Multi-Observation Blind Deconvolution with an Adaptive Sparse Prior

Haichao Zhang, *Student Member, IEEE,* David Wipf, *Member, IEEE,* and Yanning Zhang, *Senior Member, IEEE*

Abstract—This paper describes a robust algorithm for estimating a single latent sharp image given multiple blurry and/or noisy observations. The underlying multi-image blind deconvolution problem is solved by linking all of the observations together via a Bayesian-inspired penalty function, which couples the unknown latent image along with a separate blur kernel and noise variance associated with each observation, all of which are estimated jointly from the data. This coupled penalty function enjoys a number of desirable properties, including a mechanism whereby the relative-concavity or sparsity is adapted as a function of the intrinsic quality of each corrupted observation. In this way, higher quality observations may automatically contribute more to the final estimate than heavily degraded ones, while troublesome local minima can largely be avoided. The resulting algorithm, which requires no essential tuning parameters, can recover a sharp image from a set of observations containing potentially both blurry and noisy examples, without knowing *a priori* the degradation type of each observation. Experimental results on both synthetic and real-world test images clearly demonstrate the efficacy of the proposed method.

Index Terms—Multi-observation blind deconvolution, blind image deblurring, sparse priors, sparse estimation

1 Introduction

Multi-observation blind deconvolution problems exist under various guises in fields such as signal/image processing, computer vision, communications, and controls. For example, in communications it is frequently known as *multi-channel blind equalization*, where the objective is to estimate an unknown input signal that drives the output of several observed channels without knowledge of the source signal or the channel [20]. Similarly, many control theory applications require the *blind identification* of a multi-channel plant model [37], while multi-channel blind deconvolution for single-input and multiple-output (SIMO) systems is an integral part of signal processing tasks including speech dereverberation [10]. Finally, from a computer vision and image processing perspective, numerous scenarios taken from photography [23] and microscopy [26] present us with multiple captures of the same physical scene under different imaging conditions. In all of these examples, the core estimation problem involves the reconstruction of some intrinsic source signal of interest, along with a series of observation-dependent convolution operators, from a series of degraded observations.

Although the algorithms and attendant analysis apply in a general setting, this paper will present a robust blind deconvolution approach derived in the specific context of multi-image deblurring. Here the goal is to recover a single sharp, latent image from multiple blurry and/or noisy observations obtained using, for example, the exposure bracketing or burst-mode functionality found on many consumer cameras. Fig. 1 presents a representative example. A typical factor causing blur is the relative motion between camera and scene during the exposure period, which may arise from hand jitter [9], [22].

Given L corrupted versions of a latent sharp image $\mathbf{x}$, the uniform convolutional blur model [9] assumes the observation process

$$\mathbf{y}_l = \mathbf{k}_l * \mathbf{x} + \mathbf{n}_l, \qquad \forall l \in \{1, \ldots, L\}, \tag{1}$$

where l is the observation index, $\mathbf{k}_l$ is a point spread function (PSF) or blur kernel, $*$ denotes the convolution operator, and $\mathbf{n}_l$ is a zero-mean Gaussian noise term with covariance $\lambda_l \mathbf{I}$. Within this context, the ultimate goal of multi-image blind deblurring is to estimate the sharp and clean image $\mathbf{x}$ given only the blurry and noisy observations $\{\mathbf{y}_l\}_{l=1}^{L}$, without any prior knowledge regarding the unknown kernels $\mathbf{k}_l$ or noise levels λ_l. By combining the complementary information from multiple images, it is often possible to generate higher quality estimates of the scene $\mathbf{x}$ than in the single-observation, blind deconvolution case [21].

While a number of successful multi-image blind deconvolution methods exist, e.g., [4], [6], [21], [23], [39], there remains room for practical improvements and additional theoretical understanding. In this regard, we present

- *H. Zhang and Y. Zhang are with the School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China. E-mail: hczhang1@gmail.com; ynzhang@nwpu.edu.cn.*
- *D. Wipf is with Microsoft Research Asia, Beijing 100080, China. E-mail: davidwipf@gmail.com.*

Manuscript received 25 Apr. 2013; revised 15 Oct. 2013; accepted 12 Nov. 2013. Date of current version 10 July 2014.
Recommended for acceptance by M. S. Brown.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier 10.1109/TPAMI.2013.241

Fig. 1. Dual motion deblurring examples. Full images are shown in Fig. 10.

a principled energy-minimization algorithm that can handle a flexible number of degraded observations without requiring that we know the nature (e.g., blurry vs. noisy) or extent of the degradation for each observation. The underlying cost function relies on a coupled penalty function, which combines the latent sharp image estimate with a separate blur kernel and noise variance associated with each observed image. Theoretical analysis reveals that this penalty provides a useful agency for adaptively balancing the effects of observations with varying quality, while at the same time avoiding suboptimal local minima. All unknown quantities are optimized using a majorization-minimization algorithm that requires no tuning parameters. Additionally, when only a single observation is present, the method reduces to a principled, single-image blind deconvolution algorithm with an image penalty that adaptively interpolates between the ℓ_0 and ℓ_1 norms without any heuristic hyperparameter. Experimental results on both synthetic and real-world test images validate the proposed method relative to the current state-of-the-art. Finally, some brief discussions on non-uniform blur extension is presented. A preliminary version of this work appeared in [35]; however, the conference version contains no proofs, algorithm derivations, detailed analysis, nor the extensive empirical results contained here.

The remainder of the paper is organized as follows. Section 2 briefly reviews some related single image deblurring work as well as existing multi-image blind deconvolution algorithms; we then introduce our alternative algorithm in Section 3. Theoretical properties and analysis related to the proposed adaptive penalty function are presented in Section 4. Extensive empirical comparisons are conducted in Section 5. Some algorithm details, including the derivations of the cost function and the associated minimizing algorithm as well as a discussion on kernel prior extensions are provided in Section 6, followed by the proofs of the Theorems in Section 7.

2 Related Work

Blind deconvolution techniques are too numerous to exhaustively survey here. Therefore, we will only present relevant state-of-the-art algorithms and briefly mention some of their limitations.

2.1 Single-Image Blind Deblurring

Blind deblurring using only a single image has been addressed using a number of different techniques [2], [7], [9], [13], [15], [22], [32], [36]. Many of these can be viewed as maximum a posterior (MAP) estimators with different prior/penalty terms [7], [13], [22], [32], [36], leading to optimization problems of the form

$$\min_{\mathbf{x},\mathbf{k}} \|\mathbf{k} * \mathbf{x} - \mathbf{y}\|_2^2 + \mu E_x(\mathbf{x}) + \tau E_k(\mathbf{k}). \tag{2}$$

Here $E_x(\mathbf{x})$ and $E_k(\mathbf{k})$ regularize the latent image and blur kernel respectively while μ and τ are standard weighting factors. The MAP formulation is relatively straightforward, but frequently requires additional heuristics such as structure selection and sharp edge prediction for generating good deblurring results [7], [22], [32].

Variational Bayesian (VB) techniques provide an alternative to MAP estimation that attempt to make more thorough use of the underlying posterior distribution of the latent image [2], [9], [15], [17], [27]. Although mitigating some of the shortcomings of MAP, VB nonetheless depends on certain posterior factorial assumptions (the so-called mean-field approximation) and cost function that has function-valued arguments and multiple integrations, making transparent analysis difficult. Moreover, a pre-specified, decreasing sequence of noise values is required to achieve reasonable results with some of the most successful recent VB algorithms [2], [15]. Consequently, rigorous analysis leading to principled, multi-image extensions does not currently exist.

2.2 Multi-Image Blind Deblurring

Rav-Acha and Peleg use two motion blurred images with different blur directions and show that restoration quality is superior than when using only a single image [21]. Since this work, many other multi-image blind deblurring algorithms have been developed [4], [6], [23], [39]. Most of these assume that only two blurry observations $\{\mathbf{y}_1, \mathbf{y}_2\}$ are present. In addition to other standard regularizers common to single-image blind deconvolution algorithms, a 'cross-blur' penalty function given by

$$E_k(\mathbf{k}_1, \mathbf{k}_2) = \|\mathbf{k}_2 * \mathbf{y}_1 - \mathbf{k}_1 * \mathbf{y}_2\|_2^2, \tag{3}$$

is often included [6], [23]. The rationale here is that, given the commutative convolutional model from (1), $E_k(\mathbf{k}_1, \mathbf{k}_2)$ should be nearly zero if the noise levels are low and the correct kernels have been estimated. This penalty also implicitly relies on the coprimeness assumption, meaning that the blur kernels can only share a scalar constant [23]. Once the unknown kernels are estimated, the sharp image $\mathbf{x}$ may be recovered using a separate non-blind step if necessary.

Although computationally efficient, inclusion of this quadratic energy term does not always produce good kernel estimation results [6], [39]. One reason is that if the noise level is relatively high, it can dominate the minimization of $E_k(\mathbf{k}_1, \mathbf{k}_2)$, leading to kernel estimates that are themselves blurry, which may then either produce ringing artifacts or loss of detail in the deblurred image [39]. Another issue is solution ambiguity, meaning that for a given optimal solution $\{\tilde{\mathbf{k}}_1, \tilde{\mathbf{k}}_2\}$, there exists a family of solutions $\{\tilde{\mathbf{k}}_1 * \mathbf{h}, \tilde{\mathbf{k}}_2 * \mathbf{h}\}$ that also minimize (3) [6], [39]. Finally, a practical limitation of $E_k(\mathbf{k}_1, \mathbf{k}_2)$ is that it only applies to images pairs, and hence would expand combinatorially as the number of observations grows.

To mitigate some of these problems, a sparse penalty on the blur kernel may be integrated into the estimation objective directly [6], [23] or applied via post-processing [39]. For example, Chen *et al.* propose a modified version of (3) that regularizes the kernel estimates using a sparse prior E_s, a continuity (smoothness) prior E_c, and a robust Lorentzian factor φ leading to the cost function

$$E_k(\mathbf{k}_1, \mathbf{k}_2) = \varphi(\mathbf{k}_2 * \mathbf{y}1 - \mathbf{k}_1 * \mathbf{y}_2) + \alpha \sum_{l=1}^{2} E_s(\mathbf{k}_l) + \beta \sum_{l=1}^{2} E_c(\mathbf{k}_l), \quad (4)$$

where α and β are trade-off parameters [6]. Similarly, Šroubek *et al.* modified (3) and incorporated a sparsity-promoting kernel prior based on a rectified ℓ_1-norm [23].

In contrast, Zhu *et al.* proposed a two-step approach for dual-image deblurring [39]. The blur kernels are first estimated using [23] with (3) incorporated. The resulting 'blurry' estimates $\{\hat{\mathbf{k}}_1, \hat{\mathbf{k}}_2\}$ are then refined in a second, sparsifying step. For this purpose, $\{\hat{\mathbf{k}}_1, \hat{\mathbf{k}}_2\}$ are treated as two blurry images whose sharp analogues are produced by minimizing

$$E_k(\mathbf{k}_1, \mathbf{k}_2, \mathbf{h}) = \sum_{l=1}^{2} \|\mathbf{k}_l * \mathbf{h} - \hat{\mathbf{k}}_i\|_2^2 + \alpha \sum_{l=1}^{2} \|\mathbf{k}_l\|_p^p \quad (5)$$

over $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{h}$, with $p \leq 1$ producing a sparse ℓ_p norm over the kernels. This helps to remove the spurious factor $\mathbf{h}$ mentioned above while producing sparser kernel estimates. Although these approaches are all effective to some extent, the sparsity level of the blur kernels may require tuning.

