

# MULTI-FRAME IMAGE DENOISING AND STABILIZATION

*Marius Tico*

Nokia Research Center  
955 Page Mill Road, Palo Alto, CA, USA  
Email: marius.tico@nokia.com

## ABSTRACT

We propose an image restoration approach based on fusing visually similar image blocks located in a single or in multiple image frames of the same scene. The proposed approach copes with misalignment between frames, as well as with the presence of outliers represented by moving objects in the scene. The main application targeted by the proposed approach is multi-frame image stabilization, which reduces the effect of unwanted camera motion during the image integration time by fusing multiple short exposed image frames of the scene. Due to their short exposure, the individual frames are noisy, but they are less corrupted by motion blur than it would be a single long exposed frame. The proposed approach is demonstrated through a series of experiments and comparisons. The results reveal the ability of the proposed method to improve the image quality by reducing noise and simulating longer exposure times.

## 1. INTRODUCTION

Multi-frame image processing solutions can be utilized to improve the visual quality provided by small cameras embedded in mobile devices, by exploiting the variability in image degradation captured in multiple images of the same scene. One typical application of multi-frame image processing is image stabilization.

The objective of image stabilization is to reduce the visual effect of any unwanted camera and object motion during the image integration (exposure) time. Any such motion may result in an image degradation known as motion blur. The need for robust and efficient image stabilization solutions in mobile devices is driven by two main factors. On one hand, due to the miniaturization and resolution increase, the small cameras embedded in mobile devices have a limited ability to collect light requiring longer integration times. On the other hand, it is the difficulty in avoiding unwanted motion during the integration time when using high zoom, and/or small hand-held devices.

Image stabilization solutions are either aiming to correct or to prevent the motion blur degradation. Correcting the motion blur degradation from an input image requires knowledge about the motion that took place during the exposure time. In the absence of such knowledge the only chance is to adopt a blind de-convolution approach [1, 2]. Unfortunately, most of these methods rely on rather simple motion models, e.g. linear constant speed motion, and hence their potential use in consumer products is rather limited. Measurements of the camera motion during the exposure time could help in estimating the motion blur PSF and eventually to restore the original image of the scene. Such an approach have been introduced in [3], where a secondary video camera is used for estimating the motion during the exposure time of the principal camera.

Another way to estimate the PSF has been proposed in [4, 5], where a second image of the scene is taken with a short exposure. Although noisy, the secondary image is unaffected by the motion blur and it can be used as a reference for estimating the motion blur PSF which degraded the principal image.

The second category of image stabilization solutions are aiming to prevent the motion blur for happening in the first place. In this category are included all optical image stabilization (OIS) solutions adopted nowadays by many camera manufactures. These solutions are utilizing inertial sensors (gyroscopes) in order to measure the camera motion, following then to cancel the effect of this motion by moving either the image sensor [6], or some optical element [7] in the opposite direction. Due to the fact the inertial sensors are less sensitive to low frequency motions, the OIS solutions are effective only for relatively small exposure times. As the exposure time increases the mechanism may drift, producing motion blurred images. A different method, based on specially designed high-speed CMOS sensors has been proposed in [8]. The method utilizes the possibility to independently control the exposure time of each image pixel. In order to prevent motion blur the integration is stopped selectively in those pixels where motion is detected.

Multi-frame image stabilization, is another approach to prevent motion blur, and it relies on dividing a long exposure time in shorter intervals following to capture multiple short exposed image frames of the same scene. Due to their short exposure, the individual frames are corrupted by sensor noises [9], but fortunately they are less affected by motion blur. Consequently, a long exposed and motion blur free picture could be synthesized by registering and fusing the available short exposed image frames. In this paper we introduce a novel and efficient approach to multi-frame image fusion for image stabilization. The method extends our previous work [10], by allowing multiple correspondences for each image block in both spatial and temporal dimensions. One input frame is selected as the reference following then to be improved based on visual information present in itself as well as in all other image frames captured from the same scene.

## 2. THE PROPOSED ALGORITHM

We assume the following model for the  $K$  observed irradiance images:

$$g_k(\mathbf{x}) = f_k(\mathbf{x}) + n_k(\mathbf{x}), \quad (1)$$

where where  $\mathbf{x} = [x \ y]^T$  denotes the spatial position of an image pixel,  $g_k$  is the  $k$ -th observed image frame,  $n_k$  denotes a zero mean additive noise, and  $f_k$  denotes the latent image of the scene at the moment the  $k$ -th input frame was captured.

We emphasize the fact that the scene may change between the moments when different input frames are captured. Such changes could be caused by unwanted motion of the camera and/or by the motion of different objects in the scene. Consequently, the algorithm can provide a number of  $K$  different estimates of the latent scene image each of them corresponding to a different reference moment.

