Handling Motion Blur in Multi-Frame Super-Resolution

Ziyang Ma[1] Renjie Liao[2] Xin Tao[2] Li Xu[2] Jiaya Jia[2] Enhua Wu[1,3]

[1]University of Chinese Academy of Sciences &
State Key Lab. of Computer Science, Inst. of Software, CAS

[2]The Chinese University of Hong Kong
[3]FST, University of Macau

Abstract

Ubiquitous motion blur easily fails multi-frame super-resolution (MFSR). Our method proposed in this paper tackles this issue by optimally searching least blurred pixels in MFSR. An EM framework is proposed to guide residual blur estimation and high-resolution image reconstruction. To suppress noise, we employ a family of sparse penalties as natural image priors, along with an effective solver. Theoretical analysis is performed on how and when our method works. The relationship between estimation errors of motion blur and the quality of input images is discussed. Our method produces sharp and higher-resolution results given input of challenging low-resolution noisy and blurred sequences.

1. Introduction

Multi-frame super-resolution (MFSR) refers to the process of estimating a high-res image from a sequence of low-res observations. It is a fundamental task in computer vision and image processing. It is also of great value in revealing important information, such as text or fine details, from low-quality surveillance or mobile phone videos.

Most previous MFSR methods made assumptions on noise and point spread function (PSF), and may not handle well images under severe quality degradation. One important type of degradation is motion blur caused by camera shake or fast object motion especially in dim light. When the region of interest (ROI), e.g., text and logo, is with a small size, even slight motion blur can be sufficiently influential. An example is shown in Fig. 1.

Albeit common in videos, motion blur has not been sufficiently discussed for MFSR in literatures. Straightforward preprocessing of each input image using existing single/multiple image deblurring techniques [19, 14, 6, 7, 32, 30] could introduce visual artifacts (Fig. 1(a)) and/or yield incomplete blur removal (Fig. 1(b)). Further, the low resolution of input images makes blind deconvolution [5, 28, 30] hardly find enough strong edges for its kernel estimation, as illustrated in Fig. 1(a). Therefore, image deblurring strategies do not directly suit MFSR.

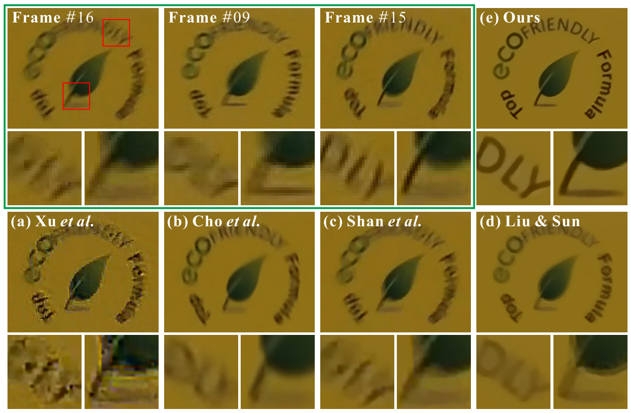


Figure 1. Multi-frame super-resolution (SR) results on a real video sequence. Green box: Input frames (150 × 120) directly cropped from a video captured by an iPhone. Three clearest frames are shown. Motion blur and compression artifacts are present. (a) Result of single image deblurring [30]. (b) Result of video deblurring [7]. (c) Result of video upsampling [19]. (d) MFSR result [16]. (e) Our result (×3).

In addition, state-of-the-art MFSR methods [26, 16] model blurriness for anti-aliasing, but not the inherent motion blur. Most SR methods, either single- or multiple-image ones, also assume that the underlying blur kernel is known, or has a simple analytic form, which is hardly satisfied in real-world cases. In Fig. 1(d), when motion blur is not well handled, the result is accordingly blurred, and contains false edges near the leaf.

To deal with motion blur, another intuitive option is to run MFSR only on manually selected 'clear' frames. We produce the result in Fig. 1(d) where false edges still exist because there is no completely motion-blur-free frame in the sequence.

Our solution Our technical contribution is threefold. First, we propose a system to estimate motion blur and the high-res image with quality feedback and control. Second, based on the observation that in typical motion blurred videos, the same region is not equally blurred across frames, a temporal region selection scheme is devised to select informative structure from each frame. These regions are modeled using latent variables and are effectively solved for in an EM framework. Third, to suppress noise, we pursue spatial sparsity based on a family of penalties, along with an effective solver. Experiments show that our method yields not only reasonable blur estimate, but also visually more compelling restoration results on challenging sequences that cannot be well handled in previous work (Fig. 1(e)).

On the theory side, based on the Cramer-Rao lower bound [12], we provide understanding on the relationship between the restoration error and motion blur. The encouraging conclusion is that more images, even degraded by blur, could induce better SR results.

2. Related Work

Single/Multi-frame SR Single-frame SR aims to recover a high-res image from one low-res input. Extensive research has been done. Most methods focused on developing image priors [19, 25]. Recently, Efrat *et al.* [8] demonstrated the importance of estimating anti-aliasing kernels. Michaeli and Irani [17] proposed estimating the optimal convolution kernel using internal statistics of natural images [33].

Multi-frame SR was also studied [18] since the seminal work of Tsai and Huang [27]. Early registration [11] and image priors [9] were adopted. The parameters of convolution and registration are either assumed known or in parametric forms, which is simplistic for many natural videos.

For generalization and with advanced strategies, Takeda *et al.* [26] used 3D kernel regression to avoid explicit motion estimation. Liu and Sun [16] proposed a Bayesian approach via jointly estimating optical flow, blur kernel, noise, and latent high-res frames. Sunkavalli *et al.* [22] generated a single high-quality image from a video clip in an importance-based framework that weights the contribution of each pixel. Several hand-crafted weights were proposed. These methods do not consider the influence of motion blur. Zhang *et al.* [31] jointly performed image alignment, deblurring and SR via an assumption of projective motion path.

Single/Multi-image deblurring In single image blind deconvolution, several methods [5, 28, 30] employed prediction of sharp edges in early stages of kernel estimation. Tai *et al.* [24] discussed the influence of noise in blur estimation, and deblurred noisy images in a synergistic manner. Cho *et al.* [6] developed a probabilistic approach to handle outliers (*e.g.*, saturated pixels) in single-image deconvolution. A pixel-wise latent binary mask was constructed and distributed according to a spatially independent prior. We in this paper consider a temporally relative sharpness prior for selecting clear regions.

