Low Memory Image Reconstruction Algorithm from RAW Images

Thibaud Briand

Low Memory Image Reconstruction Algorithm from RAW Images

Thibaud Briand
LIGM (UMR CNRS 8049), Université Paris-Est, ENPC, F-77455 Marne-la-Vallée, France
CMLA (UMR CNRS 8536), Université Paris-Saclay, ENS Cachan, 94235 Cachan, France
thibaud.briand@{enpc,ens-cachan}.fr

Abstract—**In this paper, we present a fast and low memory image reconstruction algorithm from a burst of RAW images. Existing algorithms produce high-quality images but the number of input images is limited by severe computational and memory costs. Our algorithm processes the images sequentially so that the memory cost only depends on the size of the output image. Data are combined using classical kernel regression and blur is removed by applying the inverse of the corresponding asymptotic equivalent filter that we introduce. In addition, we propose an accurate and efficient registration method for mosaicked images. We verify the performance of our algorithm on synthetic and real images. For a large amount of data, the results are similar to slower memory greedy methods.**

Index Terms—**Image reconstruction, kernel regression, denoising, demosaicking, registration.**

I. Introduction

Many computational imaging methods produce high-quality images from a set of lower quality images. They compensate for the shortcomings of the imaging system by an adequate digital processing of the accumulated data. This idea has been sucessfully exploited for increasing resolution [1], [2], denoising [3], [4], reducing blur [5] or extending depth-of-field [6]. A particularly interesting problem is image reconstruction from a burst of RAW images since it solves simultaneously demosaicking and denoising, and eventually super-resolution, problems [7], [8]. The reconstructed image does not suffer from artefacts introduced by uncontrolled processing such as quantization, JPEG compression or demosaicking. However in practice the use of accumulated data is limited by huge computational and memory costs.

In this work, we focus on the low memory image reconstruction problem from a burst of RAW images. By burst we mean a set of digital images taken from the same camera, in the same state, and quasi instantaneously. The images differ because of small motions of the camera, noise and illumination variations. The scene is assumed to be static. The color intensity values are available through a color filter array (CFA), commonly the Bayer filter array [9].

Building an image from a burst is generally decomposed into two main steps: the registration, where the images are expressed in a common system of coordinates, and the combination, where the data are combined to form an image. An accurate subpixel motion estimation for CFA images is necessary to reconstruct a high-quality image. According to [7] it can be done by applying a registration method designed for grayscale images on the luminance part of the luminance and chrominance decomposition of [10]. However, the proposed Fourier-based method [11] for estimating a rigid transformation is not accurate enough. In addition, a homographic transformation model may be preferable.

In burst denoising methods [3], [4] aligned RGB images are computed from the estimated transformations by interpolation and a denoised image is obtained by averaging. In a demosaicking or super-resolution context [2], the aliased images cannot be interpolated. A color image, expressed on a regular grid, is generally computed from the irregularly sampled data obtained by gathering all the pixel values in the reference system of coordinates. A band-limited color image can be computed by applying the adaptive weights conjugate gradient Toeplitz (ACT) method [12] independently on the three channels. The method proposed by [8] is based on a maximum *a posteriori* estimation technique by minimizing a multiterm cost function, which handles the channel correlation. In [7] the luminance and chrominance parts of the reconstructed image are computed using an adaptive normalized convolution [13]. This is a robust variant of normalized convolution [14] using polynomial bases and taking into account local linear structures. It can also be described as an iterative steering kernel regression method [15], [16].

While registration is performed sequentially on the images, all the above mentioned combination methods use an iterative scheme and require the availability of all the data in memory at once. A fast and low memory method requires a simpler combination. For instance a blurry image can be computed pixel-by-pixel without any iterative scheme using classical kernel regression, which takes into account the spatial distance but not the photometric distance. As shown in [17], each pixel value is a local weighted averaging of the data, i.e., a local linear filtering. The equivalent filter depends on the data spatial repartition.

This paper proposes a fast and low memory image reconstruction algorithm from a burst of RAW images. The combination, using classical kernel regression of second order with Gaussian kernel, is split into an accumulation part and an image computation part. Using the particular structure of the linear systems involved for computing the pixel values, the input images are processed sequentially during the accumulation. An asympotic equivalent filter is introduced and its inverse is applied during a sharpening step. In this paper, the reconstructed image has the same size as the input CFA images so that the super-resolution potential of the algorithm is not discussed. The main contributions of this work are: (1) an accurate registration method for CFA images, (2) the introduction of the asymptotic equivalent filter used to remove blur and (3) a fast and low memory combination method whose performances are similar to more complex and costly methods.

