Multiframe Demosaicing and Super-Resolution of Color Images

Sina Farsiu, Michael Elad, and Peyman Milanfar, *Senior Member, IEEE*

***Abstract*—In the last two decades, two related categories of problems have been studied independently in image restoration literature: super-resolution and demosaicing. A closer look at these problems reveals the relation between them, and, as conventional color digital cameras suffer from both low-spatial resolution and color-filtering, it is reasonable to address them in a unified context. In this paper, we propose a fast and robust hybrid method of super-resolution and demosaicing, based on a maximum *a posteriori* estimation technique by minimizing a multiterm cost function. The L_1 norm is used for measuring the difference between the projected estimate of the high-resolution image and each low-resolution image, removing outliers in the data and errors due to possibly inaccurate motion estimation. Bilateral regularization is used for spatially regularizing the luminance component, resulting in sharp edges and forcing interpolation along the edges and not across them. Simultaneously, Tikhonov regularization is used to smooth the chrominance components. Finally, an additional regularization term is used to force similar edge location and orientation in different color channels. We show that the minimization of the total cost function is relatively easy and fast. Experimental results on synthetic and real data sets confirm the effectiveness of our method.**

***Index Terms*—Color enhancement, demosaicing, image restoration, robust estimation, robust regularization, super-resolution.**

I. Introduction

SEVERAL distorting processes affect the quality of images acquired by commercial digital cameras. Some of the more important distorting effects include warping, blurring, color-filtering, and additive noise. A common image formation model for such imaging systems is illustrated in Fig. 1[1]. In this model, a real-world scene is seen to be warped at the camera lens because of the relative motion between the scene and camera. The imperfections of the optical lens results in the blurring of this warped image which is then subsampled and color-filtered at the CCD. The additive readout noise at the CCD will further degrade the quality of captured images.

Manuscript received June 21, 2004; revised December 15, 2004. This work was supported in part by the National Science Foundation under Grant CCR-9984246, in part by the U.S. Air Force under Grant F49620-03-1-0387, and in part by the National Science Foundation Science and Technology Center for Adaptive Optics, managed by the University of California, Santa Cruz, under Cooperative Agreement AST-9876783. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Robert D. (G. E.) Nowak.

S. Farsiu and P. Milanfar are with the Electrical Engineering Department, University of California, Santa Cruz CA 95064 USA (e-mail: farsiu@ee.ucsc.edu; milanfar@ee.ucsc.edu).

M. Elad is with the Computer Science Department, The Technion—Israel Institute of Technology, Haifa, Israel (e-mail: elad@cs.technion.ac.il).

Digital Object Identifier 10.1109/TIP.2005.860336

[1]This paper (with all color pictures and a MATLAB-based software package for resolution enhancement) is available at http://www.ee.ucsc.edu/~milanfar.

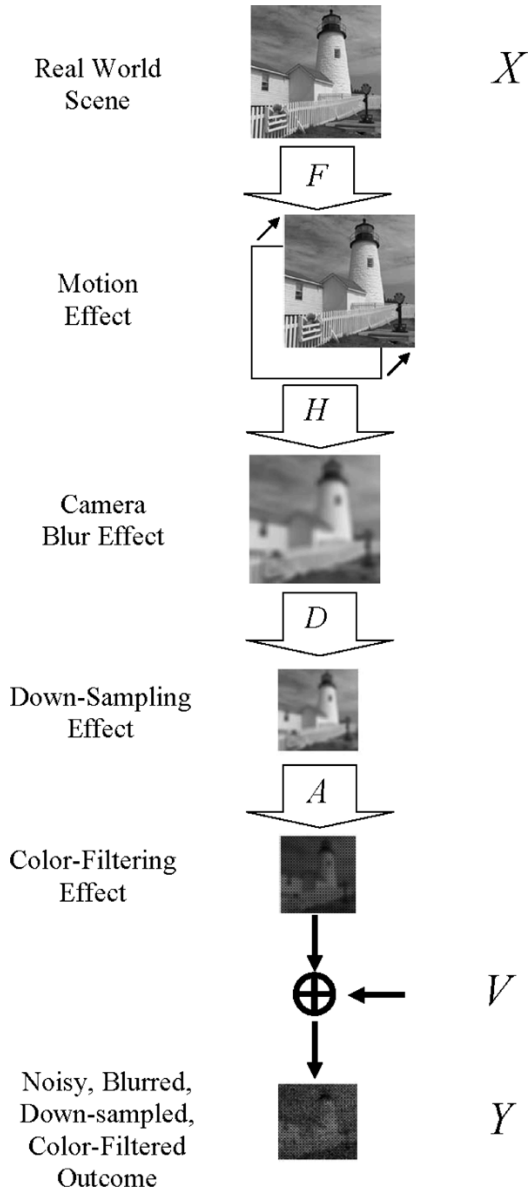


Fig. 1. Block diagram representing the image formation model considered in this paper, where X is the intensity distribution of the scene, V is the additive noise, and Y is the resulting color-filtered low-quality image. The operators F, H, D, and A are representatives of the warping, blurring, down-sampling, and color-filtering processes, respectively.

There is a growing interest in the multiframe image reconstruction algorithms that compensate for the shortcomings of the imaging system. Such methods can achieve high-quality images using less expensive imaging chips and optical components by capturing multiple images and fusing them.

1057-7149/$20.00 © 2006 IEEE

In digital photography, two image reconstruction problems have been studied and solved independently—super-resolution (SR) and demosaicing. The former refers to the limited number of pixels and the desire to go beyond this limit using several exposures. The latter refers to the color-filtering applied on a single CCD array of sensors on most cameras, that measures a subset of red (R), green (G), and blue (B) values, instead of a full RGB field.[2] It is natural to consider these problems in a joint setting because both refer to resolution limitations at the camera. Also, since the measured images are mosaiced, solving the super-resolution problem using preprocessed (demosaiced) images is suboptimal and, hence, inferior to a single unifying solution framework. In this paper, we propose a fast and robust method for joint multiframe demosaicing and color super-resolution.

The organization of this paper is as follows. In Section II, we review the super-resolution and demosaicing problems and the inefficiency of independent solutions for them. In Section III, we formulate and analyze a general model for imaging systems applicable to various scenarios of multiframe image reconstruction. We also formulate and review the basics of the maximum *a posteriori* (MAP) estimator, robust data fusion, and regularization methods. Armed with material developed in earlier sections, in Section IV, we present and formulate our joint multiframe demosaicing and color-super-resolution method. In Section V, we review two related methods of multiframe demosaicing. Simulations on both synthetic and real data sequences are given in Section VI, and concluding remarks are drawn in Section VII.

II. Overview of Super-Resolution and Demosaicing Problems

In this section, we study and review some of the previous work on super-resolution and demosaicing problems. We show the inefficiency of independent solutions for these problems and discuss the obstacles to designing a unified approach for addressing these two common shortcomings of digital cameras.

A. Super-Resolution

Digital cameras have a limited spatial resolution, dictated by their utilized optical lens and CCD array. Surpassing this limit can be achieved by acquiring and fusing several low-resolution (LR) images of the same scene, producing high-resolution (HR) images; this is the basic idea behind super-resolution techniques [1]–[4].

In the last two decades, a variety of super-resolution methods have been proposed for estimating the HR image from a set of LR images. Early works on SR showed that the aliasing effects in the LR images enable the recovery of the HR-fused image, provided that a relative subpixel motion exists between the under-sampled input images [5]. However, in contrast to the clean and practically naive frequency domain description of SR in that early work, in general, SR is a computationally complex and numerically ill-behaved problem in many instances [6]. In recent years, more sophisticated SR methods were developed (see [3] and [6]–[10] as representative works).

[2]Three CCD cameras which measure each color field independently tend to be relatively more expensive.

Note that almost all super-resolution methods to date have been designed to increase the resolution of a single channel (monochromatic) image. A related problem, color SR, addresses fusing a set of previously demosaiced color LR frames to enhance their spatial resolution. To date, there is very little work addressing the problem of color SR. The typical solution involves applying monochromatic SR algorithms to each of the color channels independently [11], [12], while using the color information to improve the accuracy of motion estimation. Another approach is transforming the problem to a different color space, where chrominance layers are separated from luminance, and SR is applied only to the luminance channel [7]. Both of these methods are suboptimal as they do not fully exploit the correlation across the color bands.

In Section VI, we show that ignoring the relation between different color channels will result in color artifacts in the super-resolved images. Moreover, as we will advocate later in this paper, even a proper treatment of the relation between the color layers is not sufficient for removing color artifacts if the measured images are mosaiced. This brings us to the description of the demosaicing problem.

B. Demosaicing

A color image is typically represented by combining three separate monochromatic images. Ideally, each pixel reflects three data measurements; one for each of the color bands.[3] In practice, to reduce production cost, many digital cameras have only one color measurement (red, green, or blue) per pixel.[4] The detector array is a grid of CCDs, each made sensitive to one color by placing a color-filter array (CFA) in front of the CCD. The Bayer pattern shown on the left hand side of Fig. 3 is a very common example of such a color-filter. The values of the missing color bands at every pixel are often synthesized using some form of interpolation from neighboring pixel values. This process is known as color *demosaicing*.

Numerous demosaicing methods have been proposed through the years to solve this under-determined problem, and, in this section, we review some of the more popular ones. Of course, one can estimate the unknown pixel values by linear interpolation of the known ones in each color band independently. This approach will ignore some important information about the correlation between the color bands and will result in serious color artifacts. Note that the red and blue channels are down-sampled two times more than the green channel. It is reasonable to assume that the independent interpolation of the green band will result in a more reliable reconstruction than the red or blue bands. This property, combined with the assumption that the red/green and blue/green ratios are similar for the neighboring pixels, make the basics of the smooth hue transition method first discussed in [13].

Note that there is a negligible correlation between the values of neighboring pixels located on the different sides of an edge. Therefore, although the smooth hue transition assumption is logical for smooth regions of the reconstructed image, it is not successful in the high-frequency (edge) areas. Considering this

[3]This is the scenario for the more expensive 3-CCD cameras.

[4]This is the scenario for cheaper 1-CCD cameras.

fact, gradient-based methods, first addressed in [14], do not perform interpolation across the edges of an image. This noniterative method uses the second derivative of the red and blue channels to estimate the edge direction in the green channel. Later, the green channel is used to compute the missing values in the red and blue channels.

A variation of this method was later proposed in [15], where the second derivative of the green channel and the first derivative of the red (or blue) channels are used to estimate the edge direction in the green channel. The smooth hue and gradient-based methods were later combined in [16]. In this iterative method, the smooth hue interpolation is done with respect to the local gradients computed in eight directions about a pixel of interest. A second stage using anisotropic inverse diffusion will further enhance the quality of the reconstructed image. This two-step approach of interpolation followed by an enhancement step has been used in many other publications. In [17], spatial and spectral correlations among neighboring pixels are exploited to define the interpolation step, while adaptive median filtering is used as the enhancement step. A different iterative implementation of the median filters is used as the enhancement step of the method described in [18] that take advantage of a homogeneity assumption in the neighboring pixels.

Iterative MAP methods form another important category of demosaicing methods. A MAP algorithm with a smooth chrominance prior is discussed in [19]. The smooth chrominance prior is also used in [20], where the original image is transformed to YIQ representation. The chrominance interpolation is preformed using isotropic smoothing. The luminance interpolation is done using edge directions computed in a steerable wavelet pyramidal structure.

Other examples of popular demosaicing methods available in published literature are [21]–[27]. Almost all of the proposed demosaicing methods are based on one or more of these following assumptions.

1) In the constructed image with the mosaicing pattern, there are more green sensors with regular pattern of distribution than blue or red ones (in the case of Bayer CFA, there are twice as many greens than red or blue pixels and each is surrounded by four green pixels).
2) Most algorithms assume a Bayer CFA pattern, for which each red, green, and blue pixel is a neighbor to pixels of different color bands.
3) For each pixel, one, and only one, color band value is available.
4) The pattern of pixels does not change through the image.
5) The human eye is more sensitive to the details in the luminance component of the image than the details in chrominance component [20].
6) The human eye is more sensitive to chromatic changes in the low spatial frequency region than the luminance change [24].
7) Interpolation should be preformed along and not across the edges.
8) Different color bands are correlated with each other.
9) Edges should align between color channels.

Note that even the most popular and sophisticated demosaicing methods will fail to produce satisfactory results when severe aliasing is present in the color-filtered image. Such severe aliasing happens in cheap commercial still or video digital cameras, with small number of CCD pixels. The color artifacts worsen as the number of CCD pixels decreases. The following example shows this effect.

Fig. 2(a) shows a HR image captured by a 3-CCD camera. If for capturing this image, instead of a 3-CCD camera a 1-CCD camera with the same number of CCD pixels was used, the inevitable mosaicing process will result in color artifacts. Fig. 2(d) shows the result of applying demosaicing method of [16] with some negligible color-artifacts on the edges.

