be equivalently represented in 2 forms as:

$$\boldsymbol{I}_B = \boldsymbol{A}\boldsymbol{I}_C + \epsilon, or \tag{3.10}$$

$$\boldsymbol{I}_B = \boldsymbol{B}\boldsymbol{\omega} + \epsilon \tag{3.11}$$

where each row of the $N \times N$ matrix A is a local blur filter which acts on the clean image $\boldsymbol{I}_c$ to obtain the corresponding blurry image pixel at that row location and each column of the $N \times K$ matrix $\boldsymbol{B}$ contains a projectively transformed version of $\boldsymbol{I}_c$ for some rotational pose corresponding to the column location. The matrix A is very sparse as the blur filters only act on a small set of pixels of the clean image.

3.2 Latent image and kernel optimization

Since Eq. 3.10 and 3.11 are linear in the sharp image and the MDF respectively, it allows for efficient optimization of the two unknowns. However, optimizing for the best sharp image and MDF is ill-posed because of fewer equations than parameters. There can be multiple pairs of the sharp image estimate $\hat{\boldsymbol{I}}_C$ and the MDF estimate $\hat{\boldsymbol{\omega}}$ which yield the same blurry image $\boldsymbol{I}_B$. Therefore, in order to estimate the right solution, regularization and/or constraints are used on both the sharp image and kernel. Firstly, the kernel elements are restricted to be between 0 and 1 (as they represent fraction of exposure time) and sum up to 1. Additionally, handheld shake causes the camera to pass through only a small subset of all possible oritentations while the camera shutter is open. Thus, $\boldsymbol{\omega}$ will have very few non-zero elements and can have a sparsity prior imposed on it. For the case of image, typically, a hyper laplacian prior is imposed on the image gradients enforcing sparsity.

Representing the blurry image by g, the clean image $\mathbf{f}$ and the MDF by $\boldsymbol{\omega}$, we wish to get estimates $\hat{\mathbf{f}}$ and $\hat{\boldsymbol{\omega}}$ of the clean image and the MDF. This is obtained by maximizing the posterior distribution,

$$p(\mathbf{f}, \boldsymbol{\omega}|g) \propto p(g|\mathbf{f}, \boldsymbol{\omega})p(\mathbf{f})p(\boldsymbol{\omega}) \tag{3.12}$$

where $p(\mathbf{f})$ and $p(\boldsymbol{\omega})$ are the respective prior distributions for image and kernel, and the likelihood $p(g|\mathbf{f}, w)$ is obtained by assuming noise, ϵ, to be isotropic Gaussian:

$$p(g|\mathbf{f}, \boldsymbol{\omega}) \propto \prod_i exp\left(-\frac{(\hat{g}_i(\mathbf{f}, \boldsymbol{\omega}) - g_i)^2}{2\sigma^2}\right) \tag{3.13}$$

where σ is the noise standard deviation and $\hat{g}$ can be obtained by current estimates, $\hat{\mathrm{f}}$ and $\hat{\boldsymbol{\omega}}$.

A simple alternating minimization approach of Eq. 3.12 with respect to $\mathbf{f}$ and $\boldsymbol{\omega}$ does not yield the desired results. A series of nonlinear filtering and thresholding steps are added in order to ensure that the priors come into effect.

In the first step, the current estimate of clean image, $\hat{\mathbf{f}}$, is bilateral filtered and shock filtered in order to enhance the edges of the image (as homogeneous regions carry no blur information). Sparse gradient maps are then computed for this filtered image by finding its derivatives (upto the second order), $\{\mathbf{f}_x, \mathbf{f}_y, \mathbf{f}_{xx}, \mathbf{f}_{yy}, \mathbf{f}_{xy}\}$ and thresholding, so that only a small number of non-zero gradients are retained. However, some parts of the image might have smaller gradients. In order to ensure all regions are equally represented, the image is split into a 3×3 grid and thresholding is done independently for each cell of the grid. The steps above are also done for the blurry image g to get the gradient maps $\{g_x, g_y, g_{xx}, g_{yy}, g_{xy}\}$. The gradient maps obtained are then added to the following cost function which is minimized with respect to the kernel, $\boldsymbol{\omega}$ along with some constraints:

$$\begin{aligned} C(\boldsymbol{\omega}) &= \sum_{(\mathbf{f}_j, \boldsymbol{g}_j)} \|\hat{g}(\mathbf{f}_j, \boldsymbol{\omega}) - \boldsymbol{g}_j\|_2^2 + \beta \sum_k \boldsymbol{\omega}_k \\ \text{such that } \forall k &= 1, 2, \ldots, K, \boldsymbol{\omega}_k \geq 0 \end{aligned} \tag{3.14}$$

where each partial derivative $(\mathbf{f}_j, \boldsymbol{g}_j) \in \left\{(\mathbf{f}_x, \boldsymbol{g}_x), (\mathbf{f}_y, \boldsymbol{g}_y), (\mathbf{f}_{xx}, \boldsymbol{g}_{xx}), (\mathbf{f}_{yy}, \boldsymbol{g}_{yy}), (\mathbf{f}_{xy}, \boldsymbol{g}_{xy})\right\}$. Eq. 3.14 applies ℓ_1 regularization combined with non-negativity constraints allowing for the kernel to be sparse. Minimization of this cost function yields a MDF estimate $\hat{\boldsymbol{\omega}}$. Subsequently, we use $\hat{\boldsymbol{\omega}}$ to estimate the sharp image $\hat{\boldsymbol{I}}_c$ by inverting Eq. 3.10.

Image reconstruction can be done by various methods. One of the methods [10] does MAP estimation of the sharp image using the hyper-Laplacian prior on the image gradients. This is equivalent to maximizing the following posterior over $\mathbf{f}$:

$$p(\mathbf{f}|\boldsymbol{g}, \boldsymbol{\omega}) \propto p(f|\mathbf{g}, \boldsymbol{\omega}) p(\mathbf{f}) \tag{3.15}$$

where

$$p(\mathbf{f}) = \prod_i exp(-\lambda |f_i^x|^p) exp(-\lambda |f_i^y|^p) \tag{3.16}$$

where $p = 0.5$ to ensure sparsity of image gradients. Repeated optimizations of quadratic cost functions are done as in the work of [10].

Additionally, the method of Richardson-Lucy can be used for deconvolution which works especially well around saturated pixel regions where the linear model is not valid. The algorithm runs for multiple iterations while improving the current estimate $\hat{\mathbf{f}}$ as:

$$\hat{\mathbf{f}}_{i+1} = \hat{\mathbf{f}}_i \odot \left(\boldsymbol{A}^T (g \otimes \boldsymbol{A} \hat{\mathbf{f}}_i) \right) \tag{3.17}$$

where i is the iteration number, $\odot$ is elementwise multiplication, $\otimes$ is elementwise division and matrix $\boldsymbol{A}$ depends on the estimated non-uniform blur as in Eq. 3.10.

CHAPTER 4

Dual-Lens Blind Motion Deblurring

4.1 Blur Model for unconstrained Dual-Lens setups

(Sections 4.1 and 4.2 were written by Mahesh Mohan and are mentioned here for the sake of completeness.) This section explains how the single-lens model can be extended to a general DL setup. DL setups in most mobile phones consist of a single narrow-FOV camera (higher focal length) and a single wide-FOV camera separated by some baseline $\boldsymbol{t}_b$. Thus, at any instant of time, one camera will perceive the world, shifted by the stereo baseline. with respect to the other camera. Both the cameras need not have same image resolutions or exposure times. Denoting the parameters of the narrow angle and the wide angle camera by superscripts n and w respectively, the change in the world coordinate system for the two cameras can be represented by

$$\boldsymbol{X}^n = \boldsymbol{R}(\boldsymbol{X} - \boldsymbol{t}_c) + \boldsymbol{t}_c + \boldsymbol{t}_b \tag{4.1}$$

$$\boldsymbol{X}^w = \boldsymbol{R}(\boldsymbol{X} - \boldsymbol{t}_c) + \boldsymbol{t}_c \tag{4.2}$$

Thus a DL motion blurred image-pair ($\boldsymbol{I}_B^w$ and $\boldsymbol{I}_B^n$ can be represented by the following equations (based on Eq. 3.6):

$$\begin{aligned} \boldsymbol{I}_B^n &= \int_{\boldsymbol{\theta} \in \mathbb{P}^3} P^n(\boldsymbol{R}(\boldsymbol{X} - \boldsymbol{t}_c) + \boldsymbol{t}_c + \boldsymbol{t}_b)\omega_{\boldsymbol{\theta}}^n d\boldsymbol{\theta} + \epsilon^n \\ \boldsymbol{I}_B^w &= \int_{\boldsymbol{\theta} \in \mathbb{P}^3} P^w(\boldsymbol{R}(\boldsymbol{X} - \boldsymbol{t}_c) + \boldsymbol{t}_c)\omega_{\boldsymbol{\theta}}^w d\boldsymbol{\theta} + \epsilon^w \end{aligned} \tag{4.3}$$

(a) Scene depth (b) Blur kernel (PSF) (c) Blurred image (d) Image-patches

Figure 4.1: DL configuration warrants a *depth-variant* transformation (Note the variation of PSF in Fig. (b) with respect to scene depth in Fig. (a)). As the single-lens motion blur model is *depth-invariant*, the model optimized for a fixed depth can fail for other depths, leading to *ineffective* deblurring across depths (Fig. (d)).