2.3 Multi-image With Special Capturing

In addition to motion-blurred observation processes, deblurring has also been attempted using multiple images captured under different imaging conditions such as varied exposure lengths [1], [34] and coded apertures [38]. For example, an image pair composed of a short exposure observation, which is typically sharp but noisy, and a long exposure observation, which will generally be blurry but with limited noise, contains complementary information that can be merged to form a single, high-quality image estimate [34]. These assumptions simplify kernel estimation substantially. One strategy is to first independently denoise the short exposure image. A second linear regression step can then be used to estimate the blur kernel of the long exposure image by treating the output of the first step as the latent sharp image [34].

Beyond exposure time length, other points of the image acquisition process can be exploited to implement multi-image blind deblurring. Relevant examples are numerous: a prism system based upon special alignments in motion-blurred image pairs [16]; a procedure that employs flash sequences to obtain complementary high-resolution detail and ambient illumination components of image pairs [40]; coded exposure images captured with a fluttered shutter to preserve high-frequency information [1]; image pairs obtained using coded apertures that also reflect high-frequency components while facilitating depth recovery [38]; and deblurring with the assistant of an auxiliary video camera with low spatial resolution but high frame rate [24].

While using multiple images generally has the potential to outperform the single-image methods by fusing complementary information [6], [21], [23], [34], a principled approach that applies across a wide range of scenarios with little user-involvement or parameter tuning is still somewhat lacking. Our algorithm, which applies to any number of both noisy or blurry images without explicit trade-off parameters, is one attempt to fill this void.

3 A New Multi-Observation Blind Deconvolution Algorithm

We will work in the derivative domain of images for ease of modeling and better performance [9], [15], meaning that $\mathbf{x}$ and $\mathbf{y}$ will denote the lexicographically ordered image derivatives of sharp and blurry images respectively obtained via a particular derivative filter.[1] Because convolution is a commutative operator, the blur kernels are unaltered.

Now consider the case where we have a single observation $\mathbf{y}$. The observation model from (1) defines a Gaussian likelihood function $p(\mathbf{y}|\mathbf{x}, \mathbf{k})$; however, maximum likelihood estimation of $\mathbf{x}$ and $\mathbf{k}$ is obviously ill-posed and hence we need a prior to regularize the solution space. In this regard, it is well-known that the gradients of sharp natural images tend to exhibit sparsity [9], [15], [17], meaning that many elements equal (or nearly equal) zero, while a few values remain large. With roots in convex analysis [18], it can be shown that essentially all iid distributions that favor such sparse solutions can be expressed as a maximization over zero-mean Gaussians with different variances. Mathematically, this implies that $p(\mathbf{x}) = \prod_{i=1}^{m} p(x_i)$ where m is the size of $\mathbf{x}$ ($\mathbf{y}$ is of size $n < m$) and

$$p(x_i) = \max_{\gamma_i \geq 0} \mathcal{N}(x_i; 0, \gamma_i) \exp\left[-\frac{1}{2} f(\gamma_i)\right]. \quad (6)$$

Here f is an arbitrary energy function. The hyperparameter variances $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_m]^T$ provide a convenient way of implementing several different estimation strategies [18]. For example, perhaps the most direct is a form of MAP estimation given by

$$\max_{\mathbf{x};\boldsymbol{\gamma},\mathbf{k} \geq 0} p(\mathbf{y}|\mathbf{x}, \mathbf{k}) \prod_i \mathcal{N}(x_i; 0, \gamma_i) \exp\left[-\frac{1}{2} f(\gamma_i)\right], \quad (7)$$

where simple update rules are available via coordinate ascent over $\mathbf{x}$, $\boldsymbol{\gamma}$, and $\mathbf{k}$ (a prior can also be included on $\mathbf{k}$ if desired). However, recently it has been argued that an alternative estimation procedure may be preferred for canonical sparse linear inverse problems [30]. The basic idea, which naturally extends to the blind deconvolution problem, is to first integrate out $\mathbf{x}$, and then optimize over $\mathbf{k}$, $\boldsymbol{\gamma}$, as well as the noise level λ. The final latent sharp image can then be recovered using the estimated kernel and noise level with standard non-blind multi-image deblurring algorithms.

1. The derivative filters used in this work are $\{[-1, 1], [-1, 1]^T\}$. Other choices are open.

Using the framework from [30], [31], it can be shown that this alternative estimator is formally equivalent to solving

$$\min_{\mathbf{x};\mathbf{k},\lambda\geq 0} \frac{1}{\lambda}\|\mathbf{y}-\mathbf{k}*\mathbf{x}\|_2^2 + \mathfrak{g}(\mathbf{x},\mathbf{k},\lambda), \tag{8}$$

where $\mathfrak{g}(\mathbf{x},\mathbf{k},\lambda) \triangleq \min_{\boldsymbol{\gamma}\geq 0} \mathbf{x}^T\boldsymbol{\Gamma}^{-1}\mathbf{x} + \log|\lambda\mathbf{I}+\mathbf{H}\boldsymbol{\Gamma}\mathbf{H}^T|$, and $\mathbf{H}$ is the convolution matrix of $\mathbf{k}$. The derivation of (8) is summarized in Section 6.1. Note that this expression assumes that f is a constant; rigorous justification for this selection can be found in [31].

Optimization of (8) is difficult in part because of the high-dimensional determinants involved with realistic sized images. To alleviate this problem, we use determinant identities and a diagonal approximation to $\mathbf{H}^T\mathbf{H}$ as motivated in [15]. This leads to the simplified penalty function

$$\mathfrak{g}(\mathbf{x},\mathbf{k},\lambda) = \min_{\boldsymbol{\gamma}\geq 0} \sum_i \left[\frac{x_i^2}{\gamma_i} + \log\left(\lambda+\gamma_i\|\bar{\mathbf{k}}\|_2^2\right)\right], \tag{9}$$

with an extra $(n-m)\log\lambda$ term. Here $\|\bar{\mathbf{k}}\|_2^2 \triangleq \sum_j k_j^2\bar{I}_{ji}$ and $\bar{\mathbf{I}}$ is an indicator matrix with the j-th row recording the (column) positions where the j-th element of $\mathbf{k}$ appears in $\mathbf{H}$. $\|\bar{\mathbf{k}}\|_2^2$ can be viewed as the squared norm of $\mathbf{k}$ accounting for edge effects, or equivalently, as the squared norm of each respective column of $\mathbf{H}$. While technically then $\|\bar{\mathbf{k}}\|_2^2$ should depend on i, the column index of $\mathbf{H}$, we omit explicit referencing for simplicity.[2]

In addition to many desirable attributes as described in [31], the cost function (8) provides a transparent entry-point for multi-image deblurring. Assuming that all observations $\mathbf{y}_l$ are blurry and/or noisy measurements of the same underlying image $\mathbf{x}$, then we may justifiably postulate that $\boldsymbol{\gamma}$ is shared across all l. This then leads to the revised, multi-image optimization problem

$$\begin{aligned}\min_{\mathbf{x},\{\mathbf{k}_l,\lambda_l\geq 0\}} &\sum_{l=1}^{L}\frac{1}{\lambda_l}\|\mathbf{y}_l-\mathbf{k}_l*\mathbf{x}\|_2^2 + (n-m)\log\lambda_l \\ &+ \mathfrak{g}(\mathbf{x},\{\mathbf{k}_l,\lambda_l\}),\end{aligned} \tag{10}$$

where the multi-image penalty function is now naturally defined as

$$\begin{aligned}&\mathfrak{g}(\mathbf{x},\{\mathbf{k}_l,\lambda_l\}) \triangleq \\ &\min_{\boldsymbol{\gamma}\geq 0}\sum_{l=1}^{L}\sum_{i=1}^{m}\left[\frac{x_i^2}{\gamma_i}+\log(\lambda_l+\gamma_i\|\bar{\mathbf{k}}_l\|_2^2)\right].\end{aligned} \tag{11}$$

The proposed cost function (10) can be minimized using coordinate descent (similar to MAP) outfitted with convenient upper bounds that decouple the terms embedded in (11). The resulting majorization-minimization approach, which is summarized in Algorithm 1, is guaranteed to reduce or leave unchanged (10) at each iteration, with similar convergence properties to the EM algorithm. Detailed derivation of the proposed algorithm is provided in the Section 6.2.

While admittedly simple, the proposed model has a number of desirable features:

2. If $\mathbf{H}$ is assumed to be a circulant matrix, then in fact all i-dependency is removed.

Algorithm 1: Multi-Observation Blind Deconvolution

Input: blurry images $\{\mathbf{y}_l\}_{l=1}^L$
Initialize: blur kernels $\{\mathbf{k}_l\}$, noise levels $\{\lambda_l\}$
While stopping criteria is not satisfied, do

- **Update $\mathbf{x}$:**
$$\mathbf{x} \leftarrow \left[\sum_{l=1}^{L}\frac{\mathbf{H}_l^T\mathbf{H}_l}{L\lambda_l}+\boldsymbol{\Gamma}^{-1}\right]^{-1}\sum_{l=1}^{L}\frac{\mathbf{H}_l^T\mathbf{y}_l}{L\lambda_l}$$
where $\mathbf{H}_l$ is the convolution matrix of $\mathbf{k}_l$
- **Update $\boldsymbol{\gamma}$:**
$$\gamma_i \leftarrow x_i{}^2 + \frac{\sum_{l=1}^{L} z_{li}}{L}, \quad \boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$$
$$z_{li} \triangleq \left((\textstyle\sum_j k_{lj}^2\bar{I}_{ji})\lambda_l^{-1}+\gamma_i^{-1}\right)^{-1}$$
- **Update $\mathbf{k}_l$:**
$$\mathbf{k}_l \leftarrow \arg\min_{\mathbf{k}_l\geq 0}\ \frac{1}{\lambda_l}\|\mathbf{y}_l-\mathbf{W}\mathbf{k}_l\|_2^2 + \sum_j k_{lj}^2(\sum_i z_{li}\bar{I}_{ji})$$
with $\mathbf{W}$ the convolution matrix of $\mathbf{x}$
- **Update noise levels λ_l:**
$$\lambda_l \leftarrow \frac{\|\mathbf{y}_l-\mathbf{x}*\mathbf{k}_l\|_2^2 + \sum_{i=1}^{m}\sum_j k_{lj}^2 z_{li}\bar{I}_{ji}}{n}$$

End

- It can handle a flexible number of degraded observations without requiring an extra 'cross-blurring' term, which generally limits the number of observations.
- The input can be a set of blurry or noisy observations without specifying the degradation type of each example; the algorithm will automatically estimate the blur kernel and the noise level for each one. We note that in the case of a single observation, the proposed method reduces to a robust single image blind deblurring model given some minor modifications.[3]
- The penalty function $\mathfrak{g}$ couples the latent image, blur kernels, and noise levels in a principled way. This leads to a number of interesting properties, including an inherent mechanism for scoring the relative quality of each observed image during the recovery process and using this score to adaptively adjust the sparsity of the image regularizer. Section 4 is devoted to these developments.
- The resulting algorithm (Algorithm 1) is tuning parameter free thus requires minimal user involvement.

