Camera Shake Removal With Multiple Images Via Weighted Fourier Burst Accumulation

Ms. Snehal P. Kharche
TSSM's Bhivarabai Sawant
College of Engg & Research, Narhe
Pune, India
snehal.narkhede@gmail.com

Prof. B.H. Pansambal
TSSM's Bhivarabai Sawant
College of Engg & Research, Narhe
Pune, India
bpansambal@gmail.com

Abstract— Blur introduced in an image from camera shake is mostly due to the 3D rotation of the camera. This results in a blur kernel which is non uniform throughout the image. Hence each image in the burst is blurred differently. Various experiments were done to find the deblurred image either with single image or with multiple image. In this paper we analyze multiple image approaches, which capture and combine multiple frames in order to make deblurring more robust and tractable. If the photographer takes many images known as burst, we show that a clear and sharp image can be obtained by combining these multiple images. Also for this work the blurring kernel is unknown (blind) and also it is not found. The methodology used here is Fourier Burst Accumulation which performs a weighted average in Fourier Domain where the weights depend on Fourier spectrum magnitude. In simple words the method can be generalized as Align and Average procedure. Experiments with real camera data and extensive comparisons, show that the proposed burst accumulation algorithm achieves results faster.

Keywords—*Multi-image deblurring, burst fusion, camera shake, low light photography, blind deconvolution, motion blur.*

*****

I. INTRODUCTION

The basic principle of photography is: When light is incident on an object to be captured, the reflected photons get accumulated on the sensor of the camera, an image is formed. The image will have good quality when more photons will reach the surface of the sensor within the exposure time. One can experience that if there is movement of scene or camera, the photons will be accumulated in the neighboring pixels resulting in a blurred photograph. An image is said to be blurred if one can notice the shaky effect in the image. The shaky eff ect to an image is due to motion of the subject or the imaging device. Also blurring can be caused due to incorrect focus. Motion blur is the apparent streaking of rapidly moving objects or a movie or animation. It results when the image being recorded moves rapidly in a single exposure, or due to long exposure. Under reasonable hypotheses, the mathematical model of result due to camera shake can be represented as

$$v = u * k + n \qquad (1)$$

where v is noisy blurred observation, u is the latent sharp image k is a unknown blurring kernel n is additive noise. Basically the camera can move in three directions i.e. x, y, z directions. Depending upon the position of camera, out of these one will be optical axis and other two will form planes of rotation. If the camera movement is in its optical axis with negligible in-plane rotation [1], the above model will be accurate. There are many sources which give rise to blurring kernel. For example light diff racted due to the finite aperture, or out of focus, light accumulation in the photo-sensor and relative motion between the camera and the scene during the exposure. In the situation, the scene is static and the user/camera has correctly set to focus, the blurring kernel may result from hand tremors or due to vibrations or movements of the device on which camera is mounted e.g. the camera mounted on satellite may capture blurred images. Now a day there is setting in cameras as well as mobile phones to take burst of images. Thus the camera shake originated from vibrations is essentially random. Thus we can say that the movement of the camera in each image of the burst is different and independent of each other. Thus the blur in one frame will be different from the one in another image of the burst. [2]. This is the basic concept used in this paper.

Here an algorithm is discussed in which a burst of images is aggregated. While aggregating, the contents which are less blurred of each frame are taken into consideration to build an image that is sharper and less noisy as compared to other all the images in the burst. The algorithm used here is very simple to understand and implement. Its takes a series of images as input which are already registered and computes the weighted average of the Fourier coefficients of the images from the burst. Garrel et al. [3] explored same ideas in perspective of astronomical images where he obtained a sharp clean image from a video affected by atmospheric random blur.

In phone cameras, with the availability of accurate gyroscope and accelerometers, the image registration can be obtained for free which makes the whole algorithm very efficient for on-board implementation.

The contents of the remaining paper are organized as follows.

