

Removing Camera Shake via Weighted Fourier Burst Accumulation

Sowmyan.V¹, Vasanth.M², Vignesh.M.R³, Vinothkumar.S⁴, Dr.Sumathi.S⁵

UG Scholar, Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India^{1,2,3,4}

Professor, Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India⁵

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 can be seen as a generalization of the align and average procedure, with a weighted average, motivated by hand-shake physiology and theoretically supported, taking place in the Fourier domain. 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, and extensive comparisons, 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. Finally, we also present experiments in real high dynamic range (HDR) scenes, showing how the method can be straightforwardly extended to HDR photography.

KEYWORDS: Multi-image blurring, burst fusion, camera shake, lowlight photography, high dynamic range.

I. INTRODUCTION

One of the most challenging experience 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 dimlight conditions. 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). However, this principle is disturbed if the images in the burst have blurred. 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, 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. 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 different 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 (>1/100s), 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 artefacts 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.

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). The review by Kundur and Hatzinakos, 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. This variation 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.

II. LITERATURE SURVEY

A. A Note On Multi-Image Denoising:

Taking photographs under low light conditions with a handheld camera is problematic. A long exposure time can cause motion blur due to the camera shaking and a short exposure time gives a noisy image. We consider the new technical possibility offered by cameras that take image bursts. Each image of the burst is sharp but noisy. In this preliminary investigation, we explore a strategy to efficiently denoise multi-images or video. The proposed algorithm is a complex image processing chain involving accurate registration, video equalization, noise estimation and the use of state-of-the-art denoising methods. Yet, we show that this complex chain may become risk free thanks to a key feature: the noise model can be estimated accurately from the image burst. Preliminary tests will be presented. On the technical side, the method can already be used to estimate a non-parametric camera noise model from any image burst.

B. Multi-Image Blind Deblurring Using A Coupled Adaptive Sparse Prior:

This paper presents 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, blur kernels, and noise levels together in a unique way. This coupled penalty function enjoys a number of desirable properties, including a mechanism whereby the relative-concavity or shape is adapted as a function of the intrinsic quality of each blurry observation. In this way, higher quality observations may automatically contribute more to the final estimate than heavily degraded ones. The resulting algorithm, which requires no essential tuning parameters, can recover a high quality 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.

C. Camera-Motion and Effective Spatial Resolution:

The booming mobile imaging market is aggressively pursuing CMOS imagers with higher spatial resolution and smaller form factors. Given a fixed die size, reducing pixel size seems like a straightforward way to increase spatial resolution. The reduced sensitivity of smaller pixels, however, must be compensated by longer exposure durations. Longer exposure durations result in camera motion blur when the camera is hand-held, effectively reducing spatial resolution. We empirically measured how the exposure duration, camera mass and subject's skill level affect camera motion. We then used these measurements to calculate the effect that camera motion has upon the system MTF for CMOS imagers with different pixel sizes at different scene luminance levels. We described the scene luminance conditions for which imaging sensors with larger pixels produce sharper images than imagers with smaller pixel sizes.

D. Removing Camera Shake from a Single Photograph

Camera shakes during exposure leads to objectionable image blur and ruins many photographs. Conventional blind deconvolution methods typically assume frequency-domain constraints on images, or overly simplified parametric forms for the motion path during camera shake. Real camera motions can follow convoluted paths, and a spatial domain prior can better maintain visually salient image characteristics. We introduce a method to remove the effects of camera shake from seriously blurred images. The method assumes a uniform camera blur over the image and negligible in-plane camera rotation. In order to estimate the blur from the camera shake, the user must specify an image region without saturation effects. We show results for a variety of digital photographs taken from personal photo collections.

E. High Resolution Imaging in the Visible from the Ground without Adaptive Optics: New Techniques and Results

Many astronomical imaging studies, such as those of weak gravitational lensing, call for better angular resolution than is normally possible from the ground over wide fields of view. For many of these studies astronomers need images which show a consistent point spread function across the field even if this comes at the expense of the ultimate in angular resolution at the center of the field. Adaptive Optics does not show any prospect of being able to achieve fields of view as large as are needed at visible wavelengths and therefore a new technique of delivering high resolution images from the ground must be developed. Electron multiplying CCDs are available that allow images to be taken at high speed without the usual penalty of read noise. We have developed a new technique called Lucky Imaging which achieves high resolution by selecting the better images from a sequence of images, then shifting and adding each to give a much higher resolution output image. Resolutions in the range 0.1-0.2 arc seconds can be obtained routinely under relatively good conditions on a 2.5 meter telescope working in I band (850 nanometers) and using as much as 30% of the images taken. Even under poorer conditions we find that image selection allows the final resolution to be better than the traditional seeing value by a factor of as much as three. This paper describes the technique and some of the results obtained using this method.

