Burst Deblurring: Removing Camera Shake Through Fourier Burst Accumulation

Mauricio Delbracio Guilermo Sapiro
ECE, Duke University
{mauricio.delbracio, guillermo.sapiro}duke.edu

Abstract

Numerous recent approaches attempt to remove image blur due to camera shake, either with one or multiple input images, by explicitly solving an inverse and inherently ill-posed deconvolution problem. If the photographer takes a burst of images, a modality available in virtually all modern digital cameras, we show that it is possible to combine them to get a clean sharp version. This is done without explicitly solving any blur estimation and subsequent inverse problem. The proposed algorithm is strikingly simple: it performs a weighted average in the Fourier domain, with weights depending on the Fourier spectrum magnitude. The method's rationale is that camera shake has a random nature and therefore each image in the burst is generally blurred differently. Experiments with real camera data show that the proposed Fourier Burst Accumulation algorithm achieves state-of-the-art results an order of magnitude faster, with simplicity for on-board implementation on camera phones.

1. Introduction

One of the most challenging experiences in photography is taking images in low-light environments. The basic principle of photography is the accumulation of photons in the sensor during a given exposure time. In general, the more photons reach the surface the better the quality of the final image, as the photonic noise is reduced. However, this basic principle requires the photographed scene to be static and that there is no relative motion between the camera and the scene. Otherwise, the photons will be accumulated in neighboring pixels, generating a loss of sharpness (blur). This problem is significant when shooting with hand-held cameras, the most popular photography device today, in dim light conditions.

Under reasonable hypotheses, the camera shake can be modeled mathematically as a convolution,

$$v = u \star k + n, \tag{1}$$

where v is the noisy blurred observation, u is the latent sharp image, k is an unknown blurring kernel and n is additive white noise. For this model to be accurate, the camera movement has to be essentially a rotation in its optical axis with negligible in-plane rotation, e.g., [26]. The kernel k results from several blur sources: light diffraction due to the finite aperture, out-of-focus, light integration in the photo-sensor, and relative motion between the camera and the scene during the exposure. To get enough photons per pixel in a typical low light scene, the camera needs to capture light for a period of tens to hundreds of milliseconds. In such a situation (and assuming that the scene is static and the user/camera has correctly set the focus), the dominant contribution to the blur kernel is the camera shake — mostly due to hand tremors.

Current cameras can take a burst of images, this being popular also in camera phones. This has been exploited in several approaches for accumulating photons in the different images and then forming an image with less noise (mimicking a longer exposure time a posteriori, see e.g., [2]). However, this principle is disturbed if the images in the burst have blur. The classical mathematical formulation as a multi-image deconvolution, seeks to solve an inverse problem where the unknowns are the multiple blurring operators and the underlying sharp image. This procedure, although produces good results [30], is computationally very expensive, and very sensitive to a good estimation of the blurring kernels. Furthermore, since the inverse problem is ill-posed it relies on priors either or both for the calculus of the blurs and the latent sharp image.

Camera shake originated from hand tremor vibrations is essentially random [4, 11, 27]. This implies that the movement of the camera in an individual image of the burst is independent of the movement in another one. Thus, the blur in one frame will be differ-

ent from the one in another image of the burst. There are mainly three sources of tremor: arm (< 5Hz), wrist (5-20Hz), and fingers (20-30Hz) (Fig. 2 in [11]). This implies that the correlation between successive frames is typically low, unless the frames are acquired at a very fast shutter speed ($>1/100$s), but of course, in such scenario there will be no camera shake blur.

Our work is built on this basic principle. We present an algorithm that aggregates a burst of images taking what is less blurred of each frame to build an image that is sharper and less noisy than all the images in the burst. The algorithm is straightforward to implement and conceptually simple. It takes as input a series of registered images and computes a weighted average of the Fourier coefficients of the images in the burst. With the availability of accurate gyroscope and accelerometers in, for example, phone cameras, the registration can be obtained "for free," rendering the whole algorithm very efficient for on-board implementation. We also completely avoid the explicit computation of the blurring kernel (as commonly done in the literature), which is not only an unimportant hidden variable for the task at hand, but as mentioned above, still leaves the ill-posed and computationally very expensive task of solving the inverse problem.

Evaluation through synthetic and real experiments shows that the final image quality is significantly improved. This is done without explicitly performing deconvolution, which generally introduces artifacts and also a significant overhead. Comparison to state-of-the-art multi-image deconvolution algorithms shows that our approach produces similar or better results while being orders of magnitude faster and simpler. The proposed algorithm does not assume any prior on the latent image; exclusively relying on the randomness of hand tremor.

The remaining of the paper is organized as follows. Section 2 discusses the related work and the substantial differences to what we propose. Section 3 explains how a burst can be combined in the Fourier domain to recover a sharper image, while in Section 4 we present an empirical analysis of the algorithm performance through the simulation of camera shake kernels. Section 5 details the algorithm implementation and in Section 6 we present results of the proposed aggregation procedure in real data, comparing the algorithm to state-of-the-art multi-image deconvolution methods. We finally close in Section 7.

