FRAGMENTED APERTURE IMAGING FOR MOTION AND DEFOCUS DEBLURRING

Manuel Martinello, Paolo Favaro

School of EPS, Heriot-Watt University, Edinburgh, EH14 4AS, United Kingdom

ABSTRACT

In this paper we present a space-varying deblurring algorithm from a single defocused and motion-blurred image obtained with a fragmented aperture. We show that, for the same overall incoming light, a fragmented aperture leads to better motion and defocus deblurring than a (compact) conventional circular aperture. We demonstrate that not only fragmented apertures preserve more spectrum of an image of the scene than traditional circular apertures, but they also allow a better identification of blur scale. Our algorithm estimates both motion blur magnitude and direction as well as defocus blur scale at each pixel. The estimation of the blur parameters is addressed by using local projections on subspaces and L_1 regularization, while deblurring is posed as a variational minimization problem and solved via linearization of the Euler-Lagrange equations. The technique produces convincing results on real scenario.

Index Terms— motion-blur, defocus, single image, space-varying deblurring, coded aperture photography.

1. INTRODUCTION

With the increasing resolution of the camera sensors, a small motion of either an object in the scene (e.g., the bus in Fig.1) or the camera is enough to loose important details when capturing an image. Moreover, when one brings an object into focus, the texture placed at other locations in the scene may result out-of-focus (e.g., see the shops in the background in Fig.1). Hence, to recover the sharp texture from degraded images, one has to remove both defocus and motion-blur.

This paper presents two main contributions: 1) We show experimentally that defocus and motion-blur identification benefits from fragmenting the lens aperture; 2) We introduce for the first time an efficient technique to identify and perform space-varying defocus and motion deblurring from a single image.

1.1. Prior Work

Defocus Deblurring from Single Image. It has been recently demonstrated that a mask in the lens aperture can improve the defocus deblurring from a single image [1, 2, 3, 4]. Optimal patterns are obtained by imposing that the blur kernel be wide band and by incorporating natural image statistics [2, 3]. The results are quite impressive as not only one can restore a sharp image, but also a depth map can be estimated. These methods deal with the space-varying nature of real images, but do not investigate how motion blur affects the deblurring.

This work has been partly supported by SELEX Galileo grant SELEX/HWU/2010/SOW3.

Fig. 1. **Challenging scene.** Example of a blurred image captured with a conventional camera, where the degradation is due to both defocus (background) and motion (bus).

Motion Deblurring from Single Image. Motion and defocus deblurring from a single image when the scene is approximately a fronto-parallel plane has been long known in the field of signal processing as blind deconvolution [5, 6]. Recently, it has received renewed attention due to progress achieved by using natural image priors [7, 8, 9, 10]. For this choice of priors, currently [11] achieves the best results and can deal with very large (although uniform) blurs. Other recent methods deal with non-uniform motion-blur, but they assume that the scene is rigid and the motion is due to the camera shake [12]. An analysis of blind deconvolution algorithms in [13] finds that recovering blur first and then performing deblurring is a key ingredient. It also shows that the shift invariance assumption in all existing algorithms is often violated in real imagery. Our two-step approach for space-varying deblurring is somewhat inspired by these conclusions.

Alternative approaches to motion-deblurring are shown in [14], where a prototype camera moves with a parabolic motion during the exposure, and in [3, 15] where exposure is coded to facilitate the inversion of the motion-blur kernel. These techniques, however, have not been tested yet on images affected by space-varying defocus. Furthermore, as coding the exposure results in limiting the amount of incoming light, a longer exposure is needed.

2. MOTION & DEFOCUS DEBLURRING

When imaging moving objects, the image undergoes a degradation that is made of both defocus, which depends on the

aperture and the location $\boldsymbol{d}$ of the object in space, and motion blur, which depends on the object motion $\boldsymbol{m}$. Since objects in the scene may be placed at different locations and/or moving with different motions, the degradation (or blur) may be different at each pixel in the image. Hence, we use a general linear model: We indicate with $\boldsymbol{g}$ the observed blurred image and with $\boldsymbol{k}_i$ the *kernel* (or blur) of the i-th pixel in the sharp image $\boldsymbol{f}$; if both $\boldsymbol{k}_i$ and $\boldsymbol{f}$ are ordered as two column vectors, the model can be written as a product of matrices

$$\boldsymbol{g} = \boldsymbol{K}_{\boldsymbol{d},\boldsymbol{m}}\boldsymbol{f} \tag{1}$$