In the following, we assume that  $g_r$ , ( $r \in \{1, \dots, K\}$ ) is the reference observation, and hence the objective of the algorithm is to recover an estimate of the latent scene image at moment  $r$ , i.e.  $f = f_r$ .

The restoration process is carried out based on a spatiotemporal block processing. Assuming a division of  $g_r$  in non-overlapping blocks of size  $B \times B$  pixels, the restored version of each block is obtained as a weighted average of all blocks located in a specific search range, inside all observed images.

Let  $\mathbf{X}_x^B$  denote the sub-set of spatial locations included into a block of  $B \times B$  pixels centered in the pixel  $\mathbf{x}$ , i.e.:

$$\mathbf{X}_x^B = \{\mathbf{y} \in \Omega \mid [-B \ -B]^T < 2(\mathbf{y} - \mathbf{x}) \leq [B \ B]^T\}, \quad (2)$$

where the inequalities are componentwise, and  $\Omega$  stands for the image support. Also, let  $g(\mathbf{X}_x^B)$  denote the  $B^2 \times 1$  column vector comprising the values of all pixels from an image  $g$  that are located inside the block  $\mathbf{X}_x^B$ .

The restored image is calculated block by block as follows

$$\hat{f}(\mathbf{X}_x^B) = \frac{1}{Z} \sum_{k=1}^K \sum_{\mathbf{y} \in \mathbf{X}_x^S} w_k(\mathbf{x}, \mathbf{y}) g_k(\mathbf{X}_y^B), \text{ for all } \mathbf{X}_x^B, \quad (3)$$

where  $Z = \sum_{k=1}^K \sum_{\mathbf{y} \in \mathbf{X}_x^S} w_k(\mathbf{x}, \mathbf{y})$ , is a normalization value,  $\mathbf{X}_x^S$  denotes the spatial search range of size  $S \times S$  centered in  $\mathbf{x}$ , and  $w_k(\mathbf{x}, \mathbf{y})$  is a scalar weight value assigned to an input block  $\mathbf{X}_y^B$  from image  $g_k$ . The weigh values are emphasizing the input blocks that are more similar with the reference block. At the limit, using only the most similar block from each input image, we obtain the solution proposed in [10].

We use exponential like weighted functions in the form

$$w_k(\mathbf{x}, \mathbf{y}) = \exp \left[ -\frac{\alpha(\mathbf{x})}{Q^2 \sigma_{r,k}^2} \text{dist}(g_k(\mathbf{X}_y^Q), g_r(\mathbf{X}_x^Q)) \right], \quad (4)$$

where  $\sigma_{r,k}^2$  is the sum of noise variances in the reference and the  $k$ -th observed image, and  $\mathbf{X}_x^Q$  is a block of size  $Q \times Q$ , ( $Q \geq B$ ), called here *outer-block*, that includes the actual  $B \times B$  image block in the center (Fig. 1). Thus, in accordance to (4), the similarity between blocks can be calculated based on matching larger neighborhoods (i.e. outer-blocks) that include the actual image blocks in the middle.

The parameter  $\alpha(\mathbf{x})$  determines the filtering strength for the reference block  $\mathbf{X}_x^B$ . In order to avoid over-smoothing the edges,  $\alpha(\mathbf{x})$  is defined as an increasing function of the mean absolute gradient value inside the block.

The vectorial distance function "dist", used in (4) is defined as

$$\text{dist}(\mathbf{a}, \mathbf{b}) = \sum_i \rho(a_i - b_i), \quad (5)$$

where  $\mathbf{a}, \mathbf{b}$  are vectors of the same length, and  $\rho(u)$  is equal with  $u^2$  if  $|u| > \sigma_{r,k}$ , and zero otherwise.

The proposed process of spatiotemporal block based image stabilization is illustrated in Fig. 1, where image blocks are shown with gray in the center of their corresponding outer-blocks.

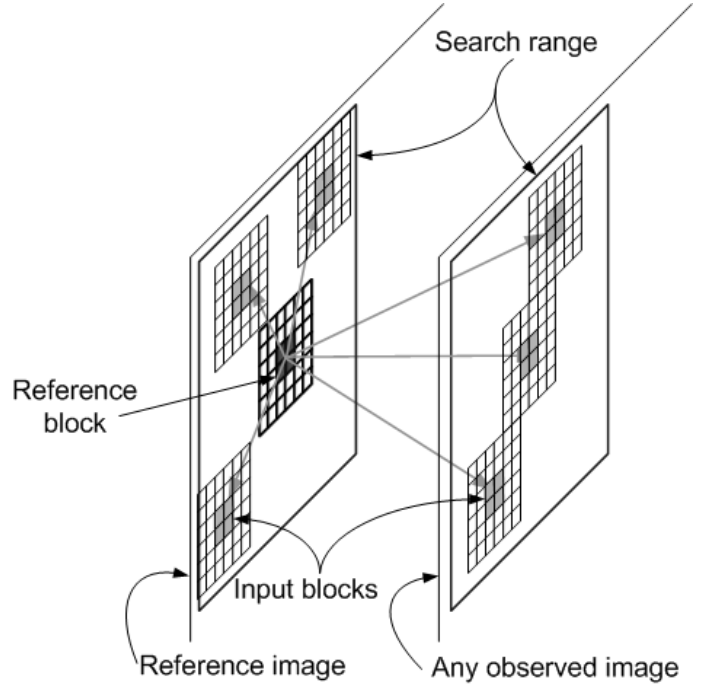


Figure 1: The proposed block based processing.