Note that many commercial digital video cameras can only be used in lower spatial resolution modes while working in higher frame rates. Fig. 2(b) shows a same scene from a 3-CCD camera with a down-sampling factor of 4 and Fig. 2(e) shows the demosaiced image of it after color-filtering. Note that the color artifacts in this image are much more evident than Fig. 2(d). These color artifacts may be reduced by low-pass filtering the input data before color-filtering. Fig. 2(c) shows a factor of four down-sampled version of Fig. 2(a), which is blurred with a symmetric Gaussian low-pass filter of size 4×4 with standard deviation equal to one, before down-sampling. The demosaiced image shown in Fig. 2(f) has less color artifacts than Fig. 2(e); however, it has lost some high-frequency details.

The poor quality of single-frame demosaiced images stimulates us to search for multiframe demosaicing methods, where the information of several low-quality images are fused together to produce high-quality demosaiced images.

C. Merging Super-Resolution and Demosaicing Into One Process

Referring to the mosaic effects, the geometry of the single-frame and multiframe demosaicing problems are fundamentally different, making it impossible to simply cross apply traditional demosaicing algorithms to the multiframe situation. To better understand the multiframe demosaicing problem, we offer an example for the case of translational motion. Suppose that a set of color-filtered LR images is available (images on the left in Fig. 3). We use the two-step process explained in Section IV to fuse these images. The shift-and-add image on the right side of Fig. 3 illustrates the pattern of sensor measurements in the HR image grid. In such situations, the sampling pattern is quite arbitrary depending on the relative motion of the LR images. This necessitates different demosaicing algorithms than those designed for the original Bayer pattern.

Fig. 3 shows that treating the green channel differently than the red or blue channels, as done in many single-frame demosaicing methods before, is not useful for the multiframe case. While, globally, there are more green pixels than blue or red pixels, locally, any pixel may be surrounded by only red or blue colors. So, there is no general preference for one color band over the others (the first and second assumptions in Section II-B are not true for the multiframe case).

Another assumption, the availability of one and only one color band value for each pixel, is also not correct in the multiframe case. In the under-determined cases,[5] there are

[5]Where the number of nonredundant LR frames is smaller than the square of resolution enhancement factor; a resolution enhancement factor of r means that LR images of dimension $M \times M$ produce a HR output of dimension $rM \times rM$.

Fig. 2. HR image (a) captured by a 3-CCD camera is (b) down-sampled by a factor of four. In (c), the image in (a) is blurred by a Gaussian kernel before down-sampling by a factor of 4. The images in (a)–(c) are color-filtered and then demosaiced by the method of [16]. The results are shown in (d)–(f), respectively. (a) Original. (b) Down-sampled. (c) Blurred and down-sampled. (d) Demosaiced (a). (e) Demosaiced (b). (f) Demosaiced (c).

not enough measurements to fill the HR grid. The symbol "?" in Fig. 3 represents such pixels. On the other hand, in the over-determined cases,[6] for some pixels, there may in fact be more than one color value available.

The fourth assumption in the existing demosaicing literature described earlier is not true because the field of view (FOV) of real world LR images changes from one frame to the other, so the center and the border patterns of red, green, and blue pixels differ in the resulting HR image.

III. Mathematical Model and Solution Outline

A. Mathematical Model of the Imaging System

Fig. 1 illustrates the image degradation model that we consider. We represent this approximated forward model by the following equation:

$$\begin{aligned}\underline{Y}_i(k) &= D_i(k)H(k)F(k)\underline{X}_i + \underline{V}_i(k)\\ &= T_i(k)\underline{X}_i + \underline{V}_i(k)\\ k &= 1,\ldots,N \quad i = R,G,B\end{aligned} \tag{1}$$

[6]Where the number of nonredundant LR frames is larger than the square of resolution enhancement factor.

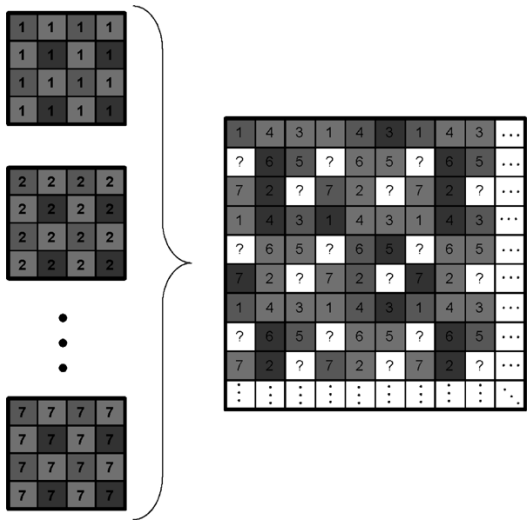


Fig. 3. Fusion of seven Bayer pattern LR images with relative translational motion (the figures in the left side of the accolade) results in a HR image ($\widehat{Z}$) that does not follow Bayer pattern (the figure in the right side of the accolade). The symbol "?" represents the HR pixel values that were undetermined (as a result of insufficient LR frames) after the shift-and-add step (the shift-and-add method is extensively discussed in [3], and briefly reviewed in Section III-F).

which can be also expressed as

$$\underline{Y} = T\underline{X} + \underline{V},$$
$$\underline{Y} = \begin{bmatrix} \underline{Y}_R(1) \\ \underline{Y}_G(1) \\ \underline{Y}_B(1) \\ \underline{Y}_R(2) \\ \vdots \\ \underline{Y}_B(N) \end{bmatrix}, \underline{V} = \begin{bmatrix} \underline{V}_R(1) \\ \underline{V}_G(1) \\ \underline{V}_B(1) \\ \underline{V}_R(2) \\ \vdots \\ \underline{V}_B(N) \end{bmatrix}$$
$$T = \begin{bmatrix} \underline{T}_R(1) \\ \underline{T}_G(1) \\ \underline{T}_B(1) \\ \underline{T}_R(2) \\ \vdots \\ \underline{T}_B(N) \end{bmatrix}, \underline{X} = \begin{bmatrix} \underline{X}_R \\ \underline{X}_G \\ \underline{X}_B \end{bmatrix}. \quad (2)$$

The vectors $\underline{X}_i$ and $\underline{Y}_i(k)$ are representing the i^{th} band (R, G, or B) of the HR color frame and the k^{th} LR frame after lexicographic ordering, respectively. Matrix $F(k)$ is the geometric motion operator between the HR and LR frames. The camera's point spread function (PSF) is modeled by the blur matrix $H(k)$. The matrix $D_i(k)$ represents the down-sampling operator, which includes both the color-filtering and CCD down-sampling operations.[7] Geometric motion, blur, and down-sampling operators are covered by the operator $T_i(k)$, which we call the system matrix. The vector $\underline{V}_i(k)$ is the system noise and N is the number of available LR frames.

The HR color image ($\underline{X}$) is of size $[12r^2M^2 \times 1]$), where r is the resolution enhancement factor. The size of the vectors $\underline{V}_G(k)$ and $\underline{Y}_G(k)$ is $[2M^2 \times 1]$ and vectors $\underline{V}_R(k)$, $\underline{Y}_R(k)$, $\underline{V}_B(k)$, and $\underline{Y}_B(k)$ are of size $[M^2 \times 1]$. The geometric motion and blur matrices are of size $[4r^2M^2 \times 4r^2M^2]$. The down-sampling and system matrices are of size $[2M^2 \times 4r^2M^2]$ for the green band and of size $[M^2 \times 4r^2M^2]$ for the red and blue bands.[8]

Considered separately, super-resolution and demosaicing models are special cases of the general model presented above. In particular, in the super-resolution literature the effect of color-filtering is usually ignored [3], [9], [10] and, therefore, the model is simplified to

$$\underline{Y}(k) = D(k)H(k)F(k)\underline{X} + \underline{V}(k) \quad k = 1, \ldots, N. \quad (3)$$

In this model, the LR images $\underline{Y}(k)$ and the HR image $\underline{X}$ are assumed to be monochromatic. On the other hand, in the demosaicing literature only single frame reconstruction of color images is considered, resulting in a simplified model

$$\underline{Y}_i = D_i\underline{X}_i + \underline{V}_i \qquad i = R, G, B. \quad (4)$$

As such, the classical approach to the multiframe reconstruction of color images has been a two-step process. The first step is to solve (4) for each image (demosaicing step) and the second step is to use the model in (3) to fuse the LR images resulting from the first step, reconstructing the color HR image (usually each R, G, or B bands is processed individually). Of course, this two-step method is a suboptimal approach to solving the overall problem. In Section IV, we propose a MAP estimation approach to directly solve (1).

B. MAP Approach to MultiFrame Image Reconstruction

Following the forward model of (1), the problem of interest is an inverse problem, wherein the source of information (HR image) is estimated from the observed data (LR images). An inherent difficulty with inverse problems is the challenge of inverting the forward model without amplifying the effect of noise in the measured data. In many real scenarios, the problem is worsened by the fact that the system matrix T is singular or ill-conditioned. Thus, for the problem of super-resolution, some form of regularization must be included in the cost function to stabilize the problem or constrain the space of solutions.

From a statistical perspective, regularization is incorporated as a priori knowledge about the solution. Thus, using the MAP estimator, a rich class of regularization functions emerges, enabling us to capture the specifics of a particular application. This can be accomplished by way of Lagrangian type penalty terms as in

$$\widehat{\underline{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[\rho(\underline{Y}, T\underline{X}) + \lambda\Gamma(\underline{X})\right] \quad (5)$$

where ρ, the data fidelity term, measures the "distance" between the model and measurements, and Γ is the regularization cost function, which imposes a penalty on the unknown X to direct it to a better formed solution. The regularization parameter λ is a scalar for properly weighting the first term (data fidelity cost) against the second term (regularization cost). Generally speaking, choosing λ could be either done manually, using visual inspection, or automatically using methods like generalized cross-validation [29], [30], L-curve [31], or other techniques. How to choose such regularization parameters is in itself a vast topic, which we will not treat in the present paper.

C. Monochromatic Spatial Regularization

Tikhonov regularization, of the form $\Gamma(\underline{X}) = \|\Lambda\underline{X}\|_2^2$, is a widely employed form of regularization [6], [9], where Λ is a matrix capturing some aspects of the image such as its general smoothness. Tikhonov regularization penalizes energy in the higher frequencies of the solution, opting for a smooth and, hence, blurry image.

To achieve reconstructed images with sharper edges, in the spirit of the total variation criterion [32], [33] and a related method called the bilateral filter[9] [34], [35], a robust regularizer called bilateral-TV (B-TV) was introduced in [3]. The B-TV regularizing function looks like

$$\Gamma(\underline{X}) = \sum_{l=-P}^{P} \sum_{m=-P}^{P} \alpha^{|m|+|l|} \left\|\underline{X} - S_x^l S_y^m \underline{X}\right\|_1 \quad (6)$$

[7]It is convenient to think of $D_i(k) = A_i(k)D(k)$, where $D(k)$ models the incoherent down-sampling effect of the CCD and $A_i(k)$ models the color-filter effect [28].

[8]Note that color super resolution by itself is a special case of this model, where vectors $\underline{V}_i(k)$ and $\underline{Y}_i(k)$ are of size $[4M^2 \times 1]$ and matrices $T_i(k)$ and $D_i(k)$ are of size $[4M^2 \times 4r^2M^2]$ for any color band.

[9]Note that by adopting a different realization of the bilateral filter, [27] has proposed a successful single frame demosaicing method.

where S_x^l and S_y^m are the operators corresponding to shifting the image represented by $\underline{X}$ by l pixels in horizontal direction and m pixels in vertical direction, respectively. This cost function in effect computes derivatives across multiple scales of resolution (as determined by the parameter P). The scalar weight $0 < \alpha < 1$ is applied to give a spatially decaying effect to the summation of the regularization term. The parameter "P" defines the size of the corresponding bilateral filter kernel. The bilateral filter and its parameters are extensively discussed in [3], [34], and [35].

The performance of B-TV and Tikhonov priors are thoroughly studied in [3]. The B-TV regularization is used in Section IV to help reconstruct the luminance component of the demosaiced images. Note that these two regularization terms in the presented form do not consider the correlation of different color bands.

D. Color Regularization

To reduce color artifacts, a few MAP-based demosaicing algorithms have adapted regularization terms for color channels. Typically, the color regularization priors are either applied on the chrominance component of an image (after transforming to a suitable color space such as YIQ representation [20]), or directly on the RGB bands [19]. While the former can be easily implemented by some isotropic smoothing priors such as Tikhonov regularization, the latter is computationally more complicated.