Section II narrates the literature survey and the important differences to what we propose. Section III describes the methodology by which the images in the burst are combined to get better sharper image whereas section IV elaborates in detail the concepts in algorithm and how they are implemented. Section V shows the experimental results which are implemented as of now. There is implementation of part of the algorithm in this paper. Further work is in progress.

II. LITERATURE SURVEY

To deblur a blurred image is the most challenging work in Image Processing. Many restoration algorithms for deblurred images have been introduced since last few years. The deblurred result is obtained by deconvolution of blurred image with the blurring kernel. The blurring kernel can be a known (non-blind) or unknown (blind). Kundur and Hatzinakos [4] have reviewed over most classical methods for the same.

Here, we are considering the blurring kernel to be unknown and the survey is done regarding deblurring of image/images with blind deconvolution only.

A. Single Image Blind Deconvolution

Many blind deconvolution algorithms work with the single blurred image as an input. Using this condition, the work is the one by Fergus et al. [5]. Various methods were discussed by competitors considering many factors like assumption on blurring operators, complex optimization frameworks, estimation of blurring kernel and sharp image simultaneously. In [6] Shan, Jia and Agarwal exploited the use of sparse priors for both the sharp image and the blurring kernel. Cai et al. [7] worked on proposal of a joint optimization framework using curvelet and a framelet systems which maximizes the sparsity of the blur kernel and the sharp image simultaneously. The unnatural sparse representation of the image was discussed in [8], which mainly retains the edge information which is further used to estimate the blurring kernel. In [9], Michaeli and Irani proposed that the recurrence of small natural image patches across different scales can be used as an image prior. For sharp images, the cross-scale patch occurrence should be maximum.

Later study trend was to find the blurring operator, and then use it in non-blind deconvolution algorithm. [10] speeded up the estimation of blurring kernel by using fast image filters to detect and restore strong edges in the latent sharp image.
It was concluded that it is better to estimate the blurring kernel [11] and [12] highlighted that it is better to estimate the blurring kernel first than to find the latent image and the kernel simultaneously.

B. Multi-Image Blind Deconvolution

After the work on single image for deblurring, it was concluded that two or more input images can improve the estimation of bot blurred image and the blurring kernel. In [13] Rav-Acha and Peleg appealed that two motion blurred images are better than one if the directions of blurs are different. In [14] Yuan, Sun suggested two specific types of qualities of blurred images. Those are: one with short exposure time, noisy but sharp and other with long exposure time, blurred with less noise. These two images are complementary. The basic idea is the sharp image is used to estimate the motion kernel of the blurred image.

Now we will discuss some work close to our proposed work. In [15] Cai, Ji, Shen and Liu claimed that sparsity of image is a good measurement of clearness of the recovered latent image. Most of the multi-image algorithms introduce cross-blur penalty function between image pairs. Obviously the number of pairs will increase with the number of images in the burst.

In [16] Zang suggested, combining all the blurring kernels and to find the latent image from a single resultant kernel. This has good mathematical properties but the optimization is very slow. Recently Park and Levoy [17] trust on gyroscope which is now a days present in many pones and tablets. The gyroscope helps directly to align the input images and to estimate the blurring kernel. Then simply a multi-image non-blind deconvolution algorithm is applied.

All these papers studied proposed kernel estimation first and then applying a deconvolution algorithm. These may result into ringing effect as there might be some errors in convolution model or estimated kernel.

C. Lucky Imaging

Mostly in astronomical photography, series of thousands of short-exposure images are taken and resultant is obtained by fusing only the sharper ones [18]. In general, when the series of photographs or the video from which the frames are to be taken are too long enough, the probability of getting such images is higher. Traditionally, the selection techniques have brightness (of the brightest frame) as the important selection criteria. The balance between sharpness and the signal to noise ratio in the application is managed by the selection of number of frames.

