

Joint Burst Denoising and Demosaicking via Regularization and an Efficient Alignment

R. Azizi¹ and AM. Latif^{2*}

1. Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
2. Computer Engineering Department, Engineering Faculty, Yazd University, Yazd, Iran

Received 19 December 2019; Revised 23 April 2020; Accepted 01 August 2020

* Corresponding author: alatif@yazd.ac.ir (AM. Latif)

Abstract

In this work, we show that an image reconstruction from a burst of individually demosaicked RAW captures propagates demosaicking artifacts throughout the image processing pipeline. Hence, we propose a joint regularization scheme for burst denoising and demosaicking. We model the burst alignment functions and the color filter array sampling functions into one linear operator. Then, we formulate the individual burst reconstruction and the demosaicking problems into a three-color-channel optimization problem. We introduce a cross-channel prior to the solution of this optimization problem and develop a numerical solver via the alternating direction method of multipliers. Moreover, our proposed method avoids the complexity of alignment estimation as a pre-processing step for burst reconstruction. It relies on a phase correlation approach in the Fourier's domain to efficiently find the relative translation, rotation, and scale among the burst captures and to perform warping accordingly. As a result of these steps, the proposed joint burst denoising and demosaicking solution improves the quality of the reconstructed images by a considerable margin compared to the existing image model-based methods.

Keywords: *Burst Imaging, Image Demosaicking, Alternating Direction Method of Multipliers, Efficient Alignment.*

1. Introduction

One of the most challenging problems in manufacturing an imaging pipeline is noise reduction. The most intuitive solution is, of course, to collect as much light as possible when taking a photograph. This can be partially addressed in larger standalone cameras, e.g. digital single-lens reflex (DSLR) cameras through the use of a large aperture lens, sensors with large photosites, and high-quality analog to the digital (A/D) conversion. However, modern mobile cameras are very limited with regards to such hardware elements. Collecting more light is also possible with the use of a longer exposure so that each photosite integrates light over a longer period of time. This technique is also limited since it requires placing the camera on a tripod or taking advantage of advanced hardware stabilizers to avoid motion and consequently motion blur in the final image.

An efficient yet effective computational approach towards solving the issues above is burst photography that allows collecting light over a longer period of time by capturing a sequence of photos. The process is performed assuming that a more accurate estimate of the true underlying signal is obtained by averaging independent noisy samples of the signal. However, due to the inevitable camera and/or scene motion in burst captures taken with hand-held cameras, simple temporal averaging yields poor results. Many techniques attempt to align the burst images before averaging or include some notion of translation-invariance

within the denoising operator itself [1]. Many of the most successful single image denoising techniques [2, 3] rely on combining multiple aligned image patches via taking into account the self-similarity of a single image rather than using just a few neighbouring pixels via filtering approaches [4]. They group similar patches within the image and jointly filter them under the assumption that multiple noisy observations can be averaged to better estimate the true underlying signal. Such ideas have been retargeted towards the task of burst denoising of images captured using mobile phones [1, 5]. These methods align similar image patches within a few pixels across the burst set and then perform a robust averaging.

Achieving a high-quality image by combining multiple image formation steps with a single linear operator can be generalized to burst image denoising [6, 7]. Such methods use modern optimization techniques to solve the associated inverse problem. However, they require a costly alignment as part of the forward model. The proposed technique in this paper lies on the line of such work that combines burst image reconstruction and the transformation of RAW captures into high-resolution color images, i.e. demosaicking, into a single linear operation. Burst reconstruction and demosaicking as individual stages are mathematically ill-posed and rely heavily on heuristics and prior information to produce good results. Performing one after the other requires the latter to treat the results of these heuristics as a ground truth input. This will cause aggregating the mistakes through the pipeline. We propose to

perform burst reconstruction jointly with the demosaicking in a single process as an attempt to avoid propagating the demosaicking artifacts throughout the burst denoising process. Fig. 1 highlights the issue with performing demosaicking followed by burst reconstruction as two individual steps.

Many single image demosaicking approaches assume noise-free color filter array (CFA) images [8, 9, 10, 11, 12]. Such sophisticated non-linear methods generally fail due to the inevitable presence of noise in real captures. Hence, denoising is required to be incorporated in the process of demosaicking as carried out in the methods proposed in [13, 14, 15, 16, 17, 18], for instance. A comprehensive review of the classic demosaicking approaches is presented in [19] and [20]. Employing regularization to address the ill-posed problem of image reconstruction has become an effective solution. Exploiting some prior knowledge about the image often derived from some statistics about natural images is quite fulfilling in producing high-quality images. Employing the total-variation (TV) regularization scheme for each color channel [21], or using the TV term for color differences [22] rather than inter-channel data in order to better fix the demosaicking color artifacts are examples of such regularization-based techniques. Menon et al. [23] have proposed to perform regularization via a quadratic penalty term over the difference of gradients of color channels to account for cross-channel dependencies. Tan et al. [24] have proposed a combination of TV, a cross-channel prior, and the block-matching and 3D filtering (BM3D) denoiser, to perform the reconstruction. Although this method outperforms single-image demosaicking methods, the use of BM3D as the prior along with the iterative solver make it computationally expensive. A regularization-based approach to solve a joint demosaicking and super-resolution problem has been proposed by Farsiu et al. [25]. The proposed method first generates a high-resolution luminance component via an edge preserving bilateral regularization strategy. Then the Tikhonov regularizer is employed to reconstruct the chrominance components of the image.

The recent data-driven approaches have significantly outperformed the traditional model-based methods [18, 26, 27, 28]. However, their results have been reported with neural networks (NNs) trained on special datasets tailored to perform demosaicking and burst imaging. Such NNs are not lightweight and required a lot of training data. Moreover, they are not yet deployable to the current mobile hardware systems. Hence, we deviate from such methods in this work and aim at an efficient approach towards this imaging problem.

In this paper, we propose a regularization scheme for a burst denoising that incorporates demosaicking in the process. We first, carefully combine the burst alignment functions and the CFA sampling functions into a single linear operator. Then, we model the individual burst reconstruction and the demosaicking problems into a single imaging inverse problem. This is modeled as a 3-channel (RGB) optimization problem that we approach via least-squares regularization. We extend our recently proposed cross-channel prior [29] as a regularizer to this least-squares optimization problem and develop a numerical solution via Alternating Direction Method of Multipliers (ADMM).

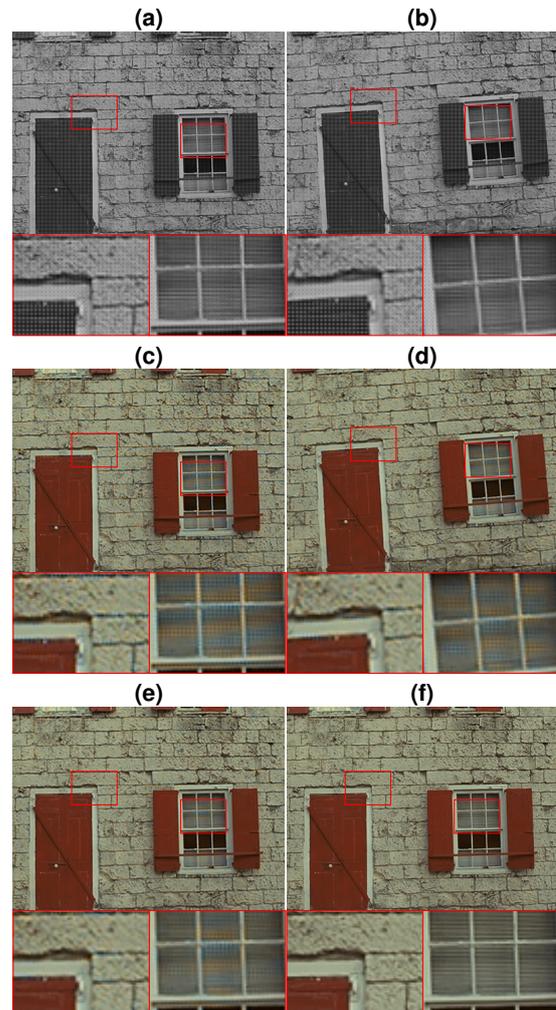


Figure 1 . Two simulated RAW images of a synthetic burst set (a) and (b) were demosaicked independently, (c) and (d), then aligned and averaged (e). The demosaicking zippering artifacts that appear in the individually demosaicked images (c) and (d) still exist in the aligned and averaged result (e). Demosaicking artifacts do not happen randomly and are, in fact, visible around the same edge in different captures of a same scenes. Therefore, they are not cancelled out through averaging. The ground truth high-resolution image is shown in (f).

