Intra-Frame Deblurring by Leveraging Inter-Frame Camera Motion

Haichao Zhang
Duke University, NC
hczhang1@gmail.com

Jianchao Yang
Adobe Research, CA
jiayang@adobe.com

Abstract

Camera motion introduces motion blur, degrading the quality of video. A video deblurring method is proposed based on two observations: (i) camera motion within *capture of each individual frame leads to motion blur; (ii) camera motion* between *frames yields inter-frame misalignment that can be exploited for blur removal. The proposed method effectively leverages the information distributed across multiple video frames due to camera motion, jointly estimating the motion between consecutive frames and blur within each frame. This joint analysis is crucial for achieving effective restoration by leveraging temporal information. Extensive experiments are carried out on synthetic data as well as real-world blurry videos. Comparisons with several state-of-the-art methods verify the effectiveness of the proposed method.*

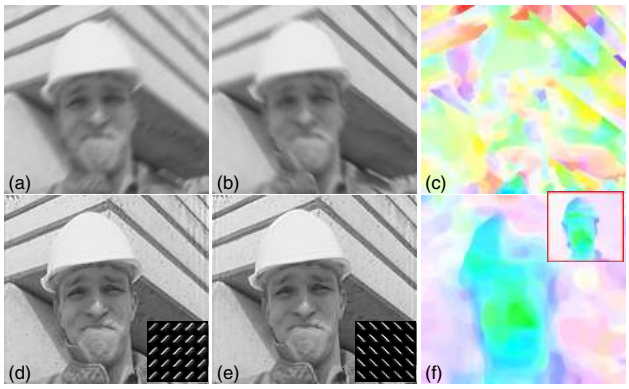

Figure 1. **Illustration of the proposed approach**. (a-b) two blurry video frames (c) optical flow estimated directly on the blurry images (d-f) the output of the proposed method: (d-e) deblurred video frames and blur kernels (f) optical flow (with color encoding [22]) jointly estimated with blur using the proposed method (for comparison, the ground-truth flow estimated on the original sharp images is shown on the top-right of (f)).

1. Introduction

Camera shake is a typical problem encountered during video capture, especially when using light-weight hand-held devices, such as cell-phones. Conventional video stabilization techniques compensate for the irregular inter-frame motion caused by camera shake, but such techniques do not help with camera-shake-induced intra-frame blur. Even worse, the problem of intra-frame motion blur often becomes more pronounced after stabilization, because of inconsistencies with the modified stabilization-induced motion path [16]. Blurry frames often cause a flickering effect when viewed in real time, leading to degraded quality in terms of visual perception. In this work, we aim to present a video deblurring method to handle this problem.

The task of video deblurring is constrained by several challenges. The blur introduced by the camera shake is typically spatially varying, and therefore estimating the blur from a single frame is challenging. Also, recovering a deblurred frame from only a single frame typically introduces ringing artifacts. Therefore, to achieve robust video deblurring, it is desirable to go beyond a single frame and exploit information contained in multiple frames. Because of blur variation due to camera motion, different frames may contain complementary information about the latent scene, which can be exploited for robust restoration, as show in Figure 1. However, direct application of multi-image-based deblurring methods typically does not always produce desirable results, mainly due to artifacts caused by inter-frame camera and/or object motion [5]. While this often requires accurate motion estimation to better exploit complementary information across frames, establishing the correspondence between different frames is challenging due to the inherent matching ambiguity caused by the motion blur.

We argue that camera motion is both an enemy and a friend for restoration, particularly in the context of video. On the one hand, camera motion occurring during the exposure period (within capture of a single frame) introduces blur, which smears detailed structures. On the other hand, due to inter-frame motion, different observations (frames) typically have different blurs, thus containing complementary information. This therefore provides more constraints for the ill-posed inverse problem, and hence is beneficial both in blur-kernel estimation and sharp-video-frame recov-

ery. The main contributions of this paper are as follows:

1. An approach for removing the camera-shake-induced blur in video is proposed, which can handle *non-uniform blur* with *non-rigid inter-frame motions*.
2. The complementary information distributed across frames is harnessed for blur removal, by estimating motion and blur jointly.
3. An effective algorithm is developed, taking advantage of recent progress in both blur estimation and flow estimation.

The remainder of the paper is organized as follows. We first briefly review related work in Section 2. We then present our video deblurring method in Section 3. Extensive experiments are conducted in Section 4 and results are compared with state-of-the-art methods. A discussion on the limitations of the proposed method is also provided in this section. We conclude the paper in Section 5.

2. Related Work

Single-/Multi-Image Blind Deblurring Blind deblurring techniques using a single image can be classified into maximum *a posterior* (MAP) estimators with different prior/penalty terms [20, 4, 28], and variational Bayesian (VB) based techniques, which attempt to make more thorough use of the underlying posterior distribution of the latent image [6, 13]. While initial emphasis has been on uniform deblurring [6, 20, 4, 28, 13], more recent efforts are on robust non-uniform deblurring [27, 8, 29, 32, 34], for which the blur kernel can vary across the image.

In many scenarios, we have access to multiple related observations, and it is advantageous to leverage the multiple observations for robust recovery. Most multi-image deblurring methods incorporate a two-observation-based 'cross-blur' penalty function, in addition to other standard regularizers [18, 3, 19, 35]. Although computationally efficient, inclusion of such a quadratic energy term does not always produce better kernel estimation [3, 35], leading to kernel estimates that are themselves blurry, producing ringing artifacts or loss of details in the deblurred image [35]. To mitigate some of these problems, a sparseness-promoting penalty on the blur kernel [3, 19, 35], or a strong sparseness-promoting prior on the latent image has been exploited [33, 31]. These methods assume a shared latent sharp image for all the observations, and therefore are only applicable to images of a static scene with small global rigid displacements. Image deblurring with the aid of a sharp reference image of the same scene in the presence of non-rigid displacements has been investigated recently in [7].