The paper is organized as follows: Section II presents the computation of pixel values using classical kernel regression and introduces the corresponding asymptotic equivalent filter. Our low memory image reconstruction algorithm from a burst of RAW images is presented in Section III and is experimentally validated in Section IV. The work is concluded in Section V.

II. Classical Kernel regression

Let $\mathcal{S}$ be a set of irregularly sampled data and $\mathbf{x}^0 = (x^0, y^0)$ be a pixel position. Denote by w the two-dimensional Gaussian function of scale $\sigma_\mathrm{s} > 0$, i.e.,

$$w(\mathbf{x}) = \frac{1}{2\pi\sigma_\mathrm{s}^2} \exp\left(-\frac{\|\mathbf{x}\|^2}{2\sigma_\mathrm{s}^2}\right). \tag{1}$$

Pixel value computation. Within a local neighborhood $B(\mathbf{x}^0, r)$ of radius $r = 4\sigma_\mathrm{s}$ around $\mathbf{x}^0$, the intensity value at position $\mathbf{x} = (x + x^0, y + y^0)$ is approximated by a second order polynomial expansion

$$P(\mathbf{x}, \mathbf{v}) = v_1 + v_2 x + v_3 y + v_4 x^2 + v_5 xy + v_6 y^2. \tag{2}$$

Denote by $(I_i, \mathbf{x}_i)_{1 \leq i \leq N}$ the samples of $\mathcal{S}$ located in $B(\mathbf{x}^0, r)$ at position $\mathbf{x}_i = (x_i + x^0, y_i + y^0)$ and with intensity value I_i. Write to simplify $w_i = w(\mathbf{x}_i - \mathbf{x}^0)$. The parameter $\mathbf{v} = (v_1, \ldots, v_6)^T$ is chosen as the solution to a weighted linear regression on the intensities I_i, which amounts to minimizing the energy

$$E(\mathbf{x}^0, \mathbf{v}) = \sum_{i=1}^{N} w_i \left(I_i - P(\mathbf{x}_i, \mathbf{v})\right)^2. \tag{3}$$

Set $X_i = (1, x_i, y_i, x_i^2, x_i y_i, y_i^2)^T$ and $A_i = X_i X_i^T$. The minimization problem can be interpreted as a weighted least squares problem whose solution satisfies the 6-by-6 symmetric linear system $A\mathbf{v} = \mathbf{b}$ where

$$A = \sum_{i=1}^{N} w_i A_i \quad \text{and} \quad \mathbf{b} = \sum_{i=1}^{N} w_i I_i X_i. \tag{4}$$

In a non-degenerate case A is positive semi-definite and the system admits a unique solution. Finally, the pixel value at position $\mathbf{x}^0$ is given by $P(\mathbf{x}^0, \mathbf{v}) = v_1$.

Asymptotic equivalent filter. Consider the ideal case where the data in $\mathcal{S}$ correspond to the sampling, with a uniform spatial distribution, of a bounded function $I : \mathbb{R}^2 \to \mathbb{R}$. As N and r go to infinity, the energy in (3) is approximated, after renormalization by $\frac{1}{N}$, by

$$\mathcal{E}(\mathbf{x}^0, \mathbf{v}) = \int_{\mathbb{R}^2} w(\mathbf{x} - \mathbf{x}^0) \left(I(\mathbf{x}) - P(\mathbf{x}, \mathbf{v})\right)^2 d\mathbf{x}. \tag{5}$$

One can easily show that the above asymptotic energy has a unique minimum $\mathbf{v}$. The corresponding polynomial expansion verifies

$$P(\mathbf{x}^0, \mathbf{v}) = (k * I)(\mathbf{x}^0) \tag{6}$$

where k is called the asymptotic equivalent filter and is defined by

$$k(\mathbf{x}) = w(\mathbf{x}) \left(2 - \frac{\|\mathbf{x}\|^2}{2\sigma_\mathrm{s}^2}\right). \tag{7}$$

This kernel does not depend on $\mathbf{x}^0$ and its Fourier transform is given by

$$\mathcal{F}(k)(\mathbf{x}) = \left(1 + \frac{\sigma_\mathrm{s}^2}{2}\|\mathbf{x}\|^2\right) \exp\left(-\sigma_\mathrm{s}^2 \frac{\|\mathbf{x}\|^2}{2}\right). \tag{8}$$

To summarize, an image computed by classical kernel regression can be asymptotically interpreted as the convolution between the ideal unknown target image and k. The image may seem blurry since k attenuates the high-frequencies. This justifies the application of an enhancement filter reverting the blur.