4 Analysis

This section will examine several theoretical aspects of the penalty function (11). These properties help to explain the

3. In the special case of a single image, extra care must be taken when learning the noise. In this sense the multi-image scenario enjoys a significant advantage in that with each additional image, the number of unknown parameters increases modestly (since the latent sharp image is unchanged) relative to the total number of observations. Hence learning the noise accurately for one image is actually much harder than learning multiple noise levels from multiple observations. On the positive side though, with only a single image our model collapses into a more simplified form rendering additional theoretical analyses and extensions possible. These contingencies unique to the $L=1$ case will be considered in a future publication.

success of our algorithm and hopefully demystify, at least to some extent, what otherwise may appear to be a somewhat non-standard, coupled regularizer that differs substantially from typical MAP estimators.

4.1 Penalty Function Properties

For convenience, we first define

$$h(x, \boldsymbol{\rho}) \triangleq \min_{\gamma \geq 0} \sum_{l=1}^{L} \left[\frac{x^2}{\gamma} + \log(\rho_l + \gamma) \right]. \tag{12}$$

where $\boldsymbol{\rho} \triangleq [\rho_1, \ldots, \rho_L]^T$ with $\rho_l \triangleq \lambda_l / \|\bar{\mathbf{k}}_l\|_2^2$.[4] Then by noting the separability across pixels, (11) can be re-expressed as

$$\mathfrak{g}(\mathbf{x}, \{\mathbf{k}_l, \lambda_l\}) = \sum_{i=1}^{m} h(x_i, \boldsymbol{\rho}) + m \sum_{l=1}^{L} \log \|\bar{\mathbf{k}}_l\|_2^2, \tag{13}$$

which partitions image and kernel penalties into a more familiar form. The second term in (13) penalizes the kernels via a logarithmic transformation of the respective ℓ_2 norms. However, we can reformulate this term into an alternative form that may be more familiar. Let $\mathbf{k}_l^*$ denote a set of kernels that minimize (10). Consequently, solving (10) is equivalent to minimizing

$$\begin{aligned} \min_{\mathbf{x}, \{\mathbf{k}_l, \lambda_l \geq 0\}} & \sum_{l=1}^{L} \frac{1}{\lambda_l} \|\mathbf{y}_l - \mathbf{k}_l * \mathbf{x}\|_2^2 + \sum_{i=1}^{m} h(x_i, \boldsymbol{\rho}), \\ \text{s.t.} \quad & \log \|\mathbf{k}_l\|_2^2 \leq \log \|\mathbf{k}_l^*\|_2^2 \qquad \forall l. \end{aligned} \tag{14}$$

Because the logarithm is a smooth, continuous function over the constraint set, it can be removed from the above constraint. Consequently, the corresponding Langrangian for (14) can be expressed as

$$\begin{aligned} \min_{\mathbf{x}, \{\mathbf{k}_l, \lambda_l \geq 0\}} & \sum_{l=1}^{L} \frac{1}{\lambda_l} \|\mathbf{y}_l - \mathbf{k}_l * \mathbf{x}\|_2^2 + \sum_{i=1}^{m} h(x_i, \boldsymbol{\rho}) \\ & + \sum_{l=1}^{L} C_l \|\bar{\mathbf{k}}_l\|_2^2, \end{aligned} \tag{15}$$

for some constants C_l. Thus, the proposed cost function can be viewed as utilizing weighted quadratic kernel penalties, a typical choice in many blind deblurring algorithms, e.g., [7]. Importantly though, while existing algorithms must heuristically select each C_l, our algorithm implicitly computes these trade-off parameters automatically. Additionally, with minor modifications, it is possible to incorporate non-quadratic kernel factors into this framework as discussed in Section 6.3.

We now turn to the image penalty $h(x, \boldsymbol{\rho})$ in (13), which is quite different from existing image regularizers, and evaluate some of its relevant properties via the Theorems below followed by further discussion and analysis.

Theorem 1 (Concavity). *The penalty function $h(x, \boldsymbol{\rho})$ is a concave non-decreasing function of $|x|$.*

Proofs will be deferred to the Section 7. See Fig. 2 for a graphical illustration of the penalty function. Theorem 1 explicitly stipulates that a strong, sparsity promoting $\mathbf{x}$ penalty is produced by our framework, since concavity with respect to coefficient magnitudes is a well-known, signature property of sparse penalties [30]. Yet while this attribute may anchor our approach as a legitimate sparse estimator in the image (filter) domain, it does not explain precisely why it often produces superior results compared to more traditional MAP (or penalized regression) approaches, which also frequently possess a similar attribute (e.g., ℓ_1 norm-based penalties). For this purpose we must look deeper and examine how $\boldsymbol{\rho}$ modulates the effective penalty on $\mathbf{x}$.

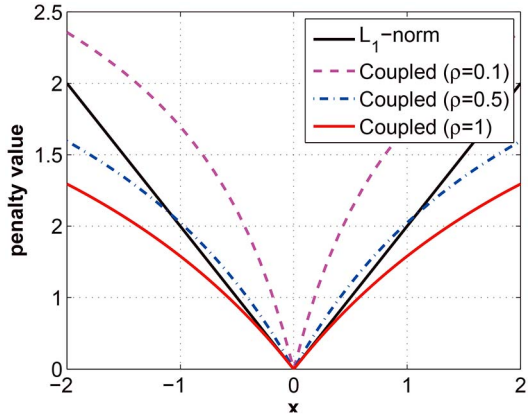

Fig. 2. Plot of the penalty function $h(x, \rho)$ with $L = 1$ (normalized).

First, for two values of the vector $\boldsymbol{\rho}$, e.g., $\boldsymbol{\rho}^1$ and $\boldsymbol{\rho}^2$, we use $\boldsymbol{\rho}^2 \succ \boldsymbol{\rho}^1$ to denote elementwise '$\geq$' with at least one element where the inequality is strict. We also define the function $h_{\boldsymbol{\rho}^\alpha}: \mathbb{R}^+ \rightarrow \mathbb{R}$ as $h_{\boldsymbol{\rho}^\alpha}(z) = h(z, \boldsymbol{\rho} = \boldsymbol{\rho}^\alpha)$, with domain $z \geq 0$. Note that because h is a symmetric function with respect to the origin, we may conveniently examine its concavity/curvature properties considering only the positive half of the real line.

Theorem 2 (Relative Sparsity). *The penalty function $h(x, \boldsymbol{\rho})$ is such that:*

1) *For all $\boldsymbol{\rho}^1$ and $\boldsymbol{\rho}^2$, $h_{\boldsymbol{\rho}^2}(z) - h_{\boldsymbol{\rho}^1}(z) \rightarrow 0$ as $z \rightarrow \infty$. Therefore, $h_{\boldsymbol{\rho}^1}$ and $h_{\boldsymbol{\rho}^2}$ penalize large magnitudes of x equally.*
2) *Let $\boldsymbol{\rho}^2 \succ \boldsymbol{\rho}^1$. Then if $z < z'$, we have $h_{\boldsymbol{\rho}^2}(z) - h_{\boldsymbol{\rho}^1}(z) > h_{\boldsymbol{\rho}^2}(z') - h_{\boldsymbol{\rho}^1}(z')$. Therefore, as $z \rightarrow 0$, $h_{\boldsymbol{\rho}^2}(z) - h_{\boldsymbol{\rho}^1}(z)$ is maximized, implying that $h_{\boldsymbol{\rho}^1}$ favors zero-valued coefficients more heavily than $h_{\boldsymbol{\rho}^2}$.*

4.2 Discussion

From a more intuitive standpoint, $\boldsymbol{\rho}$ represents a form of shape parameter that modulates the concavity, or sparsity favorability, of the image penalty $\sum_i h(x_i, \boldsymbol{\rho})$. Moreover, each element of $\boldsymbol{\rho}$ can be viewed as a measure of the relative quality of a given observation, with larger values indicative of lower quality. This is justified by the fact that larger values of some λ_l (meaning a higher noise level), or small values of some $\|\bar{\mathbf{k}}_l\|_2^2$ (meaning a more difficult, distributed kernel[5]), imply that $\rho_l = \lambda_l / \|\bar{\mathbf{k}}_l\|_2^2$ will be large.

4. Because of boundary effects, technically $\boldsymbol{\rho}$ will depend on i; however we omit this dependency to simplify notation.

5. For a given value of $\sum_j k_{lj}$, a delta kernel maximizes $\|\bar{\mathbf{k}}_l\|_2^2$, while a kernel with equal-valued elements provides the minimum.

Thus we may conclude that the degree of sparsity promotion in $\mathbf{x}$ is ultimately determined jointly by the quality of the constituent observations.

More difficult cases (elements of $\boldsymbol{\rho}$ are all large) occur for one of two reasons: (i) Either the underlying images are really corrupted by complex, diffuse blur kernels and/or high noise, or (ii) in the initial stages the algorithm has not been able to converge to a desirable, low-noise, low-blur solution. In both cases, it can be shown by analyzing the variational expression for h that it becomes flat and nearly convex. This represents a highly desirable adaptation because it helps prevent premature convergence to potentially suboptimal local solutions allowing coarse structures to be identified accurately (structures which generally do not require a highly sparse, non-convex prior to identify per the analysis in [14]).

In contrast, for cases where at least one image has a small ρ_l value, $h(x, \boldsymbol{\rho})$ becomes highly concave in $|x|$ (sparsity favoring), even approaching a scaled (approximate) version of the ℓ_0 norm. To see this, note that whenever $\rho_l \approx 0$, the $\log(\gamma + \rho_l)$ term associated with the l-th image will dominate the variational formation of $h(x, \boldsymbol{\rho})$ leading to the approximation

$$h(x, \boldsymbol{\rho}) \approx \min_{\gamma \geq 0} \left[\frac{x^2}{\gamma} + \log \gamma \right] + \text{constant} \equiv \log |x|. \quad (16)$$

We then obtain a scaled approximation to the maximally-sparse, non-convex ℓ_0 norm since

$$\sum_i \log |x_i| = \lim_{p \to 0} \frac{1}{p} \sum_i (|x_i|^p - 1) \propto \|\mathbf{x}\|_0. \quad (17)$$

Therefore, a single small element in $\boldsymbol{\rho}$ implies that the image penalty will heavily favor sparse solutions, allowing fine-grained kernel structures to be resolved.[6]

Crucially, the existence of a single good kernel/noise estimation pair during the estimation process (meaning the associated ρ_l is small) necessitates that in all likelihood a good overall solution is nearby (otherwise ρ_l being small would mean the cost function value is high). This remains true even if some blur kernel/noise pairs associated with other observations are large. Consequently there is now relatively little danger of local minima with a non-convex penalty since we presumably must be in the neighborhood of a good solution.

The shape-adaptiveness of the coupled penalty function is the key factor leading the algorithm to success in the general case. Both the noise and blur dependency allow the algorithm to naturally possess a 'coarse-to-fine' estimation strategy, recovering large scale structures using a less aggressive (more convex) sparse penalty in the beginning, while later increasing its aggressiveness for recovering the small details.[7] This helps to avoid being trapped in a local minima while recovering the blur kernel progressively.[8]

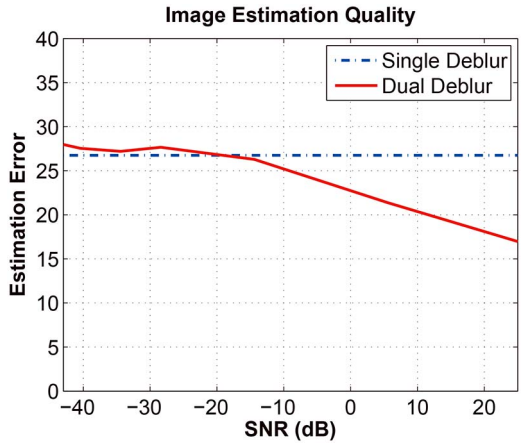

Fig. 3. Image estimation quality under different SNR levels.