In this work, we also revisit the classic Fourier-based phase correlation approaches in order to efficiently find the translation between burst images. We extend phase correlation to the cases where the 2D rotation and scaling occur. This results in an accurate yet efficient alignment method. We propose to employ this alignment method on the sub-sampled mosaicked images rather than utilizing the complicated over-kill alignment methods used in the state-of-the-art burst reconstruction methods. As a result of these contributions, the proposed method outperforms the existing model-based burst image reconstruction techniques.

The rest of this paper is organized as what follows. Section 2 presents the problem statement and the proposed solution to the burst reconstruction problem using ADMM. The image alignment technique is also discussed in this

section. The experimental results are reported in Section 3. Finally, the paper is concluded in Section 4.

2. Problem Formulation of Burst Demosaicking

Given a camera sensor whose resolution is $N \times M$, and denoting Bayer's sampling by $s_r(\cdot)$, $s_g(\cdot)$, and $s_b(\cdot)$, the color image acquisition through a CFA can be modeled as

$$I_{cfa} = s_r(R) + s_g(G) + s_b(B) + \eta, \quad (1)$$

where I_{cfa} denotes a Bayer's mosaicked image sampled from the discretized $N \times M$ color components R, G, B . In (1), zero-mean Gaussian additive noise is denoted by η . For each color channel, the corresponding sampling operator acts as a mask on the pixel coordinates that correspond to the other two channels taking into account the CFA's pattern. The color sampling functions for a GRBG patterned CFA can be defined as:

$$s_r(I) = \frac{I(x,y)}{4} (1 + (-1)^x)(1 - (-1)^y), \quad (2)$$

$$s_g(I) = \frac{I(x,y)}{2} (1 + (-1)^{x+y}), \quad (3)$$

$$s_b(I) = \frac{I(x,y)}{4} (1 - (-1)^x)(1 + (-1)^y). \quad (4)$$

where (x, y) denote the pixel coordinates of an $N \times M$ image I , $\forall x = 1 \dots M, y = 1 \dots N$.

In the process of burst imaging, a set of L captures of the scene is acquired. This set includes L number of CFA images as $\{I_{cfa_1}, \dots, I_{cfa_L}\}$. Due to the inevitable hand-shake, these captures are not geometrically aligned. Therefore, in order to relate all these captures with regards to a single scene, i.e., a single high-resolution R, G, B capture, we need to account for the geometry mapping of the captures as:

$$I_{cfa_l} = s_r(a_l(R)) + s_g(a_l(G)) + s_b(a_l(B)) + \eta_l \quad (5)$$

where $a_l(\cdot)$ denotes the warping function that maps the single high-resolution RGB capture to the CFA image I_{cfa_l} . Reconstructing a high-resolution RGB image from the observations I_{cfa_l} given the linear operations a_l, s_r, s_g , and s_b is an ill-posed problem due to the presence of the unknown noise η_l . This problem can be formulated as a least-squares regression problem [30], in a same way done for other inverse imaging problems [31, 32, 33].

In practice, one of the CFA images is typically considered as the reference image and the warping functions are estimated based on that image through homography estimation. It should be noted that each burst capture has a unique alignment with respect to the reference capture that is required to be estimated independently from the other captures. For example, if the first capture of the set is taken as the reference, by applying the warping functions on the red component of other images in the set, red components of the entire burst set relate to the reference as $R_2 = a_2(R_1), R_3 = a_3(R_1), \dots, R_L = a_L(R_1)$. High-resolution color channels, R, G , and B , can be denoted in vector form by $\mathbf{c}_1, \mathbf{c}_2$, and $\mathbf{c}_3 \in \mathbb{R}^{NM \times 1}$, respectively. For more clarity, the sampling functions are modified so that they down-sample the corresponding color channels

into the low-resolution space. They are also required to separate color channels from the underlying CFA pattern. Therefore, the sampling operators in matrix form are defined as $\mathbf{S}_1, \mathbf{S}_3 \in \mathbb{R}^{\frac{NM}{4} \times NM}$, $\mathbf{S}_2 \in \mathbb{R}^{\frac{NM}{2} \times NM}$, corresponding to s_r, s_b, s_g , respectively. The human visual system (HSV) is more responsive to the green spectrum. Therefore, the amount of green light passed through the CFA is twice more than that of the other two color components. Denoting the warping functions $a_l(\cdot)$ as $\mathbf{A}^l \in \mathbb{R}^{\frac{NM}{4} \times \frac{NM}{4}}$ ($l = 1 \dots L$) in the matrix form, we propose to minimize the following convex cost function in order to reconstruct high-resolution color channels:

$$\begin{aligned} & \text{minimize}_{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3} \frac{1}{2} \left\| \begin{bmatrix} \hat{\mathbf{c}}_1^1 \\ \hat{\mathbf{c}}_2^1 \\ \hat{\mathbf{c}}_3^1 \\ \vdots \\ \hat{\mathbf{c}}_1^L \\ \hat{\mathbf{c}}_2^L \\ \hat{\mathbf{c}}_3^L \end{bmatrix} - \underbrace{\begin{bmatrix} \mathbf{A}^1 & 0 & 0 & 0 \\ 0 & \mathbf{A}^1 & 0 & 0 \\ 0 & 0 & \mathbf{A}^1 & 0 \\ 0 & 0 & 0 & \mathbf{A}^1 \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{A}^L & 0 & 0 & 0 \\ 0 & \mathbf{A}^L & 0 & 0 \\ 0 & 0 & \mathbf{A}^L & 0 \\ 0 & 0 & 0 & \mathbf{A}^L \end{bmatrix}}_{\text{data-fidelity}} \begin{bmatrix} \mathbf{S}_1 & 0 & 0 \\ 0 & \mathbf{S}_2 & 0 \\ 0 & 0 & \mathbf{S}_3 \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{c}_3 \end{bmatrix} \right\|_2^2 \\ & + \lambda_1 \sum_{j=1}^2 \left\| \begin{bmatrix} \partial_j \mathbf{c}_1 \\ \partial_j \mathbf{c}_2 \\ \partial_j \mathbf{c}_3 \end{bmatrix} \right\|_1 \\ & + \lambda_2 \underbrace{\sum_{j=1}^2 \left(\|\partial_j \mathbf{c}_1 - \partial_j \mathbf{c}_2\|_1 + \|\partial_j \mathbf{c}_1 - \partial_j \mathbf{c}_3\|_1 + \|\partial_j \mathbf{c}_2 - \partial_j \mathbf{c}_3\|_1 \right)}_{\text{cross-channel prior}}, \end{aligned} \quad (6)$$

where ∂_1 and $\partial_2 \in \mathbb{R}^{NM \times NM}$ denote the horizontal $[-1 \ 1]$ and vertical $[-1 \ 1]^T$ derivative operators in matrix form, respectively. The ℓ_p norm function is represented by $\|\cdot\|_p$ and the matrix transpose operation is denoted by T . Moreover, $\hat{\mathbf{c}}_n^l$ corresponds to the n -th color channel ($n = 1, 2, 3$) of the l -th CFA image, i.e. $\mathbf{S}_n I_{cfa_l}$ denoted in vector form $\hat{\mathbf{c}}_1^l, \hat{\mathbf{c}}_3^l \in \mathbb{R}^{\frac{NM}{4} \times 1}$, $\hat{\mathbf{c}}_2^l \in \mathbb{R}^{\frac{NM}{2} \times 1}$. In the block-matrices in (6), 0 at every row denotes a null matrix whose size is equivalent to that of the non-zero sub-block on the same row of the matrix. For example, 0 on the rows that include \mathbf{A}^1 is an $\frac{NM}{4} \times \frac{NM}{4}$ null matrix. Similarly, 0 on the row that includes \mathbf{S}_1 is an $\frac{NM}{4} \times NM$ null matrix.