III. Algorithm

From a burst of N_im RAW images of size $W \times H$, our image reconstruction algorithm builds an RGB image of size $W \times H$. An additional reference RAW image is used during the preprocessing and registration steps. The main steps are summarized in Fig. 1(b).

Preprocessing. A variance stabilizing transformation (VST) is applied to the RAW images to approximate a homoscedastic noise. For the channel c, using the classical assumption that the variance V_c of the noise is an affine function of the intensity $V_c(u) = a_c u + b_c$, the VST is given by $f_c(u) = \sqrt{a_c u + b_c}$. The noise curve of the reference RAW image is estimated using the Ponomarenko et al. method [18], [19] and the parameters a_c and b_c are computed by linear least squares.

Then, a multiplicative mean equalization of the reference image's channels is done and its maximal value is arbitrarily set to 255. Finally, multiplicative mean equalizations between all the images and the reference are performed.

Registration. First, the burst and reference CFA images are converted to grayscale images by applying the 11×11 filter of [10, Fig. 4(b)]. Then, homographic transformations linking the images to the reference are estimated by applying the inverse compositional algorithm [20] to the corresponding filtered images. More precisely, the registration is done using a robust error function and a multiscale coarse-to-fine approach as described in [21].

Accumulation. Computing pixel values by classical kernel regression is equivalent to solving a symmetric linear system $A\mathbf{v} = \mathbf{b}$. Thanks to the redundancies in A the system is determined by only 21 coefficients (instead of 42). The accumulation step consists in computing and storing the 21 system coefficients for each output pixel. It takes as input the images and their corresponding homographic transformations.

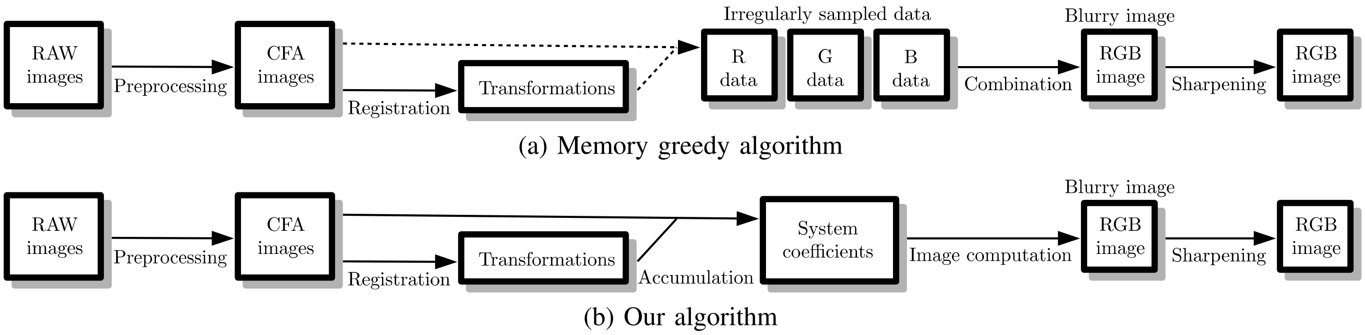


Fig. 1. Main steps of our image formation method. (a) and (b) are theoretically equivalent.

Thanks to the summative expressions in (4) the contribution of a sample to the system coefficients can be computed by accumulation. The system coefficients are initialized to 0. Then, for each of the N_{im} CFA images the contributions of its $W \times H$ samples are added to the system coefficients. Note that a sample only contributes to pixels of the same channel and at a distance lower than r.

In the following the value $\sigma_{\text{s}} = 1/\sqrt{2}$ is used. It insures a non-negligible contribution to a sample at distance $2\sigma_{\text{s}} = \sqrt{2}$, which is approximately the distance separating red (or blue) samples of a CFA image. Each sample contributes to at most 20 surrounding pixels, which represents 420 updates.

Image computation and sharpening. For each pixel of the color regular grid of size $W \times H$, the symmetric linear system, corresponding to the computed system coefficients, is solved using a Cholesky decomposition. The blur of the resulting RGB image is corrected in the sharpening step. It consists in applying the inverse of the asymptotic equivalent filter k to the RGB image. This is done using the inverse of (8) and the discrete cosine transform (DCT) convolution method.

