Invertible Motion Blur in Video

Amit Agrawal* Yi Xu Ramesh Raskar
Mitsubishi Electric Research Labs (MERL), Cambridge, MA MIT Media Lab, Cambridge, MA

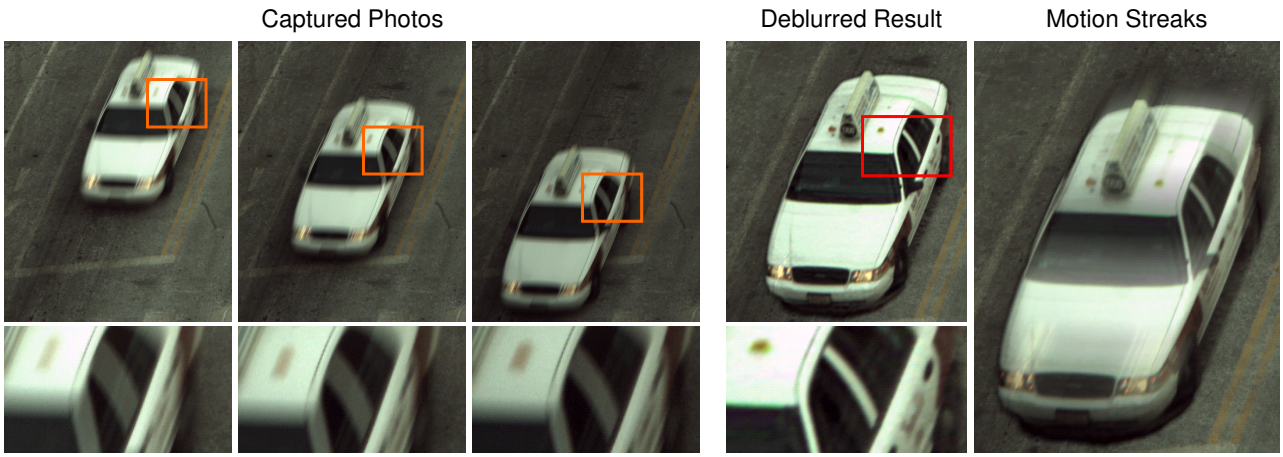

Figure 1: By simply varying the exposure time for video frames, multi-image deblurring can be made invertible. (Left) Varying exposure photos of a moving car. Notice the change in illumination and the blur size in the captured photos. (Right) The foreground object is automatically rectified, segmented, deblurred, and composed onto the background using the varying exposure video. Novel renderings, such as motion streaks, can be generated by linear combination of the deblurred image and the captured photos.

Abstract

We show that motion blur in successive video frames is invertible even if the point-spread function (PSF) due to motion smear in a single photo is non-invertible. Blurred photos exhibit nulls in the frequency transform of the PSF, leading to an ill-posed deconvolution. Hardware solutions to avoid this require specialized devices such as the coded exposure camera or accelerating sensor motion. We employ ordinary video cameras and introduce the notion of null-filling along with joint-invertibility of multiple blur-functions. The key idea is to record the same object with varying PSF's, so that the nulls in the frequency component of one frame can be filled by other frames. The combined frequency transform becomes null-free, making deblurring well-posed. We achieve jointly-invertible blur simply by changing the exposure time of successive frames. We address the problem of automatic deblurring of objects moving with constant velocity by solving its critical components: preservation of all spatial frequencies, segmentation and motion estimation of moving parts, and non-degradation of the static parts of the scene. We demonstrate several challenging cases of object motion blur including textured backgrounds and partial occluders.

CR Categories: I.4.3 [Image Processing and Computer Vision]: Enhancement—Sharpening and Deblurring

Keywords: Computational Photography, Smart Cameras, Motion Deblurring, PSF Estimation, PSF Invertibility

1. Introduction

Motion blur is a common problem in photographing fast moving objects. Consider deblurring a fast moving object in front of a static background. Automatic deblurring involves three critical components: (a) maintaining invertible PSF, (b) estimating the motion of the moving parts, and (c) segmenting the moving objects from the static background. In addition, the fidelity of the static parts of the scene should also be preserved. Previous approaches have attempted to solve one or more of these challenges individually, but none addresses them all. Although, solving them for a single photo is challenging, we observe that the solution for video is more promising. In this paper, we propose a unique approach based on ordinary cameras and show joint-invertibility of blurs in video frames via the concept of frequency domain null-filling.

Maintaining invertible motion PSF is typically not possible in ordinary photos. The box function due to the finite exposure time corresponds to convolution with a low pass filter and hence the frequency transform of the PSF contains zeros (nulls). The corresponding spatial frequencies in the captured photo are lost and hence the deblurring process becomes ill-posed. Previous approaches have used specialized devices to engineer the motion PSF. Raskar et al. [2006] open and close the shutter within the exposure time using a broadband binary code. The code does not have any nulls in the frequency transform; thus making the resulting PSF *invertible*. However, they assume constant background and require manual PSF estimation and object segmentation. Motion invariant photography (MIP) [Levin et al. 2008b] moves the camera with a constant acceleration while capturing the photo. The idea is to make the motion PSF *invariant* to object speed within a certain range. This makes segmentation and PSF estimation unnecessary. However, it needs a priori knowledge of motion direction and the static parts of the scene are blurred due to camera motion.

In this paper, we show that automatic deblurring is possible by using information from multiple frames. The key idea is to record the same object with varying PSF's, so that the nulls in the frequency components of one frame can be filled by other frames. We refer to this as *null-filling*. The combined frequency transform becomes

*e-mails: agrawal@merl.com, xu43@cs.purdue.edu (Currently at Purdue University). Web: http://raskar.info, http://amitkagrawal.com

Figure 2: One can use widely available auto exposure bracketing (AEB) mode on consumer grade digital SLR cameras to achieve PSF null-filling. Top row shows three photos of a moving truck captured using Canon Digital Rebel XT with AEB mode (exposures: 1/50, 1/80, and 1/30 sec). Bottom row shows the manually rectified blurred regions and the deblurred output. Notice that sharp features such as text are recovered in the deblurred image.

null-free, making the deblurring well-posed. We achieve jointly-invertible blur simply by changing the exposure time of successive frames. Our technique does not require camera motion or any coding within the exposure time. It can be implemented on a standard video camera with auto-exposure feature which typically varies the exposure time to compensate for scene brightness.

