Multi-Shot Imaging: Joint Alignment, Deblurring and Resolution-Enhancement

Haichao Zhang   Lawrence Carin

Duke University

hczhang1@gmail.com   lcarin@duke.edu

Abstract

The capture of multiple images is a simple way to increase the chance of capturing a good photo with a lightweight hand-held camera, for which the camera-shake blur is typically a nuisance problem. The naive approach of selecting the single best captured photo as output does not take full advantage of all the observations. Conventional multi-image blind deblurring methods can take all observations as input but usually require the multiple images are well aligned. However, the multiple blurry images captured in the presence of camera shake are rarely free from mis-alignment. Registering multiple blurry images is a challenging task due to the presence of blur while deblurring of multiple blurry images requires accurate alignment, leading to an intrinsically coupled problem. In this paper, we propose a blind multi-image restoration method which can achieve joint alignment, non-uniform deblurring, together with resolution enhancement from multiple low-quality images. Experiments on several real-world images with comparison to some previous methods validate the effectiveness of the proposed method.

1. Introduction

Capturing high-quality images is usually a central goal in many imaging scenarios. Considerable improvements have been made on the hardware side (e.g., the increasing resolutions of the camera sensor) to aid the users in achieving this goal. However, some factors such as camera shake during the photo capture will still lead to degradations in the captured images, especially with light-weight hand-held cameras, such as those in smart phones.

One important effect of camera shake during exposure is motion blur, preventing one from obtaining high-quality images. *Lucky imaging* is a simple strategy to obtain a high quality image under this scenario, taking multiple photos of the same scene and picking the best as output; this is a scheme embedded in many recent digital cameras. While simple, this technique discards much information that is potentially useful for constructing a high-quality image. Fur-

thermore, it relies on the assumption that there is a single 'good' quality photo that has been captured. Blind image deblurring techniques have also been developed to recover a high-quality sharp image from blurry, compromised observations. Extensive efforts have been devoted to the uniform blur (shift-invariant) case [16, 5, 22, 17], while increasing efforts have been devoted to the task of deblurring in the presence of non-uniform blur recently, a more realistic and challenging task [18, 20, 7, 9, 23]. Direct extension of these methods to multiple images usually faces the challenging problem of mis-alignment among the multiple observations, caused by camera shake.

The work in this paper is inspired by the intimate relationship between motion and blur as well as mis-alignment: *the camera motion, when integrated within-frame, will produce intra-frame motion blur while causing inter-frame mis-alignment when it appears between frames.* The removal of the within-frame motion is known as the task of **deblurring**, while the removal of inter-frame motion is known as **alignment** or **registration** (see Figure 1). The objective of this work is to develop an approach for recovering the high-quality sharp image from several observations with *unknown* (non-uniform) blur and spatial mis-alignment caused by camera motion. This kind of technique is especially useful due to the recent popular burst imaging and exposure bracketing functionalities of hand-held digital cameras, making multi-image capture a popular approach. The main contributions of this paper are:

1. We present a joint formulation for the tasks of alignment, deblurring and resolution enhancement for flexible multi-image restoration. The proposed model is mathematically principled and requires no empirical tuning or specific restrictions on the input images to achieve success.

2. The novel formulation allows us to solve the problem by taking advantage of an existing algorithm [26] with a natural extension. Various experiments are conducted compared with *state-of-the-art* methods to valid the effectiveness of the proposed method.

The remainder of the paper is structured as follows. In Section 2, we make briefly review previous work on restora-

tion using multiple images, with a special focus on the connection between image alignment and restoration. Section 3 presents the proposed blind multi-image restoration method, including a discussion of its connection with several previous methods, as well as implementation details. Experiments are carried out in Section 4, and results are compared with *state-of-the-art* methods in the literature. Finally, we conclude the paper in Section 5.

2. Alignment and Restoration: Prior Art

Camera motion is manifested as intra-frame blur and inter-frame mis-alignment among multiple observations. Resolution enhancement utilizing alignment/motion among multiple low-resolution observations has been exploited extensively since the seminal work of Tsai and Huang [19], where the blur kernel is typically assumed to be known. Blur-kernel estimation techniques using camera motion provided by auxiliary video cameras [1], or inertial measurement sensors [10], have been developed.

In the presence of only blurry observations without knowing the blur kernel or additional camera motion information, most previous multi-image restoration methods take well-aligned images as input for restoration [25, 20, 27]. Yuan *et al.* utilized a well-aligned blurry-noisy image pair for estimating the uniform convolutional blur kernel [25]. Li *et al.* explored well-aligned blurry image pairs captured with a specially designed device [12]. Zhang *et al.* utilized well-aligned training images as external domain-knowledge for guiding face deblurring [28].