2. Related Work

Removing camera shake blur is one of the most challenging problems in image processing. Although in the last decade several image restoration algorithms have emerged giving outstanding performance, its success is still very dependent on the scene. Most image deblurring algorithms cast the problem as a deconvolution with either a known (non-blind) or an unknown blurring kernel (blind). See e.g., the review by Kundur and Hatzinakos [14], where a discussion of the most classical methods is presented.

Most blind deconvolution algorithms try to estimate the latent image without any other input than the noisy blurred image itself. A representative work is the one by Fergus *et al.* [9]. This variational method sparked many competitors seeking to combine natural image priors, assumptions on the blurring operator, and complex optimization frameworks, to simultaneously estimate both the blurring kernel and the sharp image [13, 17, 20, 24, 28].

Others attempt to first estimate the degradation operator and then applying a non-blind deconvolution algorithm. For instance, [6] accelerates the kernel estimation step by using fast image filters for explicitly detecting and restoring strong edges in the latent sharp image. Since the blurring kernel has typically a very small support, the kernel estimation problem is better conditioned than estimating the kernel and the sharp image simultaneously [16, 17]. However, even in non-blind deblurring, i.e., when the blurring kernels are known, the problem is generally ill-posed, because the blur introduces zeros in the frequency domain. Thereby avoiding explicit inversion, as here proposed, becomes critical.

Two or more input images can improve the estimation of both the underlying image and the blurring kernels. Rav-Acha and Peleg [23] claimed that *"Two motion-blurred images are better than one,"* if the direction of the blurs are different. In [29] the authors proposed to capture two photographs: one having a short exposure time, noisy but sharp, and one with a long exposure, blurred but with low noise. The two acquisitions are complementary, and the sharp one is used to estimate the motion kernel of the blurred one.

The closest to our work are papers on multi-image blind deconvolution [3, 5, 25, 30, 32]. In [3] the authors introduce a prior on the sparsity of the motion blur kernel to constraint the blind deblurring problem. Most of these multi-image algorithms introduce cross-blur penalty functions between image pairs. However this has the problem of growing combinatorially with the number of images in the burst. This idea is extended in [30] using a Bayesian framework for coupling all the unknown blurring kernels and the latent image in a unique prior. Although this prior has numerous good mathematical properties, its optimization is very slow. The algorithm produces very good results but it may take several minutes or even hours for a typical

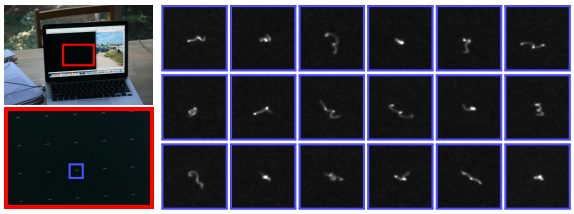


Figure 1: When the camera is set to a burst mode, several photographs are captured sequentially. Due to the random nature of hand tremor, camera shake blur is mostly independent from one frame to the other. An image consisting of white dots was photographed with a handheld camera at 1/4" to depict the camera motion kernels. The kernels are mainly unidimensional regular trajectories that are not completely random (perfect random walk) nor uniform.

burst of 8-10 images of several megapixels. The very recent work by Park and Levoy [22] relies on an attached gyroscope, now present in many phones and tablets, to align all the input images and to get an estimation of the blurring kernels. Then, a multi-image non-blind deconvolution algorithm is applied. By taking a burst of images, the multi-image deconvolution problem becomes less ill-posed allowing the use of simpler priors. This is explored in [12] where the authors introduce a total variation prior on the underlying sharp image.

All these papers propose kernel estimation and to solve an inverse problem of image deconvolution. The main inconvenience of tackling this problem as a deconvolution, on top of the computational burden, is that if the convolution model is not accurate or the kernel is not accurately estimated, the restored image will contain strong artifacts (such as ringing). Our approach is radically different. The idea is to fuse all the images in the burst without explicitly estimating the blurring kernels and subsequent inverse problem approach, but taking the information that is less degraded from each image in the burst. In that sense our work is closer to multi-image denoising [2], however we consider that the input images may also be blurred.

3. Fourier Burst Accumulation

Camera shake originated from hand tremor vibrations has undoubtedly a random nature [4, 11, 27]. The independent movement of the photographer hand causes the camera to be pushed randomly and unpredictably, generating blurriness in the captured image. Figure 1 shows several photographs taken with a DSLR handheld camera. The photographed scene consists of a laptop displaying a black image with white dots. The captured picture of the white dots illustrates the trace of the camera movement in the image plane. If the dots are very small —mimicking Dirac masses— their photographs represent the blurring kernels themselves. The kernels mostly consist of unidimensional regular random trajectories. This stochastic behavior will be the key ingredient in our proposed approach.