1.1. Contributions

We propose an automatic deblurring solution by solving the critical components of the deblurring process and exploiting the temporal variations in the blur for invertibility. The contributions of our paper are as follows:

- We propose PSF null-filling that combines several non-invertible PSFs to form a jointly-invertible PSF for motion blur.
- We show that by varying the exposure of each frame within a video, PSF null-filling can be achieved for object motion.
- We demonstrate automatic PSF estimation and object segmentation for motions with constant velocity.

1.2. Benefits and Limitations

Our technique can be used with off-the-shelf machine vision cameras and does not require specialized hardware as in [Raskar et al. 2006]. It can also be implemented on traditional cameras with auto-exposure feature or exposure bracketing in burst mode. Multiple frames simplify critical components of the process to support automatic deblurring. Static parts of the scene are not degraded and a priori knowledge of object motion is not required as opposed to [Levin et al. 2008b].

Our method has several limitations which are shared by typical motion deblurring techniques. We assume linear constant velocity motions that lead to spatially-invariant blur in photos. Non-constant motions (e.g. acceleration) break the spatially-invariant PSF assumption. However, we can still handle spatially varying motions that lead to spatially-invariant blur after image rectification. We cannot handle view dependent effects such as out of plane rotation, specularities and non-diffuse BRDF, along with transparent or translucent objects. We assume the moving object to be in sharp focus and allow out of focus background. We also assume constant illumination. Multiple moving objects can be handled only if they do not occlude each other in the photos. In addition, attached shadows are considered part of the foreground and show noise in the deblurred output due to low signal to noise ratio (SNR).

1.3. Related Work

PSF Manipulation: Specialized capture devices employ two important classes of techniques for engineering the PSF to make it (a) invertible and/or (b) invariant. For defocus PSF, wavefront coding [Dowski and Cathey 1995] uses cubic phase plate in front of the lens to make the PSF invariant to scene depths. This can also be achieved by lateral sensor motion [Nagahara et al. 2008]. However, these approaches result in defocus blur on the scene parts originally in focus. Coded exposure [Raskar et al. 2006] flutters the shutter with a broadband binary code to make the motion PSF invertible. Agrawal and Xu [2009] proposed optimized invertible codes for coded exposure that also helps in PSF estimation. Accelerating camera motion [Levin et al. 2008b] makes the motion PSF invariant to the speed of the object (requiring a priori knowledge of motion direction), at the cost of blurring static parts. Coded aperture [Veeraraghavan et al. 2007] makes the defocus blur invertible for full resolution digital refocusing. Our approach does not modify the camera but indirectly engineers the joint-PSF across frames by carefully choosing the exposure times.

PSF Estimation and Deblurring: Motion deblurring has been an active area of research over last few decades. Blind deconvolution [Jansson 1997] attempts to estimate the PSF from the given image itself. Since deblurring is typically ill-posed, regularization algorithms [Richardson 1972; Lucy 1974] are used to reduce noise. Recent interest in computational photography have spurred significant research in PSF estimation and deblurring algorithms. Fergus et al. [2006] use natural image statistics to estimate the PSF from a single blurred image. Recent papers [Jia 2007; Joshi et al. 2008; Dai and Wu 2008; Yuan et al. 2008; Shan et al. 2008] have shown excellent results on PSF estimation and/or deblurring. While these techniques use natural image statistics and image priors for PSF estimation and deblurring from a single image, the information lost in a single photo due to non-invertible PSF is hallucinated. Our approach uses multiple images for motion blur estimation and invertible deblurring.

Using two images with different PSFs has been studied for PSF estimation [Chen et al. 1996; Sellent et al. 2008]. Rav-Acha and Peleg [2005] and Cho et al. [2007] propose to use two blurs in different directions for deblurring. In contrast, our approach utilize blurs in same direction with different magnitudes. Yuan et al. [2007] use a small exposure image to estimate the motion PSF and use it to deblur a long exposure image to handle camera shake. Chen et al. [2008] perform iterative blur kernels estimation and dual-image deblurring to infer the sharp image from complex motion blurs caused by camera shake. Synthetic shutter speed imaging [Telleen et al. 2007] captures a set of sharp images using small exposure

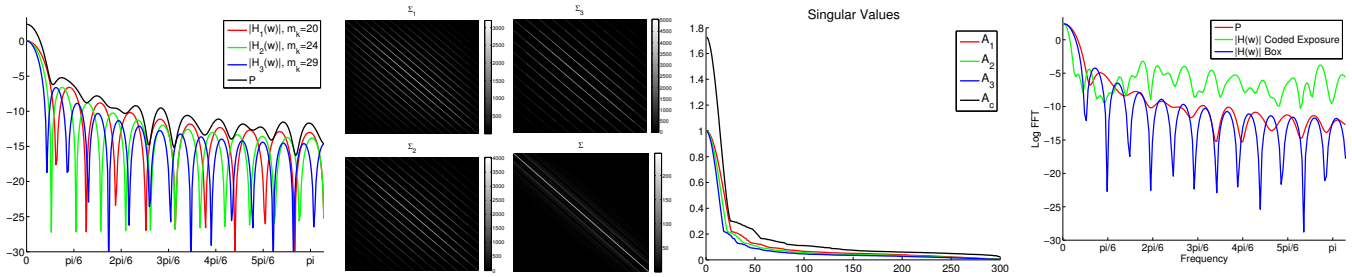

Figure 3: By combining non-invertible PSF's with no common zeros, one can obtain an invertible PSF. (Left) While each $|H_k(w)|$ has zeros, the combined operator $P(w)$ does not have zeros due to null-filling. (Middle left) Each noise covariance matrix (Σ_k) has large diagonal and sub-diagonal entries which result in ringing and noise amplification compared to the covariance matrix Σ for combined deblurring. (Middle right) Singular values of motion blur matrices show that the combined system is better conditioned. (Right) Comparison with coded exposure. Even though coded exposure performs better for high spatial frequencies, it does not benefit PSF estimation and object segmentation.

times and composites the noisy images together by estimating the relative camera motion. These methods do not analyze the problem as PSF null filling. In addition, they are impractical for object motion deblurring observed with static cameras.

