

Correction of Spatially Varying Image and Video Motion Blur Using a Hybrid Camera

Yu-Wing Tai, *Member, IEEE*, Hao Du, *Student Member, IEEE*,
Michael S. Brown, *Member, IEEE*, and Stephen Lin, *Member, IEEE*

Abstract—We describe a novel approach to reduce spatially varying motion blur in video and images using a hybrid camera system. A hybrid camera is a standard video camera that is coupled with an auxiliary low-resolution camera sharing the same optical path but capturing at a significantly higher frame rate. The auxiliary video is temporally sharper but at a lower resolution, while the lower frame-rate video has higher spatial resolution but is susceptible to motion blur. Our deblurring approach uses the data from these two video streams to reduce spatially varying motion blur in the high-resolution camera with a technique that combines both deconvolution and super-resolution. Our algorithm also incorporates a refinement of the spatially varying blur kernels to further improve results. Our approach can reduce motion blur from the high-resolution video as well as estimate new high-resolution frames at a higher frame rate. Experimental results on a variety of inputs demonstrate notable improvement over current state-of-the-art methods in image/video deblurring.

Index Terms—Motion deblurring, spatially varying motion blur, hybrid camera.

1 INTRODUCTION

THIS paper introduces a novel approach to reduce spatially varying motion blur in video footage. Our approach uses a hybrid camera framework first proposed by Ben-Ezra and Nayar [6], [7]. A hybrid camera system simultaneously captures a high-resolution video together with a low-resolution video that has denser temporal sampling. The hybrid camera system is designed such that the two videos are synchronized and share the same optical path. Using the information in these two videos, our method has two aims: 1) to deblur the frames in the high-resolution video and 2) to estimate new high-resolution video frames at a higher temporal sampling.

While high-resolution, high-frame-rate digital cameras are becoming increasingly more affordable (e.g., $1,960 \times 1,280$ at 60 fps are now available at consumer prices), the hybrid camera design remains promising. Even at 60 fps, high-speed photography/videography is susceptible to motion blur artifacts. In addition, as the frame rate of high-resolution cameras increases, low-resolution camera frame-rate speeds increase accordingly with cameras available now

with over 1,000 fps at lower resolution. Thus, our approach has application to ever increasing temporal imaging. In addition, the use of hybrid cameras and hybrid camera-like designs have been demonstrated to offer other advantages over single-view cameras including object segmentation and matting [7], [35], [36], depth estimation [31], and high dynamic range imaging [1]. The ability to perform object segmentation is key in deblurring moving objects, as demonstrated by the authors of [7] and our own work in Section 5.

The previous work in [6], [7] using a hybrid camera system focused on correcting motion blur in a single image under the assumption of globally invariant motion blur. In this paper, we address the broader problem of correcting spatially varying motion blur and aim to deblur temporal sequences. In addition, our work achieves improved deblurring performance by more comprehensively exploiting the available information acquired in the hybrid camera system, including optical flow, back-projection constraints between low-resolution and high-resolution images, and temporal coherence along image sequences. In addition, our approach can be used to increase the frame rate of the high-resolution camera by estimating intermediate frames.

The central idea in our formulation is to combine the benefits of both deconvolution and super-resolution. Deconvolution of motion-blurred, high-resolution images yields high-frequency details, but with ringing artifacts due to the lack of low-frequency components. In contrast, super-resolution-based reconstruction from low-resolution images recovers artifact-free low-frequency results that lack high-frequency detail. We show that the deblurring information from deconvolution and super-resolution is complementary to each other and can be used together to improve deblurring performance. In video deblurring applications, our method further capitalizes on additional deconvolution constraints that can be derived from consecutive video frames. We

- Y.-W. Tai is with the Korea Advanced Institute of Science and Technology (KAIST), Korea. E-mail: yuwing@gmail.com.
- H. Du is with the Department of Computer Science and Engineering, University of Washington, Box 352350, Seattle, WA 98195-2350. E-mail: duhao@cs.washington.edu.
- M.S. Brown is with the School of Computing, National University of Singapore, Computing 1, 13 Computing Drive, Singapore 117417, Republic of Singapore. E-mail: brown@comp.nus.edu.sg.
- S. Lin is with Microsoft Research Asia, Beijing Sigma Center, No. 49, Zhichun Road, Beijing 100190, P.R. China. E-mail: stevelin@microsoft.com.

Manuscript received 18 Apr. 2008; revised 4 Nov. 2008; accepted 30 Mar. 2009; published online 29 Apr. 2009.

Recommended for acceptance by K. Kutulakos.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2008-04-0224.

Digital Object Identifier no. 10.1109/TPAMI.2009.97.

demonstrate that this approach produces excellent results in reducing spatially varying motion blur. In addition, the availability of the low-resolution imagery and subsequently derived motion vectors further allows us to estimate new temporal frames in the high-resolution video, which we also demonstrate.

A shorter version of this work appeared in [47]. This journal version extends our conference work with greater discussion of the deblurring algorithm, further technical details of our implementation, and additional experimentations. In addition, a method to estimate new temporal frames in the high-resolution video is presented in Section 6, together with supporting experiments in Section 7.

The processing pipeline of our approach is shown in Fig. 2, which also relates process components to their corresponding section in the paper. The remainder of the paper is organized as follows: Section 2 discusses related work, Section 3 describes the hybrid camera setup and the constraints on deblurring available in this system, Section 4 describes our overall deconvolution formulation expressed in a maximum a posteriori (MAP) framework, Section 5 discusses how to extend our framework to handle moving objects, Section 6 describes how to perform temporal upsampling with our framework, Section 7 provides results and comparisons with other current work, followed by a discussion and summary in Section 8.

2 RELATED WORK

Motion deblurring can be cast as the deconvolution of an image that has been convolved with either a global motion point spread function (PSF) or a spatially varying PSF. The problem is inherently ill-posed as there are a number of unblurred images that can produce the same blurred image after convolution. Nonetheless, this problem is well studied given its utility in photography and video capture. The following describes several related works.

Traditional deblurring. The majority of related work involves traditional blind deconvolution, which simultaneously estimates a global motion PSF and the deblurred image. These methods include well-known algorithms such as Richardson-Lucy [40], [33] and Wiener deconvolution [50]. For a survey on blind deconvolution, readers are referred to [20], [19]. These traditional approaches often produce less than desirable results that include artifacts such as ringing.

PSF estimation and priors. A recent trend in motion deblurring is to either constrain the solution of the deblurred image or to use auxiliary information to aid in either the PSF estimation or the deconvolution itself (or both). Examples include work by Fergus et al. [17], which used natural image statistics to constrain the solution to the deconvolved image. Raskar et al. [38] altered the shuttering sequence of a traditional camera to make the PSF more suitable for deconvolution. Jia [23] extracted an alpha mask of the blurred region to aid in PSF estimation. Dey et al. [15] modified the Richardson-Lucy algorithm by incorporating total variation regularization to suppress ringing artifacts. Levin et al. [28] introduced gradient sparsity constraints to reduce ringing artifacts. Yuan et al. [53] proposed a multi-scale nonblind deconvolution approach to progressively

recover motion-blurred details. Shan et al. [41] studied the relationship between estimation errors and ringing artifacts, and proposed the use of a spatial distribution model of image noise together with a local prior that suppresses ringing to jointly improve global motion deblurring.

Other recent approaches use more than one image to aid in the deconvolution process. Bascle et al. [5] processed a blurry image sequence to generate a single unblurred image. Yuan et al. [52] used a pair of images, one noisy and one blurred. Rav-Acha and Peleg [39] consider images that have been blurred in orthogonal directions to help estimate the PSF and constrain the resulting image. Chen and Tang [11] extend the work of Rav-Acha and Peleg [39] to remove the assumption of orthogonal blur directions. Bhat et al. [8] proposed a method that uses high-resolution photographs to enhance low-quality video, but this approach is limited to static scenes. Most closely related to ours is the work of Ben-Ezra and Nayar [6], [7], which used an additional imaging sensor to capture high-frame-rate imagery for the purpose of computing optical flow and estimating a global PSF. Li et al. [31] extend the work of Ben-Ezra and Nayar [6], [7] by using parallel cameras with different frame rates and resolutions, for the purpose of depth map estimation and not deblurring.

The aforementioned approaches assume the blur to arise from a global PSF. Recent work addressing spatially varying motion blur includes that of Levin [27], which used image statistics to correct a single motion blur on a stable background. Bardsley et al. [4] segmented an image into regions exhibiting similar blur, while Cho et al. [12] used two blurred images to simultaneously estimate local PSFs as well as deconvolve the two images. Ben-Ezra and Nayar [7] demonstrated how the auxiliary camera could be used to separate a moving object from the scene and apply deconvolution to this extracted layer. These approaches [27], [4], [12], [7], however, assume the motion blur to be globally invariant within each separated layer. Work by Shan et al. [42] allows the PSF to be spatially varying; however, the blur is constrained to that from rotational motion. Levin et al. [30] proposed a parabolic-motion camera designed for deblurring images with 1D object motion. During exposure, the camera moves in a manner that allows the resulting image to be deblurred using a single deconvolution kernel.

Super-resolution and upsampling. The problem of super-resolution can be considered as a special case of motion deblurring in which the blur kernel is a low-pass filter that is uniform in all motion directions. High-frequency details of a sharp image are, therefore, completely lost in the observed low-resolution image. There are two main approaches to super-resolution: image hallucination based on training data and image super-resolution computed from multiple low-resolution images. Our work is closely related to the latter approach, which is reviewed here. The most common technique for multiple image super-resolution is the back-projection algorithm proposed by Irani and Peleg [21], [22]. The back-projection algorithm is an iterative refinement procedure that minimizes the reconstruction errors of an estimated high-resolution image through a process of convolution, downsampling, and upsampling. A brief review that includes other early work



Fig. 1. Trade-off between resolution and frame rates. (a) Image from a high-resolution, low-frame-rate camera. (b) Images from a low-resolution, high-frame-rate camera.

on multiple image super-resolution is given in [10]. More recently, Patti et al. [37] proposed a method to align low-resolution video frames with arbitrary sampling lattices to reconstruct a high-resolution video. Their approach also uses optical flow for alignment and PSF estimation. These estimates, however, are global and do not consider local object motion. This work was extended by Elad and Feuer [16] to use adaptive filtering techniques. Zhao and Sawhney [55] studied the performance of multiple image super-resolution against the accuracy of optical flow alignment and concluded that the optical flows need to be reasonably accurate in order to avoid ghosting effects in super-resolution results. Shechtman et al. [43] proposed space-time super-resolution in which multiple video cameras with different resolutions and frame rates are aligned using homographies to produce outputs of either higher temporal and/or spatial sampling. When only two cameras are used, this approach can be considered a demonstration of a hybrid camera; however, this work does not address the scenario where severe motion blur is present in the high-resolution, low-frame-rate camera. Sroubek et al. [45] proposed a regularization framework for solving the multiple image super-resolution problem. This approach also does not consider local motion blur effects. Recently, Agrawal and Raskar [2] proposed a method to increase the resolution of images that have been

deblurred using a coded exposure system. Their approach can also be considered as a combination of motion deblurring and super-resolution, but is limited to translational motion.

Our work. While various previous works are related in part, our work is unique in its focus on spatially varying blur with no assumption on global or local motion paths. Moreover, our approach takes full advantage of the rich information available from the hybrid camera system, using techniques from both deblurring and super-resolution together in a single MAP framework. Specifically, our approach incorporates spatially varying deconvolution together with back-projection against the low-resolution frames. This combined strategy produces deblurred images with less ringing than traditional deconvolution, but with more detail than approaches using regularization and prior constraints. As with other deconvolution methods, we cannot recover frequencies that have been completely lost due to the motion blur and downsampling. A more detail discussion on our approach is provided in Section 4.4.