In multi-image deblurring, Zhu *et al.* [32] refined the rough PSF estimate. Tai *et al.* [23] deblurred high-res videos at a low frame rate with the help of simultaneously captured low-res videos at a high frame rate. Cho *et al.* [4] proposed a blind deconvolution algorithm, which transforms the PSF estimation problem into image registration. Cai *et al.* [3] introduced curvelet decomposition for kernels to increase robustness in blurry image alignment. Li *et al.* [13] designed a camera system to simultaneously capture two aligned images with known kernel relationship to help its estimation.

Based on videos, Li *et al.* [14] estimated a sharp panoramic image from motion-blurred frames. The blur kernels are determined from the parameterized homography. Our work takes as input a set of low-res degraded frames for super-resolution in a non-parametric manner.

For video deblurring, Cho *et al.* [7] replaced blurry regions by sharp ones in nearby frames. We deal with more challenging sequences that no clear frame exists. Instead of estimating affine blur for patch matching between blurred and sharp regions, our method is a reconstruction-based approach with newly estimated kernels in iterations.

3. SR Model

We first briefly describe the model. Given a set of low resolution images $\Omega = \{I^L_{-N}, \cdots, I^L_0, \cdots, I^L_N\}$, multi-frame SR aims to estimate a high-res image I corresponding to I^L_0. We follow the expression of [9] and write

$$I^L_i = SK_iF_{0\to i}I + n \quad \text{where } i = -N, \cdots, N. \tag{1}$$

Here, I is a vector representing the latent high-res image. $F = \{F_{0\to -N}, \cdots, F_{0\to 0}, \cdots, F_{0\to N}\}$ is a set of warping matrices corresponding to the motion from I to every other frame. Matrices S and K_i correspond to down-sampling and filtering operations. With motion blur, we denote each K_i as $K_i = K_aK_{b_i}$, where K_a is the anti-aliasing convolution and K_{b_i} is the motion blur kernel. n is the noise.

The latent variables can be estimated in the MAP framework [16] as

$$\begin{aligned}\{I, K, F\} &= \underset{I,K,F}{\arg\max}\, P(I, K, F|\Omega) \\ &= \underset{I,K,F}{\arg\max}\, P(I)P(K)P(F)P(\Omega|I, K, F).\end{aligned} \tag{2}$$

To robustly handle degenerated low-res inputs in the presence of motion blur, the key of our approach is to introduce a binary latent variable $Z = \{Z_{-N}, \cdots, Z_0, \cdots, Z_N\}$ to classify each pixel in each input image as either useful ($Z = 1$) or useless ($Z = 0$). The rationale is to exclude pixels that are largely blurred compared to other temporally corresponding ones. We will show in experiments that

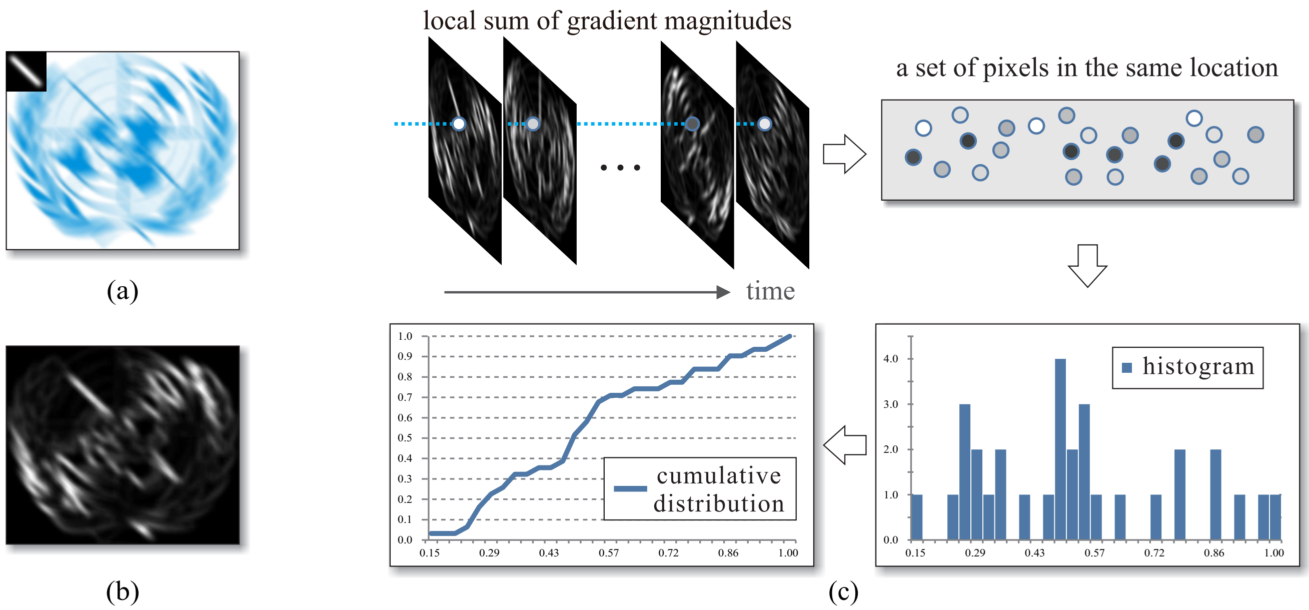


Figure 2. Illustration of the sharpness prior. (a) One synthetic input image with its blur kernel at the top-left corner. (b) Visualization of the local sharpness measure (Eq. (8)) for the image in (a). (c) Histogram of the local sharpness values for the pixels in the same location across all frames, and its corresponding cumulative distribution.

these blurred edges could mislead kernel estimation in MF-SR, while a suitable temporal selection process of clear pixels can form sharp structures beneficial to kernel estimation.