III. EXISTING METHOD

A. Blind Deconvolution:

Two or more input images can improve the estimation of both the underlying image and the blurring kernels. "Two motion-blurred images are better than one," if the direction of the blurs are different. 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

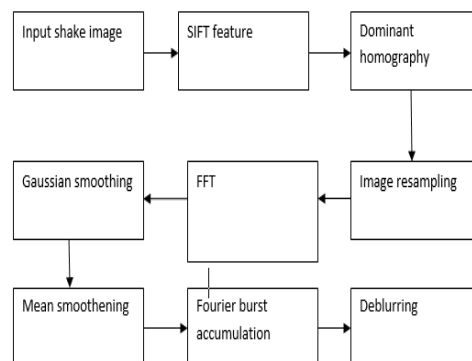
estimate the motion kernel of the blurred one. Close to our work are papers on multi-image blind deconvolution, the authors showed that given multiple observations, the sparsity of the image under a tight frame is a good measurement of the clearness of the recovered image. Having multiple input images improves the accuracy of identifying the motion blur kernels, reducing the ill-posedness of the problem. Most of these multi-image algorithms introduce cross-blur penalty functions between image pairs. However this has the problem of growing combinatorial with the number of images in the burst. 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).

B. Disadvantages:

1. The major drawback of the BD method is its lack of reliability
2. The major disadvantage of BD is the convergence to the global minimum of the cost function is slow.
3. The PSNR value is low.
4. It have poor convergence property.

IV. PROPOSED METHOD

A. Block Diagram:



B. SIFT Feature:

Scale-invariant feature transform (SIFT) is an algorithm in computer vision to detect and describe local features in images. The algorithm was patented in the US by the University of British Columbia and published by David Lowe in 1999. Applications include object recognition, robotic mapping and navigation, image stitching, 3D modeling, gesture recognition, video tracking, individual identification of wildlife and match moving.

SIFT key points of objects are first extracted from a set of reference images and stored in a database. An object is recognized in a new image by individually comparing each feature from the new image to this database and finding candidate matching features based on Euclidean distance of their feature vectors. From the full set of matches, subsets of key points that agree on the object and its location, scale, and orientation in the new image are identified to filter out good matches. The determination of consistent clusters is performed rapidly by using an efficient hash table implementation of the generalized Hough transform. Each cluster of 3 or more features that agree on an object and its pose is then subject to further detailed model verification and subsequently outliers are discarded. Finally the probability that a particular set of features indicates the presence of an object is computed, given the accuracy of fit and number of probable false matches. Object matches that pass all these tests can be identified as correct with high confidence.

C. Dominant Homography:

In the field of computer vision, any two images of the same planar surface in space are related by a homography (assuming a pinhole camera model). This has many practical applications, such as image rectification, image registration, or computation of camera motion—rotation and translation—between two images. Once camera rotation and translation have been extracted from an estimated homography matrix, this information may be used for navigation, or to insert models of 3D objects into an image or video, so that they are rendered with the correct perspective and appear to have been part of the original scene.

D. Gaussian Smoothing:

The idea of Gaussian smoothing is to use this 2-D distribution as a 'point-spread' function, and this is achieved by convolution. Since the image is stored as a collection of discrete pixels we need to produce a discrete approximation to the Gaussian function before we can perform the convolution. In theory, the Gaussian distribution is non-zero everywhere, which would require an infinitely large convolution kernel, but in practice it is effectively zero more than about three standard deviations from the mean, and so we can truncate the kernel at this point. Figure 3 shows a suitable integer-valued convolution kernel that approximates a Gaussian with an σ of 1.0. It is not obvious how to pick the values of the mask to approximate a Gaussian. One could use the value of the Gaussian at the center of a pixel in the mask, but this is not accurate because the value of the Gaussian varies non-linearly across the pixel. We integrated the value of the Gaussian over the whole pixel (by summing the Gaussian at 0.001 increments). The integrals are not integers: we rescaled the array so that the corners had the value 1. Finally, the 273 is the sum of all the values in the mask

E. Mean Smoothing:

The idea of mean filtering is simply to replace each pixel value in an image with the mean ('average') value of its neighbors, including itself. This has the effect of eliminating pixel values which are unrepresentative of their surroundings. Mean filtering is usually thought of as a convolution filter. Like other convolutions it is based around a kernel, which represents the shape and size of the neighborhood to be sampled when calculating the mean. Often a 3x3 square kernel is used, although

larger kernels (e.g. 5x5 squares) can be used for more severe smoothing. (Note that a small kernel can be applied more than once in order to produce a similar but not identical effect as a single pass with a large kernel.)

F. Fourier Burst Accumulation:

Let 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 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 $\int k(x) dx = 1$. 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.