The centre of the wide angle camera is taken as the origin. Notice the different projection functions ($P^n(.)$ and $P^w(.)$)for the two configurations as well as the different kernels ($\boldsymbol{\omega}^n$ and $\boldsymbol{\omega}^w$). From Eq. 3.2, we have the two image projection equations given by:

$$\boldsymbol{x}^n = \frac{1}{Z}\boldsymbol{K}^n(\boldsymbol{X} + \boldsymbol{t}_b) \tag{4.4}$$

$$\boldsymbol{x}^w = \frac{1}{Z}\boldsymbol{K}^w\boldsymbol{X} \tag{4.5}$$

Substituting Eq. 4.4 and Eq. 4.1 into the equation below,

$$\begin{aligned}
(\boldsymbol{x}')^n &= \frac{1}{Z'}\boldsymbol{K}^n(\boldsymbol{X}')^n \\
&= \frac{1}{Z'}\boldsymbol{K}^n\Big(\boldsymbol{R}(\boldsymbol{X} - \boldsymbol{t}_c) + \boldsymbol{t}_c + \boldsymbol{t}_b\Big) \\
&= \frac{1}{Z'}\boldsymbol{K}^n\Big(\boldsymbol{R}\Big(Z(\boldsymbol{K}^n)^{-1}\boldsymbol{x}^n - \boldsymbol{t}_b - \boldsymbol{t}_c\Big) + \boldsymbol{t}_c + \boldsymbol{t}_b\Big) \\
&= \frac{1}{Z'}\boldsymbol{K}^n\Big(Z\boldsymbol{R}\boldsymbol{K}^n)^{-1}\boldsymbol{x}^n + \boldsymbol{t}_c(\boldsymbol{I} - \boldsymbol{R}) + \boldsymbol{t}_b(\boldsymbol{I} - \boldsymbol{R})\Big) \\
&= \lambda\Big(\underbrace{\boldsymbol{K}^n\boldsymbol{R}(\boldsymbol{K}^n)^{-1}}_{\text{homography}}\boldsymbol{x}^n + \underbrace{\frac{1}{Z}\boldsymbol{K}^n(\boldsymbol{I} - \boldsymbol{R})\boldsymbol{t}_c}_{\text{center-of-rotation}} + \underbrace{\frac{1}{Z}\boldsymbol{K}^n(\boldsymbol{I} - \boldsymbol{R})\boldsymbol{t}_b}_{\text{baseline}}\Big)
\end{aligned} \tag{4.6}$$

where Z is the scene depth of the coordinate $\boldsymbol{x}^n$ and λ normalizes the third coordinate of $\boldsymbol{x}'$. Thus, Eq. 4.6 yields the pixel mapping for the narrow angle camera. A similar pixel mapping holds for the wide angle case, with the superscript n replace with w and with $\boldsymbol{t}_b = 0$. Eq. 4.6 contains a similar homography as the single-lens system but also has additional components such as the center-of-rotation $\boldsymbol{t}_c$ and the baseline $\boldsymbol{t}_b$. Unlike the single-lens pixel mapping in Eq. 3.3, Eq. 4.6 has a translational component and is dependent on the scene depth. The Point Spread Function (PSF), at the spacial coordinate $\boldsymbol{x}^n$ is obtained by super-imposing the pixel-mappings of $\boldsymbol{x}^n$ for all the poses which the camera undergoes, and weighed by the corresponding element of the MDF. PSFs quantify the motion blur, because the blurred image is obtained by convolving the space-variant PSFs with the latent image. As the pixel mapping in Eq. 4.6 is a function of COR, baseline, and depth, the MDF itself depends on these quantities unlike in the case of single lens. It is thus necessary to estimate depth and COR in order to completely characterize motion blur. To analyze this depth-variant nature of the kernel, Figs. 4.1(a-d)consider a camera trajectory and a 3D scene from [21].Fig. 4.1(b) shows the corresponding kernels (projected using Eq. (4.6)), which clearly reveals depth-dependency of blur, with lower depths exhibiting severe blurs relative to the farther ones. Figure 4.1(d) shows the deblurred image-patches for different depths employing the normal camera method [35], optimized for a given depth; it is evident that this approach is *not* quite successful due to the *depth-dependency* of the blur, which clearly necessitates a new approach for DL-BMD. While single image methods model blur as independent of depth, as recovering depth from a single blurry image is a difficult problem, pixel disparities between images can be used to obtain a depth estimate in the dual-lens case.

4.2 Costs and Prior for unconstrained DL-BMD

Similar to the cost function in Eq. 3.14, the joint cost for DL-BMD is $L = L^n + L^w$, where:

$$L^k = ||\boldsymbol{B}^k \boldsymbol{\omega}^k - \boldsymbol{I}_B^k||_2^2 + \alpha^k ||\boldsymbol{\omega}^k||_1 + \beta^k ||\nabla \boldsymbol{I}_c^k||_1, \tag{4.7}$$

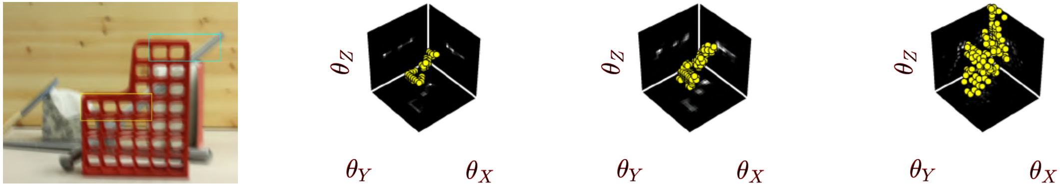
where $k \in \{n, w\}$ and $||\boldsymbol{B}^k \boldsymbol{\omega}^k - \boldsymbol{I}_B^k||_2^2 = ||\boldsymbol{A}^k \boldsymbol{I}_c^k - \boldsymbol{I}_B^k||_2^2$ is used depending on minimization of either kernel or latent image, as per Eq. 3.11 or Eq. 3.10. Note that the matrices $\boldsymbol{A}_k$ and $\boldsymbol{B}_k$ are depth and COR dependent as explained in section 4.1.

The estimated deblurred image-pair $\boldsymbol{I}_c^n, \boldsymbol{I}_c^w$ must obey the scene binocularity, where the narrow-angle camera perceives the same scene-orientation as the wide-angle camera, but displaced by baseline $\boldsymbol{t}_b$. However, directly optimizing for the DL-BMD cost L yields multiple valid solutions of the deblurred image pair, which do not adhere to the scene consistent disparities. A desired solution which minimizes the cost L is $\{P^n(\boldsymbol{X} + \boldsymbol{t}_b), P^w(\boldsymbol{X})\}$, which is referred to as the true image-pair and $\{\omega^n(p), \omega^w(p)\}$ referred to as the true MDF-pair. Now, Eq. 4.3 can be equivalently written in discrete form (akin to Eq. 3.7) as

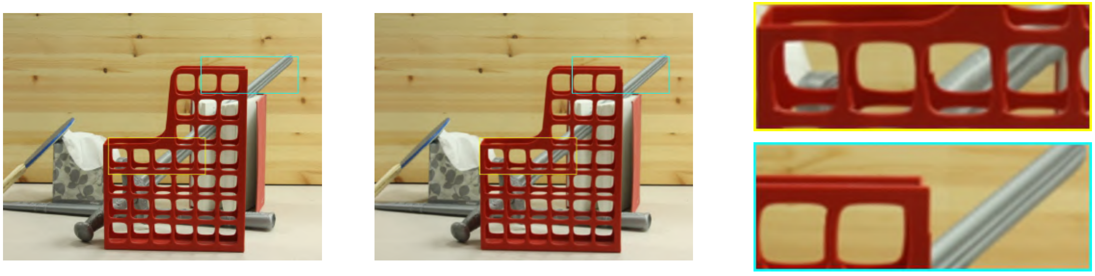
$$\begin{aligned} \boldsymbol{I}_B^n &= \sum_p \omega^n(p) P^n\Big(\boldsymbol{R}_p \boldsymbol{R}_n^{-1} \boldsymbol{R}_n(\underbrace{\;\boldsymbol{X}\;}_{true} - \boldsymbol{t}_c) + \boldsymbol{t}_c + \boldsymbol{t}_b\Big) \\ &= \sum_p \omega^n(p) P^n\Big(\boldsymbol{R}_p \boldsymbol{R}_n^{-1}\big(\underbrace{\boldsymbol{R}_n(\boldsymbol{X} - \boldsymbol{t}_c) + \boldsymbol{t}_c}_{apparent} - \boldsymbol{t}_c\big) + \boldsymbol{t}_c + \boldsymbol{t}_b\Big) \end{aligned} \tag{4.8}$$

where the new scene-orientation is $\boldsymbol{R}_n(\boldsymbol{X} - \boldsymbol{t}_c) + \boldsymbol{t}_c$, with $\boldsymbol{R}_n \neq I$. The rotation matrix $\boldsymbol{R}_n$ shifts all the true poses by an offset of $\boldsymbol{R}_n^{-1}$, which produces a shifted version of the true MDF (sparsity cost remains same). Thus, a new solution is the image-pair

(a) Wide-angle blurred (b) True MDF (c) MDF *with* prior (d) MDF *without* prior



(e) Narrow-angle (f) Wide-angle (g) Wide-angle image-patches



(1) Ground truth

(2) *With* Prior

(3) *Without* Prior

Figure 4.2: Effect of the proposed prior: The MDF estimate and the deblurred image-patches of prior-less case clearly show a significant rotational ambiguity (Figs. (g,3) d). Also, the deblurred image in the prior-less case exhibits considerable ringing artifacts and residual blur (Figs. (e-f,3)), which could be possibly due to the less accurate MDF estimate (Fig. d). In contrast, the addition of the prior successfully curbs the pose ambiguity (Figs. (g,2),c), improves the MDF accuracy (Fig. c) and produces better deblurring quality (Figs. (e-f,2)).

$\{P^n(\boldsymbol{R}_n(\boldsymbol{X} - \boldsymbol{t}_c + \boldsymbol{t}_c + \boldsymbol{t}_b), P^w(\boldsymbol{X})$, which do not obey scene binocularity (narrow angle camera perceives a different scene orientation). The TV prior cost on the image remains same as the new narrow-angle image is a warped version of the true image. The cost L^w remains same and hence, the above solution minimizes L as well. A similar ambiguity occurs in the wide angle case as well, where the resultant image pair becomes $\{P^n(\boldsymbol{R}_n(\boldsymbol{X} - \boldsymbol{t}_c) + \boldsymbol{t}_c + \boldsymbol{t}_b), P^w(\boldsymbol{R}_w(\boldsymbol{X} - \boldsymbol{t}_c) + \boldsymbol{t}_c)\}$ (in general, $\boldsymbol{R}_n \neq \boldsymbol{R}_w$). The ill-posedness exists irrespective of whether the exposure time is same or different in the two cameras. This is tackled by introducing a prior which assumes some overlap between the two exposure times. The prior is motivated by the fact that the image-pair will be scene-consistent if $\boldsymbol{R}_n = \boldsymbol{R}_w$. For identical exposure time, this criterion requires that both the MDFs completely intersect over the pose-space. For overlapping exposure time, both MDFs must intersect over the shared poses. Hence a prior of the form $||\boldsymbol{\omega}^n - \boldsymbol{\omega}^w||_2$ is included in the cost. The priorless DL BMD cost L can allow MDF pairs with significant drifts but the prior cost increases for such pairs which prevents it from being the minima of the overall cost. The proposed DL prior preserves the biconvexity property of the cost which guarantees convergence by alternative minimization. The effect of the proposed prior is highlighted in Fig. 4.2.