Note that, although different bands may have larger or smaller gradient magnitudes at a particular edge, it is reasonable to assume the same edge orientation and location for all color channels. That is to say, if an edge appears in the red band at a particular location and orientation, then an edge with the same location and orientation should appear in the other color bands. Therefore, a cost function that penalizes the difference in edge location and/or orientation of different color bands incorporates the correlation between different color bands prior. We will employ such a cost function in Section IV to remove color artifacts.

Following [19], minimizing the vector product norm of any two adjacent color pixels forces different bands to have similar edge location and orientation. The vector (outer) product of $\underline{M}$: $[m_r, m_g, m_b]^T$ and $\underline{N}$: $[n_r, n_g, n_b]^T$, which represent the color values of two adjacent pixels, is defined as

$$\begin{aligned}\|\underline{M} \times \underline{N}\|_2^2 &= [|M||N|sin(\Theta)]^2 \\ &= \left\|\vec{i}(m_g n_b - m_b n_g)\right\|_2^2 + \left\|\vec{j}(m_b n_r - m_r n_b)\right\|_2^2 \\ &\quad + \left\|\vec{k}(m_r n_g - m_g n_r)\right\|_2^2\end{aligned}$$

where Θ is the angle between these two vectors. As the data fidelity penalty term will restrict the values of $|M|$ and $|N|$, minimization of $\|M \times N\|_2^2$ will minimize $sin(\Theta)$, and, consequently, the Θ itself, where a small value of Θ is an indicator of similar orientation.

E. Data Fidelity

One of the most common cost functions to measure the closeness of the final solution to the measured data is the least-squares (LS) cost function, which minimizes the L_2 norm of the residual vector

$$\rho(\underline{Y}, T\underline{X}) = \|\underline{Y} - T\underline{X}\|_2^2 \tag{7}$$

(see [9], [10], and [36] as representative works). For the case where the noise $\underline{V}$ is additive white, zero mean Gaussian, this approach has the interpretation of providing the maximum likelihood (ML) estimate of $\underline{X}$ [9]. However, a statistical study of the noise properties found in many real image sequences used for multiframe image fusion techniques, suggests that heavy-tailed noise distributions such as Laplacian are more appropriate models (especially in the presence of the inevitable motion estimation error) [37]. In [3], an alternate data fidelity term based on the L_1 norm is recently used, which has been shown to be very robust to data outliers

$$\rho(\underline{Y}, T\underline{X}) = \|\underline{Y} - T\underline{X}\|_1. \tag{8}$$

Note that the L_1 norm is the ML estimate of data in the presence of Laplacian noise. The performance of the L_1 and L_2 norms is compared and discussed in [3]. The performance of the L_1 and L_2 norms is compared and discussed in [3]. In this paper (Section IV), we have adopted the L_1 norm (which is known to be more robust than L_2) as the data fidelity measure.

F. Speedups for the Special Case of Translation Motion and Common Space-Invariant Blur

Considering translational motion model and common[10] space-invariant PSF, the operators H and $F(k)$ are commutative ($F(k)H = HF(k)$). We can rewrite (1) as

$$\underline{Y}_i(k) = D_i(k)F(k)H\underline{X} + \underline{V}_i(k) \qquad k = 1, \ldots, N \qquad i = R, G, B. \tag{9}$$

By substituting $\underline{Z} = H\underline{X}$, the inverse problem may be separated into the much simpler subtasks of:

1) fusing the available images and estimating a blurred HR image from the LR measurements (we call this result $\widehat{\underline{Z}}$);
2) estimating the deblurred image $\widehat{\underline{X}}$ from $\widehat{\underline{Z}}$.

The optimality of this method is extensively discussed in [3], where it is shown that $\widehat{\underline{Z}}$ is the weighted mean (mean or median operators, for the cases of L_2 norm and L_1 norm, respectively) of all measurements at a given pixel, after proper zero filling and motion compensation. We call this operation shift-and-add, which greatly speeds up the task of multiframe image fusion under the assumptions made. To compute the shift-and-add image, first the relative motion between all LR frames is computed. Then, a set of HR images is constructed by up-sampling each LR frame by zero filling. Then, these HR frames are registered with respect to the relative motion of the corresponding LR frames. A pixel-wise mean or median operation on the nonzero values of these HR frames will result in the shift-and-add image.

[10] $\forall k\ H(k) = H$, which is true when all images are acquired with the same camera.

In the next section, we use the penalty terms described in this section to formulate our proposed method of multiframe demosaicing and color super-resolution.

IV. Multiframe Demosaicing

In Section II-C we indicated how the multiframe demosaicing is fundamentally different than single-frame demosaicing. In this section, we propose a computationally efficient MAP estimation method to fuse *and* demosaic a set of LR frames (which may have been color-filtered by any CFA) resulting in a color image with higher spatial resolution and reduced color artifacts. Our MAP-based cost function consists of the following terms, briefly motivated in the previous section:

1) a penalty term to enforce similarities between the raw data and the HR estimate (data fidelity penalty term);
2) a penalty term to encourage sharp edges in the luminance component of the HR image (spatial luminance penalty term);
3) a penalty term to encourage smoothness in the chrominance component of the HR image (spatial chrominance penalty term);
4) a penalty term to encourage homogeneity of the edge location and orientation in different color bands (intercolor dependencies penalty term).

Each of these penalty terms will be discussed in more detail in the following sections.

A. Data Fidelity Penalty Term

This term measures the similarity between the resulting HR image and the original LR images. As it is explained in Section III-E and [3], L_1 norm minimization of the error term results in robust reconstruction of the HR image in the presence of uncertainties such as motion error. Considering the general motion and blur model of (1), the data fidelity penalty term is defined as

$$J_0(\underline{X}) = \sum_{i=R,G,B} \sum_{k=1}^{N} \|D_i(k)H(k)F(k)\underline{X}_i - \underline{Y}_i(k)\|_1 . \tag{10}$$

Note that the above penalty function is applicable for general models of data, blur, and motion. However, in this paper, we only treat the simpler case of common space invariant PSF and translational motion. This could, for example, correspond to a vibrating camera acquiring a sequence of images from a static scene.

For this purpose, we use the two-step method of Section III-F to represent the data fidelity penalty term, which is easier to interpret and has a faster implementation potential [3]. This simplified data fidelity penalty term is defined as

$$J_0(\underline{X}) = \sum_{i=R,G,B} \left\| \Phi_i(H\underline{\widehat{X}}_i - \underline{\widehat{Z}}_i) \right\|_1 \tag{11}$$

where $\underline{\widehat{Z}}_R$, $\underline{\widehat{Z}}_G$, and $\underline{\widehat{Z}}_B$ are the three color channels of the color shift-and-add image $\underline{\widehat{Z}}$. The matrix Φ_i $(i = R, G, B)$, is a diagonal matrix with diagonal values equal to the square root of the number of measurements that contributed to make each element of $\underline{\widehat{Z}}_i$ (in the square case is the identity matrix). So, the undefined pixels of $\underline{\widehat{Z}}_B$ have no effect on the HR estimate. On the other hand, those pixels of $\underline{\widehat{Z}}_B$ which have been produced from numerous measurements, have a stronger effect in the estimation of the HR frame. The $\Phi_{i\in\{R,G,B\}}$ matrices for the multiframe demosaicing problem are sparser than the corresponding matrices in the color SR case. The vectors $\underline{\widehat{X}}_R$, $\underline{\widehat{X}}_G$, and $\underline{\widehat{X}}_B$ are the three color components of the reconstructed HR image $\widehat{X}$.

B. Spatial Luminance Penalty Term

The human eye is more sensitive to the details in the luminance component of an image than the details in the chrominance components [20]. Therefore, it is important that the edges in the luminance component of the reconstructed HR image look sharp. As explained in Section III-C, applying B-TV regularization to the luminance component will result in this desired property [3]. The luminance image can be calculated as the weighted sum $\underline{X}_L = 0.299\underline{X}_R + 0.597\underline{X}_G + 0.114\underline{X}_B$ as explained in [38]. The luminance regularization term is then defined as

$$J_1(\underline{X}) = \sum_{l=-P}^{P} \sum_{m=-P}^{P} \alpha^{|m|+|l|} \left\| \underline{X}_L - S_x^l S_y^m \underline{X}_L \right\|_1 . \tag{12}$$

C. Spatial Chrominance Penalty Term

Spatial regularization is required also for the chrominance layers. However, since the HVS is less sensitive to the resolution of these bands, we can use a simpler regularization, based on the L_2 norm [3]

$$J_2(\underline{X}) = \|\Lambda\underline{X}_{C1}\|_2^2 + \|\Lambda\underline{X}_{C2}\|_2^2 \tag{13}$$

where the images $\underline{X}_{C1}$ and $\underline{X}_{C2}$ are the I and Q layers in the YIQ color representation.[11]

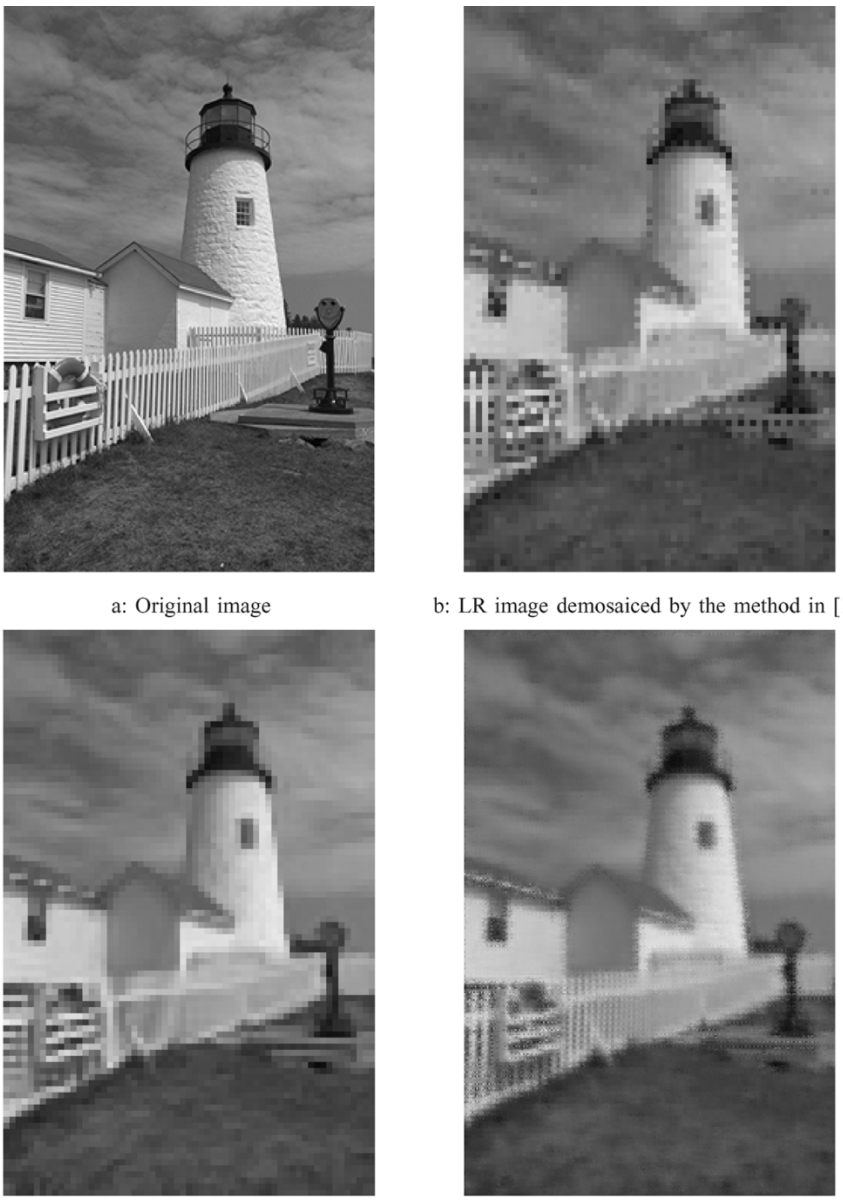
D. Intercolor Dependencies Penalty Term

This term penalizes the mismatch between locations or orientations of edges across the color bands. As described in Section III-D, the authors of [19] suggest a pixelwise intercolor dependencies cost function to be minimized. This term has the vector outer product norm of all pairs of neighboring pixels, which is solved by the finite element method. With some modifications to what was proposed in [19], our intercolor dependencies penalty term is a differentiable cost function

$$\begin{aligned} J_3(\underline{X}) &= \sum_{l=-1}^{1} \sum_{m=-1}^{1} \Big[\left\| \underline{X}_G \odot S_x^l S_y^m \underline{X}_B - \underline{X}_B \odot S_x^l S_y^m \underline{X}_G \right\|_2^2 \\ &\quad + \left\| \underline{X}_B \odot S_x^l S_y^m \underline{X}_R - \underline{X}_R \odot S_x^l S_y^m \underline{X}_B \right\|_2^2 \\ &\quad + \left\| \underline{X}_R \odot S_x^l S_y^m \underline{X}_G - \underline{X}_G \odot S_x^l S_y^m \underline{X}_R \right\|_2^2 \Big] \end{aligned} \tag{14}$$

where $\odot$ is the element by element multiplication operator.

