Multi-Frame Demosaicing and Super-Resolution from Under-Sampled Color Images

Sina Farsiu[a*], Michael Elad[b], Peyman Milanfar[a]

[a] Electrical Engineering Department, University of California, Santa Cruz CA. 95064 USA.
[b] Computer Science Department, Technion, Israel Institute of Technology, Haifa 32000, Israel

ABSTRACT

In the last two decades, two related categories of problems have been studied independently in the 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 (MAP) estimation technique by minimizing a multi-term 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 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 component. Finally, an additional regularization term is used to force similar edge 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.

Keywords: Super-Resolution, Demosaicing, Robust Estimation, Robust Regularization, Color Enhancement

1. INTRODUCTION

In digital photography two problems have been studied and solved independently - super-resolution 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 the existing sensors to measure the R (red), G (green), and B (blue) values without using a full RGB sensor array. 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 pre-processed (demosaiced) images is sub-optimal and may be inferior to a single unifying solution framework. In this section, we study these two common problems of digital cameras and review some of the previous work on these subjects. We show the inefficiency of independent solutions for these problems and in the next section provide a unified approach for addressing them.

1.1. 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, 2].

We use the Ref.[3] notation to explain the image formation process of digital cameras. Assume that $\underline{X}$ is the monochromatic HR frame (of size $[r^2M^2 \times 1]$) and $\underline{Y}_k$ is the k^{th} monochromatic LR frame (of size $[M^2 \times 1]$), which are rearranged in the lexicographic order. Then we assume that they are related to each other by the following model:

$$\underline{Y}_k = D_k H_k F_k \underline{X} + \underline{V}_k \qquad k = 1, \ldots, N \tag{1}$$

* (Corresponding Author) Email: farsiu@ee.ucsc.edu, Phone:(831)-459-4141, Fax: (831)-459-4829

where $[r^2M^2 \times r^2M^2]$ matrix F_k is the geometric motion operator between $\underline{X}$ and $\underline{Y}_k$ and r is the resolution enhancement factor. The camera's point spread function (PSF) is modelled by the $[r^2M^2 \times r^2M^2]$ blur matrix H_k, and $[M^2 \times r^2M^2]$ matrix D_k represents the decimation operator. The $[r^2M^2 \times 1]$ vector $\underline{V}_k$ is the system noise and N is the number of available LR frames.

In the last two decades a variety of super-resolution methods have been proposed for estimating the HR image $\underline{X}$ from the set of LR images $\underline{Y}_k$ (see Ref.[3–8]). These methods are either computationally expensive or sensitive to the outliers in measurements or modelling, which limit their applications to real world data. In Ref.[1, 2] the importance of robustness in the super-resolution problem is explored, and a novel robust method is proposed, being numerically efficient. In this section we briefly review the method presented in Ref.[1, 2][†] where a penalty functional is defined and the reconstructed super-resolution image is the minimizer of this functional:

$$\underline{\widehat{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[\sum_{k=1}^{N} \|D_k H_k F_k \underline{X} - \underline{Y}_k\|_1 + \lambda \underbrace{\sum_{l=-P}^{P} \sum_{m=0}^{P} \alpha^{m+l} \|\underline{X} - S_x^l S_y^m \underline{X}\|_1}_{l+m \geq 0} \right] \tag{2}$$

The L_1 norm used for measuring the first term (likelihood cost), makes the method robust to the errors in modelling and outliers in measurements. The second term (regularization cost) forces spatial smoothness, and changes the possibly ill-posed super-resolution problem to a well-posed one, and results in images with sharp edges. In (2) λ, the regularization parameter, is weighting the likelihood cost against the regularization cost, and the matrices (operators) S_x^l, and S_y^m shift $\underline{X}$ by l, and m pixels in horizontal and vertical directions respectively, presenting several scales of derivatives. The scalar weight α, $0 < \alpha < 1$, is applied to give a spatially decaying effect to the summation of the regularization terms. Note that the unsymmetric summation of the regularization terms excludes redundant computations. Assuming a common space-invariant PSF (which is a reasonable assumption when all images are acquired with the same camera) and considering translational motion, the above minimization problem can be solved in a computationally more efficient way:

$$\underline{\widehat{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[\|A(H\underline{X} - \underline{\widehat{Z}})\|_1 + \lambda' \underbrace{\sum_{l=-P}^{P} \sum_{m=0}^{P} \alpha^{m+l} \|\underline{X} - S_x^l S_y^m \underline{X}\|_1}_{l+m \geq 0} \right] \tag{3}$$

where $\underline{\widehat{Z}}$ is the blurry high resolution image obtained by motion compensation and mapping the LR pixels to the HR grid (after proper zero filling), and computing a median of values falling within the same pixel location. The matrix A is a diagonal matrix with diagonal values equal to the square root of the number of measurements that contributed to make each element of $\underline{\widehat{Z}}$.