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

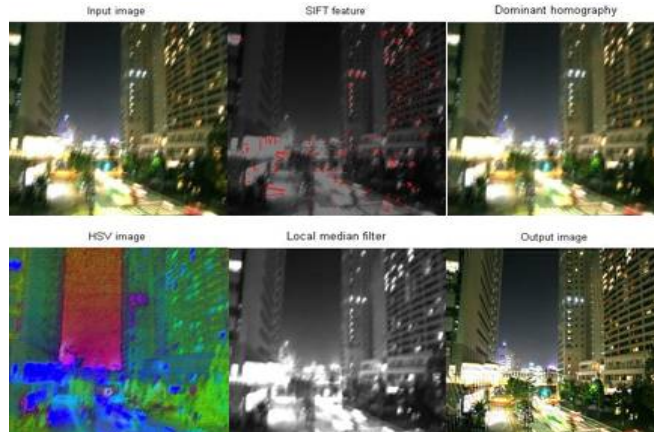
Proof:

$$|\hat{k}(\zeta)| = \left| \int k(x) e^{ix \cdot \zeta} dx \right| \leq \int |k(x)| dx = \int k(x) dx = 1.$$

Most modern digital cameras have a burst mode where the photographer is allowed to sequentially 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 ,

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

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.



G. Deblurring:

Deblurring is the process of removing blurring artifacts from images, such as blur caused by defocus aberration or motion blur. The blur is typically modeled as the convolution of a (sometimes space- or time-varying) point spread function with a hypothetical sharp input image, where both the sharp input image (which is to be recovered) and the point spread function are unknown. This is an example of an inverse problem. In almost all cases, there is insufficient information in the blurred image to uniquely determine a plausible original image, making it an ill-posed problem. In addition the blurred image contains additional noise which complicates the task of determining the original image. This is generally solved by the use of a regularization term to attempt to eliminate implausible solutions. This problem is analogous to echo removal in the signal processing domain. Nevertheless, when coherent beam is used for imaging, the point spread function can be modeled mathematically. As the figure on the right illustrates, by proper deconvolution of the point spread function and the image, the resolution can be enhanced several times.

V. ALGORITHM

Input: A series of images $\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_n$ of size $m \times n \times c^*$. An integer value p .

Output : The aggregated image u_p

```

1  $w = \text{zeros}(m, n); \tilde{u}_p = \text{zeros}(m, n, c);$ 
2 for image  $i$  in  $\{1, \dots, n\}$  do
    Burst Registration
3  $M_i = \text{SIFT}(\tilde{v}_i, \tilde{v}_1);$   $M_i$  set of corresponding points.
4  $H_i = \text{ORSA}(M_i);$   $H_i$  dominant homography in  $M_i$ .
5  $v_i = \tilde{v}_i \circ H_i;$  Image resampling.
    Fourier Burst Accumulation
6  $\hat{v}_i = \text{FFT}(v_i);$ 
7  $w_i = \frac{1}{c} \sum_{j=1}^c |\hat{v}_i^j|;$  Mean over color channels
8  $w_i = G_\sigma w_i;$  Gaussian smoothing
9  $\tilde{u}_p = \tilde{u}_p + w_i \cdot \hat{v}_i;$  Weighted Fourier accumulation
10  $w = w + w_i;$ 
11  $u_p = \text{IFFT}(\tilde{u}_p / w);$ 
    Noise Aware Sharpening (Optional)
12  $\bar{u}_p = \text{DENOISE}(u_p);$ 
13  $\bar{u}_p^s = 2\bar{u}_p - G_\rho \bar{u}_p;$  Gaussian sharpening,  $\rho \in [1, 3]$ 
14  $u_p = \bar{u}_p^s + \delta(u_p - \bar{u}_p);$  Add a fraction of removed noise,  $\delta = 0.4$ 

```

[*] c is the number of color channels, typically 3. The color channels of v are denoted by v^j , for $j = 1, \dots, c$. All the regular operations (e.g. $+$, $/$, \cdot) are point-wise.

IV. CONCLUSION

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. Experimental results showed that there constructed image is sharp and less noisy than the original ones. 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 produce similar or better results than the state-of-the- art multi- image deconvolution while being significantly faster and with lower memory footprint. Weal so presented a direct application of the Fourier Burst Accumulation algorithm to HDR imaging with a handheld camera.

REFERENCES

1. T.Buades, Y.Lou, J.-M. Morel, and Z.Tang, "Anoteonmulti-image denoising,"in*Proc.Int. Workshop Local Non-Local Approx. Image Process.(LNLA)*,Aug.2009
2. R. Fergus, B.Singh, A. Hertzmann, S.T.Roweis, and W. T.Freeman, "Removing camerashakefromasingle photograph," *ACM Trans. Graph.*,2006.
3. L.Yuan, J.Sun, L. Quan, and H.-Y. Shum, "Image deblurring with blurred/noisyimagepairs,"*ACMTrans.Graph.*,2007
4. X.Zhu,F.Šroubek, and P.Milanfar,"DeconvolvingPSFsforabetter motion deblurring using multiple images,"in*Proc.IEEE12th Eur. Conf. Comput. Vis.(ECCV)*,Oct.2012
5. A.I to, A.C .Sankaranarayanan, A.Veeraraghavan, and R.G. Baraniuk, "Blurburst: Removing blur due to cameras hake using multipleimages," *ACM Trans. Graph.* ,to be published