4.3 A practical algorithm for DL-BMD

Most single lens BMD methods work in a scale-space manner while optimizing for the MDF and the latent image. This involves alternative minimization from a coarse to fine image scale over a number of iterations each scale, in order to account for large blurs. For the first scale and first iteration, the latent images, MDFs and COR are initialized as shock-filtered blurry images, Kronecker delta and optical centre respectively.

Computational complexity is an important factor for such iterative algorithms. In order to have a fast and accurate optimization scheme, Efficient Filter Flow (EFF) is used in the deblurring pipeline (as per the work of [6]).

4.3.1 Blurring using EFF

As performing space variant convolution for every pixel in the image is computationally expensive, in this approach, motion blur is approximated as space invariant convolutions in small individual patches of the image.

Relaxing the superscript notation of n and w (both cameras will have similar equations), this is represented as

$$\boldsymbol{I}_B = \sum_{m=1}^{M} \boldsymbol{D}^{\dagger}_{m}. \Big(\boldsymbol{h}_m * (\boldsymbol{D}_m.\hat{\boldsymbol{I}}_c)\Big) \tag{4.9}$$

where M is the total number of overlapping patches in the latent image $\hat{\boldsymbol{I}}_c$, $\boldsymbol{h}$ is the blur kernel or the PSF (explained in Sec. 4.1) which is convolved with the m^{th} image patch, $\boldsymbol{D}_m.\hat{\boldsymbol{I}}_c$ is a linear operator which extracts the m^{th} patch from $\hat{\boldsymbol{I}}_c$ and $\boldsymbol{D}^{\dagger}_{m}$ inserts it back into the original position with a Barlett windowing operation. Thus, each patch is represented by a single blur kernel allowing for efficient space invariant convolution for the patch. The PSF is obtained as follows:

$$\boldsymbol{h}_m = \sum_{p \in \mathbb{P}^3} \omega(p)\boldsymbol{\delta}_m(p) \tag{4.10}$$

where $\boldsymbol{\delta}_m(p)$ is an impulse which is shifted by transforming, according to some pose p, an impulse centred at the m^{th} patch center. The transformation is based on the depth and COR dependent pixel mapping in Eq.4.6. Thus the blur kernel at the m^{th} patch center is just a superposition of these shifted impulses weighed by the MDF. The impulses are calculated only once, for each patch k, and can be used to create the blur kernel for any image. Thus, given a latent image estimate $\hat{\boldsymbol{I}}_c$ and MDF estimate $\hat{\boldsymbol{\omega}}$, the blurring process for EFF computes kernels in the M patch centres using the MDF and the precomputed $\boldsymbol{\delta}_m(p)$, convolves it with the corresponding image patch and combines it to yield the full image. Efficient convolution can be done in the frequency domain by using Fast Fourier Transforms. This eliminates the need for building large warp matrices and carrying out expensive matrix multiplication in order to obtain the blurry image.

4.3.2 Depth and COR estimation

As depth from stereo is a well-studied problem, an off-the-shelf algorithm for depth estimation is selected [13], owing to its good trade-off between accuracy and speed [12, 25]. Disparity is obtained by resizing the current wide-angle image estimate to the size of the narrow-angle image estimate, which is used as the reference. Once the disparity is computed, depth of the scene can be obtained by utilizing the baseline and focal length information of the dual-lens camera. To estimate the COR, a least squares cost is considered between the observed blurry image and the synthesized blurry image using the blur model in Eq. 4.9 and the current image and MDF estimates. The cost is framed in the gradient domain and involves gradient map estimation and thresholding as in Section. 3.2. The optimization for the COR is given by:

$$\hat{\boldsymbol{t}}_c = \underset{\boldsymbol{t}_c}{\operatorname{argmin}}(L^n_{\boldsymbol{t}_c} + L^w_{\boldsymbol{t}_c}) \tag{4.11}$$

where,

$$L_{t_c}^k = \left\| \sum_j \left(g_j - \hat{g}(\hat{\mathbf{f}}_j^k, \hat{\boldsymbol{\omega}}^k) \right) \right\|_2^2 \tag{4.12}$$

$j \in \{x, y, xx, yy, xy\}$, $k \in \{n, w\}$, $\hat{\mathbf{f}}_j^k$ is the clean image estimate and $\hat{\boldsymbol{\omega}}^k$ is the MDF estimate for the corresponding configuration. A trust-region reflective algorithm [4] is used for optimizing Eq. 4.12, which is initialized with the COR estimate of the previous iteration.

4.3.3 Kernel estimation

With respect to the cost function for kernel estimation in Eq. 3.14, we have $\hat{g}(\mathbf{f}_j, \boldsymbol{\omega}) = \boldsymbol{B}_j\boldsymbol{\omega}$ and each column of the matrix $\boldsymbol{B}_j$ contains a projectively transformed version of $\mathbf{f}_j$ for some rotational pose corresponding to the column location, as in Eq. 3.11. Taking the wide-angle camera as reference, we get the MDF objective function as

$$\hat{\boldsymbol{\omega}}^w = \underset{\boldsymbol{\omega}^w}{\operatorname{argmin}} \sum_j ||\boldsymbol{B}_j^w \boldsymbol{\omega}^w - \boldsymbol{g}_j^w||_2^2 + \beta||\boldsymbol{\omega}^w||_1 \tag{4.13}$$

where the superscript w implies the required quantities correspond to the wide angle camera. The cost function in Eq. 4.13 applies ℓ_1 regularization which helps optimize for a sparse kernel. Adding the prior to the narrow angle cost, we get

$$\hat{\boldsymbol{\omega}}^n = \underset{\boldsymbol{\omega}^n}{\operatorname{argmin}} \sum_j ||\boldsymbol{B}_j^n \boldsymbol{\omega}^n - \boldsymbol{g}_j^n||_2^2 + \beta||\boldsymbol{\omega}^n||_1 + \alpha||\boldsymbol{\omega}^n - \hat{\boldsymbol{\omega}}^w||_2^2 \tag{4.14}$$

In order to reduce ill-conditionness and to improve convergence, the cost function is minimized in the gradient domain instead. This is done by separating out the sparsity prior as a constraint and taking the derivative yielding:

$$\begin{aligned}\hat{\boldsymbol{\omega}}^w = \underset{\boldsymbol{\omega}^w}{\operatorname{argmin}} \sum_j \left\| (\boldsymbol{B}_j^w)^T \boldsymbol{B}_j^w \boldsymbol{\omega}^w - (\boldsymbol{B}_j^w)^T \boldsymbol{g}_j^w \right\|_2^2 \\ \text{such that } \|\boldsymbol{\omega}^w\|_1 \leq \beta^{'}\end{aligned} \tag{4.15}$$

$$\begin{aligned}\hat{\boldsymbol{\omega}}^n = \underset{\boldsymbol{\omega}^n}{\operatorname{argmin}} \sum_j \left\| \left((\boldsymbol{B}_j^n)^T \boldsymbol{B}_j^n + \alpha I \right) \boldsymbol{\omega}^n - \left((\boldsymbol{B}_j^n)^T \boldsymbol{g}_j^n + \alpha \hat{\boldsymbol{\omega}}^w \right) \right\|_2^2 \\ \text{such that } \|\boldsymbol{\omega}^n\|_1 \leq \beta^{'}\end{aligned} \tag{4.16}$$

where I is the identity matrix of the same shape as the square matrix $(\boldsymbol{B}_j^n)^T \boldsymbol{B}_j^n$. Eq. 4.15 and 4.16 are LASSO of the form $\operatorname{argmin}_x \|\boldsymbol{Ax} - \boldsymbol{b}\|_2, \|\boldsymbol{x}\|_1 \leq \gamma$ which can be optimized with many efficient solvers (LARS [30]).

(Relaxing the superscript notation n, w and referring to a general case).

Naively using the large dense $N \times K$ matrix $\boldsymbol{B}_j$, to obtain $\boldsymbol{B}_j^T \boldsymbol{B}_j$ has a large memory requirement as well as computational cost for performing the matrix multiplication. Employing the EFF structure, the image is split into M overlapping patches, as in Subsection. 4.3.1 and the matrices $\boldsymbol{B}_j^T(m) \boldsymbol{B}_j(m)$ are computed by direct matrix multiplication of $\boldsymbol{B}_j(m)$ with its transpose, for the m^{th} patch. Each of these M matrices can be computed independently of each other and have much smaller matrix dimensions. The results are then added to obtain the full $\boldsymbol{B}_j^T \boldsymbol{B}_j$ matrix. A similar method is also adopted to obtain the $\boldsymbol{B}_j^T \boldsymbol{g}_j$ vector.

4.3.4 Latent image estimation

Using the MDF estimates $\{\hat{\boldsymbol{\omega}}^n, \hat{\boldsymbol{\omega}}^w\}$ at the current iteration, this step estimates the latent image pair $\{\hat{\boldsymbol{I}}_c^n, \hat{\boldsymbol{I}}_c^w\}$ which ensures scene consistent disparities. Since the MDF estimation already utilizes the image gradient information, image priors are not always necessary for this step of the alternative minimization. The latent image is thus obtained by inverting the forward blurring process in Eq. 4.9. Again, employing the EFF structure, each patch of the blurred image is obtained and deconvolved with the corresponding PSF using FFT and then combined using the windowing operation, to get the latent image estimate. The equation is thus given as

$$\boldsymbol{I}_C = \sum_{m=1}^{M} \boldsymbol{D}_m^{\dagger}.\mathbb{F}^{-1}\left(\mathbb{F}(\boldsymbol{D}_m.\boldsymbol{I}_B) \oslash \mathbb{F}(\boldsymbol{h}_m)\right) \tag{4.17}$$

where $\oslash$ represents the element-wise division while ensuring numerical stability for small values of the PSF $\boldsymbol{h}_m$. $\mathbb{F}$ and $\mathbb{F}^{-1}$ represent the forward and inverse DFT operations respectively.