Note that this method and almost all other super-resolution methods are designed to increase the resolution of a single channel (monochromatic) image and a super-resolved color image could be acquired by super-resolving each color channel independently. Those algorithms that do refer to color suggest a super-resolution algorithm on the luminance, leaving the color layers to a simple interpolation[5]. In Section 3 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 if the measured images are mosaiced. This brings us to the description of the demosaicing problem.

1.2. Demosaicing

A color image is represented by combining three separate monochromatic images. Therefore, if possible, each pixel should be provided with three sensors, each measuring one color band. To reduce production cost, in many digital cameras only one color sensor (red, green, or blue) is assigned to each pixel. Thus, the detector array is a grid of CCDs, each sensitive to one certain color. Such grid is constructed by placing a color filter array (CFA)

[†]These papers are available at *http://www.ee.ucsc.edu/∼milanfar/publications.htm*

on a CCD. A very common pattern for the CFAs is the Bayer pattern which is shown in Figure 1. Demosaicing addresses the estimation of the missing color bands of a pixel, from the sensor values of neighboring pixels. Numerous demosaicing methods have been proposed through the years[9–15]. Almost all of these 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 pixel is surrounded by 4 green pixels).
2. Each red, green and blue pixel is in the neighborhood of pixels from different color bands.
3. The pattern of pixels does not change through the image.
4. For each pixel one and only one color band value is available.
5. The human eye is more sensitive to the details in the luminance component of the image than the details in chrominance component[14].
6. The human eye is more sensitive to chromatic change in the low spatial frequency region than the luminance change[15].
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 components.

1.3. Merging super-resolution and demosaicing into one process

Unfortunately, many popular single-frame demosaicing methods are based on assumptions not valid for the multi-frame cases. To have a better understanding of the multi-frame demosaicing problem, lets consider an example with the simplifying assumptions of common space-invariant PSF and translational motion. In this example, 16 Bayer filtered LR images are fused together, increasing the resolution by a factor of 4 in each direction. As explained in Ref.[1, 2], the Maximum Likelihood(ML) estimate of the HR image is nothing but the Shift and Add of all LR images. Figure 1 illustrates that the pattern of sensor distribution in the HR image does not necessarily follow the original Bayer pattern, but rather depends on the relative motion of the LR images.

Figure 1 shows that the first assumption of single-frame demosaicing, exploited in many papers to propose methods that treat green channel different than red or blue channels, is not useful for the multi-frame case. While globally there are more green pixels than the blue or red pixels, locally any pixels may be surrounded by only red or blue colors. So, there is no general preference for one color band over the others bands. A closer look at pixel B_1 in the low and high resolution images (B_1 is denoted in Figure 1 by bold fonts), illustrates this claim. In the LR frame, B_1 is surrounded by four green pixels but in the HR frame, there are no green pixels in its neighborhood. The same example shows that the second assumption is not true, either. The third assumption 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. The forth assumption is not correct, because in the under-determined case [‡], there are not enough measurements to fill the HR grid. On the other hand, in the over-determined case[§], for some pixels, there may in fact be more than one color channel available.

It should be noted that Zomet and Peleg[16] have recently proposed a novel method for combining the information from multiple sensors, which can also be used for demosaicing purposes. Their method is based on the assumption of affine relation between the intensities of different sensors in a local neighborhood. Although their method has produced successful results, since it is not exclusively designed for solving the multi-frame demosaicing problem there is much more room for improvements. In contrast to their method, our proposed

[‡] where the number of non-redundant LR frames is smaller than the square of resolution enhancement factor ($N < r^2$)

[§] where the number of non-redundant LR frames is larger than the square of resolution enhancement factor ($N > r^2$)