[11]The Y layer ($\underline{X}_L$) is treated in (12).



a: Original image

b: LR image demosaiced by the method in [14]

c: LR image demosaiced by the method in [16]

d: Shift-And-Add image.

Fig. 4. HR image (a) is passed through our model of camera to produce a set of LR images. (b) One of these LR images is demosaiced by the method in [14]. (c) The same image is demosaiced by the method in [16]. Shift-and-add on the ten input LR images is shown in (d).

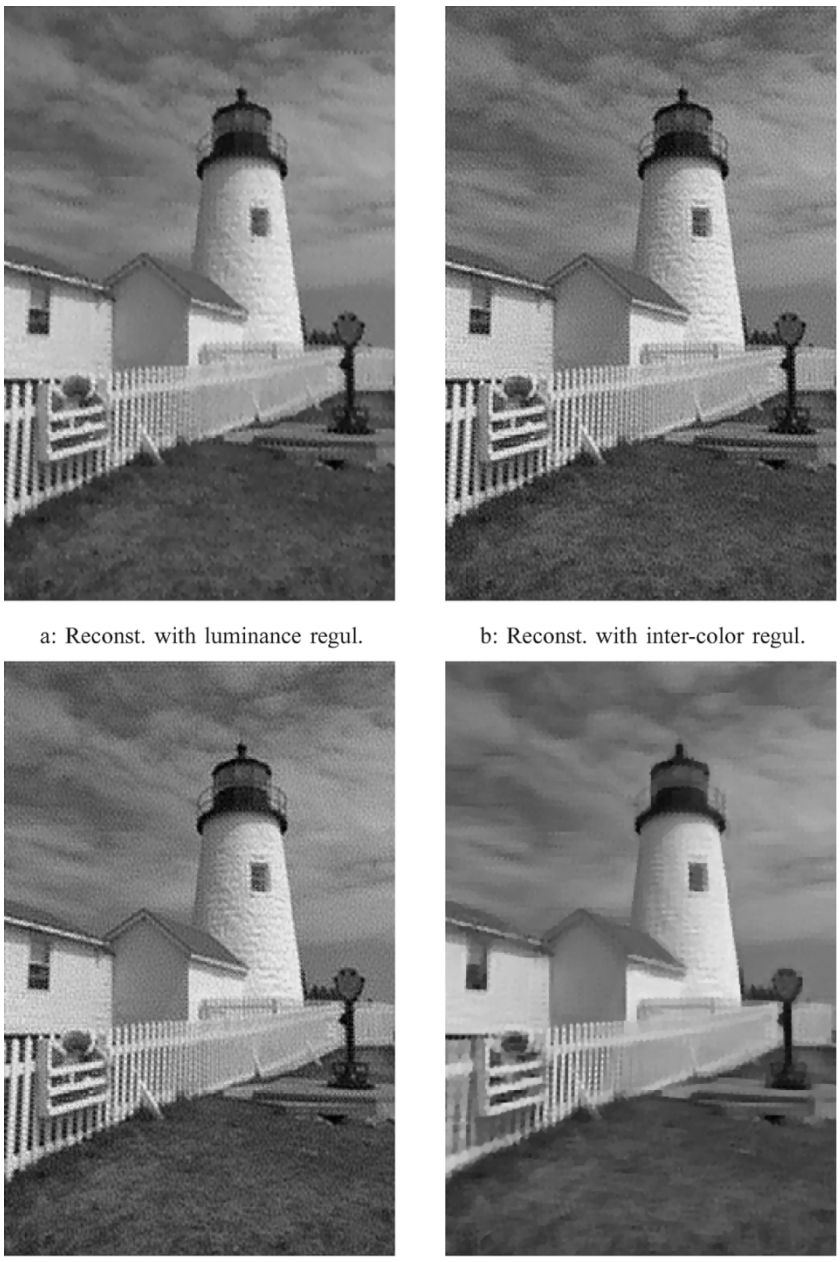
E. Overall Cost Function

The overall cost function is the summation of the cost functions described in the previous sections

$$\underline{\widehat{X}} = \underset{\underline{X}}{\mathrm{ArgMin}} \left[J_0(\underline{X}) + \lambda' J_1(\underline{X}) + \lambda'' J_2(\underline{X}) + \lambda''' J_3(\underline{X}) \right]. \tag{15}$$

Steepest descent optimization may be applied to minimize this cost function. In the first step, the derivative of (15) with respect to one of the color bands is calculated, assuming the other two color bands are fixed. In the next steps, the derivative will be computed with respect to the other color channels. For example, the derivative with respect to the green band ($\underline{X}_G$) is calculated as in (16), shown at the bottom of the next page, where S_x^{-l} and S_y^{-m} define the transposes of matrices S_x^l and S_y^m, respectively, and have a shifting effect in the opposite directions of S_x^l and S_y^m. The notation $\mathbf{X_R}$ and $\mathbf{X_B}$ stands for the diagonal matrix representations of the red and blue bands and $\mathbf{X_R^{l,m}}$ and $\mathbf{X_B^{l,m}}$ are the diagonal representations of these matrices shifted by l and m pixels in the horizontal and vertical directions, respectively. The calculation of the intercolor dependencies term derivative is explained in the Appendix I.

Matrices H, Λ, Φ, D, S_x^l, and S_y^m, and their transposes can be exactly interpreted as direct image operators such as blur, high-pass filtering, masking, down-sampling, and shift. Noting and implementing the effects of these matrices as a sequence of operators on the images directly spares us from explicitly constructing them as matrices. This property helps



a: Reconst. with luminance regul. b: Reconst. with inter-color regul.

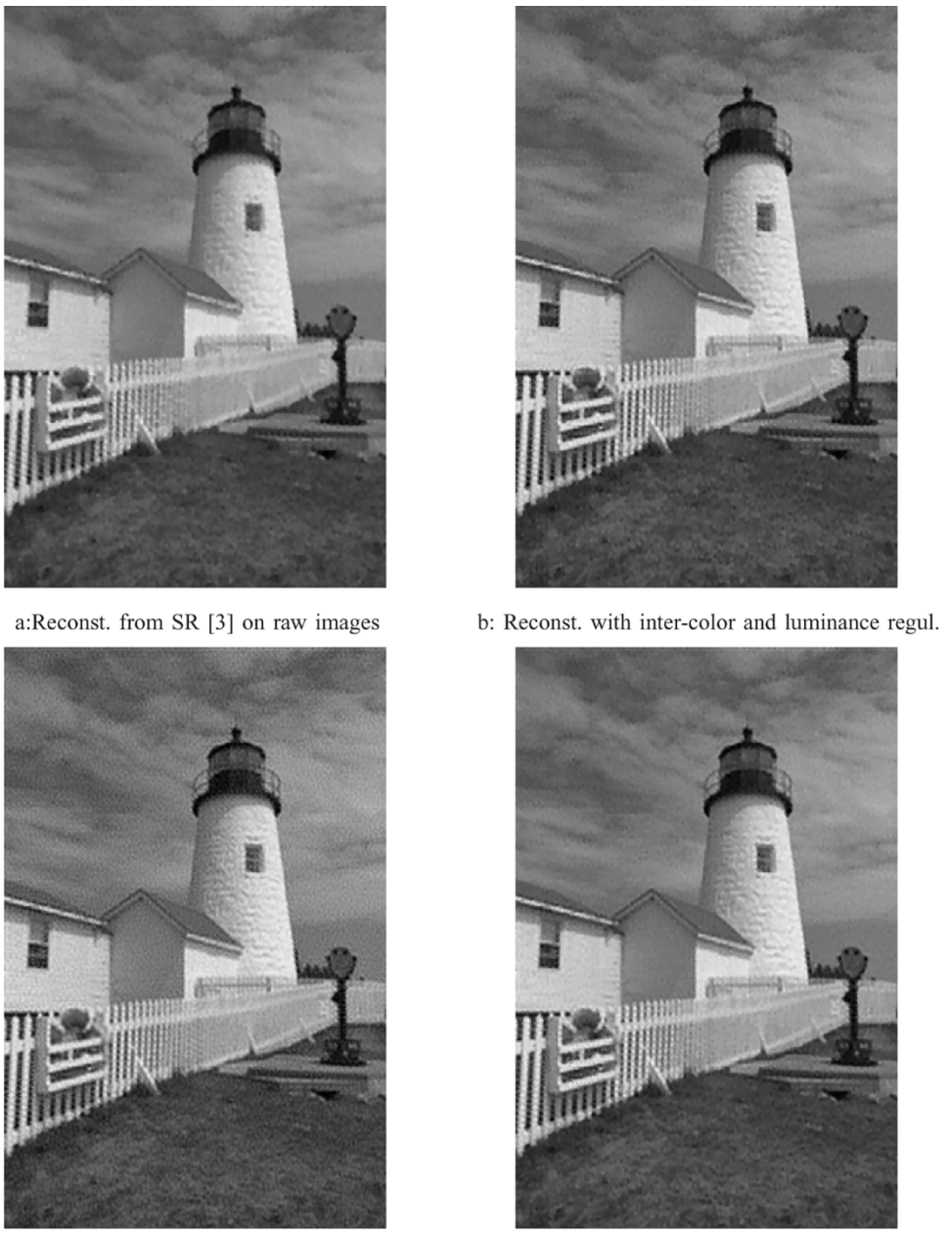
c: Reconst. with chrominance regul. d: Reconst. from LR demosaiced [16]+SR[3]

Fig. 5. Multiframe demosaicing of this set of LR frames with the help of only luminance, intercolor dependencies, or chrominance regularization terms is shown in (a)–(c), respectively. The result of applying the super-resolution method of [3] on the LR frames each demosaiced by the method [16] is shown in (d).

our method to be implemented in a fast and memory efficient way.

The gradient of the other channels will be computed in the same way, and the following steepest (coordinate) descent iter-

$$
\begin{aligned}
\nabla \widehat{\underline{X}}_G^n = {} & H^T \Phi_G^T sign\left(\Phi_G H \widehat{\underline{X}}_G^n - \Phi_G \widehat{\underline{Z}}_G\right) + \lambda' \sum_{l=-P}^{P} \sum_{m=-P}^{P} \alpha^{|m|+|l|} \times 0.5870 \times \left[I - S_y^{-m} S_x^{-l}\right] \\
& \times sign\left(0.2989\left(\underline{X}_R^n - S_x^l S_y^m \underline{X}_R^n\right) + 0.5870\left(\underline{X}_G^n - S_x^l S_y^m \underline{X}_G^n\right) + 0.1140\left(\underline{X}_B^n - S_x^l S_y^m \underline{X}_B^n\right)\right) \\
& + \lambda'' \sum_{l=-1}^{1} \sum_{m=-1}^{1} \left[2\left(\mathbf{X}_{\mathbf{B}}^{\mathbf{l,m}} - S_x^{-l} S_y^{-m} \mathbf{X}_{\mathbf{B}}\right)\left(\mathbf{X}_{\mathbf{B}}^{\mathbf{l,m}} \underline{X}_G - \mathbf{X}_{\mathbf{B}} S_x^l S_y^m \underline{X}_G\right)\right. \\
& \qquad \left. + 2\left(\mathbf{X}_{\mathbf{R}}^{\mathbf{l,m}} - S_x^{-l} S_y^{-m} \mathbf{X}_{\mathbf{R}}\right)\left(\mathbf{X}_{\mathbf{R}}^{\mathbf{l,m}} \underline{X}_G - \mathbf{X}_{\mathbf{R}} S_x^l S_y^m \underline{X}_G\right)\right] \\
& + \lambda''' \Lambda^T \Lambda(-0.1536 \times \underline{X}_R + 0.2851 \times \underline{X}_G - 0.1316 \times \underline{X}_B)
\end{aligned} \tag{16}
$$



a:Reconst. from SR [3] on raw images

b: Reconst. with inter-color and luminance regul.

c: Reconst. with chrominance and inter-color regul.

d: Reconst. from chrominance and luminance regul.

Fig. 6. Result of super-resolving each color band (raw data before demosaicing) separately considering only bilateral regularization [3] is shown in (a). Multiframe demosaicing of this set of LR frames with the help of only intercolor dependencies-luminance, intercolor dependencies-chrominance, and luminance-chrominance regularization terms is shown in (b)–(d), respectively.

ations will be set up to calculate the HR image estimate iteratively

$$\underline{\widehat{X}}_i^{n+1} = \underline{\widehat{X}}_i^{n} - \beta \nabla \underline{\widehat{X}}_i^{n} \qquad i = R, G, B \tag{17}$$

where the scalar β is the step size.