with the matrix $\boldsymbol{K}_{\boldsymbol{d},\boldsymbol{m}} = [\boldsymbol{k}_1\ \boldsymbol{k}_2\ \ldots\ \boldsymbol{k}_N] \in \mathbb{R}^{N\times N}$, where N is the number of pixels of the image. Each blurring kernel $\boldsymbol{k}_i$ can be rearranged as a 2D matrix $\boldsymbol{k}_i^{\square}$, which can be thought as the result of a convolution between two simpler kernels

$$\boldsymbol{k}_i^{\square} = \boldsymbol{k}_{d_i}^{\square} * \boldsymbol{k}_{m_i}^{\square} \tag{2}$$

where $\boldsymbol{k}_{d_i}^{\square}$ contains only the degradation due to defocus and $\boldsymbol{k}_{m_i}^{\square}$ corresponds to the motion blur.

The problem of deblurring a single image can be posed as

$$\tilde{\boldsymbol{f}}, \tilde{\boldsymbol{d}}, \tilde{\boldsymbol{m}} = \underset{\boldsymbol{f},\boldsymbol{d},\boldsymbol{m}}{\operatorname{argmin}} \left[||\boldsymbol{g} - \hat{\boldsymbol{g}}||^2 + E_{reg}(\boldsymbol{f}, \boldsymbol{d}, \boldsymbol{m})\right] \tag{3}$$

where we require the simulated blurred image $\hat{\boldsymbol{g}} = \boldsymbol{K}_{\boldsymbol{d},\boldsymbol{m}}\boldsymbol{f}$ to match in a least square sense the measured image $\boldsymbol{g}$, and we impose in the regularization term $E_{reg}(\boldsymbol{f}, \boldsymbol{d}, \boldsymbol{m})$ that all unknowns be piecewise constant. This minimization problem is a formidable challenge as we are given a single image $\boldsymbol{g}$ and we are looking for a 4-fold increase in number of parameters. Hence, to reduce the complexity of the problem we quantize the space of the scale and motion parameters so that only a finite set of possible values is allowed. Moreover, we break down the problem in two separate steps where we first identify the blur parameters at each pixel $\boldsymbol{K}_{\boldsymbol{d},\boldsymbol{m}}$ (see Sec. 3) and then estimate the sharp image $\boldsymbol{f}$ (see Sec. 5).

Notice that the above model works with any type of lens aperture. Therefore, we look for an aperture that allows a good reconstruction of both $\boldsymbol{f}$ and blur. We find that solving the above problem for conventional compact aperture yields poor results (see, for example, Table 1 in Sec. 6) due to a poor identifiability of the blur parameters and a stronger degradation of the image $\boldsymbol{f}$ (Sec.4). Our analysis shows that if the aperture is instead fragmented, both the parameter identification and the image degradation improve not only with still images, as already shown in [2, 7, 10], but also with motion-blur.

3. MOTION & DEPTH ESTIMATION

For now, assume that an aperture is given. As we consider a local patch of the input image and use constant velocity motion and constant defocus assumption, we can look for a blur identification method that does not require the simultaneous estimation of the sharp image $\boldsymbol{f}$. A successful method in blind deconvolution is the projection onto subspaces [16]. The key idea is that instead of solving problem (3) one minimizes

$$\tilde{\boldsymbol{d}}, \tilde{\boldsymbol{m}} = \underset{\boldsymbol{d},\boldsymbol{m}}{\operatorname{argmin}} \left[||\boldsymbol{K}_{\boldsymbol{d},\boldsymbol{m}}^{\perp}\,\boldsymbol{g}||^2 + \beta||\nabla\boldsymbol{d}|| + \gamma||\nabla\boldsymbol{m}||\right] \tag{4}$$

By solving this problem via subspace projections one can show that Gaussian priors on the unknown image $\boldsymbol{f}$ are implicitly used. This however, is not a severe limitation, as also noticed by [13]. For a given defocus scale d_i and motion m_i, the local kernel $\boldsymbol{K}_{d_i,m_i}^{\perp}$ is a collection of orthonormal vectors. The energy term corresponds to the projection of a patch of $\boldsymbol{g}$ to a subspace. The local kernel can be computed directly from the analytic forms of the blur kernels or learned from synthetic and real data as was shown in [16]. In our implementation we learn the local kernels by using real sharp images of size $\delta\times\delta$ synthetically motion-blurred and defocused. The procedure is rather straightforward, as one simply needs to: (1) generate a training set $\boldsymbol{G}_{\boldsymbol{d},\boldsymbol{m}} = [\boldsymbol{g}_1\ \boldsymbol{g}_2\ \ldots\ \boldsymbol{g}_M]^T \in \mathbb{R}^{\delta^2\times M}$ of images blurred with a specific parameter choice, (2) perform its singular value decomposition $\boldsymbol{G}_{\boldsymbol{d},\boldsymbol{m}} = USV^T$ where $U = [U_1\ U_2\ \ldots\ U_{\delta^2}] \in \mathbb{R}^{\delta^2\times\delta^2}$, V are orthonormal matrices, and S is diagonal with the singular values of $\boldsymbol{G}_{\boldsymbol{d},\boldsymbol{m}}$, (3) define $\boldsymbol{K}_{\boldsymbol{d},\boldsymbol{m}}^{\perp} = U_t$, with $t = 1, \ldots, T < \delta^2$.