Exact Deconvolution: Using multiple filters for exact deconvolution has been investigated by Berenstein and Patrick [1990] in the context of finding the set of convolution filters that results in *finite impulse response* (FIR) deconvolution filters. However, their motivation for obtaining FIR deconvolution filters is for faster processing on DSP chips. Forcing deconvolution filters to be FIR is restrictive and un-necessary. One could use infinite impulse response (IIR) deconvolution filters which provide more flexibility in choosing the convolution filters (corresponding to choosing the exposure time for each frame for motion blur).

Hardware Techniques: Inertial sensors are used in consumer cameras to perform image stabilization. Ben-Ezra and Nayar [2004] and Tai et al. [2008] proposed hybrid imaging systems that estimate PSF using an auxiliary low resolution high frame rate camera to deblur the high resolution primary sensor images. In contrast, our technique estimates the PSF using the high resolution images themselves and does not require an auxiliary camera for PSF estimation.

Video Analysis: Motion PSF has been estimated by combining partial information from successive video frames of a single camera [Schultz and Stevenson 1996; Bascle et al. 1996]. Multiple cameras with overlapped exposure time [Shechtman et al. 2002] and reconfigurable multi-camera array [Wilburn et al. 2005] have also been used to increase the temporal resolution of the captured video streams. Our technique uses a single ordinary camera with different exposure times to make the resulting PSF *invertible*. Background subtraction and segmentation [Piccardi 2004] using a video sequence with same exposure time for all frames is well-known in computer vision community.

Varying Camera Exposure: Several computational photography techniques combine photos with varying parameters. A high dynamic range image (HDRI) can be obtained by combining varying exposure photos [Mann and Picard 1995; Debevec and Malik 1997]. In [Grossberg and Nayar 2003], optimal exposures for HDRI were obtained.

2. Joint Invertibility of Non-Invertible PSFs

We first explain the key idea behind PSF null-filling. Let f denote the sharp image of the object if it was static. Suppose we capture N blurred images i_k of the same object with different PSFs h_k

$$i_k = f * h_k + n_k, \quad k = 1 \dots N, \tag{1}$$

where $*$ denotes the convolution operator and n_k denotes zero mean additive white Gaussian noise with variance σ_k^2. Let T_k be the exposure time for the k^{th} frame.

2.1. Frequency Domain Analysis

We denote the Fourier transform of quantities using capital letters. The Fourier transform of the captured images are given by

$$I_k(w) = F(w)H_k(w) + N_k(w) \quad k = 1 \dots N. \tag{2}$$

Assuming 1D object motion parallel to the sensor plane with constant velocity, each of the PSFs corresponds to a box filter whose width is proportional to the exposure time T_k. Let r_k be the blur size in the k^{th} frame. Then

$$h_k(x) = 1/r_k \quad 0 < x < r_k. \tag{3}$$

Single image deblurring (SID) of any individual image can be written as

$$\widehat{F}(w) = I_k(w)V_k(w) = \frac{I_k(w)}{H_k(w)} = F(w) + \frac{N_k(w)}{H_k(w)}, \tag{4}$$

where $\widehat{F}(w)$ denotes the Fourier transform of the deblurred image and $V_k(w) = \frac{1}{H_k(w)} = \frac{H_k^*(w)}{|H_k(w)|^2}$ denotes the Fourier transform of the corresponding deconvolution filter v_k. Thus, if the frequency transform of the PSF contains zeros, the deconvolution filter becomes unstable and noise is severely amplified. The Fourier transform of the box filter is a sinc function which contains zeros, thereby making deblurring ill-posed.

Now consider utilizing all N frames for multi-image deblurring (MID). The optimal deconvolution filters $V_k(w)$ can be obtained by minimizing the noise power in the resulting deblurred image given by $\sum_{k=1}^N N_k^2(w)|V_k(w)|^2$ at each frequency w. Note that $\sum_{k=1}^N V_k(w)H_k(w) = 1$ to recover the sharp image. Thus, $V_k(w)$ is obtained by

$$\min \sum_{k=1}^N N_k^2(w)|V_k(w)|^2 \quad \texttt{s.t.} \sum_{k=1}^N V_k(w)H_k(w) = 1. \tag{5}$$

Solving, after some manipulations, results in

$$V_k(w) = \frac{H_k^*(w)/N_k^2(w)}{\sum_{k=1}^N |H_k(w)|^2/N_k^2(w)}, \tag{6}$$

$$\widehat{F}(w) = \sum_{k=1}^N I_k(w)V_k(w) = F(w) + \frac{\sum_{k=1}^N H_k^*(w)/N_k(w)}{\sum_{k=1}^N |H_k(w)|^2/N_k^2(w)}.$$

The resulting noise power (NP) in the deblurred image is

$$\texttt{NP} = \sum_{k=1}^{N} N_k^2(w)|V_k(w)|^2 = \frac{1}{\sum_{k=1}^{N}|H_k(w)|^2/N_k^2(w)}. \quad (7)$$

If there are *common* zeros among all the PSFs at a certain frequency w, then $H_k(w) = 0$ for all k at that frequency and $V_k(w)$ becomes unstable. Thus, if the PSFs do not have *common* zeros, the combined deconvolution can be made well-posed, even though each PSF is non-invertible. Intuitively, if there are no common zeros in the Fourier transform, then the information lost in each individual image will be captured by some other image. The nulls in each individual PSF can be *filled* by other PSFs. For motion PSF, this requires that the exposure times should not be integer multiples of each other. In comparison, coded exposure [Raskar et al. 2006] modifies the PSF for a single image such that the corresponding $H(w)$ does not have any zeros.

Let $P^2(w)$ define the denominator in (6). For simplicity, we assume same noise power $N_k(w) = N_0(w)$ for all k. Thus, $P^2(w) = \sum_{k=1}^{N}|H_k(w)|^2$. Figure 3 (left) plots $|H(w)|$ for $N = 3$ PSFs assuming blur sizes r_k of 20, 24 and 29 pixels and object width $W = 300$ pixels along the motion line. Note that each PSF has zeros in the frequency spectrum but the zeros are not aligned. The plot for $P(w)$ shows that while each of the individual frequency transform has zeros, $P(w)$ does not have any zeros.