Finally, there is also a desirable form of scale invariance attributable to the proposed cost function, meaning that if $\mathbf{x}^*$ and $\{\mathbf{k}_l^*\}$ represent the optimal solution to (10) under the constraint $\sum_j k_{lj} = 1, \forall l$, then $\alpha^{-1}\mathbf{x}^*$ and $\{\alpha \mathbf{k}_l^*\}$ will always represent the optimal solution under the modified constraint $\sum_j k_{lj} = \alpha, \forall l$. Many previous models lack this type of scale invariance, and the exact calibration of the constraint (or related trade-off parameters) can fundamentally alter the form of the optimal solution beyond an irrelevant rescaling, thus require additional tuning.

4.3 Illustrative Example

Interestingly, one auxiliary benefit of this procedure is that, given a set of corrupted image observations, and provided that at least one of them is reasonably good, the existence of other more highly degraded observations should not in theory present a significant disruption to the algorithm. In principle, such images are effectively discounted automatically, and each estimated ρ_l can be treated as a score function.

As an example, consider the following empirical comparison. We take one sharp image (the `bridge` image) from Levin's dataset [14], and generate blurry/noisy image pairs for testing. The blurry image is generated by convolving the sharp image with a motion blur kernel at 45 degrees and with motion length of 5 pixels. The noisy image is generated by adding random Gaussian noise to the sharp image, with different standard derivations. The image estimation quality is measured with the Sum of Squared Difference (SSD) metric as defined by [14]. Fig. 3 shows the image

6. In brief, image blur decreases sparsity, hence a sparse prior is needed to favor sharp images. However, blur also reduces image variance as pointed out in [14], which can cause marginally sparse priors such as the ℓ_1 norm to actually favor the blurry solution in areas with fine structure. Fortunately, the ℓ_0 norm and close approximations are insensitive to changes in variance, hence they are indispensable for resolving fine details.

7. We note that, even with this mechanism in place, using a multi-resolution scheme can still be helpful for improving performance. Hence we apply low-resolution results to initialize higher levels for the results reported in Section 5, a standard practice in nearly all deblurring algorithms we are aware of.

8. This phenomena is in some ways similar to certain homotopy sparse estimation schemes (e.g., [5]), where additional hyperparameters are introduced to gradually introduce greater non-convexity into canonical compressive sensing problems, but without any dependence on the noise or other factors. The key difference here with our method is that penalty shape modulation is explicitly dictated by both the noise level λ and the kernels $\mathbf{k}_l$ in an entirely integrated fashion with no heuristic hyperparameters involved.

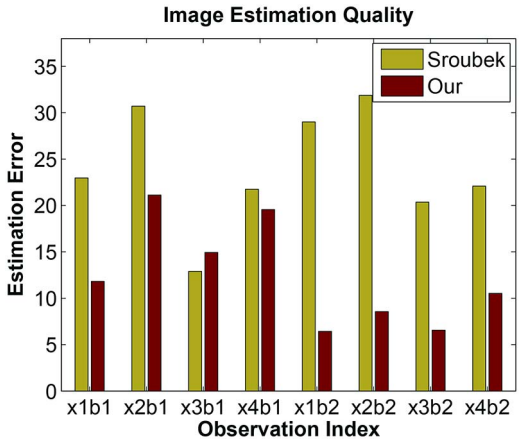

Fig. 4. Error bar plot: Comparison of Šroubek *et al.*'s method [23] and ours on Levin *et al.*'s dataset [14].

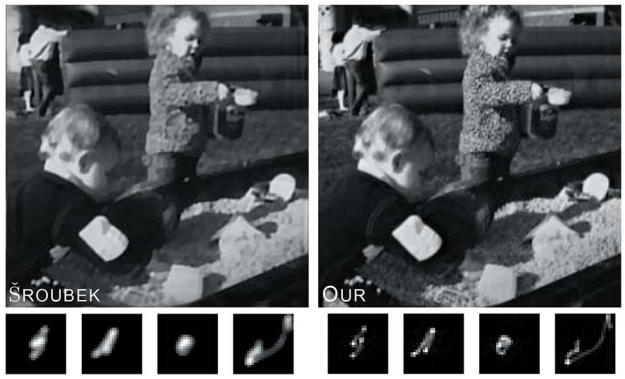

Fig. 5. Recovered image and blur kernels of Šroubek *et al.*'s method [23] and ours on $\{x1, b1\}$, *i.e.*, the first image and kernels 1–4 from Levin *et al.*'s dataset [14].

estimation error curve using the blurry/noisy image pair under different signal-to-noise-ratio (SNR) values for the noisy image. As can be observed from Fig. 3, the estimation error increases as the SNR decreases, but is generally lower than the estimation error from the single blurry image (dashed line in Fig. 3), implying that even highly noisy images can still provide some useful information up to a point. However, when the SNR dips below around -15 or -20 dB, the image estimation quality of the algorithm starts to maintain an almost constant error level, which is at nearly the same level as when using the single blurry image alone. This demonstrates that the inclusion of a highly degraded (essentially junk) image does not substantially derail the algorithm, likely because of the role of 'scoring' different observations using ρ_l as discussed above.

The underlying reason why this desirable scoring phenomena occurs becomes more clear upon closer inspection of (10). Consider the case where we have a junk image $\mathbf{y}_r$, meaning an observation that is either so noisy or so blurry that it conveys almost no information regarding the true, sharp $\mathbf{x}$. The associated data fidelity term $\frac{1}{\lambda_r}\|\mathbf{y}_r - \mathbf{k}_r * \mathbf{x}\|_2^2$ can be minimized in one or more of the following three ways:

(i) λ_r can be increased.
(ii) $\mathbf{k}_r$ can become a large, diffuse kernel allowing it to convert virtually any $\mathbf{x}$ into $\mathbf{y}_r$ (accounting for irrelevant scaling factors, this generally reduces $\|\bar{\mathbf{k}}_l\|_2^2$ for the reasons stated previously).
(iii) Or finally, $\mathbf{x}$ can be directly matched to $\mathbf{y}_r$ (with $\mathbf{k}_r$ a delta kernel).

Both (i) and (ii) naturally cause ρ_r to increase. Moreover, once ρ_r increases, $\log(\rho_r + \gamma)$ converges to something like a modestly large constant regardless of γ (because of the logarithm), and therefore the effective image penalty approximately satisfies

$$h(x, \boldsymbol{\rho}) \equiv \min_{\gamma \geq 0} \sum_{l \neq r} \left[\frac{x^2}{\gamma} + \log(\rho_l + \gamma) \right] \tag{18}$$

at every voxel. Thus, the estimation of $\mathbf{x}$ using all remaining terms is more or less independent of the r-th observation. Provided that other images are of relatively high quality, then some elements of $\boldsymbol{\rho}$ may be small while still maintaining a small value of $\sum_{l \neq r} \frac{1}{\lambda_l}\|\mathbf{y}_l - \mathbf{k}_l * \mathbf{x}\|_2^2$. Moreover, zero-valued elements of $\mathbf{x}$ will naturally drive $h(x, \boldsymbol{\rho})$ towards minus infinity, and thus the overall cost function can be reduced drastically to a large negative value.

Alternatively, suppose (iii) occurs to some extent instead. If $\mathbf{x}$ is at all matched to $\mathbf{y}_r$, then it generally cannot be sparse and $h(x, \boldsymbol{\rho})$ can only be reduced marginally. Hence it is unlikely that the overall cost function will be minimized. Consequently, the natural outcome of a junk observation is a desirable form of image pruning as we have illustrated in Fig. 3.

5 Experimental Results

Using both synthetic data and real-world images, we now compare our algorithm with several *state-of-the-art* multi-image methods from Cai *et al.* [4], Šroubek *et al.* [23], Zhu *et al.* [39] and Chen *et al.* [6] for blurry observations as well as Yuan *et al.* [34] and Whyte *et al.* [28] on blurry/noisy pairs.

5.1 Evaluation on Synthetic Data

We first use the standard test data collected by Levin *et al.* [14] for evaluation, which consists of 4 images of size 255×255 and 8 different blur kernels, giving a total of 32 blurry images. The kernel sizes range from 13×13 to 27×27. The blurry images, ground-truth images, and the ground-truth kernels are also provided. Following the experimental settings in [23], we construct multi-observation test sets with $L = 4$ blurry images by dividing the whole kernel set into two halves: $b1 = \{1 \cdots 4\}$ and $b2 = \{5 \cdots 8\}$. In so doing, 8 multi-observation sets are generated for testing. We then perform blind deblurring using different algorithms on each set. We compare our method with the recent method of Šroubek *et al.* [23], for which the `matlab` implementation is publicly available.[9]

The Sum of Squared Difference (SSD) metric defined in [14] is used for measuring the error between the deblurred and ground-truth images. Results are shown in Fig 4, where the proposed method generates deblurring results that are significantly better on most of the

9. http://zoi.utia.cas.cz/files/fastMBD.zip

Fig. 6. Uniform and non-uniform deblurring. (a) Blurry image pair [4]. (b) Results from Cai *et al.* [4]. (c) Results produced with Šroubek *et al.*'s software [23]. (d) Our results.

test images. The recovered image and blur kernels from both methods for the first test set are shown in Fig. 5. Here we observe that some fine details of the sharp image are not recovered by Šroubek *et al.*'s method (e.g., sand and sweater textures are compromised). In contrast, our approach can recover the blur kernels with high quality without using any explicit sparse prior over the kernel (the kernel will automatically become sparse if the modified curvature of the image penalty is advantageous). By incorporating a sparsity prior over the kernel, which must be distinguished from image sparsity, the results can be further improved (results not shown). Overall, the more refined kernel estimates obtained via the proposed approach translate into more details recovered in the latent images.

5.2 Evaluation on Real-world Images

Blind restoration using multiple observations is a ubiquitous problem, with many potential applications. This section investigates two common scenarios using real-world images.

- *Dual motion deblurring*: The use of two motion-blurried observations for joint blind deblurring [4], [6], [21], [23], [39], and
- *Blurry/Noisy pair restoration*: The use a short-exposure noisy and long-exposure blurry image pair for joint restoration [28], [34].

We emphasis that the reason we evaluate under these somewhat restrictive scenarios separately is primarily for ease of comparison with previous state-of-the-art algorithms that have been explicitly tailored for each specific case. In contrast, our algorithm does not require any modification and can handle both tasks seamlessly in a unified way, and is in this sense more practical.

Dual Motion Deblurring: For dual motion deblurring, we compare with the multi-image methods proposed by Cai *et al.* [4], Šroubek *et al.* [23] with parameters set via consultation with the author, Zhu *et al.* [39] as well as Chen *et al.* [6] on several different real-world images used in previous deblurring work. We first evaluate the relative performance on an image pair from [4] as shown in Fig. 6. The results of Cai *et al.*, Šroubek *et al.*, and our method are also shown in Fig. 6, with the estimated blur kernels displayed in the top-right corner of each image. We observe that the kernel estimations from Cai *et al.* are probably too sparse, as the recovered image suffers from severe ringing artifacts. The deblurred image by Šroubek *et al.* suffers less from ringing than that of Cai *et al.*, although our result has even fewer artifacts. While we do not have access to the ground-truth kernel for real-world images, our kernel estimation appears to be reasonable given the high quality of the estimated sharp image.