Let $\mathcal{F}$ denote the Fourier Transform and $\hat{k}$ the Fourier Transform of the kernel k. Images are defined in a regular grid indexed by the $2D$ position $\mathbf{x}$ and the Fourier domain is indexed by the $2D$ frequency ζ. Let us assume, without loss of generality, that the kernel k due to camera shake is normalized such that $\int k(\mathbf{x})d\mathbf{x} = 1$. The blurring kernel is nonnegative since the integration of incoherent light is always nonnegative. This implies that the motion blur does not amplify the Fourier spectrum:

Claim 1. *Let $k(\mathbf{x}) \geq 0$ and $\int k(\mathbf{x}) = 1$. Then, $|\hat{k}(\zeta)| \leq 1, \forall \zeta$. (Blurring kernels do not amplify the spectrum.)*

Proof.

$$\left|\hat{k}(\zeta)\right| = \left|\int k(\mathbf{x}) e^{i\mathbf{x}\cdot\zeta} d\mathbf{x}\right| \leq \int |k(\mathbf{x})|\, d\mathbf{x} = \int k(\mathbf{x}) d\mathbf{x} = 1.$$

□

Most modern digital cameras have a burst mode where the photographer is allowed to sequently take a series of images, one right after the other. Let us assume that the photographer takes a sequence of M images of the same scene u,

$$v_i = u \star k_i + n_i, \quad \text{for} \quad i = 1, \ldots, M. \tag{2}$$

The movement of the camera during any two images of the burst will be essentially independent. Thus, the blurring kernels k_i will be mostly different for different images in the burst. Hence, each Fourier frequency of $\hat{u}$ will be differently attenuated on each frame of the burst. The idea is to reconstruct an image whose Fourier spectrum takes for each frequency the value having the largest Fourier magnitude in the burst. Since a blurring kernel does not amplify the Fourier spectrum (Claim 1), the reconstructed image picks what is less attenuated from each image of the burst.

More formally, let p be a non-negative integer, we will call *Fourier Burst Accumulation* (FBA) to the Fourier weighted averaged image,

$$u_p(\mathbf{x}) = \mathcal{F}^{-1}\left(\sum_{i=1}^{M} w_i(\zeta) \cdot \hat{v}_i(\zeta)\right)(\mathbf{x}), \tag{3}$$

$$w_i(\zeta) = \frac{|\hat{v}_i(\zeta)|^p}{\sum_{j=1}^{M} |\hat{v}_j(\zeta)|^p},$$

where $\hat{v}_i$ is the Fourier Transform of the individual burst image v_i. The weight $w_i := w_i(\zeta)$ controls the contribution of the frequency ζ of image v_i to the final reconstruction u_p. Given ζ, for $p > 0$, the larger the value of $|\hat{v}_i(\zeta)|$, the more $\hat{v}_i(\zeta)$ contributes to the average, reflecting what we discussed above that the strongest frequency values represent the least attenuated u components.

The integer p controls the aggregation of the images in the Fourier domain. If $p = 0$, the restored image is just the average of the burst (as standard for example in the case of noise only), while if $p \to \infty$, each reconstructed frequency takes the maximum value of that frequency along the burst. This is stated in the following claim; the proof is straightforward and it is therefore omitted.

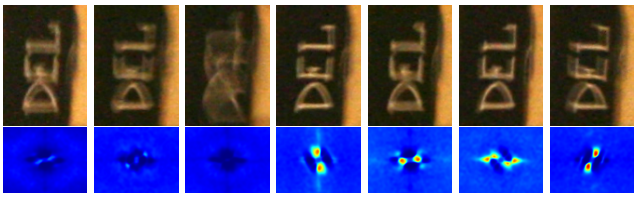
Claim 2. *Mean/Max aggregation. Suppose that* $\hat{v}_i(\zeta)$ *for* $i = 1, \dots, M$ *are such that* $|\hat{v}_{i_1}(\zeta)| = |\hat{v}_{i_2}(\zeta)| = \dots = |\hat{v}_{i_q}(\zeta)| > |\hat{v}_{i_{q+1}}(\zeta)| \geq |\hat{v}_{i_{q+2}}(\zeta)| \geq \dots \geq |\hat{v}_{i_M}(\zeta)|$ *and* $w_i(\zeta)$ *is given by* (3). *If* $p = 0$, *then* $w_i(\zeta) = \frac{1}{M}, \forall i$ *(arithmetic mean pooling), while if* $p \to \infty$, *then* $w_i(\zeta) = \frac{1}{q}$ *for* $i = i_1, \dots, i_q$ *and* $w_i(\zeta) = 0$ *otherwise (maximum pooling).*

The Fourier weights only depend on the Fourier magnitude and hence they are not sensitive to image misalignment. However, when doing the average in (3), the images v_i have to be correctly aligned to mitigate Fourier phase intermingling and get a sharp aggregation. The images in our experiments are aligned using SIFT correspondences and then finding the dominant homography between each image in the burst and the first one (implementation details are given in Section 5). This pre-alignment step can be done exploiting the camera gyroscope and accelerometer data.