2.2. Discrete Domain Solution

The convolution equation in discrete domain can be written as $\mathbf{i}_k = A_k\mathbf{f} + \mathbf{n}_k$ for each motion line, where A_k is the circulant motion smear matrix for frame k and $\mathbf{i}_k$, $\mathbf{f}$ and $\mathbf{n}_k$ are the vector of blurred object, sharp object and noise intensities along each motion line respectively [Raskar et al. 2006]. For SID, the estimated deblurred vector $\widehat{\mathbf{f}}$ can be obtained by solving a set of linear equations: $A_{\texttt{SID}}\mathbf{f} = \mathbf{b}$ where

$$A_{\texttt{SID}} = A_k^T A_k \;\; \& \;\; \mathbf{b} = A_k^T\mathbf{i}_k. \quad (8)$$

To reduce ringing artifacts of SID, natural image priors such as Gaussian prior can be used [Levin et al. 2007]:

$$A_{\texttt{SID}} = A_k^T A_k + w\left(C_{g_x}^T C_{g_x} + C_{g_y}^T C_{g_y}\right), \quad (9)$$

where C_{g_x}, C_{g_y} are the convolution matrices corresponding to the derivative filters $g_x = [1\ \text{-}1]$, $g_y = [1\ \text{-}1]^T$ and w is a weighting parameter. The system can be solved in spatial domain using Conjugate Gradient algorithm.

Similarly, for MID, the combined linear system can be written as

$$\begin{bmatrix}\mathbf{i}_1\\ \vdots \\ \mathbf{i}_k\end{bmatrix} = \begin{bmatrix}A_1\\ \vdots \\ A_k\end{bmatrix}\mathbf{f} + \begin{bmatrix}\mathbf{n}_1\\ \vdots \\ \mathbf{n}_k\end{bmatrix} = A_c\mathbf{f} + \mathbf{n}_c. \quad (10)$$

The estimated deblurred vector can be obtained by solving the following system: $A_{\texttt{MID}}\mathbf{f} = \mathbf{b}$ where

$$A_{\texttt{MID}} = \sum_{k=1}^{N} A_k^T A_k + w\left(C_{g_x}^T C_{g_x} + C_{g_y}^T C_{g_y}\right) \;\; \& \;\; \mathbf{b} = \sum_{k=1}^{N} A_k^T\mathbf{i}_k.$$

We use the Conjugate Gradient algorithm to solve the combined system as well.

2.3. Noise Analysis

For SID, using (9), the covariance matrix of the noise in the estimate $\widehat{\mathbf{f}} - \mathbf{f}$ is equal to

$$\Sigma_{\texttt{SID}} = (A_{\texttt{SID}})^{-1}A_k^T\sigma_k^2 A_k(A_{\texttt{SID}})^{-T}. \quad (11)$$

If Gaussian prior is not used, (11) simplifies to

$$\Sigma_{\texttt{SID}} = (A_k^T A_k)^{-1}A_k^T\sigma_k^2 A_k(A_k^T A_k)^{-T} = \sigma_k^2(A_k^T A_k)^{-1}. \quad (12)$$

For MID, the covariance matrix is

$$\Sigma_{\texttt{MID}} = (A_{\texttt{MID}})^{-1}(\sum_{k=1}^{N} A_k^T A_k\sigma_k^2)(A_{\texttt{MID}})^{-T}. \quad (13)$$

The mean square error (MSE) equals $\texttt{Trace}(\Sigma_{\texttt{MID}})/W$, where W is the size of the motion line. The SNR of the deblurred object is given by

$$\texttt{SNR}_{\texttt{MID}} = \frac{\sum_{k=1}^{N} i_k/N}{\sqrt{\texttt{Trace}(\Sigma_{\texttt{MID}})/W}}. \quad (14)$$

By simply averaging the captured N photos, the capture SNR is

$$\texttt{SNR}_{\texttt{Capture}} = \frac{\sum_{k=1}^{N} i_k}{\sqrt{\sum_{k=1}^{N}\sigma_k^2}}. \quad (15)$$

Thus, due to deconvolution, SNR decreases by a factor of Nf, where

$$f = \sqrt{\frac{\texttt{Trace}(\Sigma_{\texttt{MID}})/W}{\sum_{k=1}^{N}\sigma_k^2}}. \quad (16)$$

For SID, (14) simplifies to

$$\texttt{SNR}_{\texttt{SID}} = \frac{i_k}{\sqrt{\sigma_k^2\texttt{Trace}[(A_k^T A_k)^{-1}]/W}} \quad (17)$$

If all frames have the same exposure, then all i_k's and A_k's are the same for all k, and (14) simplifies to

$$\texttt{SNR}_{\texttt{MID Equal Exposure}} = \frac{i_k\sqrt{N}}{\sqrt{\sigma_k^2\texttt{Trace}[(A_k^T A_k)^{-1}]/W}} \quad (18)$$

Thus, using frames with the same exposure (traditional video) is equivalent to *averaging* the deblurred image obtained from individual frames separately, since the SNR increases by $\sqrt{N}$. In contrast, varying exposure gives significantly larger SNR improvement and noise reduction as shown in Figure 3 (middle).

2.4. Exposure Sequence Optimization

The optimal exposure sequence can be obtained by minimizing the increase in noise power given by (7) or by minimizing f. This also implies that $P(w)$ should be as large as possible for each frequency w. Since the variance of the photon generated electrons linearly increases with the measured signal (and hence exposure time), σ_k^2 is given by βT_k, where β is a camera dependent constant and T_k is the exposure time. Thus, f depends on the choice of exposure values. We use a random search method similar to [Raskar et al. 2006]. For many random sets of exposures, we compute the decrease in SNR and record the best sequence. Note that for coded exposure, the search space is of the order of 2^n, where n is the code length (e.g. 52). In contrast, the number of unknowns for MID is equal to the number of different exposures. Typically, $3-4$ different exposures yield a good balance between computational cost and deblurring performance; and thus the search space is small.