For the final iteration at the finest scale, instead of direct deconvolution in Eq. 4.17, the Richardson-Lucy deconvolution algorithm [14] is used instead as shown in Eq. 3.17.

The same equations are used for the narrow-angle and wide-angle cases, with the quantities in the equation replaced with the ones corresponding to the narrow-angle and wide-angle configuration respectively.

CHAPTER 5

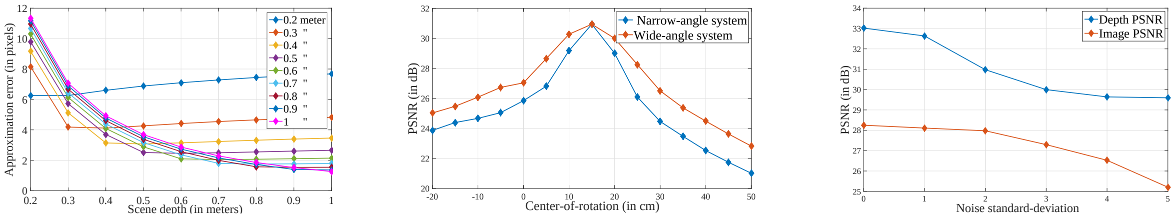
Analysis and experimental results

5.1 Sensitivity of COR

(This chapter was written by Mahesh Mohan while the results were gathered by me.) To analyze the sensitivity of COR for narrow-angle and wide-angle configurations, we considered images blurred with a common COR , and performed deblurring by perturbing the COR vector and using the true ego-motion (*identically* for both the configurations). Figure 5.1(b) compares the average PSNR of deblurred images for different COR approximations. The figure clearly shows a significant drop in deblurring performance as the approximated COR deviates from the true COR. Also, note the detrimental effect of the common COR approximation about the camera center (that is followed in single-lens BMD methods). The figure also reveals higher sensitivity of COR in narrow-angle configuration as shown by the higher rate of its performance-drop. This is due to higher focal-length, and hence larger blur inherent in narrow-angle setup which is a function of COR.

5.2 Noise analysis

To analyze the effect of noise in our DL-BMD method, we experimented with blurry images corrupted with additive white Gaussian noise. Standard-deviation of noise (in pixels) is varied from 0 (noise-less case) to 5. Fig. 5.1(c) plots the average PSNRs



(a) Model inaccuracies (b) Sensitivity-analysis of COR (c) Noise analysis

Figure 5.1: Analysis: (a) Model inaccuracies of the homography model. (b) Sensitivity of COR: Both narrow-angle and wide-angle configurations are very sensitive to COR, with the former exhibiting relatively more sensitivity. (c) Effect of image noise on our method.

of a deblurred image and depth estimate corresponding to different noise levels. The average PSNRs for deblurred image and depth-estimate is more than 25 dB and 29 dB, respectively, over the entire standard-deviation range; this clearly reveals the noise-robustness of our algorithm. Although we did *not* perform denoising in any examples, for *very* high noisy levels, the blurred image-pair need to be denoised prior to deblurring. This is because noise can deteriorate image-gradients which are required for ego-motion estimation.

5.3 Generalizability and effect of prior and COR

The theory and method presented directly apply to DL cameras with entirely different settings. They hold well for identical cameras ($f^n = f^w$) or camera arrays (multiple baselines), wherein exposures are different ($\boldsymbol{\omega}^n \neq \boldsymbol{\omega}^w$) or identical ($\boldsymbol{\omega}^n = \boldsymbol{\omega}^w$). They also work for normal camera methods ($\boldsymbol{t}_c = \boldsymbol{t}_b$). The table 5.2 shows the generalizability of the proposed algorithm for the three DL-configurations: Narrow-Narrow, Narrow-Wide, Wide-Wide. We consider the same exposure time for both cameras, 52mm focal length for narrow angle camera and 26mm focal length for the wide angle camera. The values reported in the Table 5.2 are averaged over

PSNR (dB)	Blur	W/o Prior W/o COR	W/o Prior W/ COR	W/ Prior W/o COR	W/ prior W/ COR
Image	22.39	25.69	26.59	27.28	28.88
Depth	28.33	23.35	**23.59**	29.12	**30.52**

Table 5.1: Quantitative results of our method with and without the DL prior and COR. In particular, our DL prior reduces the ill-posedness by a good margin (*i.e.*, by 7 dB, as indicated in bold).

three examples. As can be seen, our method performs consistently better than the methods of [34, 15] in all three configurations.

In order to analyze the effect of the DL prior and COR, PSNR results for image/depth are gathered by averaging over five examples by including or excluding the two quantities. Table 5.1 summarizes the results. For creating the synthetic dataset, exposure overlap and COR are randomly sampled from 10 to 100% and -30 cm to 30 cm cube, respectively. The unconstrained setup employed is narrow-FOV and wide-FOV pair, with $f^n = 52mm$, $f^w = 26mm$ and the former having twice the resolution of the latter (As in Samsung S9+). The depth information gets significantly corrupted for the prior-less case (PSNR drops by 7 dB). This underlines the importance of resolving the pose-ambiguity in dual-lens BMD. The deblurring performance also drops by 2.3 dB in the prior-less case, possibly be due to the loss of reinforcement between the narrow-angle and wide-angle costs (as discussed earlier). Further, the table reveals that both image and depth accuracies deteriorate when COR issue is not addressed, i.e., image and depth PSNRs drop by 1.6 dB and 1.3 dB, respectively.

Thus, our method can seamlessly address partial and full exposure-overlaps ([14, 29, 47, 30, 24]), without any modifications.

Configuration	Blurred	Xu *et al.*[34]	Mohan et al. [15]	Ours
Narrow-Narrow	27.27 / 1.75 / 0.23	19.90 / 1.08 / 0.22	29.21 / 2.30 / 0.36	31.03 / 3.04 / 0.43
	29.22	15.83	29.50	30.35
Narrow-Wide	27.33 / 1.78 / 0.23	19.86 / 1.13 / 0.22	26.50 / 1.95 / 0.31	30.50 / 3.10 / 0.42
	28.51	15.29	28.56	31.11
Wide-Wide	27.87 / 1.97 / 0.27	14.56 / 0.94 / 0.17	25.90 / 2.04 / 0.32	30.64 / 4.40 / 0.56
	30.15	13.88	28.56	30.62

Table 5.2: Generalizability to diverse DL set-ups: Our method consistently outperforms the methods of [34, 15] in the PSNR, IFC and VIF metrics respectively, for image (top part of rows) and the PSNR metric for depth (bottom part of rows).

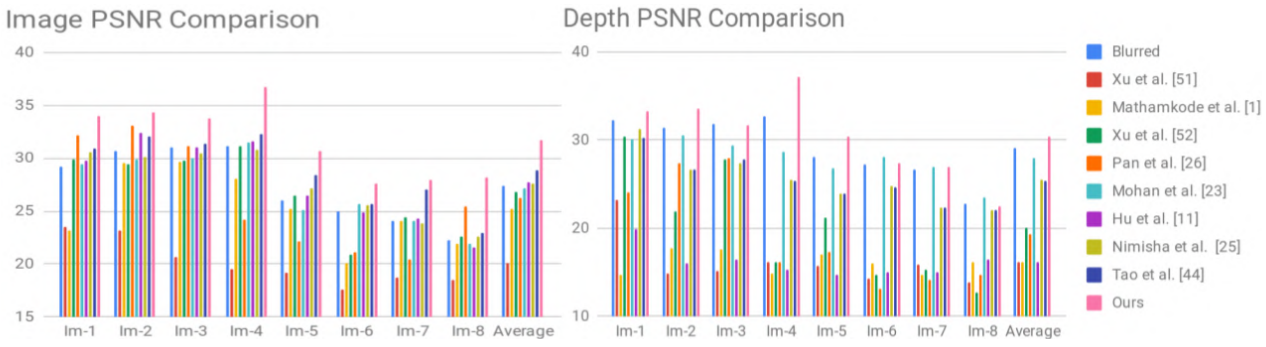


Figure 5.2: Quantitative evaluations using objective measure (PSNR). Our method performs competitively against the state-of-the-art, and produces the least depth errors.

5.4 Quantitative and Qualitative evaluations

In this section, the algorithm is evaluated on both synthetic and real examples.

Comparison methods: The works of [18, 35] are considered to represent normal camera BMD. For computational cameras, state-of-the-art stereo BMD [34] and light field BMD [15] are used as comparisons. The single-image BMD work of [7] and the multi-image method of [1] are considered for depth-aware case. Finally the deep learning works which are comparisons are the works of [17] and [29] representing auto encoder and recurrent neural networks respectively.