V. Related Methods

As mentioned earlier, there has been very little work on the problem we have posed here. One related paper is the work of Zomet and Peleg [39], who have recently proposed a novel method for combining the information from multiple sensors, which can also be used for demosaicing purposes. Although their method has produced successful results for the single frame demosaicing problem, it is not specifically posed or directed toward solving the multiframe demosaicing problem, and no multiframe demosaicing case experiment is given.

The method of [39] is based on the assumption of affine relation between the *intensities* of different sensors in a local neighborhood. To estimate the red channel, first, affine relations that project green and blue channels to the red channel are computed. In the second stage, a super-resolution algorithm (e.g., the method of [7]) is applied on the available LR images in the red channel (i.e., the original CFA data of the red channel plus the projected green and blue channels) to estimate the HR red channel image. A similar procedure estimates the HR green and

TABLE I
QUANTITATIVE COMPARISON OF THE PERFORMANCE OF DIFFERENT DEMOSAICING METHODS ON THE LIGHTHOUSE SEQUENCE. THE PROPOSED METHOD HAS THE LOWEST S-CIELAB ERROR AND THE HIGHEST PSNR VALUE

	Shift-And-Add	LR Demosaiced [16] +SR [3]	Only Lumin.	Only Orient.	Only Chromin.
S-CIELAB	1.532×10^{11}	1.349×10^{11}	7.892×10^{10}	6.498×10^{10}	4.648×10^{10}
PSNR (dB)	17.17	19.12	17.74	20.10	20.35
	SR [3] on Raw Data	Lumin.+Orient.	Orient.+Chrom.	Lumin.+Chrom.	Full
S-CIELAB	5.456×10^{10}	4.543×10^{10}	4.382×10^{10}	3.548×10^{10}	3.365×10^{10}
PSNR (dB)	19.28	20.79	20.68	21.12	21.13

blue channel images. As affine model is not always valid for all sensors or image sets, so an affine model validity test is utilized in [39]. In the case that the affine model is not valid for some pixels, those projected pixels are simply ignored.

The method of [39] is highly dependent on the validity of the affine model, which is not confirmed for the multiframe case with inaccurate registration artifacts. Besides, the original CFA LR image of a channel and the less reliable projected LR images of other channels are equally weighted to construct the missing values, and this does not appear to be an optimal solution.

In contrast to their method, our proposed technique exploits the correlation of the information in different channels explicitly to guarantee similar edge position and orientation in different color bands. Our proposed method also exploits the difference in sensitivity of the human eye to the frequency content and outliers in the luminance and chrominance components of the image.

In parallel to our work, Gotoh and Okotumi [40] are proposing another MAP estimation method for solving the same joint demosaicing/super-resolution problem. While their algorithm and ours share much in common, there are fundamental differences between our algorithm and theirs in the robustness to model errors, and prior used. Model errors, such as choice of blur or motion estimation errors, are treated favorably by our algorithm due to the L_1 norm employed in the likelihood fidelity term. By contrast, in [40], an L_2-norm data fusion term is used, which is not robust to such errors. In [3], it is shown how this difference in norm can become crucial in obtaining better results in the presence of model mismatches.

As to the choice of prior, ours is built of several pieces, giving an overall edge preserved outcome, smoothed chrominance layers, and forced edge and orientation alignment between color layers. On the contrary, [40] utilizes an unisotropic Tikhonov (L_2 norm) method of regularizing.

VI. EXPERIMENTS

Experiments on synthetic and real data sets are presented in this section. In the first experiment, following the model of (1), we created a sequence of LR frames from an original HR image [Fig. 4(a)], which is a color image with full RGB values. First we shifted this HR image by one pixel in the vertical direction. Then to simulate the effect of camera PSF, each color band of this shifted image was convolved with a symmetric Gaussian low-pass filter of size 5×5 with standard deviation equal to one. The resulting image was subsampled by the factor of 4 in each direction. The same process with different motion vectors (shifts) in vertical and horizontal directions was used to produce 10 LR images from the original scene. The horizontal shift between the low resolution images was varied between 0 to .75 pixels in the LR grid (0 to 3 pixels in the HR grid). The vertical shift between the low resolution images varied between 0 to .5 pixels in the LR grid (0 to 2 pixels in the HR grid). To simulate the errors in motion estimation, a bias equal to half a pixel shift in the LR grid was intentionally added to the known motion vector of one of the LR frames. We added Gaussian noise to the resulting LR frames to achieve a signal-to-noise ratio (SNR) equal[12] to 30 dB. Then each LR color image was subsampled by the Bayer filter.

a: Reconst. with all terms.

Fig. 7. Result of applying the proposed method (using all regularization terms) to this data set is shown in (a).

In order to show what one of those measured images looks like, one of these Bayer filtered LR images is reconstructed by the method in [14] and shown in Fig. 4(b). The above method is implemented on Kodak DCS-200 digital cameras [41], so each LR image may be thought of as one picture taken with this

[12]SNR is defined as $10\log_{10}(\sigma^2/\sigma_n^2)$, where σ^2, σ_n^2 are variance of a clean frame and noise, respectively.

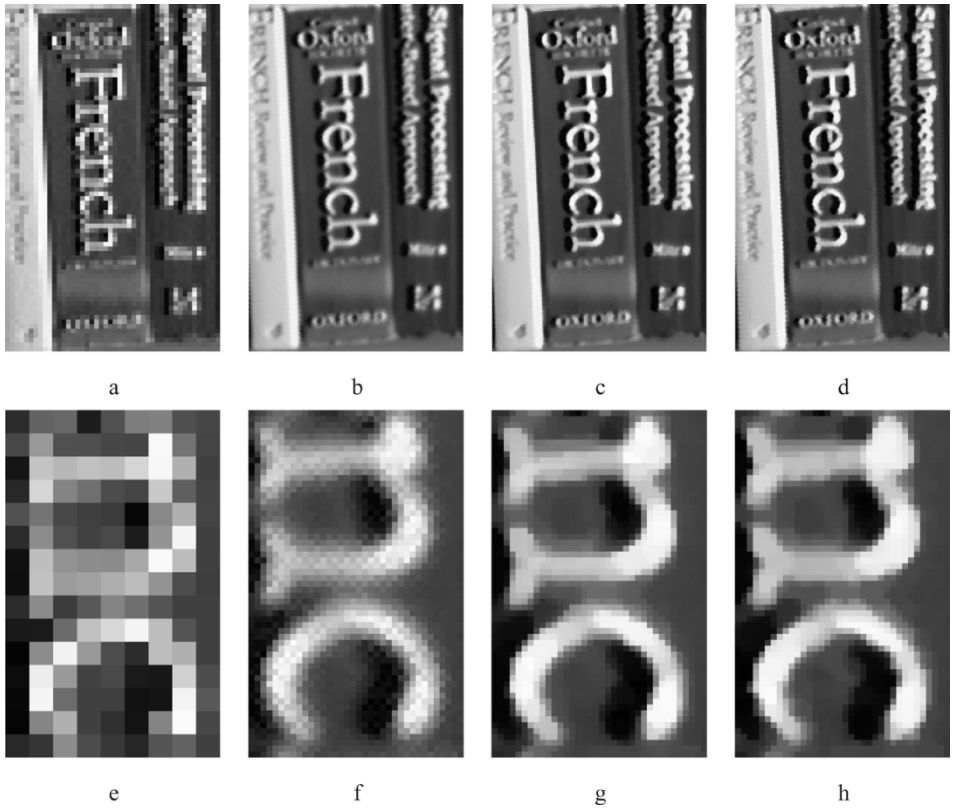


Fig. 8. Multiframe color super-resolution implemented on a real data sequence. (a) One of the input LR images and (b) shift-and-add result increasing resolution by a factor of 4 in each direction. (c) Result of the individual implementation of the super-resolution [3] on each color band. (d) Implementation of (15), which has increased the spatial resolution, removed the compression artifacts, and also reduced the color artifacts. (e)–(h) are the zoomed images of the (a)–(d), respectively.

camera brand. Fig. 4(c). shows the result of using the more sophisticated demosaicing method[13] of [16].

As the motion model for this experiment is translational and the blur kernel is space invariant, we can use the fast model of (16) to reconstruct the blurry image $(\widehat{Z})$ on the HR grid. The shift-and-add result of the demosaiced LR frames after bilinear interpolation,[14] before deblurring and demosaicing is shown in Fig. 4(d). We used the result of the shift-and-add method as the initialization of the iterative multiframe demosaicing methods. We used the original set of frames (raw data) to reconstruct a HR image with reduced color artifacts. Fig. 5(a)–(c) shows the effect of the individual implementation of each regularization term (luminance, chrominance, and intercolor dependencies), described in Section IV.

We applied the method of [16] to demosaic each of these ten LR frames individually, and then applied the robust super-resolution method of [3] on each resulting color channel. The result of this method is shown in Fig. 5(d). We also applied the robust super-resolution method of [3] on the raw (Bayer filtered) data (before demosaicing).[15] The result of this method is shown in Fig. 6(a). To study the effectiveness of each regularization term, we paired (intercolor dependencies-luminance, intercolor dependencies-chrominance, and luminance-chrominance) regularization terms for which the results are shown in Fig. 6(b)–(d), respectively. Finally, Fig. 7(a) shows the result of the implementation of (15) with all terms. The parameters used for this example are as follows:[16] $\beta = 0.002$, $\alpha = 0.9$, $\lambda' = 0.01$, $\lambda'' = 150$, $\lambda''' = 1$.

It is clear that the resulting image [Fig. 7(a)] has a better quality than the LR input frames or other reconstruction methods. Quantitative measurements confirm this visual comparison. We used PSNR[17] and S-CIELAB[18] measures to compare the performance of each of these methods. Table I compares these values in which the proposed method has the lowest S-CIELAB error and the highest PSNR values (and also the best visual quality, especially in the red lifesaver section of the image).

[13]We thank Prof. R. Kimmel of the Technion for providing us with the code that implements the method in [16].

[14]Interpolation is needed as this experiment is an under-determined problem, where some pixel values are missing.

[15]To apply the monochromatic SR method of [3] on this color-filtered sequence, we treated each color band separately. To consider the color-filtering operation, we substituted matrix A in [3, (23)] with matrix Φ in (11).

[16]The criteria for parameter selection in this example (and other examples discussed in this paper) was to choose parameters which produce visually most appealing results. Therefore, to ensure fairness, each experiment was repeated several times with different parameters and the best result of each experiment was chosen as the outcome of each method.

[17]The PSNR of two vectors $\underline{X}$ and $\underline{\widehat{X}}$ of size $[4r^2M^2 \times 1]$ is defined as

$$\mathrm{PSNR}(\underline{X}, \underline{\widehat{X}}) = 10 \log_{10} \left(\frac{255^2 \times 4r^2M^2}{\|\underline{X} - \underline{\widehat{X}}\|_2^2} \right).$$

[18]The S-CIELAB measure is a perceptual color fidelity measure that measures how accurate the reproduction of a color is to the original when viewed by a human observer [42]. In our experiments, we used the code with default parameters used in the implementation of this measure available at http://white.stanford.edu/~brian/scielab/scielab.html.

a: LR b: Shift-And-Add

c: SR [3] on LR frames d: Proposed method

Fig. 9. Multiframe color super-resolution implemented on a real data sequence. (a) One of the input LR images and (b) the shift-and-add result increasing resolution by a factor of 4 in each direction. (c) Result of the individual implementation of the super-resolution [3] on each color band. (d) Implementation of (15) which has increased the spatial resolution, removed the compression artifacts, and also reduced the color artifacts. These images are zoomed in Fig. 10.

In the second experiment, we used 30 compressed images captured from a commercial webcam (PYRO-1394). Fig. 8(a) shows one of these LR images [a selected region of this image is zoomed in Fig. 8(e) for closer examination]. Note that the compression and color artifacts are quite apparent in these images. This set of frames was already demosaiced, and no information was available about the original sensor values, which makes the color enhancement task more difficult. This example may be also considered as a multiframe color super-resolution case. The (unknown) camera PSF was assumed to be a 4×4 Gaussian kernel with standard deviation equal to one. As the relative motion between these images followed the translational model, we only needed to estimate the motion between the luminance components of these images [43]. We used the method described in [44] to computed the motion vectors.

The shift-and-add result (resolution enhancement factor of 4) is shown in Fig. 8(b) [zoomed in Fig. 8(f)]. In Fig. 8(c) [zoomed in Fig. 8(g)], the method of [3] is used for increasing the resolution by a factor of 4 in each color band, independently, and, finally, the result of applying our method on this sequence is shown in Fig. 8(d) [zoomed in Fig. 8(h)], where color artifacts are significantly reduced. The parameters used for this example are as follows: $\beta = 0.004$, $\alpha = 0.9$, $\lambda' = 0.25$, $\lambda'' = 500$, $\lambda''' = 5$.