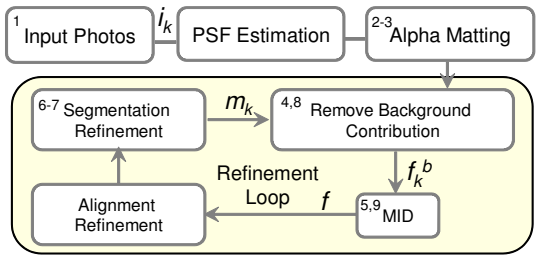

Figure 4: Overview of automatic PSF estimation, segmentation and deblurring algorithm. The numbers on the top left corner of each block correspond to the numbers in Figure 5.

3. Deblurring using Varying Exposure Video

PSF invertibility is one of the criteria for a successful deblurring method. In real situations of a moving object in front of a non-smooth background, PSF estimation and object segmentation are equally important. Now we describe the steps involved in motion deblurring using multiple frames captured with different exposures.

Let $m_k(x, y)$ be the binary mask for the object in the k^{th} frame. If $b(x, y)$ is the background image without the object, the captured motion blurred images i_k are given by [Raskar et al. 2006]

$$i_k = (f \cdot m_k) * h_k + (1 - m_k * h_k) \cdot b. \tag{19}$$

For explanation, let us assume that the object is moving perpendicular to the optical axis horizontally and PSFs are spatially-invariant for each photo. We first estimate the background b using simple median filtering along the temporal direction. We assume that the typical background subtraction [Piccardi 2004] will provide the background b in a practical scenario. The rest of the algorithm proceed according to the flowchart shown in Figure 4.

3.1. PSF Estimation

For spatially-invariant linear blur, PSF can be found by multiplying the image-space object velocity v (pixel/ms) with the exposure time (ms). Object velocity is the ratio of the inter-frame motion vector to the inter-frame time lapse. For spatially-invariant blur, the inter-frame motion vector can be computed by matching corresponding image patches in different frames. However, different exposure times lead to different sizes of blur. To facilitate matching and PSF estimation, we *repeat* the exposure sequence for capturing the video. Thus, every N^{th} frame has the same exposure, where N is the number of different exposures used ($\approx 3-4$). Motion vectors can be easily computed by matching the frames captured using the same exposure. Averaging the motion vectors for different exposures gives the estimate of object velocity v and the PSFs.

3.2. Foreground Initialization

A blurred photo has contributions from both the blurred foreground and the static background. The image blurring equation (19) can be written in terms of alpha matting equation as [Jia 2007; Dai and Wu 2008]

$$i_k = \alpha_k g + (1 - \alpha_k) \cdot b, \tag{20}$$

where $g = \frac{(f \cdot m_k) * h_k}{m_k * h_k}$ and $\alpha_k = m_k * h_k$. Ideally, deblurring of alpha maps should recover the binary segmentation mask m_k. Note that while matting is typically used for non-opaque static objects, we assume that the foreground motion blurred object is opaque and in sharp focus. Thus, the alpha map depends only on the motion blur and the matting foreground actually corresponds to the *blurred* object (not the sharp object).

To compute initial alpha maps, we first generate a crude trimap for each frame by thresholding the difference between the input photo

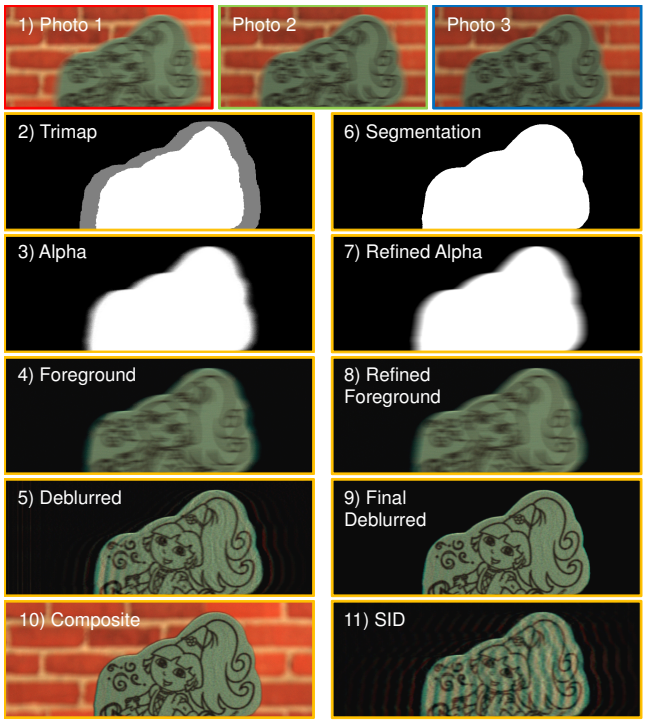

Figure 5: Deblurring procedure. Input frames (1) are automatically processed. Intermediate results (2-10) show how the alpha map, deblurring and segmentation improve after the first iteration. (11) shows the SID result using only the first photo and the corresponding extracted foreground.

i_k and the background image b (Figure 5 (2)). The trimap is 1 for the interior of the moving object, 0 for background and unknown for the blurred region. Morphological operations, such as erosion and hole-filling, are applied to reduce noise. Using the trimap, alpha matting [Levin et al. 2008a] is performed on each frame independently (Figure 5 (3)). The blurred foreground f_k^b is obtained by subtracting the background contribution from each input photo as

$$f_k^b = i_k - (1 - \alpha_k) \cdot b, \tag{21}$$

as shown in Figure 5 (4). Since a general matting algorithm does not utilize the property of motion blur, the initial alpha maps are crude.

3.3. Multi-Image Deblurring (MID)

Because the object moves in successive frames, the blurred foregrounds (f_k^b) need to be aligned before deconvolution. For linear constant velocity motion, the alignment corresponds to a shift in the image plane. Since we already computed the object velocity v, the shift between i^{th} and $(i+1)^{th}$ frame is simply given by vT_f, where T_f is the inter-frame time lapse and can be computed given the camera frame rate. The input photos are also scaled to the same intensity level using the ratios between the known exposure values assuming linear camera response. After aligning the blurred foregrounds, deblurring is performed by solving the linear system (11) using Conjugate Gradient algorithm. Due to inaccurate alpha maps, this initial estimate of the deblurred image is noisy and contains erroneous background contribution as shown in Figure 5 (5).