Metrics: The metrics used for quantitative evaluation of images are PSNR, IFC [24]

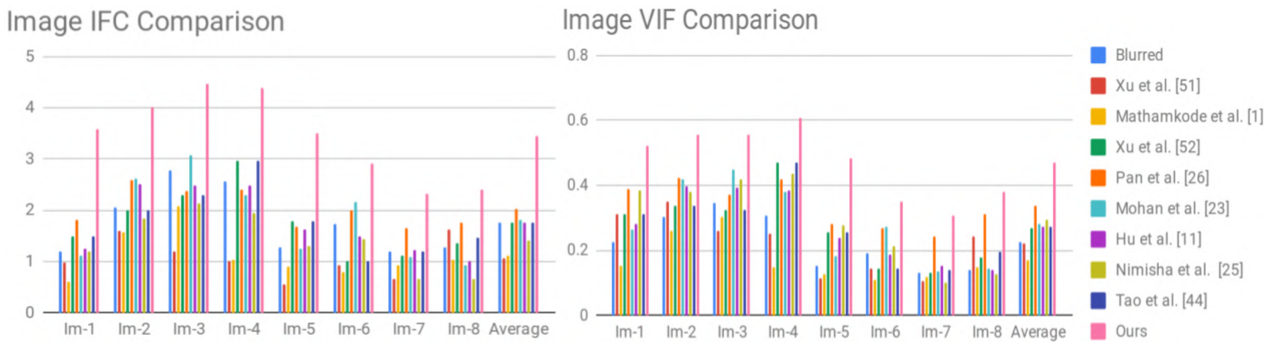
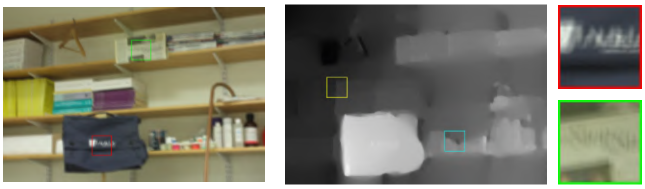


Figure 5.3: Quantitative evaluations using subjective measures (IFC, VIF). Our method performs deblurring with the best aesthetics.

and VIF [23]. IFC and VIF are shown to be the best metrics for subjective evaluation of BMD [11]. For qualitative evaluation, we provide the narrow-FOV image and (normalized) depth estimated from deblurred image-pair or by algorithms [7, 1].

Quantitative Evaluations: Figures 5.2-5.3 provide objective and subjective measures for different methods. Both the measures of the state-of-the-art DL-BMD [34] clearly reveal its high sensitivity, when it deviates from the assumptions of synchronized and identical cameras, and layered depth scenes. This once again emphasizes the need for an unconstrained DL-BMD method. For normal camera methods [18, 35], there is a perceivable drop in the depth performance, which clearly suggests their inadequacy in DL set-up. While the inferior depth performance of [1] can be attributed to its assumption of layered depth, for [7], it can also be due to its single image restriction. As compared to our method, light field BMD [15] is not quite successful (i.e., image/depth PSNR is less by 2.37/4.47 dB). This can be attributed to its lens effect and assumption of synchronized and identical camera settings. Finally, our method outperforms deep learning methods [17, 29] by 3.50 dB and 2.72 dB for image and 4.39 dB and 4.36 dB for depth, respectively. Based on the claims of [17, 29] that they generalize well for real-captured images, this performance degradation could be possibly due to the unique characteristics of unconstrained DL blur.

Qualitative evaluations: Figures 5.4-5.13 provide visual results for synthetic [21] and real experiments. We wish to highlight that ringing artifacts in deblurring are mainly caused by ego-motion error, which can be either due to inaccurate blur/ego-motion model or ineffectiveness of optimization. It can be seen that depth estimation is *also* sensitive to ringing artifacts; one reason could be that ringing deteriorates the feature matches required for depth estimation. The deblurred images of [34, 1] exhibit severe ringing artifacts (possibly due to the assumptions on scene and ego-motion and capture settings). Also, note that [7] produces erroneous layered-depth estimates (*e.g.*, nearer depths appear to be farther, as in Fig. 5.6, first row, chandelier). This is due to its sole restriction to single image cues for depth sensing. The results of [15, 18, 35] amply demonstrate the inadequacy of light field and single-lens BMD in the dual-lens setup, where the deblurring is *not* uniform over different depth levels (*e.g.*, in Fig. 5.5, fifth row, the closer books and farther windows are *not* simultaneously accounted for) and exhibits perceivable ringing artifacts, (*e.g.*, in Fig. 5.6, first row, over the chandelier). The visual results of deep learning methods [17, 29] once again prove that they are inadequate to deal with DL blur. When compared with the competing methods on all the examples, it is evident that our DL deblurring method consistently accounts for features at different depths, produces lesser ringing artifacts, and faithfully preserves consistent depth information.



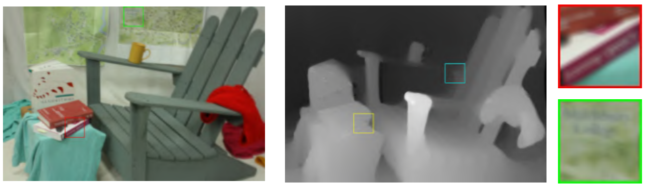
(a) Blurry image

(b) Ours (c) Mohan *et al.*[15] (d) Pan *et al.*[18]

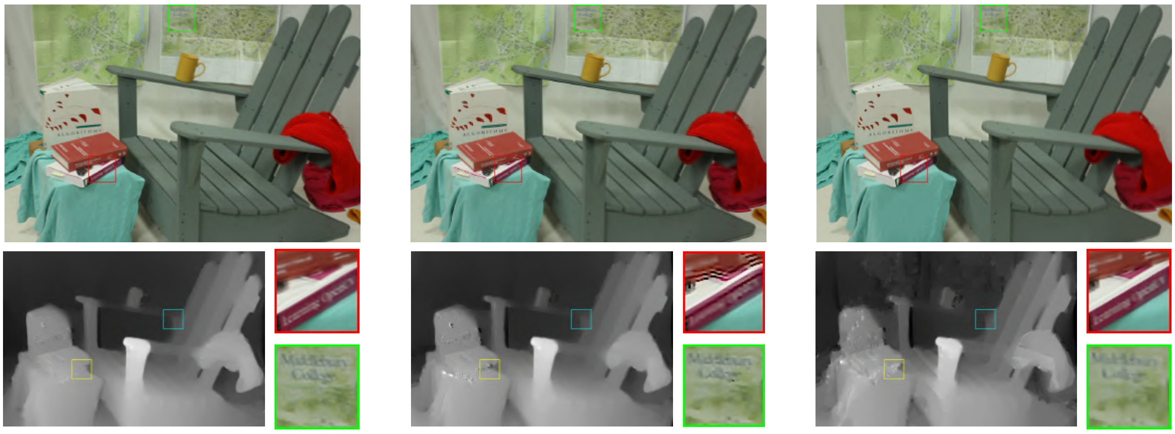
(e) Tao *et al.*[29] (f) Xu *et al.*[35] (g) Xu *et al.*[34]

(e) Mathamkode *et al.*[1] (f) Hu *et al.*[7] (g) Nimisha *et al.*[17]

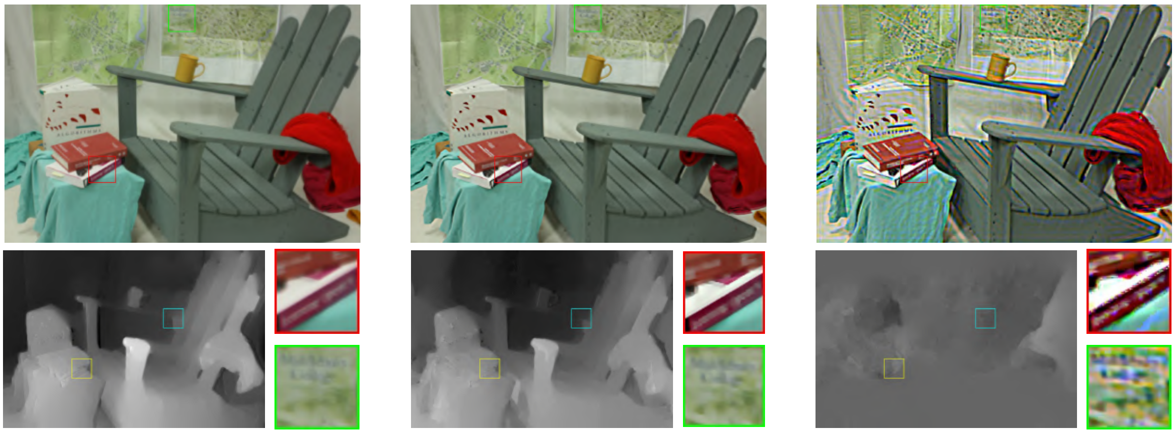
Figure 5.4: Synthetic experiments: Our method is able to retrieve the finer details at different depth levels with little ringing. The text in the patches are sharper when compared with the other methods.



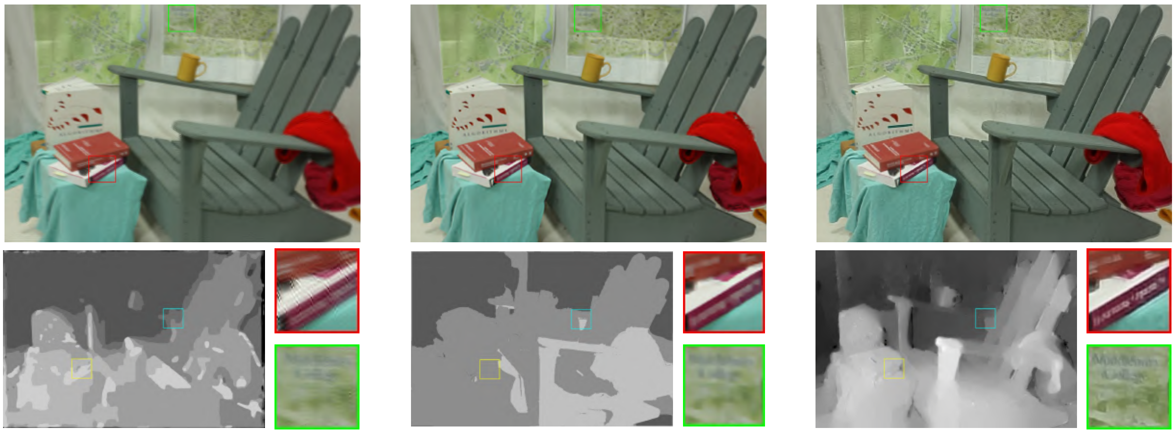
(a) Blurry image



(b) Ours (c) Mohan *et al.*[15] (d) Pan *et al.*[18]