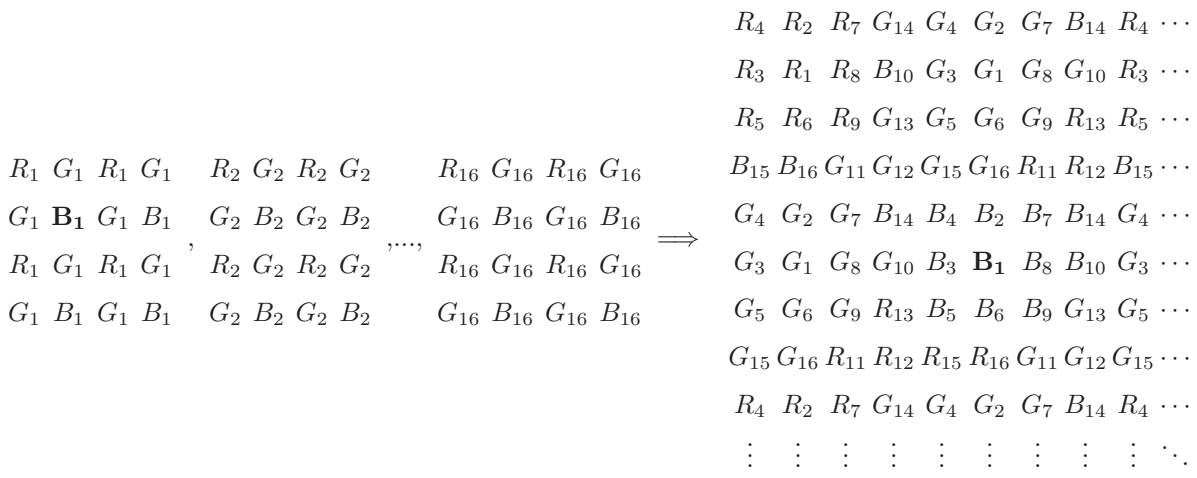


Figure 1. Fusion of 16 Bayer pattern LR images with relative translational motion (the figures in the left side of the arrow) results in a HR image that does not follow Bayer pattern (the figure in the right side of the arrow). R_n, G_n, and B_n represent the red, green, and blue sensor values of the n^{th} LR frame, respectively.

technique exploits the difference in sensitivity of the human eye to the frequency content and outliers in the luminance and chrominance components of the image. Also, the method in Ref.[16] is computationally expensive as the affine mapping coefficients of each pixel should be computed in each iteration.

2. MULTI-FRAME DEMOSAICING

In the previous section we argued that the multi-frame demosaicing is different than the single-frame demosaicing in the sense that it may only benefit from the last five assumptions of Section 1.2.

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:

1. A penalty term for the non-similarities between the raw data and the HR estimate (Likelihood Penalty Term).
2. A penalty term for the non-smoothness in the luminance component of the HR image (Luminance Penalty Term).
3. A penalty term for the non-smoothness in the chrominance component of the HR image (Chrominance Penalty Term).
4. A penalty term for the non-homogenousity of the edge orientation in different color bands (Orientation Penalty Term).

Each of these penalty terms will be discussed in the following subsections.

2.1. Likelihood Penalty Term

This term measures the similarity between the resulting HR image and the original LR images. As it is explained in Section 1.1 and Ref.[1], L_1 norm minimization of the error term results in robust reconstruction of the HR image. We use the first term in the right-hand-side of (3) to represent the data penalty term, as it is easier to interpret and has a faster implementation potential. So the likelihood penalty term is defined as:

$$Cost(data) = \|A_R(H\underline{X}_R - \widehat{\underline{Z}}_R)\|_1 + \|A_G(H\underline{X}_G - \widehat{\underline{Z}}_G)\|_1 + \|A_B(H\underline{X}_B - \widehat{\underline{Z}}_B)\|_1$$

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 $\underline{\widehat{Z}}$ and A_R, A_G, and A_B keep track of the known red, green, and blue sensor values in the shift and add image and $\underline{\widehat{X}}_R$, $\underline{\widehat{X}}_G$, and $\underline{\widehat{X}}_B$ are the three color components of the reconstructed HR image X.

More generally, considering motion models other than translational model or non-common space-variant PSF, would force us to use the first term in (2) as the likelihood penalty term:

$$Cost(data) = \sum_{k=1}^{N} \|A_{k_R}(D_k H_k F_k \underline{X}_R - \underline{Y}_{k_R})\|_1 + \sum_{k=1}^{N} \|A_{k_G}(D_k H_k F_k \underline{X}_G - \underline{Y}_{k_G})\|_1 + \sum_{k=1}^{N} \|A_{k_B}(D_k H_k F_k \underline{X}_B - \underline{Y}_{k_B})\|_1$$

where $\underline{Y}_{k_R}$, $\underline{Y}_{k_G}$, and $\underline{Y}_{k_B}$ are the sparse red, green and blue components of the k^{th} LR image (of size $[M^2 \times 1]$ after zero filling the unknown color sensor value positions) and A_{k_R}, A_{k_G}, and A_{k_B} keep track of the known sensor value positions.

Note that the above penalty function is applicable for general models of data, blur and motion. However, in this paper we only experimented with 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.