At this point we can summarize the proposed stabilization algorithm in the following steps:

1. Convert the input images to a linear color space by compensating for camera response function non-linearity. In our work we used inverse gamma correction with  $\gamma = 2.2$ .
2. Estimate the additive noise variance in each input image  $g_k$ , ( $k \in \{1, \dots, K\}$ ). Instead using a global variance value for the entire image, in our experiments we employed a linear model for the noise variance with respect to the intensity level in order to emulate the Poisson process of photon counting in each sensor pixel.
3. Select a reference image  $g_r$  either manually or automatically. Manual selection can be based on preferred scene content at the moment the image frame was captured, whereas automatic selection could be trivial (i.e. selecting the first frame), or image quality based (i.e. selecting the higher quality frame based on a quality criteria). In our work we select the reference image frame as the one that is the least affected by blur. To do this we employ a sharpness measure, that consists of the average energy of the image in the middle frequency band, calculated in the FFT domain.
4. Restore each block of the reference image in accordance to (3).
5. Convert the resulted image  $\hat{f}$ , back to the non-linear domain by gamma correction.

### 3. EXPERIMENTS

A comparison between the proposed method and different other methods applied on public domain test images is shown in Table 1. In these simulations we used a single input frame for the proposed method and the values of different parameters were as follows:  $S = 9$ , and  $B = Q = 4$ . Also,  $\alpha(\mathbf{x}) = 1$  for each reference image block where the mean absolute gradient value inside the block is below a threshold, and  $\alpha(\mathbf{x}) = 2$  otherwise. The threshold is calculated as the mean absolute gradient value in the entire reference image.

The proposed method is compared against:

- M1 the method proposed in our previous work [10].
- M2 Matlab's spatial local Wiener filtering,
- M3 hard thresholding of wavelet coefficients [11],
- M4 hard thresholding of curvelet coefficients [12].

	Noise standard deviation			
	10	15	20	25
Lenna				
Proposed	<b>34.40</b>	<b>32.58</b>	31.22	30.06
(M1)	31.87	30.02	28.77	27.69
(M2)	33.70	31.22	29.09	27.27
(M3)	30.73	28.99	27.79	26.89
(M4)	33.68	32.35	<b>31.33</b>	<b>30.42</b>
Noisy Input	28.12	24.59	22.10	20.15
Barbara				
Proposed	<b>32.45</b>	<b>30.27</b>	<b>28.84</b>	<b>27.62</b>
(M1)	30.08	27.23	25.45	24.32
(M2)	29.81	28.24	26.80	25.46
(M3)	27.28	25.01	23.68	22.90
(M4)	29.17	26.60	25.35	24.70
Noisy Input	28.14	24.62	22.10	20.19
Cameraman				
Proposed	<b>32.56</b>	<b>30.30</b>	<b>28.89</b>	<b>27.83</b>
(M1)	31.38	28.94	27.19	25.87
(M2)	30.89	29.31	27.76	26.43
(M3)	28.29	26.00	24.62	23.55
(M4)	29.49	27.59	26.40	25.60
Noisy Input	28.16	24.60	22.07	20.16

Table 1: PSNR results in decibels achieved with different approaches.

The results in Table 1 reveal that the proposed method overcomes the other approaches in almost all cases except. Two exceptions are encountered for Lenna image when the method M4 achieves slightly better PSNR than the proposed method. A visual comparison of the results for one of these two cases is shown in Fig. 4. The figure reveals that in spite of achieving better PSNR the method M4 has the tendency to introduce image artifacts visible in smooth image areas.

Two real examples using images captured with a mobile phone camera are shown in Fig. 3 and Fig. 4. In both cases the algorithm was applied onto the Bayer RAW image data before image pipeline operations. A simple linear model for the noise variance with respect to the intensity level was assumed in order to emulate the Poisson process of photon counting in each sensor pixel [9], for each color channel.

Fig. 3(a), shows an image obtained without stabilization using the mobile device set on automatic exposure. Due to unwanted camera motion the resulted image is rather blurry.