3 HYBRID CAMERA SYSTEM

The advantages of a hybrid camera system are derived from the additional data acquired by the LR-HFR camera. While

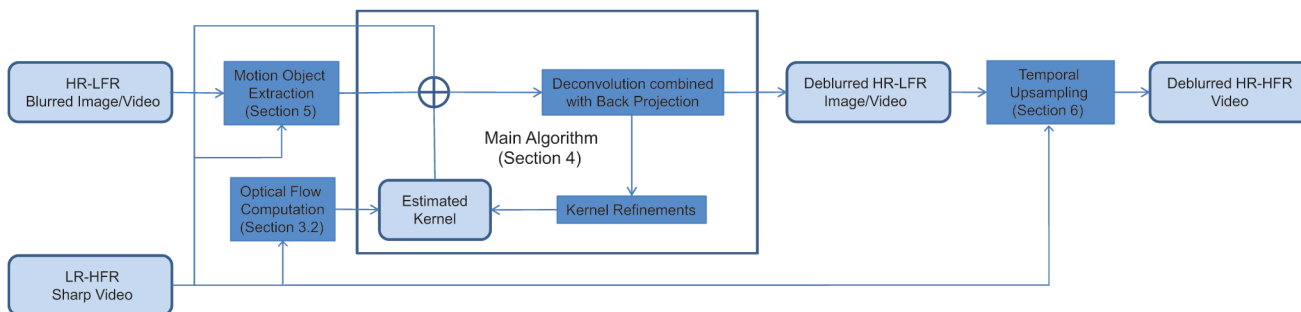


Fig. 2. The processing pipeline of our system. Optical flows are first calculated from the low-resolution, high-frame-rate (LR-HFR) video. From the optical flows, spatially varying motion blur kernels are estimated (Section 3.2). Then the main algorithm performs an iterative optimization procedure, which simultaneously deblurs the high-resolution, low-frame-rate (HR-LFR) image/video and refines the estimated kernels (Section 4). The output is a deblurred HR-LFR image/video. For the case of deblurring a moving object, the object is separated from the background prior to processing (Section 5). In the deblurring of video, we can additionally enhance the frame rate of the deblurred video to produce a high-resolution, high-frame-rate (HR-HFR) video result (Section 6).

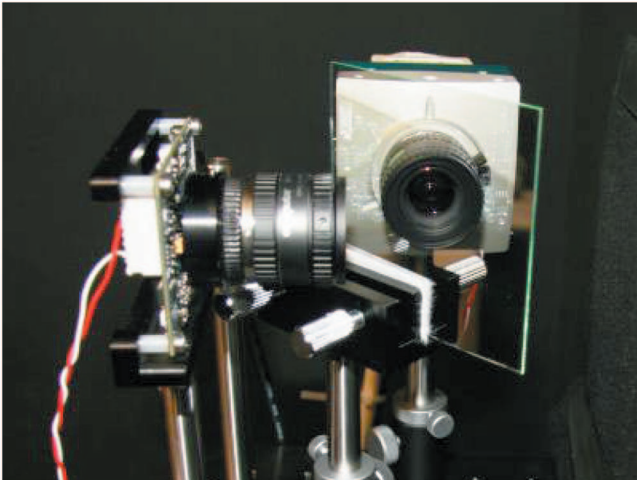


Fig. 3. Our hybrid camera combines a Point Gray Dragonfly II camera, which captures images of $1,024 \times 768$ resolution at 25 fps (6.25 fps for image deblurring examples), and a Mikrotron MC1311 camera that captures images of 128×96 resolution at 100 fps. A beamsplitter is employed to align their optical axes and respective images. Video synchronization is achieved using a 8051 microcontroller.

the spatial resolution of this camera is too low for many practical applications, the high-speed imagery is reasonably blur free and thus is suitable for optical flow computation. Fig. 1 illustrates an example. Since the cameras are assumed to be synchronized temporally and observing the same scene, the optical flow corresponds to the motion of the scene observed by the HR-LFR camera, whose images are blurred due to its slower temporal sampling. This ability to directly observe fast moving objects in the scene with the auxiliary camera allows us to handle a larger class of object motions without the use of prior motion models, since optical flow can be computed.

3.1 Camera Construction

Three conceptual designs of the hybrid camera system were discussed by Ben-Ezra and Nayar [6]. In their work, they implemented a simple design in which the two cameras are placed side by side such that their viewpoints can be considered the same when viewing a distant scene. A second design avoids the distant scene requirement by using a beam splitter to share between two sensing devices the light rays that pass through a single aperture, as demonstrated by McGuire et al. [36] for the studio matting problem. A promising third design is to capture both the HR-LFR and LR-HFR videos on a single sensor chip. According to [9], this can readily be achieved using a programmable CMOS sensing device.

In our work, we constructed a handheld hybrid camera system based on the second design as shown in Fig. 3. The two cameras are positioned such that their optical axes and pixel arrays are well aligned. Video synchronization is achieved using a 8051 microcontroller. To match the color responses of the two devices, we employ histogram mapping. In our implemented system, the exposure levels of the two devices are set to be equal, and the signal-to-noise ratios in the HR-LFR and LR-HFR images are approximately the same.

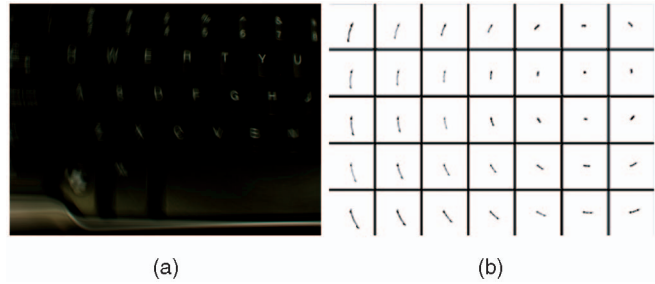


Fig. 4. Spatially varying blur kernel estimation using optical flows. (a) Motion blur image. (b) Estimated blur kernels of (a) from optical flows.

3.2 Blur Kernel Approximation Using Optical Flows

In the absence of occlusion, disocclusion, and out-of-plane rotation, a blur kernel can be assumed to represent the motion of a camera relative to objects in the scene. In [6], this relative motion is assumed to be constant throughout an image, and the globally invariant blur kernel is obtained through the integration of global motion vectors over a spline curve.

However, since optical flow is in fact a local estimation of motions, we can calculate spatially varying blur kernels from optical flows. We use the multiscale Lucas-Kanade algorithm [32] to calculate the optical flow at each pixel location. Following the brightness constancy assumption of optical flow estimation, we assume that our motion-blurred images are captured under constant illumination such that the change of pixel color of moving scene/object points over the exposure period can be neglected. The per-pixel motion vectors are then integrated to form spatially varying blur kernels, $K(x, y)$, one per pixel. This integration is performed as described by the authors of [6] for global motion. We use a spline curve with C1 continuity to fit the path of optical flow at position (x, y) . The number of frames used to fit the spline curve is 16 for image examples and 4 for video examples (Fig. 3). Fig. 4 shows an example of spatially varying blur kernels estimated from optical flows.

The estimated optical flows may contain noise that degrades blur kernel estimation. We found such noisy estimates to occur mainly in smooth or homogeneous regions that lack features for correspondence, while regions with sharp features tend to have accurate optical flows. Since deblurring artifacts are evident primarily around such features, the Lucas-Kanade optical flows are effective for our purposes. On the other hand, the optical flow noise in relatively featureless regions has little effect on deblurring results, since these areas are relatively unaffected by errors in the deblurring kernel. As a measure to heighten the accuracy and consistency of the estimated optical flows, we use local smoothing [51] as an enhancement of the multi-scale Lucas-Kanade algorithm [32].

The estimated blur kernels contain quantization errors due to the low resolution of the optical flows. Additionally, motion vector integration may provide an imprecise temporal interpolation of the flow observations. Our MAP optimization framework addresses these issues by refining the estimated blur kernels in addition to deblurring the video frames or images. Details of this kernel refinement will be discussed fully in Section 4.

3.3 Back-Projection Constraints

The capture of low-resolution frames in addition to the high-resolution images not only facilitates optical flow

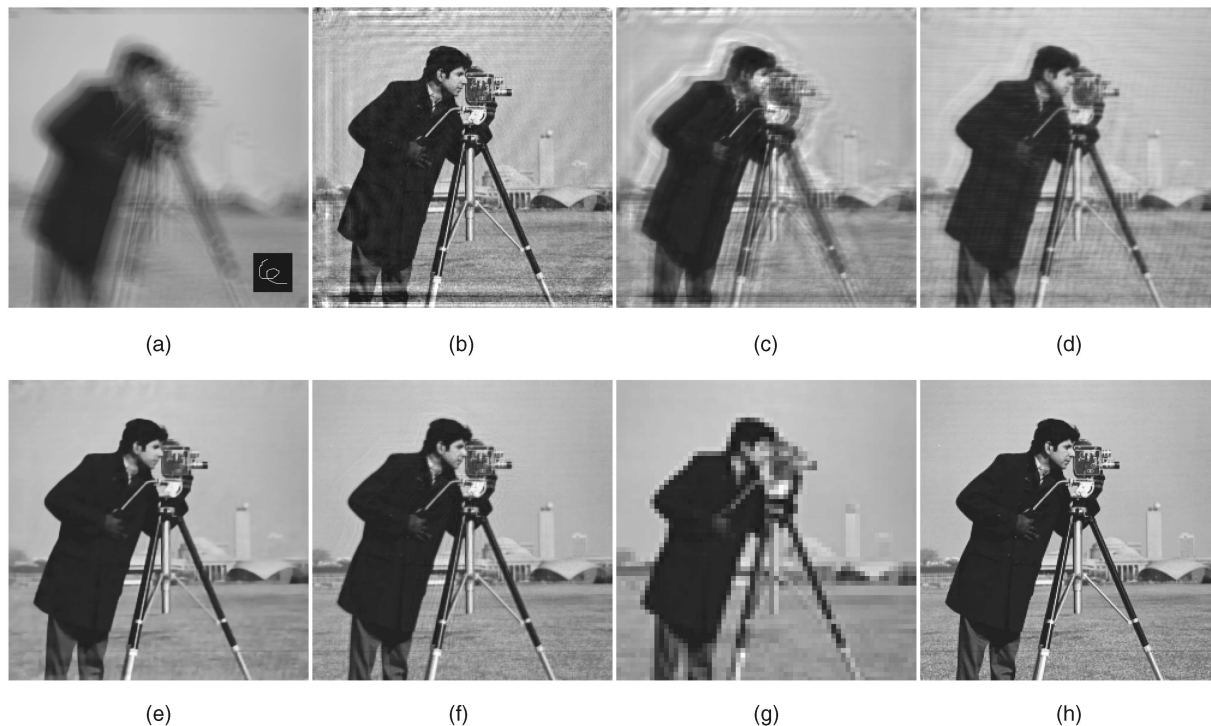


Fig. 5. Performance comparisons for different deconvolution algorithms on a synthetic example. The ground-truth motion blur kernel is used to facilitate comparison. The signal-to-noise ratio (SNR) of each result is reported. (a) A motion-blurred image (SNR(dB) = 25.62) with the corresponding motion blur kernel shown in the inset. Deconvolution results using (b) Wiener filter (SNR(dB) = 37.0), (c) Richardson-Lucy (SNR(dB) = 33.89), (d) total variation regularization (SNR(dB) = 36.13), (e) gradient sparsity prior (SNR(dB) = 46.370, and (f) our approach (SNR(dB) = 50.26 dB), which combines constraints from both deconvolution and super-resolution. The low-resolution image in (g) is eight times downsampled from the original image, shown in (h).

computation but also provides super-resolution-based reconstruction constraints [21], [22], [37], [10], [16], [3], [43] on the high-resolution deblurring solution. The back-projection algorithm [21], [22] is a common iterative technique for minimizing the reconstruction error and can be formulated as follows:

$$I^{t+1} = I^t + \sum_{j=1}^M (u(W(I_{l_j}) - d(I^t \otimes h))) \otimes p, \quad (1)$$

where M represents the number of corresponding low-resolution observations, t is an iteration index, I_{l_j} refers to the j th low-resolution image, $W(\cdot)$ denotes a warp function that aligns I_{l_j} to a reference image, \otimes is the convolution operation, h is the convolution filter before downsampling, p is a filter representing the back-projection process, and $d(\cdot)$ and $u(\cdot)$ are the downsampling and upsampling processes, respectively. Equation (1) assumes that each observation carries the same weight. In the absence of a prior, h is chosen to be a Gaussian filter with a size proportionate to the downsampling factor, and p is set equal to h .