Similar observations have also been presented in single image deblurring that sharp edges are of vital importance to reliable kernel estimation [30]. In addition, outliers such as saturated pixels are also classified as useless since they often violate the image formation model. Following [6], we write Eq. (2) as

$$\{I,K,F\} = \underset{I,K,F}{\arg\max}\, P(I) \prod_{i=-N}^{N} [P(K_i)P(F_{0\to i}) \sum_{Z_i} P(I_i^L, Z_i|I,K,F)]. \tag{3}$$

We will show in Sec. 4.3 that with a suitably guided iterative scheme, a simple Gaussian regularizer is enough to estimate kernel, expressed as

$$P(K_i) \propto \exp\{-\xi \, \|K_i\|_F^2\}. \tag{4}$$

The motion prior $P(F_{0\to i})$ adopts the classical optical flow regularizer [16, 21]:

$$P(F_{0\to i}) \propto \exp\{-\psi \|\nabla F_{0\to i}\|_1\}. \tag{5}$$

Given spatially independent noise, we perform pixel-wise decomposition:

$$\begin{aligned} P(I_i^L, Z_i|I,K,F) &= \prod\nolimits_p P(I_{i,p}^L, Z_{i,p}|I,K,F). \\ &= \prod\nolimits_p P(I_{i,p}^L|Z_{i,p},I,K,F)P(Z_{i,p}|I,K,F). \end{aligned} \tag{6}$$

Here, p indexes pixels. The reconstruction error is modeled by a Laplacian distribution [16] for those informative pixels. The rest, without preference, is just set as uniform [6]:

$$P(I_{i,p}^L|Z_{i,p},I,K,F) \propto \begin{cases} \exp\{-\lambda \, |D_{i,p}|\} & \text{if } Z_{i,p}=1 \\ 1 & \text{otherwise} \end{cases} \tag{7}$$

Here we define the error $D_i = SK_iF_{0\to i}I - I_i^L$ for notation simplification.

3.1. Temporal Relative Sharpness Prior

To exclude pixels that are severely blurred, we introduce a new prior $P(Z_{i,p}|I,K,F)$ in Eq. (6). This is based on the observation that corresponding regions are not always similarly blurred across frames. Edges might be preserved or eliminated depending on whether they are along the blur direction or not, as shown in Fig. 1 and Fig. 2(a).

All pixel values are normalized to $[0,1]$. To measure the sharpness of each pixel in image I_i^L relative to I_j^L ($j \neq i$), we first register frames. This is done via estimating homography matrices using RANSAC with a combination of SURF [1] and KLT [20] features, followed by a TV-L1 optical flow estimation [2, 21]. This scheme is robust against small or median blur. Let J_i be the i-th registered image, and $J_{i,p}$ denote its p-th pixel. We adopt the simple normalized local sum of gradient magnitudes to measure the sharpness as

$$V_{i,p} = \frac{\sum_{q\in\mathcal{N}(p)} \|\nabla J_{i,q}\|_1}{\sum_{j=-N}^{N}\sum_{q\in\mathcal{N}(p)} \|\nabla J_{j,q}\|_1 + \varepsilon}, \tag{8}$$

where $\mathcal{N}(p)$ is the set of spatially neighboring pixels of p. Here, the numerator is the local sum of gradient mag-

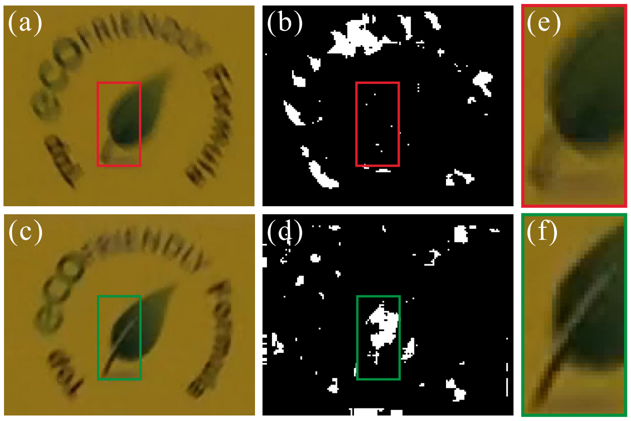


Figure 3. Intermediate latent masks. (a) and (c) are two of the input frames as in Fig. 1. (b) and (d) are the corresponding intermediate latent masks. (e) and (f) are close-ups of (a) and (c) respectively. The masks identify relatively sharp regions.

nitudes, and the denominator is to normalize values across all frames at the same pixel position. Eq. (8) yields a large value when there are sharp edges around pixel p, as shown in Fig. 2(b).

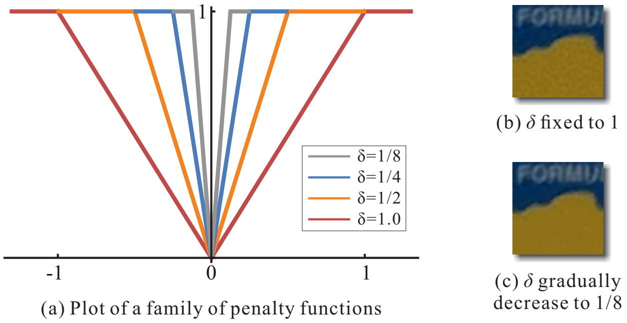
To exclude pixels with small $V_{i,p}$ relative to other $V_{j,p}$ for $j \neq i$, during MFSR, we treat the sharpness values as random variables in $[0,1]$. We investigate the empirical distribution of these sharpness values at a specific location p across all frames, as exemplified in the bottom-right figure of Fig. 2(c).

Then we denote the corresponding cumulative distribution function as $W_p(x)$ ($x \in [0,1]$), and let $W_{i,p} = W_p(V_{i,p})$ (bottom-left of Fig. 2(c)). A small $W_{i,p}$ indicates that pixel p in frame i is not as clear as the same pixel in many other frames. It suggests a small chance that pixel p is informative in SR. Hence we define the prior to follow a conditional Bernoulli distribution as

$$P(Z_{i,p}=1|I,K,F) \propto \begin{cases} \exp\{-\gamma/W_{i,p}\} & \text{if } SK_iF_{0\to i}I \in [0,1] \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

If $SK_iF_{0\to i}I \in [0,1]$, the probability is proportional to its relative sharpness. In this case, we define $P(Z_{i,p} = 0|I,K,F) \propto \exp\{-\gamma\beta\}$ for normalization, where $\beta > 0$, and γ is a weighting parameter. If $SK_iF_{0\to i}I \notin [0,1]$, the pixel is out-of-range. Hence we do not consider it. We explain parameters γ and β in the next section, and show that this prior leads to a temporal L_0 sparsity based optimal selection process.