3.4. Alignment Refinement

The alignment of the foreground layers in the previous step can be slightly off due to the violation of the constant motion assumption in real world scenarios. The misalignment causes errors in

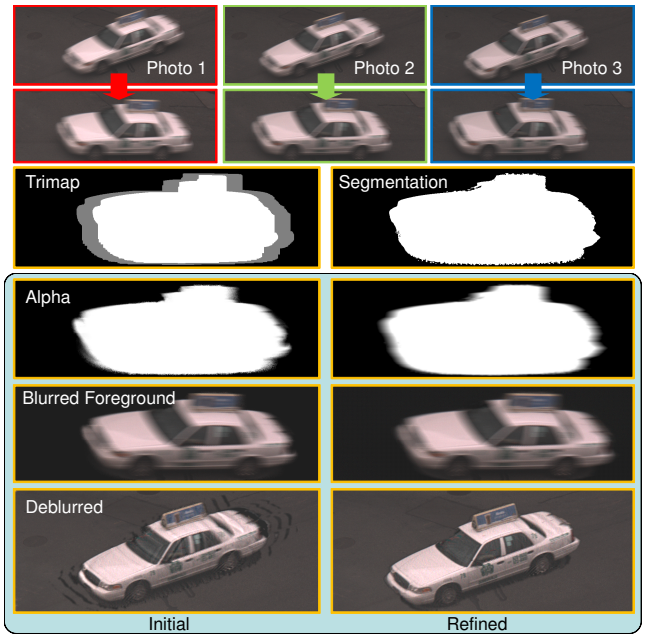

Figure 6: Outdoor car sequence. Input blurred photos are automatically rectified using the estimated motion direction. After refinement, sharp features on the car are recovered and deblurring artifacts are reduced.

deconvolution. We refine the alignment using the deblurred foreground f obtained in the previous step. Specifically, we re-blur the sharp foreground f using the estimated blur kernels h_k and find the shift between the synthetically blurred foreground and blurred foreground f_k^b computed using matting. In this way, all the blurred foreground f_k^b's can be correlated through deblurring the sharp foreground f and can be better aligned.

3.5. Segmentation Refinement

Given the initial alpha maps, a binary segmentation of the sharp object can be computed. Since the linear system is well-posed, deblurring the alpha maps using the MID algorithm and then thresholding the result gives us an approximate binary segmentation m_k of the object. The segmentation can then be refined efficiently. By using a conservative threshold (0.65 in all of our examples), we obtain an initial segmentation mask m^0 smaller than the object size. For each motion line, we grow the mask pixel by pixel and find the best estimate that minimizes the difference between the alpha values computed using matting and those computed by blurring the sharp mask. Typically, we search within 10 pixels on each side and this step finishes within 30 seconds in our experiments. Figure 5 (6) shows the mask after refinement.

Blurring the refined segmentation gives us refined alpha maps (Figure 5 (70), which are used to compute new blurred foregrounds f_k^b (Figure 5 (8)). The f_k^b's are subsequently deblurred (Figure 5 (9)) using MID. The quality of segmentation and deblurring improves significantly after the first iteration. The refinement is iterated $1-2$ times for the final result.

4. Implementation and Results

We use PointGrey Flea2 camera with maximum frame rate of 15 fps. Using the SDK provided with the camera, the exposure time for each frame could be changed easily. For indoor scenes, we placed objects on a variable speed toy train to capture the datasets. In order to find the optimal exposures, we bound each exposure within $T_{\min}$ and $T_{\max}$ to avoid saturation and unusable photos. For numerical

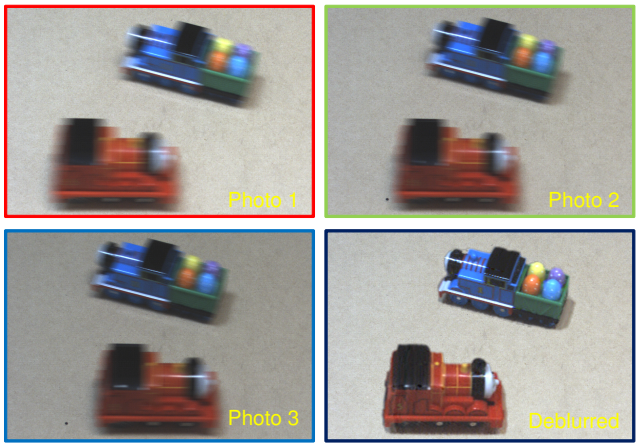

Figure 7: Multiple objects. Using k-means clustering ($k = 2$ in this case), complex-shaped objects can be segmented separately while creating trimaps. Each object is subsequently deblurred and pasted onto the background independently. Note that the view-dependent effects (highlights) cause artifacts in the final result.

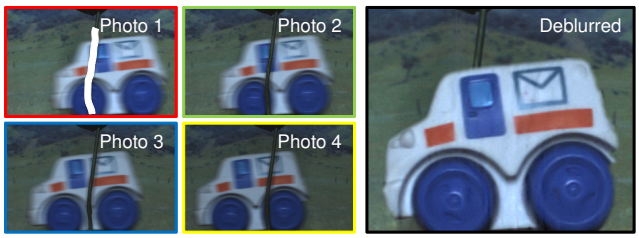

Figure 8: Partial occlusion. Our method allows deblurring objects partially occluded in the captured photos. The user specifies the occluder in one of the input photos (top left) using scribbles. These pixels are ignored during deblurring and the object is fully recovered.

stability, we enforce that the difference between any two exposure values is greater than a threshold $T_\delta = 5$ ms. For $N = 3$, $T_{\min} = 30$ ms and $T_{\max} = 50$ ms, our optimized exposures were 30, 35, and 42 ms. We capture at least $2N$ images to allow PSF estimation.

Figure 6 shows captured photos of a fast moving taxi outdoors. Since the taxi is far away, it is assumed to be moving parallel to the image plane. We automatically estimate the motion direction and magnitude using our technique. Figure 6 also shows the estimated trimap, initial alpha, initial deblurring, refined segmentation and the final deblurred image of the taxi. Figure 1 shows the deblurring result of a similar car. Shadows pose a common problem in segmentation and background subtraction for computer vision applications. Since attached shadows move with object, they are considered to be part of the foreground and show increased noise due to low SNR. Multiple objects can also be handled as long as they do not occlude each other (Figure 7).