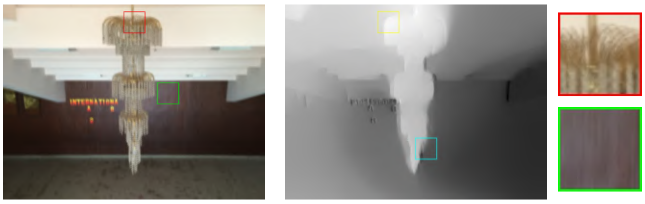


(e) Tao *et al.*[29] (f) Xu *et al.*[35] (g) Xu *et al.*[34]

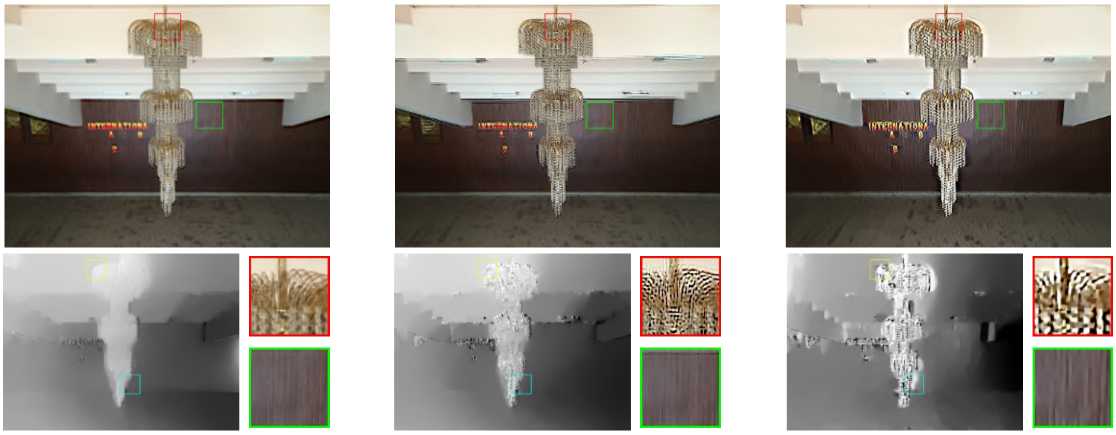


(e) Mathamkode *et al.*[1] (f) Hu *et al.*[7] (g) Nimisha *et al.*[17]

Figure 5.5: Synthetic experiments: Our method is able to retrieve the finer details at different depth levels with little ringing. The text in the patches are sharper when compared with the other methods.

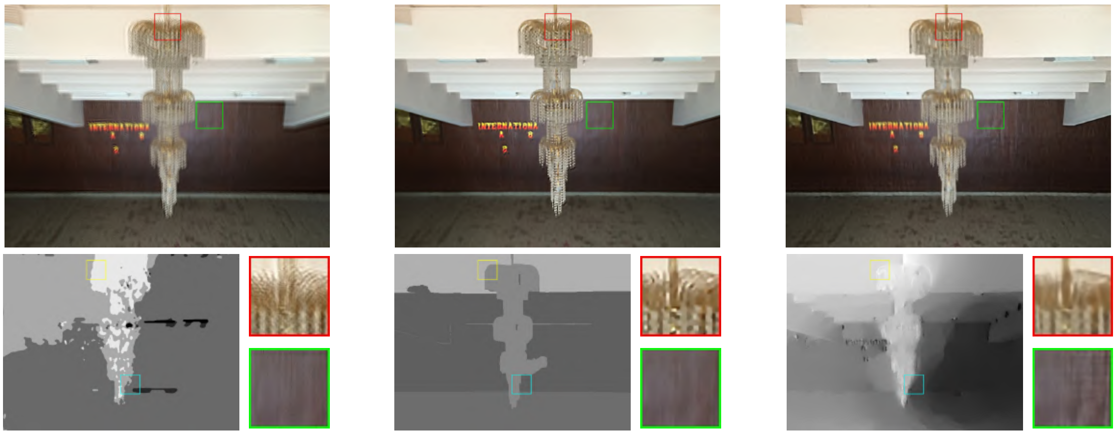


(a) Blurry image



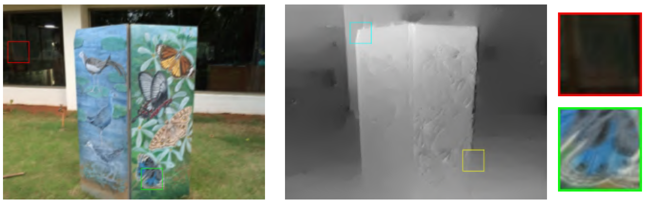
(b) Ours (c) Mohan *et al.*[15] (d) Pan *et al.*[18]

(e) Tao *et al.*[29] (f) Xu *et al.*[35] (g) Xu *et al.*[34]



(e) Mathamkode *et al.*[1] (f) Hu *et al.*[7] (g) Nimisha *et al.*[17]

Figure 5.6: Real experiments: Unlike the results of [15, 18, 34], our method is able to deblur the image with no artifacts while also recovering the finer details on the chandelier and the wall. Also, note the ineffectiveness of the single-lens methods [7, 18, 35] in DL configuration.



(a) Blurry image

(b) Ours (c) Mohan *et al.*[15] (d) Pan *et al.*[18]

(e) Tao *et al.*[29] (f) Xu *et al.*[35] (g) Xu *et al.*[34]

(e) Mathamkode *et al.*[1] (f) Hu *et al.*[7] (g) Nimisha *et al.*[17]

Figure 5.7: Real experiments: The results of the deep learning methods of [17, 29] exhibit a significant amount of residual blur. Our method is able to faithfully preserve the depth information and exhibits the least ringing artifacts.

(a) Blurry image

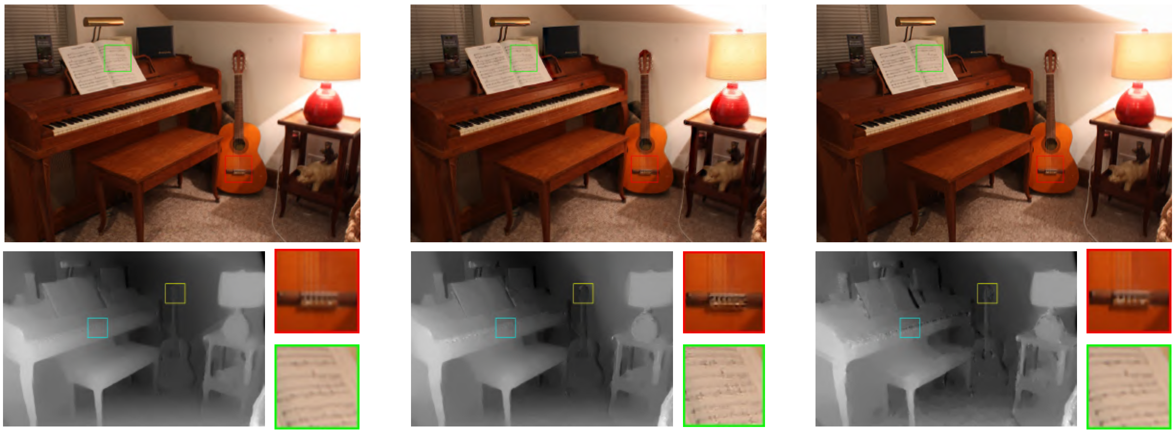
(b) Ours (c) Mohan *et al.*[15] (d) Pan *et al.*[18]

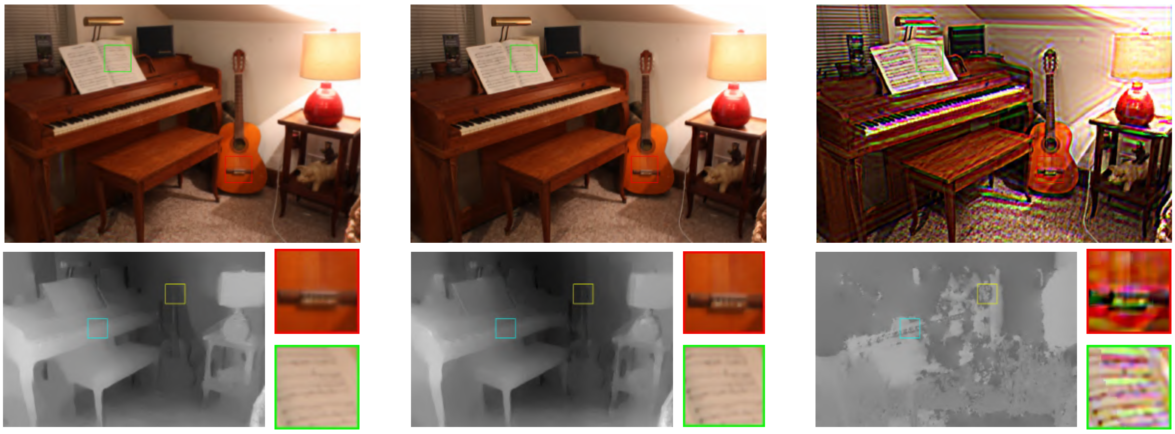
(e) Tao *et al.*[29] (f) Xu *et al.*[35] (g) Xu *et al.*[34]

(e) Mathamkode *et al.*[1] (f) Hu *et al.*[7] (g) Nimisha *et al.*[17]

Figure 5.8: Real experiments: Unlike all the other methods, our method is able to recover fine textual information on the car and also the thin branches in the background. The light field method of [15] and the dual lens method of [34] exhibit severe artifacts in the deblurred image.