Low memory requirement. Theoretically the accumulation and image computation steps are equivalent to a single combination step taking as input the three sets, one per channel, of irregularly sampled data. The structure of the resulting algorithm is shown in Fig. 1(a). The storage of the data makes this algorithm memory greedy. It requires to store three floating point numbers per sample with a total amount of $3N_{\text{im}}WH$.

On the contrary the memory cost of the proposed algorithm is constant with the number of images N_{im} since the images and their samples are processed sequentially. The main cost comes from the storage of the system coefficients, which only represents $63WH$ floating point numbers.

IV. Experiments

Our image reconstruction algorithm was compared to an algorithm using the memory greedy structure of Fig. 1(a) where the combination is done using the ACT method [12]. It uses the same preprocessing and registration but requires no sharpening.

When the reference target image is known we define $\mathcal{E}_{\text{ref}}$ as the root mean square error (RMSE) of the reconstructed image. We define $\mathcal{E}_{\text{set}}$ as the root mean square difference (RMSD) between reconstructed images from two separated subsets of the burst. In the case of an ideal denoising from n images we would have $\mathcal{E}_{\text{ref}} = \sqrt{\frac{10}{3}}\frac{\sigma}{\sqrt{n}}$ and $\mathcal{E}_{\text{set}} = \sqrt{2}\mathcal{E}_{\text{ref}}$. The factor $\sqrt{\frac{10}{3}}$ is due to the different number of samples available in the channels (two times more in the green channel). Note that before any image comparison a band of 20 boundary pixels is discarded and a mean equalization is performed.

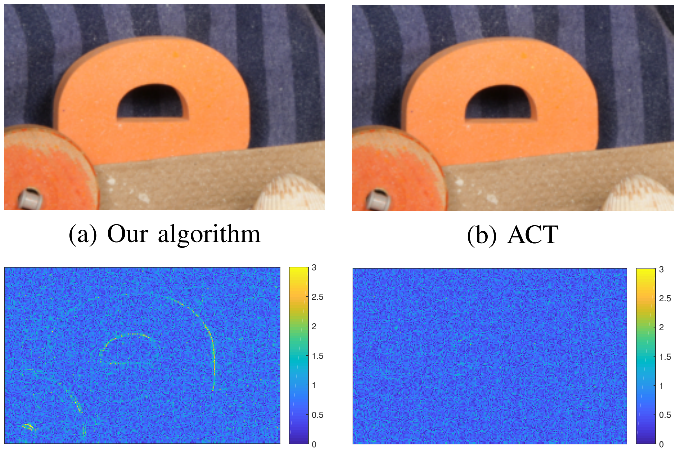
The first goal of the experiments is to show that we are able to provide similar results as the ACT with less computation time. Secondly, we verify that the decay of $\mathcal{E}_{\text{ref}}$ and $\mathcal{E}_{\text{set}}$ empirically follows the ideal denoising error decay.

In the figures, we display details on a small relevant zone. For the difference of color images we display the root mean square (RMS) over the channels. The experiments were made using C-language implementations and an Intel(R) Core(TM) i7-7820HQ CPU (2.90GHz).

On synthetic data. We generated a burst of $N_{\text{im}} = 200$ CFA images from the *Rubberwhale* image [22], which is a color image of size 584×388. The images were linked by homographies obtained by randomly moving the four corners of the input image. Each displacement was uniformly distributed in $[-3, 3]^2$. After resampling with a high-order interpolation method, the color filter was applied and Gaussian white noise of standard deviation $\sigma = 5$ was added. A noisy CFA reference image was also generated.

The registration step was performed in 30s, i.e, 150ms per image. The end-point error was on average 0.00909 pixel. With such a subpixellic precision, the results with or without estimating the transformations are similar. Therefore, we present the results using the estimated transformations.