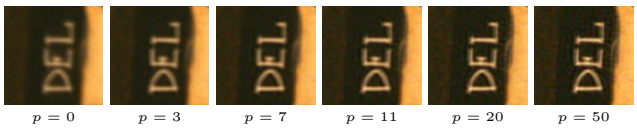
Dealing with noise. The images in the burst are blurry but also contaminated with noise. In the ideal case where the input images are not contaminated with noise, (3) is reduced to

$$w_i = \frac{|\hat{v}_i|^p}{\sum_{j=1}^M |\hat{v}_j|^p} = \frac{|\hat{k}_i \cdot \hat{u}|^p}{\sum_{j=1}^M |\hat{k}_j \cdot \hat{u}|^p} = \frac{|\hat{k}_i|^p}{\sum_{j=1}^M |\hat{k}_j|^p}, \quad (4)$$

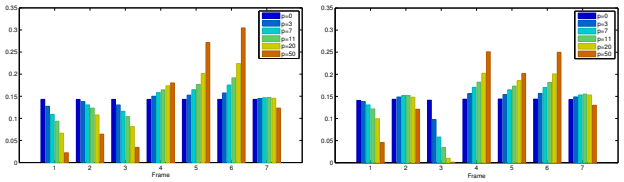
as long as $|\hat{u}| > 0$. This is what we would like to have: a procedure for weighting more the frequencies that are less attenuated by the different camera shake kernels. Since camera shake kernels have typically a small support, of maximum only a few tenths of pixels, its Fourier spectrum magnitude varies very smoothly. Thus, $|\hat{v}_i|$ can be smoothed out before computing the weights. This helps to remove noise and also to stabilize the weights (see Section 5).



(a) Frames crop 1-7 and the Fourier weights w_i for $p = 11$.



(b) Fourier Aggregation results for different p values.



(c) Weights energy distribution (d) Weighted frames energy distribution

Figure 2: Weights distribution of the Fourier Burst Aggregation when changing the value of p. As p increases, the weights are concentrated in fewer images and the aggregated image becomes sharper but also noisier.

4. Fourier Burst Accumulation Analysis

4.1. Weights Distribution

The value of p balances sharpness and noise reduction. Although, one would always prefer to get a sharp image, due to noise and the unknown Fourier phase shift introduced by the aggregation procedure, the resultant image would not necessary be better as $p \to \infty$. Figure 2 shows an example of the proposed FBA for a burst of 7 images, and the amount of contribution of each frame to the final aggregation. As p grows, the weights are concentrated in fewer images. The weights maps clearly show that different Fourier frequencies are recovered from different frames. In this example, the high frequency content is uniformly taken from all the frames in the burst. This produces a strong noise reduction behavior, in addition to the sharpening effect.

4.2. Statistical Performance

To show the statistical performance of the Fourier weighted accumulation, we carried out an empirical analysis applying the proposed aggregation with different values of p. We simulated motion kernels following [11], where the (expected value) amount of blur is controlled by a parameter related to the exposure time. We also controlled the number of images in the burst and the noise level in each frame. The kernels were generated by simulating the random shake of the camera

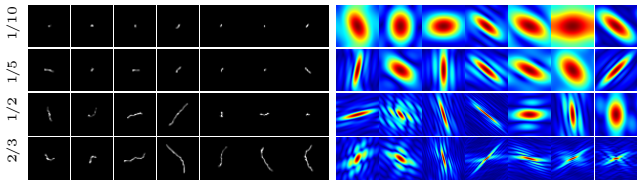


Figure 3: Simulated kernels due to hand tremor following [11]. Each row shows a set of simulated kernels (left panel) for different exposures $T_{\text{exp}} = 1/10, 1/5, 1/2, 2/3$, and the respective Fourier spectrum magnitude (right panel). The parameter T_{exp} controls the amount of expected blur.

from the power spectral density of measured physiological hand tremor [4]. All the images were aligned by pre-centering the motion kernels before blurring the underlying sharp image. Figure 3 shows several different realizations for different exposure values T_{exp}. Actually, the amount of blur not only depends on the exposure time but also on the focal distance, user expertise, and camera dimensions and mass [1]. However, for simplicity, all these variables were controlled by the single parameter T_{exp}.

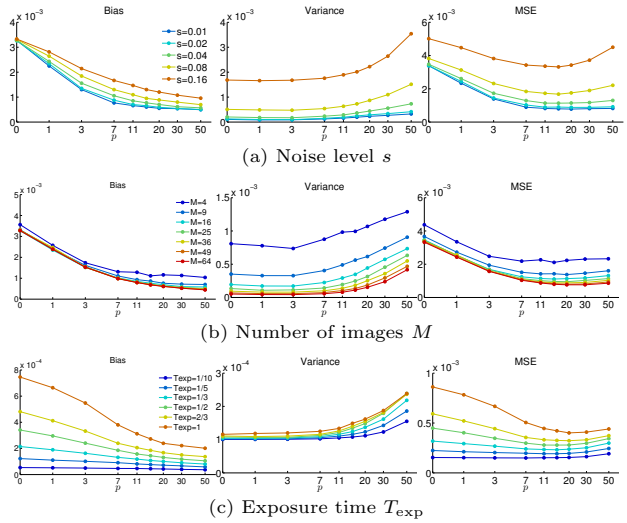
We computed the empirical mean square error (MSE) by randomly sampling hundreds of different motion kernels and Gaussian noise realizations, and then applying the Fourier aggregation procedure. The mean square error was decomposed into the bias and variance terms, namely $\text{MSE}(u_p) = \text{bias}(u_p)^2 + \text{var}(u_p)$, to help visualize the behavior of the algorithm.

