LUCKY DCT AGGREGATION FOR CAMERA SHAKE REMOVAL

Sanjay Ghosh, Satyajit Naik, and Kunal N. Chaudhury

Department of Electrical Engineering, Indian Institute of Science

ABSTRACT

We consider the task of removing the effect of camera shake during a long exposure. Technically, this is a blind deconvolution problem in which both the image and the motion blur have to be jointly inferred. Several algorithms have been proposed till date for removing camera shake that work with one or more images. However, most of these algorithms are computationally expensive and hence cannot be used in real-time. In this work, we propose a simple and cheap algorithm that can effectively recover the original sharp image from multiple *burst* images (captured using the burst modality of modern cameras). In summary, we pick selected images from the burst (using ideas from lucky imaging), which are then aggregated using the discrete cosine transform (similar to the idea of Fourier burst accumulation). We present some preliminary results and comparisons to demonstrate the effectiveness of the proposal.

Index Terms— Burst imaging, camera shake removal, discrete cosine transform, lucky imaging.

1. INTRODUCTION

Low-light imaging with hand-held cameras is a challenging task in photography. In a dimly lit environment, a short exposure is not sufficient to accumulate enough photons, which can result in noisy captured images. One solution is to use a longer exposure time. However, images obtained with a long exposure are often blurred due to the movement of the photographer's hand. The blurring can be fixed by applying a standard blind deconvolution algorithm. In one class of algorithms, the blur is first estimated from the image, which is then used as an input to the deblurring algorithm [1]. An alternative approach is to simultaneously estimate the sharp image and the blurring kernel [2]. We note that the problem of estimating the sharp image is generally ill-posed even in non-blind deblurring (i.e., when the blurring kernel is known).

In recent years, the concept of multiple-image blind deconvolution has gained popularity. It was shown in [3, 4, 5] that multiple blurred images (with various exposures) can be used to generate a sharp image. A well-known technique in astronomical imaging called *lucky imaging* captures a series of short-exposure images and then fuses the sharper images to obtain a single high-quality image. Applying inversion-based methods on large astronomical images is computationally infeasible, and hence the need for a cheap algorithm. The sharp images are selected using the brightest pixel [6], the maximum intensity of the Laplacian [7], or the image gradient [8]. A robust sharpness measure that is widely used in astronomical imaging is the so-called Dirichlet energy [9]. In a recent method, a sharp image was reconstructed using a weighted average of multiple input images, where the weights are determined using the Fourier spectrum [10]. The idea was later extended to video deblurring [11].

Most modern digital cameras come with a burst mode for capturing a series of images in quick succession. In this work, we consider the problem of recovering a sharp image from such a burst. The mathematical model is that we have blurred versions $y_1, y_2, ..., y_N$ of a sharp image x:

$$y_i = k_i * x + \sigma n_i \qquad (i = 1, \ldots, N), \tag{1}$$

where (k_i) are the blurring kernels, (n_i) are i.i.d. $\mathcal{N}(0, 1)$, and σ is the noise level. The problem is to recover the unknown image x from $y_1, y_2, ..., y_N$. The kernels are used to model the camera shake. The primary source of camera shake in hand-held photography is hand tremor. It has been observed that the kernels follow random trajectories (cf. Figure 1). The random nature of a camera shake plays a crucial role in [10]. In this paper, we combine ideas from lucky imaging [6] and Fourier Burst Accumulation [10] to develop a simple yet efficient technique to fuse burst images. In particular, our algorithm is based on the idea of rejecting "outlier" images used in the former and that of Fourier aggregation used in the latter. We empirically demonstrate that by combining these ideas, we can improve the deblurring and also reduce the computation time.

The rest of the paper is organized as follows. The proposed algorithm is described in Section 2. Simulation results on various datasets and comparisons with [10] are provided in Section 3. We conclude the paper in Section 4.