In the hybrid camera system, a number of low-resolution frames are captured in conjunction with each high-resolution image. To exploit this available data, we align these frames according to the computed optical flows, and use them as back-projection constraints in (1). The number of low-resolution image constraints M is determined by the relative frame rates of the cameras. In our implementation, we choose the first low-resolution frame as the reference frame to which the estimated blur kernel and other low-resolution frames are aligned. Choosing a different low-resolution frame than the

reference frame would lead to a different deblurred result, which is a property that can be used to increase the temporal samples of the deblurred video as later discussed in Section 6.

The benefit of using multiple such back-projection constraints is illustrated in Fig. 5. Each of the low-resolution frames presents a physical constraint on the high-resolution solution in a manner that resembles how each offset image is used in a super-resolution technique. The effectiveness of incorporating the back-projection constraint to suppress ringing artifacts is demonstrated in Fig. 5 in comparison to several other deconvolution algorithms.

4 OPTIMIZATION FRAMEWORK

Before presenting our deblurring framework, we briefly review the Richardson-Lucy deconvolution algorithm, as our approach is fashioned in a similar manner. For the sake of clarity, our approach is first discussed for use in correcting global motion blur. This is followed by its extension to spatially varying blur kernels.

4.1 Richardson-Lucy Image Deconvolution

The Richardson-Lucy algorithm [40], [33] is an iterative maximum likelihood deconvolution algorithm derived from Bayes' theorem that minimizes the following estimation error:

$$\arg \min_I n(\|I_b - I \otimes K\|^2), \quad (2)$$

where I is the deblurred image, K is the blur kernel, I_b is the observed blur image, and $n(\cdot)$ is the image noise

distribution. A solution can be obtained using the iterative update algorithm defined as follows:

$$I^{t+1} = I^t \times K * \frac{I_b}{I^t \otimes K}, \quad (3)$$

where $*$ is the correlation operation. A blind deconvolution method using the Richardson-Lucy algorithm was proposed by Fish et al. [18] which iteratively optimizes I and K in alternation using (3) with the positions of I and K switched during optimization iterations for K . The Richardson-Lucy algorithm assumes image noise $n(\cdot)$ to follow a Poisson distribution. If we assume image noise to follow a Gaussian distribution, then a least-squares method can be employed [21]:

$$I^{t+1} = I^t + K * (I_b - I^t \otimes K), \quad (4)$$

which shares the same iterative back-projection update rule as (1).

From video input with computed optical flows, multiple blurred images I_b and blur kernels K may be acquired by reversing the optical flows of neighboring high-resolution frames. These multiple observation constraints can be jointly applied in (4) [39] as

$$I^{t+1} = I^t + \sum_{i=1}^N w_i K_i * (I_{b_i} - I^t \otimes K_i), \quad (5)$$

where N is the number of aligned observations.

4.2 Optimization for Global Kernels

In solving for the deblurred images, our method jointly employs the multiple deconvolution and back-projection constraints available from the hybrid camera input. For simplicity, we assume in this section that the blur kernels are spatially invariant. Our approach can be formulated into an MAP estimation framework as follows:

$$\begin{aligned} & \arg \max_{I, K} P(I, K | I_b, K_o, I_l) \\ &= \arg \max_{I, K} P(I_b | I, K) P(K_o | I, K) P(I_l | I) P(I) P(K) \\ &= \arg \min_{I, K} L(I_b | I, K) + L(K_o | I, K) + L(I_l | I) + L(I) + L(K), \end{aligned} \quad (6)$$

where I and K denote the sharp images and the blur kernels we want to estimate; I_b , K_o , and I_l are the observed blurred images, estimated blur kernels from optical flows, and the low-resolution, high-frame-rate images, respectively; and $L(\cdot) = -\log(P(\cdot))$. In our formulation, the priors $P(I)$ and $P(K)$ are taken to be uniformly distributed. Assuming that $P(K_o | I, K)$ is conditionally independent of I , that the estimation errors of likelihood probabilities $P(I_b | I, K)$, $P(K_o | I, K)$, and $P(I_l | I)$ follow Gaussian distributions, and that each observation of I_b , K_o , and I_l is independent and identically distributed, we can then rewrite (6) as

$$\begin{aligned} & \arg \min_{I, K} \sum_{i=1}^N \|I_{b_i} - I \otimes K_i\|^2 \\ &+ \lambda_B \sum_{j=1}^M \|I_{l_j} - d(I \otimes h)\|^2 + \lambda_K \sum_{i=1}^N \|K_i - K_{o_i}\|^2, \end{aligned} \quad (7)$$

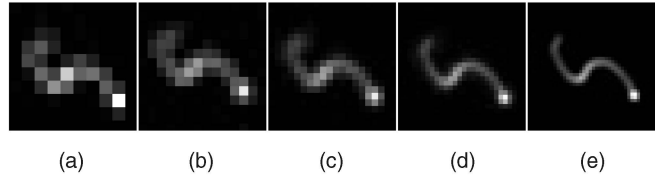


Fig. 6. Multiscale refinement of a motion blur kernel for the image in Figs. 11a, b, c, d, and e exhibits refined kernels at progressively finer scales. Our kernel refinement starts from the coarsest level. The result of each coarser level is then upsampled and used as an initial kernel estimate for the next level of refinement.

where λ_K and λ_B are the relative weights of the error terms. To optimize the above equation for I and K , we employ alternating minimization. Combining (1) and (5) yields our iterative update rules as follows:

1. update

$$\begin{aligned} I^{t+1} = I^t + \sum_{i=1}^N K_i^t * (I_{b_i} - I^t \otimes K_i^t) \\ + \lambda_B \sum_{j=1}^M h \otimes (u(W(I_{l_j}) - d(I^t \otimes h))), \end{aligned}$$

2. update

$$K_i^{t+1} = K_i^t + \tilde{I}^{t+1} * (I_{b_i} - I^{t+1} \otimes K_i^t) + \lambda_K (K_{o_i} - K_i^t),$$

where $\tilde{I} = I / \sum_{(x,y)} I(x,y)$, $I(x,y) \geq 0$, $K_i(u,v) \geq 0$, and $\sum_{(u,v)} K_i(u,v) = 1$. The two update steps are processed in alternation until the change in I falls below a specified level or until a maximum number of iterations is reached. The term $W(I_{l_j})$ is the warped aligned observations. The reference frame to which these are aligned to can be any of the M low-resolution images. Thus, for each deblurred high-resolution frame, we have up to M possible solutions. This will later be used in the temporal upsampling described in Section 6. In our implementation, we set $N = 3$ in correspondence to the current, previous, and next frames, and M is set according to the relative camera settings (4/16 for video/image deblurring in our implementation). We also initialize I^0 as the currently observed blurred image I_b , K_i^0 as the estimated blur kernel K_{o_i} from optical flows, and set $\lambda_B = \lambda_K = 0.5$.

For more stable and flexible kernel refinement, we refine the kernel in a multiscale fashion as done in [17], [52]. Fig. 6 illustrates the kernel refinement process. We estimate PSFs from optical flows of the observed low-resolution images and then downsample to the coarsest level. After refinement at a coarser level, kernels are then upsampled and refined again. The multiscale pyramid is constructed using a downsampling factor of $1/\sqrt{2}$ with five levels. The likelihood $P(K_o | K)$ is applied at each level of the pyramid with a decreasing weight so as to allow more flexibility in refinement at finer levels. We note that starting at a level coarser than the low-resolution images allows our method to recover from some error in PSF estimation from optical flows.

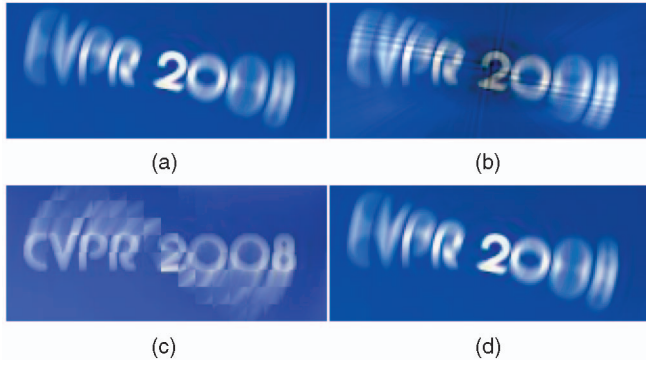


Fig. 7. Convolution with kernel decomposition. (a) Convolution result without kernel decomposition, where full blur kernels are generated on the fly per pixel using optical flow integration. (b) Convolution using 30 PCA-decomposed kernels. (c) Convolution using a patch-based decomposition. (d) Convolution using delta function decomposition of kernels, with at most 30 delta functions per pixel.

4.3 Spatially Varying Kernels

A spatially varying blur kernel can be expressed as $K(x, y, u, v)$, where (x, y) is the image coordinate and (u, v) is the kernel coordinate. For large-sized kernels, e.g., 65×65 , this representation is impractical due to its enormous storage requirements. Recent work has suggested ways to reduce the storage size, such as by constraining the motion path [42]; however, our approach places no constraints on possible motion. Instead, we decompose the spatially varying kernels into a set of P basis kernels k_l whose mixture weights a_l are a function of image location:

$$K(x, y, u, v) = \sum_{l=1}^P a_l(x, y) k_l(u, v). \quad (8)$$

The convolution equation then becomes

$$I(x, y) \otimes K(x, y, u, v) = \sum_{l=1}^P a_l(x, y) (I(x, y) \otimes k_l(u, v)). \quad (9)$$

In related work [26], principal components analysis (PCA) is used to determine the basis kernels. PCA, however, does not guarantee positive kernel values, and we have found in our experiments that PCA-decomposed kernels often lead to unacceptable ringing artifacts, exemplified in Fig. 7b. The ringing artifacts in the convolution

result resemble the patterns of basis kernels. Another method is to use a patch representation which segments images into many small patches such that the local motion blur kernel is the same within each small patch. This method was used Joshi et al. [25], but their blur kernels are defocus kernels with very small variations within local areas. For large object motion, blur kernels in the patch-based method would not be accurate, leading to discontinuity artifacts as shown in Fig. 7c. We instead choose to use a delta function representation, where each delta function represents a position (u, v) within a kernel. Since a motion blur kernel is typically sparse, we store only 30-40 delta functions for each image pixel, where the delta function positions are determined by the initial optical flows. From the total of 65×65 possible delta functions in the spatial kernel at each pixel in the image, we find, in practice, that we only use about 500-600 distinct delta functions to provide a sufficient approximation of the spatially varying blur kernels in the convolution process. Examples of basis kernel decomposition using PCA and the delta function representation are shown in Fig. 8. The delta function representation also offers more flexibility in kernel refinement, while refinements using the PCA representation are limited to the PCA subspace.