Image reconstruction is, in fact, modeled as an optimization problem in (6). The objective function is regularized by the traditional TV and a cross-channel prior. Sparsity per reconstructed channel is enforced by the first regularization term, while the second term signifies frequency components shared between color channels. In fact, in camera systems, due to the unavoidable lens dispersion, known as chromatic aberration, and/or due to the structure of the color filter arrays, there is some amount of apparent color fringe in the images that is more visible near edges, as shown in Fig. 2. Hence, the preserved details in one channel are utilized in the reconstruction of another channel in order to correct for the spatial displacements in the color channels.

We propose to solve this problem by alternating among color channels iteratively. Thus, at every iteration $i + 1$,

we find $\mathbf{c}_n^{(i+1)}$ via solving the following problem:

$$\mathbf{c}_n^{(i+1)} = \arg \min_{\mathbf{c}_n} \frac{1}{2} \left\| \begin{bmatrix} \hat{\mathbf{c}}_n^1 \\ \vdots \\ \hat{\mathbf{c}}_n^L \end{bmatrix} - \begin{bmatrix} \mathbf{A}^1 \\ \vdots \\ \mathbf{A}^L \end{bmatrix} \mathbf{S}_n \mathbf{c}_n \right\|_2^2 + \lambda_1 \sum_{j=1}^2 \|\partial_j \mathbf{c}_n\|_1 + \lambda_2 \sum_{m=1}^3 \sum_{j=1}^2 \|\partial_j \mathbf{c}_n - \partial_j \mathbf{c}_m^{(i)}\|_1. \quad (7)$$

It has to be noted that the warping operator matrix \mathbf{A}^l in (7) is, in fact, $\begin{bmatrix} \mathbf{A}^l & 0 \\ 0 & \mathbf{A}^l \end{bmatrix}$ for the green channel problem $\mathbf{c}_2^{(i+1)}$. For brevity, we use the \mathbf{A}^l notation in all the channel optimization problems.

It is more efficient to estimate the warping functions from the sub-sampled channels of the mosaicked data than estimating them in the super-resolved space. Therefore, the warping operator \mathbf{A}^l is estimated from the sub-sampled channels of the mosaicked CFA images and performed on the sub-sampled color channels in (6). The process to estimate the warping operators is detailed in Sec. 2.2.

2.1. Burst Demosaicking Solution

The warping operators can be calculated from the mosaicked data before the reconstruction process. However, they are required to be applied in the reconstruction process at every iteration of the optimization solution of (7), which is quite costly. One can take the advantage of the fact that warping is invertible, i.e. $\mathbf{A}^l \mathbf{x} = \mathbf{y}$, $(\mathbf{A}^l)^{-1} \mathbf{y} = \mathbf{x}$ and reformulate (7) as follows:

$$\mathbf{c}_n^{(i+1)} = \arg \min_{\mathbf{c}_n} \frac{1}{2} \left\| \begin{bmatrix} \mathbf{H}^1 \hat{\mathbf{c}}_n^1 \\ \vdots \\ \mathbf{H}^L \hat{\mathbf{c}}_n^L \end{bmatrix} - \begin{bmatrix} \mathbf{S}_n \\ \times L \end{bmatrix} \mathbf{c}_n \right\|_2^2 + \lambda_1 \sum_{j=1}^2 \|\partial_j \mathbf{c}_n\|_1 + \lambda_2 \sum_{m=1}^3 \sum_{j=1}^2 \|\partial_j \mathbf{c}_n - \partial_j \mathbf{c}_m^{(i)}\|_1, \quad (8)$$

where $(\mathbf{A}^l)^{-1}$ is denoted by \mathbf{H}^l , and $\begin{bmatrix} \mathbf{a} \\ \vdots \\ \times L \end{bmatrix}$ denotes L

copies of a matrix \mathbf{a} stacked vertically. Thus $\mathbf{H}^l \hat{\mathbf{c}}_n^l$ can be treated as a constant through the entire process.

A variable splitting scheme, particularly an approach based on the ADMM [34, 35] can lead to a suitable solution to the multi-functional convex problems like (8). Accordingly, for each sub-problem \mathbf{c}_n , we introduce auxiliary variables and corresponding update variables \mathbf{z}_n , $\mathbf{w}_{n,j}$, $\mathbf{y}_{n,j,m}$, δ_n , $\mathbf{w}_{n,j}$, $\eta_{n,j,m}$, $n = 1, 2, 3$, $j = 1, 2$, $m = 1 \dots 3$, $m \neq n$. We then propose an ADMM numerical solver to further split the sub-problems and solve smaller convex sub-problems. Thus, considering every \mathbf{c}_n , we can define:

$$\mathbf{c}_n^{(i+1)} = \left(\mathbf{L} \mathbf{S}_n^T \mathbf{S}_n + 3 \sum_j^2 (\partial_j^T \partial_j) \right)^{-1} \left([\mathbf{S}_n^T \times L] (\mathbf{z}_n^{(i)} + \delta_n^{(i)}) + \sum_j^2 \partial_j^T (\mathbf{w}_{n,j}^{(i)} + \mathbf{w}_{n,j}^{(i)}) + \sum_{m=1, m \neq n}^3 \partial_j^T (\mathbf{y}_{n,j,m}^{(i)} + \eta_{n,j,m}^{(i)}) \right). \quad (9)$$

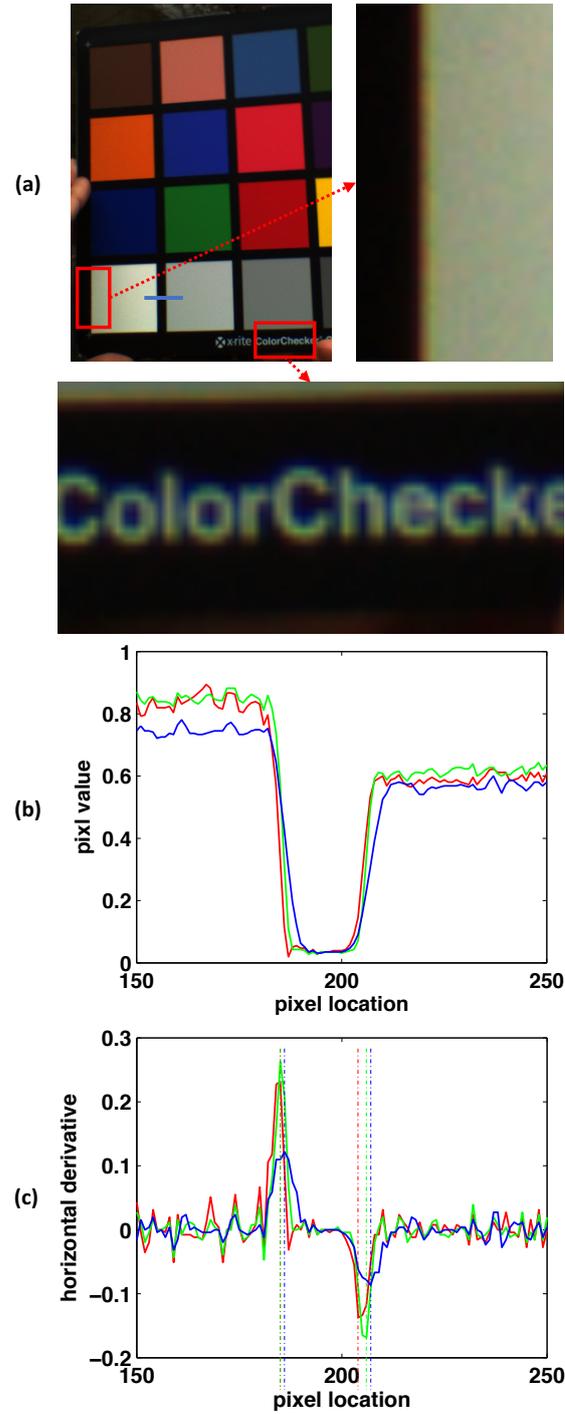


Figure 2 . Color fringing due to chromatic aberration. (a) An RGB image and a scan-line on the image. The insets show close-up displays of color artifacts that appear near edges. (b) RGB intensity profiles of the blue scan-line marked in (a). (c) RGB horizontal gradients of the scan-line showing that the pixel locations (peaks of the gradients) of the edges do not match from one channel to another.