. The number of selected frames is chosen to optimize the tradeoff between sharpness and signal to-noise ratio required in the application. Others propose to measure the local sharpness from the norm of the gradient or the image Laplacian. Joshi and Cohen [19] engineered a weighting scheme to balance noise reduction and sharpness preservation. The sharpness is measured through the intensity of the image Laplacian. They also proposed a local selectivity weight to reflect the fact that more averaging should be done in smooth regions. Haro and colleagues explored similar ideas to fuse diff erent acquisitions of painting images. The weights for combining the input images rely on a local sharpness measure based on the energy of the image gradient. The main disadvantage of these approaches is that they only rely on sharpness measures and do not profit the fact that camera shake blur can be in diff erent directions in diff erent frames. Garrel et al [20]. introduced a selection scheme for astronomic images, based on the relative strength.

In a burst mode, several photographs are captured sequentially. Due to the random nature of hand tremor, the camera shake blur is mostly independent from one frame to the other. An image consisting of white dots was photographed with a DSLR handheld camera to depict the camera motion kernels. The kernels are mainly unidimensional irregular impressions that are non uniform signal for each spatial frequency in the Fourier domain. From a series of realistic image simulations, the authors showed that this procedure produces images of higher resolution and better signal to noise ratio than traditional lucky image fusion schemes. This procedure makes a much more efficient use of the information contained in each frame.

This paper is based on similar ideas but in a diff erent scenario. The basic principle is to get resultant of all the images in the burst without explicitly estimating the blurring kernels. While doing so the proposed system takes the information that is less degraded from each image in the burst.

In all the estimation of the less degraded information is done in a very simple manner as explained further here.

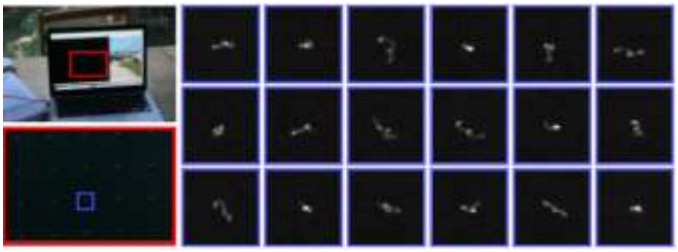

Fig1: Different kernels in the burst

III. METHODOGOY

The methodology "Fourier Burst Accumulation" is used here which is faster as compared to the previous ones. The basic steps are explained below.

A. *Concept*

The nature of shake of camera caused due to vibrations is random [2]. When the camera is handheld, the photographer will have independent movement of the hand which will causes the camera to move randomly generating blurriness in the captured image. Let us consider a photograph consists of a laptop with a black image with white dots. If the image is captured with white dot only, and the white dot represents a dirac mass then its shaken photograph will be the blurring kernel. Basically, the kernels mostly consist of unidimensional regular random shapes. This property will be the key point in our proposed approach.

Now we will discuss in short the behavior of kernel. Let F denote the Fourier Transform of image and denotes the Fourier Transform of the kernel k. Images are defined in spatial domain indexed by the 2D position x whereas their Fourier Transforms are defined in Fourier domain indexed by 2D frequency ζ. Ideally i.e. without any deviation in kernel the kernel due to camera shake is normalized such that

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

The blurring kernel is nonnegative as the combination of disjointed light is constantly nonnegative. This suggests that the motion blur does not amplify the Fourier spectrum.

Lemma: Let $k(x) \geq 0$ and $\int k(x) = 1$. Then, $k\zeta \leq 1$, $\forall \ \zeta$ (i.e. Blurring kernel does not increase the spectrum)

As the blurring kernel for each image will be different, the M images in the burst of same scene, u will be

$$v_i = u * k_i + n_i \, , for \ i = 1, \ldots M \qquad (2)$$

Thus the blurring kernel for each frame of the burst will be different and each image will be blurred differently. Thus the Fourier Transform of each frame of the burst will be different.

In the proposed system an image is to be reconstructed whose Fourier spectrum takes the largest valued Fourier magnitude in the burst. Thus the reconstructed image picks up what is less attenuated.