By combining (9) and (7), our optimization function becomes

$$\arg \min_{I, K} \left\| \sum_i^N I_{b_i} - \sum_l^P a_{il} (I \otimes k_{il}) \right\|^2 + \lambda_B \sum_j^M \|I_{l_j} - d(I \otimes h)\|^2 + \lambda_K \sum_i^N \sum_l^P \|a_{il} k_{il} - a_{oi} k_{il}\|^2. \quad (10)$$

The corresponding iterative update rules are then

1. update

$$I^{t+1} = I^t + \sum_{i=1}^N \sum_{l=1}^P a_{il}^t k_{il} * \left(I_{b_i} - \sum_l^P a_{il}^t (I^t \otimes k_{il}) \right) + \lambda_B \sum_{j=1}^M h \otimes (u(W(I_{l_j}) - d(I^t \otimes h))),$$

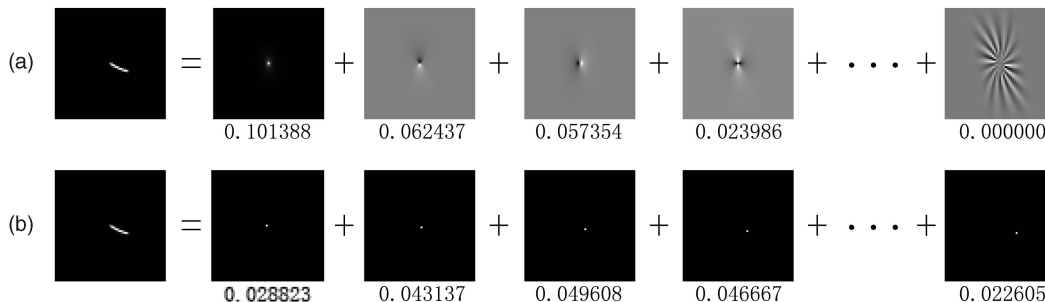


Fig. 8. PCA versus the delta function representation for kernel decomposition. The top row illustrates the kernel decomposition using PCA and the bottom row shows the decomposition using the delta function representation. The example kernel is taken from among the spatially varying kernels of Fig. 7 from which the basis kernels are derived. Weights are displayed below each of the basis kernels. The delta function representation not only guarantees positive values of basis kernels but also provides more flexibility in kernel refinement.

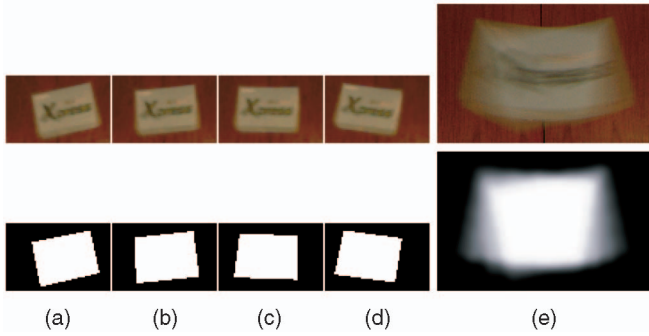


Fig. 9. Layer separation using a hybrid camera: (a)-(d) low-resolution frames and their corresponding binary segmentation masks. (e) High-resolution frame and the matte estimated by compositing the low-resolution segmentation masks with smoothing.

2. update

$$a_{il}^{t+1} = a_{il}^t + \left(\tilde{I}^{t+1} * \left(I'_{b_i} - \sum_l^P a_{il}^t (I^{t+1} \otimes k_{il}) \right) \right) \cdot k_{il} + \lambda_K (a_{o_{il}} - a_{il}^t),$$

where I' and I'_b are local windows in the estimated result and the blur image. This kernel refinement can be implemented in a multiscale framework for greater flexibility and stability.

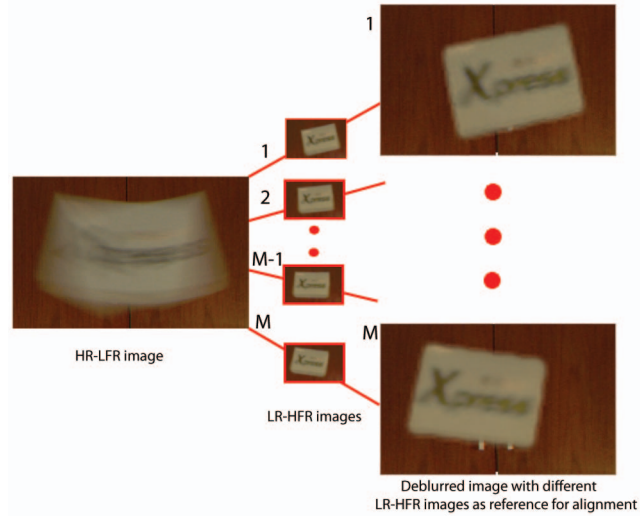


Fig. 10. Relationship of high-resolution deblurred result to corresponding low-resolution frame. Any of the low-resolution frames can be selected as a reference frame for the deblurred result. This allows up to M deblurred solutions to be obtained.

The number of delta functions k_{il} stored at each pixel position may be reduced when an updated value of a_{il} becomes insignificant. For greater stability, we process each update rule five times before switching to the other.

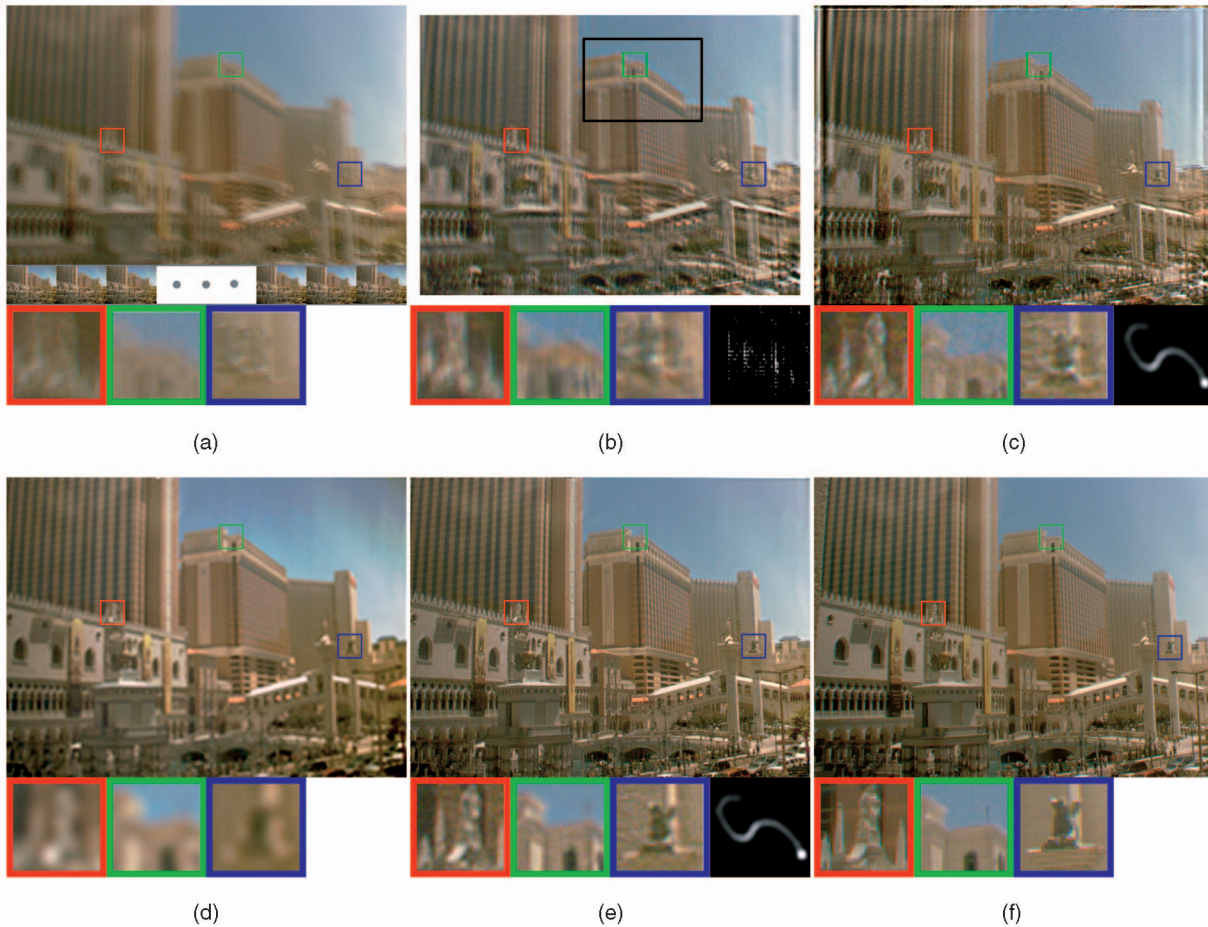


Fig. 11. Image deblurring using globally invariant kernels. (a) Input. (b) Result generated with the method of [17], where the user-selected region is indicated by a black box. (c) Result generated by Ben-Ezra and Nayar [6]. (d) Result generated by back projection [21]. (e) Our results. (f) The ground-truth sharp image. Close-up views and the estimated global blur kernels are also shown.



Fig. 12. Image deblurring with spatial varying kernels from rotational motion. (a) Input. (b) Result generated with the method of [42] (obtained courtesy of the authors of [42]). (c) Result generated by Ben-Ezra and Nayar [6] using spatially varying blur kernels estimated from optical flow. (d) Result generated by back projection [21]. (e) Our results. (f) The ground-truth sharp image. Close-ups are also shown.

4.4 Discussion

Utilizing both deconvolution of high-resolution images and back projection from low-resolution images offers distinct advantages because the deblurring information from these two sources tends to complement each other. This can be intuitively seen by considering a low-resolution image to be a sharp high-resolution image that has undergone motion blurring with a Gaussian PSF and bandlimiting. Back projection may then be viewed as a deconvolution with a Gaussian blur kernel that promotes recovery of lower frequency image features without artifacts. On the other hand, deconvolution of high-resolution images with the high-frequency PSFs typically associated with camera and object motion generally supports reconstruction of higher frequency details, especially those orthogonal to the motion direction. While some low-frequency content can also be restored from motion blur deconvolution, there is often significant loss due to the large support regions for motion blur kernels, and this results in ringing artifacts. As discussed in [39], the joint use of images having such different blur functions and deconvolution information favors a better deblurring solution.

Multiple motion blur deconvolutions and multiple back projections can further help to generate high-quality results. Differences in motion blur kernels among neighboring frames provide different frequency information, and

multiple back-projection constraints help to reduce quantization and the effects of noise in low-resolution images. In some circumstances, there exists redundant information from a given source, such as when high-resolution images contain identical motion blur or when low-resolution images are offset by integer pixel amounts. This makes it particularly important to utilize as much deblurring information as can be obtained.

Our current approach does not utilize priors on the deblurred image or the kernels. With constraints from the low-resolution images, we have found these priors to be unneeded. Fig. 5 compares our approach with other deconvolution algorithms. Specifically, we compare our approach with Total Variation regularization [15] and Sparsity Priors [28], which have recently been shown to produce better results than traditional Wiener filtering [50] and the Richardson-Lucy [40], [33] algorithm. Both Total Variation regularization and Sparsity Priors produce results with less ringing artifacts. There are almost no ringing artifacts with Sparsity Priors, but many fine details are lost. In our approach, most medium to large-scale ringing artifacts are removed using the back-projection constraints, while fine details are recovered through deconvolution.