Fig. 7 provides further comparison with Šroubek *et al.* on an image pair from [23], as well as with a standard

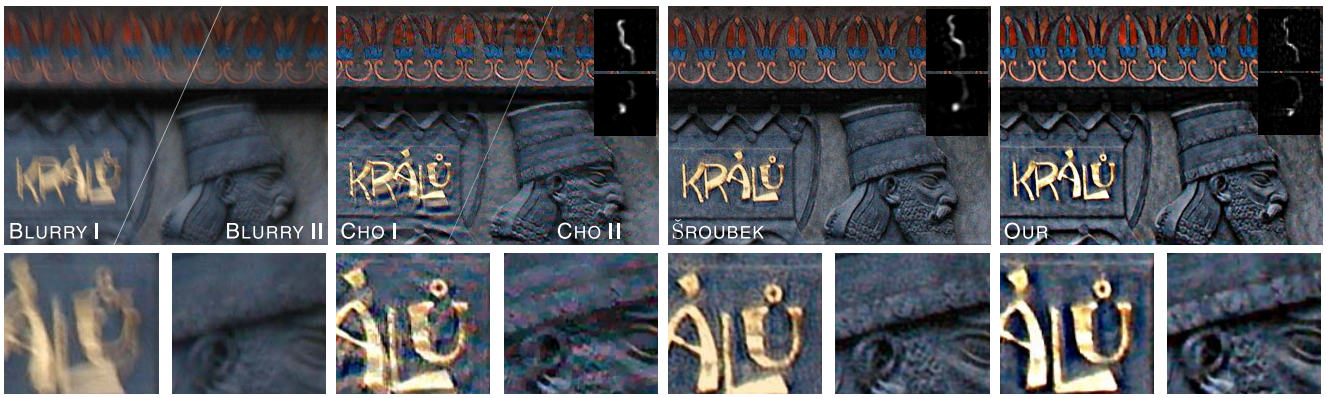

Fig. 7. Dual motion deblurring results. (a) Blurry image pair [23]. (b) Results produced with Cho *et al.*'s software [7]. (c) Results produced with Šroubek *et al.*'s software [23]. (d) Our results.

Fig. 8. Dual exposure deblurring results. (a) Blurry/Noisy image pair [34]. (b) Results produced with Cho *et al.*'s software using the blurry image [7]. (c) Result from Yuan *et al.* [34]. (d) Our results.

single-image method from Cho *et al.* [7] as a benchmark. The kernel estimates from Šroubek *et al.* appear similar to those from Cho *et al.*; however, the associated deblurred image has fewer artifacts. The kernels estimated from the proposed method are more sparse than those of Šroubek *et al.* and the deblurred image has less significant artifacts.

The method of Zhu *et al.* [39] attempts to refine the estimated blur kernels from Šroubek *et al.* [23] via an explicit sparsity penalty. Comparisons between Zhu *et al.*, Šroubek *et al.*, and our approach on an image pair from [39] are shown in Fig. 10. While the kernel estimates from Zhu *et al.* are indeed more compact than those from Šroubek *et al.*, the accuracy is likely still below that of our method. For example, some fine details such as the text on the book cover are not properly recovered. One potential reason for this is that the kernel refining step of Zhu *et al.* relies purely on the kernels estimated via Šroubek *et al.*, without using the observed data. Therefore, although the estimated blur kernels do become less diffuse, they are not necessarily consistent with the observed data, as any error generated in the original kernel estimation step will be inevitably transferred during the kernel refining process. In contrast, our approach can implicitly determine the proper kernel sparsity directly from the data without any secondary rectifications or an explicit sparse prior for the kernel; it therefore appears to be more reliable on these test images.

Further results on another set of blurry images used by Chen *et al.* [6] are shown in Fig. 11. Here we observe that the kernels from Šroubek *et al.* [23] are perhaps not accurate enough, as the final deblurred image has severe ringing artifacts. The kernels estimated by Chen *et al.* [6] are more accurate as the deblurred image has less ringing. However, the recovered image appears blurry and over-smoothed. In contrast, the recovered image by our approach is clean and sharp without any visible artifacts.

Restoration from Blurry/Noisy Pairs: As mentioned previously, our algorithm can be seamlessly applied to images with differing types of degradation extending beyond the typical dual-motion deblurring tasks, e.g., restoration based on blurry/noisy pairs. Although the existing dual-motion deblurring algorithms tested above are no longer directly applicable, alternative approaches have been specifically tailored to work only with a blurry and noisy pair [28], [34], and hence provide a benchmark for comparison.

We first compare with Yuan *et al.* [34] on the blurry/noisy image pair previously used in their paper. The results, including Cho *et al.*'s method [7] as a single-blurry-image baseline, are shown in Fig. 8. Not surprisingly, Yuan *et al.* can generate a restoration result that is of higher quality compared to the result obtained from a single blurry image and Cho *et al.*'s algorithm. Yet the image recovered via our approach is of relatively similar quality to that of Yuan *et al.*; however, we emphasize that our method is at a substantial disadvantage because it has no knowledge that we are dealing with a blurry/noisy pair and it has received no special design for this situation. It is also interesting to point out that the blur kernel estimated for the noisy image is a delta kernel as would be expected if the correct solution were to be found. This reflects the strong generalization ability of our method.

Fig. 9. Dual exposure deblurring results. (a) Blurry/Noisy image pair [28]. (b) Uniform deblurring results from Whyte *et al.* [28]. (c) Non-uniform deblurring result from Whyte *et al.* [28]. (d) Our results.

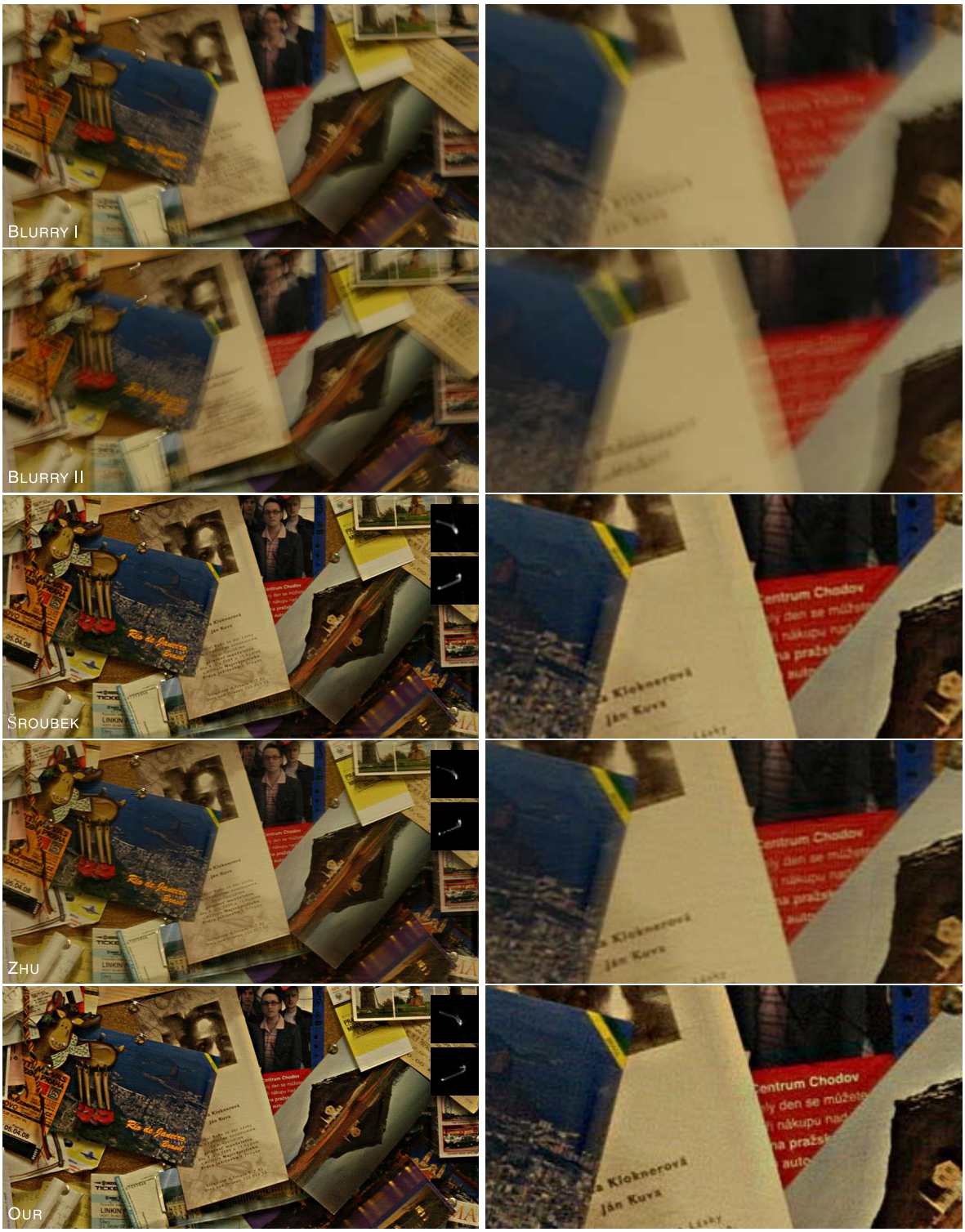

Fig. 10. Dual motion deblurring results. From top to bottom: (a) and (b) Blurry image pair [39]. (c) Results produced with Šroubek *et al.*'s software [23]. (d) Results from Zhu *et al.* [39]. (e) Our results. Zoomed parts of the images from the left column are shown on the right.

Then we compare with two recent blurry/noisy pair-based methods from Whyte *et al.* [28] using images from their paper. Results are shown in Fig. 9. Note that Whyte *et al.*'s non-uniform method does not produce a typical 2D kernel per the standard convolutional model (1), and hence no blur kernel is shown. Again, we observe that our algorithm, without resorting to more complicated observation models or special tuning, performs competitively with algorithms specifically designed to work with a known blurry and noisy pair.