However, in many practical scenarios it is unrealistic to assume the multiple blurry observations are always well-aligned, e.g., when capturing multiple images of the same scene using the exposure bracketing or burst-mode functionality of a hand held camera, where the handshake is almost inevitable. In the case of multiple blurry images with mis-alignment, a pre-alignment step using standard registration techniques is usually introduced [16, 17, 24, 20]. S-roubek *et al.* tackled the problem of image mis-alignment between the multiple blurry images with general 2D blurs using a pre-alignment step and then used the registered blurry images as input to a multi-channel blind deblurring algorithm [17]. Yuan *et al.* proposed a two-step approach for deblurring using blurry/noisy image pairs, where the two images are first aligned to each other using a coarse-to-fine greedy search and then the final blur kernel can be estimated straightforwardly from the well-aligned image pairs [24]. Cho *et al.* proposed a non-uniform deblurring method using registration [2], with the requirement that the dominant blur directions of the two blurry images should be approximately orthogonal to each other to help with blur estimation.

Real-world applications typically suffer from non-uniform blur and inter-frame mis-alignment at the same time, which has either been simplified (such as using unifor-

m blur model [16, 17]) or addressed using a greedy-type approach [2]. While the gross mis-alignment component can be removed relatively easily via a standard pre-alignment process, the remaining mis-alignment residue due to the blur may still pose a great challenge for the subsequent deblurring process, which typically causes ringing artifacts. This paper presents an approach that can perform joint alignment and non-uniform blur estimation.

3. Approach

In the presence of camera-motion induced blur, much previous work assumes a spatially invariant convolutional process for generating the blurry image $\mathbf{y}$ from a latent sharp image $\mathbf{x}$ [16, 17, 24, 5, 22]. However, typical blur due to camera shake often deviates from the standard uniform convolutional assumption, in part because of problematic rotations which create greater blurring away from some unknown center point [20, 7]. To explicitly account for realistic camera motion, the *projective motion path* (PMP) model [18] treats a blurry image as the weighted sum of projectively transformed sharp images, leading to the revised observation model

$$\mathbf{y} = \sum_j w_j \mathbf{P}_j \mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{P}_j$ is the j-th projection (a combination of rotations and translations) and w_j is the corresponding combination weight representing the proportion of time spent at that particular camera pose during exposure, with the constraint $w_j \geq 0$ and $\sum_j w_j = 1$; $\mathbf{n}$ denotes a noise term. The uniform convolutional model can be obtained by restricting the general projection operators $\{\mathbf{P}_j\}$ to be translations. In this regard, (1) represents a more general model and has been used in many recent non-uniform deblurring work [18, 20, 7, 26]. Our work is built on this non-uniform observation model and a recently introduced deblurring framework using an empirical Bayesian technique [21, 26], as will be reviewed briefly in the sequel.

3.1. Empirical Bayesian Deblur: A Brief Review

The model is derived in the derivative domain, meaning that $\mathbf{x} \in \mathbb{R}^m, \mathbf{y} \in \mathbb{R}^n$ will now denote the lexicographically ordered sharp and blurry image derivatives, respectively. [1] The observation model (1) leads to the following likelihood

$$p(\mathbf{y}|\mathbf{x}, \mathbf{w}, \lambda) \propto \exp\left[-\frac{\|\mathbf{y} - \sum_j w_j \mathbf{P}_j \mathbf{x}\|_2^2}{2\lambda}\right], \tag{2}$$

where $\mathbf{w} = [w_1, w_2, \cdots]$ and λ denotes the noise variance. A straightforward approach would place a prior over the latent high-quality image $\mathbf{x}$ and compute the maximum a pos-

---
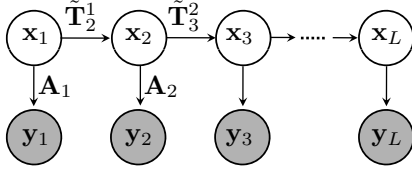
[1] The filter set $\{[-1, 1], [-1, 1]^T\}$ is used in this work.

Figure 1. **Graphical model for the observation process.** $\mathbf{A}_l \triangleq \sum_j w_{(j,l)} \mathbf{DP}_j$ denotes the low-quality observation operator (non-uniform blur and decimation) for the observation process and $\tilde{\mathbf{T}}_l^{l-1}$ denotes the relative motion between the latent high-quality images of the two consecutive observations. Both $\mathbf{A}_l$ and $\tilde{\mathbf{T}}_l^{l-1}$ are caused by camera shakes.

terior (MAP) solution, which typically requires empirical tuning for achieving success.

An alternative approach has been advocated recently in [21, 26], adapted from empirical Bayesian sparse estimation technique in the machine learning literature [15]. Mathematically, this alternative scheme requires that we solve

$$\max_{\boldsymbol{\gamma},\mathbf{w},\lambda \geq 0} \int p(\mathbf{y}|\mathbf{x},\mathbf{w},\lambda) \prod_i \mathcal{N}(x_i;0,\gamma_i) d\mathbf{x} \quad (3)$$
$$\equiv \min_{\boldsymbol{\gamma},\mathbf{w},\lambda \geq 0} \mathbf{y}^T \boldsymbol{\Sigma}^{-1} \mathbf{y} + \log|\boldsymbol{\Sigma}|,$$

where $p(\mathbf{x}) = \prod_i p(x_i) = \prod_i \max_{\gamma_i \geq 0} \mathcal{N}(x_i;0,\gamma_i)$ is a sparseness-promoting prior on $\mathbf{x}$ (in the derivative domain), $\boldsymbol{\Sigma} \triangleq (\mathbf{H}\boldsymbol{\Gamma}\mathbf{H}^T + \lambda\mathbf{I})$, $\mathbf{H} \triangleq \sum_j w_j \mathbf{P}_j$ and $\boldsymbol{\Gamma} \triangleq \mathrm{diag}[\boldsymbol{\gamma}]$.