Eq. (9) can be considered as a large linear system, i.e. $\mathbf{b} = \mathbf{K} \mathbf{x}$, where $\mathbf{x} = \mathbf{c}_n^{(i+1)}$, $\mathbf{K} = \mathbf{L} \mathbf{S}_n^T \mathbf{S}_n + 3 \sum_j^2 (\partial_j^T \partial_j)$, and \mathbf{b} is the rest of the equation. Needless to say, \mathbf{K} is an extremely large matrix, and its implementation is not feasi-

ble in practice. The fact that \mathbf{K} is composed of multiplication of matrices with their own transposes makes it a symmetric positive-semidefinite matrix. Thus (9) can be addressed by the standard linear solvers such as the conjugate gradient (CG) one, quite efficiently. Let the constant \mathbf{u}_n^l denote $\mathbf{H}^l \mathbf{c}_n^l$ for each \mathbf{z}_n . We can define the following normal equation:

$$\mathbf{z}_n^{(i+1)} = \underset{\mathbf{z}_n}{\operatorname{argmin}} \frac{1}{2} \left\| \begin{bmatrix} \mathbf{u}_n^1 \\ \vdots \\ \mathbf{u}_n^L \end{bmatrix} - \mathbf{z}_n \right\|_2^2 + \frac{\rho}{2} \left\| \mathbf{z}_n - \begin{bmatrix} \mathbf{S}_n \\ \vdots \\ \times L \end{bmatrix} \mathbf{c}_n^{(i+1)} + \mathfrak{z}_n^{(i)} \right\|_2^2, \quad (10)$$

and solve it as

$$\mathbf{z}_n^{(i+1)} = \left(\begin{bmatrix} \mathbf{u}_n^1 \\ \vdots \\ \mathbf{u}_n^L \end{bmatrix} + \rho \left(\begin{bmatrix} \mathbf{S}_n \\ \vdots \\ \times L \end{bmatrix} \mathbf{c}_n^{(i+1)} - \mathfrak{z}_n^{(i)} \right) \right) / (\rho + 1). \quad (11)$$

In a similar fashion, we define:

$$\mathbf{w}_{n,j}^{(i+1)} = \underset{\mathbf{w}_{n,j}}{\operatorname{argmin}} \lambda_1 \|\mathbf{w}_{n,j}\|_1 + \frac{\rho}{2} \|\mathbf{w}_{n,j} - \partial_j \mathbf{c}_n^{(i+1)} + \mathbf{w}_{n,j}^{(i)}\|_2^2, \quad (12)$$

$$\mathbf{y}_{n,j,m}^{(i+1)} = \underset{\mathbf{y}_{n,j,m}}{\operatorname{argmin}} \lambda_2 \|\mathbf{y}_{n,j,m} - \partial_j \mathbf{c}_m^{(i)}\|_1 + \frac{\rho}{2} \|\mathbf{y}_{n,j,m} - \partial_j \mathbf{c}_n^{(i+1)} + \mathbf{y}_{n,j,m}^{(i)}\|_2^2, \quad m \neq n \quad (13)$$

for $\mathbf{w}_{n,j}$, $\mathbf{y}_{n,j,m}$. The problems defined in (12) and (13) can be seen as the classic shrinkage regression problem. Their objectives allow a fast solution via soft-thresholding. Given the arbitrary vectors \mathbf{x} and \mathbf{b} , and γ as a positive scalar, the shrinkage problem with respect to \mathbf{x} and its solution are defined as:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \gamma \|\mathbf{x}\|_1 + \frac{1}{2} \|\mathbf{x} - \mathbf{b}\|_2^2$$

$$= \mathcal{S}_\gamma(\mathbf{b}) = \begin{cases} \mathbf{b} - \gamma, & \text{if } \mathbf{b} > \gamma \\ \mathbf{b} + \gamma, & \text{if } \mathbf{b} < -\gamma \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

where soft-thresholding is denoted by $\mathcal{S}_\gamma(\cdot)$, and the threshold value is denoted by γ . The two objective functions defined in (12) and (13) share similar characteristics with (14). Thus, we adopt the soft-thresholding (14) approach [36] to find their optimal solutions. In case of using isotropic TV in (6), $\mathbf{w}_{n,j}$ turns into a 2D vector. Therefore, a 2D soft-thresholding is required, as in [35, 36].

A solution to every sub-problem is now developed. Thus we can solve the main problem by alternating among the sub-problems, iteratively. The steps of the proposed burst demosaicking method are outlined in Algorithm 1. It is worth mentioning that implementation of the matrix form operators is not feasible since they are extremely large and memory inefficient. However, the corresponding linear operators, i.e. sampling, convolution can be used more efficiently. Also it should be noted that in the ADMM algorithm, ρ is slightly increased per iteration to guaranty a convergence [37].

Algorithm 1 Proposed demosaicking.

Require: $\mathbf{I}_{cfa1} \dots \mathbf{I}_{cfaL}$, $\mathbf{H}^1 \dots \mathbf{H}^L$, λ_1 , λ_2 , ρ

- 1: $\mathbf{c}_1^l = \mathbf{S}_1 \mathbf{I}_{cfa1}$, $\mathbf{c}_2^l = \mathbf{S}_2 \mathbf{I}_{cfa1}$, $\mathbf{c}_3^l = \mathbf{S}_3 \mathbf{I}_{cfa1}$, for $l = 1 \dots L$
- 2: $i = 1$, initiate elements of $\mathbf{c}_1^{(0)}$, $\mathbf{c}_2^{(0)}$, $\mathbf{c}_3^{(0)}$ with 1, and other auxiliary and update variables with 0
- 3: **while** stop criterion is not satisfied **do**
- 4: **for** $n = 1$ to 3 **do**
- 5: find $\mathbf{c}_n^{(i+1)}$ using CG applied on (9)
- 6: find $\mathfrak{z}_n^{(i+1)}$ using (11)
- 7: $\mathfrak{z}_n^{(i+1)} = \mathfrak{z}_n^{(i)} - \left(\begin{bmatrix} \mathbf{S}_n \\ \vdots \\ \times L \end{bmatrix} \mathbf{c}_n^{(i+1)} - \mathfrak{z}_n^{(i+1)} \right)$
- 8: **for** $j = 1$ to 2 **do**
- 9: $\mathbf{w}_{n,j}^{(i+1)} = \mathcal{S}_{\lambda_1/\rho}(\partial_j \mathbf{c}_n^{(i+1)} - \mathbf{w}_{n,j}^{(i)})$
- 10: $\mathbf{w}_{n,j}^{(i+1)} = \mathbf{w}_{n,j}^{(i)} - (\partial_j \mathbf{c}_n^{(i+1)} - \mathbf{w}_{n,j}^{(i+1)})$
- 11: **for** $m = 1$ to 3, $m \neq n$ **do**
- 12: $\mathbf{y}_{n,j,m}^{(i+1)} = \mathcal{S}_{\lambda_2/\rho}(\partial_j \mathbf{c}_n^{(i+1)} - \mathbf{y}_{n,j,m}^{(i)} - \partial_j \mathbf{c}_m^{(i)}) + \partial_j \mathbf{c}_m^{(i)}$
- 13: $\mathbf{y}_{n,j,m}^{(i+1)} = \mathbf{y}_{n,j,m}^{(i)} - (\partial_j \mathbf{c}_n^{(i+1)} - \mathbf{y}_{n,j,m}^{(i+1)})$
- 14: **end for**
- 15: **end for**
- 16: **end for**
- 17: $\rho = \rho \times 1.1$, $i = i + 1$
- 18: **end while**
- 19: **return** $r = \mathbf{c}_1^{(i)}$, $g = \mathbf{c}_2^{(i)}$, $b = \mathbf{c}_3^{(i)}$

The proposed solution is so scalable that one can add other priors such as ringing suppressors, non-local-means denoiser, etc. to the problem (8) and add their solutions [6, 38, 39] to Algorithm 1 straightforwardly. Moreover, the proposed method can take other linear imaging pipeline operations such as a correction of the sensor's dead pixels or saturated pixels, lens distortion warping, color correction [6], and even calibrated lens blur functions [40] into account as in [29].