Actually, the framework developed in this work can be naturally extended to handle non-uniform blur (*e.g.*, due to camera shake) to further enhance its ability, by representing $\mathbf{H}_l$ using appropriate basis functions (*e.g.*, projection/homography operator [25]) as we mentioned in our

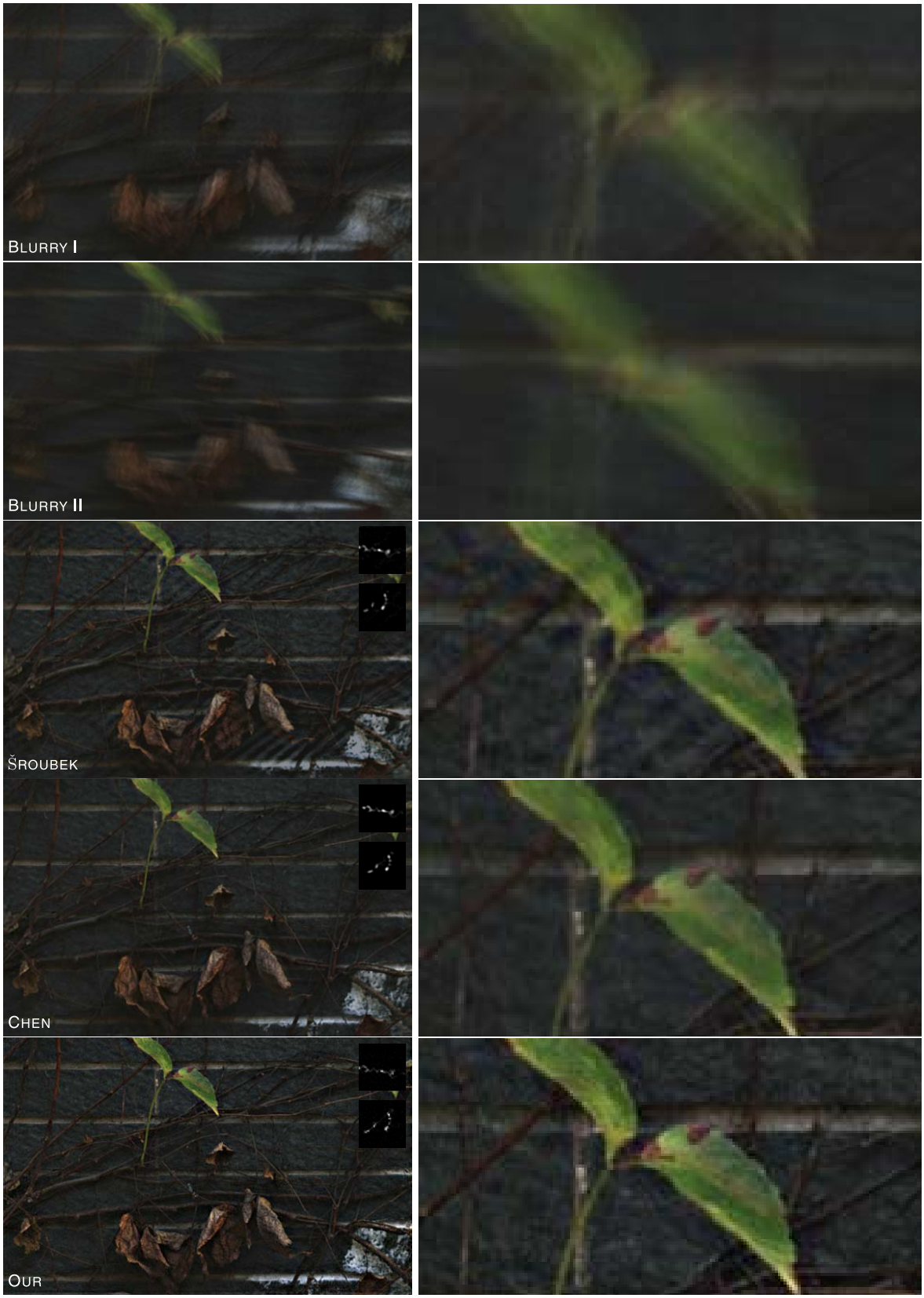

Fig. 11. Dual motion deblurring results. From top to bottom: (a) and (b) Blurry image pair [39]. (c) Results produced with Šroubek *et al.*'s software [23]. (d) Results from Chen *et al.* [6]. (e) Our results. Zoomed parts of the images from the left column are shown on the right.

previous work [35]. The advantage of this generalization can be observed from an example shown in Fig. 12. It is shown that by using the non-uniform blur model, the deblurred image has even fewer artifacts than the result from our model with uniform blur assumption. The kernel patterns recovered reveal that the actual

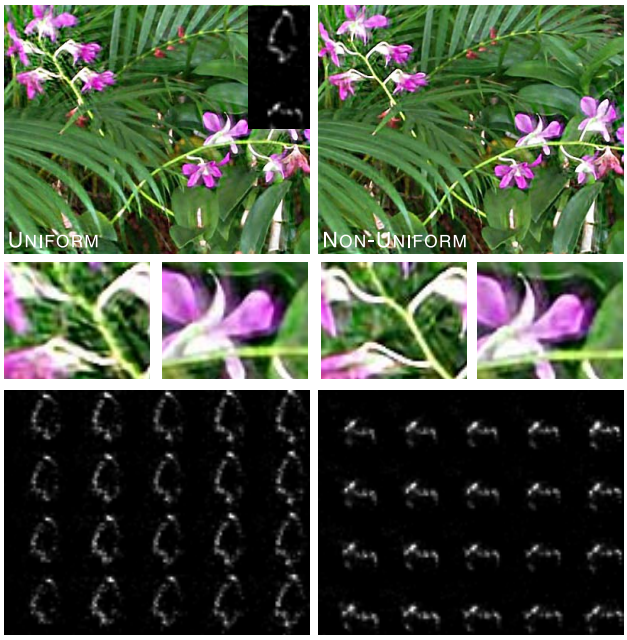

Fig. 12. Uniform vs. non-uniform deblurring (a) Uniform deblurring results (same as Fig. 6-(d)) (b) Non-Uniform deblurring results (c) and (d) Non-uniform blur kernel patterns for the two blurry images (Fig. 6-(a)). The recovered blur kernel patterns indicate that the blurs of the two blurry images are actually non-uniform.

blurring effects for the two blurry images are spatially variant.

6 Algorithm Details

This section provides additional information regarding the origins of the proposed cost function, as well as detailed derivation of the associated update rules for minimizing Algorithm 1. We conclude by discussing a simple modification of the algorithm to handle alternative blur kernel penalty functions.

6.1 Cost Function Derivation

Here we provide a brief derivation of the cost function from (8). Mathematically, the marginalization scheme described in Section 3 requires that we solve

$$\begin{aligned}&\max_{\boldsymbol{\gamma},\mathbf{k},\lambda\geq 0}\int p(\mathbf{y}|\mathbf{x},\mathbf{k},\lambda)\prod_i \mathcal{N}(x_i;0,\gamma_i)d\mathbf{x}\\ &\equiv \min_{\boldsymbol{\gamma},\mathbf{k},\lambda\geq 0}\mathbf{y}^T\left(\lambda\mathbf{I}+\mathbf{H}\boldsymbol{\Gamma}\mathbf{H}^T\right)^{-1}\mathbf{y}+\log\left|\lambda\mathbf{I}+\mathbf{H}\boldsymbol{\Gamma}\mathbf{H}^T\right|\end{aligned} \tag{19}$$

where $\boldsymbol{\Gamma}\triangleq \text{diag}[\boldsymbol{\gamma}]$, where the required integration involves a standard convolution of Gaussians for which closed-form solutions are available. It can be shown [30] using basic linear algebra techniques that

$$\mathbf{y}^T\left(\lambda\mathbf{I}+\mathbf{H}\boldsymbol{\Gamma}\mathbf{H}^T\right)^{-1}\mathbf{y}=\min_{\mathbf{x}}\frac{1}{\lambda}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2+\mathbf{x}^T\boldsymbol{\Gamma}^{-1}\mathbf{x}. \tag{20}$$

Plugging (20) into (19) we have

$$\begin{aligned}&\min_{\boldsymbol{\gamma},\mathbf{k},\lambda\geq 0}\mathbf{y}^T\left(\lambda\mathbf{I}+\mathbf{H}\boldsymbol{\Gamma}\mathbf{H}^T\right)^{-1}\mathbf{y}+\log\left|\lambda\mathbf{I}+\mathbf{H}\boldsymbol{\Gamma}\mathbf{H}^T\right|\\ &=\min_{\mathbf{x},\boldsymbol{\gamma},\mathbf{k},\lambda\geq 0}\frac{1}{\lambda}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2+\mathbf{x}^T\boldsymbol{\Gamma}^{-1}\mathbf{x}+\log\left|\lambda\mathbf{I}+\mathbf{H}\boldsymbol{\Gamma}\mathbf{H}^T\right|\end{aligned}$$

$$=\min_{\mathbf{x},\mathbf{k},\lambda\geq 0}\frac{1}{\lambda}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2+\underbrace{\min_{\boldsymbol{\gamma}\geq 0}\mathbf{x}^T\boldsymbol{\Gamma}^{-1}\mathbf{x}+\log\left|\lambda\mathbf{I}+\mathbf{H}\boldsymbol{\Gamma}\mathbf{H}^T\right|}_{\mathfrak{g}(\mathbf{x},\mathbf{k},\lambda)},$$

directly leading to (8).

6.2 Update Rule Derivation

Algorithm 1 from Section 3 is designed to minimize the objective

$$\begin{aligned}\mathcal{L}(\mathbf{x},\{\mathbf{k}_l,\lambda_l\})&=\sum_{l=1}^{L}\frac{1}{\lambda_l}\|\mathbf{y}_l-\mathbf{k}_l*\mathbf{x}\|_2^2\\ &\quad+(n-m)\log\lambda_l+\mathfrak{g}(\mathbf{x},\{\mathbf{k}_l,\lambda_l\}),\\ \text{s.t.}\quad &\mathbf{k}_l\geq 0,\forall l\in\{1,\ldots,L\},\end{aligned} \tag{21}$$

where

$$\mathfrak{g}(\mathbf{x},\mathbf{k},\lambda)\triangleq\min_{\boldsymbol{\gamma}\geq 0}\sum_{l=1}^{L}\sum_{i=1}^{m}\left[\frac{x_i^2}{\gamma_i}+\log(\lambda_l+\gamma_i\|\bar{\mathbf{k}}_l\|_2^2)\right]. \tag{22}$$

For this purpose we employ a majorization-minimization technique by constructing upper bounds on some of the terms embedded in $\mathfrak{g}$. This conveniently decouples relevant factors and leads naturally to an alternating minimization approach by iteratively solving a series of simple subproblems leading to Algorithm 1.

To begin, we remove the minimization over $\boldsymbol{\gamma}$ from the penalty term in $\mathcal{L}(\mathbf{x},\{\mathbf{k}_l,\lambda_l\})$, resulting in a rigorous upper bound, denoted $\mathcal{L}(\mathbf{x},\boldsymbol{\gamma},\{\mathbf{k}_l,\lambda_l\})$, of the original cost function since

$$\mathcal{L}(\mathbf{x},\boldsymbol{\gamma},\{\mathbf{k}_l,\lambda_l\})\geq\mathcal{L}(\mathbf{x},\{\mathbf{k}_l,\lambda_l\}), \tag{23}$$

where equality is achieved whenever $\boldsymbol{\gamma}=\boldsymbol{\gamma}^{\text{opt}}$. Therefore, to minimize the original objective, we may instead minimize $\mathcal{L}(\mathbf{x},\boldsymbol{\gamma},\{\mathbf{k}_l,\lambda_l\})$ over $\mathbf{x}$, $\boldsymbol{\gamma}$, $\mathbf{k}_l$, and λ_l for all l. This can be accomplished via coordinate descent, meaning we optimize one variable while keeping the others fixed.

x-subproblem: Isolating relevant terms, the latent image $\mathbf{x}$ can be computed using the weighted least squares problem

$$\min_{\mathbf{x}}\sum_{l=1}^{L}\frac{1}{\lambda_l}\|\mathbf{y}_l-\mathbf{k}_l*\mathbf{x}\|_2^2+L\sum_{i=1}^{m}\frac{x_i^2}{\gamma_i}, \tag{24}$$

where the optimal solution is given by the closed-form solution

$$\mathbf{x}^{\text{opt}}=\left[\sum_{l=1}^{L}\frac{\mathbf{H}_l^T\mathbf{H}_l}{L\lambda_l}+\boldsymbol{\Gamma}^{-1}\right]^{-1}\sum_{l=1}^{L}\frac{\mathbf{H}_l^T\mathbf{y}_l}{L\lambda_l}. \tag{25}$$

Here $\mathbf{H}_l$ denotes the convolution matrix corresponding with $\mathbf{k}_l$ and $\boldsymbol{\Gamma}\triangleq\text{diag}(\boldsymbol{\gamma})$.