Although our approach can acquire and utilize a greater amount of data, high-frequency details that have been lost by both motion blur and downsampling cannot be recovered. This is a fundamental limitation of any

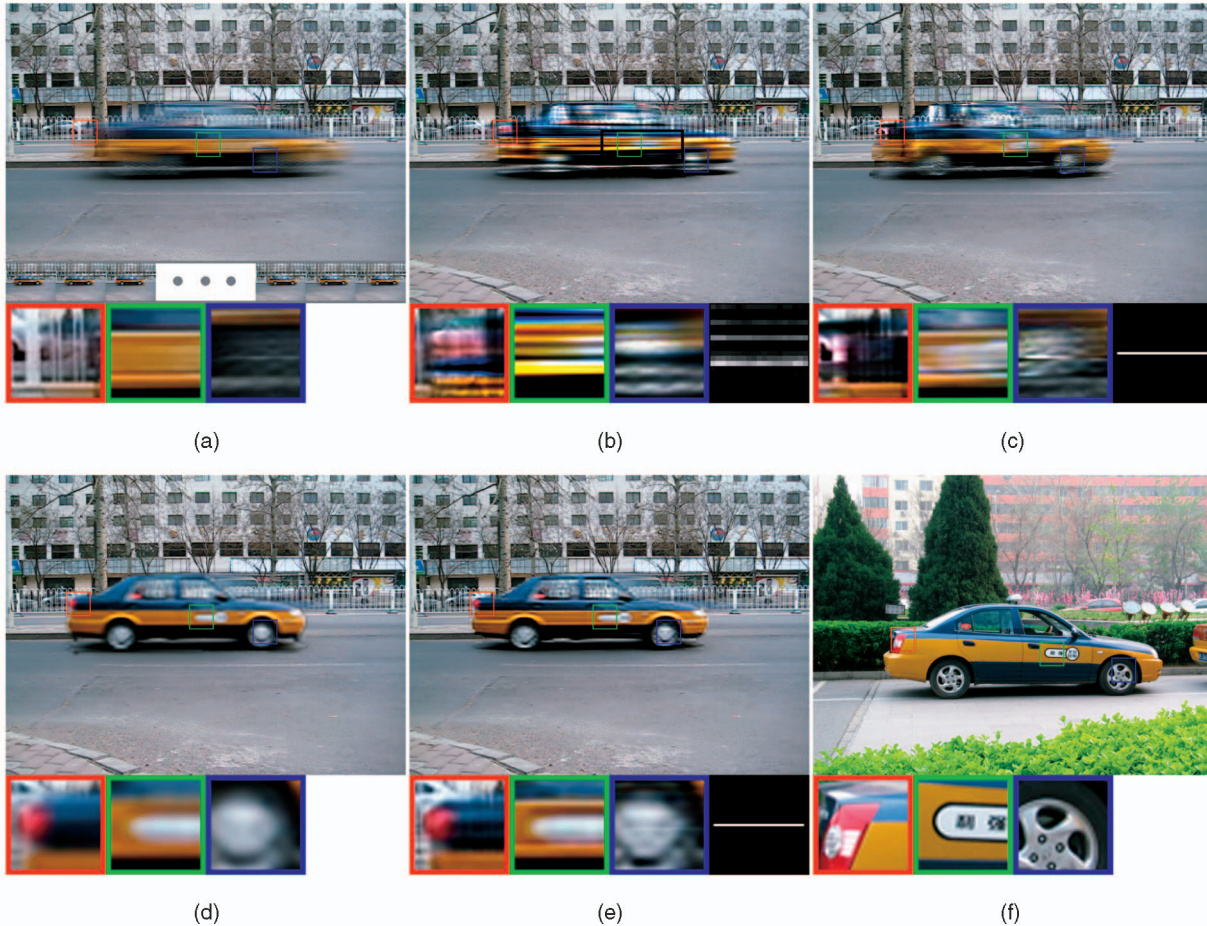


Fig. 13. Image deblurring with translational motion. In this example, the moving object is a car moving horizontally. We assume that the motion blur within the car is globally invariant. (a) Input. (b) Result generated by Fergus et al. [17], where the user-selected region is indicated by the black box. (c) Result generated by Ben-Ezra and Nayar [6]. (d) Result generated by back projection [21]. (e) Our results. (f) The ground-truth sharp image captured from another car of the same model. Close-up views and the estimated global blur kernels within the motion blur layer are also shown.

deconvolution algorithm that does not hallucinate detail. We also note that reliability in optical flow cannot be assumed beyond a small time interval. This places a restriction on the number of motion blur deconvolution constraints that can be employed to deblur a given frame.

Finally, we note that the iterative back-projection technique incorporated into our framework is known to have convergence problems. Empirically, we have found that stopping after no more than 50 iterations of our algorithm produces acceptable results.

5 DEBLURRING OF MOVING OBJECTS

To deblur a moving object, a high-resolution image needs to be segmented into different layers because pixels on the blended boundaries of moving objects contain both foreground and background components, each with different relative motion to the camera. This layer separation is inherently a matting problem that can be expressed as

$$I = \alpha F + (1 - \alpha)B, \quad (11)$$

where I is the observed image intensity, F , B , and α are the foreground color, background color, and alpha value of the fractional occupancy of the foreground. The matting

problem is an ill-posed problem since the number of unknown variables is greater than the number of observations. Single-image approaches require user assistance to provide a trimap [14], [13], [46] or scribbles [49], [29], [48] for collecting samples of the foreground and background colors. Fully automatic approaches, however, have required either a blue background [44], multiple cameras with different focus [35], polarized illumination [36], or a camera array [24]. In this section, we propose a simple solution to the layer separation problem that takes advantage of the hybrid camera system.

Our approach assumes that object motion does not cause motion blur in the high-frame-rate camera such that the object appears sharp. To extract the alpha matte of a moving object, we perform binary segmentation of the moving object in the low-resolution images and then compose the binary segmentation masks with smoothing to approximate the alpha matte in the high-resolution image. We note that Ben-Ezra and Nayar [7] used a similar strategy to perform layer segmentation in their hybrid camera system. In Fig. 9, an example of this matte extraction is demonstrated together with the moving object separation method of Zhang et al. [54]. The foreground color F must also be estimated for deblurring. This can be done by assuming a local color smoothness

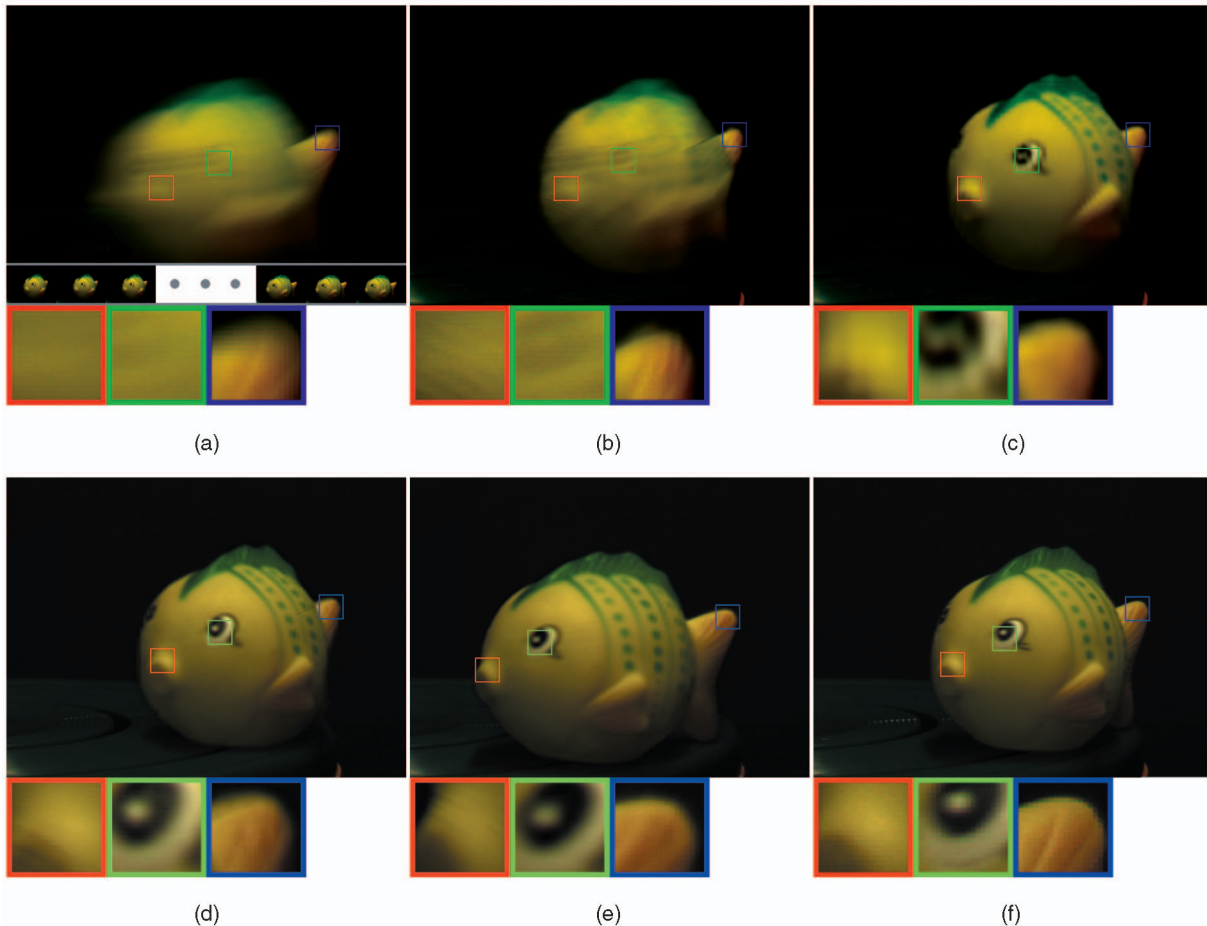


Fig. 14. Image deblurring with spatially varying kernels. In this example, the moving object contains out-of-plane rotation with both occlusion and disocclusion at the object boundary. (a) Input. (b) Result generated by Ben-Ezra and Nayar [6]. (c) Result generated by back projection [21]. (d) Our results using the first low-resolution frame as the reference frame. (e) Our results using the last low-resolution frame as the reference frame. (f) The ground-truth sharp image. Close-ups are also shown.

prior on F and B and solving for their values with Bayesian matting [14]:

$$\begin{aligned} & \begin{bmatrix} \Sigma_F^{-1} + \mathbf{I}\alpha^2/\sigma_I^2 & \mathbf{I}\alpha(1-\alpha)/\sigma_I^2 \\ \mathbf{I}\alpha(1-\alpha)/\sigma_I^2 & \Sigma_B^{-1} + \mathbf{I}(1-\alpha)^2/\sigma_I^2 \end{bmatrix} \begin{bmatrix} F \\ B \end{bmatrix} \\ & = \begin{bmatrix} \Sigma_F^{-1}\mu_F + \mathbf{I}\alpha/\sigma_I^2 \\ \Sigma_B^{-1}\mu_B + \mathbf{I}(1-\alpha)/\sigma_I^2 \end{bmatrix}, \end{aligned} \quad (12)$$

where (μ_F, Σ_F) and (μ_B, Σ_B) are the local color mean and covariance matrix (Gaussian distribution) of the foreground and background colors, \mathbf{I} is a 3×3 identity matrix, and σ_I is the standard derivation of I , which models estimation errors of (11). Given the solution of F and B , the α solution can be refined by solving (11) in closed form. Refinements of F , B , and α can be done in alternation to further improve the result.

Once moving objects are separated, we deblur each layer separately using our framework. The alpha mattes are also deblurred for compositing, and the occluded background areas revealed after alpha mask deblurring can then be filled in either by back projection from the low-resolution images or by the motion inpainting method of [34].

6 TEMPORAL UPSAMPLING

Unlike deblurring of images, videos require deblurring of multiple consecutive frames in a manner that preserves temporal consistency. As described in Section 4.2, we can jointly use the current, previous, and subsequent frames to deblur the current frame in a temporally consistent way. However, after sharpening each individual frame, temporal discontinuities in the deblurred high-resolution, low-frame-rate video may become evident through some jumpiness in the sequence. In this section, we describe how our method can alleviate this problem by increasing the temporal sampling rate to produce a deblurred high-resolution, high-frame-rate video.

As discussed by Shechtman et al. [43], temporal super-resolution results when an algorithm can generate an output with a temporal rate that surpasses the temporal sampling of any of the input devices. While our approach generates a high-resolution video at greater temporal rate than the input high-resolution, low-frame-rate video, its temporal rate is bounded by the frame rate of the low-resolution, high-frame-rate camera. We therefore refrain from the term super-resolution and refer to this as *temporal upsampling*.