Video-Enhancement via Transferring Complementary information contained in a video clip has also been exploited for removing blur from video [16, 30, 9, 5, 23, 10].

Figure 2. **Blurry video observation model**. For video deblurring, to estimate a high-quality video frame $\mathbf{x}$ (indicated with red rectangle), we can exploit information from two sources: (i) direct observation $\mathbf{y}$, which is linked to $\mathbf{x}$ via the blur operator $\mathbf{H}$ (ii) consecutive frames ($\mathbf{y}_{-1}$ and $\mathbf{y}_1$), which are connected to $\mathbf{x}$ via a motion based transformation and blurring. Both sources of information can be described by the observation model (1).

Matsushita *et al.* [16] proposed to increase the sharpness of the blurry frame by transferring sharper image pixels from nearby frames, using interpolation. Another information-transfer-based video enhancement method is presented in [9], using projection onto convex sets (POCS). These two methods do not involve deconvolution, and enhance the blurry frame by information transfer only, relying heavily on the existence of sharp adjacent frames as the source of high-frequency information. Cho *et al.* [5] proposed a video deblurring method by transferring high-frequency information from sharp frames/regions by patch matching, which also relies on the existence of sharp frames (regions). Another related approach for extracting a snapshot from a video clip is proposed in [23].

Video-Enhancement based on Deblurring All of the aforementioned video-enhancement methods do not include an explicit deblurring process. Liu *et al.* [15] recently proposed a video supperresolution method, estimating the blur kernel and optical flow jointly. However, the blur kernel is assumed to be an isometric and invariant Gaussian kernel, and no motion blur is modeled. Lee *et al.* [10] take advantage of an accurate blur kernel estimation technique using blurry-sharp frame pairs. While this method tries to incorporate a deblurring process explicitly, it uses a uniform-blur assumption that is often inaccurate for real-world camera motion blur [27]. It also relies on the presence of a corresponding sharp frame for each blurry frame. An approach for generating a sharp panorama from a motion-blurred video is proposed in [14], with however a static-scene assumption.

3. Exploiting Inter-Frame Motion for Intra-Frame Deblurring

The task of video deblurring is challenging mainly due to the *coupled nature of motion and blur*. On the one hand, accurate estimation of the blur using multiple frames

requires reliable motion estimation, to compensate for the inter-frame camera motion as well as object motion. On the other hand, reliable motion estimation requires robust blur removal for avoiding matching ambiguities caused by smeared structures due to blur [18]. We aim to tackle the problem of motion and blur estimation *jointly*, for removing camera-motion blur in video, in the presence of non-rigid inter-frame motion. Without loss of generality, we omit the specific temporal location of the frame under consideration and assume it is at the canonical position $t = 0$ in the sequel, i.e., denoting the latent high quality frame at a specified location as $\mathbf{x}$ and the set of low quality observations around that location as $\{\mathbf{y}_l\}_{l=-T}^{T}$ (T frames before, and T frames after).

3.1. The Model

Video records a spatial-temporal volume of a time-varying scene, with each video frame a sample along the temporal axis. In the presence of camera shake, the captured video frame is blurred due to the integration of motion over time (see Figure 2), leading to the following observation model for the l-th frame ($l \in \{-T, \cdots, 0, \cdots, T\}$)

$$\mathbf{y}_l = \mathbf{H}_l \mathbf{x}_l + \mathbf{n}_l, \tag{1}$$

where $\mathbf{x}_l$ is the latent sharp frame corresponding to the blurry video frame $\mathbf{y}_l$, $\mathbf{H}_l$ is the blurring operator, and $\mathbf{n}_l$ is an additive noise term (or model mis-match residual). Constructing $\mathbf{H}_l$ as the convolution matrix of a spatially invariant blur kernel is too restrictive for modeling camera motion blur [10]. To model the general camera-motion-induced blur, we adopt the recent Projective Motion Path (PMP) model [25], i.e., representing the blurring operator as $\mathbf{H}_l \triangleq \sum_{j=1}^{J} w_{jl} \mathbf{P}_j$, where $\{\mathbf{P}_j\}$ denotes a set of projection operators induced by discretizing the motion space (e.g., a certain range of translation and rotation) [25], and $w_{jl} \geq 0$ denotes the contribution of $\mathbf{P}_j$ to $\mathbf{H}_l$.

A straightforward approach for recovering $\mathbf{x}$ ($l = 0$) is based on the observed frame $\mathbf{y}$ and (1) directly, by solving a standard single-image-based blind deblurring problem:

$$\min_{\mathbf{w} \geq 0, \mathbf{x}} \frac{1}{\lambda} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + R(\mathbf{x}, \mathbf{w}, \lambda), \tag{2}$$

where $R(\cdot)$ is a regularization term, employed to alleviate the ill-posedness of the inverse problem; the regularizer is either separable [20, 29] or coupled over the unknown variables [33].[1] This approach, however, relies purely on a single observed frame, not exploiting the complementary information distributed in the observed video cube due to motion (across multiple frames). To exploit this complementary information for improved estimation, we connect two latent sharp frames as $\mathbf{x}_l \approx \mathbf{F}_l \mathbf{x}$, where $\mathbf{F}_l$ is the non-rigid spatial transformation operator induced by the flow field that relates two video frames. While it is possible to establish this connection for each pair of frames, and recover all the latent sharp frames jointly, we concentrate on a specified frame $\mathbf{x}$ without loss of generality. We solve the following problem incorporating the temporally relevant frames

$$\begin{aligned} \min_{\mathbf{x}, \{\mathbf{w}_l, \lambda_l\} \geq 0, \{\mathbf{F}_l\}} & \underbrace{\frac{1}{\lambda} \|\mathbf{y} - \sum_j w_j \mathbf{P}_j \mathbf{x}\|_2^2}_{\text{data term}} \\ & + \sum_{l \neq 0} \underbrace{\frac{1}{\lambda_l} \|\mathbf{y}_l - \sum_j w_{jl} \mathbf{P}_j \mathbf{F}_l \mathbf{x}\|_2^2}_{\text{temporal term}} + R(\mathbf{x}, \{\mathbf{w}_l, \lambda_l, \mathbf{F}_l\}) \end{aligned} \tag{3}$$

where $\mathbf{w}_l = [w_{1l}, w_{2l}, \cdots]^T$. Each term in (3) is as follows:

- The first term in (3) is a standard data-fidelity term, based directly on the observation model, and is referred to as *data* term in the sequel.
- The second term is a *motion-aware* regularizer, which exploits the motion-induced complementary information across multiple temporal frames, by relating $\mathbf{x}$ to $\mathbf{x}_l$ via $\mathbf{F}_l$, for further improving the estimation; it is referred to as a *temporal* term.
- The temporal term is helpful only if $\mathbf{y}_l$ contains additional information compared to $\mathbf{y}$ in the data term, meaning $\sum_j w_{jl} \mathbf{P}_j \mathbf{F}_l$ is different from $\mathbf{H} \triangleq \sum_j w_j \mathbf{P}_j$, which is satisfied in general due to the random nature of camera shake.
- The choice of the generic regularization term $R(\cdot)$ is detailed later.

The proposed joint approach (3) has several advantages:

- This method can estimate non-uniform blur, non-rigid motion, and sharp video frames from the blurry video itself, without introducing simplified assumptions (e.g., uniform blur [19, 33, 10], rigid motion [16]) or hardware assistance [2, 24].
- The proposed method can effectively fuse information from multiple frames, and does not rely on the existence of sharp video frames to achieve enhancement. It is applicable to the scenario where all the video frames are blurred, which is a significant difference from most previous methods [5, 9, 10]. However, the existence of sharp frames will help to further improve the restoration quality.
- While using video deblurring as a motivational example, the proposed method is generic and applicable to other scenarios involving multiple *non-uniformly* blurred observations with possibly *non-rigid* inter-frame motions.

However, (3) is difficult to solve, as we want to infer the set of blur operators, flow operators as well as the latent

[1] Here the λ in $R(\cdot)$ can either be regarded as a parameter [13] or a joint variable [33].

sharp frame from only a set of blurry frames. To reduce the complexity of the problem and increase the robustness of the algorithm, we approach the problem in two phases: in the *blind phase*, we recover the spatially-variant blur kernels and motion fields jointly, using an image penalty that has a strong ability to promote sparsity, for enhancing kernel estimation and reduce the ambiguity in flow estimation. In the second, *non-blind phase*, sharpened video frames are recovered using the estimated blur kernels and motion fields from the first phase, with a natural-image prior [11]. A similar strategy has proven to be effective for blind image deblurring [6, 29, 33]. We outline our optimization in the following section. More complete details can be found in supplementary material.

3.2. Blind Phase

In this phase, we work in the derivative domain of images, for simplicity of modeling and effectiveness of blur kernel estimation [13]. We use $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ to denote the (vectorized) derivatives of $\mathbf{x}$ and $\mathbf{y}$ respectively.[2] Flow is still estimated in the pixel domain. The blur kernel is estimated by solving a regularized regression problem based on (3). An additional advantage of working in the derivative domain is that the simple quadratic fidelity term can be used, without resorting to more-advanced robust functions, simplifying the optimization process.[3] We therefore derive from (3) two subproblems as follows, and solve them iteratively to estimate blur and flow:

$$\texttt{Blur} \quad \min_{\{\mathbf{w}_l,\lambda_l\}\geq 0,\bar{\mathbf{x}}} \sum_l \frac{1}{\lambda_l}\|\bar{\mathbf{y}}_l - \mathbf{H}_l\mathbf{F}_l\bar{\mathbf{x}}\|_2^2 + R_1(\bar{\mathbf{x}},\{\mathbf{w}_l,\lambda_l\}) \quad (4)$$

$$\texttt{Flow} \quad \min_{\mathbf{F}_l} \frac{1}{\lambda_l}\|\mathbf{y}_l - \mathbf{H}_l\mathbf{F}_l\mathbf{x}\|_2^2 + R_2(\mathbf{F}_l), \forall l \in \{-T,\cdots,T\} \quad (5)$$

R_1 is a penalty function coupled over $\bar{\mathbf{x}}$, $\mathbf{w}_l$ and λ_l [33], employing a sparsity-promoting ability with respect to $\bar{x}_i$, modulated by $\mathbf{w}_l$ and λ_l[4]

$$R_1(\bar{\mathbf{x}},\{\mathbf{w}_l,\lambda_l\}) \triangleq \min_{\boldsymbol{\gamma}\geq 0} \sum_{l,i}\left[\frac{\bar{x}_i^2}{\gamma_i} + \log(\lambda_l + \gamma_i\|\mathbf{h}_{il}\|_2^2)\right],$$

where $\mathbf{h}_{il} = \mathbf{B}_{il}\mathbf{w}_l$ is the local blur kernel at site i, $\mathbf{B}_{il} = [\mathbf{P}_1,\cdots,\mathbf{P}_i,\cdots]\mathbf{F}_l\mathbf{e}_i$, and $\mathbf{e}_i$ denotes the vectorized delta image with all zero except at site i. R_2 is a penalty term for $\mathbf{F}_l$ that promotes sparseness as detailed in the flow update.