2.2. 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[14]. Therefore, it is important that the edges in the luminance component of the reconstructed HR image look sharp. Applying bilateral regularization to the luminance component will result in this desired property[1]. The luminance regularization term is defined as:

$$Cost(luminance) = \underbrace{\sum_{l=-P}^{P} \sum_{m=0}^{P}}_{l+m\geq 0} \alpha^{m+l} \|0.299(\underline{X}_R - S_x^l S_y^m \underline{X}_R) + 0.587(\underline{X}_G - S_x^l S_y^m \underline{X}_G) + 0.114(\underline{X}_B - S_x^l S_y^m \underline{X}_B)\|_1$$

where (0.299,0.587,0.114) are the appropriate coefficients for extracting the luminance component of the RGB representation of the reconstructed HR image[17] X. The variables α, S_x, S_y, and P are defined in (2).

2.3. Chrominance Penalty Term

As the human eye is less sensitive to the chrominance channel resolution, it can be smoothed more aggressively. Therefore, Tikhonov regularization is an appropriate method for smoothing the Chrominance term:

$$Cost(Chrominance) = \|\Lambda(-0.169 \times \underline{X}_R - 0.331 \times \underline{X}_G + 0.5 \times \underline{X}_B)\|_2^2 + \|\Lambda(0.5 \times \underline{X}_R - 0.419 \times \underline{X}_G - 0.081 \times \underline{X}_B)\|_2^2$$

where (-0.169,-0.331,0.5) and (0.5,-0.419,-0.081) are the appropriate coefficients for extracting the chrominance component of the reconstructed HR image X[17], and Λ is the matrix realization of a high-pass operator such as the Laplacian.

2.4. Orientation Penalty Term

Although different bands may have larger or smaller gradients at a particular edge, it is reasonable to assume that all color bands have the same edge orientation[¶]. Many papers have addressed the correlation of the RGB color bands and exploited it for enhancing their proposed single-frame demosaicing algorithms. Most of these methods are not applicable for the multi-frame demosaicing cases. In this paper, we follow Ref.[11][‖] to force different

[¶]That is, if a vertical (or horizontal) edge appears in the red band, a vertical (or horizontal) edge in the same location is likely to appear in the green and blue bands. Note that this does not imply the equality of the edge magnitude in different color bands.

[‖]The proposed regularization term in Ref.[11] unlike many other methods is independent of the pattern of sensor distribution in the HR image and therefor applicable to the multi-frame demosaicing problem.

bands to have similar edge 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:

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

where Θ is the angle between these two vectors. As the likelihood 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. Note that small Θ is an indicator of similar orientation.

Ref.[11] suggests a pixelwise orientation cost function which is solved by the finite element method. With some modifications our orientation penalty term is a differentiable cost function:

$$\begin{aligned} Cost(Orientation) = \underbrace{\sum_{l=-1}^{1}\sum_{m=0}^{1}}_{l+m\geq 0} \big(&\|\underline{X}_G \odot S_x^l S_y^m \underline{X}_B - \underline{X}_B \odot S_x^l S_y^m \underline{X}_G\|_2^2 + \\ &\|\underline{X}_B \odot S_x^l S_y^m \underline{X}_R - \underline{X}_R \odot S_x^l S_y^m \underline{X}_B\|_2^2 + \|\underline{X}_R \odot S_x^l S_y^m \underline{X}_G - \underline{X}_G \odot S_x^l S_y^m \underline{X}_R\|_2^2\big) \end{aligned} \quad (4)$$

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

2.5. Overall Cost Function

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

$$\begin{aligned} \widehat{\underline{X}} = \underset{\underline{X}_R,\underline{X}_G,\underline{X}_B}{\text{ArgMin}} \Big[&\|A_R(H\underline{X}_R - \widehat{\underline{Z}}_R)\|_1 + \|A_G(H\underline{X}_G - \widehat{\underline{Z}}_G)\|_1 + \|A_B(H\underline{X}_B - \widehat{\underline{Z}}_B)\|_1 + \\ &\lambda' \underbrace{\sum_{l=-P}^{P}\sum_{m=0}^{P}}_{l+m\geq 0} \big(\alpha^{m+l}\|0.2989(\underline{X}_R - S_x^l S_y^m \underline{X}_R) + 0.5870(\underline{X}_G - S_x^l S_y^m \underline{X}_G) + 0.1140(\underline{X}_B - S_x^l S_y^m \underline{X}_B)\|_1\big) + \\ &\lambda'' \underbrace{\sum_{l=-1}^{1}\sum_{m=0}^{1}}_{l+m\geq 0} (\|\underline{X}_G \odot S_x^l S_y^m \underline{X}_B - \underline{X}_B \odot S_x^l S_y^m \underline{X}_G\|_2^2 + \|\underline{X}_B \odot S_x^l S_y^m \underline{X}_R - \underline{X}_R \odot S_x^l S_y^m \underline{X}_B\|_2^2 + \\ &\|\underline{X}_R \odot S_x^l S_y^m \underline{X}_G - \underline{X}_G \odot S_x^l S_y^m \underline{X}_R\|_2^2) + \\ &\lambda''' (\|\Lambda(-0.169 \times \underline{X}_R - 0.331 \times \underline{X}_G + 0.5 \times \underline{X}_B)\|_2^2 + \|\Lambda(0.5 \times \underline{X}_R - 0.419 \times \underline{X}_G - 0.081 \times \underline{X}_B)\|_2^2)\Big] \end{aligned} \quad (5)$$