Auto Exposure Bracketing (AEB): PSF null-filling can also be achieved using the AEB mode commonly available on digital SLR cameras. Figure 2 shows such an example. Since the AEB mode only allows three frames and the camera does not provide accurate time-stamp information, we manually segmented the blurred region and measured the motion PSF for this example. For AEB mode, the change in exposure is limited to multiples of $\frac{1}{3}$ f-stops. Even though this choice may not be optimal, visual deblurring quality is good as shown by the recovered text in Figure 2. In our experience, the choice of exposure values is not highly critical for deblurring as long as they are not close to integer multiples of each other.

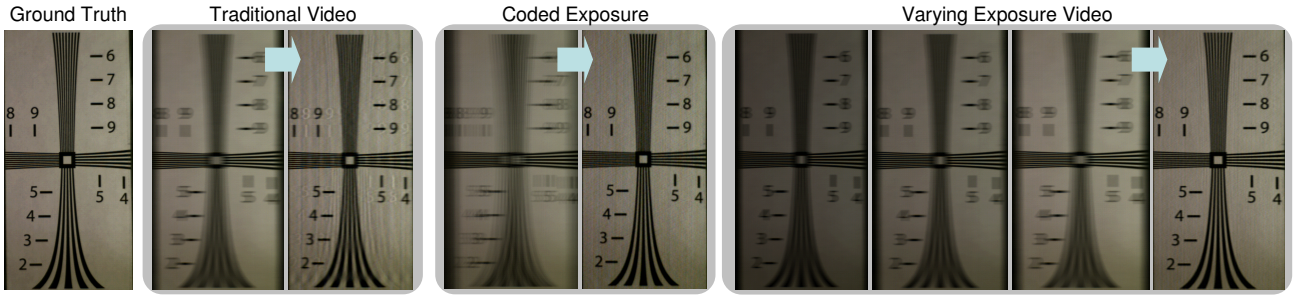

Figure 9: Comparison of the deblurring performance of varying exposure video, traditional video (same number of frames) and coded exposure (for light level corresponding to T_m). The blurred photos are generated by adding high speed camera photos of a moving resolution chart. Note that the visual deblurring quality using our technique is similar to coded exposure and significantly better than traditional video.

Partial Occlusion: Deblurring in presence of partial occlusion using a single image has been demonstrated in [Raskar et al. 2006], assuming the size of the occluder is smaller than the blur size. Interestingly, it becomes much easier for multiple image deblurring, since the occluder now occludes *different* parts of the object in multiple images. In our system, the user simply marks the occluder using a scribble in one of the blurred photos as shown in Figure 8. While deblurring, the pixels on the occluder are ignored in the least square system. The foreground regions occluded in every frame are accurately recovered in the deblurred image.

Comparisons: Let T_m be the maximum exposure used in our technique. We compare our technique with coded exposure and traditional video having same number of frames at light level corresponding to T_m. Since coded exposure loses half the light, it needs twice the exposure ($2T_m$) for the same light level. Figure 9 compares the deconvolution result on a moving resolution chart using the three techniques. Note that deblurring using varying exposure recovers significantly high spatial frequencies on the resolution chart compared to the traditional video. The visual quality of the deblurring is similar to coded exposure output. Coded exposure allows more degrees of freedom in manipulating the PSF (equal to the code length), while our approach combines box filters of different widths and thus has less degrees of freedom. Figure 3 (right) shows that coded exposure performs better than varying exposure for high spatial frequencies. In general, varying exposure gives $4-5$ dB more deconvolution noise than coded exposure.

5. Discussions

While our technique combines box filters to achieve PSF null-filling, several variations are possible. For motion blur, coded exposure could be combined with our method at the expense of hardware modification to improve the SNR of the deblurred image. In several imaging applications, the captured photo loses high spatial frequency content during the capture process. Without any hardware modifications, it might be possible to change capture parameters to preserve different frequencies in different photos and combine them for invertible deconvolution. We hope video cameras can include this feature by default. For an uninterested consumer, simple normalization of the photos will show smooth changes in intensity per frame. But motion blurred photos can be processed afterwards if desired. Since coded exposure can be used to achieve super resolution [Agrawal and Raskar 2007], it may be possible to resolve objects at higher resolution from multiple photos using varying exposure video.

Conclusions: We showed that motion blur in video can be made invertible by combining non-invertible PSFs that do not have common zeros. PSF null-filling can be easily achieved on machine vision cameras as well as off-the-shelf digital SLR's using exposure bracketing, without requiring additional hardware or camera motion. For a complete deblurring solution, segmentation and PSF estimation are as important as PSF invertibility. We demonstrated that repeated sequence of varying exposure frames can be used for automatic PSF estimation and segmentation for challenging object motion blur scenarios.

Acknowledgements

We thank the anonymous reviewers and several members of the Mitsubishi Electric Research Labs for their suggestions. We also thank Daniel Aliaga, Jay Thornton, Keisuke Kojima, Joseph Katz, Ashok Veeraraghavan and Brandon Taylor, along with Haruhisa Okuda and Kazuhiko Sumi, Mitsubishi Electric, Japan for their help and support.

References

AGRAWAL, A., AND RASKAR, R. 2007. Resolving Objects at Higher Resolution from a Single Motion-blurred Image. In *Proc. Conf. Comp. Vision and Pattern Recognition*, 1–8.

AGRAWAL, A., AND XU, Y. 2009. Coded Exposure Deblurring: Optimized Codes for PSF Estimation and Invertibility. In *Proc. Conf. Comp. Vision and Pattern Recognition*, 1–8.

BASCLE, B., BLAKE, A., AND ZISSERMAN, A. 1996. Motion Deblurring and Super-resolution from an Image Sequence. In *Proc. European Conf. Computer Vision*, vol. 2, 573–582.

BEN-EZRA, M., AND NAYAR, S. 2004. Motion-based Motion Deblurring. *IEEE Trans. Pattern Anal. Machine Intell. 26*, 6 (Jun), 689–698.

BERENSTEIN, C., AND PATRICK, E. 1990. Exact Deconvolution for Multiple Convolution Operators-an Overview, plus Performance Characterizations for Imaging Sensors. *Proc. of the IEEE 78* (Apr.), 723–734.