In the third experiment, we used 40 compressed images of a test pattern from a surveillance camera, courtesy of Adyoron Intelligent Systems Ltd., Tel Aviv, Israel. Fig. 9(a) shows one of these LR images [a selected region of this image is zoomed in Fig. 10(a) for closer examination]. Note that the compression and color artifacts are quite apparent in these images. This set of frames was also already demosaiced, and no information was available about the original sensor values, which makes the color enhancement task more difficult. This example may be also considered as a multiframe color super-resolution case. The (unknown) camera PSF was assumed to be a 6×6 Gaussian kernel with standard deviation equal to two.

We used the method described in [44] to compute the motion vectors. The shift-and-add result (resolution enhancement factor of 4) is shown in Fig. 9(b) [zoomed in Fig. 10(b)]. In Fig. 9(c) [zoomed in Fig. 10(c)], the method of [3] is used for increasing the resolution by a factor of 4 in each color band, independently, and, finally, the result of applying the proposed method on this sequence is shown in Fig. 9(d) [zoomed in Fig. 10(d)], where color artifacts are significantly reduced. Moreover, compared to Fig. 9(a)–(d), the compression errors have been removed more effectively in Fig. 9(d). The parameters used for this example are as follows: $\beta = 0.004$, $\alpha = 0.9$, $\lambda' = 0.25$, $\lambda'' = 500$, $\lambda''' = 5$.

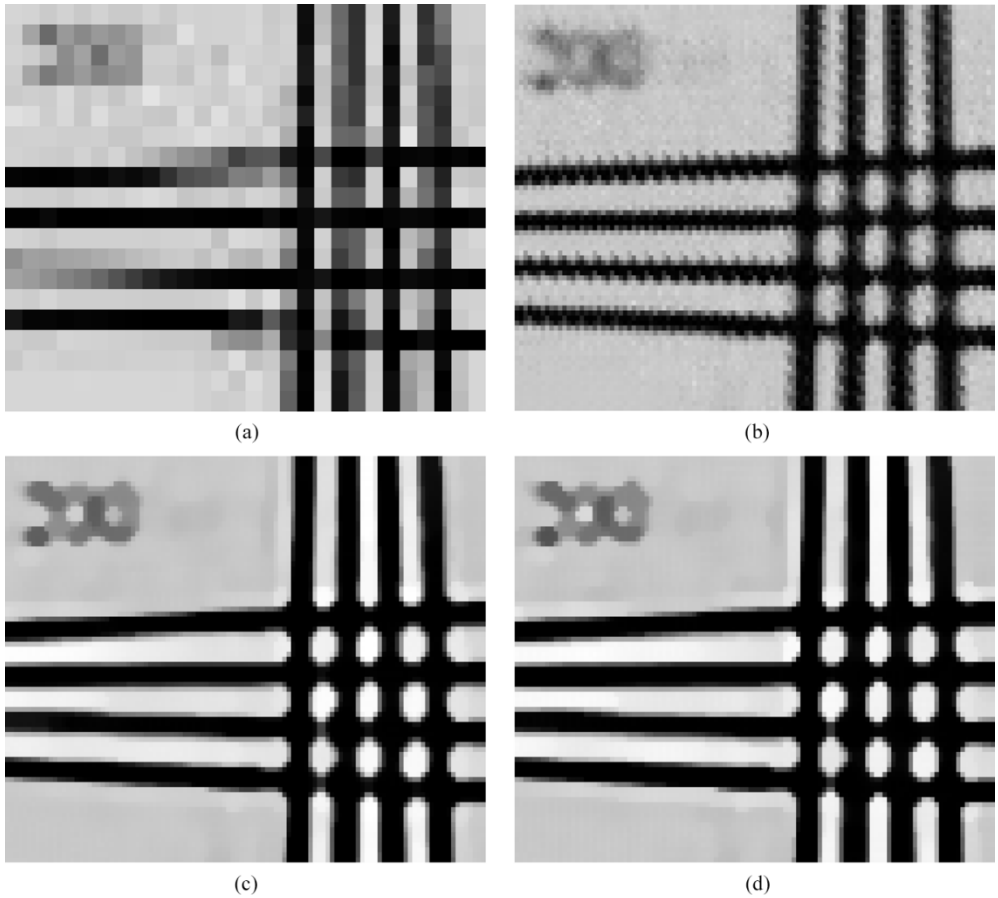


Fig. 10. Multiframe color super-resolution implemented on a real data sequence. A selected section of Fig. 9(a)–(d) are zoomed in (a)–(d), respectively. In (d), almost all color artifacts that are present on the edge areas of (a)–(c) are effectively removed. (a) LR. (b) Shift-and-add. (c) SR [3] on LR frames. (d) Proposed method.

In the fourth, fifth, and sixth experiments (girl, bookcase, and window sequences), we used 31 uncompressed, raw CFA images (30 frames for the window sequence) from a video camera (based on Zoran 2MP CMOS sensors). We applied the method of [14] to demosaic each of these LR frames, individually. Fig. 11(a) [zoomed in Fig. 12(a)] shows one of these images from the girl sequence [corresponding image of the bookcase sequence is shown in Fig. 13(a), and the corresponding image of the window sequence is shown in Fig. 15(a)]. The result of the more sophisticated demosaicing method of [16] for girl sequence is shown in Fig. 11(b) [zoomed in Fig. 12(b)]. Fig. 13(b) shows the corresponding image for the bookcase sequence, and Fig. 15(b) shows the corresponding image for the window sequence.

To increase the spatial resolution by a factor of three, we applied the proposed multiframe color super-resolution method on the demosaiced images of these two sequences. Fig. 11(c) shows the HR color super-resolution result from the LR color images of girl sequence demosaiced by the method of [14] [zoomed in Fig. 12(c)]. Fig. 13(c) shows the corresponding image for the bookcase sequence, and Fig. 15(c) shows the corresponding image for the window sequence. Similarly, Fig. 11(d) shows the result of resolution enhancement of the LR color images from girl sequence demosaiced by the method of [16] [zoomed in Fig. 12(d)]. Fig. 13(d) shows the corresponding image for the bookcase sequence, and Fig. 15(d) shows the corresponding image for the window sequence.

Finally, we directly applied the proposed multiframe demosaicing method on the raw CFA data to increase the spatial resolution by the same factor of three. Fig. 11(e) shows the HR result of multiframe demosaicing of the LR raw CFA images from the girl sequence without using the inter color dependence term $[J_3(\underline{X})]$ [zoomed in Fig. 12(e)]. Fig. 14(a) shows the corresponding image for the bookcase sequence, and Fig. 15(e) shows the corresponding image for the window sequence. Fig. 11(f) shows the HR result of applying the multiframe demosaicing method using all proposed terms in (15) on the LR raw CFA images from the girl sequence [zoomed in Fig. 12(f)]. Fig. 14(b) shows the corresponding image for the bookcase sequence and Fig. 15(f) shows the corresponding image for the window sequence.

These experiments show that single frame demosaicing methods such as [16] (which in effect implement anti-aliasing filters) remove color artifacts at the expense of making the images more blurry. The proposed color super-resolution algorithm can retrieve some high frequency information and further remove the color artifacts. However, applying the proposed multiframe demosaicing method directly on raw CFA data produces the sharpest results and effectively removes color artifacts. These experiments also show the importance of the in-

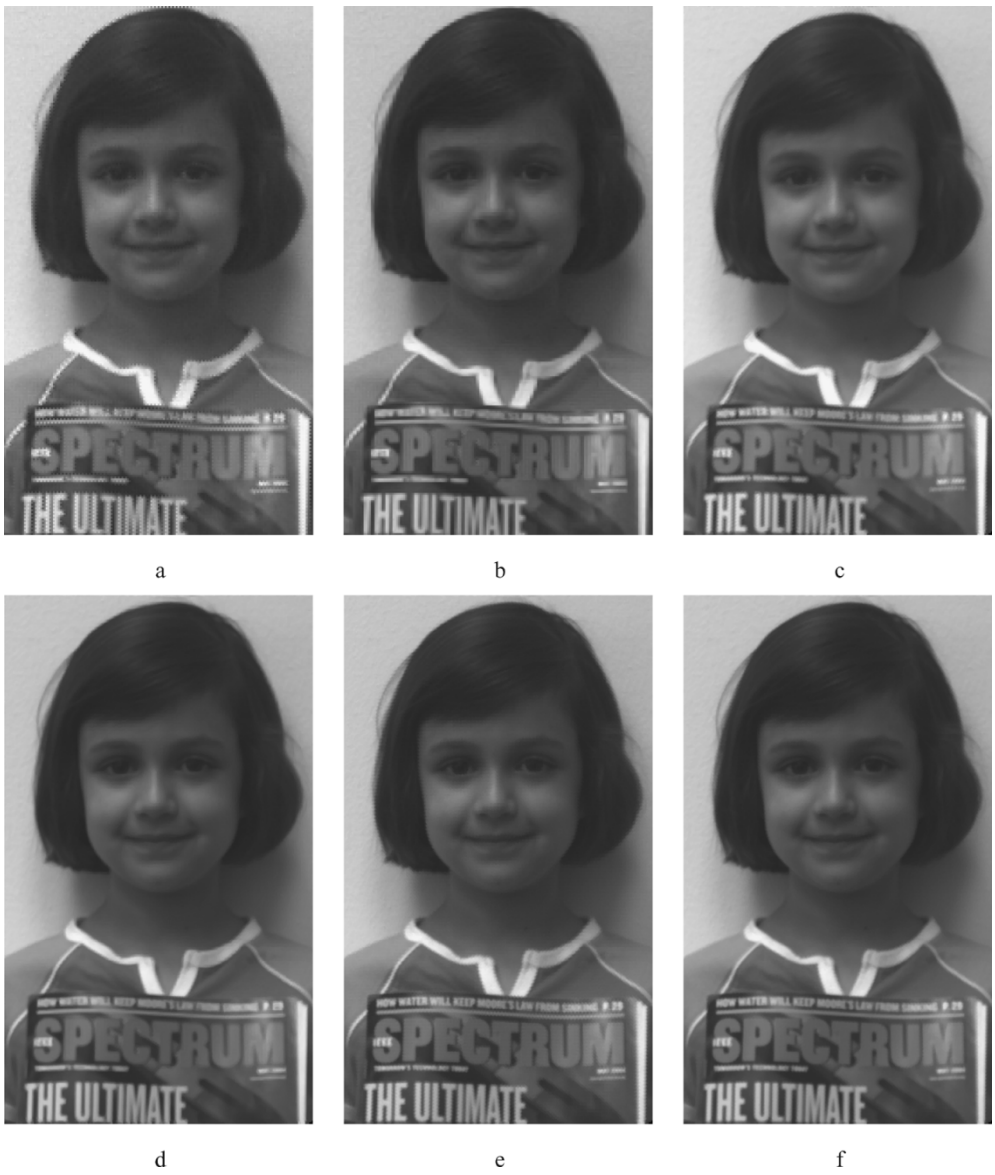


Fig. 11. Multiframe color super-resolution implemented on a real data sequence. (a) One of the input LR images demosaiced by [14] and (b) one of the input LR images demosaiced by the more sophisticated [16]. (c) Result of applying the proposed color-super-resolution method on 31 LR images each demosaiced by [14] method. (d) Result of applying the proposed color-super-resolution method on 31 LR images each demosaiced by [16] method. The result of applying our method on the original un-demosaiced raw LR images (without using the inter color dependence term) is shown in (e). (f) Result of applying our method on the original un-demosaiced raw LR images.

tercolor dependence term which further removes color artifacts. The parameters used for the experiments on girl, bookcase, and window sequences are as follows: $\beta = 0.002$, $\alpha = 0.9$, $\lambda' = 0.1$, $\lambda'' = 250$, $\lambda''' = 25$. The (unknown) camera PSF was assumed to be a tapered 5×5 disk PSF.[19]

VII. Discussion and Future Work

In this paper, based on the MAP estimation framework, we proposed a unified method of demosaicing and super-resolution, which increases the spatial resolution and reduces the color artifacts of a set of low-quality color images. Using the L_1 norm for the data error term makes our method robust to errors in data and modeling. Bilateral regularization of the luminance term results in sharp reconstruction of edges, and the chrominance and intercolor dependencies cost functions remove the color artifacts from the HR estimate. All matrix-vector operations in the proposed method are implemented as simple image operators. As these operations are locally performed on pixel values on the HR grid, parallel processing may also be used to further increase the computational efficiency. The computational complexity of this method is on the order of the computational complexity of the popular iterative super-resolution algorithms, such as [9]. Namely, it is linear in the number of pixels.