Steepest (coordinate) descent optimization may be applied to minimize this cost function. In the first step, the derivative of (5) 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 follows:

$$\begin{aligned} \nabla \widehat{\underline{X}}_G^n = \; & H^T A_G^T sign(A_G H \widehat{\underline{X}}_G^n - A_G \widehat{\underline{Z}}_G) + \\ & \lambda' \underbrace{\sum_{l=-P}^{P}\sum_{m=0}^{P}}_{l+m\geq 0} \alpha^{m+l} \times 0.5870 \times [I - S_y^{-m} S_x^{-l}] sign\big(0.2989(\underline{X}_R^n - S_x^l S_y^m \underline{X}_R^n) + \\ & 0.5870(\underline{X}_G^n - S_x^l S_y^m \underline{X}_G^n) + 0.1140(\underline{X}_B^n - S_x^l S_y^m \underline{X}_B^n)\big) + \\ & \lambda'' \underbrace{\sum_{l=-1}^{1}\sum_{m=0}^{1}}_{l+m\geq 0} \Big[2(\mathbf{X_B^{l,m}} - S_x^{-l} S_y^{-m} \mathbf{X_B})(\mathbf{X_B^{l,m}} \underline{X}_G - \mathbf{X_B} S_x^l S_y^m \underline{X}_G) + \\ & 2(\mathbf{X_R^{l,m}} - S_x^{-l} S_y^{-m} \mathbf{X_R})(\mathbf{X_R^{l,m}} \underline{X}_G - \mathbf{X_R} S_x^l S_y^m \underline{X}_G)\Big] + \\ & \lambda''' \Lambda^T \Lambda(-0.1536 \times \underline{X}_R + 0.2851 \times \underline{X}_G - 0.1316 \times \underline{X}_B) \end{aligned} \quad (6)$$

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. $\mathbf{X_R}$, and $\mathbf{X_B}$ are 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 orientation term derivative is explained in Appendix.A.

Matrices H, Λ, 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, decimation, and shift. Noting and implementing the effects of these matrices as a sequence of operators spares us from explicitly constructing them as matrices. This property helps our method to be implemented in an extremely fast and memory efficient way.

The gradient of the other channels will be computed in the same way, and the following steepest (coordinate) descent iterations will be set up to calculate the HR image estimate iteratively.

$$\begin{aligned}\underline{\widehat{X}}_R^{n+1} &= \underline{\widehat{X}}_R^n - \beta\nabla\underline{\widehat{X}}_R^n \\ \underline{\widehat{X}}_G^{n+1} &= \underline{\widehat{X}}_G^n - \beta\nabla\underline{\widehat{X}}_G^n \\ \underline{\widehat{X}}_B^{n+1} &= \underline{\widehat{X}}_B^n - \beta\nabla\underline{\widehat{X}}_B^n\end{aligned} \tag{7}$$

where the scalar β is the step size.

3. EXPERIMENTS

Experiments on synthetic and real world 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 (Figure 2.a). First we shifted this HR image by a pixel in 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 4×4 with standard deviation equal to one. The resulting image was subsampled by the factor of 4 in each direction. The same approach with different motion vectors (shifts) in vertical and horizontal directions was used to produce 10 LR images from the original scene. Then each LR image was subsampled by the Bayer filter. In order to show how one of those measured images looks like, one of these Bayer filtered LR images is reconstructed by the method in Ref.[9] and shown in Figure 2.b. The above method is implemented on Kodak DCS-200 digital cameras[18], so each LR image may be considered as one output image of this camera brand.