B. *Fourier Magnitude Weights*

Assume that p is a non-negative integer. The resultant image will be the collection of all Fourier weights of the images in the burst.

$$u_p(x) = F^{-1}\left(\sum_{i=1}^{M} w_i(\zeta).\hat{v}_i(\zeta)\right)(x) \qquad (3)$$

$$w_i(\zeta) = \frac{|\hat{v}_i(\zeta)|p}{\sum_{j=1}^{M}|\hat{v}_j(\zeta)|p} \qquad (4)$$

where $\hat{v}_i$ is the Fourier Transform of individual burst image v_i
w_i is the Fourier weight which is the frequency dependent factor.

The integer p plays an important role in these calculations. It panels the collection of the images in Fourier domain. For the value of p equal to 0, the restored image is just the arithmetic average of the burst, whereas if p approaches to infinity, each recreated frequency takes the maximum value of the frequency from the burst.

C. *Equivalent Point Spread Function*

The point spread function (PSF) defines the accuracy of an image system at a point level. It can be described as a system's impulse response. In a blur image, if we consider any point, it is moved little bit from its original position. T measure of tis point position spread is called as Point Spread Function.

In functional terms it is the spatial domain version of the transfer function of the image system. The extent of spreading (blurring) of the point in the image is a measure for the quality of an image. In single blur image, a resultant blur image is the convolution of latent clear image with blurring kernel along with some additive white noise. Here a burst of images is considered, hence the overall blurring kernel will be as in (6).

$$u_p = u * k_{FBA} + n \qquad (5)$$

where

$$k_{FBA} = F^{-1}\left(\sum_{i=1}^{M} w_i(\zeta).\hat{k}_i(\zeta)\right)(x) \qquad (6)$$

As value of the k_{FBA} kernel approaches to a Dirac function, the better the Fourier aggregation. In this case PSF will be equal to the k_{FBA}.

IV. ALGORITHM IMPLEMENTATION

A. *Input Images*

The input given to the algorithm is in the form of burst of images. Now a days there is burst mode available in various cameras. In burst mode many photographs are clicked successively with very small shutter speed. In some cameras there is multi-image burst function with which many images can be captured with a single click of shutter button. In DSLR cameras, for burst of images the shutter button is long held and photographs will be captured till the shutter button is held down. Tis is useful mostly in sport photography as the scene is continuously moving. Among the group of photographs, better imaes can be considered. But ere we are considerin burst of images of stationary scene. Hence all the images in the burst will be of same scene but differently blurred. The photograph capturing speed of camera mainly depends on processing power of the camera.

B. SIFT (Sift Invariant Feature Transform)

The object recognition is carried out with two basic features, those are detection of image and description of image. The SIFT features of the images are local features. These SIFT features are based on a particular location and are independent and invariant to image scale and rotation. Along with these properties they are very prominent and easy to extract. They are helpful for correct object identification with less mismatch probability. In a large database of local features the SIFT features are easy to match. These SIFT features are called as key points'. These key points specify the 2D location and orientation. If there is comparison of two images then for each image key points are specified and then they are matched. Each matched key point has its parameter record in the database.

Algorithm: Aggregation of blurred images

INPUT: A series of images $v_1, v_2, \ldots\ldots, v_n,$ of size *m* x *n* x *c* .

An integer value *p*

OUTPUT: An aggregated image u_p

1 w = zeros(m,n); $\hat{u}_p$= zeros(m, n, c);

2 **for** image i in {1,…..,n} **do**

Burst Registration

3 M_i = SIFT(v_i, v_1) ; M_i *set of corresponding points*

4 H_i = ORSA(M_i); H_i *dominant homography in* M_i

5 $v_i = \widetilde{v_i} \circ H_i$; *Image resampling*

Fourier Burst Accumulation

6. $\hat{v}_i$ = 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. $\hat{u}_p = \hat{u}_p + w_i \,.\, \hat{v}_i$; *Weighted Fourier accumulation*
10. w = w + w_i;

11 u_p= IFFT ($\hat{u}_p$ / w);

---

which successive photographs can be captured depends on several factors, but mainly on the processing power of the camera. Disabling certain features such as post processing which the camera applies automatically after capturing each image will usually allow a faster rate of capture. While some cheaper point and shoot cameras may have a multi-image burst function which allows them to capture a number of frames within a second with a single shutter button press, most film and digital SLR cameras will continue to actuate the shutter for as long as the button is held down, until the memory card fills or the battery runs out, although the rate of capture may significantly slow when the camera's data buff er becomes full.