The reconstructed images using our algorithm and ACT are displayed in Fig. 2(a)-(b). It is hardly possible to detect differences to the naked eye. As shown in Fig. 2(c)-(d), the residuals using both methods are mainly composed of noise. Our result has higher errors at the discontinuities. This effect is less important with ACT since it is Fourier based. As shown in Fig. 2(e), both reconstruction errors are close to the ideal



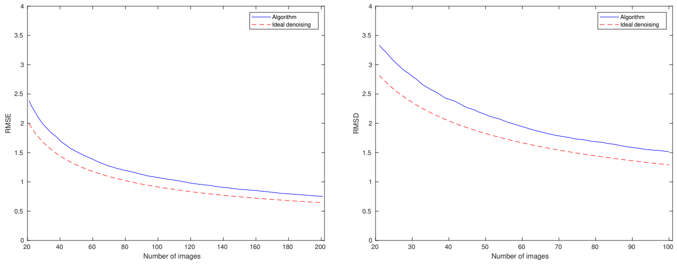
(a) Our algorithm (b) ACT

(c) Residual of (a) (d) Residual of (b)

	Denoising	ACT	Our algorithm
$\mathcal{E}_{\text{ref}}$	0.645	0.660	0.760
Time (s)	ND	560	65

(e) $\mathcal{E}_{\text{ref}}$ and computation time

Fig. 2. Image reconstruction on synthetic data. The computation times in (e) do not take into account the registration.



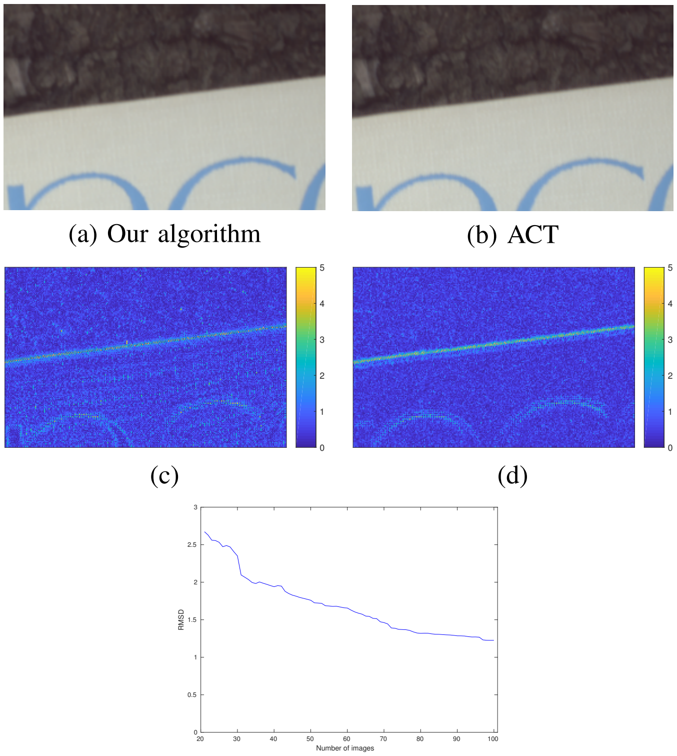
(a) $\mathcal{E}_{\text{ref}}$ (b) $\mathcal{E}_{\text{set}}$

Fig. 3. Error evolution with the number of input images using our algorithm on synthetic data.

denoising error. The error $\mathcal{E}_{\text{ref}}$ using our algorithm is slightly higher than using ACT but the RGB image was computed 8.5 times faster using about $\frac{N_{\text{im}}}{21} \simeq 10$ times less memory. The evolutions of $\mathcal{E}_{\text{ref}}$ and $\mathcal{E}_{\text{set}}$ with the number of input images are shown in Fig. 3. In both cases the curve shape is similar to the ideal denoising one.

On real data. We built images from a burst of $N_{\text{im}} = 200$ RAW images and an additional reference RAW image. The images were acquired using an Olympus E-M5 Mark II camera set in sequential mode and then cropped to have a size of 1024×1024.

The registration was performed in 167s, i.e., 835ms per image. The combination and sharpening were done in 367s using our algorithm and in 2705s using ACT. The constructed images are shown in Fig. 4(a)-(b) and compared in Fig. 4(c). The result using ACT is slightly better at the discontinuities but our result is computed 7 times faster and required about $\frac{N_{\text{im}}}{21} \simeq 10$ times less memory.



(a) Our algorithm (b) ACT

(c) (d)

(e) $\mathcal{E}_{\text{set}}$

Fig. 4. Image reconstruction on real data. (c) is the difference between (a) and (b), and (d) is the difference between reconstructed images built using our algorithm on two separated sets of 100 images.

The evolution of $\mathcal{E}_{\text{set}}$ with the number of images is shown in Fig 4(e). It confirms that the error decay is close to the ideal denoising one. Indeed, the error between 20 and 100 images is divided by a factor 2.18 for an ideal denoising decay factor of $\sqrt{5} \simeq 2.23$. The difference image corresponding to the fusion of 100 images is shown in Fig 4(d).