Blur Update From (4), $\mathbf{H}_l \triangleq \sum_{j=1}^J w_{jl}\mathbf{P}_j$ and the definition of R, we derive the following cost function for blur ($\mathbf{w}_l$) update of each frame l [32]:

$$\min_{\mathbf{w}_l\geq 0} \|\bar{\mathbf{y}}_l - \mathbf{D}_l\mathbf{w}_l\|_2^2 + R_1(\bar{\mathbf{x}},\{\mathbf{w}_l,\lambda_l\}) \quad (6)$$

with $\mathbf{D}_l = [\mathbf{P}_1,\cdots,\mathbf{P}_j,\cdots]\mathbf{F}_l\bar{\mathbf{x}}$ denoting the dictionary constructed by projectively transforming $\mathbf{F}_l\bar{\mathbf{x}}$ using a set of transformation operators. (7) can be solved by bounding the non-convex R_1 with the introduction of laten variables z_{il}[5]

$$\min_{\mathbf{w}_l\geq 0} \|\bar{\mathbf{y}}_l - \mathbf{D}_l\mathbf{w}_l\|_2^2 + \mathbf{w}_l^T(\sum_i z_{il}\mathbf{B}_{il}^T\mathbf{B}_{il})\mathbf{w}_l \quad (7)$$

reducing (6) to alternating between solving standard non-negative quadratic programming problem (7), and updating the latent variables in closed-forms [32]:

$$z_{il} \leftarrow (\lambda_l^{-1}\|\mathbf{h}_{il}\|_2^2 + \gamma_i^{-1})^{-1},\ \gamma_i \leftarrow \bar{x}_i^2 + \sum_l z_{il}/(2T+1).$$

The latent image (derivatives) $\bar{\mathbf{x}}$ is used for constructing the dictionary $\mathbf{D}_l$ and plays the role of an auxiliary variable for kernel estimation in (7). With other variables fixed, $\bar{\mathbf{x}}$ is updated by solving the following regularized-least-square problem derived from (4):

$$\min_{\bar{\mathbf{x}}} \frac{1}{\lambda}\|\bar{\mathbf{y}} - \mathbf{H}\bar{\mathbf{x}}\|_2^2 + \sum_{l\neq 0}\frac{1}{\lambda_l}\|\bar{\mathbf{y}}_l - \mathbf{H}_l\mathbf{F}_l\bar{\mathbf{x}}\|_2^2 + \bar{\mathbf{x}}^T\boldsymbol{\Gamma}^{-1}\bar{\mathbf{x}}, \quad (8)$$

which has closed-form solution

$$\bar{\mathbf{x}} = \Big(\sum_l \frac{1}{\lambda_l}\mathbf{F}_l^T\mathbf{H}_l^T\mathbf{H}_l\mathbf{F}_l + \boldsymbol{\Gamma}^{-1}\Big)^{-1}\sum_l \frac{1}{\lambda_l}\mathbf{F}_l^T\mathbf{H}_l^T\bar{\mathbf{y}}_l, \quad (9)$$

where $\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$. In practice, (9) is calculated via the conjugate gradient (CG) method.

Flow Update After updating the blur ($\mathbf{w}_l$), we get a deblurred image $\hat{\mathbf{x}}_l$ of $\mathbf{y}_l$ using a standard non-blind deblurring method adapted to non-uniform deblurring [11]. The operator $\mathbf{F}_L$ is then updated by solving (5) via

$$\min_{\mathbf{u},\mathbf{v}} \frac{1}{\lambda_l}\|\hat{\mathbf{x}}_l - \mathbf{F}_l\mathbf{x}\|_2^2 + R_2(\mathbf{F}_l). \quad (10)$$

We implement $\mathbf{F}_l$ with an optical-flow-driven motion field, represented locally in terms of the horizontal and vertical motion vectors as $\mathbf{F}_l\mathbf{x} = \mathbf{x} + \frac{\partial\mathbf{x}}{\partial u}\mathbf{u}_l + \frac{\partial\mathbf{x}}{\partial v}\mathbf{v}_l$. Substituting this into (10) and using the Charbonnier-penalty ($R_c(\mathbf{x}) = \sum_i\sqrt{x_i^2+\epsilon^2}$) for the motion vectors, we have

$$\min_{\mathbf{u},\mathbf{v}} \|\hat{\mathbf{x}}_l - \mathbf{x} - \frac{\partial\mathbf{x}}{\partial u}\mathbf{u}_l - \frac{\partial\mathbf{x}}{\partial v}\mathbf{v}_l\|_2^2 + \gamma R_c(\mathbf{u}) + \gamma R_c(\mathbf{v}), \quad (11)$$

which can be solved effectively using the iteratively reweighted least squares (IRLS) method [22]. Note that λ_l in (10) has been absorbed into γ in (11).

Noise Level Update The effective noise level reflects the accuracy of the model estimation. By retaining only the terms relevant to λ_l in (4), we obtain the cost function:

$$\min_{\lambda_l\geq 0} \frac{1}{\lambda_l}\|\bar{\mathbf{y}}_l - \mathbf{H}_l\mathbf{F}_l\bar{\mathbf{x}}\|_2^2 + \sum_i \log(\lambda_l + \gamma_i\|\mathbf{h}_{il}\|_2^2) \quad (12)$$

[2] Filter set $\{\mathbf{q}_1 = [-1,1], \mathbf{q}_2 = \mathbf{q}_1^T\}$ is used in our implementation.

[3] However, robustness term is preferred in the non-blind phase, as detailed in Section 3.3.

[4] This property is presented in Theorem 2 from [33].

[5] Please refer to the supplementary file for the derivation.