2.2. Efficient Alignment

The stack of burst images is very likely misaligned due to a small camera motion. We handle this by aligning all L observations \mathbf{I}_{cfa1} to a reference (e.g. the first one). This is performed by computing a sub-pixel-accurate registration on the sub-sampled color channels.

Considering just a translation between two images $f(x, y)$ and $g(x - x_0, y - y_0)$ whose pixel coordinates are denoted by (x, y) and the amount of translate is denoted by (x_0, y_0) , the usual technique to address this problem is to compute an up-sampled cross-correlation between the image to register and a reference image by means of a fast Fourier transform (FFT) and locating its peak. More formally, we minimize the normalized root-mean-square error (NRMSE) between $f(x, y)$ and $g(x, y)$, defined in [41]

$$\min_{x_0, y_0, \alpha} \frac{\sum_{x,y} |\alpha g(x - x_0, y - y_0) - f(x, y)|^2}{\sum_{x,y} |f(x, y)|^2} \quad (15)$$

whose minimizer is equivalent to the maximizer of the cross-correlation of $f(x, y)$ and $g(x, y)$ as:

$$\sum_{x,y} f(x, y) g^*(x - x_0, y - y_0) \quad (16)$$

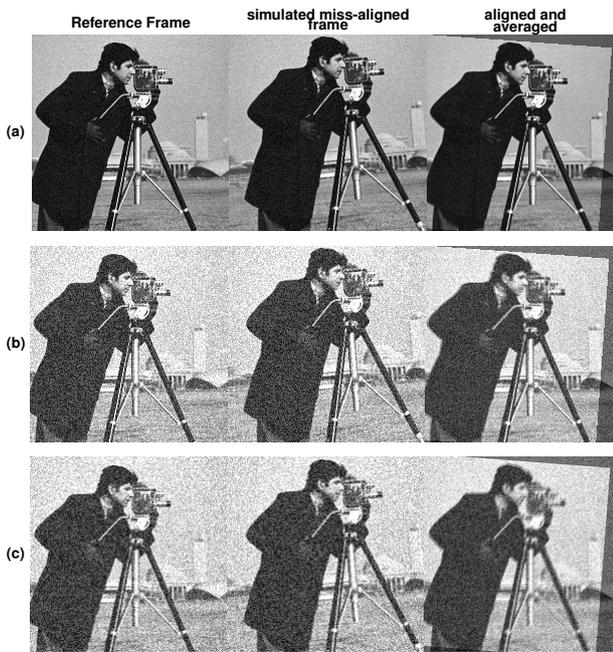


Figure 3 . (a), (b) Efficient alignment under different noise levels. (a) Noise $\sigma = 0.1$. (b) Noise $\sigma = 0.25$. (c) FFT-based alignment using log-scaled values with noise $\sigma = 0.25$.

which can be found efficiently using its Fourier counterparts as:

$$\sum_{u,v} F(u,v)G^*(u,v) \exp(i2\pi(ux_0/M + vy_0/N)) \quad (17)$$

where i denotes the imaginary number and $*$ denotes complex-conjugate. In (17), $F(u,v)$ and $G(u,v)$ denote $f(x,y)$ and $g(x,y)$ in the Fourier domain, respectively. Hence, finding x_0 and y_0 within a fraction, $1/\tau$, of a pixel is [42] (i) compute $F(u,v)$ and $G(u,v)$, (ii) embed the product $F(u,v)G^*(u,v)$ in a larger array of zeros of dimensions $\tau M \times \tau N$, (iii) compute an inverse Fourier transform to obtain an up-sampled cross-correlation, and (iv) locate its peak. Hereafter, we refer to this process as Fourier-based translation registration.

Consider the polar coordinate system (r, θ) , where r denotes radial distance from the center (x_c, y_c) and θ denotes angle:

$$r = \sqrt{(x - x_c)^2 + (y - y_c)^2}, \theta = \tan^{-1} \frac{(y - y_c)}{(x - x_c)}. \quad (18)$$

The benefit of the polar coordinate transformation of $f(x,y)$ and $g(x,y)$ denoted by $f_p(r, \theta)$ and $g_p(r, \theta)$ is that the scale and rotation changes in the images can be induced by modifying (r, θ) . Hence, in order to find the amount of rotation and scale between the two images, one can consider the notation $f_p(r, \theta)$ and $g_p(r - r_0, \theta - \theta_0)$ and use the Fourier-based translation registration discussed before. The estimated r_0 and θ_0 correspond the amount of scale and rotation, respectively, in the original Cartesian space $f(x,y)$ and $g(x,y)$.

We assume a 2-D affine transform between the burst of images, i.e. the transform consists of translation and scaling along the horizontal and vertical axes and a planar rotation. This is a reasonable assumption since the burst images are typically captured in a short period of time with

minimal handshake. Let \hat{c}_{ref}^1 and \hat{c}_i^1 denote the 2-D sub-sampled red channel of the reference CFA image and the i -th one. The 2-D transform is obtained as:

$$h^l = \begin{bmatrix} r_0 \cos(\theta_0) & r_0 \sin(\theta_0) & x_0 \\ -r_0 \sin(\theta_0) & r_0 \cos(\theta_0) & y_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (19)$$

where r_0 and θ_0 are obtained using the Fourier-based translation registration of the polar coordinate transformation of \hat{c}_{ref}^1 and \hat{c}_i^1 . Then the warped \hat{c}_i^1 with regards to r_0 and θ_0 in the Cartesian space and \hat{c}_{ref}^1 are again gone through the translation registration, this time in the Cartesian coordinate system, in order to find x_0 and y_0 . The obtained transform matrix h^l can be used for other color channels of the CFA images. We denote the process of alignment, i.e. warping using the transform h^l , as a linear operator \mathbf{H}^l in (9).

Fig. 3 shows the alignment results using the proposed method given two sets of simulated miss-aligned frames. The right column in each row shows the alignment result averaged with the reference frame. In an experiment, the additive noise follows a Gaussian distribution with standard-deviation 0.1, and in the other, the standard-deviation of the noise is 0.25. For both cases, we used one randomly generated 2D affine homography with $\theta_0 = 4.4$, $r_0 = 1.02$, $x_0 = 5.6$, $y_0 = 2.3$. The classic methods such as [43] apply such phase correlation techniques on the log-scaled Fourier transformed images. This helps in locating the peak in the Fourier's domain in case of noise-free images. However, in case of images contaminated with a large amount of noise, the accuracy degrades as the random noise affects the peak in the frequency domain. Fig. 3(c) shows an example of such a poor alignment, compared to using FFT directly in Fig. 3(b), in the presence of random noise with standard deviation of 0.25 where blur artifacts appear near edges in the averaging result.

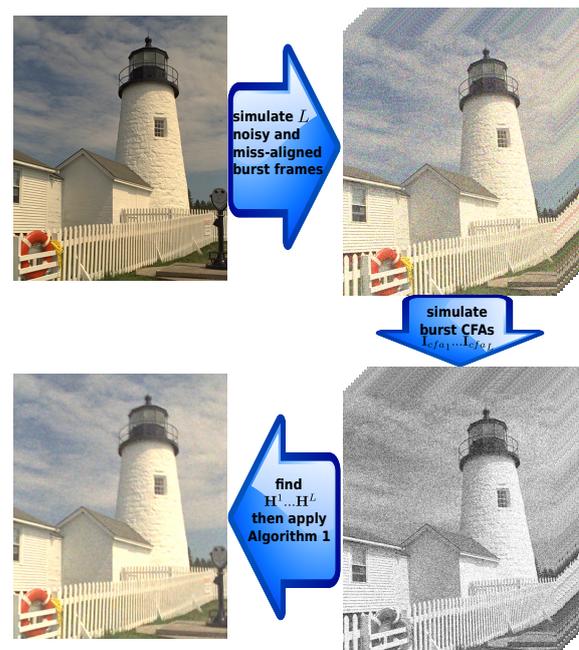


Figure 4 . Simulation of burst reconstruction.