Our solution to temporal upsampling derives directly from our deblurring algorithm. In our scenario, we have M high-frame-rate low-resolution frames corresponding to

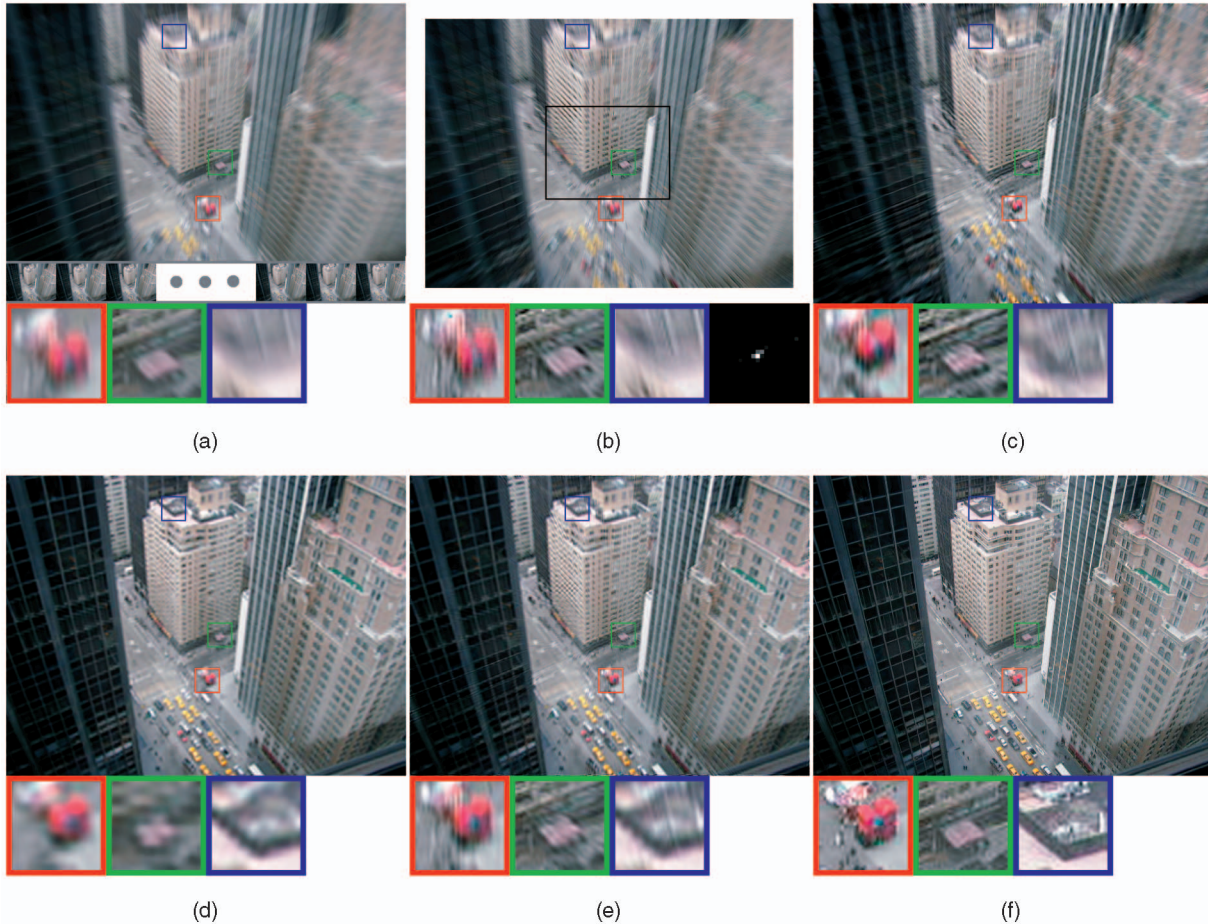


Fig. 15. Image deblurring with spatially varying kernels. In this example, the camera is zooming into the scene. (a) Input. (b) Result generated by Fergus et al. [17]. (c) Result generated by Ben-Ezra and Nayar [6]. (d) Result generated by back projection [21]. (e) Our results. (f) The ground-truth sharp image. Close-ups are also shown.

each high-resolution, low-frame-rate motion-blurred image. Fig. 10 shows an example. With our algorithm, we therefore have the opportunity to estimate M solutions using each one of the M low-resolution frames as the basic reference frame. While the ability to produce multiple deblurred frames is not a complete solution to temporal upsampling, here the use of these M different reference frames leads to a set of deblurred frames that is consistent with the temporal sequence. This unique feature of our approach is gained through the use of the hybrid camera to capture low-resolution, high-frame-rate video in addition to the standard high-resolution, low-frame-rate video. The low-resolution, high-frame-rate video not only aids in estimating the motion blur kernels and provides back-projection constraints, but can also help to increase the deblurred video frame rate. The result is a high-resolution, high-frame-rate deblurred video.

7 RESULTS AND COMPARISONS

We evaluate our deblurring framework using real images and videos. In these experiments, a ground-truth blur-free image is acquired by mounting the camera on a tripod and capturing a static scene. Motion blurred images are then obtained by moving the camera and/or introducing a dynamic scene object. We show examples of several different cases: **globally invariant motion blur** caused by camera shake, **in-plane rotational motion** of a scene object,

translational motion of a scene object, **out-of-plane rotational motion** of an object, **zoom-in motion** caused by changing the focal length (i.e., camera’s zoom setting), a combination of translational motion and rotational motion with **multiple frames** used as input for deblurring one frame, **video deblurring with out-of-plane rotational motion**, **video deblurring with complex in-plane motion**, and **video deblurring with a combination of translational and zoom-in motion**.

Globally invariant motion blur. In Fig. 11, we present an image deblurring example with globally invariant motion, where the input is one high-resolution image and several low-resolution images. Our results are compared with those generated by the methods of Fergus et al. [17], Ben-Ezra and Nayar [6], and back projection [21]. Fergus et al.’s approach is a state-of-the-art blind deconvolution technique that employs a natural image statistics constraint. However, when the blur kernel is not correctly estimated, an unsatisfactory result shown in (b) will be produced. Ben-Ezra and Nayar use the estimated optical flow as the blur kernel and then perform deconvolution. Their result in (c) is better than that in (b) as the estimated blur kernel is more accurate, but ringing artifacts are still unavoidable. Back projection produces a super-resolution result from a sequence of low-resolution images as shown in (d). Noting that motion blur removal is not the intended application of back projection, we can see that its results are blurry since

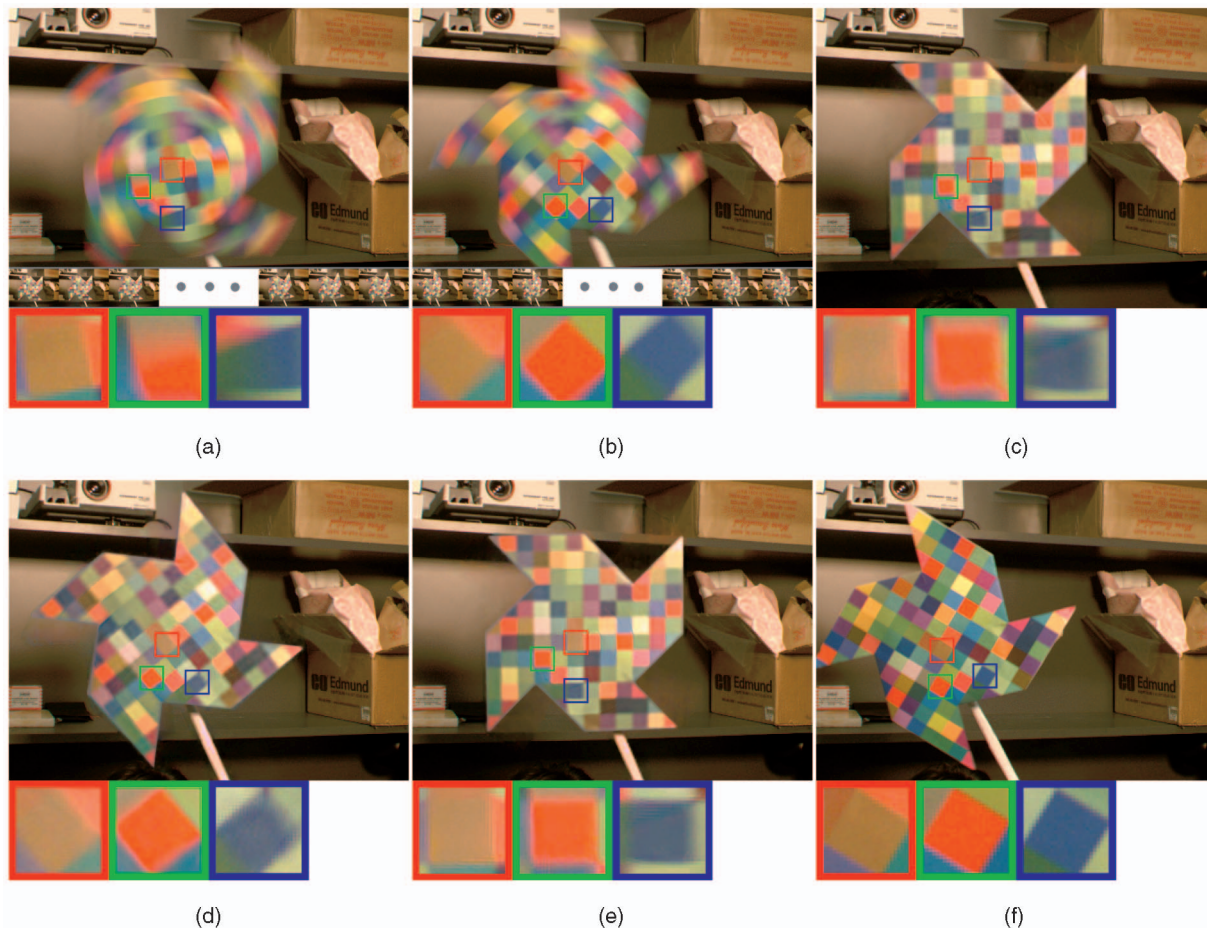


Fig. 16. Deblurring with and without multiple high-resolution frames. (a) and (b) Input images containing both translational and rotational motion blur. (c) Deblurring using only (a) as input. (d) Deblurring using only (b) as input. (e) Deblurring of (a) using both (a) and (b) as inputs. (f) Ground-truth sharp image. Close-ups are also shown.

the high-frequency details are not sufficiently captured in the low-resolution images. The result of our method and the refined kernel estimate are displayed in (e). The ground truth is given in (f) for comparison.

In-plane rotational motion. Fig. 12 shows an example with in-plane rotational motion. We compared our result with those by Shan et al. [42], Ben-Ezra and Nayar [6], and back-projection [21]. Shan et al.'s [42] is a recent technique that targets deblurring of in-plane rotational motion. Our approach is seen to produce less ringing artifacts compared to [42] and [6], and it generates greater detail than [21].

Translational motion. Fig. 13 shows an example of a car translating horizontally. We assume that the motion blur within the car region is globally invariant and thus techniques for removing globally invariant motion blur can be applied after layer separation of the moving object. We use the technique proposed in Section 5 to separate the moving car from the static background. Our results are compared with those generated by Fergus et al. [17], Ben-Ezra and Nayar [6], and back projection [21]. In this example, the moving car is severely blurred with most of the high-frequency details lost. We demonstrate in (c) the limitation of using deconvolution alone even with an accurate motion blur kernel. In this example, the super-resolution result in (d) is better than the deconvolution result, but there are some high-frequency details that are not recovered. Our result is shown in (e), which maintains most low-frequency details recovered by

super-resolution and also high-frequency details recovered by deconvolution. Some incorrect high-frequency details from the static background are incorrectly retained in our final result because of the presence of some high-frequency background details in the separated moving object layer. We believe that a better layer separation algorithm would lead to improved results. This example also exhibits a basic limitation of our approach. Since there is significant car motion during the exposure time, most high-frequency detail is lost and cannot be recovered by our approach. The ground truth in (f) shows a similar, parked car for comparison.