These local kernels can then be used to perform the discrete minimization in eq. (4) for all possible parameters via graph cuts [17]. Notice that the second and third terms are standard total variation penalization terms involving pairwise interactions between neighboring pixels.

4. ANALYSIS OF APERTURE FRAGMENTATION

In this section we devise analysis and a procedure to determine what aperture is most suitable for the purpose of motion and defocus deblurring. Similarly to [2], we perform a frequency analyses of the aperture, in order to find a fragmentation patter that allows to preserve more frequencies than the compact aperture for different motions and defocus scales.

4.1. Combinatorics of Aperture Fragmentation

Consider partitioning a conventional aperture in a regular grid and moving the partitions within the chosen grid. Fragmentation can be seen as an "n choose m" allocation where m holes are assigned among n possible locations. The number of possible combinations can be readily obtained as $\frac{n!}{m!(n-m)!}$, which grows rather quickly as we increase the number of partitions. Hence, exhaustive search for the optimal fragmentation becomes rapidly impractical. Fortunately, diffraction poses a limit to the number of possible partitions by introducing a lower bound on the smallest diameter that we can consider before blur starts increasing rather than decreasing. By using the Rayleigh criterion we can obtain the smallest slit Δ_{min} on the lens such that a point will be reproduced clearly on the image sensor. We have that $\Delta_{min} = 1.22\frac{F\lambda}{q}$, where F denotes the focal length, λ is the light wavelength, and q is the pixel resolution. In our case we use a CANON EOS 5D (i.e., $q = 8.2\mu m$) with a $50mm$ lens, and take the worst-case light wavelength $\lambda = 750nm$ (red visible light). As we downsample the input images once, which corresponds to simulating a pixel resolution of $q = 16.4\mu m$, we have $\Delta_{min} \approx 2.79mm$.

By imposing that compact aperture and fragmented aperture must allow the same incoming light, one obtains that the

maximum number m of possible square apertures is

$$m = \left\lfloor \pi \left(\frac{F}{2F/\#} \right)^2 \frac{1}{\Delta_{min}^2} \right\rfloor \quad (5)$$

where $\lfloor a \rfloor$ denotes rounding to the largest integer not exceeding a. Conversely, given the number m of possible square apertures, one obtains that the side of each square must be

$$\Delta = \sqrt{\frac{\pi}{m}} \frac{F}{2F/\#}. \quad (6)$$

Let us illustrate these formulas with two examples. If we fix the aperture of the conventional camera to, for instance, F-number $F/9$ (i.e., an aperture with diameter $5.6mm$ for a $50mm$ focal length lens), then the area of the aperture is $24.2mm^2$. In the fragmented aperture we aim at covering the same area with apertures of at least $\Delta_{min} \times \Delta_{min} = 7.8mm^2$ area and this yields that no more than 3 square apertures are possible, and therefore a modest 84 combinations in a 3×3 grid. Vice versa, suppose that we use as F-number $F/7.1$ and we are interested in allocating 3 square apertures, then each square must have sides of $3.6mm$ (which is the dimension that we use in our experiments). Clearly, by using larger apertures, grids with more combinations are possible.

4.2. Frequency Analysis

Now that we have reduced the search space, we need to define a metric to compare different apertures and establish how much degradation they introduce. The analysis is carried out in the frequency domain of each fragmented aperture. A small patch of $\boldsymbol{f}$ (e.g., 64×64 pixels) is represented via the complex Fourier series: $f(i_1, i_2) = \sum_{n_1,n_2} \hat{f}_{n_1,n_2} e^{j(n_1 i_1 + n_2 i_2)}$, where $\hat{f}_{n_1,n_2}$ are the Fourier coefficients, and i_1 and i_2 are pixels of the image. If we assume that the signal $\boldsymbol{f}$ is corrupted by additional noise bounded in absolute value by ω, we will not be able to recover frequencies corresponding to Fourier coefficients below the noise level. Hence, we can define the number of Fourier coefficients above a given noise level as a metric for the degradation introduced by a certain aperture across several motion blur and defocus scale parameters. If $\hat{k}^{d,m}_{n_1,n_2}$ denotes the Fourier coefficients of the blurring kernel $\boldsymbol{k}_{d,m}$, we can define the degradation metric M_ω as

$$M_\omega = \sum_{d,m} \sum_{n_1,n_2} (|\hat{k}^{d,m}_{n_1,n_2} \hat{f}_{n_1,n_2}| > w). \quad (7)$$