2. LUCKY DCT AGGREGATION

The proposed method is built upon two distinct lines of prior research on multi-image deblurring – lucky imaging using Dirichlet energy [9] and Fourier Burst Accumulation (FBA) [10, 11]. The latter approach exploits the randomness of hand

Correspondence: sghosh@ee.iisc.ernet.in. This work was supported by a Startup Grant from IISc Bangalore and an EMR Grant SERB/F/6047/2016-2017 from the Department of Science and Technology, Government of India.

tremors. Moreover, it is assumed that camera movements in individual images of the burst are independent. The entire set of images are aggregated using weights computed in the Fourier domain. On the other hand, only the sharp images are aggregated in [9], where the integral Dirichlet energy is used as the sharpness criteria. The Dirichlet energy for an image y is defined as

$$\mathcal{E} = \sum_{\ell \in \text{support}(y)} \sum_{\Omega_\ell} \|\nabla y(\ell)\|^2. \tag{2}$$

where ∇y is (discrete) gradient of y and Ω_ℓ is an $n \times n$ window around pixel ℓ. A typical value of n is 100.

The main difference with [10] is that we take the weighted average of select sharp images from the burst. In particular, we sort the images according to decreasing Dirichlet energy, and aggregate the first N_0 images. The weighted averaging is performed using the type-II discrete cosine transform (DCT) [12]. Unlike the discrete Fourier transform (DFT) used in [10], the DCT coefficients are real valued. Moreover, the DCT is known to exhibit better energy compaction than the DFT [12]. As an aside, the use of DCT also makes the approach more hardware efficient.

Let Y_i denote the Gaussian-smoothed version of the DCT of burst image y_i, that is, $Y_i = \mathcal{G}(\mathcal{D}(y_i))$, where $\mathcal{G}$ denotes the two-dimensional Gaussian filter and $\mathcal{D}$ is the forward DCT. For some non-negative integer p, we define the weights

$$\bar{w}_i(\nu) = \frac{|Y_i(\nu)|^p}{\sum_{j=1}^{N_0} |Y_i(\nu)|^p}, \tag{3}$$

where ν is the frequency index for the DCT. The integer p controls the nature of the aggregation process. When $p \to \infty$, we essentially select the coefficient with largest magnitude; on the other hand, when $p = 0$, the reconstructed image is just the arithmetic average of the burst. Prior to the aggregation step, each $\bar{w}_i$ in (3) is Gaussian filtered, that is, we compute $w_i = \mathcal{G}(\bar{w}_i)$. The DCT of the aggregated image is set to be

$$\hat{X}(\nu) = \frac{\sum_{i=1}^{N_0} w_i(\nu) Y_i(\nu)}{\sum_{i=1}^{N_0} w_i(\nu)}. \tag{4}$$

In other words, the aggregated image is given by $\hat{x} = \mathcal{D}^{-1}(\hat{X})$, where $\mathcal{D}^{-1}$ stands for the inverse DCT. The complete process is summarized in Algorithm 1. The symbols $\oplus, \otimes$, and $\oslash$ in Algorithm 1 denote pixelwise addition, multiplication, and division performed on each of the c channels.

Algorithm 1: Lucky DCT Aggregation

Input: Images $y_1, y_2, \ldots, y_N$ of size $M_1 \times M_2 \times c$.
Parameters: Integers $p \geq 0$ and $N_0 \leq N$.
Output: Output image $\hat{x}$.
Initialize: Null images w, Y, Z of size $M_1 \times M_2 \times c$;
for $i = 1, 2, \ldots, N$ **do**
  Compute $\mathcal{E}_i$ using (2);
end
Rank images according to decreasing $\mathcal{E}_i$ values;
Select first N_0 images;
for $i = 1, 2, \ldots, N_0$ **do**
  $\bar{Y}_i = \mathcal{D}(y_i)$; % DCT
  $Y_i = \mathcal{G}(\bar{Y}_i)$; % Smoothing
  $Z = Z \oplus |Y_i|^p$;
end
for $i = 1, 2, \ldots, N_0$ **do**
  $\bar{w}_i = |Y_i|^p \oslash Z$;
  $w_i = \mathcal{G}(\bar{w}_i)$; % Smoothing
  $Y = Y \oplus \big(w_i \otimes Y_i\big)$ % Aggregation
  $w = w \oplus w_i$;
end
$\hat{X} = Y \oslash w$; % Normalization
$\hat{x} = \mathcal{D}^{-1}(\hat{X})$. % Inverse DCT