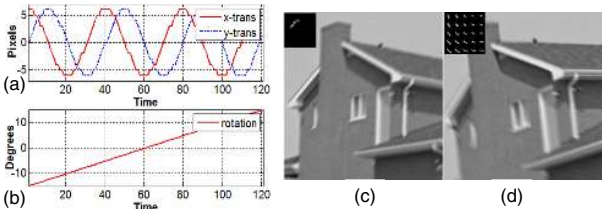

Figure 3. **Synthetic blurry video generation**. (a) Translation parameters plot. (b) Rotation parameter plot. (c) A blurry video frame generated from the `house` image driven by the motion parameters in (a), with uniform-blur. (d) A blurry frame with non-uniform blur, i.e., both the translation and rotation are used for producing the video. The corresponding blur kernels are shown on the top-left of (c) and (d) respectively.

which leads to the following update rule for λ by using the conjugate bounding [32]:

$$\lambda_l \leftarrow (\|\bar{\mathbf{y}}_l - \mathbf{H}_l\mathbf{F}_l\bar{\mathbf{x}}\|_2^2 + \eta_l)/n, \tag{13}$$

where n is the dimensionality of $\bar{\mathbf{y}}_l$ and $\eta_l \leftarrow \sum_i z_{il}\|\mathbf{h}_{il}\|_2^2$.

3.3. Non-Blind Phase

There are several reasons for requiring a robust non-blind video deblurring approach. First, different frames capture different observations of the scene, due to a change of viewpoint caused by camera motion. Therefore, some parts of the scene that may be visible in one frame might become invisible in another, due to occlusion. Secondly, the scene may contain dynamic (time-varying) objects, which will have a different appearance in different frames. Naive application of multi-image non-blind deblurring might produce results with severe ringing artifacts due to these factors (see Figure 4). Here we enhance the original model in (3) with a robust function and a natural image prior [11]:

$$\min_{\mathbf{x}} \frac{\rho(\mathbf{y}-\mathbf{H}\mathbf{x})}{\lambda} + \sum_{l\neq 0}\frac{\rho(\mathbf{M}_l(\mathbf{y}_l-\mathbf{H}_l\mathbf{F}_l\mathbf{x}))}{\lambda_l} + \tau\sum_j \|\mathbf{q}_j * \mathbf{x}\|^{0.8} \tag{14}$$

where $\rho(\mathbf{x}) = \sum_i \log(1+0.5(x_i/\delta)^2)$ is the robust Lorentzian function and $\{\mathbf{q}_j\}$ is a set of derivative filters. The expression in (14) can be solved again with the iteratively reweighted least squares method, by iteratively updating the following two steps:

$$\mathbf{x} = \Big(\sum_l \frac{\mathbf{F}_l^T\mathbf{H}_l^T\mathbf{W}_l\mathbf{H}_l\mathbf{F}_l}{\lambda_l} + \tau\sum_j \mathbf{Q}_j^T\mathbf{V}_j\mathbf{Q}_j\Big)^{-1}\sum_l \frac{\mathbf{F}_l^T\mathbf{H}_l^T\mathbf{W}_l\mathbf{y}_l}{\lambda_l}$$

$$\mathbf{W}_l = \mathbf{M}_l\mathrm{diag}\big((\mathbf{e}_l^2+2\delta^2)^{-0.5}\big), \quad \mathbf{V}_j = \mathrm{diag}\big(|\mathbf{Q}_j\mathbf{x}|^{-1.2}\big),$$

where $\mathbf{Q}_j\mathbf{x} = \mathbf{q}_j * \mathbf{x}$ and $\mathbf{e}_l = \mathbf{y}_l - \mathbf{H}_l\mathbf{F}_l\mathbf{x}$. $\mathbf{M}_l = \mathrm{diag}\big(\exp(-\frac{|\mathbf{H}\mathbf{y}_l-\mathbf{F}_l^T\mathbf{H}_l\mathbf{y}|^2}{0.01^2})\big)$ is a diagonal matrix with diagonal elements indicating the visibility of the corresponding observation with respect to the current frame. We fix the number of iterations to 3 for IRLS in our experiments.

3.4. Implementation Details

For the blind phase, we used a standard multi-scale estimation technique, first performing estimation on a low-resolution image, and using the solution to initialize the solution at the next higher resolution level using bilinear interpolation [6, 22]. For each level of resolution, we iteratively update blur, flow and noise level for 5 times. The length of the temporal window is typically set as 3, meaning the two frames closest to the specified one are used jointly with the current frame for restoration. Local convolutional approximation is used for reducing the computational expense of evaluating $\mathbf{H}_l\mathbf{x}$ [8]. We set $\epsilon = 0.001$, $\gamma = 3$ for flow estimation in (11) and $\delta = 0.001$, $\tau = 0.2$ for robust non-blind deblurring in (14). Pixel values are scaled to $[0, 1]$.

4. Experimental Results

We conduct several experiments using both synthetic data and real-world blurry videos. Experimental results are compared with several state-of-the-art methods, including the single-image-based uniform deblurring method by Xu *et al.* [28], the non-uniform deblurring algorithm [29],[6] the multi-image-based deblurring method by Sroubek *et al.* [19] (with motion correction using the optical-flow from the blurry frames),[7] and the recent video deblurring method by Cho *et al.* [5].

4.1. Blurry Video of Static Scenes

The synthetic blurry video is generated based on the observation model (1), by transforming a sharp image according to the time-varying motion parameter and integrating over certain time period. Additive Gaussian noise with standard derivation of 2 (pixel intensity range $[0, 255]$) is added to each video frame. The evolution curves of the used motion parameters are shown in Figure 3(a)-(c). We generate two types of blurry videos driven by the motion parameters: (i) **Uniform Blur**: the video is generated using the translational motion only (Figure 3(a)); (ii) **Non-Uniform Blur**: the video is generated using both translation and rotation (Figure 3(a) + (b)). We average over every 9 consecutive frames to simulate the accumulation process during the exposure period. Example blur kernels are also shown in Figure 3 (d)-(e). The method of Cho *et al.* [5] is not compared for this experiment, as it is only applicable to videos with some sharp frames.

The experimental results are shown in Figure 5 in terms of Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) measured with the same scheme as in [13]. As observed, for the uniform-blur case, the single-frame-based method [28] can improve the video quality.