As the motion model for this experiment is translational and the blur kernel is space invariant, we can use the fast model of (3) to reconstruct the blurred HR grid. The shift and add result of the demosaiced LR frames after bilinear interpolation**, before deblurring and demosaicing is shown in Figure 2.c. We used the original set of frames (raw data) to reconstruct a HR image with reduced color artifacts. Figures 2.d-f show the effect of the individual implementation of each regularization term, and Figure 2.e shows the result of the implementation of (5) with all terms. The parameters used for this example are as follows: $\beta = 0.04$, $\alpha = 0.9$, $\lambda^{'} = 0.005$, $\lambda^{''} = 0.5$, $\lambda^{'''} = 100$. It is clear that the resulting image has a better quality than the LR input frames or other reconstruction methods.

In the second experiment, we used 40 compressed images from a commercial video camera; courtesy of Adyoron Intelligent Systems Ltd., Tel Aviv, Israel. Figure 3.a shows one of these LR images (a selected region of this image is zoomed in Figure 3.b for close examinations). 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 multi-frame 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[19]. We used the method described in Ref.[20] to computed the motion vectors. In Figure 3.c (which is zoomed in Figure 3.d) the method of Ref.[1] 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 Figure 3.e (which is zoomed in Figure 3.f), where color artifacts are significantly reduced. Moreover, comparing to the Figures 3.a-d, the compression errors have been removed more effectively in Figures 3.e-f.

**Interpolation is needed as this experiment is an under-determined problem.

4. DISCUSSION AND FUTURE WORK

In this paper, based on the MAP estimation technique, we proposed a hybrid 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 L_1 norm for the data error term makes our method robust to errors in data and modelling. Bilateral regularization of the luminance term results in sharp reconstruction of edges, and the chrominance and orientation cost functions remove the color outliers from the HR estimate. All matrix-vector operations in the proposed method are implemented as simple image operators. Parallel processing may also be used to further increase the computational efficiency.

We noticed that the bilateral luminance regularization term had the highest effect on the reconstructed image. The chrominance penalty term had the least effect, and therefore can be ignored to save time.

The non-linearity of the orientation term (4) results in the non-convexity of the overall penalty function. Therefore, the steepest decent optimization of (5) may reach a local rather than the global minimum of the overall function. The non-convexity does not impose a serious problem if a reasonable initial guess is used for the steepest decent method, as many experiments showed effective multi-frame demosaicing results. In our experiments we noticed that a good initial guess is the shift and add of the individually demosaiced low-resolution images (Figures 2-3).

Accurate subpixel motion estimation is an essential part of any image fusion process such as multi-frame 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 relative issues. Ref.[19] has addressed the problem of color motion estimation, where the information of different color channels are incorporated by simply using alternative color representations such as HSV or normalized RGB. Moreover, a few papers have suggested different methods of concurrent motion estimation and super-resolution[21–25]. Simulation results show the effectiveness of these methods. Therefore an important extension of our paper includes incorporation of motion estimation algorithms in the proposed multi-frame demosaicing method.

APPENDIX A. DERIVATION OF THE ORIENTATION PENALTY TERM

In this appendix, we illustrate the differentiation of the first term in (4), which we call $\underline{L}$, with respect to $\underline{X}_G$. From (4) we have:

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

We can substitute the element by element multiplication operator "$\odot$", with the differentiable dot product by rearranging $\underline{X}_B$ as the diagonal matrix[††] $\mathbf{X_B}$ 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.

$$\underline{L} = \|\mathbf{X_B^{l,m}} \underline{X}_G - \mathbf{X_B} S_x^l S_y^m \underline{X}_G\|_2^2 \tag{8}$$

Using the identity:

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

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

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

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:

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

Differentiation of the second term in (4), and also differentiation with respect to the other color bands follow the same technique.

ACKNOWLEDGMENTS

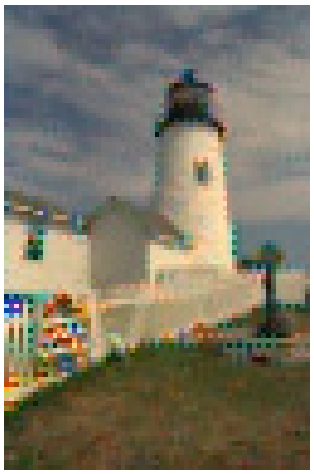
This work was supported in part by the National Science Foundation Grants CCR-9984246, CCR-9971010, AFOSR award F49620-03, and by the National Science Foundation Science and Technology Center for Adaptive Optics, managed by the University of California at Santa Cruz under Cooperative Agreement No. AST - 9876783. Sina Farsiu would like to thank Dirk Robinson, for useful discussions.