Fig. 3 shows the intermediate latent masks generated in our experiments. They cover relatively sharp regions. Small regions and structures could be mistaken as rich details due to noise in the input frames. We remedy this problem through a family of sparse image priors.



(a) Plot of a family of penalty functions

(b) δ fixed to 1

(c) δ gradually decrease to 1/8

Figure 4. Effect of sparse image priors. (a) Plot of Eq. (12) in 1D. (b) When $\delta = 1$, Eq. (12) is a truncated L_1 penalty. (c) When δ gradually decreases to $1/8$, L_0 penalty is approximated and noise is better suppressed.

3.2. Family of Sparse Image Priors

We employ a family of sparse image priors to keep useful salient structures while suppressing noise. The penalty functions are

$$-\log P(I) \propto \eta \cdot \phi_\delta(\nabla I), \tag{10}$$

where $\phi_\delta(\nabla I)$ is defined as a sum over all single-pixel penalties as

$$\phi_\delta(\nabla I) = \sum\nolimits_p \psi(\nabla I_p). \tag{11}$$

$\psi(\nabla I_p)$ is set as

$$\psi(\nabla I_p) = \begin{cases} \|\nabla I_p\|_2/\delta & \text{if } \|\nabla I_p\|_2 < \delta \\ 1 & \text{otherwise} \end{cases} \tag{12}$$

This is a family of piecewise functions that concatenate a linear penalty with a constant, as shown in Fig. 4(a). As δ goes from 1 to 0, the penalty varies from truncated L_1 to the most sparse L_0 function. Fig. 4(c) shows as the penalty gets sparser (by gradually decreasing δ) in later iterations when solving for the high-res image, it effectively suppresses noise without attenuating salient edges. We show in the supplementary file[1] that this process is equivalent to spatial L_0 regularized image reconstruction.

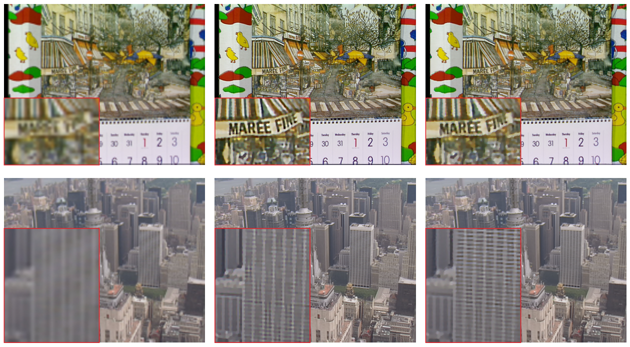
4. Inference

With Eq. (3), we estimate F, I and K iteratively.

4.1. Motion Estimation

I and K are fixed in this step, the optical flow F can be computed as $F = \arg\max_F P(\Omega|I,K,F)P(F)$. In this equation, the flow should be established upon the high-res image I. However, it costs much computation. We instead adopt a simple approximation of using interpolated classical TV-L1 flow [2, 21, 15] on the low-res images. The simple modification is 60 times faster and does not notably lower result quality, as shown in Fig. 5.

[1] http://www.cse.cuhk.edu.hk/leojia/projects/mfsr



(a) Bicubic ×4

(b) Results with high-res flow

(c) Results with interpolated low-res flow

Figure 5. Using TV-L1 based optical flow on the low-res grid yields reasonable results. (a) One input frame with bicubic ×4. (b) Results of using high-res flow [16]. (c) Results of using the interpolated low-res TV-L1 flow [2].

4.2. Image Reconstruction

K and F are fixed in this step. Directly marginalizing $P(\Omega, Z|I, K, F)$ over Z is intractable due to the large number of possible states. We adopt the EM method. It iterates between updating the expectation of $\log P(\Omega, Z|I, K, F)$ using the estimate of $P(Z|I^0, K, F, \Omega)$ and I^0, and revising the high-res image I^0.

E step: This step computes the posterior distribution $P(Z|I^0, K, F, \Omega)$ given the estimate of I^0 in the previous iteration, and uses this posterior to update the expectation of the complete-data log likelihood $\log P(\Omega, Z|I, K, F)$ as

$$Q(I|I^0) = E_{P(Z|I^0,K,F,\Omega)}[\log P(\Omega|Z, I, K, F) + \log P(Z|I, K, F)]. \tag{13}$$

Substituting Eqs. (7) and (9) into Eq. (13), we get

$$Q(I|I^0) = \begin{cases} -\lambda \sum_{i=-N}^{N} \sum_{p} E[Z_{i,p}]\,|D_{i,p}| & \text{if } (SK_iF_{0\to i}I)_p \in [0,1] \\ -\infty & \text{otherwise} \end{cases} \tag{14}$$

where $E[Z_{i,p}] = P(Z_{i,p} = 1|I^0, K, F, \Omega)$ is derived as

$$E[Z_{i,p}] = \frac{\exp\{-\lambda\,|D_{i,p}|\}\exp\{-\gamma/W_{i,p}\}}{\exp\{-\lambda\,|D_{i,p}|\}\exp\{-\gamma/W_{i,p}\} + \exp\{-\gamma\beta\}} \tag{15}$$

when $(SK_iF_{0\to i}I^0)_p \in [0, 1]$. Otherwise, it is equal to 0.

M step: This step updates the estimate I^0 to be the minimum of the complete-data negative log posterior, *i.e.*,

$$I^0 = \arg\min_{I} \sum_{i=-N}^{N} \lambda \left\| E[Z_i](SK_iF_{0\to i}I - I_i^L) \right\|_1 + \eta \cdot \phi_\delta(\nabla I). \tag{16}$$

Due to the non-convex and non-differentiable regularizer, our solution is based on the variable splitting scheme [29]. We rewrite the objective as

$$I^0 = \arg\min_{I,g} \sum_{i=-N}^{N} \lambda \left\| E[Z_i](SK_iF_{0\to i}I - I_i^L) \right\|_1 + \eta(\frac{1}{\delta}\|\nabla I - g\|_2 + \|g\|_0). \tag{17}$$

The proof is provided in our supplementary file. For each penalty obtained from a δ, we solve for image I via alternatively updating I and g as follows.