(a) Blurry image


(b) Ours (c) Mohan *et al.*[15] (d) Pan *et al.*[18]


(e) Tao *et al.*[29] (f) Xu *et al.*[35] (g) Xu *et al.*[34]

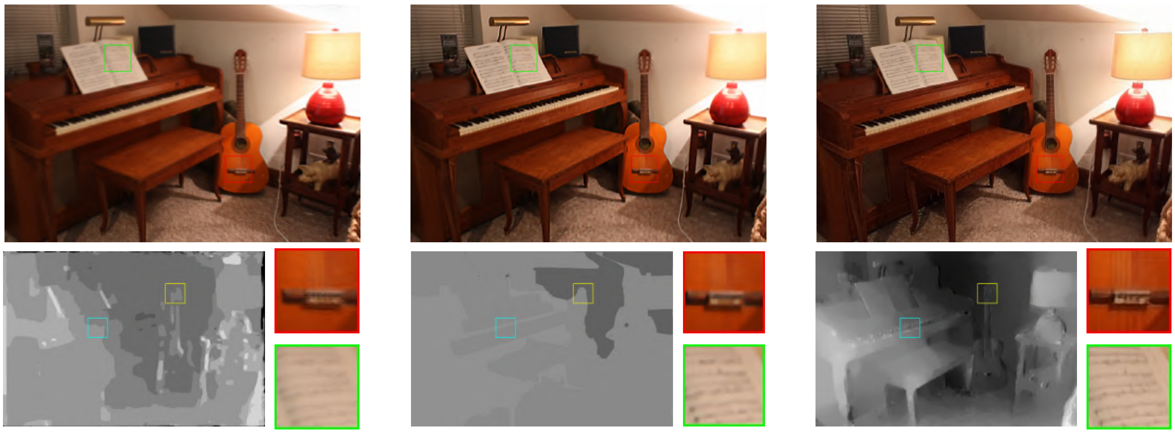
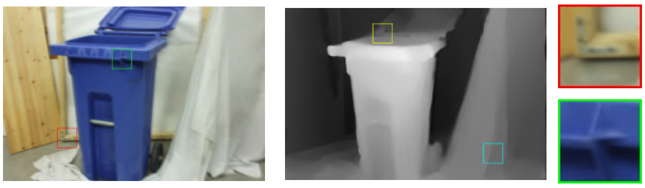

(e) Mathamkode *et al.*[1] (f) Hu *et al.*[7] (g) Nimisha *et al.*[17]

Figure 5.9: Synthetic experiments: Our method recovers the sharp details of the image such as on the guitar and the book without any erroneous depth values.



(a) Blurry image

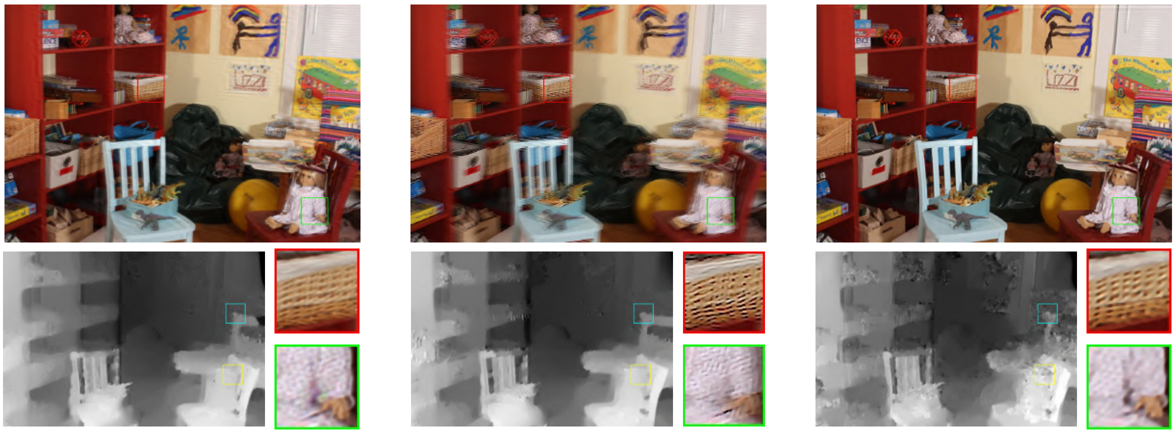
(b) Ours (c) Mohan *et al.*[15] (d) Pan *et al.*[18]

(e) Tao *et al.*[29] (f) Xu *et al.*[35] (g) Xu *et al.*[34]

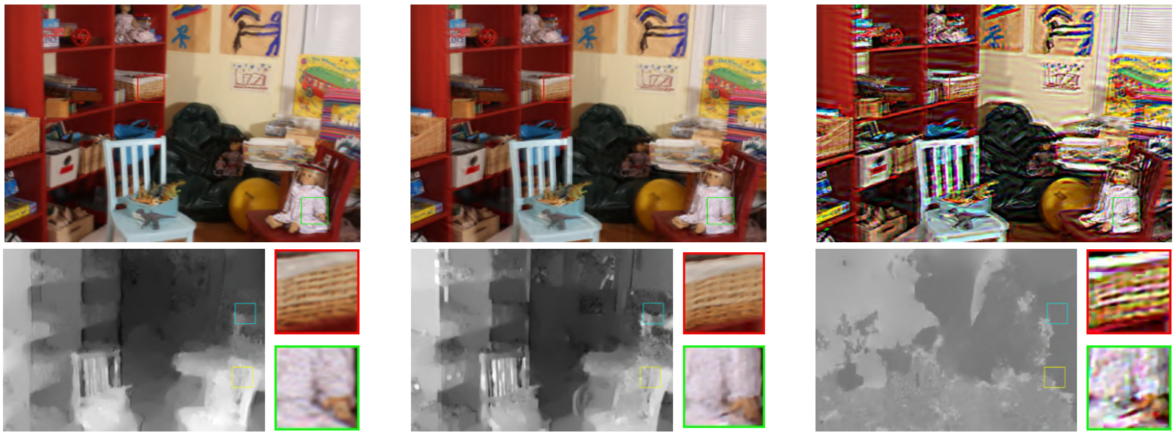
(e) Mathamkode *et al.*[1] (f) Hu *et al.*[7] (g) Nimisha *et al.*[17]

Figure 5.10: Synthetic experiments: Our method retrieves scene features without introducing aritificial structures (unlike the deep learning method [17]), e.g., the features in the highlighted patches in [17] are hallucinated by the deep learning N/W.

(a) Blurry image



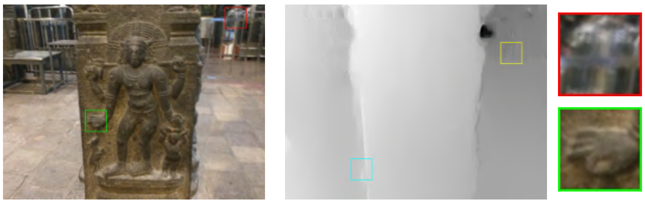
(b) Ours
(c) Mohan *et al.*[15]
(d) Pan *et al.*[18]



(e) Tao *et al.*[29]
(f) Xu *et al.*[35]
(g) Xu *et al.*[34]

(e) Mathamkode *et al.*[1]
(f) Hu *et al.*[7]
(g) Nimisha *et al.*[17]

Figure 5.11: Synthetic experiments: Our method has considerably lesser ringing and sharper features (such as on the closer doll and the farther basket) leading to good depth estimates compared to the other methods.



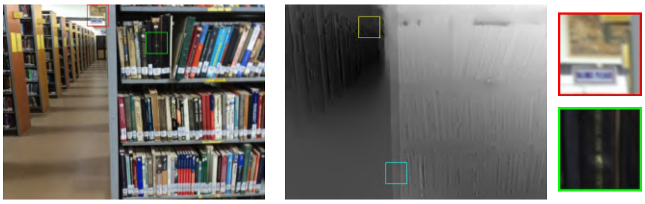
(a) Blurry image

(b) Ours (c) Mohan *et al.*[15] (d) Pan *et al.*[18]

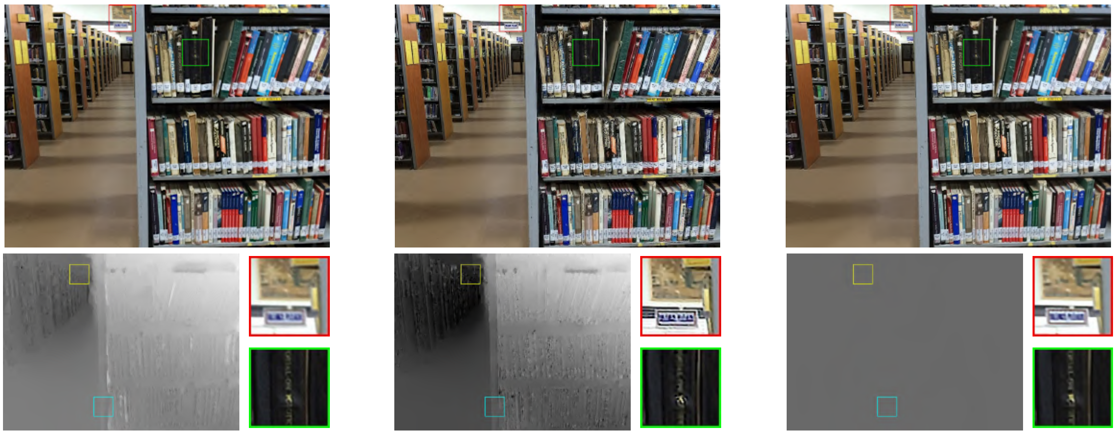
(e) Tao *et al.*[29] (f) Xu *et al.*[35] (g) Xu *et al.*[34]

(e) Mathamkode *et al.*[1] (f) Hu *et al.*[7] (g) Nimisha *et al.*[17]

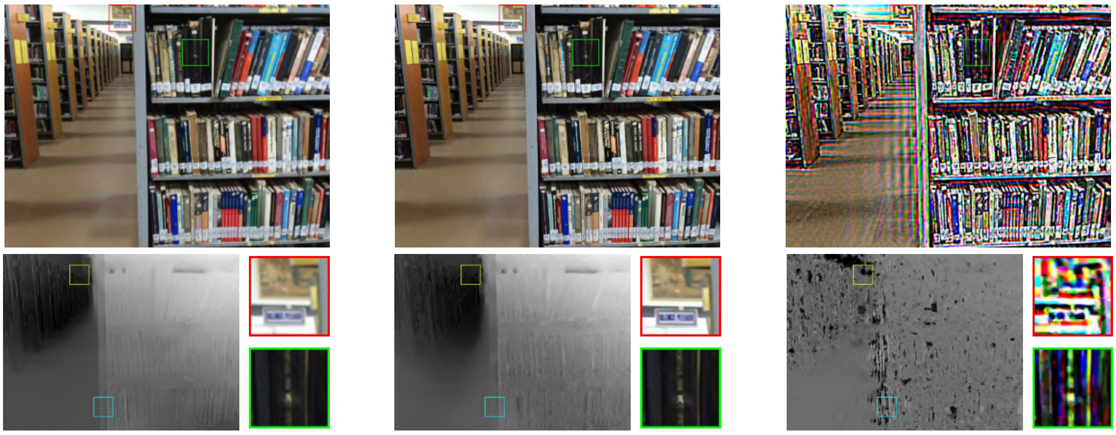
Figure 5.12: Real experiments: Even in a low-light (noisy) scenario, the uniform deblurring performance of our method over different depth levels reveals the noise-robustness of our method.