While optimizing (3) enjoys a strong Bayesian motivation, the $\boldsymbol{\gamma}$-dependent cost function is not as intuitive as the standard penalized regression models (or MAP) that are typically employed for blind deblurring [9, 22]. Furthermore, optimizing (3) using general EM techniques is possible but is computationally expensive in part because of the high-dimensional determinant $\boldsymbol{\Sigma}$ [15]. Therefore, minimizing a convenient upper bound with the following approximation of $\boldsymbol{\Sigma}$ allows one to circumvent this issue [26]

$$\log|\boldsymbol{\Sigma}| \leq \sum_i \log\left(\lambda + \gamma_i\|\mathbf{h}_i\|_2^2\right) + (n-m)\log\lambda, \quad (4)$$

where $\mathbf{h}_i$ denotes the i-th column of $\mathbf{H}$. Using standard linear algebra together with the bound (4), it can be shown that (3) can be reformulated as [21, 26]

$$\min_{\mathbf{x};\mathbf{w},\lambda \geq 0} \frac{1}{\lambda}\|\mathbf{y}_l - \sum_j w_j \mathbf{P}_j \mathbf{x}\|_2^2 + \mathfrak{g}_s(\mathbf{x},\mathbf{w},\lambda) + (n-m)\log\lambda,$$

$$\text{with} \quad \mathfrak{g}_s(\mathbf{x},\mathbf{w},\lambda) \triangleq \sum_i g(x_i,\mathbf{h}_i,\lambda) \quad \text{and}$$

$$g(x_i,\mathbf{h}_i,\lambda) \triangleq \min_{\gamma_i} \frac{x_i^2}{\gamma_i} + \log(\lambda + \gamma_i\|\mathbf{h}_i\|_2^2). \quad (5)$$

The cost function for handling multiple degraded observations can be obtained by simply accumulating the cost function over all the observations, without considering the mis-alignment [27].

3.2. High-Quality Image from Multiple Captures in Presence of Camera Shake

We will derive the mathematical model for the task of *blind multi-image restoration*[2] in the presence of unknown blur and mis-alignment. We propose the following model for the low-quality-images observation process in the presence of camera motion, which is a generalization of (1):

$$\mathbf{y}_l = \sum_j w_{(j,l)} \mathbf{DP}_j \mathbf{x}_l + \mathbf{n}_l, \quad (6)$$

$$\mathbf{x}_l = \tilde{\mathbf{T}}_l^{l-1} \mathbf{x}_{l-1}, \quad (7)$$

where $\mathbf{x}_l \in \mathbb{R}^n$ is the latent canonical high-quality image for the l-th observation and $\mathbf{D}$ is the sub-sampling operator. The observation process (6) models the low quality (blurry and/or low resolution) observation $\mathbf{y}_l \in \mathbb{R}^m$ as a weighted sum of projectively transformed and decimated high-quality image $\mathbf{x}_l$. The latent high-quality images of the two consecutive captures are linked via the the *relative projective motion* operator $\tilde{\mathbf{T}}_l^{l-1}$ between them.[3] A graphical illustration of the observation model is given in Figure 1. With the relative motion model in (7), it is easy to derive the following relationship between two observations:

$$\mathbf{x}_l = \tilde{\mathbf{T}}_l^{l-1} \mathbf{x}_{l-1} = \prod_{k=1}^l \tilde{\mathbf{T}}_k^{k-1} \mathbf{x}_0 \triangleq \mathbf{T}_l \mathbf{x}, \quad (8)$$

where $\mathbf{T}_l \triangleq \prod_{k=1}^l \tilde{\mathbf{T}}_k^{k-1}$ and $\mathbf{x} \triangleq \mathbf{x}_0$. Substituting (8) into (6), we have the following observation model:

$$\mathbf{y}_l = \sum_j w_{(j,l)} \mathbf{DP}_j \mathbf{T}_l \mathbf{x} + \mathbf{n}_l. \quad (9)$$

With (9), the l-th low quality image $\mathbf{y}_l$ is generated by first transforming the latent high-quality image $\mathbf{x}$ with $\mathbf{T}_l$, and then passing this transformed high-quality image through the low-quality observation process (6) (see Figure 1).