In comparing different apertures we fix $\hat{f} = 1$ at all frequencies and look for the highest M_ω. This analysis results in three optimal apertures shown in Fig. 2, where we have examined 10 noise levels for ω between 10^{-2} and 10^{-1}. In Fig. 2 we show 1D slices corresponding to noise levels $\omega = 0.04, 0.05, 0.1$ of the normalized 2D frequency domain, to illustrate that fragmentation better distributes degradation across the frequency domain. The frequency response of a conventional aperture (a disk with diameter $6.8mm$) is shown with a dashed red plot. We now consider fragmenting the conventional disk aperture in a collection of smaller apertures, thus retaining the same overall incoming light. All apertures have the same noise levels. The corresponding response of the three best fragmented apertures for each noise level are shown in solid blue and the noise level (constant across all frequencies) is shown in solid green. We find that fragmentation results in more frequencies above the given noise level.

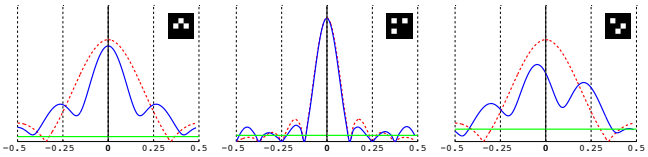

Fig. 2. The dashed red graph corresponds to different 1D slices of the frequency response of a compact aperture, while the solid blue corresponds to a fragmented aperture; the green threshold indicates the noise level. Left: Best fragmentation (evaluated over the entire 2D spectrum, not just a 1D slice) for noise levels $\omega = 0.01 - 0.04$. Middle: Best one for noise levels $\omega = 0.05 - 0.08$. Right: Best one for noise levels $\omega = 0.09 - 0.10$.

5. SPACE-VARYING DEBLURRING

Given the blur parameters provided by the procedure in Sec. 3, the space-varying deblurring task is a simpler problem. Indeed, the image formation model is linear in the unknown sharp image (although not a convolution) and efficient and stable schemes for piecewise constant regularization exist.

As a first step we compute the first-order variation of the cost functional in eq. (3) and obtain a discrete linearized version of the Euler-Lagrange equations

$$\boldsymbol{K}^T_{\tilde{\boldsymbol{d}},\tilde{\boldsymbol{m}}} (\boldsymbol{g} - \hat{\boldsymbol{g}}) + \alpha \boldsymbol{C} \cdot \boldsymbol{f} = 0, \quad (8)$$

where $\boldsymbol{C}$ is a matrix operator which performs a discretization of $\boldsymbol{f}$ based on the previous estimate, as described in [18]. As we reduced the cost functional minimization in eq. (3) to solving a linear system, standard numerical solvers can be used. Unfortunately, because the linear system involves blur, it is not diagonally dominant and fast solvers such as Gauss-Seidel or successive overrelaxation cannot be employed. We resort to conjugate gradient descent which does not have such limitations: It converges in a finite number of steps and it is fairly efficient ($\sim$ 1 minute for a 640×480 image with a Matlab implementation under a MacPro 2.6GHz quad-core CPU).

6. EXPERIMENTS

6.1. Performance

Before testing our algorithm with real images, we run a simulation to compare the performance of both defocus and motion estimation with the conventional aperture and the optimal patterns we have found in the frequency analysis (Sec. 4.2). The performance is evaluated under the same overall aperture incoming light over a set of 10 possible depth (defocus) and 64 different motions (8 directions by 8 sizes). For each level we take the same image (70x70 pixels) of random texture and we simulate both the defocus process and the motion blur using eq. (1). Then we apply the local kernel, learnt

Fig. 3. **Real Data**. Left: input image when using the fragmented aperture (pattern C); Center: estimated image when *only* defocus blur is removed; Right: sharp image when *both* defocus and motion blurs are removed.