3. EXPERIMENTS

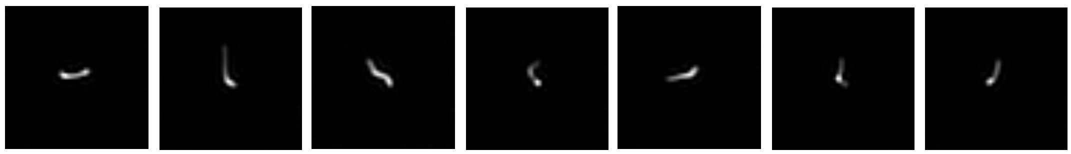
We now present some results for the proposed method. For the simulations, we have used the synthetic kernels[1] shown in Figure 1. These were generated by simulating the random nature of camera shake. For simplicity, we have assumed that the images in the burst are perfectly aligned; this is indeed the case when the support of the kernel is small [10]. Figure 2 shows the blurred images obtained via model (1) along with the Dirichlet energies. The Dirichlet energy is relatively smaller for images that appear to be more blurred. We aggregate the first N_0 images with largest Dirichlet energies. Visual comparisons between the proposed method and FBA [10] are provided in Figures 3 and 4. The results obtained using the proposed method exhibit better visual appearance (cf. the zoomed sections in Fig. 4). Moreover, the proposed method is faster by a factor of about N_0/N, neglecting the overhead of computing the Dirichlet energies and ranking them. The effect of N_0 parameter on the performance is depicted in Figure 5. Notice that the sharpest image (one with largest Dirichlet energy) is quantitatively better than the image obtained by aggregating the full set of images. However, by aggregating two of the sharpest images, we get an optimal result. The performance under different noise levels is shown in Figure 6. Our method outperforms FBA [10] by a margin of at least 0.5 dB for a wider range of noise levels. The gain in performance is mainly due to the outlier rejection.

[1] The kernels were obtained from the site `http://dev.ipol.im/~mdelbra/fba/`

4. CONCLUSION

In this paper, we demonstrated how by combining ideas from lucky imaging and Fourier Burst Accumulation, we can obtain

(a) k_1. (b) k_2. (c) k_3. (d) k_4. (e) k_5. (f) k_6. (g) k_7.



(h) k_8. (i) k_9. (j) k_{10}. (k) k_{11}. (l) k_{12}. (m) k_{13}. (n) k_{14}.

Fig. 1. Camera shake kernels.

(a) $y_1, 2.26 \times 10^{12}$. (b) $y_2, 1.26 \times 10^{12}$. (c) $y_3, 1.81 \times 10^{12}$. (d) $y_4, 1.92 \times 10^{12}$. (e) y_5 3.19×10^{12}. (f) $y_6, 1.59 \times 10^{12}$. (g) $y_7, 1.28 \times 10^{12}$.

(h) $y_8, 2.25 \times 10^{12}$. (i) $y_9, 2.60 \times 10^{12}$. (j) $y_{10}, 2.01 \times 10^{12}$. (k) $y_{11}, 2.72 \times 10^{12}$. (l) $y_{12}, 2.42 \times 10^{12}$. (m) $y_{13}, 3.47 \times 10^{12}$. (n) $y_{14}, 2.76 \times 10^{12}$.

Fig. 2. Images obtained via model (1) where x is the sharp image in Fig. 3(a), k_i are the kernels shown in Fig. 1; and the additive Gaussian noise of $\sigma = 5$. The corresponding Dirichlet energies are shown in the caption.

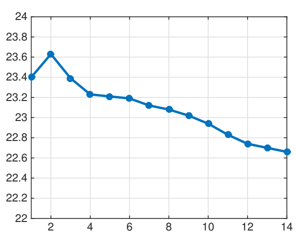
(a) Input image (x) of size (1024×1024). (b) FBA [10], **22.88 / 88.11**, 4.95 sec. (c) Proposed, **23.37 / 90.51**, 1.74 sec.

Fig. 3. Deblurring results for the images in Fig. 2 using FBA [10] and the proposed method. In particular, we used just images y_5 and y_{13} in the aggregation process. The quality indices (PSNR / SSIM) and the run-times are mentioned in the caption.