[6] http://www.cse.cuhk.edu.hk/~leojia/research.html
[7] http://zoi.utia.cas.cz/files/fastMBD.zip

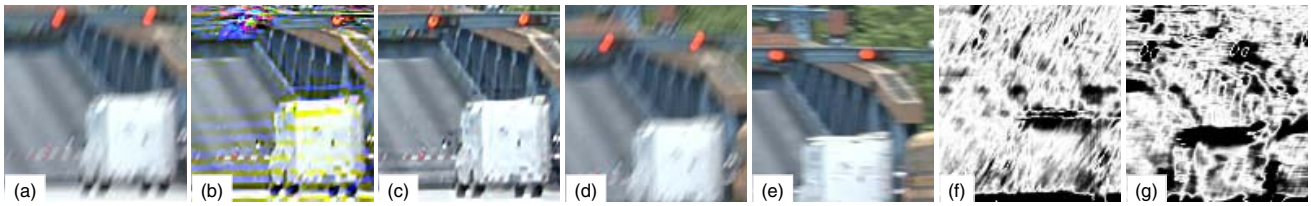

Figure 4. **Robust non-blind phase**. (a) A blurry frame to be deblurred. Deblurred frame via (b) non-robust method (without using $\mathbf{M}_l$ and set $\rho(\cdot)$ as ℓ_2-norm), and (c) proposed robust method. (c)-(d) Two frames around the specified frame in (a), which are themselves blurry. (f)-(g) Robust masks ($\mathbf{M}_l$) for (d) and (e).

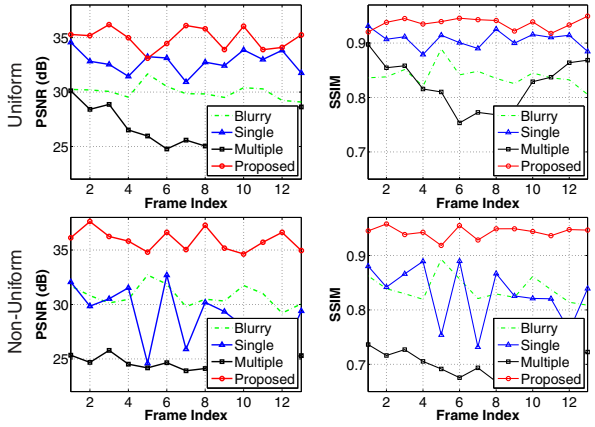

Figure 5. **Video deblurring results on `house` sequence**. PSNR and SSIM plots for uniform (top) and non-uniform (bottom) cases.

However, in the presence of non-uniform blur, the single-frame-based non-uniform deblurring method [29] does not always possible to improve the quality of the video. The multi-image-based method, with pre-motion correction, does not perform well, mainly due to the inaccuracy in motion estimation using the blurry frames.

4.2. Blurry Video of Non-Static Scenes

In the second experiment, we use a standard video sequence containing object motion (`Foreman`) for evaluating the effectiveness of the proposed method on video deblurring in the presence of both camera motion and object motion. The camera-motion blur is produced using a spatially and temporally invariant blur kernel (motion blur kernel with size 7 and motion direction $45°$). As the kernel is stationary across all the frames, it is easier to observe *the gain from inter-frame motion* under this setting.

The PSNR and SSIM plots of several methods on this test sequence are shown in Figure 6. As observed, the proposed method performs better than other methods in terms of both PSNR and SSIM. We also present deblurring results of different methods in Figure 6. As the blur kernel is a simple one, the single-image-based method [28] achieves fairly good restoration quality. The multi-image method [19] introduces some artifacts mainly because of the object motion, performing worse both visually and quantitatively. The proposed method, by exploiting complementary information introduced by the inter-frame camera motion, produces high-quality deblurred frames with more fine details than other methods. Moreover, the flow estimation from the proposed method is also more faithful to the ground-truth flow than that estimated from the sharp frames directly (see Figure 1 for an illustration). This result indicates the benefits of exploiting information contained in multiple frames, and the effectiveness of the proposed method in handling both inter-frame camera motion and objet motion.

4.3. Video Deblurring on Real-World Videos

Experiments on real-world videos from the literature are conducted in this section. We first use the `bridge` sequence from Cho *et al.* [5], and compare our results with the results from the state-of-the-art video deblurring method by Cho *et al.* [29, 5], as shown in Figure 7. Deblurring results using a single-image-based deblurring method [29] are also presented in Figure 7 for comparison. The single-image-based method [29] does not produce good image restoration on this video frame, and the deblurred image has severe ringing artifacts. The method of Cho *et al.* [5] produces visually sharp results and avoids ringing artifacts, by using a patch-based synthesis technique (and shock filtering) rather than deconvolution. However, note that the result of Cho *et al.* [5] has a different type of artifact, manifested as *structure distortion*, which is most apparent in structured regions (for example, the *person* in the frame and the *light pole* on the right part of the frame). One possible reason for this is that the local patch-matching strategy might introduce erroneous matches in the presence of severe blur and thus cause artifacts in the subsequent patch-based synthesis step. Therefore, while the individual deblurred frame is sharp, it might not be consistent with the observed image, as the reconstruction constraint is not enforced explicitly. For our approach, in contrast, the matching between frames is established using a global optimization framework, with the data fidelity term explicitly incorporated, and thus it does not suffer from the structure distortion problem of Cho *et al.* [5], as seen from Figure 7. Moreover, it can recover more

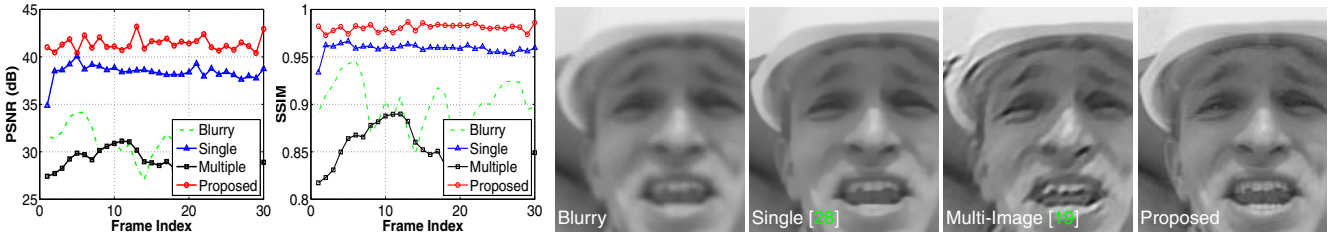

Figure 6. **Video deblurring results on the `Foreman` sequence**. Top: PSNR and SSIM plots. Bottom: blurry and deblurred frames.