The intercolor dependencies term (14) results in the nonconvexity of the overall penalty function. Therefore, the steepest decent optimization of (15) may reach a local rather than the global minimum of the overall function. The nonconvexity does

[19]MATLAB command fspecial('disk', 2) creates such blurring kernel.

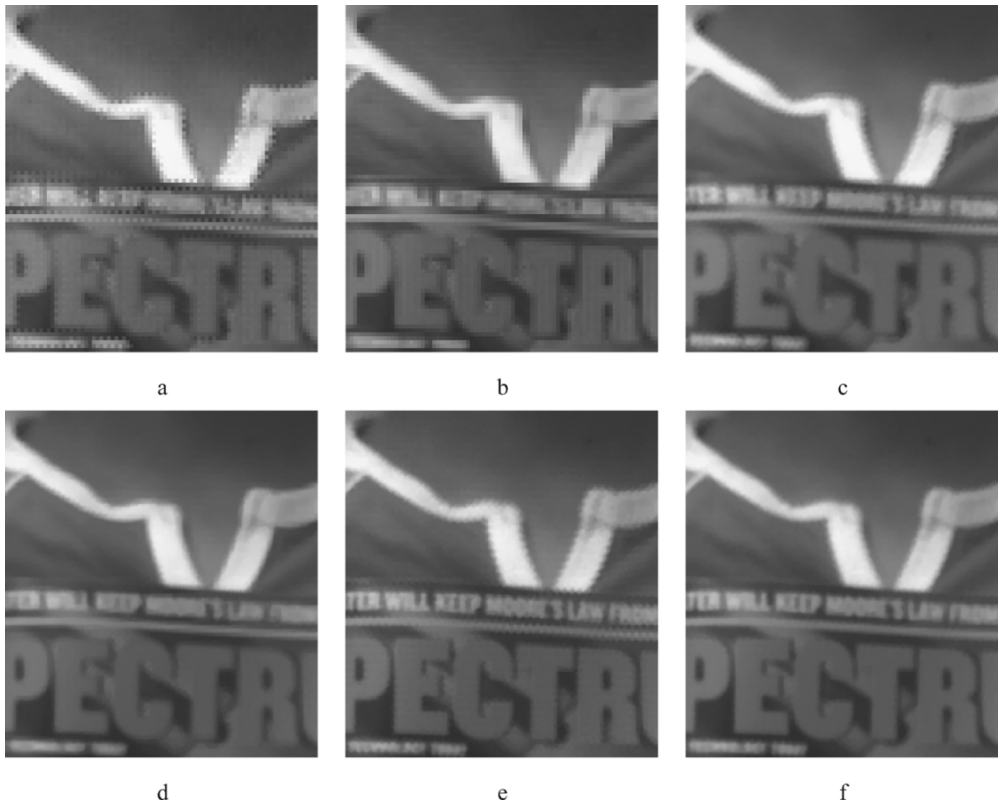


Fig. 12. Multiframe color super-resolution implemented on a real data sequence (zoomed). (a) One of the input LR images demosaiced by [14] and (b) one of the input LR images demosaiced by the more sophisticated [16]. (c) Result of applying the proposed color-super-resolution method on 31 LR images each demosaiced by [14] method. (d) Result of applying the proposed color-super-resolution method on 31 LR images each demosaiced by [16] method. The result of applying our method on the original un-demosaiced raw LR images (without using the inter color dependence term) is shown in (e). (f) Result of applying our method on the original un-demosaiced raw LR images.

not impose a serious problem if a reasonable initial guess is used for the steepest decent method, as many experiments showed effective multiframe demosaicing results. In our experiments, we noticed that a good initial guess is the shift-and-add result of the individually demosaiced LR images.

Accurate subpixel motion estimation is an essential part of any image fusion process such as multiframe super-resolution or demosaicing. To the best of our knowledge, no paper has addressed the problem of estimating motion between Bayer filtered images. However, a few papers have addressed related issues. Reference [43] has addressed the problem of color motion estimation, where information from different color channels are incorporated by simply using alternative color representations such as HSV or normalized RGB. More work remains to be done to fully analyze subpixel motion estimation from colored filtered images.

Appendix I
Derivation of the Intercolor Dependencies Penalty Term

In this appendix, we illustrate the differentiation of the first term in (14), which we call $\underline{L}$, with respect to $\underline{X}_G$. From (14), we have

$$\underline{L} = \left\| \underline{X}_G \odot S_x^l S_y^m \underline{X}_B - \underline{X}_B \odot S_x^l S_y^m \underline{X}_G \right\|_2^2 \xrightarrow{\odot \text{ is commutative}}$$

$$\underline{L} = \left\| S_x^l S_y^m \underline{X}_B \odot \underline{X}_G - \underline{X}_B \odot S_x^l S_y^m \underline{X}_G \right\|_2^2.$$

We can substitute the element by element multiplication operator "$\odot$," with the differentiable dot product by rearranging $\underline{X}_B$ as the diagonal matrix[20] $\mathbf{X_B}$ and $S_x^l S_y^m \underline{X}_B$ as $\mathbf{X_B^{l,m}}$, which is the diagonal form of shifted $\underline{X}_B$ by l, m pixels in horizontal and vertical directions

$$\underline{L} = \left\| \mathbf{X_B^{l,m}} \underline{X}_G - \mathbf{X_B} S_x^l S_y^m \underline{X}_G \right\|_2^2. \tag{18}$$

Using the identity

$$\frac{\partial \|\mathbf{Q}\underline{C}\|_2^2}{\partial \underline{C}} = \frac{\partial (\underline{C}^T \mathbf{Q^T Q} \underline{C})}{\partial \underline{C}} = 2\mathbf{Q^T Q}\underline{C}$$

and noting that $\mathbf{X_B^{l,m}}$ and $\mathbf{X_B}$ are symmetric matrices, the differentiation with respect to green band will be computed as follows:

$$\frac{\partial \underline{L}}{\partial \underline{X}_G} = 2\left(\mathbf{X_B^{l,m}} - S_x^{-l} S_y^{-m} \mathbf{X_B}\right)\left(\mathbf{X_B^{l,m}} \underline{X}_G - \mathbf{X_B} S_x^l S_y^m \underline{X}_G\right).$$

[20]We are simply denoting a vector $\underline{Q}$ to its diagonal matrix representation $\mathbf{Q}$ such that

$$\underbrace{\begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_{4r^2M^2} \end{pmatrix}}_{\underline{Q}} \longrightarrow \underbrace{\begin{pmatrix} q_1 & 0 & \cdots & 0 \\ 0 & q_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & q_{4r^2M^2} \end{pmatrix}}_{\mathbf{Q}}.$$

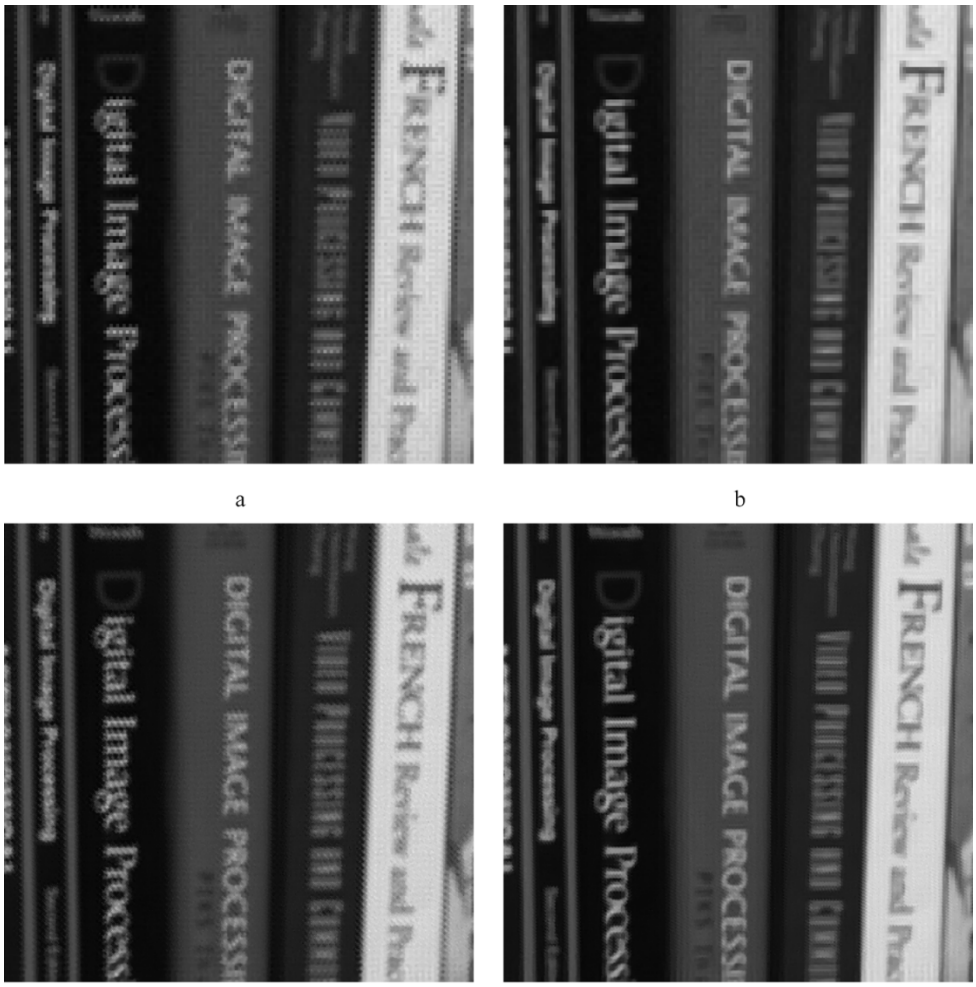


Fig. 13. Multiframe color super-resolution implemented on a real data sequence. (a) One of the input LR images demosaiced by [14] and (b) one of the input LR images demosaiced by the more sophisticated [16]. (c) Result of applying the proposed color-super-resolution method on 31 LR images each demosaiced by [14] method. (d) Result of applying the proposed color-super-resolution method on 31 LR images each demosaiced by [16] method.

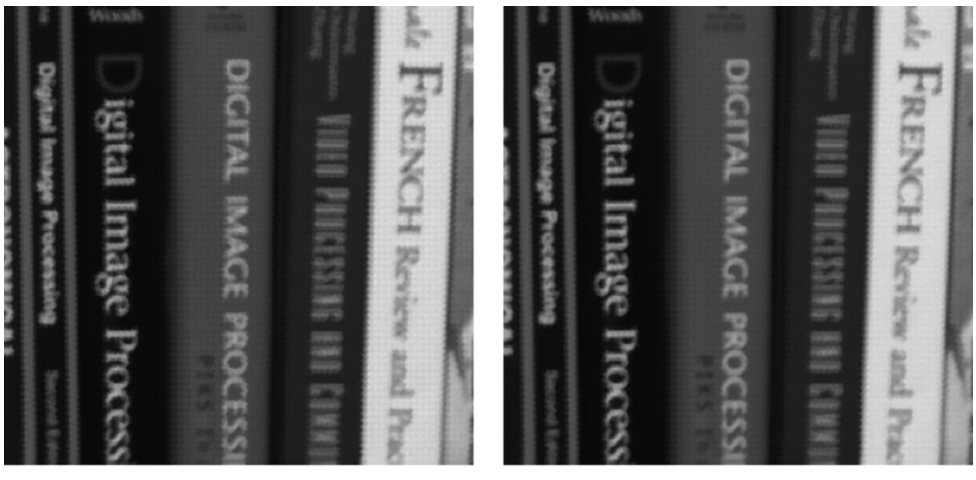


Fig. 14. Multiframe color super-resolution implemented on a real data sequence. The result of applying our method on the original un-demosaiced raw LR images (without using the inter color dependence term) is shown in (a). (b) Result of applying our method on the original un-demosaiced raw LR images.

Differentiation of the second term in (14), and also differentiation with respect to the other color bands follow the same technique.