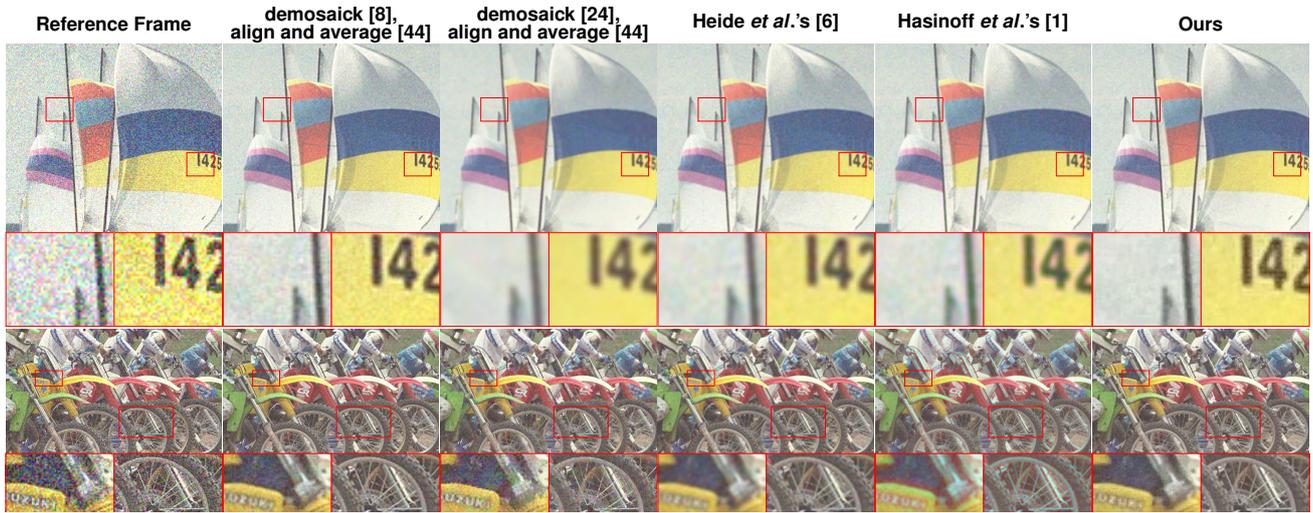


Figure 5 .Burst reconstruction of simulated RAW Kodak dataset by means of individually demosaicking using [8] and [24] followed by the method in [44], methods introduced in [6] and [1], and the proposed method. Noise level is set to $\sigma = 0.25$ in RAW images with pixel values normalized to $[0, 1]$.

Table 1 . Image quality measured on the Kodak data-set after reconstruction using different methods. Zero-mean Gaussian noise $\sigma = 0.25$ was added to each mosaicked image (for normalized pixel values in $[0, 1] \subset \mathbb{R}$). M1 and M2 S-CIELAB ΔE^* and CPSNR, respectively.

Image	Demosaick [8] and [44]		Demosaick [24] and [44]		[6]		[1]		Ours	
	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
1	12.28	28.15	12.12	29.60	11.89	29.27	12.28	28.13	11.10	28.20
2	12.65	28.84	11.52	29.93	12.73	29.03	12.10	29.05	12.10	30.23
3	12.96	28.90	11.72	30.88	12.85	28.97	11.91	30.97	10.95	29.29
4	12.09	28.00	11.46	29.76	11.98	29.13	12.11	28.77	11.75	31.15
5	12.05	28.42	12.25	29.57	12.44	29.19	12.25	29.11	11.92	29.59
6	11.91	27.82	12.25	28.86	11.56	29.06	11.50	29.14	10.66	30.33
7	12.57	27.91	11.42	28.51	12.05	28.99	11.68	28.54	12.45	30.12
8	12.85	28.76	11.25	28.62	11.77	29.33	12.46	30.79	11.11	28.09
9	11.69	28.51	11.68	30.49	11.28	29.38	11.01	30.88	10.96	30.43
10	12.51	28.64	12.56	29.01	10.90	28.12	11.42	29.98	10.97	28.16
11	12.96	28.65	11.91	28.95	12.68	29.04	11.59	29.14	11.80	30.65
12	11.53	27.98	11.92	28.45	12.76	29.15	11.07	29.85	12.31	28.26
13	12.28	28.14	12.36	29.20	12.05	28.96	11.93	30.26	11.09	30.56
14	11.67	29.01	12.27	29.83	12.73	29.20	11.86	30.88	11.26	30.90
15	12.79	27.87	11.78	30.54	11.36	29.28	12.47	30.82	11.10	31.48
16	12.75	28.23	12.42	29.45	12.55	29.32	12.78	29.09	10.93	28.61
17	11.92	28.03	11.05	28.68	11.60	28.86	11.94	30.12	10.85	29.67
18	12.88	28.45	11.52	29.75	11.57	29.13	12.53	29.29	12.07	29.05
19	11.97	28.71	11.97	30.54	12.53	29.17	12.47	28.06	12.32	28.96
20	12.52	27.96	12.58	29.07	11.47	29.09	12.36	29.90	10.57	29.31
21	11.81	28.02	11.70	30.94	11.40	29.08	12.61	28.34	10.35	30.91
22	12.43	28.01	12.77	28.04	12.05	29.06	12.34	29.71	11.50	29.12
23	12.36	27.92	12.41	29.94	12.80	28.91	11.18	28.64	11.09	28.93
24	12.53	28.02	11.20	29.15	12.83	28.97	12.71	30.60	11.38	31.23
Mean	12.33	28.28	11.92	29.48	12.07	29.07	12.02	29.58	11.35	29.71

3. Experimental Results

We first quantitatively assess our method on a synthetic test set. We then qualitatively evaluate our method and demonstrate its ability to generalize on real bursts captured by a mobile phone and compare against several recent techniques. The convergence profile for Algorithm 1 was analyzed empirically and $\rho = 0.001$ was found as an appropriate initiation for a fast convergence. This value leads to 10 to 12 iterations. Also $0.001 \leq \lambda_1 \leq 0.007$ was used to avoid over-smoothing or amplifying noise and ar-

tifacts in the reconstructed images from the test dataset. Finally, $0.001 \leq \lambda_2 \leq 0.01$ was used as a good choice to prevent color fringing from appearing in the results.

3.1. Results on synthetic test set

We tested the proposed method on the Kodak image dataset. This dataset contains 24 digitized film-captured images often used as test images for evaluating demosaicking methods and similar image construction meth-

ods. For each image in the dataset, we first generated 10 ($L = 10$) random 2D affine homographies (rotation, scaling, and translation) and generated 10 different warped versions of the original image. Each image was then contaminated with additive zero-mean Gaussian noise with a reasonable standard deviation σ . Finally, Bayer’s sampling was applied on each image to simulate a burst of I_{cfa} . Fig. 4 depicts the process of synthesizing a burst set for each image. Each burst set was then processed using the proposed method. It has to be noted that since the purpose of burst reconstruction is generally to remove strong noise induced in low light conditions, we selected a high value for the standard deviation of the noise i.e. $\sigma = 0.25$ in our simulations.

The typical burst reconstruction techniques [5, 44, 45] include an alignment of already demosaicked images and fusing the resulting images through averaging. We also applied the alignment and averaging technique of Adams et al. [44] on the dataset to provide a comparison with the proposed joint demosaicking and burst denoising. To this end, we used the methods in [8] and also in [24] in order to demosaick the images before applying the burst reconstruction of Adams et al. [44]. Besides, we processed the synthetic bursts with other techniques [1, 6] that took demosaicking into account during the burst reconstruction. Fig. 5 shows the sample images reconstructed using the proposed method, averaging aligned demosaicked set [44], and the reconstruction methods proposed in [6] and [1]. Inaccurate alignment and consequently burst reconstruction lead to blurred edges in the image. Also since the methods are also meant to carry-out demosaicking, one should expect color artifacts in the final result, in case a method does not perform well.