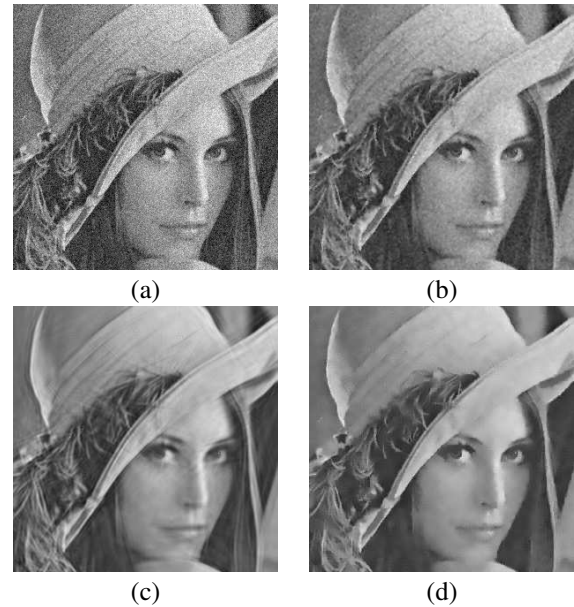


Figure 2: Visual comparison on a fragment from Lenna image: (a) noisy input (noise standard deviation 20), (b) method (M2) Wiener filter (PSNR 29.09 dB), (c) method (M4) curvelet de-noising (PSNR 31.33 dB), and (d) the proposed method (PSNR 31.22 dB).

For comparison, Fig. 3(b), shows the image obtained with our proposed stabilization algorithm by fusing several short exposed images of the same scene. An example when the proposed algorithm is applied onto a single input image is shown in Fig. 4. In this case the algorithm acts as a noise filtering method delivering the image Fig. 3(b), by reducing the noise present in the input image Fig. 3(a).

### 4. CONCLUSIONS

In this paper we introduced a novel approach to multi-frame image stabilization. In order to avoid motion blur the output image is synthesized from multiple short exposed input frames of the same scene, by identifying and fusing visually similar image blocks. The proposed method tolerates some degree of misalignment between the input frames and hence it does not require an accurate global registration of the input image frames. The method has been demonstrated through a series of numerical simulations and experiments on artificial and real image examples.

### REFERENCES

- [1] Tony F. Chan and Chiu-Kwong Wong, "Total Variation Blind Deconvolution," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 370–375, 1998.
- [2] Yu-Li You and M. Kaveh, "A regularization approach to joint blur identification and image restoration," *IEEE Trans. on Image Processing*, vol. 5, no. 3, pp. 416–428, Mar. 1996.
- [3] Moshe Ben-Ezra and Shree K. Nayar, "Motion-Based Motion Deblurring," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 689–698, 2004.



(a)



(b)

Figure 3: Real imaging examples: (a) auto-exposed image taken with a camera phone (exposure time: 1.8 sec), (b) stabilized image by fusing four frames with exposure time of 0.3 sec each.



(a)



(b)

Figure 4: Applying the proposed algorithm onto a single input image (a), delivers a noise filtered version (b), of the input image.

- [4] Marius Tico, Mejdi Trimeche, and Markku Vehvilainen, "Motion blur identification based on differently exposed images," in *Proc. of the IEEE International Conference of Image Processing (ICIP)*, Atlanta, GA, USA, Oct. 2006, pp. 2021–2024.
- [5] Marius Tico and Markku Vehvilainen, "Image stabilization based on fusing the visual information in differently exposed images," in *Proc. of the IEEE International Conference of Image Processing (ICIP)*, San Antonio, TX, USA, Sep. 2007, vol. 1, pp. 117–120.
- [6] Konika Minolta Inc., "Anti-Shake Technology," [www.konicaminolta.com/products/consumer/digital\\_camera/dimage/dimage-a2/02.html](http://www.konicaminolta.com/products/consumer/digital_camera/dimage/dimage-a2/02.html), 2003.
- [7] Canon Inc., "Shift-Method Optical Image Stabilizer," [www.canon.com/technology/dv/02.html](http://www.canon.com/technology/dv/02.html), 2006.
- [8] Xinqiao Liu and Abbas El Gamal, "Synthesis of high dynamic range motion blur free image from multiple captures," *IEEE Transaction on Circuits and Systems-I*, vol. 50, no. 4, pp. 530–539, 2003.
- [9] Junichi Nakamura, "Basics of image sensors," in *Image Sensors and Signal Processing for Digital Still Cameras*, Junichi Nakamura, Ed., pp. 53–94. CRC Press, 2006.
- [10] Marius Tico and Markku Vehvilainen, "Robust image fusion for image stabilization," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Honolulu, USA, 2007.
- [11] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
- [12] Jean-Luc Starck, Emmanuel J. Candes, and David L. Donoho, "The Curvelet Transform for Image Denoising," *IEEE Trans. on Image Processing*, vol. 11, no. 6, pp. 670–684, 2002.