CHEN, W.-G., NANDHAKUMAR, N., AND MARTIN, W. N. 1996. Image Motion Estimation from Motion Smear-A New Computational Model. *IEEE Trans. Pattern Anal. Mach. Intell. 18*, 4 (Apr.), 412–425.

CHEN, J., YUAN, L., TANG, C.-K., AND QUAN, L. 2008. Robust Dual Motion Deblurring. In *Proc. Conf. Comp. Vision and Pattern Recognition*, 1–8.

CHO, S., MATSUSHITA, Y., AND LEE, S. 2007. Removing Non-Uniform Motion Blur from Images. In *Proc. Int'l Conf. Computer Vision*, 1–8.

DAI, S., AND WU, Y. 2008. Motion from Blur. In *Proc. Conf. Comp. Vision and Pattern Recognition*, 1–8.

DEBEVEC, P. E., AND MALIK, J. 1997. Recovering High Dynamic Range Radiance Maps from Photographs. In *Proc. SIGGRAPH 97*, 369–378.

DOWSKI, E. R., AND CATHEY, W. 1995. Extended Depth of Field through Wavefront Coding. *Appl. Optics 34*, 11 (Apr.), 1859–1866.

FERGUS, R., SINGH, B., HERTZMANN, A., ROWEIS, S. T., AND FREEMAN, W. T. 2006. Removing Camera Shake from a Single Photograph. *ACM Trans. Graph. 25*, 3 (jul), 787–794.

GROSSBERG, M., AND NAYAR, S. 2003. High Dynamic Range from Multiple Images: Which Exposures to Combine? In *ICCV Workshop on Color and Photometric Methods in Computer Vision (CPMCV)*.

JANSSON, P. 1997. *Deconvolution of Image and Spectra*, 2nd ed. Academic Press.

JIA, J. 2007. Single Image Motion Deblurring using Transparency. In *Proc. Conf. Comp. Vision and Pattern Recognition*, 1–8.

JOSHI, N., SZELISKI, R., AND KRIEGMAN, D. 2008. PSF Estimation using Sharp Edge Prediction. In *Proc. Conf. Comp. Vision and Pattern Recognition*, 1–8.

LEVIN, A., FERGUS, R., DURAND, F., AND FREEMAN, W. T. 2007. Image and Depth from a Conventional Camera with a Coded Aperture. *ACM Trans. Graph. 26*, 3 (Jul.), 70.

LEVIN, A., LISCHINSKI, D., AND WEISS, Y. 2008. A Closed-Form Solution to Natural Image Matting. *IEEE Trans. Pattern Anal. Mach. Intell. 30*, 2, 228–242.

LEVIN, A., SAND, P., CHO, T. S., DURAND, F., AND FREEMAN, W. T. 2008. Motion-Invariant Photography. *ACM Trans. Graph. 27*, 3 (Aug.), 71.

LUCY, L. 1974. An iterative technique for the rectification of observed distributions. *J. Astronomy 79*, 745–754.

MANN, S., AND PICARD, R. W. 1995. Being Undigital with Digital Cameras: Extending Dynamic Range by Combining Differently Exposed Pictures. In *Proc. of IS&T 48th Annual Conference*, 422–428.

NAGAHARA, H., KUTHIRUMMAL, S., ZHOU, C., AND NAYAR, S. 2008. Flexible Depth of Field Photography. In *Proc. European Conf. Computer Vision*, 60 – 73.

PICCARDI, M. 2004. Background Subtraction Techniques: a Review. In *Proc. IEEE SMC Intl. Conf. Systems, Man and Cybernetics*.

RASKAR, R., AGRAWAL, A., AND TUMBLIN, J. 2006. Coded Rxposure Photography: Motion Deblurring using Fluttered Shutter. *ACM Trans. Graph. 25*, 3 (Jul.), 795–804.

RAV-ACHA, A., AND PELEG, S. 2005. Two Motion-blurred Images are Better than One. *Pattern Recognition Letters 26*, 3, 311 – 317.

RICHARDSON, W. 1972. Bayesian-based iterative method of image restoration. *J. Opt. Soc. of America 62*, 1, 55–59.

SCHULTZ, R. R., AND STEVENSON, R. L. 1996. Extraction of High-Resolution Frames from Video Sequences. *IEEE Trans. Image Processing 5* (jun), 996–1011.

SELLENT, A., EISEMANN, M., AND MAGNOR, M. 2008. Calculating Motion Fields from Images with Two Different Exposure Times. Tech. rep., Computer Graphics Lab, Technical University of Braunschweig, 5.

SHAN, Q., JIA, J., AND AGARWALA, A. 2008. High-Quality Motion Deblurring from a Single Image. *ACM Trans. Graph. 27*, 3 (Aug.), 73.

SHECHTMAN, E., CASPI, Y., AND IRANI, M. 2002. Increasing Space-Time Resolution in Video. In *Proc. European Conf. Computer Vision*, 753–768.

TAI., W. Y., HAO, D., BROWN, M. S., AND LIN, S. 2008. Image/Video Deblurring using a Hybrid Camera. In *Proc. Conf. Comp. Vision and Pattern Recognition*, 1–8.

TELLEEN, J., SULLIVAN, A., YEE, J., GUNAWARDANE, P., WANG, O., COLLINS, I., AND DAVIS, J. 2007. Synthetic Shutter Speed Imaging. In *Proc. Eurographics*, 591–598.

VEERARAGHAVAN, A., RASKAR, R., AGRAWAL, A., MOHAN, A., AND TUMBLIN, J. 2007. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph. 26*, 3, 69.

WILBURN, B., JOSHI, N., VAISH, V., TALVALA, E.-V., ANTUNEZ, E., BARTH, A., ADAMS, A., HOROWITZ, M., AND LEVOY, M. 2005. High Performance Imaging using Large Camera Arrays. *ACM Trans. Graph. 24*, 3 (Jul.), 765–776.

YUAN, L., SUN, J., QUAN, L., AND SHUM, H.-Y. 2007. Image Deblurring with Blurred/Noisy Image Pairs. *ACM Trans. Graph. 26*, 3 (Jul.), 1.

YUAN, L., SUN, J., QUAN, L., AND SHUM, H.-Y. 2008. Progressive Inter-Scale and Intra-Scale Non-Blind Image Deconvolution. *ACM Trans. Graph. 27*, 3 (Aug.), 74.