Out-of-plane rotational motion. Fig. 14 shows an example of out-of-plane rotation, where occlusion/disocclusion occurs at the object boundary. Our result is compared to that of Ben-Ezra and Nayar [6] and back projection [21]. One major advantage of our approach is that we can detect the existence of occlusions/disocclusions of the motion-blurred moving object. This not only helps to estimate the alpha mask for layer separation but also aids in eliminating irrelevant low-resolution reference frame constraints for back projection. We show our result by choosing the first frame and the last frame as the reference frame. Both occlusion and disocclusion are contained in this example.

Zoom-in motion. Fig. 15 shows another example of motion blur from zoom-in effects. Our result is compared to Fergus et al. [17], Ben-Ezra and Nayar [6], and back



Fig. 17. Video deblurring with out-of-plane rotational motion. The moving object is a vase with a center of rotation approximately aligned with the image center. (a) Input video frames. (b) Close-ups of a motion-blurred region. (c) Deblurred video. (d) Close-ups of deblurred video using the first low-resolution frames as the reference frames. (e) Close-ups of deblurred video frames using the fifth low-resolution frames as the reference frames. The final video sequence has higher temporal sampling than the original high-resolution video and is played with frames ordered according to the red lines.

projection [21]. We note that the method of Fergus et al. [17] is intended for globally invariant motion blur and is shown here to demonstrate the effects of using only a single blur kernel to deblur spatially varying motion blur. Again, our approach produces better results with less ringing artifacts and richer detail.

Deblurring with multiple frames. The benefit of using multiple deconvolutions from multiple high-resolution frames is exhibited in Fig. 16 for a pinwheel with both translational and rotational motions. The deblurring result in (c) was computed using only (a) as input. Likewise, (d) is the deblurred result from only (b). Using both (a) and (b) as inputs yields the improved result in (e). This improvement can be attributed to the difference in high-frequency detail that can be recovered from each of the differently blurred images. The ground truth is shown in (f) for comparison.

Video deblurring with out-of-plane rotational motion. Fig. 17 demonstrates video deblurring of a vase with out-of-plane rotation. The center of rotation is approximately aligned with the image center. The top row displays five consecutive input frames. The second row shows close-ups of a motion-blurred region. The middle row shows our results with the first low-resolution frames as the reference frames. The fourth and fifth rows show close-ups of our results with respect to the first and fifth low-resolution frames as the reference frames.

This example also demonstrates the ability to produce multiple deblurring solutions as described in Section 6. For temporal upsampling, we combine the results together in

the order indicated by the red lines in Fig. 17. With our method, we can increase the frame rate of deblurred high-resolution videos up to the same rate as the low-resolution, high-frame-rate video input.

Video deblurring with complex in-plane motion. Fig. 18 presents another video deblurring result of a tossed box with complex (in-plane) motion. The top row displays five consecutive input frames. The second row shows close-ups of the motion-blurred moving object. The middle row shows our separated mattes for the moving object, and the fourth and fifth rows present our results with the first and third low-resolution frames as reference. The text on the tossed box is recovered to a certain degree by our video deblurring algorithm. As in the previous video deblurring example, our output is a high-resolution, high-frame-rate deblurred video. This result also illustrates a limitation of our method, where the shadow of the moving object is not deblurred and may appear inconsistent. This problem is a direction for future investigation.

Video deblurring with a combination of translational and zoom-in motions. Our final example is shown in Fig. 19. The moving object of interest is a car driving toward the camera. Both translational effects and zoom-in blur effects exist in this video deblurring example. The top row displays five consecutive frames of input. The second row shows close-ups of the motion-blurred moving object. The middle row shows our extracted mattes for the moving object, and the fourth and fifth rows present our results with the first and fifth low-resolution frames as reference.

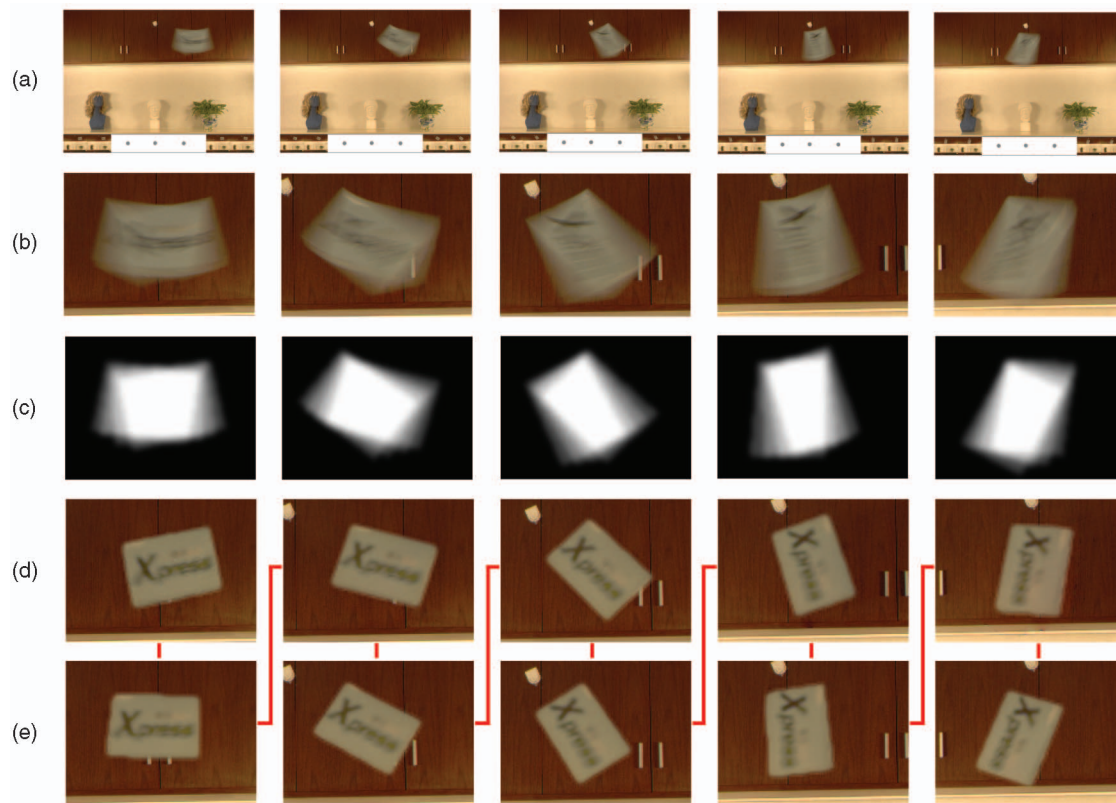


Fig. 18. Video deblurring with a static background and a moving object. The moving object is a tossed box with arbitrary (in-plane) motion. (a) Input video frames. (b) Close-up of the motion-blurred moving object. (c) Extracted alpha mattes of the moving object. (d) The deblurred video frames using the first low-resolution frames as the reference frames. (e) The deblurred video frames using the third low-resolution frames as the reference frames. The final video with temporal super-resolution is played with frames ordered as indicated by the red lines.

8 CONCLUSION

We have proposed an approach for image/video deblurring using a hybrid camera. Our work has formulated the deblurring process as an iterative method that incorporates optical flow, back projection, kernel refinement, and frame coherence to effectively combine the benefits of both deconvolution and super-resolution. We demonstrate that this approach can produce results that are sharper and cleaner than state-of-the-art techniques.

While our video deblurring algorithm exhibits high-quality results on various scenes, there exist complicated forms of spatially varying motion blur that can be difficult for our method to handle (e.g., motion blur effects caused by object deformation). The performance of our algorithm is also bounded by the performance of several of its components, including optical flow estimation, layer separation, and also the deconvolution algorithm. Despite these limitations, we have proposed the first work to handle spatially varying motion blur with arbitrary in-plane/out-of-plane rigid motion. This work is also the first to address video deblurring and increase video frame rates using a deblurring algorithm.

Future research directions for this work include how to improve the deblurring performance through incorporating priors into our framework. Recent deblurring methods have demonstrated the utility of priors, such as the natural image statistics prior and the sparsity prior, for reducing ringing

artifacts and for kernel estimation. Another research direction is to improve layer separation by more fully exploiting the available information in the hybrid camera system. Additional future work may also be done on how to recover the background partially occluded by a motion-blurred object.

REFERENCES

- [1] M. Aggarwal and N. Ahuja, "Split Aperture Imaging for High Dynamic Range," *Int'l J. Computer Vision*, vol. 58, no. 1, pp. 7-17, 2004.
- [2] A. Agrawal and R. Raskar, "Resolving Objects at Higher Resolution from a Single Motion-Blurred Image," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2007.
- [3] S. Baker and T. Kanade, "Limits on Super-Resolution and How to Break Them," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167-1183, Sept. 2002.
- [4] J. Bardsley, S. Jefferies, J. Nagy, and R. Plemmons, "Blind Iterative Restoration of Images with Spatially-Varying Blur," *Optics Express*, pp. 1767-1782, 2006.
- [5] B. Basclé, A. Blake, and A. Zisserman, "Motion Deblurring and Super-Resolution from an Image Sequence," *Proc. European Conf. Computer Vision*, pp. 573-582, 1996.
- [6] M. Ben-Ezra and S. Nayar, "Motion Deblurring Using Hybrid Imaging," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 657-664, June 2003.
- [7] M. Ben-Ezra and S. Nayar, "Motion-Based Motion Deblurring," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 689-698, June 2004.
- [8] P. Bhat, C.L. Zitnick, N. Snavely, A. Agarwala, M. Agrawala, B. Curless, M. Cohen, and S.B. Kang, "Using Photographs to Enhance Videos of a Static Scene," *Proc. Eurographics Symp. Rendering*, pp. 327-338, 2007.
- [9] M. Bigas, E. Cabruja, J. Forest, and J. Salvi, "Review of cmos Image Sensors," *Microelectronics J.*, vol. 37, no. 5, pp. 433-451, 2006.