Figure 4 shows the average algorithm performance when changing the acquisition conditions. In general, the larger the value of p the smaller the bias and the larger the variance. There is a minimum of the mean square error for $p \in [7, 30]$. This is reasonable since there exists a tradeoff between variance reduction and bias. Although both the bias and the variance are affected by the noise level, the qualitative performance of the algorithm remains the same. The bias is not altered by the number of frames in the burst but the variance is reduced as more images are used, implying a gain in the expected MSE as more images are used. On the other hand, the exposure time mostly affects the bias, being much more significant for larger exposures as expected.

5. Algorithm Implementation

The burst restoration algorithm is built on three main blocks: Burst Registration, Fourier Burst Accumulation, and Noise Aware Sharpening as a post-processing. These are described in what follows.

Burst Registration. There are several ways of registering images (see [33] for a survey). In this work, we



(a) Noise level s

(b) Number of images M

(c) Exposure time T_{exp}

Figure 4: Bias-Variance tradeoff. Average algorithm performance with respect to p when changing (a) the amount of noise in the input images s, (b) the number of images in the burst M, and (c) the exposure time T_{exp}. The rest of the parameters are set to $M = 16$, $s = 0.04$ and $T_{\text{exp}} = 1/3$, unless other specified. With short exposures, the arithmetic average ($p = 0$) produces the best MSE since the images are not blurred. The bias does not depend on M, but the variance can be significantly reduced by taking more images (light accumulation procedure). Noise affects the bias and the variance terms (with the exception of $p = 0$ where the bias is unaffected). The MSE plots show the existence of a minimum for $p \in [7, 30]$, indicating that the best is to balance a perfect average and a max pooling.

use image correspondences to estimate the dominant homography relating every image of the burst and a reference image (the first one in the burst). Although restrictive, the homography assumption is valid if the scene is planar (or far from the camera) or the viewpoint location is fixed, e.g., the camera only rotates around its optical center. Image correspondences are found using SIFT features [19] and then filtered out through the ORSA algorithm [21], a variant of the so called RANSAC method [10]. Recall that as in prior art, e.g., [22], the registration can be done with the gyroscope and accelerometer information from the camera.

Fourier Burst Accumulation. Given the registered images $\{v_i\}_{i=1}^M$ we directly compute the corresponding Fourier transforms $\{\hat{v}_i\}_{i=1}^M$. Since camera shake motion kernels have a small spatial support, their Fourier spectrum magnitudes vary very smoothly. Thus, $|\hat{v}_i|$ can be lowpass filtered before computing the weights, that is, $|\bar{\hat{v}}_i| = G_\sigma |\hat{v}_i|$, where G_σ is a Gaussian filter of standard deviation σ. The strength of the low pass filter (controlled by the parameter σ) should depend on

the assumed kernel size (the smaller the kernel the more regular its Fourier spectrum magnitude). In our implementation we set $\sigma = {}^{\min(m_h,m_w)}/_{k_s}$, where $k_s = 50$ pixels and the image size is $m_h \times m_w$ pixels. Although this low pass filter is important, the results are not too sensitive to the value of σ.

The final Fourier burst aggregation is (note that the smoothing is only applied to the weights calculation)

$$u_p = \mathcal{F}^{-1}\left(\sum_{i=1}^{M} w_i \cdot \hat{v}_i\right), \qquad w_i = \frac{|\bar{\hat{v}}_i|^p}{\sum_{j=1}^{M} |\bar{\hat{v}}_j|^p}. \quad (5)$$

The extension to color images is straightforward. The accumulation is done channel by channel using the same Fourier weights for all channels. The weights are computed by arithmetically averaging the Fourier magnitude of the channels before the low pass filtering.

Noise Aware Sharpening. While the results of the Fourier burst accumulation are already very good, considering that the process so far has been computationally non-intensive, one can optionally apply a final sharpening step if resources are still available. The sharpening must contemplate that the reconstructed image may have some remaining noise. Thus, we first apply a denoising algorithm (we used NLBAYES [15] [1]), then on the filtered image we apply a Gaussian sharpening. To avoid removing fine details we finally add back a percentage of what has been removed during the denoising step.

Memory and Complexity Analysis. Once the images are registered, the algorithm runs in $O(M \cdot m \cdot \log m)$, where $m = m_h \times m_w$ is the number of image pixels and M the number of images in the burst. The heaviest part of the algorithm is the computation of M FFT, very suitable and popular in VLSI implementations. This is the reason why the method has a very low complexity. Regarding memory consumption, the algorithm does not need to access all the images simultaneously and can proceed in an online fashion. This keeps the memory requirements to only three buffers: one for the current image, one for the current average, and one for the current weights sum.

6. Experimental Results

We captured several handheld bursts with different number of images using a Canon 400D DSLR camera and the back camera of an iPad tablet. The full restored images and the details of the camera parameters are shown in Figure 5. The photographs contain complex structure, texture, noise and saturated pixels, and were acquired under different lighting conditions. All the results were computed using $p = 11$.