REFERENCES

1. S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super-resolution," *Submitted to IEEE Trans. Image Processing* , June 2003.
2. S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Robust shift and add approach to super-resolution," *Proc. of the 2003 SPIE Conf. on Applications of Digital Signal and Image Processing* , Aug. 2003.
3. M. Elad and A. Feuer, "Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images," *IEEE Trans. Image Processing* **6**, pp. 1646–1658, Dec. 1997.
4. T. S. Huang and R. Y. Tsai, "Multi-frame image restoration and registration," *Advances in computer vision and Image Processing* **1**, pp. 317–339, 1984.
5. M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP:Graph. Models Image Process* **53**, pp. 231–239, 1991.
6. S. Peleg, D. Keren, and L. Schweitzer, "Improving image resolution using subpixel motion," *CVGIP:Graph. Models Image Processing* **54**, pp. 181–186, March 1992.
7. D. Capel and A. Zisserman, "Super-resolution enhancement of text image sequences," in *Proceedings International Conference on Pattern Recognition*, pp. 1600–1605, 2000.
8. N. Nguyen, P. Milanfar, and G. H. Golub, "A computationally efficient image superresolution algorithm," *IEEE Trans. Image Processing* **10**, pp. 573–583, Apr. 2001.
9. C. Laroche and M. Prescott, "Apparatus and method for adaptive for adaptively interpolating a full color image utilizing chrominance gradients." United States Patent 5,373,322, 1994.
10. R. Kimmel, "Demosaicing: Image reconstruction from color ccd samples," *IEEE Trans. Image Processing* **8**, pp. 1221–1228, Sept. 1999.
11. D. Keren and M. Osadchy, "Restoring subsampled color images," *Machine Vision and applications* **11**(4), pp. 197–202, 1999.
12. X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Processing* **10**, pp. 1521–1527, Oct. 2001.
13. D. D. Muresan and T. W. Parks, "Optimal recovery demosaicing," in *IASTED Signal and Image Processing*, Aug. 2002.
14. Y. Hel-Or and D. Keren, "Demosaicing of color images using steerable wavelets," Tech. Rep. HPL-2002-206R1 20020830, HP Labs Israel, 2002.
15. S. C. Pei and I. K. Tam, "Effective color interpolation in ccd color filter arrays using signal correlation," *IEEE Trans. Image Processing* **13**, pp. 503–513, june 2003.
16. A. Zomet and S. Peleg, "Multi-sensor super resolution," in *Proceedings of the IEEE Workshop on Applications of Computer Vision*, pp. 27–31, Dec. 2001.
17. W. K. Pratt, *Digital image processing.* John Wiely & Sons, INC., 2001.
18. R. Ramanath, W. E. Snyder, G. L. Bilbro, and W. A. Sander III, "Demosaicking methods for Bayer color arrays," *Electronic Imaging* **11**, pp. 306–315, July 2002.

19. P. Golland and A. M. Bruckstein, "Motion from color," *Computer Vision and Image Understanding* **68**, pp. 346–362, Dec. 1997.
20. J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierachical model-based motion estimation," *Proceedings European Conference on Computer Vision* , pp. 237–252, 1992.
21. R. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint map registration and high-resolution image estimaion using a sequence of undersampled images," *IEEE Trans. Image Processing* **6**, pp. 1621–1633, Dec. 1997.
22. R. R. Schultz, L. Meng, and R. Stevenson, "Subpixl motion estimation for super-resolution image sequence enhancement," *Journal of visal communication and imag representation* **9**, pp. 38–50, Mar. 1998.
23. N. R. Shah and A. Zakhor, "Resolution enhancement of color video sequences," *IEEE Trans. Image Processing* **8**, pp. 879–885, June 1999.
24. C. A. Segall, R. Molina, A. Katsaggelos, and J. Mateos, "Bayesian high-resolution reconstruction of low-resolution compressed video," in *IEEE Int. Conf. on Image Processing*, **2**, pp. 25–28, Oct. 2001.
25. B. C. Tom and A. Katsaggelos, "Resolution enhancement of monochrome and color video using motion compensation," *IEEE Trans. Image Processing* **10**, pp. 278–287, Feb. 2001.

a: Original image



b: Low resolution image

c: Shift and add image

d: Reconst. with luminance regul.

e: Reconst. with orientation regul.

f: Reconst. with chrominance regul.

g: Reconst. from single channel S.R.

h: Reconst. with all terms.