Many previous approaches attempted to align the observations in hope of removing the factor $\mathbf{T}_l$ prior to deblurring, e.g., [16, 17, 24]. However, the task of accurate alignment in the presence of blur itself is very challenging [24]. We are inspired by the fact that the mis-alignment and blur (within our context) are all caused by the relative motion between the camera and scene due to camera shake, and therefore it is desirable that they are handled jointly. This observation inspires us to develop the approach for joint blur recovery and mis-alignment removal.

The objective of blind multi-image restoration is to estimate the latent high-quality image $\mathbf{x}$ as well as the non-uniform blur and alignment given a set of low-quality observations $\{\mathbf{y}_l\}$ with mis-alignment. We propose the following

---

[2]Here we use this term to denote the general task of recovering a high-quality (high-resolution, sharp) image from a set of (single or multiple) low-quality (low-resolution, blurry) images.

[3]The relative motion model (7) is presented in error-free form for simplicity but the alignment errors can be absorbed into the noise term of $\boldsymbol{\theta}$).

model for joint blind restoration (alignment, deblurring and resolution enhancement) given multiple low-quality observations ($l \in (1, \cdots, L)$):

$$\min_{\boldsymbol{\gamma}, \{\mathbf{w}_l, \lambda_l\} \geq 0} \sum_l \mathbf{y}_l^T \boldsymbol{\Sigma}_l^{-1} \mathbf{y}_l + \log \prod_l |\boldsymbol{\Sigma}_l|, \qquad (10)$$

where $\boldsymbol{\Sigma}_l \triangleq \left(\mathbf{H}_l \boldsymbol{\Gamma} \mathbf{H}_l^T + \lambda_l \mathbf{I}\right)$, $\mathbf{H}_l \triangleq \sum_j w_{(j,l)} \mathbf{DP}_j \mathbf{T}_l$. (10) can be rewritten into a penalized regression form by utilizing a diagonal upper-bounding technique similar to (4):

$$\min_{\mathbf{x}; \{\mathbf{w}_l, \lambda_l\} \geq 0} \sum_l \frac{1}{\lambda_l} \|\mathbf{y}_l - \sum_j w_{(j,l)} \mathbf{DP}_j \mathbf{T}_l \mathbf{x}\|_2^2$$
$$+ \mathfrak{g}(\mathbf{x}, \{\mathbf{w}_l, \lambda_l\}) + (n - m) \log \lambda \qquad (11)$$

where

$$\mathfrak{g}(\mathbf{x}, \{\mathbf{w}_l, \lambda_l\}) = \sum_{il} \min_{\gamma_i} \left[\frac{x_i^2}{\gamma_i} + \log(\lambda + \gamma_i \|\mathbf{h}_{(i,l)}\|_2^2)\right]. \quad (12)$$

The local kernel $\mathbf{h}_{(i,l)}$ can now be calculated by

$$\mathbf{h}_{(i,l)} = \sum_j w_{(j,l)} \mathbf{DP}_{(j,l)} \mathbf{T}_l \mathbf{e}_i \triangleq \mathbf{H}_l \mathbf{e}_i \triangleq \mathbf{B}_{(i,l)} \mathbf{w}_l, \quad (13)$$

where $\mathbf{e}_i$ denotes an all-zero image with a 1 at site i. Consequently we have $\|\mathbf{h}_{(i,l)}\|_2^2 = \mathbf{w}_l^T (\mathbf{B}_{(i,l)}^T \mathbf{B}_{(i,l)}) \mathbf{w}_l$. We have several comments regarding the model in (11):

- While the global transformation operator $\mathbf{T}$ may not be necessary in the presence of a single observation (as it can be absorbed into the recovered sharp image as $\mathbf{Tx}$), it is necessary for capturing the geometric relation between different observations in the presence of multiple observations, as illustrated in Figure 3.
- The proposed model can achieve image enhancement in the presence of unknown non-uniform blur and mis-alignment from low-resolution images, which can be regarded as a unified model for conventional blind deblurring and super-resolution tasks.
- The proposed method is flexible in the number of inputs. It reduces to a single image based blind restoration method with only a single observation, while can take full advantage of multiple images when available.
- The resulting algorithm (Algorithm 1) is virtually tuning-parameter free, requiring no empirical tuning of the weight parameters for balancing the contributions from different terms or observations.

3.3. Relationship with Existing Methods

The proposed model (11) is a flexible framework especially appropriate for the task of blind multi-image restoration, which has been investigated from different perspectives under different assumptions in the literature. For example, super-resolution techniques using multiple images

---

Algorithm 1: Blind Multi-Image Restoration.