Comparison to multi-image blind deblurring. Since this problem is typically addressed by multi-image blind deconvolution techniques, we selected two state-of-the-art algorithms for comparison [25, 30]. Both algorithms are built on variational formulations and estimate first the blurring kernels using all the frames in the burst and then do a step of multi-image non-blind deconvolution, requiring significant memory for normal operation. We used the code provided by the authors. The algorithms rely on parameters that were manually tuned to get the best possible results. We also compare to the simple align and average algorithm (which indeed is the particular case $p = 0$).

Figures 6 and 7 show some crops of the restored images by all the algorithms. In addition, we show two input images for each burst: the best one in the burst and a typical one in the series. The full sequences are available in the supplementary material. The proposed algorithm obtains similar or better results than the one by Zhang *et al.* [30], at a significantly lower computational and memory cost. Since this algorithm explicitly seeks to deconvolve the sequence, if the convolution model is not perfectly valid or there is misalignment, the restored image will have deconvolution artifacts. This is clearly observed in the bookshelf sequence where [30] produces a slightly sharper restoration but having ringing artifacts (see *Jonquet* book). Also, it is hard to read the word "Women" in the spine of the red book. Due to the strong assumed priors, [30] generally leads to very sharp images but it may also produce overshooting/ringing in some regions like in the brick wall (parking night).

The proposed method clearly outperforms [25] in all the sequences. This algorithm introduces strong artifacts that degraded the performance in most of the tested bursts. Tuning the parameters was not trivial since this algorithm relies on 4 parameters that the authors have linked to a single one (named γ). We swept the parameter γ to get the best possible performance.

Our approach is conceptually similar to a regular align and average algorithm, but it produces significantly sharper images while keeping the noise reduction power of the average principle. In some cases with numerous images in the burst (e.g., see the parking night sequence), there might already be a relatively sharp image in the burst (lucky imaging). Our algorithm does not need to explicitly detect such "best" frame, and naturally uses the others to denoise the frequencies not containing image information but noise.

[1] A variant of this is already available on camera phones, so we stay at the level of potential on-board implementations.

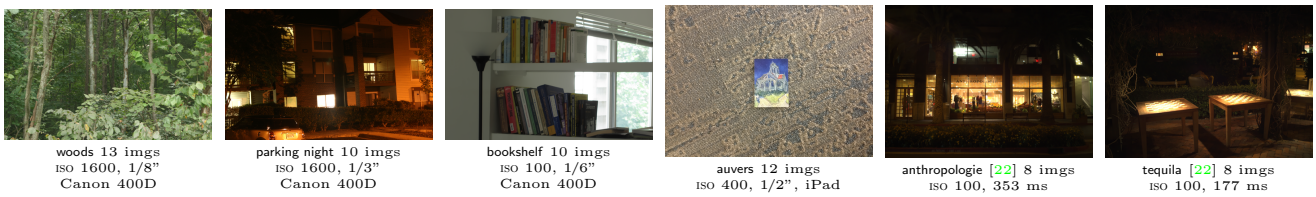


Figure 5: Restoration of image bursts captured using different cameras. Full images and additional results given in the supplementary material [8].

Execution time. Once the images are registered, the proposed approach runs in only a few seconds in our Matlab experimental code, while [30] needs several hours for bursts of 8-10 images. Even if the estimation of the blurring kernels is done in a cropped version (i.e., 200×200 pixels region), the multi-image non-blind deconvolution step is very slow, taking several hours for 6-8 megapixel images.

Multi-image non-blind deconvolution. Figure 8 shows the algorithm results in two sequences provided in [22]. The algorithm proposed in [22] uses gyroscope information present in new camera devices to register the burst and also to have an estimation of the local blurring kernels. Then a more expensive multi-image non-blind deconvolution algorithm is applied to recover the latent sharp image. Our algorithm produces similar results without explicitly solving any inverse problem nor using any information about the motion kernels.

7. Conclusions

We presented an algorithm to remove the camera shake blur in an image burst. The algorithm is built on the idea that each image in the burst is generally differently blurred; this being a consequence of the random nature of hand tremor. By doing a weighted average in the Fourier domain, we reconstruct an image combining the least attenuated frequencies in each frame.

This algorithm has several advantages. First, it does not introduce typical ringing or overshooting artifacts present in most deconvolution algorithms. This is avoided by not formulating the deblurring problem as an inverse problem of deconvolution. The algorithm produces similar or better results than the state-of-the-art multi-image deconvolution while being significantly faster and with lower memory footprint. As a future work, we would like to incorporate a gyroscope registration technique, e.g., [22], to create a real-time system for removing camera shake in image bursts. We are also exploring the use of the framework presented in this paper for HDR [18, 31] and video deblurring [7], with excellent preliminary results.