Figure 7. **Video deblurring results on the `bridge` sequence** from Cho *et al.* [5]. The result of Cho *et al.* [5] has fewer ringing artifacts than the single image-based method [28], but has some structure distortions (e.g., the *light pole*). The proposed method does not have this problem while can recover more fine details than other compared methods. Please see the supplementary file for more comparisons.

fine details than the other methods considered (e.g., see the enlarged *window* region), which validates the effectiveness of the proposed approach in *exploiting motion for motion deblurring*. We also show the blur kernel patterns estimated by our method for the blurry frame in Figure 7. It is observed that the recovered kernel pattern resembles the blurry light sticks in the frame, implying the plausibility of the recovered blur kernels.

Results on two additional video clips [5, 23] are shown in Figure 8. On the `car` sequence, the proposed method produces deblurred video frames better than that of the single-frame method using [29], and is comparable to that from the video deblurring method of Cho *et al.* [5]. On the `mural` sequence, the proposed method outperforms all other compared methods, demonstrating the effectiveness of the proposed method.

4.4. Blur-Aware Motion Estimation

Although not the main focus of this paper, the proposed method can obtain blur-aware flow estimation while performing video deblurring. A recent method for flow estimation in the presence of motion blur is proposed by Travis *et al.* [17].[8] While relevant to our approach, the method of Travis *et al.* [17] focuses on flow estimation only but not on the task of deblurring, thus it is complementary to our approach. Furthermore, although it might be possible to integrate this method into an alternating estimation framework similar to ours, the fact that it requires two extra frames for parameterizing the blur in order to estimate the blur-aware flow between any two frames makes this method less general. The proposed method, in contrast, can obtain the flow directly from two blurry frames, requiring no additional auxiliary frames.

To make a comparison, we use a dataset from Travis *et al.* [17] and construct a two-step approach by using the flow estimated from the method of Travis *et al.* [17] followed with a standard multi-image deblurring method [19]. The results are shown in Figure 9. It is noted that the two-step approach using the blur-aware flow estimation [17] can generate improved results compared to the baseline method using the flow estimated from the blurry frames directly. However, the result from this two-step method still has some ringing artifacts, as well as some distorted structures (e.g., the character '`J`'). While these results may be improved by iteratively performing the two steps multiple times, the distorted structures are not corrected. This indicates that sequentially concatenating two existing approaches cannot achieve high-quality deblurring results, compared to the proposed joint estimation approach. In terms of flow estimation, both the method from Travis *et al.* [17] and the proposed method provide improvements over the baseline method, as indicated by the improved restoration results.

4.5. Discussions and Limitations

For the proposed method, although no special treatment is incorporated, the temporal consistency is implicitly maintained due to the sharing of the latent image through the *motion-aware temporal term*. For smooth regions, as there

[8] http://pages.cs.wisc.edu/~lizhang/projects/blurflow

Figure 8. **Video deblurring results** on the `car` (top) and `mural` (bottom) sequences. The deblurred frame from the proposed method has more fine details and fewer ringing artifacts compared to the results from single frame image based deblurring method such as [29] and [32], and is sharper than the result from [5].

Figure 9. **Flow estimation in presence of motion blur**. Left to right: a blurry frame, deblurred image using standard optical flow method followed with the multi-image deblurring method of [19] (baseline), deblurred image by using [17] for optical flow estimation followed with the multi-image deblurring method of [19] (two-step), deblurred image by the proposed method. Flow estimations are shown on the top right of each deblurred image.

is no salient structural information for robustly establishing the correspondence, the flow estimation might not be accurate. This, however, will not cause problems for sharp-frame recovery, as most patches within the smooth region are similar to each other, and the inaccurate matches are still valid for restoration.

There are several limitations for the proposed method. First, our approach for flow estimation is based on the deblurred frames using the current blur estimation. While simplifying the optimization, a better approach would be to solve (5) directly by embedding the blur operator into the IRLS-based flow updating process, similar to [15]. Secondly, the proposed model can only remove the blur caused by camera motion and does not handle blur caused by object-motion. However, this can be achieved by temporal deconvolution after temporal interpolation using our flow estimation [21, 26]. Thirdly, the current model does not model the scene depth explicitly, and thus cannot handle scenes with large depth variations. Another scenario challenging to all video deblurring methods, including ours, is when all the frames are severely blurred. In this case, it might be advantageous to combine our method with an example-based approach [5, 7].

5. Conclusion

We have proposed a new approach for removing camera-shake-induced motion blur in video, taking advantage of complementary information distributed across frames due to camera motion. The proposed method jointly estimates the camera-motion-induced intra-frame blur and inter-frame motion, producing high-quality video deblurring results. The inter-frame motion allows a given region in the scene to be viewed from the perspective of different intra-frame blurring, providing important complementary information. Future work includes investigating the 'optimal' camera motion for high quality imaging in certain scenarios (i.e., *purposely* introducing inter-frame camera motion), in a similar spirit to designing patterns for coded aperture [12] and coded exposure [1]. Moreover, the recovered inter-frame motion can be leveraged to achieve temporal super-resolution and object motion deblurring. Extending the proposed method for achieving joint video stablization and deblurring is another interesting research direction.