Fix g and estimate I: We perform iterative reweighted least squares (IRLS).

Fix I and solve for g: The optimal solution is:

$$g = \nabla I \cdot \max(\text{sign}(\|\nabla I\|_2 - \delta), 0), \tag{18}$$

according to the shrinkage formula.

Discussion From the MAP point of view, intermediate selection of pixels can be determined by

$$Z_{i,p} = \arg\max_{Z_{i,p}} P(\Omega|Z_{i,p}, I^0, K, F)P(Z_{i,p}|I^0, K, F). \tag{19}$$

Through a few derivations, the solution can be written as

$$Z_{i,p} = \begin{cases} 0 & \text{if } \lambda\,|D_{i,p}| + \gamma/W_{i,p} \geq \gamma\beta \text{ or } (SK_iF_{0\to i}I^0)_p \notin [0,1] \\ 1 & \text{otherwise} \end{cases} \tag{20}$$

Intuitively, Eq. (20) yields value 0 if the pixel is out of range, the reconstruction error is too large, or the pixel does not carry useful structure information. In another point of view, Eq. (20) is actually the solution of the following L_0 regularized problem:

$$Z_{i,p} = \arg\min_{Z_{i,p}\in[0,1]} \frac{\lambda}{\gamma} \left\| Z_{i,p}(SK_iF_{0\to i}I - I_i^L)_p \right\|_1 + \beta\|1 - Z_{i,p}\|_0 + Z_{i,p}^2/W_{i,p}. \tag{21}$$

Here $Z_{i,p}$ is relaxed to a real number in $[0, 1]$. This equation also explains the parameters γ and β. As β gets larger, the number of frames to keep also increases due to the L_0 term. In the extreme case when $\beta = +\infty$, $Z_{i,p} \equiv 1$, which means all frames will be kept for each pixel. Thus β takes the role of adjusting the number of frames to maintain. This is important because we need a reasonable number of informative frames for reconstruction.

4.3. Guided Kernel Estimation

Once we have the image I obtained from the reconstruction step and optical flow field F, the blur kernel estimation can then be refined. We estimate composition of the anti-aliasing and motion kernels for each input low-res image by solving the L_2 regularized problem of

$$K_i = \operatorname{argmin}_{K_i} \left\| SK_iF_{0\to i}I - I_i^L \right\|_1 + \xi \left\| K_i \right\|_F^2 . \tag{22}$$

$E[Z_{i,p}]$ is not taken into consideration at this stage because it is only used to estimate a high-res image I. This approximation not only simplifies computation, but also has numerical advantages. It can avoid trivial solutions especially when one image contains many small $E[Z_{i,p}]$. In experiments, this approximation works well, as demonstrated in Section 6.

We use IRLS to minimize Eq. (22). The gradient of the linear system in each reweighting step w.r.t. K_i is given by

$$(A_i^T S^T W S A_i + \xi I_d) K_i - A_i^T S^T W I_i^L, \tag{23}$$

where $W \triangleq diag([(SK_i F_{0\rightarrow i} I - I_i^L)^2 + \varepsilon]^{-\frac{1}{2}})$ is the weight in each reweighting iteration. A_i is a matrix with each row the vector form of image $F_{0\rightarrow i} I$, which corresponds to one element of the filter K_i – that is, A_i satisfies $A_i K_i \equiv K_i F_{0\rightarrow i} I$.

Directly computing $A_i^T S^T W S A_i K_i$ is time consuming due to the large size of I and K_i. To accelerate it, we use $\mathcal{F}^{-1}[\overline{\mathcal{F}(A_i)}\mathcal{F}(S^T W S \mathcal{F}^{-1}(\mathcal{F}(A_i)\mathcal{F}(K_i)))]$ instead. Here, $\mathcal{F}$ and $\mathcal{F}^{-1}$ are FFT and inverse FFT operations.

5. Analysis

We analyze in 1D the relationship between the restoration error of high-res signals and motion blur kernels, as well as the impact of motion blur estimation. Similar conclusion also applies to higher dimensions.

Suppose the latent high-res signal takes the form of $I(n) = \frac{A}{N_L}\exp(\frac{i2\pi\omega_0 n}{N_H})$ in one frequency component after decomposition. Here N_L and N_H are lengths of the low-res and latent high-res signals respectively. A is the complex amplitude. The DFT of $I(n)$ is $\tilde{I}(\omega) = MA\delta(\omega - \omega_0)$, where $M = N_H/N_L$ is the ratio of downsampling. We can write DFT of the low-res signal as

$$\tilde{J}(\omega) = FG\tilde{I}(\omega) + \tilde{E}(\omega), \tag{24}$$

where F, G, and $\tilde{E}(\omega)$ are the DFT of the motion blur kernel, anti-aliasing filter, and additive Gaussian noise respectively. We assume G is Gaussian – that is, $G(\omega) = e^{-\omega^2\sigma_k^2/2}$ where σ_k is the standard deviation.

The negative log likelihood function for the input signal can be written as

$$-\log P(\tilde{J}\,|A) = \frac{1}{2\sigma_n^2}\left\|\tilde{J}(\omega_0) - FGA\right\|_2^2. \tag{25}$$

With this equation, we derive the Fisher information matrix for parameters $\theta = [\mathrm{Re}\{A\}, \mathrm{Im}\{A\}]$ as

$$I_\theta = \frac{F^*FG^2}{\sigma_n^2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{26}$$

where $*$ denotes the complex conjugate. Finally, the Cramer-Rao bound for recovering signal A is formulated as

$$\mathrm{var}\left(\hat{A}\right) \geq I_\theta^{-1}(1,1) + I_\theta^{-1}(2,2) = \frac{2\sigma_n^2}{F^*F}e^{\omega^2\sigma_k^2}, \tag{27}$$

where $\hat{A}$ is the unbiased estimation of A. This bound indicates that a small frequency component in the motion blur kernel F causes a larger error of the corresponding frequency component in the high-res image estimate.