1: **Input:** blurry images $\{\mathbf{y}_l\}$
2: **Initialize:** blur parameters $\{\mathbf{w}_l\}$, noise levels $\{\lambda_l\}$
3: **While** stopping criteria is not satisfied, do

- ***Image Estimation:***
  $\mathbf{x} \leftarrow \sum_l [\frac{\mathbf{H}_l^T \mathbf{H}_l}{\lambda_l} + \boldsymbol{\Gamma}^{-1}]^{-1} \sum_l \frac{\mathbf{H}_l^T \mathbf{y}_l}{\lambda_l}$

- ***Latent Variable Estimation:***
  $\gamma_i \leftarrow x_i^2 + \sum_l z_{il}/L$, with $z_{li} = \frac{1}{\frac{\|\mathbf{h}_{(i,l)}\|_2^2}{\lambda} + \gamma_i^{-1}}$

- ***Blur/Registration Parameter Estimation:***
  $\min_{\mathbf{w}_l \geq 0} \|\mathbf{y}_l - \mathbf{U}_l \mathbf{w}_l\|_2^2 + \mathbf{w}_l^T \sum_i z_{li} \mathbf{B}_{(i,l)}^T \mathbf{B}_{(i,l)} \mathbf{w}_l$
  where $\mathbf{U}_l = [\mathbf{DP}_{(1,l)} \mathbf{T}_l, \mathbf{DP}_{(2,l)} \mathbf{T}_l, \cdots] \mathbf{x}$ and
  $\mathbf{B}_{(i,l)} = [\mathbf{DP}_{(1,l)} \mathbf{T}_l, \mathbf{DP}_{(2,l)} \mathbf{T}_l, \cdots] \mathbf{e}_i$.

- ***Noise Estimation:***
  $\lambda_l = \frac{\|\mathbf{y}_l - \mathbf{H}_l \mathbf{x}\|_2^2 + \beta_l}{n}$, with $\beta_l = \sum_i z_{li} \|\mathbf{h}_{(i,l)}\|_2^2$

4: **End**

---

typically take the mis-alignment into consideration but treat the blur kernel either as known [4], or estimated but with simple form, without modeling motion blur [13]. On the other hand, multi-image deblurring methods can estimate the blur kernels but typically assume that all images are well-aligned [12, 27]. Our model, in contrast, can handle multiple images with unknown alignments jointly with complex motion blurs, more general than both kinds of methods. Related task on single-image super-resolution, joint with blur kernel estimation, has been recently investigated in [14]. In the case of a single observation and discarding the transformation factor $\mathbf{T}$ and decimation operator $\mathbf{D}$, the proposed model reduces to the recent single image non-uniform deblurring method proposed in [26].

The proposed model differs significantly from standard multi-channel blind deconvolution methods (e.g., [17, 27]), with a uniform convolutional model without decimation, and without joint handling of mis-alignment. Although this appears similar to standard MAP formulations used for blind deblurring [22, 9, 2], a crucial difference is that the penalty term $\mathfrak{g}$ in the proposed model (11) jointly regularizes $\mathbf{x}$, $\mathbf{w}_l$, and λ_l. This coupled penalty introduces a natural way to avoid premature convergence to a bad local minima, by progressively introducing more concavity (sparsity) of the penalty function automatically, according to the noise level and blur estimation [26]. Conventional MAP formulations, using separate penalty terms, do not have this property. Moreover, the proposed approach can estimate the noise level for each observation, thus avoiding the empirical selection of the balancing parameters. For the determination of the projection operator set $\mathbf{P}_{(j,l)}$, the proposed method uses a potentially over-complete candidate projection oper-

IMAGE I IMAGE II AVERAGE PROPOSED

Figure 2. **Image Alignment**. Left to right: (a)-(b) two sharp images of the same scene (c) the composite result by averaging (a) and (b). The composite image is blurry as the two images are not aligned. (d) the result image using the proposed method, where even some resolution improvements are noticeable. The learned transformation and kernel patterns are shown in Figure 3.

ator set and then estimates the combination weights over this set, with (close to) zero values for the irrelevant basis, thus selecting relevant transformation basis for usage automatically, requiring no greedy matching schemes as in [2].

3.4. Implementation Details

Joint alignment and non-uniform blind restoration is achieved by minimizing the cost function in (11), which can be accomplished by minimizing its rigorous upper bound via the majorization-minimization technique [15], leading to the algorithm summarized in Algorithm 1.

As mentioned above, the transformation operator $\mathbf{T}_l$ is essential for compensating the geometric mis-alignments when more than one observations are present. Furthermore, it can be combined with the projection set $\{\mathbf{P}_j\}$, producing transformed basis set $\mathbf{P}_{(j,l)}$ for each observation. For simplicity, in practice we have only used projection operators $\mathbf{P}_{(j,l)}$ involving in-plane translations and rotations similar to [7] for modeling the camera shake as well as handling the spatial mis-alignment. To alleviate the computational cost due to the evaluation of the matrix-vector product $\mathbf{Hx}$, the Efficient Filter Flow (EFF) technique is used, which can be combined with the PMP model to produce the local blur kernels for local convolutional approximation [9]. The decimation operator $\mathbf{D}$ is implemented via subsampling the input for every s rows and columns for the zooming factor of s. A standard multi-scale estimation scheme is incorporated consistent with most recent blind deblurring work [5].

Finally, we emphasize that Algorithm 1 only provides an estimate of the latent image in the derivative domain. Consequently, consistent with other methods, we use the estimated blur parameters $\{\mathbf{w}_l\}$ in a final non-blind multi-image restoration step to recover the latent high quality image using a modified approach of [11].