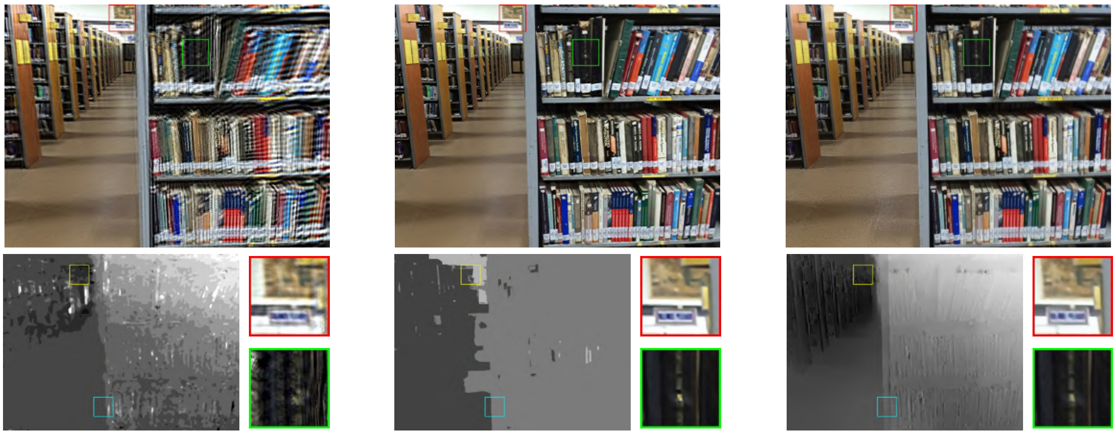
(a) Blurry image



(b) Ours (c) Mohan *et al.*[15] (d) Pan *et al.*[18]



(e) Tao *et al.*[29] (f) Xu *et al.*[35] (g) Xu *et al.*[34]



(e) Mathamkode *et al.*[1] (f) Hu *et al.*[7] (g) Nimisha *et al.*[17]

Figure 5.13: Real experiments (Well-lit scenario): The uniform deblurring over different depth levels yet again proves the effectiveness of our proposed method. Notably, the depth estimate is more accurate and finer in our approach as compared to the competing methods.

5.5 Conclusions and Future Work

In this thesis, we addressed the problem of blind motion deblurring for unconstrained dual-camera set-ups. Our algorithm allows for any arbitrary COR in the blurring process and is incorporated in the optimization pipeline. We revealed an inherent ambiguity in the BMD problem which hampers the scene-consistent depth cues embedded in the image-pair. Towards this end, we introduced a convex and computationally efficient prior. We showed the efficacy of the proposed prior which enforces scene consistent disparities leading to improved deblurring. Comprehensive comparisons with existing state-of-the-art methods amply demonstrate the superiority and need of our method. As an increasing number of modern cameras are employing dual-lens configurations, our theory and method will be very relevant for steering further research in this field.

REFERENCES

[1] **Arun, M.**, **A. N. Rajagopalan**, and **G. Seetharaman**, Multi-shot deblurring for 3d scenes. *In IEEE International Conference on Computer Vision (ICCV) Workshops*. 2015.

[2] **Bätz, M.**, **T. Richter**, **J.-U. Garbas**, **A. Papst**, **J. Seiler**, and **A. Kaup** (2014). High dynamic range video reconstruction from a stereo camera setup. *Signal Processing: Image Communication*, **29**(2), 191–202.

[3] **Cho, S.** and **S. Lee**, Fast motion deblurring. *In ACM Transactions on Graphics (TOG)*, volume 28. ACM, 2009.

[4] **Coleman, T. F.** and **Y. Li** (1996). An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on optimization*, **6**(2), 418–445.

[5] **Gupta, A.**, **N. Joshi**, **C. L. Zitnick**, **M. Cohen**, and **B. Curless**, Single image deblurring using motion density functions. *In European Conference on Computer Vision (ECCV)*. Springer, 2010.

[6] **Hirsch, M.**, **S. Sra**, **B. Schölkopf**, and **S. Harmeling**, Efficient filter flow for space-variant multiframe blind deconvolution. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010.

[7] **Hu, Z.**, **L. Xu**, and **M.-H. Yang**, Joint depth estimation and camera shake removal from single blurry image. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2014.

[8] **Iyer, G.**, **J. Krishna Murthy**, **G. Gupta**, **M. Krishna**, and **L. Paull**, Geometric consistency for self-supervised end-to-end visual odometry. *In IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018.

[9] **Jeon, D. S.**, **S.-H. Baek**, **I. Choi**, and **M. H. Kim**, Enhancing the spatial resolution of stereo images using a parallax prior. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018.

[10] **Krishnan, D.** and **R. Fergus**, Fast image deconvolution using hyper-laplacian priors. *In Advances in neural information processing systems*. 2009.

[11] **Lai, W.-S.**, **J.-B. Huang**, **Z. Hu**, **N. Ahuja**, and **M.-H. Yang**, A comparative study for single image blind deblurring. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016.

[12] **Li, B.**, **C.-W. Lin**, **B. Shi**, **T. Huang**, **W. Gao**, and **C.-C. J. Kuo**, Depth-aware stereo video retargeting. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018.

[13] **Liu, C.** *et al.* (2009). *Beyond pixels: exploring new representations and applications for motion analysis*. Ph.D. thesis, Massachusetts Institute of Technology.

[14] **Lucy, L. B.** (1974). An iterative technique for the rectification of observed distributions. *The astronomical journal*, **79**, 745.

[15] **Mahesh Mohan, M.** and **A. Rajagopalan**, Divide and conquer for full-resolution light field deblurring. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018.

[16] **Mo, J.** and **J. Sattar** (2018). Dsvo: Direct stereo visual odometry. *arXiv preprint arXiv:1810.03963*.

[17] **Nimisha, T.**, **A. K. Singh**, and **A. Rajagopalan**, Blur-invariant deep learning for blind-deblurring. *In IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2017.

[18] **Pan, J.**, **D. Sun**, **H. Pfister**, and **M.-H. Yang**, Blind image deblurring using dark channel prior. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016.

[19] **Pan, L.**, **Y. Dai**, **M. Liu**, and **F. Porikli**, Simultaneous stereo video deblurring and scene flow estimation. *In Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*. IEEE, 2017.

[20] **Park, W.-J.**, **S.-W. Ji**, **S.-J. Kang**, **S.-W. Jung**, and **S.-J. Ko** (2017). Stereo vision-based high dynamic range imaging using differently-exposed image pair. *Sensors*, **17**(7), 1473.

[21] **Scharstein, D.** and **R. Szeliski** (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, **47**(1-3), 7–42.

[22] **Sellent, A.**, **C. Rother**, and **S. Roth**, Stereo video deblurring. *In European Conference on Computer Vision (ECCV)*. Springer, 2016.

[23] **Sheikh, H. R.** and **A. C. Bovik** (2006). Image information and visual quality. *IEEE Transactions on image processing (TIP)*, **15**(2), 430–444.

[24] **Sheikh, H. R.**, **A. C. Bovik**, and **G. De Veciana** (2005). An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on image processing (TIP)*, **14**(12), 2117–2128.

[25] **Shih, K.-T.** and **H. H. Chen** (2018). Generating high-resolution image and depth map using a camera array with mixed focal lengths. *IEEE Transactions on Computational Imaging*.

[26] **Sroubek, F.** and **P. Milanfar** (2012). Robust multichannel blind deconvolution via fast alternating minimization. *IEEE Transactions on Image processing*, **21**(4), 1687–1700.

[27] **Su, S.** and **W. Heidrich**, Rolling shutter motion deblurring. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2015.

[28] **Sun, N.**, **H. Mansour**, and **R. Ward**, Hdr image construction from multi-exposed stereo ldr images. *In IEEE International Conference on Image Processing*. IEEE, 2010.

[29] **Tao, X.**, **H. Gao**, **X. Shen**, **J. Wang**, and **J. Jia**, Scale-recurrent network for deep image deblurring. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018.

[30] **Tibshirani, R.** (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

[31] **Wang, F.**, **T. Li**, and **Y. Li**, Dual deblurring leveraged by image matching. *In IEEE International Conference on Image Processing*. IEEE, 2013.

[32] **Wang, J.**, **T. Xue**, **J. Barron**, and **J. Chen** (2019). Stereoscopic dark flash for low-light photography. *arXiv preprint arXiv:1901.01370*.

[33] **Whyte, O.**, **J. Sivic**, **A. Zisserman**, and **J. Ponce** (2012). Non-uniform deblurring for shaken images. *International journal of computer vision (IJCV)*, **98**(2), 168–186.

[34] **Xu, L.** and **J. Jia**, Depth-aware motion deblurring. *In IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2012.

[35] **Xu, L.**, **S. Zheng**, and **J. Jia**, Unnatural l0 sparse representation for natural image deblurring. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2013.

[36] **Yan, Y.**, **W. Ren**, **Y. Guo**, **R. Wang**, and **X. Cao**, Image deblurring via extreme channels prior. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.

[37] **Yuan, X.**, **Z. Xu**, **H. Wang**, **Y. Liu**, and **L. Fang**, Cascaded image deblurring with combined image prior. *In IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2017.

[38] **Zhang, H.**, **D. Wipf**, and **Y. Zhang**, Multi-image blind deblurring using a coupled adaptive sparse prior. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2013.

[39] **Zhu, X.**, **F. Šroubek**, and **P. Milanfar**, Deconvolving psfs for a better motion deblurring using multiple images. *In European Conference on Computer Vision (ECCV)*. Springer, 2012.