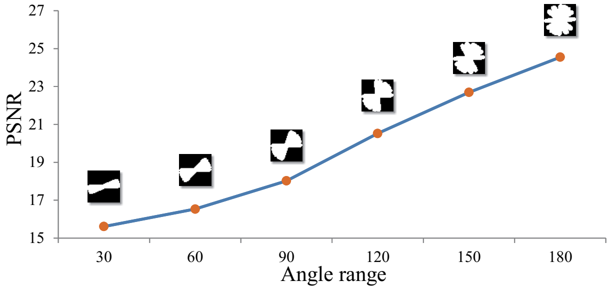


Figure 6. Diversity of blur directions vs. result quality. The diversity of directions is visualized by merging non-zero elements of the blur kernels as shown above the curve.

We take for example the case of unidirectional blur, which happens for videos when the region to magnify is small. The magnitude of the kernel DFT is large only in one direction. It attenuates quickly along the blur direction. Hence, to get a stable estimate, it is desired that the directions span a large range. We randomly generate directional kernels to synthesize 6 low-res sequences and estimate the high-res image using our framework. Fig. 6 plots the resulting PSNR vs. diversity of blur directions. The more diverse the blur directions are, the better results we get, consistent with our analysis.

If A is fixed, the Cramer-Rao bound for motion blur F can be derived as

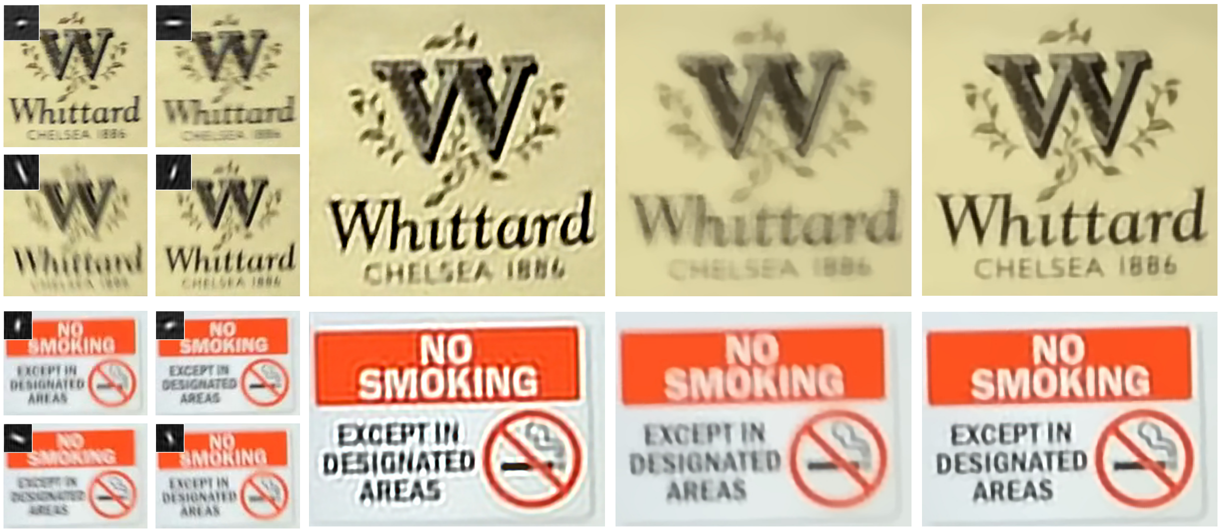
$$\mathrm{var}\left(\hat{F}\right) \geq \frac{2\sigma_n^2}{A^*A}e^{\omega^2\sigma_k^2}. \tag{28}$$

This indicates that the error of blur kernel estimation increases if the image becomes more blurry (as σ_k increases). Our framework to construct a less blurred high-res image is thus beneficial to blur kernel estimation.

6. Experiments

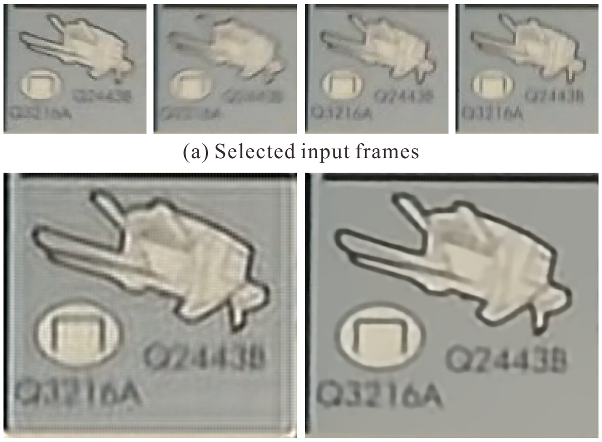
We evaluated the proposed method on several sequences captured by cameras. Our implementation is in MATLAB on an Intel Core i5 PC with 8GB RAM. For ×4 super-resolution, it takes about 20 minutes to construct one 720×480 image using 30 neighboring frames.

The parameters are set as follows. $\lambda = 1$ in Eq. (7). $\xi = 2$ in Eq. (4). The patch-size for defining the sharpness measure is fixed to 11×11. $\gamma = 10$ in Eq. (9). $\eta = 0.2$



(a) Selected input frames (b) Deblur + SR (c) Results w/o sharpness mask (d) Our final results

Figure 7. More results and comparison (×3). (a) Four input frames from each sequence. (b) Results of multi-image deblurring [32] followed by super-resolution [16]. (c) Results without sharpness mask ($\beta = \infty$). (d) Our results with the sharpness masks ($\beta = 1.4$).