4. Experimental Results

This section provides experimental results, with the proposed method compared with several multi-image restoration methods, including the multi-image deblurring methods (Sroubek *et al.* [17], Zhang *et al.*[27] and Cho *et*
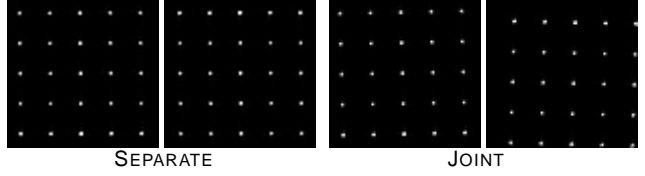

SEPARATE JOINT

Figure 3. **Estimated blur kernel patterns** for the images in Figure 2. Left: blur kernel patterns estimated separately for each image. Right: blur kernel patterns estimated jointly using the proposed method. As can be observed, when estimated separately, the blur kernel patterns are in their canonical form, while when estimated jointly, the relative transformation is captured by the recovered blur kernel patterns.

al. [2, 3]) as well as several super-resolution methods (Farsiu *et al.* [4], Glasner *et al.* [6], Michaeli and Irani [14]). More results and comparisons are not shown here due to space limitation and can be found at https://sites.google.com/site/hczhang1/.

4.1. Image Alignment

We first show an experiment on image alignment. Figure 2 shows two sharp images of the same scene taken consecutively with a hand-held camera. The two captures are not aligned, as can be easily observed from their super-composed images as shown in Figure 2 (c). The proposed method can learn the transformation automatically and generate an output image fusing all two observations, as shown in Figure 2 (d). It can be observed that the output image contains more details even than the original sharp images. Figure 3 shows the learned kernel patterns for the two input images. As can be observed, when estimated separately, the blur kernel patterns are in their canonical form and do not reflect the alignment information, while when estimated jointly, the relative transformation between observations is captured by the recovered blur kernel patterns. All the local kernels are very close to a delta kernel, and the relative position transformation of the kernel patterns encodes the learned transformation between the observed images.

4.2. Multi-Image Deblurring

In this section we evaluate the performance of the proposed method on the task of multi-image deblurring. For the first experiment, we use three blurry text images with mis-alignment from Sroubek *et al.* [17], as shown in Figure 4. These images are generated by attaching a text label to a vibrating machine, and using a standard digital camera mounted on a tripod for image capturing, with an exposure time of $1/15$s that is comparable to the irregular variation period of the machine [17]. It is clear that the captured images are not aligned with each other due to the relative motion between the text label and camera caused by the machine variation. We first use the multi-image deblurring
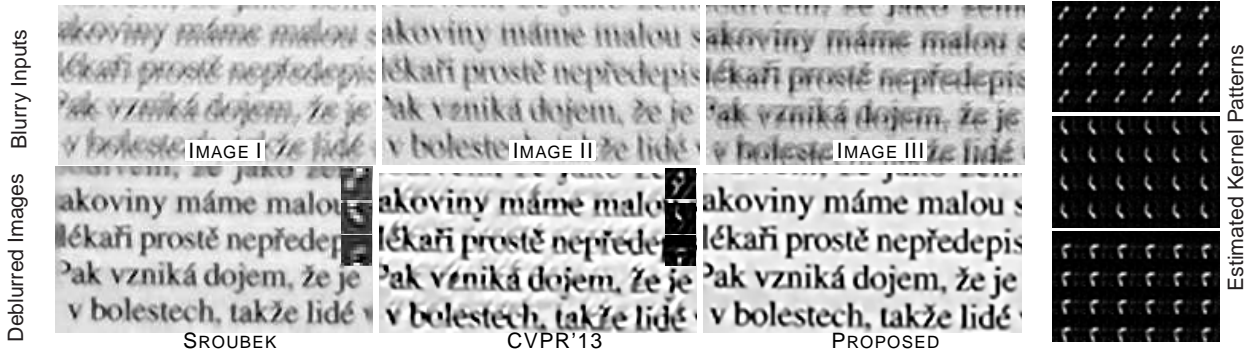
Figure 4. **Text deblurring**. Top left: three blurry observations. Bottom left: deblurred images using different methods (Sroubek *et al.* [17], Zhang *et al.* [27] and the proposed method). Right column: estimated blur kernel patterns using the proposed method.