B. SIFT (Scale Invariant Feature Transform)

The detection and description of local image features can help in object recognition. The SIFT features are local and based on the appearance of the object at particular interest points, and are invariant to image scale and rotation. They are also robust to changes in illumination, noise, and minor changes in viewpoint. In addition to these properties, they are highly distinctive, relatively easy to extract and allow for correct object identification with low probability of mismatch. They are relatively easy to match against a (large) database of local features. Each of the SIFT key points specifies 2D location, scale and orientation and each matched key point in the database has a record of its parameters relative to the training image in which it was found.

C. Homography (Using ORSA)

As discussed earlier, any two images ere are having some difference regarding allocation. This difference is termed as residual error. Tis residual error can be modelled with ORSA (Optimized RANSAC) algorithm using a contrario approach. This expected residual error model can be obtained while evaluating point correspondences between two images. Uniformly distributed points in the images and random coupling of those images are the bases of above discussed model. Mostly the contrario method is used for homography which eliminates the outliers are eliminated.

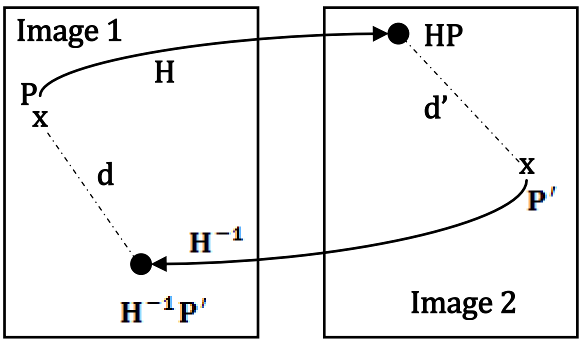

Figure 2: Residual error terms for homography: d' (transfer error in second image) or max(d,d') (symmetric transfer error)

Thus, in computer vision any two images in the same plane are related by homography. Homography can be applied while registration of two images of same scene. It can be modelled from 8 parameters under two situations wit pinhole camera, first one is wen scene is planer and other one is rotation around the optical center. The corresponding points in each image are found wit SIFT algorithm.

V. EXPERIMENTAL RESULTS

The proposed burst restoration algorithm is built on three main blocks: Burst Registration, Fourier Burst Accumulation and Noise Aware Sharpening as a post processing. In burst of images we can take many images but here the work has been started with 7 images as shown in Fig 3. Each image in the burst is differently blurred.

Burst Registration

There are several ways of registering image. In this work, we use image correspondences to estimate the dominant homography relating every image of the burst and a reference image (the first one in the burst). 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 and then filtered out through the ORSA algorithm, a variant of the so called RANSAC method. To mitigate the eff ect of the camera shake blur we only detect SIFT features having a larger scale than $\sigma min = 1.8$. Recall that as in prior art, the registration can be done with the gyroscope and accelerometer information from the camera. All the above figures are diff erently blurred.

The very first step is to extract features from the images. That was done through SIFT method for a single pair of image for now. The result of feature extraction is as below. After the feature extraction process, The ORSA algorithm (which is a type of RANSAC algorithm only) is applied to the 1st pair of images i.e. image 1 and image 2. The alignment result is shown as below.

Fig 3: Burst of Images

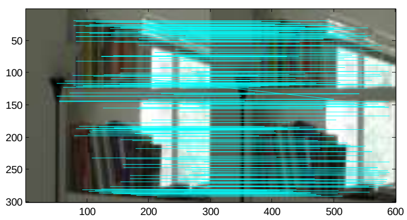

Fig 4: Result of Feature Alignment

The algorithm extracts the feature points of each pair of images. First image is considered as the reference image and it is compared with each following image. For each pair of images keypoints are found i.e. prominent points are decided and marked as feature points. The algorithm then performs alignment of extracted feature points as shown in Fig 4.