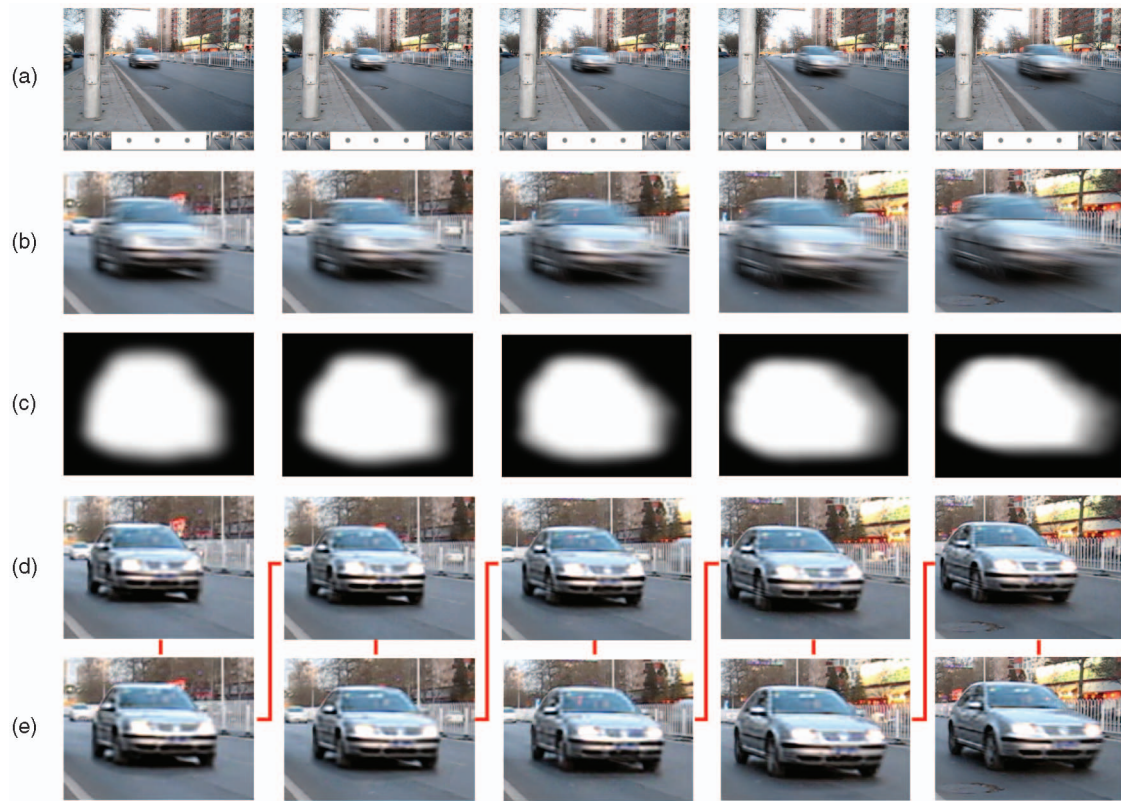


Fig. 19. Video deblurring in an outdoor scene. The moving object is a car driving toward the camera, which produces both translation and zoom-in blur effects. (a) Input video frames. (b) Close-ups of the moving car. (c) The extracted alpha mattes of the moving object. (d) The deblurred video frames using the first low-resolution frames as the reference frames. (e) The deblurred video frames using the third low-resolution frames as the reference frames. The final video consists of frames ordered as indicated by the red lines. By combining results from using different low-resolution frames as reference frames, we can increase the frame rate of the deblurred video.

- [10] S. Borman and R. Stevenson, "Super-Resolution from Image Sequences—a Review," *Proc. Midwest Symp. Circuits and Systems*, p. 374, 1998.
- [11] J. Chen and C.K. Tang, "Robust Dual Motion Deblurring," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008.
- [12] S. Cho, Y. Matsushita, and S. Lee, "Removing Non-Uniform Motion Blur from Images," *Proc. Int'l Conf. Computer Vision*, 2007.
- [13] Y. Chuang, A. Agarwala, B. Curless, D.H. Salesin, and R. Szeliski, "Matting of Complex Scenes," *ACM Trans. Graphics*, pp. 243-248, 2002.
- [14] Y. Chuang, B. Curless, D.H. Salesin, and R. Szeliski, "A Bayesian Approach to Digital Matting," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 264-271, 2001.
- [15] N. Dey, L. Blanc-Fraud, C. Zimmer, Z. Kam, P. Roux, J. Olivo-Marin, and J. Zerubia, "A Deconvolution Method for Confocal Microscopy with Total Variation Regularization," *Proc. IEEE Int'l Symp. Biomedical Imaging: Nano to Macro*, 2004.
- [16] M. Elad and A. Feuer, "Superresolution Restoration of an Image Sequence: Adaptive Filtering Approach," *IEEE Trans. Image Processing*, vol. 8, no. 3, pp. 387-395, 1999.
- [17] R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis, and W.T. Freeman, "Removing Camera Shake from a Single Photograph," *ACM Trans. Graphics*, vol. 25, no. 3, pp. 787-794, 2006.
- [18] D. Fish, A. Brinicombe, E. Pike, and J. Walker, "Blind Deconvolution by Means of the Richardson-Lucy Algorithm," *J. Optical Soc. Am.*, vol. 12, 1995.
- [19] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, second ed. Prentice Hall, 2002.
- [20] P.C. Hansen, J.G. Nagy, and D.P. OLeary, *Deblurring Images: Matrices, Spectra, and Filtering*. SIAM, 2006.
- [21] M. Irani and S. Peleg, "Improving Resolution by Image Registration," *Proc. Conf. Computer Vision, Graphics and Image Processing*, vol. 53, no. 3, pp. 231-239, 1991.
- [22] M. Irani and S. Peleg, "Motion Analysis for Image Enhancement: Resolution, Occlusion and Transparency," *J. Visual Comm. Image Representation*, vol. 4, pp. 324-335, 1993.
- [23] J. Jia, "Single Image Motion Deblurring Using Transparency," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2007.
- [24] N. Joshi, W. Matusik, and S. Avidan, "Natural Video Matting Using Camera Arrays," *ACM Trans. Graphics*, vol. 25, pp. 779-786, 2006.
- [25] N. Joshi, R. Szeliski, and D. Kriegman, "PSF Estimation Using Sharp Edge Prediction," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008.
- [26] T. Lauer, "Deconvolution with a Spatially-Variant PSF," *Astronomical Data Analysis II*, vol. 4847, pp. 167-173, 2002.
- [27] A. Levin, "Blind Motion Deblurring Using Image Statistics," *Proc. Conf. Neural Information Processing Systems*, pp. 841-848, 2006.
- [28] A. Levin, R. Fergus, F. Durand, and W.T. Freeman, "Image and Depth from a Conventional Camera with a Coded Aperture," *ACM Trans. Graphics*, 2007.
- [29] A. Levin, D. Lischinski, and Y. Weiss, "A Closed Form Solution to Natural Image Matting," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2006.
- [30] A. Levin, P. Sand, T.S. Cho, F. Durand, and W.T. Freeman, "Motion-Invariant Photography," *ACM Trans. Graphics*, 2008.
- [31] F. Li, J. Yu, and J. Chai, "A Hybrid Camera for Motion Deblurring and Depth Map Super-Resolution," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008.
- [32] B. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Proc. Imaging Understanding Workshop*, pp. 121-130, 1981.
- [33] L. Lucy, "An Iterative Technique for the Rectification of Observed Distributions," *Astronomical J.*, vol. 79, p. 745, 1974.
- [34] Y. Matsushita, E. Ofek, W. Ge, X. Tang, and H. Shum, "Full-Frame Video Stabilization with Motion Inpainting," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 1150-1163, July 2006.
- [35] M. McGuire, W. Matusik, H. Pfister, J.F. Hughes, and F. Durand, "Defocus Video Matting," *ACM Trans. Graphics*, vol. 24, pp. 567-576, 2005.

- [36] M. McGuire, W. Matusik, and W. Yezazunis, "Practical, Real-Time Studio Matting Using Dual Imagers," *Proc. Eurographics Symp. Rendering*, 2006.
- [37] A. Patti, M. Sezan, and A.M. Tekalp, "Superresolution Video Reconstruction with Arbitrary Sampling Lattices and Nonzero Aperture Time," *IEEE Trans. Image Processing*, vol. 6, no. 8, pp. 1064-1076, Aug. 1997.
- [38] R. Raskar, A. Agrawal, and J. Tumblin, "Coded Exposure Photography: Motion Deblurring Using Fluttered Shutter," *ACM Trans. Graphics*, vol. 25, no. 3, pp. 795-804, 2006.
- [39] A. Rav-Acha and S. Peleg, "Two Motion Blurred Images Are Better than One," *Pattern Recognition Letters*, vol. 26, pp. 311-317, 2005.
- [40] W. Richardson, "Bayesian-Based Iterative Method of Image Restoration," *J. Optical Soc. Am.*, vol. 62, no. 1, pp. 55-59, 1972.
- [41] Q. Shan, J. Jia, and A. Agarwala, "High-Quality Motion Deblurring from a Single Image," *ACM Trans. Graphics*, 2008.
- [42] Q. Shan, W. Xiong, and J. Jia, "Rotational Motion Deblurring of a Rigid Object from a Single Image," *Proc. Int'l Conf. Computer Vision*, 2007.
- [43] E. Shechtman, Y. Caspi, and M. Irani, "Space-Time Super-Resolution," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 531-544, Apr. 2005.
- [44] A. Smith and J.F. Blinn, "Blue Screen Matting," *Proc. ACM SIGGRAPH*, 1996.
- [45] F. Sroubek, G. Cristobal, and J. Flusser, "A Unified Approach to Superresolution and Multichannel Blind Deconvolution," *IEEE Trans. Image Processing*, vol. 16, no. 9, pp. 2322-2332, Sept. 2007.
- [46] J. Sun, J. Jia, C. Tang, and H. Shum, "Poisson Matting," *ACM Trans. Graphics*, 2004.
- [47] Y. Tai, H. Du, M. Brown, and S. Lin, "Image/Video Deblurring Using a Hybrid Camera," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008.
- [48] J. Wang, M. Agrawala, and M. Cohen, "Soft Scissors: An Interactive Tool for Realtime High Quality Matting," *ACM Trans. Graphics*, 2007.
- [49] J. Wang and M. Cohen, "An Iterative Optimization Approach for Unified Image Segmentation and Matting," *Proc. Int'l Conf. Computer Vision*, 2005.
- [50] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Wiley, 1949.
- [51] J. Xiao, H. Cheng, H. Sawhney, C. Rao, and M. Isnardi, "Bilateral Filtering-Based Optical Flow Estimation with Occlusion Detection," *Proc. European Conf. Computer Vision*, 2006.
- [52] L. Yuan, J. Sun, L. Quan, and H. Shum, "Image Deblurring with Blurred/Noisy Image Pairs," *ACM Trans. Graphics*, p. 1, 2007.
- [53] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum, "Progressive Inter-Scale and Intra-Scale Non-Blind Image Deconvolution," *ACM Trans. Graphics*, 2008.
- [54] G. Zhang, J. Jia, W. Xiong, T. Wong, P. Heng, and H. Bao, "Moving Object Extraction with a Hand-Held Camera," *Proc. Int'l Conf. Computer Vision*, 2007.
- [55] W. Zhao and H.S. Sawhney, "Is Super-Resolution with Optical Flow Feasible?" *Proc. European Conf. Computer Vision*, pp. 599-613, 2002.



Yu-Wing Tai received the BEng (first class honors) and MPhil degrees in computer science from the Hong Kong University of Science and Technology (HKUST) in 2005 and 2003, respectively, and the PhD degree from the National University of Singapore in June 2009. From September 2007 to June 2008, he worked as a full-time student intern at Microsoft Research Asia (MSRA). He joined the Korea Advanced Institute of Science and Technology (KAIST) University as an assistant professor in Fall 2009. His research interests include computer vision and image/video processing. He is a member of the IEEE.



student member of the IEEE

Hao Du received the BS and MS degrees from Fudan University, Shanghai, China, in 2005 and 2008, respectively. He is currently working toward the PhD degree in the Department of Computer Science and Engineering at the University of Washington. He was a visiting student at Microsoft Research Asia from 2007 to 2008. His recent research in the computer graphics and vision area includes computational photography and 3D reconstruction. He is a student member of the IEEE



vision, image processing, and computer graphics. He is a member of the IEEE.

Michael S. Brown received the BS and PhD degrees in computer science from the University of Kentucky in 1995 and 2001, respectively. He is currently an associate professor in the School of Computing at the National University of Singapore. He regularly serves on the program committees of the major computer vision conferences and has served as an area chair for IEEE Computer Vision and Pattern Recognition 2009. His research interests include computer vision, image processing, and computer graphics. He is a member of the IEEE.

Stephen Lin received the BSE degree from Princeton University and the PhD degree from the University of Michigan. He is currently a lead researcher in the Internet Graphics Group of Microsoft Research Asia. His research interests include computer vision and computer graphics. He has served as a program chair for the Pacific-Rim Symposium on Image and Video Technology 2009, a general chair for the IEEE Workshop on Color and Photometric Methods in Computer



Vision 2003, and as an area chair for the IEEE International Conference on Computer Vision 2007 and 2009. He is a member of the IEEE.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.