Figure 5. **Deblurring results**. (a) blurry images. (b) Zhang and Wipf [26] (c) Cho *et al.* [2]. (d) Our result.

method proposed in [27] for restoration, which uses a convolutional blur model and does not handle mis-alignment, apart from small translational ones. As can be observed from Figure 4, the deblurred image has severe ringing artifacts, possibly due to the non-translational mis-alignments which cannot be compensated by the translation of the kernels for the method in [27]. Sroubek *et al.* proposed to handle the mis-alignment problem by using a pre-registration step to align the blurry observation and then used a standard multi-image blind deblurring technique with a uniform blur assumption for recovering the final restoration result [17], as shown in Figure 4. As both models in [27] and [17] use uniform blur model, the main difference lies in the additional pre-alignment step of [17]. By comparing the results from both methods, it is clear that the deblurring method with a pre-alignment process [17] outperforms the most recent method without alignment handling [27], indicating the importance of the alignment. The proposed method, with alignment and deblurring integrated together, can generate a deblurred image with better quality than that of the two-step approach by Sroubek *et al.* [17]. As shown in Figure 4, the recovered kernel patterns and their relative positions encode the blur (intra-frame camera motion) and transformation (inter-frame camera motion) respectively. While the blur is relatively uniform in this case, the advantage of the proposed joint method would be more pronounced in the presence of both mis-alignment and non-uniform blur, as shown by the examples below.

Another set of experiments using test images from [2] are shown in Figure 5. It is clear that the two blurry images are not aligned. The deblurring result by the single-image-based deblurring method in [26], using the first blurry image, is imported for comparison. As can be observed, the single-image-based method [26] produces a deblurred image with apparent ghosting artifacts (see the zoomed patch for face area), as well as ringing artifacts (see the zoomed patch for the texts). The registration based method by Cho *et al.* [2] produces a deblurred image with less artifacts than the single-image-based method, but the deblurred image is blurry and with some fine details missing. The result from the proposed method is clear and sharp, with the least ringing and ghosting artifacts, while it is sharper and with more fine details than the result of Cho *et al.* [2].

Figure 6 shows two blurry photographs of a poster captured with a hand-held camera. The deblurring result using the method of [27] has severe ringing artifacts and the detailed structures are not recovered. The proposed method can recover more detailed structure and produce a deblurred image with less ringing artifacts.

Furthermore, simple video deblurring can be achieved with the proposed method by taking several consecutive video frames as input (Figure 8). Higher quality results are expected with more advanced alignment/matching scheme (e.g. [8, 13]), which will be our future work.
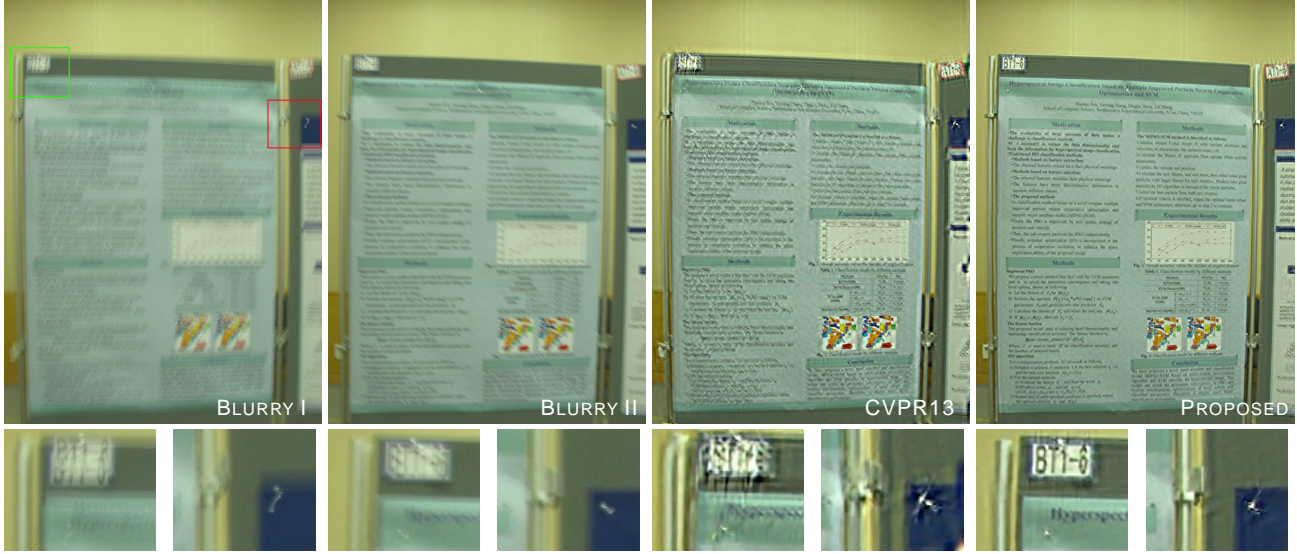
Figure 6. **Deblurring results**. (a) The first blurry image (b) The second blurry image (c) Zhang *et al.* [27]. (d) Our result.

Figure 7. **Blind multi-image super resolution**($2\times$). (a) low-res and blurry images. (b) result by Farsiu *et al.* [4] (c) Our result.

4.3. Blind Super-Resolution

With the subsampling operator integrated, the proposed model can be applied to the *blind super-resolution* (BSR) problem, where the blur kernel is also required to be estimated apart from the high-resolution image [14]. We first demonstrate the application of the proposed method on the task of multi-image blind super-resolution in Figure 7, where six low-quality images with moderate motion blur are used as input. It can be observed that the proposed method can generate a restored result with more details than the conventional multi-frame super resolution method [4]. Further, we compare with the very recent work of Michaeli and Irani [14] on single image BSR task using test images from their paper, as shown in Figure 9. As can be seen from the comparison, the proposed method, even without using a sophisticated image prior (e.g., multi-scale self-similarity [6, 14]), can outperform the method of Glasner *et al.* [6] and performs similar to Michaeli and Irani's recent method [14], both of which incorporated the multi-scale self-similarity image prior.