Fig 5. Align and Average Result

VI. CONCLUSIONS

We presented an algorithm which removes the camera shake blur from multiple images. The basic concept used here is though each image is differently blurred, the results are found without calculating the blurring kernels.

Main advantage of this algorithm is that it does not introduce any ringing effects which is generally found in many decovolution algorithm.

Once the images are registered, given algorithm executes the code in few seconds. Usually the multi-image non-blind deconvolution process takes hours for this data.

REFERENCES

[1] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce, “Non-uniform deblurring for shaken images,” Int. J. Comput. Vis., vol. 98, no. 2, pp. 168–186, 2012,

[2] F. Gavant, L. Alacoque, A. Dupret, and D. David, “A physiological camera shake model for image stabilization systems,” in Proc. IEEE Sensors, Oct. 2011, pp. 1461–1464.

[3] V. Garrel, O. Guyon, and P. Baudoz, “A highly efficient lucky imaging algorithm: Image synthesis based on Fourier amplitude selection,” Pub. Astron. Soc. Pacific, vol. 124, no. 918, pp. 861–867, 2012.

[4] D. Kundur and D. Hatzinakos, “Blind image deconvolution,” IEEE Signal Process. Mag., vol. 13, no. 3, pp. 43–64, May 1996.

[5] 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, 2006.

[6] Q. Shan, J. Jia, and A. Agarwala, “High-quality motion deblurring from a single image,” ACM Trans. Graph., vol. 27, no. 3, 2008, Art. ID 73.

[7] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, “Blind motion deblurring from a single image using sparse approximation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2009, pp. 104–111.

[8] L. Xu, S. Zheng, and J. Jia, “Unnatural L0 sparse representation for natural image deblurring,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2013, pp. 1107–1114.

[9] T. Michaeli and M. Irani, “Blind deblurring using internal patch recurrence,” in Proc. IEEE Eur. Conf. Comput. Vis. (ECCV), Sep. 2014, pp. 783–798.

ISSN: 2321-8169
665 - 670

[10] S. Cho and S. Lee, “Fast motion deblurring,” ACM Trans. Graph., vol. 28, no. 5, 2009, Art. ID 145.
[11] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, “Understanding and evaluating blind deconvolution algorithms,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2009, pp. 1964–1971.
[12] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, “Efficient marginal likelihood optimization in blind deconvolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2011, pp. 2657–2664.
[13] A. Rav-Acha and S. Peleg, “Two motion-blurred images are better than one,” Pattern Recognit. Lett., vol. 26, no. 3, pp. 311–317, 2005.
[14] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum, “Image deblurring with blurred/noisy image pairs,” ACM Trans. Graph., vol. 26, no. 3, 2007, Art. 1.
[15] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, “Blind motion deblurring using multiple images,” J. Comput. Phys., vol. 228, no. 14, pp. 5057–5071, 2009.
[16] H. Zhang, D. Wipf, and Y. Zhang, “Multi-image blind deblurring using a coupled adaptive sparse prior,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2013, pp. 1051–1058
[17] S. H. Park and M. Levoy, “Gyro-based multi-image deconvolution for removing handshake blur,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2014, pp. 3366–3373
[18] N. M. Law, C. D. Mackay, and J. E. Baldwin, “Lucky imaging: High angular resolution imaging in the visible from the ground,” Astron. Astrophys., vol. 446, no. 2, pp. 739–745, 2006
[19] N. Joshi and M. F. Cohen, “Seeing Mt. Rainier: Lucky imaging for multi-image denoising, sharpening, and haze removal,” in Proc. IEEE Int. Conf. Comput. Photogr. (ICCP), Mar. 2010, pp. 1–8.
[20] V. Garrel, O. Guyon, and P. Baudoz, “A highly efficient lucky imaging algorithm: Image synthesis based on Fourier amplitude selection,” Pub. Astron. Soc. Pacific, vol. 124, no. 918, pp. 861–867, 2012.