V. Conclusion

This work proposes a fast and low memory image reconstruction algorithm from a burst of RAW images. Thanks to a simple combination method and a sequential processing of the input images, the memory cost only depends on the size of the output image. The registration method used for CFA images is proven experimentally to be efficient and accurate. Classical kernel regression has a particular structure that allows for the accumulation of the data and admits an asymptotic equivalent filter, which is successfully inverted to remove the blur. For a large amount of data our algorithm provides similar results as slower and memory greedy methods. The residual noise decreases as expected.

The super-resolution potential will be discussed in a coming paper. This efficient algorithm opens the way to real time image formation.

REFERENCES

[1] R. Tsai, "Multiframe image restoration and registration," *Advance Computer Visual and Image Processing*, vol. 1, pp. 317–339, 1984.

[2] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 47–57, 2004.

[3] A. Buades, Y. Lou, J. Morel, and Z. Tang, "A note on multi-image denoising," in *Int. Workshop on Local and Non-Local Approximation in Image Processing*, 2009, pp. 1–15.

[4] Z. Liu, L. Yuan, X. Tang, M. Uyttendaele, and J. Sun, "Fast burst images denoising," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 6, p. 232, 2014.

[5] M. Delbracio and G. Sapiro, "Removing camera shake via weighted Fourier burst accumulation," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3293–3307, 2015.

[6] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," in *Computer Graphics Forum*, vol. 28, no. 1. Wiley Online Library, 2009, pp. 161–171.

[7] P. Vandewalle, K. Krichane, D. Alleysson, and S. Süsstrunk, "Joint demosaicing and super-resolution imaging from a set of unregistered aliased images," in *Electronic Imaging 2007*. International Society for Optics and Photonics, 2007, pp. 65 020A–65 020A.

[8] S. Farsiu, M. Elad, and P. Milanfar, "Multiframe demosaicing and super-resolution of color images," *IEEE transactions on image processing*, vol. 15, no. 1, pp. 141–159, 2006.

[9] B. Bayer, "Color imaging array," Jul. 20 1976, US Patent 3,971,065.

[10] D. Alleysson, S. Susstrunk, and J. Herault, "Linear demosaicing inspired by the human visual system," *IEEE Transactions on Image Processing*, vol. 14, no. 4, pp. 439–449, April 2005.

[11] P. Vandewalle, S. Süsstrunk, and M. Vetterli, "A frequency domain approach to registration of aliased images with application to super-resolution," *EURASIP journal on advances in signal processing*, vol. 2006, no. 1, p. 071459, 2006.

[12] H. Feichtinger, K. Gröchenig, and T. Strohmer, "Efficient numerical methods in non-uniform sampling theory," *Numerische Mathematik*, vol. 69, no. 4, pp. 423–440, 1995.

[13] T. Pham, L. J. Van Vliet, and K. Schutte, "Robust fusion of irregularly sampled data using adaptive normalized convolution," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, pp. 1–12, 2006.

[14] H. Knutsson and C. Westin, "Normalized and differential convolution," in *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on*. IEEE, 1993, pp. 515–523.

[15] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel Regression for Image Processing and Reconstruction," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 349–366, Feb 2007.

[16] P. Milanfar, "A Tour of Modern Image Filtering: New Insights and Methods, Both Practical and Theoretical," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 106–128, Jan 2013.

[17] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics New York, 2001, vol. 1.

[18] M. Colom and A. Buades, "Analysis and Extension of the Ponomarenko et al. Method, Estimating a Noise Curve from a Single Image," *Image Processing On Line*, vol. 3, pp. 173–197, 2013.

[19] N. Ponomarenko, V. Lukin, M. Zriakhov, A. Kaarna, and J. Astola, "An automatic approach to lossy compression of AVIRIS images," in *2007 IEEE International Geoscience and Remote Sensing Symposium*, July 2007, pp. 472–475.

[20] S. Baker, R. Gross, I. Matthews, and T. Ishikawa, "Lucas-Kanade 20 Years On: A Unifying Framework: Part 2," Pittsburgh, PA, Tech. Rep. CMU-RI-TR-03-01, February 2003.

[21] J. Sánchez, "The Inverse Compositional Algorithm for Parametric Registration," *Image Processing On Line*, vol. 6, pp. 212–232, 2016.

[22] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A Database and Evaluation Methodology for Optical Flow," *International Journal of Computer Vision*, vol. 92, no. 1, pp. 1–31, Mar 2011.