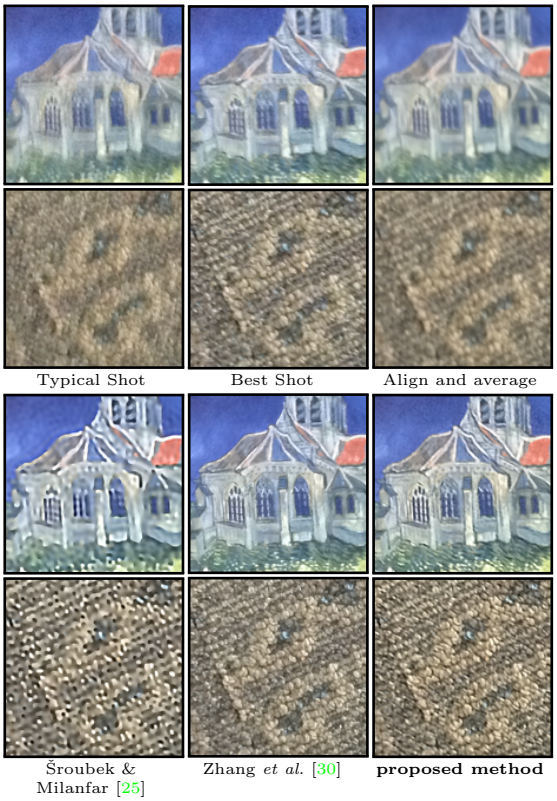


Figure 6: Real data burst deblurring results and comparison with multi-image blind deconvolution methods (auvers).

Figure 8: Restoration results with the data provided in [22] (anthropologie and tequila sequences).

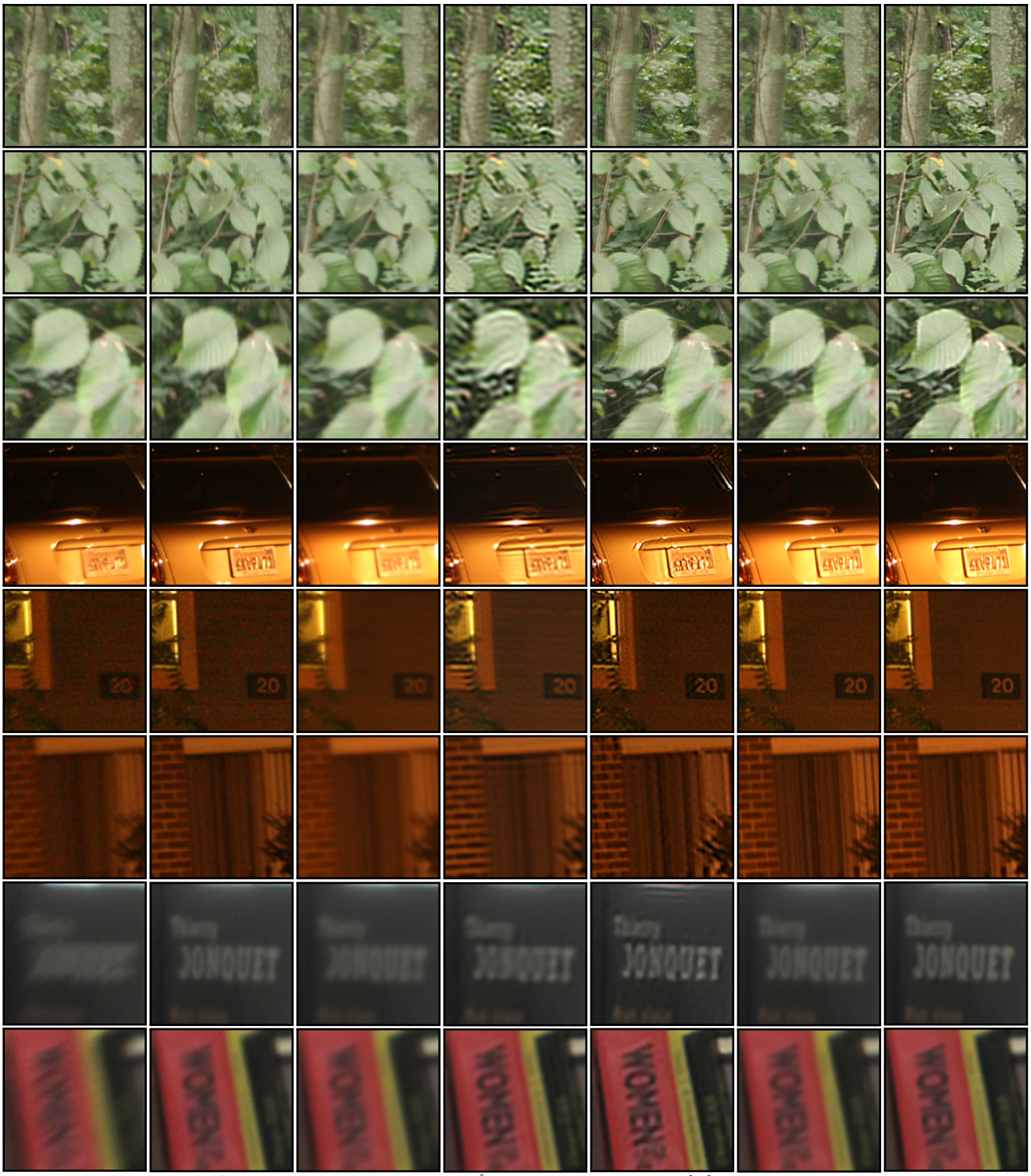


Figure 7: Real data burst deblurring results and comparison to state-of-the-art multi-image blind deconvolution algorithms (woods, parking night, bookshelf sequences).