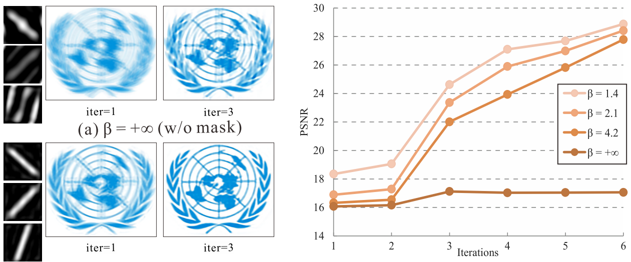
(a) Selected input frames

(b) Result of [31] (c) Our result

Figure 8. More results and comparison (×3). (a) Four input frames from a sequence. (b) Multi-shot image result [31]. (c) Our result.

in Eq. (10). $\psi = 0.02$ in Eq. (5). The number of EM iterations is 2 and the number of outer iterations for image reconstruction is 3. β starts at 1.4, and doubles itself in each outer iteration. In each M-step, we decrease δ from $1/2$ to $1/8$ by a factor of 2 to gradually suppress noise.

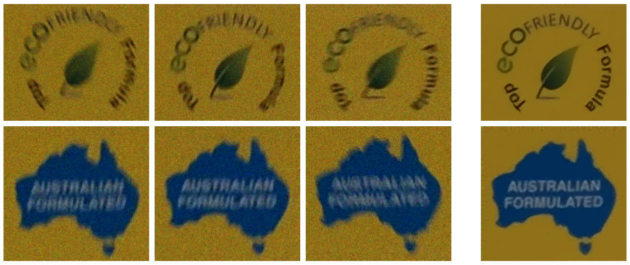
Comparison with State-of-the-arts Fig. 7(b) and (d) compare our results with those produced by multi-image deblurring [32] followed by multi-frame super-resolution (SR) [16]. Because this scheme estimates the low-res blur kernels in a blind way, the resulting ringing artifacts are magnified during the SR process [16]. Our method leverages sharp structures in the sequence and produces clearer results. Fig. 8 shows another comparison with the multi-frame SR method of [31] that does consider motion blur. More results are in our project website.



(a) β = +∞ (w/o mask)

(b) β = 1.4 (w/ mask) (c) PSNR vs. Iterations

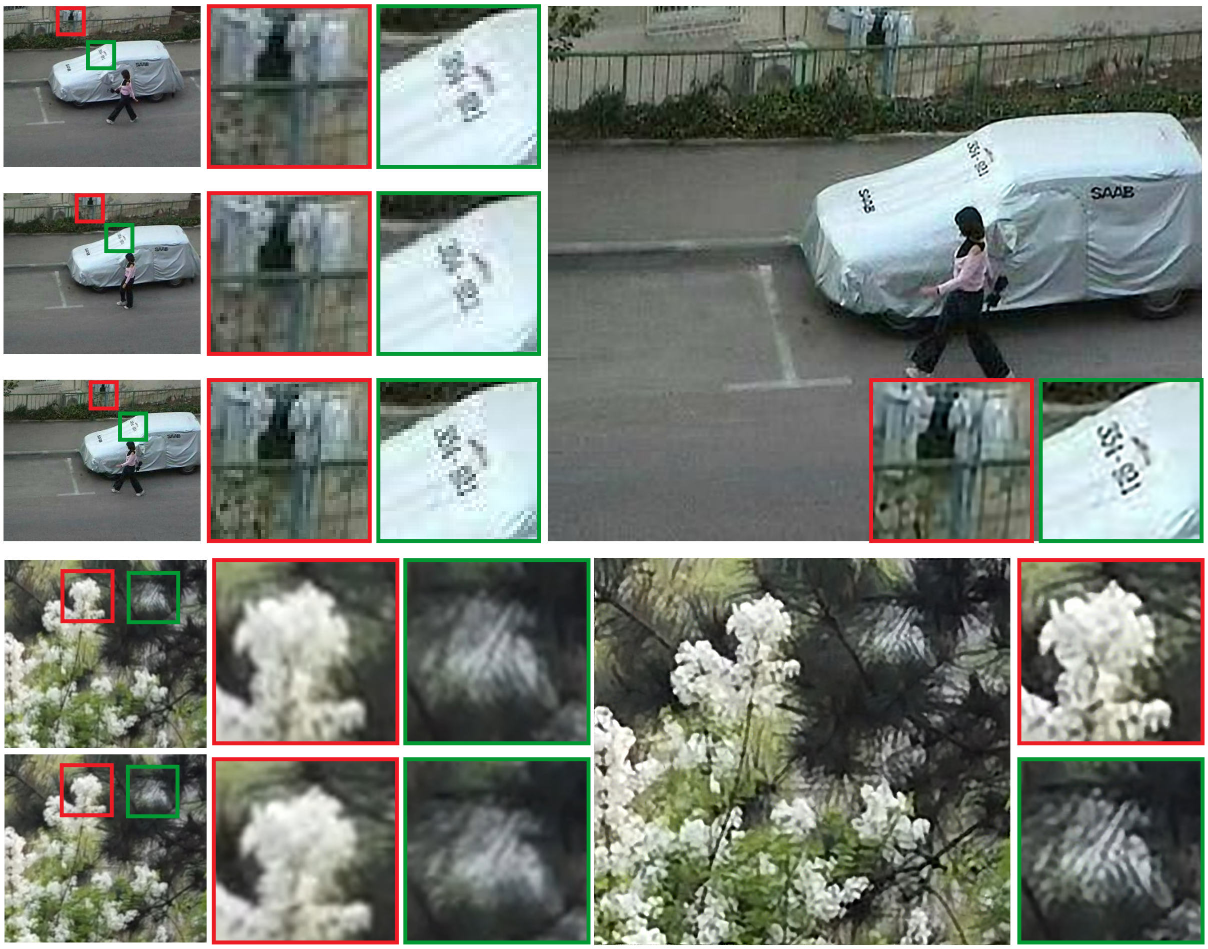
Figure 9. Quantitative evaluation. (a)-(b) Sample intermediate kernels and high-res images in different iterations. (c) PSNRs in each iteration for different β.



(a) Selected input frames (b) Our results

Figure 10. Our method is robust to noise. (a) A few relative sharp input frames. (b) Our results (×3).