$\boldsymbol{\gamma}$-subproblem: With other variables fixed, the optimization over each γ_i is separable and thus can be solved independently via

$$\min_{\gamma_i\geq 0}\sum_{l=1}^{L}\left[\frac{x_i^2}{\gamma_i}+\log\left(\lambda_l+\gamma_i\|\bar{\mathbf{k}}_l\|_2^2\right)\right]. \tag{26}$$

Using the fact that

$$\lambda_l + \gamma_i \|\bar{\mathbf{k}}_l\|_2^2 = \lambda_l \gamma_i \left(\frac{1}{\gamma_i} + \frac{\|\bar{\mathbf{k}}_l\|_2^2}{\lambda} \right), \tag{27}$$

we can rewrite (26) equivalently as

$$\min_{\gamma_i \geq 0} \sum_{l=1}^{L} \left[\frac{x_i^2}{\gamma_i} + \log \gamma_i + \log \left(\frac{\|\bar{\mathbf{k}}_l\|_2^2}{\lambda_l} + \gamma_i^{-1} \right) \right], \tag{28}$$

where the irrelevant $\log \lambda_l$ term has been omitted. Because no closed-form solution for (28) is available, we instead use basic principles from convex analysis to form a strict upper bound that will facilitate subsequent optimization. In particular, we use

$$\frac{z_{li}}{\gamma_i} - \phi^*(z_{li}) \geq \log \left(\frac{\|\bar{\mathbf{k}}_l\|_2^2}{\lambda_l} + \gamma_i^{-1} \right), \tag{29}$$

which holds for all $z_{li} \geq 0$ when ϕ^* is defined as the concave conjugate [3] of the concave function $\phi(a) \triangleq \log(\frac{\|\bar{\mathbf{k}}_l\|_2^2}{\lambda_l} + \alpha)$. It can be shown that equality in (29) is achieved when

$$z_{li}^{\text{opt}} = \left. \frac{\partial \phi}{\partial \alpha} \right|_{\alpha = \gamma_i^{-1}} = \frac{1}{\frac{\|\bar{\mathbf{k}}_l\|_2^2}{\lambda_l} + \gamma_i^{-1}}, \forall i, l. \tag{30}$$

Substituting (29) into (28), we obtain the revised $\boldsymbol{\gamma}$-subproblem

$$\min_{\gamma_i \geq 0} \sum_{l=1}^{L} \left[\frac{x_i^2 + z_{li}}{\gamma_i} + \log \gamma_i \right], \tag{31}$$

which gives the following update equation

$$\gamma_i^{\text{opt}} = {x_i}^2 + \frac{\sum_{l=1}^{L} z_{li}}{L}. \tag{32}$$

Although we can always cyclically update each γ_i and z_{li} until (26) is minimized, it is only necessary to update each once to ensure that $\mathcal{L}(\mathbf{x}, \boldsymbol{\gamma}, \{\mathbf{k}_l, \lambda_l\})$ is reduced.

k-subproblem: We will omit the subscript l to simplify notation, recognizing that the following updates must be repeated independently for all observations. Based on (21) and (28), $\mathbf{k}$ can be optimized using the constrained quadratic optimization problem

$$\min_{\mathbf{k} \geq 0} \frac{1}{\lambda} \|\mathbf{y} - \mathbf{W}\mathbf{k}\|_2^2 + \sum_{i=1}^{m} \log \left(\frac{\|\bar{\mathbf{k}}\|_2^2}{\lambda} + \gamma_i^{-1} \right), \tag{33}$$

where $\mathbf{W}$ is the convolution matrix of $\mathbf{x}$. Because there is no closed-form solution, we resort to similar bounding techniques as before, adopting

$$\|\bar{\mathbf{k}}\|_2^2 v_i - \psi^*(v_i) \geq \log \left(\frac{\|\bar{\mathbf{k}}\|_2^2}{\lambda} + \gamma_i^{-1} \right), \tag{34}$$

where the bound holds for all $v_i \geq 0$ when ψ^* is the concave conjugate of the concave function $\psi(\alpha) \triangleq \log(\frac{\alpha}{\lambda} + \gamma_i^{-1})$. Similar to the z updates, equality (34) is achieved with

$$v_i^{\text{opt}} = \left. \frac{\partial \psi_i}{\partial \alpha} \right|_{\alpha = \|\bar{\mathbf{k}}\|_2^2} = \frac{z_i}{\lambda}, \forall i. \tag{35}$$

Plugging (34) into (33), we obtain the revised optimization problem

$$\begin{aligned} \mathbf{k}^{\text{opt}} &= \arg\min_{\mathbf{k} \geq 0} \frac{1}{\lambda} \|\mathbf{y} - \mathbf{W}\mathbf{k}\|_2^2 + \sum_{i=1}^{m} v_i \|\bar{\mathbf{k}}\|_2^2 \\ &= \arg\min_{\mathbf{k} \geq 0} \|\mathbf{y} - \mathbf{W}\mathbf{k}\|_2^2 + \sum_j k_j^2 \left(\sum_{i=1}^{m} z_i \bar{I}_{ji} \right). \end{aligned} \tag{36}$$

The second equality follows once we reincorporate boundary conditions (see Section 3), meaning that $\|\bar{\mathbf{k}}\|_2^2 \triangleq \sum_j k_j^2 \bar{I}_{ji}$ is now an i-dependent quantity. This then reveals that (36) possesses a standard, weighted ℓ_2-norm penalty on $\mathbf{k}$. As a simple convex program, there exist many high-performance algorithms for solving (36).

λ-subproblem: As for $\mathbf{k}$ above, we omit the subscript l and adopt a related bounding strategy for estimating the noise level across all observed images. With other variables fixed, the required optimization over λ is

$$\min_{\lambda \geq 0} \frac{1}{\lambda} \|\mathbf{y} - \mathbf{k} * \mathbf{x}\|_2^2 + n \log \lambda + \sum_{i=1}^{m} \log(\frac{\|\bar{\mathbf{k}}\|_2^2}{\lambda} + \gamma_i^{-1}). \tag{37}$$

As there is no closed form solution, we use the bound

$$\frac{\beta}{\lambda} - \phi^*(\beta) \geq \sum_{i=1}^{m} \log \left(\beta \|\bar{\mathbf{k}}\|_2^2 + \gamma_i^{-1} \right) \tag{38}$$

which holds for all $\beta \geq 0$ when ϕ^* is the concave conjugate of $\phi(\theta) \triangleq \sum_{i=1}^{n} \log \left(\theta \|\bar{\mathbf{k}}\|_2^2 + \gamma_i^{-1} \right)$. Equality is achieved with

$$\beta^{\text{opt}} = \left. \frac{\partial \phi}{\partial \beta} \right|_{\beta = \lambda^{-1}} = \sum_{i=1}^{m} \frac{\|\bar{\mathbf{k}}\|_2^2}{\frac{\|\bar{\mathbf{k}}\|_2^2}{\lambda} + \gamma_i^{-1}}. \tag{39}$$

Plugging (38) into (37), we obtain

$$\min_{\lambda \geq 0} \frac{1}{\lambda} \|\mathbf{y} - \mathbf{k} * \mathbf{x}\|_2^2 + n \log \lambda + \frac{\beta}{\lambda}, \tag{40}$$

leading to the noise level update

$$\lambda^{\text{opt}} = \frac{\|\mathbf{y} - \mathbf{k} * \mathbf{x}\|_2^2 + \beta}{n}. \tag{41}$$

By iteratively cycling through each of the above subproblems, we arrive at Algorithm 1. From a practical standpoint, we also find that a multi-scale estimation scheme is beneficial following nearly all recent deblurring work, *e.g.*, [9], [15], [23].

6.3 Extensions

The proposed framework is very general, and we can easily place various, possibly structured sparse priors over both $\mathbf{x}$ and $\mathbf{k}$ [29]. As a simple illustrative example, we can replace the ℓ_2 kernel norm with $\|\mathbf{k}\|_p^p$ and $0 < p < 2$. Implementation is straightforward and only requires incorporation of the additional bound

$$\|\mathbf{k}\|_p^p \leq \sum_j \left[\frac{k_j^2}{\phi_j} + \frac{(2-p)}{p} \left(\frac{p}{2} \right)^{\frac{2}{2-p}} \phi_j^{\frac{p}{2-p}} \right], \tag{42}$$

with equality if and only if $\phi_j = k_j^{2-p}2/p$. The boundary problem is not considered in here for simplicity; however, this can be easily handled using the previously introduced diagonal indicator matrix $\bar{\mathbf{I}}$. More carefully designed prior models for the kernel can also be incorporated, *e.g.*, [8], [11], [12].

Note that although our framework nominally includes a quadratic penalty on $\mathbf{k}$, which in isolation would favor a non-sparse kernel estimate, the coupling mechanism intrinsic to our model provides a strong, sparsity-inducing counter-effect allowing sparse kernels to be estimated nonetheless if necessary. Simply put, when $\mathbf{k}$ becomes sparse, while the $\|\mathbf{k}\|_2^2$ factor may increase somewhat, there is the potential to reduce the overall cost dramatically as a consequence of Theorem 2 provided the sparse kernel can yield a sparse image $\mathbf{x}$. However, if no sparse image is possible, then a more diffuse kernel will be favored, which is why the no-blur delta kernel is not a significant risk factor for our algorithm.

7 Proofs

This section will provide the proofs of the theorems given in Section 4, which helps to understand further some theoretical aspects of the proposed model.

7.1 Proof of Theorem 1

The proof follows by noting that each $\log(\rho_l + \gamma)$ represents a concave, non-decreasing function of γ, and therefore the sum over l of these terms is also concave and non-decreasing. Functions of the variational form $\min_{\gamma>0} z^2/\gamma + f(\gamma)$, where $f(\gamma)$ is a concave, non-decreasing function of γ can be shown to be concave functions of $|z|$ via a small extension of Theorem 3 in [30]. □

7.2 Proof of Theorem 2

Property (1) is very straightforward. As $z \to \infty$, the optimizing γ will become arbitrarily large regardless of the value of $\boldsymbol{\rho}$. In the regime where γ is sufficiently large, the difference between the terms $\log(\gamma + \rho_l^1)$ and $\log(\gamma + \rho_l^2)$ must converge to zero. It then follows that the difference between the corresponding minimizing γ values, and therefore the cost function difference, converges to zero.