Acknowledgment

This work was partially funded by: ONR, ARO, NSF, NGA, and AFOSR.

References

[1] G. Boracchi and A. Foi. Modeling the performance of image restoration from motion blur. *IEEE Trans. Image Process.*, 21(8):3502–3517, 2012. 5

[2] T. Buades, Y. Lou, J.-M. Morel, and Z. Tang. A note on multi-image denoising. In *LNLA*, 2009. 1, 3

[3] J.-F. Cai, H. Ji, C. Liu, and Z. Shen. Blind motion deblurring using multiple images. *J. Comput. Phys.*, 228(14):5057–5071, 2009. 2

[4] B. Carignan, J.-F. Daneault, and C. Duval. Quantifying the importance of high frequency components on the amplitude of physiological tremor. *Exp. Brain Res.*, 202(2):299–306, 2010. 1, 3, 5

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

[6] S. Cho and S. Lee. Fast motion deblurring. *ACM Trans. Graph.*, 28(5):145:1–145:8, 2009. 2

[7] S. Cho, J. Wang, and S. Lee. Vdeo deblurring for hand-held cameras using patch-based synthesis. *ACM Trans. Graph.*, 31(4):64:1–64:9, 2012. 7

[8] M. Delbracio and G. Sapiro. Burst Deblurring: Removing Camera Shake Through Fourier Burst Accumulation (Supplementary Material). http://dev.ipol.im/~mdelbra/fba, 2015. 7

[9] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. *ACM Trans. Graph.*, 25(3):787–794, 2006. 2

[10] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 24(6):381–395, 1981. 5

[11] F. Gavant, L. Alacoque, A. Dupret, and D. David. A physiological camera shake model for image stabilization systems. In *IEEE Sensors*, 2011. 1, 2, 3, 4, 5

[12] A. Ito, A. C. Sankaranarayanan, A. Veeraraghavan, and R. G. Baraniuk. Blurburst: Removing blur due to camera shake using multiple images. *ACM Trans. Graph.*, Submitted. 3

[13] D. Krishnan, T. Tay, and R. Fergus. Blind deconvolution using a normalized sparsity measure. In *CVPR*, 2011. 2

[14] D. Kundur and D. Hatzinakos. Blind image deconvolution. *IEEE Signal Process. Mag.*, 13(3):43–64, 1996. 2

[15] M. Lebrun, A. Buades, and J.-M. Morel. Implementation of the "Non-Local Bayes" Image Denoising Algorithm. *IPOL*, 3:1–42, 2013. 6

[16] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Understanding and evaluating blind deconvolution algorithms. In *CVPR*, 2009. 2

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

[18] M. Levoy. HDR+: Low Light and High Dynamic Range photography in the Google Camera App. http://googleresearch.blogspot.fr/2014/10/hdr-low-light-and-high-dynamic-range.html. 7

[19] D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vision*, 60:91–110, 2004. 5

[20] T. Michaeli and M. Irani. Blind deblurring using internal patch recurrence. In *ECCV*, 2014. 2

[21] L. Moisan, P. Moulon, and P. Monasse. Automatic Homographic Registration of a Pair of Images, with A Contrario Elimination of Outliers. *IPOL*, 2:56–73, 2012. 5

[22] S. H. Park and M. Levoy. Gyro-based multi-image deconvolution for removing handshake blur. *CVPR*, 2014. 3, 5, 7

[23] A. Rav-Acha and S. Peleg. Two motion-blurred images are better than one. *Pattern Recogn. Lett.*, 26(3):311–317, 2005. 2

[24] Q. Shan, J. Jia, and A. Agarwala. High-quality motion deblurring from a single image. *ACM Trans. Graph.*, 27(3), 2008. 2

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

[26] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce. Non-uniform deblurring for shaken images. *Int. J. Comput. Vision*, 98(2):168–186, 2012. 1

[27] F. Xiao, A. Silverstein, and J. Farrell. Camera-motion and effective spatial resolution. In *ICIS*, 2006. 1, 3

[28] L. Xu, S. Zheng, and J. Jia. Unnatural l0 sparse representation for natural image deblurring. In *CVPR*, 2013. 2

[29] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum. Image deblurring with blurred/noisy image pairs. *ACM Trans. Graph.*, 26(3), 2007. 2

[30] H. Zhang, D. Wipf, and Y. Zhang. Multi-image blind deblurring using a coupled adaptive sparse prior. In *CVPR*, 2013. 1, 2, 6, 7, 8

[31] L. Zhang, A. Deshpande, and X. Chen. Denoising vs. deblurring: Hdr imaging techniques using moving cameras. In *CVPR*, 2010. 7

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

[33] B. Zitova and J. Flusser. Image registration methods: a survey. *Image Vision Comput.*, 21(11):977–1000, 2003. 5