The availability of the original (ground-truth) versions of the burst images made, it possible to evaluate the performance of different methods using reference metrics. The color peak signal-to-noise ratio (CPSNR) and ΔE^* in the S-CIELAB colour space are the commonly used metrics to evaluate demosaicking techniques [23, 46]. We measured CPSNR and S-CIELAB ΔE^* on the reconstructed

images to evaluate the methods quantitatively. Table 1 lists the quality measurements for the best results obtained using the demosaicking methods [8] and [24], followed by averaging [44], the reconstruction methods of [6] and [1], and the proposed method. This table indicates that the proposed burst demosaicking method results in a better average CPSNR and S-CIELAB ΔE^* compared to the other methods. It should be noted that in the reconstruction method introduced in [6] employs non-local means (NLM) denoising at every iteration, which is very costly and not practical. Therefore, we replaced it by the standard TV prior.

3.2. Generalizing to real data

The proposed burst reconstruction method was also tested on the real raw images captured by mobile phones. We conducted our tests on a sub-set of the raw burst dataset provided by the authors of [1]. The dataset consists of 3640 bursts of full resolution raw images captured using a variety of Android mobile cameras. Examples of such images and their reconstruction results using 4 different methods are shown in Fig. 6. As this figure suggests, the proposed method results in high-quality images compared to the other techniques in terms of a low amount of noise and reduced false color.

In these experiments, we treated the camera’s image signal processor (ISP) as a black box where we did not have any control on its processing blocks. Therefore, we by-passed it and only performed a static white-balancing on the final burst-demosaicked results shown in Fig. 6.

A non-optimized Matlab implementation of the proposed method considering 8 captures using a 4 Mega-pixel sensor takes around 4 minutes to run. On a same platform, the methods proposed in [1] and in [6] take 3 minutes and 8 minutes, respectively. Hence, while the proposed method results in a better quality, specially in terms of reduced color artifacts, it performs efficiently compared to the other two methods.

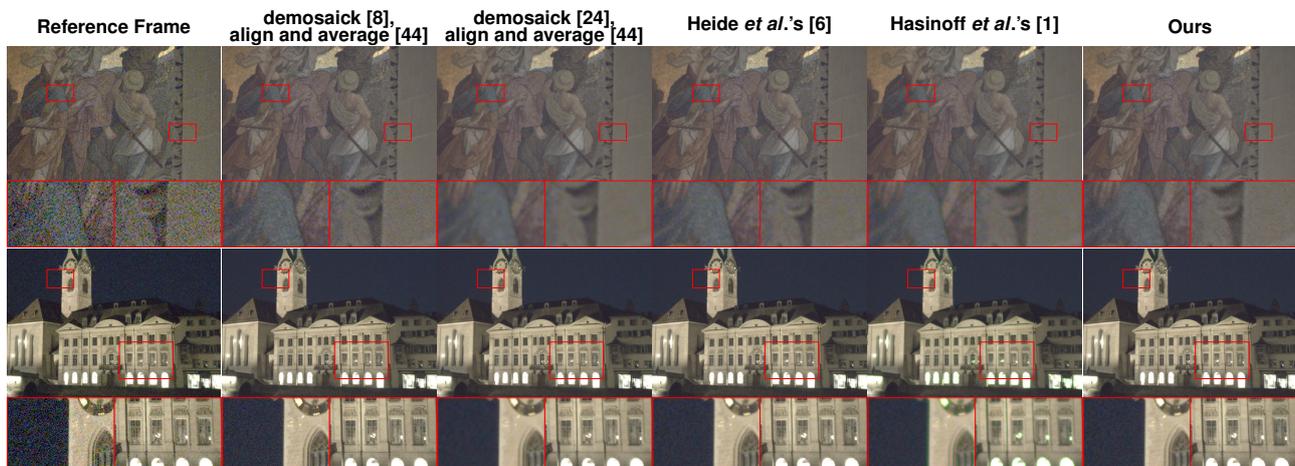


Figure 6 . Burst reconstruction of real RAW data by means of individually demosaicking using [8] and [24] followed by the method in [44], methods introduced in [6] and [1], and the proposed method.

4. Conclusion and Discussion

We formulated the burst denoising and demosaicking into a single problem and proposed an effective approach to this joint problem. The method takes advantage of the facts that a more accurate estimate of the underlying image is obtained by minimizing a least-squares regularization problem given a burst of RAW captures and that the color channels in the ground-truth image share high-frequency details, spatially. We also showed that the cumbersome process of burst alignment can be performed quite efficiently in the Fourier's domain by measuring the amount of phase difference in the Cartesian and polar coordinate systems to account for the amount of pixel translation, rotation, and also scaling. Our experiments performed on synthetic data as well as real RAW captures indicate a considerable performance for the proposed joint burst denoising and demosaicking approach in terms of reduced noise, color fringing, and demosaicking artifacts compared to similar approaches.

Burst photography is a common feature in today's cameras not only for the purpose of denoising, but also generating high-dynamic range (HDR) photographs. Hence, the ultimate goal of burst denoising is not only replacing the classic single-image denoising techniques. Instead, it can be considered as a complimentary enhancement in the commonly used HDR photography.

References

- [1] Hasinoff, S. W., Sharlet, D., Geiss, R., Adams, A., Barron, J. T., Kainz, F., Chen, J., and Levoy, M. (2016). Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (SIGGRAPH)*, vol. 35, no. 6, p. 192, 2016.
- [2] Buades, A., Coll, B., and Morel, J.-M. (2005). A non-local algorithm for image denoising. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2005, pp. 60–65.
- [3] Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2007). Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.
- [4] Saeedzarandi, M., Nezamabadi-pour, H., and Saryazdi, S. (2020). Statistical wavelet-based image denoising using scale mixture of normal distributions with adaptive parameter estimation. *Journal of AI and Data Mining*, vol. 8, no. 2, pp. 289–301, 2020.
- [5] Liu, Z., Yuan, L., Tang, X., Uyttendaele, M., and Sun, J. (2014). Fast burst images denoising. *ACM Transactions on Graphics (SIGGRAPH)*, vol. 33, no. 6, 2014.
- [6] Heide, F., Steinberger, M., Tsai, Y.-T., Rouf, M., Pająk, D., Reddy, D., Gallo, O., Liu, J., Heidrich, W., Egiazarian, K. *et al.* (2014). Flexisp: a flexible camera image processing framework. *ACM Transactions on Graphics (SIGGRAPH Asia)*, vol. 33, no. 6, p. 231, 2014.
- [7] Heide, F., Diamond, S., Nießner, M., Ragan-Kelley, J., Heidrich, W., and Wetzstein, G. (2016). Proximal: Efficient image optimization using proximal algorithms. *ACM Transactions on Graphics (SIGGRAPH)*, vol. 35, no. 4, p. 84, 2016.
- [8] Malvar, H. S., He, L.-w., and Cutler, R. (2004). High-quality linear interpolation for demosaicing of bayer-patterned color images. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, 2004, pp. iii–485.
- [9] Hirakawa, K. and Parks, T. W. (2005). Adaptive homogeneity-directed demosaicing algorithm. *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 360–369, 2005.
- [10] Zhang, L. and Wu, X. (2005). Color demosaicking via directional linear minimum mean square-error estimation. *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2167–2178, 2005.
- [11] Dubois, E. (2005). Frequency-domain methods for demosaicking of bayer-sampled color images. *IEEE Signal Processing Letters*, vol. 12, no. 12, p. 847, 2005.
- [12] Kwan, C., Chou, B., Kwan, L.-Y. M., Larkin, J., Ayhan, B., Bell, J. F., and Kerner, H. (2018). Demosaicing enhancement using pixel-level fusion. *Signal, Image and Video Processing*, vol. 12, no. 4, pp. 749–756, 2018.
- [13] Zhang, L., Wu, X., and Zhang, D. D. (2007). Color reproduction from noisy cfa data of single sensor digital cameras. *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2184–2197, 2007.
- [14] Hirakawa, K. and Parks, T. W. (2006). Joint demosaicing and denoising. *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp. 2146–2157, 2006.
- [15] Paliy, D., Katkovnik, V., Bilcu, R., Alenius, S., and Egiazarian, K. (2007). Spatially adaptive color filter array interpolation for noiseless and noisy data. *International Journal of Imaging Systems and Technology*, vol. 17, no. 3, pp. 105–122, 2007.
- [16] Zhang, L., Lukac, R., Wu, X., and Zhang, D. (2009). PCA-based spatially adaptive denoising of cfa images for single-sensor digital cameras. *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 797–812, 2009.
- [17] Hirakawa, K., Meng, X.-L., and Wolfe, P. J. (2007). A framework for wavelet-based analysis and processing of color filter array images with applications to denoising and demosaicing. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 1, 2007, pp. 1–597.
- [18] Mildenhall, B., Barron, J. T., Chen, J., Sharlet, D., Ng, R., and Carroll, R. (2018). Burst denoising with kernel prediction networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 2502–2510.
- [19] Gunturk, B. K., Glotzbach, J., Altunbasak, Y., Schafer, R. W., and Mersereau, R. M. (2005). Demosaicking: color filter array interpolation. *IEEE Signal processing magazine*, vol. 22, no. 1, pp. 44–54, 2005.
- [20] Li, X., Gunturk, B., and Zhang, L. (2008). Image demosaicing: A systematic survey. *Proc. SPIE*, vol. 6822, no. 2008, 2008, p. 6822.
- [21] Condat, L. and Mosaddegh, S. (2012). Joint demosaicking and denoising by total variation minimization. *IEEE International Conference on Image Processing (ICIP)*, 2012, pp. 2781–2784.
- [22] Saito, T. and Komatsu, T. (2008). Demosaicing approach based on extended color total-variation regularization. *IEEE International Conference on Image Processing (ICIP)*, 2008, pp. 885–888.
- [23] Menon, D. and Calvagno, G. (2009). Regularization approaches to demosaicking. *IEEE Transactions on Image Processing*, vol. 18, no. 10, pp. 2209–2220, Oct 2009.
- [24] Tan, H., Zeng, X., Lai, S., Liu, Y., and Zhang, M. (2017). Joint demosaicing and denoising of noisy bayer images with ADMM. *Image Processing (ICIP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2951–2955.
- [25] Farsiu, S., Elad, M., and Milanfar, P. (2006). Multiframe demosaicing and super-resolution of color images. *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 141–159, 2006.
- [26] Khashabi, D., Nowozin, S., Jancsary, J., and Fitzgibbon, A. W. (2014). Joint demosaicing and denoising via learned nonparametric random fields. *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 4968–4981, 2014.
- [27] Ehret, T., Davy, A., Arias, P., and Facciolo, G. (2019). Joint demosaicking and denoising by fine-tuning of bursts of raw images. *IEEE International Conference on Computer Vision (CVPR)*, 2019, pp. 8868–8877.
- [28] Kokkinos, F. and Lefkimmiatis, S. (2019). Iterative residual cnns for burst photography applications. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 5929–5938.
- [29] Azizi, R. and Latif, A. (2018). Cross-channel regularisation for joint demosaicking and intrinsic lens deblurring. *IET Image Processing*, vol. 12, no. 9, pp. 1683–1691, 2018.
- [30] Gonzales, R. C. and Woods, R. E. (2008). *Digital image processing (3rd edition)*. pp. 352–361, 2008.