ACKNOWLEDGMENT

The authors would like to thank Prof. R. Kimmel for sharing his demosaicing software with them, the associate editor and

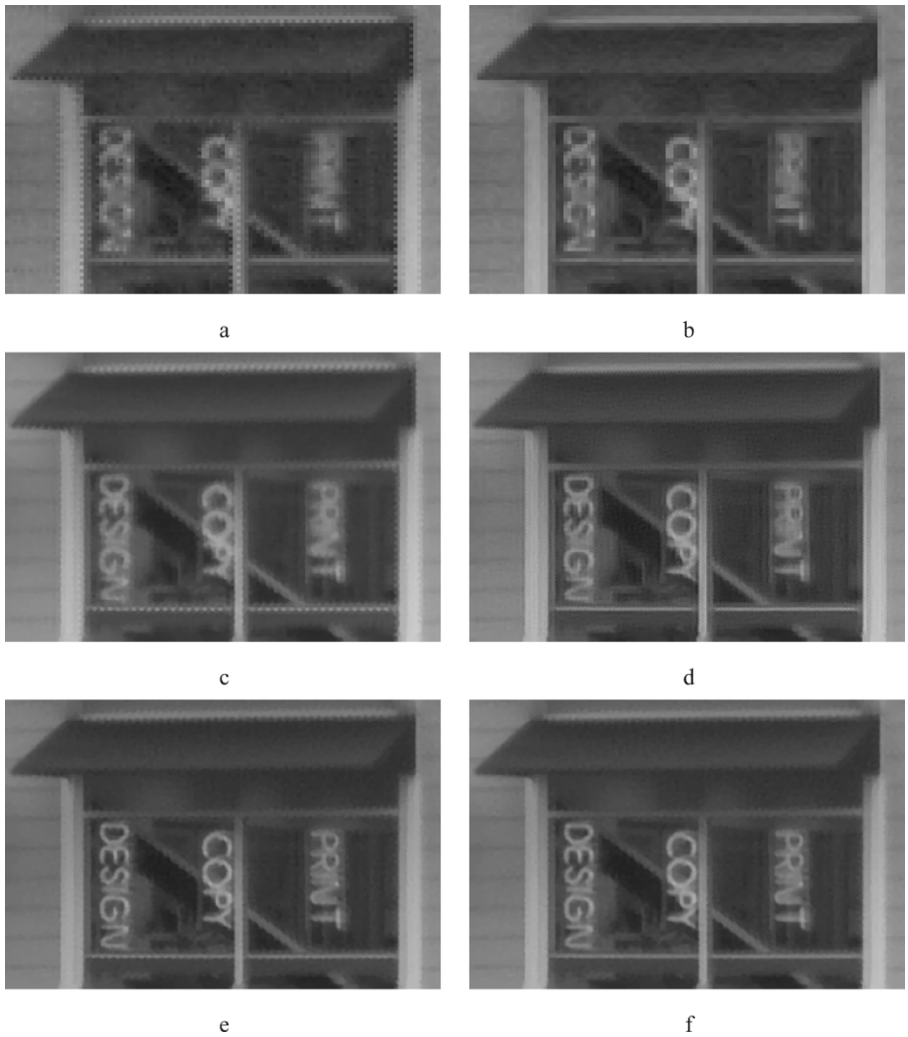


Fig. 15. Multiframe color super-resolution implemented on a real data sequence. (a) One of the input LR images demosaiced by [14] and (b) one of the input LR images demosaiced by the more sophisticated [16]. (c) Result of applying the proposed color-super-resolution method on 30 LR images each demosaiced by [14] method. (d) Result of applying the proposed color-super-resolution method on 30 LR images each demosaiced by [16] method. The result of applying our method on the original un-demosaiced raw LR images (without using the inter color dependence term) is shown in (e). (f) Result of applying our method on the original un-demosaiced raw LR images.

the reviewers for their valuable comments that helped improve the clarity of presentation of this paper, D. Robinson for useful discussions, and L. Zimet and E. Galil from Zoran Corporation for providing the camera used to produce the raw CFA images of experiments 4, 5, and 6.

REFERENCES

[1] S. Borman and R. L. Stevenson, "Super-resolution from image sequences—a review," presented at the Midwest Symp. Circuits and Systems, vol. 5, Apr. 1998.

[2] S. Park, M. Park, and M. G. Kang, "Super-resolution image reconstruction, a technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.

[3] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super-resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.

[4] ——, "Advances and challenges in super-resolution," *Int. J. Imag. Syst. Technol.*, vol. 14, no. 2, pp. 47–57, Aug. 2004.

[5] T. S. Huang and R. Y. Tsai, "Multi-frame image restoration and registration," *Adv. Comput. Vis. Image Process.*, vol. 1, pp. 317–339, 1984.

[6] N. Nguyen, P. Milanfar, and G. H. Golub, "A computationally efficient image superresolution algorithm," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 573–583, Apr. 2001.

[7] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graph. Models Image Process.*, vol. 53, pp. 231–239, 1991.

[8] S. Peleg, D. Keren, and L. Schweitzer, "Improving image resolution using subpixel motion," *CVGIP: Graph. Models Image Process.*, vol. 54, pp. 181–186, Mar. 1992.

[9] M. Elad and A. Feuer, "Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1646–1658, Dec. 1997.

[10] A. Zomet and S. Peleg, "Efficient super-resolution and applications to mosaics," in *Proc. Int. Conf. Pattern Recognition*, Sep. 2000, pp. 579–583.

[11] N. R. Shah and A. Zakhor, "Resolution enhancement of color video sequences," *IEEE Trans. Image Process.*, vol. 8, no. 6, pp. 879–885, Jun. 1999.

[12] B. C. Tom and A. Katsaggelos, "Resolution enhancement of monochrome and color video using motion compensation," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 278–287, Feb. 2001.

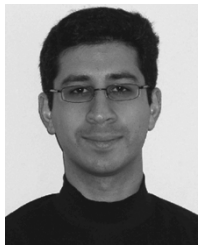
[13] D. R. Cok, "Signal Processing Method and Apparatus for Sampled Image Signals," U.S. Patent 4 630 307, 1987.

[14] C. Laroche and M. Prescott, "Apparatus and Method for Adaptive for Adaptively Interpolating a Full Color Image Utilizing Chrominance Gradients," U.S. Patent 5 373 322, 1994.

[15] J. Hamilton and J. Adams, "Adaptive Color Plan Interpolation in Single Sensor Color Electronic Camera," U.S. Patent 5 629 734, 1997.

[16] R. Kimmel, "Demosaicing: image reconstruction from color ccd samples," *IEEE Trans. Image Process.*, vol. 8, no. 9, pp. 1221–1228, Sep. 1999.

[17] L. Chang and Y.-P. Tan, "Color filter array demosaicking: new method and performance measures," *IEEE Trans. Image Process.*, vol. 12, no. 10, pp. 1194–1210, Oct. 2002.
[18] K. Hirakawa and T. Parks, "Adaptive homogeneity-directed demosaicing algorithm," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, Sep. 2003, pp. 669–672.
[19] D. Keren and M. Osadchy, "Restoring subsampled color images," *Mach. Vis. Appl.*, vol. 11, no. 4, pp. 197–202, 1999.
[20] Y. Hel-Or and D. Keren, "Demosaicing of Color Images Using Steerable Wavelets," HP Labs, Israel, Tech. Rep. HPL-2002-206R1 20 020 830, 2002.
[21] D. Taubman, "Generalized Wiener reconstruction of images from color sensor data using a scale invariant prior," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3, Sep. 2000, pp. 801–804.
[22] D. D. Muresan and T. W. Parks, "Optimal recovery demosaicing," presented at the IASTED Signal and Image Processing, Aug. 2002.
[23] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Color plane interpolation using alternating projections," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 997–1013, Sep. 2002.
[24] S. C. Pei and I. K. Tam, "Effective color interpolation in CCD color filter arrays using signal correlation," *IEEE Trans. Image Process.*, vol. 13, no. 6, pp. 503–513, Jun. 2003.
[25] D. Alleysson, S. Süsstrunk, and J. Hrault, "Color demosaicing by estimating luminance and opponent chromatic signals in the Fourier domain," in *Proc. IS&T/SID 10th Color Imaging Conf.*, Nov. 2002, pp. 331–336.
[26] X. Wu and N. Zhang, "Primary-consistent soft-decision color demosaic for digital cameras," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, Sep. 2003, pp. 477–480.
[27] R. Ramanath and W. Snyder, "Adaptive demosaicking," *J. Electron. Imag.*, vol. 12, no. 4, pp. 633–642, Oct. 2003.
[28] S. Farsiu, M. Elad, and P. Milanfar, "Multi-frame demosaicing and super-resolution from under-sampled color images," presented at the Proc. IS&T/SPIE 16th Annu. Symp. Electronic Imaging, Jan. 2004.
[29] M. A. Lukas, "Asymptotic optimality of generalized cross-validation for choosing the regularization parameter," *Numer. Math.*, vol. 66, no. 1, pp. 41–66, 1993.
[30] N. Nguyen, P. Milanfar, and G. Golub, "Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement," *IEEE Trans. Image Process.*, vol. 10, no. 9, pp. 1299–1308, Sep. 2001.
[31] P. C. Hansen and D. P. O'Leary, "The use of the 1-curve in the regularization of ill-posed problems," *SIAM J. Sci. Comput.*, vol. 14, no. 6, pp. 1487–1503, Nov. 1993.
[32] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259–268, Nov. 1992.
[33] T. F. Chan, S. Osher, and J. Shen, "The digital TV filter and nonlinear denoising," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 231–241, Feb. 2001.
[34] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE Int. Conf. Computer Vision*, Jan. 1998, pp. 836–846.
[35] M. Elad, "On the bilateral filter and ways to improve it," *IEEE Trans. Image Process.*, vol. 11, no. 10, pp. 1141–1151, Oct. 2002.
[36] M. E. Tipping and C. M. Bishop, "Bayesian image super-resolution," *Adv. Neural Inf. Process. Syst.*, vol. 15, pp. 1303–1310, 2002.
[37] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Robust shift and add approach to super-resolution," in *Proc. SPIE Conf. Applications of Digital Signal and Image Processing*, Aug. 2003, pp. 121–130.
[38] W. K. Pratt, *Digital Image Processing*, 3rd ed. New York: Wiley, 2001.
[39] A. Zomet and S. Peieg, "Multi-sensor super resolution," in *Proc. IEEE Workshop on Applications of Computer Vision*, Dec. 2001, pp. 27–31.
[40] T. Gotoh and M. Okutomi, "Direct super-resolution and registration using raw CFA images," in *Proc. Int. Conf. Computer Vision and Pattern Recognition*, vol. 2, Jul. 2004, pp. 600–607.
[41] R. Ramanath, W. Snyder, G. Bilbro, and W. Sander, "Demosaicking methods for the Bayer color arrays," *J. Electron. Imag.*, vol. 11, no. 3, pp. 306–315, Jul. 2002.
[42] X. Zhang, D. A. Silverstein, J. E. Farrell, and B. A. Wandell, "Color image quality metric s-cielab and its application on halftone texture visibility," in *Proc. IEEE COMPCON Symp. Dig.*, May 1997, pp. 44–48.
[43] P. Golland and A. M. Bruckstein, "Motion from color," *Comput. Vis. Image Understand.*, vol. 68, no. 3, pp. 346–362, Dec. 1997.
[44] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierachical model-based motion estimation," in *Proc. Eur. Conf. Computer Vision*, May 1992, pp. 237–252.



Sina Farsiu received the B.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 1999, and the M.Sc. degree in biomedical engineering from the University of Tehran, Tehran, in 2001. He is currently pursuing the Ph.D. degree in electrical engineering at the University of California, Santa Cruz.

His technical interests include signal and image processing, adaptive optics, and artificial intelligence.

Michael Elad received the B.Sc, M.Sc., and D.Sc. degrees from the Department of Electrical Engineering at The Technion—Israel Institute of Technology (IIT), Haifa, in 1986, 1988, and 1997, respectively.

From 1988 to 1993, he served in the Israeli Air Force. From 1997 to 2000, he worked at Hewlett-Packard Laboratories, Israel, as an R&D Engineer. From 2000 to 2001, he headed the research division at Jigami Corporation, Israel. From 2001 to 2003, he was a Research Associate with the Computer Science Department, Stanford University (SCCM program), Stanford, CA. In September 2003, he joined the Department of Computer Science, IIT, as an Assistant Professor. He was also a Research Associate at IIT from 1998 to 2000, teaching courses in the Electrical Engineering Department. He works in the field of signal and image processing, specializing, in particular, on inverse problems, sparse representations, and over-complete transforms.

Dr. Elad received the Best Lecturer Award twice (in 1999 and 2000). He is also the recipient of the Guttwirth and the Wolf fellowships.

Peyman Milanfar (SM'98) received the B.S. degree in electrical engineering and mathematics from the University of California, Berkeley, and the S.M., E.E., and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1988, 1990, 1992, and 1993, respectively.

Until 1999, he was a Senior Research Engineer at SRI International, Menlo Park, CA. He is currently Associate Professor of electrical engineering, University of California, Santa Cruz. He was a Consulting Assistant Professor of computer science at Stanford University, Stanford, CA, from 1998 to 2000, where he was also a Visiting Associate Professor from June to December 2002. His technical interests are in statistical signal and image processing and inverse problems.

Dr. Milanfar won a National Science Foundation CAREER award in 2000 and he was Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 1998 to 2001.