Effect of the Sharpness Mask Fig. 7(d) and (c) show the results with/without using our latent sharpness masks. To



(a) Selected input frames (with zoom-in)

(b) Our results

Figure 11. More natural video results. (a) Sample input frames. (b) Our results estimated using 31 low-res frames each.

quantitatively evaluate the effectiveness of using the binary latent mask Z for selecting pixels, we synthesize a low-res sequence with randomly generated unidirectional blur. We show in Fig. 9 the intermediate high-res images at each outer iteration and the estimated blur kernels. In (a), when $\beta = +\infty$ (without mask), the algorithm fails in estimating sharp structures in early iterations. The resulting blurred edges then mislead the kernel estimation. Fig. 9(b) shows when $\beta = 1.4$ (with mask), the mask Z has the effect of selecting sharp structures, which then benefit kernel estimation. The PSNRs of the restored frames are listed in (c).

Influence of Input Quality Our method is also robust to noise. We add Poisson noise to the sequences, as shown in Fig. 10. The resulting text is still readable.

More Natural Video Results In most of our examples, we focus on magnify text and signs because they are rich in details and important to human. In Fig. 11 we show that our method also generates reasonable results on general natural sequences. Please visit the project website for more results.

7. Concluding Remarks

We have presented a system for handling motion blur in MFSR and demonstrated a few results on challenging natural sequences. Our method is robust to both motion blur and noise. Our model currently assumes that the motion blur is uniform. It is easy to generalize our region selection to non-uniform blur as in [10]. Our future work will be developing more priors for special applications, like text and face MFSR.

Acknowledgements

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region (Project No. 413113), NSFC (61272326) and Grant of University of Macau (MYRG202(Y1-L4)-FST11-WEH).

References

[1] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *ECCV*, pages 404–417, 2006.

[2] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *ECCV*, pages 25–36, 2004.

[3] J.-F. Cai, H. Ji, C. Liu, and Z. Shen. High-quality curvelet-based motion deblurring from an image pair. In *CVPR*, pages 1566–1573, 2009.

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

[5] S. Cho and S. Lee. Fast motion deblurring. *ACM Transactions on Graphics (TOG)*, 28(5):145, 2009.

[6] S. Cho, J. Wang, and S. Lee. Handling outliers in non-blind image deconvolution. In *ICCV*, pages 495–502, 2011.

[7] S. Cho, J. Wang, and S. Lee. Video deblurring for hand-held cameras using patch-based synthesis. *ACM Transactions on Graphics (TOG)*, 31(4):64, 2012.

[8] N. Efrat, D. Glasner, A. Apartsin, B. Nadler, and A. Levin. Accurate blur models vs. image priors in single image super-resolution. In *ICCV*, pages 2832–2839, 2013.

[9] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *TIP*, 13(10):1327–1344, 2004.

[10] M. Hirsch, S. Sra, B. Scholkopf, and S. Harmeling. Efficient filter flow for space-variant multiframe blind deconvolution. In *CVPR*, pages 607–614, 2010.

[11] M. Irani and S. Peleg. Improving resolution by image registration. *CVGIP: Graphical models and image processing*, 53(3):231–239, 1991.

[12] S. M. Kay. *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.

[13] W. Li, J. Zhang, and Q. Dai. Exploring aligned complementary image pair for blind motion deblurring. In *CVPR*, pages 273–280, 2011.

[14] Y. Li, S. B. Kang, N. Joshi, S. M. Seitz, and D. P. Huttenlocher. Generating sharp panoramas from motion-blurred videos. In *CVPR*, pages 2424–2431, 2010.

[15] C. Liu et al. *Beyond pixels: exploring new representations and applications for motion analysis*. PhD thesis, Massachusetts Institute of Technology, 2009.

[16] C. Liu and D. Sun. A bayesian approach to adaptive video super resolution. In *CVPR*, pages 209–216, 2011.

[17] T. Michaeli and M. Irani. Nonparametric blind super-resolution. In *ICCV*, pages 945–952, 2013.

[18] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: a technical overview. *Signal Processing Magazine, IEEE*, 20(3):21–36, 2003.

[19] Q. Shan, Z. Li, J. Jia, and C.-K. Tang. Fast image/video upsampling. *ACM Transactions on Graphics (TOG)*, 27(5):153, 2008.

[20] J. Shi and C. Tomasi. Good features to track. In *CVPR*, pages 593–593, 1994.

[21] D. Sun, S. Roth, and M. J. Black. Secrets of optical flow estimation and their principles. In *CVPR*, pages 2432–2439, 2010.

[22] K. Sunkavalli, N. Joshi, S. B. Kang, M. F. Cohen, and H. Pfister. Video snapshots: creating high-quality images from video clips. *TVCG*, 18(11):1868–1879, 2012.

[23] Y.-W. Tai, H. Du, M. S. Brown, and S. Lin. Image/video deblurring using a hybrid camera. In *CVPR*, pages 1–8, 2008.

[24] Y.-W. Tai and S. Lin. Motion-aware noise filtering for deblurring of noisy and blurry images. In *CVPR*, pages 17–24, 2012.

[25] Y.-W. Tai, S. Liu, M. S. Brown, and S. Lin. Super resolution using edge prior and single image detail synthesis. In *CVPR*, pages 2400–2407, 2010.

[26] H. Takeda, P. Milanfar, M. Protter, and M. Elad. Super-resolution without explicit subpixel motion estimation. *TIP*, 18(9):1958–1975, 2009.

[27] R. Tsai and T. S. Huang. Multiframe image restoration and registration. *Advances in computer vision and Image Processing*, 1(2):317–339, 1984.

[28] L. Xu and J. Jia. Two-phase kernel estimation for robust motion deblurring. In *ECCV*, pages 157–170, 2010.

[29] L. Xu, C. Lu, Y. Xu, and J. Jia. Image smoothing via l0 gradient minimization. *ACM Transactions on Graphics (TOG)*, 30(6):174, 2011.

[30] L. Xu, S. Zheng, and J. Jia. Unnatural l0 sparse representation for natural image deblurring. In *CVPR*, pages 1107–1114, 2013.

[31] H. Zhang and L. Carin. Multi-shot imaging: Joint alignment, deblurring, and resolution-enhancement. In *CVPR*, pages 2925–2932, 2014.

[32] X. Zhu, F. Šroubek, and P. Milanfar. Deconvolving psfs for a better motion deblurring using multiple images. In *ECCV*, pages 636–647, 2012.

[33] M. Zontak and M. Irani. Internal statistics of a single natural image. In *CVPR*, pages 977–984, 2011.