Mask	Mean error / Accuracy		
	defocus scale	motion direction	motion size
Disk	2.53 / 13.5%	0.22 / 88.8%	1.48 / 27.2%
Pattern A	0.41 / 80.8%	0.15 / 94.0%	1.48 / 31.0%
Pattern B	0.63 / 77.3%	0.24 / 89.5%	1.53 / 30.8%
Pattern C	**0.35 / 85.0%**	**0.11 / 95.0%**	**1.39 / 33.2%**

Table 1. Aperture performance.

as described in Sec. 3, and obtain a blur estimation (defocus scale, motion direction, and motion size). The output of the algorithm is then compared with the groundtruth in order to compute the error at each pixels. Table 1 reports the mean error and the accuracy (percentage of correct estimated pixels) for each type of aperture: the first one (disk) is the aperture of a conventional camera, while masks A, B, and C are the patterns shown in Fig. 2 from top to bottom respectively. Notice that the fragmented apertures can reach an higher performance than a compact aperture.

6.2. Real Data

We capture real images with the aperture pattern C (the rightmost pattern in Fig. 2) as it gives the best performance in the synthetic analysis. The size of each of the 3 square apertures is $3.6mm$, which corresponds to a compact (circular) aperture of about $7mm$ diameter (F/7.1 with a Canon $50mm$ lens). In Fig. 3 we captured a typical scenario (the same picture with the relative compact aperture is shown in Fig. 1), where the camera brings into focus an area close to the foreground (the red bus in this scene), leaving the shops in the background out-of-focus. At the same time, the bus is moving from right to left, while the rest of the scene is still. The maximum motion blur magnitude in this dataset is $\sim$ 12 pixels. We show the reconstructed sharp texture when only defocus blur is removed and when both defocus and motion-blur are corrected.

7. CONCLUSIONS

The task of deblurring a single image degraded by space-varying motion blur and defocus is extremely ill-posed: a small variation in the data (for instance, due to noise) results in large variations of the blur parameters and the restored image. We show that by fragmenting the aperture, blur parameters and details of the original sharp image can be recovered more easily than in conventional apertures. Based on this analysis we propose an algorithm where blur parameters are first identified by using local projections onto subspaces and deblurring is then performed as a separate step given all the blur parameters. This procedure is then successfully tested on a real scenario. Although we consider a parametric representation of motion, we believe that this is the first solution for a space-varying deblurring algorithm from a single image.

8. REFERENCES

[1] Shree K. Nayar, "Computational cameras: Redefining the image," *IEEE Computer Society*, vol. 39, no. 8, pp. 30–38, Aug 2006.

[2] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 70, Aug 2007.

[3] A. Veeraraghavan, R. Raskar, A.K. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 69, Aug 2007.

[4] A. Levin, S. Hasinoff, , P. Green, F. Durand, and W. T. Freeman., "4d frequency analysis of computational cameras for depth of field extension," *ACM Trans. Graph.*, vol. 28, no. 3, Aug 2009.

[5] A. C. Likas and N. P. Galatsanos, "A variational approach for bayesian blind image deconvolution," *IEEE Trans. on Signal Processing*, vol. 52, pp. 2222–2233, 2004.

[6] R. Molina, A. K. Katsaggelos, J. Abad, and J. Mateos, "A bayesian approach to blind deconvolution based on dirichlet distributions," *ICASSP*, vol. 4, pp. 2809 – 2812, Apr 1997.

[7] Jiaya Jia, "Single image motion deblurring using transparency," *CVPR*, pp. 1–8, Jun 2007.

[8] A. Levin, "Blind motion deblurring using image statistics," *Advances in Neural Information Processing Systems*, pp. 841–848, Dec 2006.

[9] Q. Shan, W. Xiong, and J. Jia, "Rotational motion deblurring of a rigid object from a single image," *ICCV*, pp. 1–8, Oct 2007.

[10] R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis, and W.T. Freeman., "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 787–794, Aug 2006.

[11] L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," *ECCV*, Oct 2010.

[12] O. Whyte, J.Sivic, A. Zisserman, and J. Ponce, "Non-uniform deblurring for shaken images," *CVPR*, pp. 491–498, Jun 2010.

[13] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman., "Understanding and evaluating blind deconvolution algorithms," *CVPR*, pp. 1964–1971, Jun 2009.

[14] A. Levin, P. Sand, T. S. Cho, F. Durand, and W. T. Freeman, "Motion-invariant photography," *ACM Trans. Graph.*, vol. 27, no. 3, Aug 2008.

[15] A. Agrawal and R. Raskar, "Optimal single image capture for motion deblurring," *CVPR*, pp. 2560–2567, Jun 2009.

[16] P. Favaro and S. Soatto, "A geometric approach to shape from defocus," *TPAMI*, vol. 27, no. 3, pp. 406–417, Mar 2005.

[17] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?," *TPAMI*, vol. 26, no. 2, pp. 147 – 159, Jan 2004.

[18] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," *ECCV*, vol. 4, pp. 25–36, May 2004.