Acknowledgement We would like to thank the reviewers and the area chair for their valuable comments. We thank Dr. Lawrence Carin for proofreading the paper.

References

[1] A. K. Agrawal and Y. Xu. Coded exposure deblurring: Optimized codes for PSF estimation and invertibility. In *CVPR*, 2009. 8

[2] M. Ben-Ezra and S. Nayar. Motion-based Motion Deblurring. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(6):689–698, Jun 2004. 3

[3] J. Chen, L. Yuan, C.-K. Tang, and L. Quan. Robust dual motion deblurring. In *CVPR*, 2008. 2

[4] S. Cho and S. Lee. Fast motion deblurring. In *SIGGRAPH ASIA*, 2009. 2

[5] S. Cho, J. Wang, and S. Lee. Video deblurring for hand-held cameras using patch-based synthesis. *ACM Trans. Graph.*, 31(4):64, 2012. 1, 2, 3, 5, 6, 7, 8

[6] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. In *SIGGRAPH*, 2006. 2, 4, 5

[7] Y. HaCohen, E. Shechtman, and D. Lischinski. Deblurring by example using dense correspondence. In *ICCV*, 2013. 2, 8

[8] M. Hirsch, C. J. Schuler, S. Harmeling, and B. Schölkopf. Fast removal of non-uniform camera shake. In *ICCV*, 2011. 2, 5

[9] Y. Huang and N. Fan. Inter-frame information transfer via projection onto convex set for video deblurring. *J. Sel. Topics Signal Processing*, 5(2):275–284, 2011. 2, 3

[10] D.-B. Lee, S.-C. Jeong, Y.-G. Lee, and B. C. Song. Video deblurring algorithm using accurate blur kernel estimation and residual deconvolution based on a blurred-unblurred frame pair. *IEEE Trans. Image Process.*, 22(3):926–940, 2013. 2, 3

[11] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Deconvolution using natural image priors. Technical report, MIT, 2007. 4, 5

[12] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.*, 26(3):70, 2007. 8

[13] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Efficient marginal likelihood optimization in blind deconvolution. In *CVPR*, 2011. 2, 3, 4, 5

[14] Y. Li, S. B. Kang, N. Joshi, S. M. Seitz, and D. P. Huttenlocher. Generating sharp panoramas from motion-blurred videos. In *CVPR*, 2010. 2

[15] C. Liu and D. Sun. On Bayesian adaptive video super resolution. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(2):346–360, 2014. 2, 8

[16] Y. Matsushita, E. Ofek, X. Tang, and H.-Y. Shum. Full-frame video stabilization. In *CVPR*, 2005. 1, 2, 3

[17] T. Portz, L. Zhang, and H. Jiang. Optical flow in the presence of spatially-varying motion blur. In *CVPR*, 2012. 7, 8

[18] A. Rav-Acha and S. Peleg. Two motion blurred images are better than one. *Pattern Recognition Letters*, 26:311–317, 2005. 2, 3

[19] F. Šroubek and P. Milanfar. Robust multichannel blind deconvolution via fast alternating minimization. *IEEE Trans. on Image Process.*, 21(4):1687–1700, 2012. 2, 3, 5, 6, 7, 8

[20] Q. Shan, J. Jia, and A. Agarwala. High-quality motion deblurring from a single image. In *SIGGRAPH*, 2008. 2, 3

[21] E. Shechtman, Y. Caspi, and M. Irani. Space-time super-resolution. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(4):531–545, 2005. 8

[22] D. Sun, S. Roth, and M. J. Black. Secrets of optical flow estimation and their principles. In *CVPR*, 2010. 1, 4, 5

[23] K. Sunkavalli, N. Joshi, S. B. Kang, M. F. Cohen, and H. Pfister. Video snapshots: Creating high-quality images from video clips. *IEEE Trans. Vis. Comput. Graph.*, 18(11):1868–1879, 2012. 2, 7

[24] Y.-W. Tai, H. Du, M. S. Brown, and S. Lin. Correction of spatially varying image and video motion blur using a hybrid camera. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(6):1012–1028, 2010. 3

[25] Y.-W. Tai, P. Tan, and M. S. Brown. Richardson-Lucy deblurring for scenes under a projective motion path. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(8):1603–1618, 2011. 3

[26] H. Takeda and P. Milanfar. Removing motion blur with space-time processing. *IEEE Tran. on Image Process.*, 20(10):2990–3000, 2011. 8

[27] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce. Non-uniform deblurring for shaken images. In *CVPR*, 2010. 2

[28] L. Xu and J. Jia. Two-phase kernel estimation for robust motion deblurring. In *ECCV*, 2010. 2, 5, 6, 7

[29] L. Xu, S. Zheng, and J. Jia. Unnatural L0 sparse representation for natural image deblurring. In *CVPR*, 2013. 2, 3, 4, 5, 6, 7, 8

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

[31] H. Zhang and L. Carin. Multi-shot imaging: Joint alignment, deblurring and resolution enhancement. In *CVPR*, 2014. 2

[32] H. Zhang and D. Wipf. Non-uniform camera shake removal using a spatially adaptive sparse penalty. In *NIPS*, 2013. 2, 4, 5, 8

[33] H. Zhang, D. P. Wipf, and Y. Zhang. Multi-image blind deblurring using a coupled adaptive sparse prior. In *CVPR*, 2013. 2, 3, 4

[34] H. Zhang and J. Yang. Scale adaptive blind deblurring. In *NIPS*, 2014. 2

[35] X. Zhu, F. Šroubek, and P. Milanfar. Deconvolving PSFs for a better motion deblurring using multiple images. In *ECCV*, 2012. 2