For property (2), we will assume for simplicity of exposition that for all $l \neq j$, $\rho_l^2 = \rho_l^1$, and that $\rho_j^2 > \rho_j^1$. The more general scenario naturally follows. We note that it will always be the case that $h_{\boldsymbol{\rho}^2}(z) - h_{\boldsymbol{\rho}^1}(z)$ for any z. This occurs because for all γ, $\log(\gamma + \rho_j^2) > \log(\gamma + \rho_j^1)$. Therefore if

$$\gamma_2^* \triangleq \arg\min_{\gamma} \frac{z}{\gamma} + \log(\gamma + \rho_j^2) + \psi(\gamma), \tag{43}$$

where $\psi(\gamma) \triangleq \sum_{l\neq j} \log(\gamma + \rho_l)$ (the superscript 1 or 2 is irrelevant here since the values are equal), then

$$h_{\boldsymbol{\rho}^2}(z) > \frac{z}{\gamma_2^*} + \log(\gamma_2^* + \rho_j^1) + \psi(\gamma_2^*) > h_{\boldsymbol{\rho}^1}(z). \tag{44}$$

The minimizing value of γ_1^* needed to produce the second inequality will always satisfy $\gamma_1^* < \gamma_2^*$. This occurs because

$$\begin{aligned}\gamma_1^* &= \arg\min_{\gamma} \frac{z}{\gamma} + \log(\gamma + \rho_j^1) + \psi(\gamma) \\ &= \arg\min_{\gamma} \frac{z}{\gamma} + \log(\gamma + \rho_j^2) + \psi(\gamma) + \log\left(\frac{\gamma + \rho_j^1}{\gamma + \rho_j^2}\right).\end{aligned}$$

The last term, which is monotonically increasing from $\log(\rho_j^1/\rho_j^2) < 0$ to zero, implies that there is always an extra monotonically increasing penalty on γ, when $\rho_j^1 < \rho_j^2$. Since we are dealing with continuous functions here, the minimizing γ will therefore necessarily be smaller. Using basic results from convex analysis and conjugate duality, it can be shown that the minimizing $(\gamma_1^*)^{-1}$ represents the gradient of $h_{\boldsymbol{\rho}^1}(z)$ with respect to z (and likewise for γ_2^*), and we know that this gradient will always be a positive, non-increasing function. We may therefore also infer that $h'_{\boldsymbol{\rho}^1}(z) > h'_{\boldsymbol{\rho}^2}(z)$ at any point z.

We now consider a second point $z' > z$. Because the gradient at every intermediate point moving from $h_{\boldsymbol{\rho}^1}(z)$ to $h_{\boldsymbol{\rho}^1}(z')$ is greater than the associated gradients moving from $h_{\boldsymbol{\rho}^2}(z)$ to $h_{\boldsymbol{\rho}^2}(z')$, it must be the case that $h_{\boldsymbol{\rho}^1}$ increased at a faster rate than $h_{\boldsymbol{\rho}^2}$, and so it follows that

$$h_{\boldsymbol{\rho}^2}(z) - h_{\boldsymbol{\rho}^1}(z) > h_{\boldsymbol{\rho}^2}(z') - h_{\boldsymbol{\rho}^1}(z'), \tag{45}$$

thus completing the proof. □

8 Conclusion

By utilizing a novel penalty function that couples the latent sharp image, blur kernels, and noise variances in a theoretically well-motivated way, this paper describes a unified multi-image blind deconvolution algorithm applicable for recovering a latent, high-quality image from a given set of degraded (blurry, noisy) observations, without any specific modifications for different types of degradations. Moreover, it automatically adapts to the quality of each observed image, allowing higher quality images to dominate the estimation process when appropriate. Experimental evaluations validate the proposed method in different multi-image restoration scenarios.

Several generalizations of our algorithm also hold promise. Currently, translation mis-alignments between observed images are already handled seemlessly since each learned kernel can optimally adapt its position to compensate for any shifts. For more complex mis-alignments such as general rotations, we can either rectify the observations up to a translation before restoration [23], [33] or potentially embed an alignment process directly into Algorithm 1 using techniques similar to [19]. Actually, with the non-uniform extension, the alignment and deblurring can be achieved jointly, which offers much promise for handling multi-image deblurring problem. Furthermore, some of the analysis we conducted for blind deconvolution may well apply to other related problems such as sparse dictionary learning, convolutional factor analysis, and feature learning. We will investigate these topics in our future work.

ACKNOWLEDGMENTS

We would like to thank the Associate Editor and all the reviewers for their useful comments and suggestions. We would also like to thank F. Šroubek for the help in producing some results of his method included in this paper. This work was supported in part by NSF-China (61231016).

REFERENCES

[1] A. K. Agrawal, Y. Xu, and R. Raskar, "Invertible motion blur in video," *ACM Trans. Graphics*, vol. 28, no. 3, Article 95, 2009.
[2] S. D. Babacan, R. Molina, M. N. Do, and A. K. Katsaggelos, "Bayesian blind deconvolution with general sparse image priors," in *Proc. ECCV*, Florence, Italy, 2012.
[3] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.
[4] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "Blind motion deblurring using multiple images," *J. Comput. Phys.*, vol. 228, no. 14, pp. 5057–5071, 2009.
[5] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. ICASSP*, Las Vegas, NV, USA, 2008.
[6] J. Chen, L. Yuan, C.-K. Tang, and L. Quan, "Robust dual motion deblurring," in *Proc. IEEE CVPR*, Anchorage, AK, USA, 2008.
[7] S. Cho and S. Lee, "Fast motion deblurring," in *Proc. SIGGRAPH Asia*, New York, NY, USA, 2009.
[8] T. S. Cho, S. Paris, B. K. P. Horn, and W. T. Freeman, "Blur kernel estimation using the Radon transform," in *Proc. IEEE CVPR*, Providence, RI, USA, 2011.
[9] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," in *Proc. SIGGRAPH*, New York, NY, USA, 2006.
[10] K. Furuya and A. Kataoka, "Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 15, no. 5, pp. 1579–1591, Jul. 2007.
[11] W. Hu, J. Xue, and N. Zheng, "PSF estimation via gradient domain correlation," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 386–392, Jan. 2012.
[12] T. Kenig, Z. Kam, and A. Feuer, "Blind image deconvolution using machine learning for three-dimensional microscopy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2191–2204, Dec. 2010.
[13] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Proc. IEEE CVPR*, Providence, RI, USA, 2011.
[14] A. Levin, Y. Weiss, F. Durand, and W. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Proc. IEEE CVPR*, Miami, FL, USA, 2009.
[15] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Efficient marginal likelihood optimization in blind deconvolution," in *Proc. IEEE CVPR*, Providence, RI, USA, 2011.
[16] W. Li, J. Zhang, and Q. Dai, "Exploring aligned complementary image pair for blind motion deblurring," in *Proc. IEEE CVPR*, Providence, RI, USA, 2011.
[17] J. W. Miskin and D. J. C. MacKay, "Ensemble learning for blind image separation and deconvolution," in *Proc. Adv. ICA*, 2000.
[18] J. A. Palmer, D. P. Wipf, K. Kreutz-Delgado, and B. D. Rao, "Variational EM algorithms for non-Gaussian latent variable models," in *Proc. NIPS*, 2006.
[19] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2233–2246, Nov. 2012.
[20] H. Pozidis and A. P. Petropulu, "Cross-correlation based multichannel blind equalization," in *Proc. 8th IEEE SSAP*, Corfu, Greece, 1996.
[21] A. Rav-Acha and S. Peleg, "Two motion blurred images are better than one," *Pattern Recognit. Lett.*, vol. 26, no. 3, pp. 311–317, Feb. 2005.
[22] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," in *Proc. SIGGRAPH*, New York, NY, USA, 2008.
[23] F. Sroubek and P. Milanfar, "Robust multichannel blind deconvolution via fast alternating minimization," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1687–1700, Apr. 2012.
[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.*, vol. 32, no. 6, pp. 1012–1028, Jun. 2010.
[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.*, vol. 33, no. 8, pp. 1603–1618, Aug. 2011.
[26] M. Temerinac-Ott *et al.*, "Multiview deblurring for 3-D images from light-sheet-based fluorescence microscopy," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1863–1873, Apr. 2012.
[27] D. Tzikas, A. Likas, and N. P. Galatsanos, "Variational Bayesian sparse kernel-based blind image deconvolution with Student's-t priors," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 753–764, Apr. 2009.
[28] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce, "Non-uniform deblurring for shaken images," *Int. J. Comput. Vis.*, vol. 98, no. 2, pp. 168–186, 2012.
[29] D. P. Wipf, "Sparse estimation with structured dictionaries," in *Proc. NIPS*, 2011, pp. 2016–2024.
[30] D. P. Wipf, B. D. Rao, and S. S. Nagarajan, "Latent variable Bayesian models for promoting sparsity," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6236–6255, Sept. 2011.
[31] D. P. Wipf and H. Zhang, *Revisiting Bayesian Blind Deconvolution* [Online]. Available: http://arxiv.org/abs/1305.2362, 2013.
[32] L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," in *Proc. ECCV*, Heraklion, Greece, 2010.
[33] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum, "Blurred/non-blurred image alignment using sparseness prior," in *Proc. IEEE ICCV*, Rio de Janeiro, Brazil, 2007.
[34] L. Yuan, J. Sun, and H.-Y. Shum, "Image deblurring with blurred/noisy image pairs," in *Proc. SIGGRAPH*, New York, NY, USA, 2007.
[35] H. Zhang, D. P. Wipf, and Y. Zhang, "Multi-image blind deblurring using a coupled adaptive sparse prior," in *Proc. IEEE CVPR*, Portland, OR, USA, 2013.
[36] 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 *Proc. IEEE ICCV*, Barcelona, Spain, 2011.
[37] Y. Zhang and H. H. Asada, "Blind system identification of non-coprime multichannel systems and its application to noninvasive cardiovascular monitoring," *J. Dyn. Syst., Meas., Control*, vol. 126, no. 4, pp. 834–847, 2004.
[38] C. Zhou, S. Lin, and S. Nayar, "Coded aperture pairs for depth from defocus and defocus deblurring," *Int. J. Comput. Vis.*, vol. 93, no. 1, pp. 53–72, 2011.
[39] X. Zhu, F. Sroubek, and P. Milanfar, "Deconvolving PSFs for a better motion deblurring using multiple images," in *Proc. 12th ECCV*, Berlin, Germany, 2012.
[40] S. Zhuo, D. Guo, and T. Sim, "Robust flash deblurring," in *Proc. IEEE CVPR*, San Francisco, CA, USA, 2010.

Haichao Zhang received the B.S. degree in computer science from the Northwestern Polytechnical University, Xi'an, China. He is currently pursuing the Ph.D. degree with the School of Computer Science, Northwestern Polytechnical University. From 2009 to 2011, he was with the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL, USA, as a visiting student. His current research interests include sparse modeling techniques and their applications in image/video restoration and recognition tasks. He received the ICCV 2011 Best Student Paper Award.

David Wipf received the B.S. degree with highest honors in electrical engineering from the University of Virginia, Charlottesville, VA, USA, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of California, San Diego, CA, USA. For the latter, he was an NSF Fellow in Vision and Learning in Humans and Machines. Later, he was an NIH Post-Doctoral Fellow at the Biomagnetic Imaging Lab, University of California, San Francisco, CA, USA. Since 2011, he has been with the Visual Computing Group at Microsoft Research in Beijing, China. His current research interests include applying Bayesian learning techniques to sparse estimation and rank minimization with application to problems in signal/ image processing and computer vision. He is the recipient of numerous awards including the 2012 Signal Processing Society Best Paper Award, the Biomag 2008 Young Investigator Award, and the 2006 NIPS Outstanding Student Paper Award. He is currently on the Editorial Board of the *Journal of Machine Learning Research*.

Yanning Zhang received the B.S. degree from the Department of Electronic Engineering, Dalian University of Technology, Dalian, China, in 1988, the M.S. degree from the School of Electronic Engineering, and the Ph.D. degree from the School of Marine Engineering, Northwestern Polytechnical University, Xian, China, in 1993 and 1996, respectively. She is currently a Professor at the School of Computer Science, Northwestern Polytechnical University. Her current research interests include computer vision and pattern recognition, image and video processing, and intelligent information processing. Dr. Zhang was the Organization Chair of the Asian Conference on Computer Vision 2009, and served as the program committee chairs of several international conferences.