5. Conclusion

A blind multi-image restoration method is presented in this paper, which can achieve joint alignment, non-uniform deblurring as well as resolution enhancement. Currently, the scene involved in imaging is assumed to be static and the transformations relating the multiple observations are globally rigid. Future work would include generalizing the proposed model to handle scenes involving moving objects as well as non-rigid transformations using techniques similar to [8, 13]. Another interesting extension is to exploit external domain knowledge for improved blur estimation and image recovery [28]. Extending the proposed method for joint video stabilization and deblurring is also an interesting future research direction.

References

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

[2] S. Cho, H. Cho, Y.-W. Tai, and S. Lee. Registration based non-uniform motion deblurring. *Comput. Graph. Forum*, 31(7-2):2183–2192, 2012. 2, 4, 5, 6

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

[4] S. Farsiu, M. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *IEEE Trans. Image Process.*, 513(10):1327–1344, 10 2004. 4, 5, 7

[5] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. In *SIGGRAPH*, 2006. 1, 2, 5

[6] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *ICCV*, 2009. 5, 7, 8

[7] A. Gupta, N. Joshi, C. L. Zitnick, M. Cohen, and B. Curless. Single image deblurring using motion density functions. In *ECCV*, 2010. 1, 2, 5

Figure 8. **Video deblurring**. (a)-(b) Two consecutive video frames from [3] (c) Cho *et al.* [3] (d) Our result.

Figure 9. **Blind SR**(top:2×, bottom: 3×). (a) Low-res (zoomed-up by NN) (b) Glasner *et al.* [6] (c) Michaeli and Irani [14] (d) Our result.

[8] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski. N-RDC: Non-rigid dense correspondence with applications for image enhancement. In *SIGGRAPH*, 2011. 6, 7

[9] M. Hirsch, C. J. Schuler, S. Harmeling, and B. Schölkopf. Fast removal of non-uniform camera shake. In *ICCV*, 2011. 1, 3, 4, 5

[10] N. Joshi, S. B. Kang, C. L. Zitnick, and R. Szeliski. Image deblurring using inertial measurement sensors. *SIGGRAPH*, 2010. 2

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

[12] W. Li, J. Zhang, and Q. Dai. Exploring aligned complementary image pair for blind motion deblurring. In *CVPR*, 2011. 2, 4

[13] C. Liu and D. Sun. On bayesian adaptive video super resolution. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(2):346–360, 2014. 4, 6, 7

[14] T. Michaeli and M. Irani. Nonparametric blind super-resolution. In *ICCV*, 2013. 4, 5, 7, 8

[15] J. A. Palmer, D. P. Wipf, K. Kreutz-Delgado, and B. D. Rao. Variational EM algorithms for non-gaussian latent variable models. In *NIPS*, 2006. 3, 5

[16] A. Rav-Acha and S. Peleg. Two motion blurred images are better than one. *Pattern Recognition Letters*, 26:311–317, 2005. 1, 2, 3

[17] F. Sroubek and J. Flusser. Multichannel blind deconvolution of spatially misaligned images. *IEEE Trans. Image Process.*, 14(7):874–883, 2005. 1, 2, 3, 4, 5, 6

[18] 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. 1, 2

[19] R. Tsai and T. Huang. Multiframe image restoration and registration. *Advances in Computer Vision and. Image Processing*, 1. 2

[20] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce. Non-uniform deblurring for shaken images. In *CVPR*, 2010. 1, 2

[21] D. P. Wipf and H. Zhang. Revisiting Bayesian blind deconvolution. Technical report, MSRA, Feb. 2013. 2, 3

[22] L. Xu and J. Jia. Two-phase kernel estimation for robust motion deblurring. In *ECCV*, 2010. 1, 2, 3, 4

[23] L. Xu, S. Zheng, and J. Jia. Unnatural L0 sparse representation for natural image deblurring. In *CVPR*, 2013. 1

[24] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum. Blurred/non-blurred image alignment using sparseness prior. In *ICCV*, 2007. 2, 3

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

[26] H. Zhang and D. Wipf. Non-uniform camera shake removal using a spatially-adaptive sparse penalty. In *NIPS*, 2013. 1, 2, 3, 4, 6

[27] H. Zhang, D. Wipf, and Y. Zhang. Multi-image blind deblurring using a coupled adaptive sparse prior. In *CVPR*, 2013. 2, 3, 4, 5, 6, 7

[28] H. Zhang, J. Yang, Y. Zhang, N. M. Nasrabadi, and T. S. Huang. Close the loop: Joint blind image restoration and recognition with sparse representation prior. In *ICCV*, 2011. 2, 7