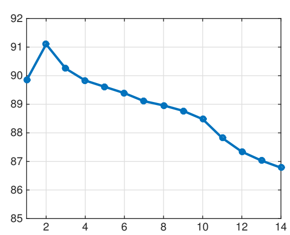
(a) Input image (512 × 768). (b) FBA, **23.32 / 83.95**, 1.76 sec. (c) Proposed, **24.22 / 84.20**, 0.62 sec.

(d) Zoomed section of (b). (e) Zoomed section of (c).

Fig. 4. Results obtained using the kernels in Fig. 1 and when $\sigma = 5$. We used $N_0 = 2$ for our method (that is, just two images with largest Dirichlet energies were selected). The quality indices (PSNR / SSIM) and the run-times are mentioned in the caption.



(a) PSNR.



(b) SSIM.

Fig. 5. Variation of PSNR and SSIM with N_0. The burst was obtained using model (1), where x is the sharp image in Fig. 3(a), k_i are the kernels in Fig. 1, and $\sigma = 0$.

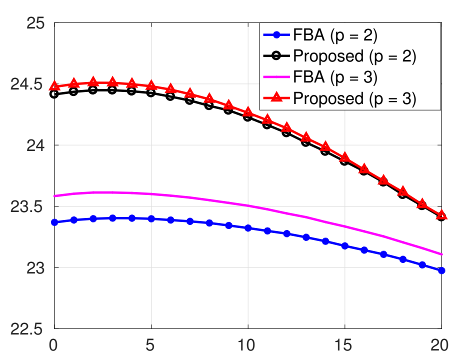


Fig. 6. PSNR *vs* σ for different algorithms. The input images (burst) were obtained using model (1), where x is the sharp image in Fig. 4(a), k_i are the kernels in Fig. 1, and $\sigma \in [0, 50]$.

an improved method for camera shake correction. In particular, our algorithm was shown to achieve better and faster recovery. The proposed approach can be extended to remove camera shakes from videos. This will investigated in future work.

5. REFERENCES

[1] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "Blind motion deblurring from a single image using sparse approximation," in Proc. *IEEE Conference in Computer Vision & Pattern Recognition (CVPR),* Jun. 2009, pp. 104-111.

[2] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Efficient marginal likelihood optimization in blind deconvolution," in Proc. *IEEE Conference in Computer Vision & Pattern Recognition (CVPR),* Jun. 2011, pp. 2657-2664.

[3] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "Blind motion deblurring using multiple images," *Journal of Comput. Phys.,* vol. 228, no. 14, pp. 5057-5071, 2009.

[4] F. Sroubek and P. Milanfar, "Robust multichannel blind deconvolution via fast alternating minimization," *IEEE Transactions in Image Processing,* vol. 21, no. 4, pp. 1687-1700, Apr. 2012.

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

[6] N. M. Law, C. D. Mackay, and J. E. Baldwin, “Lucky imaging: High angular resolution imaging in the visible from the ground,” *Astron. Astrophysics,* vol. 446, no. 2, pp. 739-745, 2006.

[7] N. Joshi and M. F. Cohen, “Seeing Mt. Rainier: Lucky imaging for multi-image denoising, sharpening, and haze removal,” in Proc. *IEEE International Conference in Computational Photography (ICCP)*, Mar. 2010, pp. 1-8.

[8] S. John and M. A. Vorontsov, “Multiframe selective information fusion from robust error estimation theory,” *IEEE Transactions in Image Processing,* vol. 14, no. 5, pp. 577-584, May 2005.

[9] G. Haro, A. Buades, and J.-M. Morel, “Photographing paintings by image fusion,” *SIAM Journal in Imaging Science,* vol. 5, no. 3, pp. 1055-1087, 2012.

[10] M. Delbracio and G. Sapiro, “Removing camera shake via weighted fourier burst accumulation,” *IEEE Transactions in Image Processing,* vol. 24, no. 11, pp. 3293-3307, Nov. 2015.

[11] M. Delbracio and G. Sapiro,“Hand-held video deblurring via efficient Fourier aggregation,” *IEEE Transactions on Computational Imaging,* vol. 1, no. 4, pp. 270-283, 2015.

[12] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.