- [31] Osher, S., Burger, M., Goldfarb, D., Xu, J., and Yin, W. (2005). An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 460–489, 2005.
- [32] Wang, Y., Yang, J., Yin, W., and Zhang, Y. (2008). A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.
- [33] Farsiu, S., Robinson, M. D., Elad, M., and Milanfar, P. (2004). Fast and robust multiframe super resolution. *IEEE transactions on image processing*, vol. 13, no. 10, pp. 1327–1344, 2004.
- [34] Afonso, M. V., Bioucas-Dias, J. M., and Figueiredo, M. A. (2010). Fast image recovery using variable splitting and constrained optimization. *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2345–2356, 2010.
- [35] Chan, S. H., Wang, X., and Elgendy, O. A. (2017). Plug-and-play ADMM for image restoration: Fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 84–98, 2017.
- [36] Yang, J., Yin, W., Zhang, Y., and Wang, Y. (2009). A fast algorithm for edge-preserving variational multichannel image restoration. *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 569–592, 2009.
- [37] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [38] Mosleh, A., Sola, Y. E., Zargari, F., Onzon, E., and Langlois, J. P. (2018). Explicit ringing removal in image deblurring. *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 580–593, 2018.
- [39] Bouzari, H. (2012). An improved regularization method for artifact rejection in image super-resolution. *Signal, Image and Video Processing*, vol. 6, no. 1, pp. 125–140, 2012.
- [40] Mosleh, A., Green, P., Onzon, E., Begin, I., and Langlois, J.-M.-P. (2015). Camera intrinsic blur kernel estimation: A reliable framework. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 4961–4968.
- [41] Fienup, J. R. (1997). Invariant error metrics for image reconstruction. *Applied optics*, vol. 36, no. 32, pp. 8352–8357, 1997.
- [42] Guizar-Sicairos, M., Thurman, S. T., and Fienup, J. R. (2008). Efficient subpixel image registration algorithms. *Optics letters*, vol. 33, no. 2, pp. 156–158, 2008.
- [43] Reddy, B. S. and Chatterji, B. N. (1996). An fft-based technique for translation, rotation, and scale-invariant image registration. *IEEE Transactions on Image Processing*, vol. 5, no. 8, pp. 1266–1271, 1996.
- [44] Adams, A., Gelfand, N., and Pulli, K. (2008). Viewfinder alignment. *Computer Graphics Forum*, 2008, pp. 597–606.
- [45] Delbracio, M. and Sapiro, G. (2015). Burst deblurring: Removing camera shake through fourier burst accumulation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 2385–2393.
- [46] Leung, B., Jeon, G., and Dubois, E. (2011). Least-squares luma-chroma demultiplexing algorithm for bayer demosaicking. *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1885–1894, 2011.

تلفیق موزاییک‌زدایی و حذف نویز با استفاده از رشته تصاویر از طریق رگولاریزاسیون و یک فرآیند تراز

بهبینه

رضا عزیزی^۱ و علی محمد لطیف^{۲*}

^۱ گروه مهندسی کامپیوتر، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.

^۲ گروه مهندسی کامپیوتر، دانشگاه یزد، یزد، ایران.

ارسال ۲۰۱۹/۱۲/۱۹؛ بازنگری ۲۰۲۰/۰۴/۲۳؛ پذیرش ۲۰۲۰/۰۸/۰۱

چکیده:

در این مقاله نشان می‌دهیم که بازسازی تصویر با استفاده از یک رشته تصاویر (burst) که با فاصله زمانی بسیار کمی توسط یک دوربین بدست آمده اند و هر کدام مستقل از دیگری موزاییک‌زدایی شده‌اند باعث انتشار خطای موزاییک‌زدایی در سرتاسر خط لوله‌ی فرایند پردازش تصویر می‌شود. برای رفع این نقص ما ابتدا فرایند تراز (alignment) تصاویر و نمونه‌گیری آرایه‌ی فیلترهای رنگ سنسور را در یک اپراتور خطی مدل می‌کنیم. سپس فرایند بازسازی تصویر از رشته تصاویر و فرایند موزاییک‌زدایی تصویر خام را به صورت یک مسئله بهینه‌سازی واحد مطرح می‌کنیم. برای این منظور یک توزیع احتمال پیشین (prior) تحت عنوان تقاطع کانالهای رنگی معرفی می‌کنیم و از آن در یک راه حل عددی بر اساس روش جهت متناوب مضارب (ADMM) جهت بازسازی تصویر استفاده می‌کنیم. همچنین روش ارائه شده در این مقاله از پیچیدگی‌های تخمین تراز که در پیش پردازش روشهای بازسازی معمول با استفاده از رشته تصاویر وجود دارد اجتناب می‌کند و بجای آن از یک روش همبستگی فازی در حوزه فوریه تصاویر جهت تخمین میزان تغییر جابجایی حول محور افقی و عمودی، میزان چرخش نسبی و میزان بزرگنمایی نسبی بهره می‌گیرد. نتیجه حاصل از این مراحل این است که در مقایسه با روشهای مشابه موجود، روش ارائه شده برای نویززدایی با استفاده از رشته تصاویر و موزاییک‌زدایی از تصاویر خام به صورت همزمان کیفیت تصاویر بازسازی شده را به طور قابل توجهی افزایش می‌دهد.

کلمات کلیدی: تصویرسازی رشته‌ای، موزاییک‌زدایی تصویر، روش جهت متناوب مضارب، تراز موثر.