Burst Deblurring: Removing Camera Shake Through Fourier Burst Accumulation

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

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.

Under reasonable hypotheses, camera shake blur can be modeled mathematically as a convolution, $v = u \star k + n$, 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. 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 k 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 (see e.g., [1]). 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 [5], 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 [2]. This implies that the blur in one frame will be different from the one in another image of the burst. 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. We also completely avoid the explicit computation of the blurring kernel, 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.

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:

Claim 1. *Blurring kernels do not amplify the spectrum. That is,* $|\hat{k}(\zeta)| \leq 1$.

Let us assume that the photographer takes a burst of M images of the same scene u,

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

The movement of the camera during any two images of the burst is essentially independent, making the blurring kernels k_i to be 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, the reconstructed image picks what is less attenuated from each image of the burst.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

Figure 1: Comparison to state-of-the-art multi-image blind deconvolution algorithms (real data).

Fourier Burst Accumulation. We call *Fourier Burst Accumulation (*FBA*)* to the Fourier weighted averaged image,

$$u_p(\mathbf{x}) = \mathcal{F}^{-1}\left(\sum_{i=1}^{M} w_i(\zeta) \cdot \hat{v}_i(\zeta)\right)(\mathbf{x}), \qquad w_i(\zeta) = \frac{|\hat{v}_i(\zeta)|^p}{\sum_{j=1}^{M} |\hat{v}_j(\zeta)|^p}, \tag{2}$$

where p is a non-negative integer and $\hat{v}_i$ is the Fourier Transform of the individual burst image v_i. The weight $w_i(\zeta)$ 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 (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.

The Fourier weights only depend on the Fourier magnitude and hence they are not sensitive to image misalignment. However, when doing the average in (2), the images v_i have to be correctly aligned to mitigate Fourier phase intermingling and get a sharp aggregation. In our experiments, images were aligned by estimating the dominant homography between each image and the first one in the set. This pre-alignment step can be done exploiting the camera gyroscope and accelerometer data (e.g., [3]).

The proposed method has several advantages. First, it does not introduce typical ringing or overshooting artifacts present in most deconvolution algorithms (Fig. 1). 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.

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

[2] F. Gavant, L. Alacoque, A Dupret, and D. David. A physiological camera shake model for image stabilization systems. In *IEEE Sens.*, 2011.

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

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

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