Figure 2. A HR image (a) is passed through our model of camera to produce a set of LR images. One of these LR images is shown in (b). Shift and add of each LR image is resulted in (c). Multi-frame demosaicing of this set of LR frames with the help of only luminance, orientation or chrominance regularization terms is shown in (d), (e), and (f), respectively. The result of super-resolving each color band separately considering only bilateral regularization, is shown in (g). And finally (h) is the result of applying the proposed method to this data set. This paper (with all color pictures) is available at *http://www.ee.ucsc.edu/∼milanfar/publications.htm*

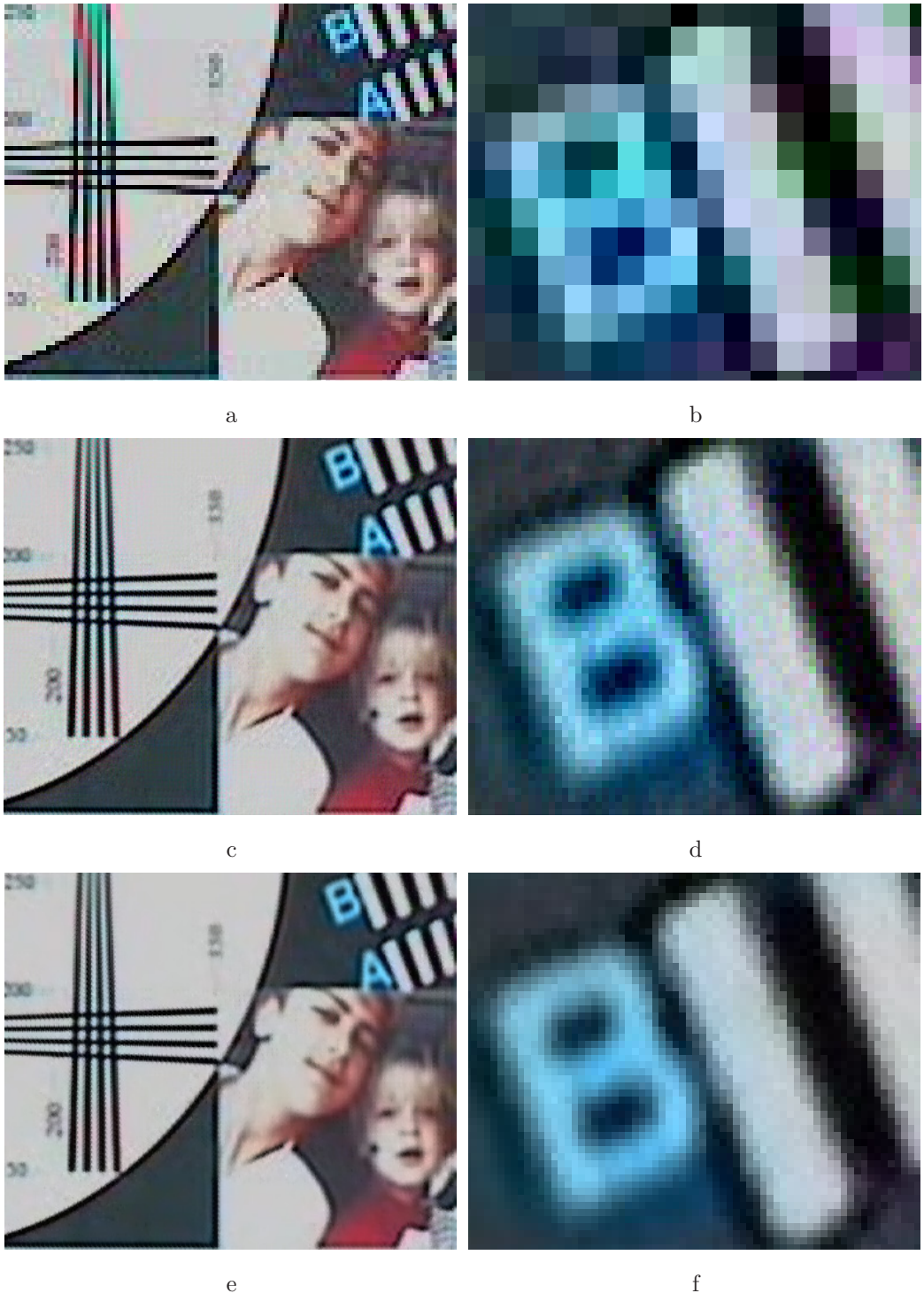


Figure 3. Multi-frame color super-resolution implemented on a real-world data sequence. (a) shows one of the input LR images and (c) is the result of implementing super-resolution on three color channels, using bilateral regularization without considering color band correlation. (e) is the result of implementing the proposed method which has increased the spatial resolution, removed the compression artifacts, and also reduced the color artifacts. Figures (b), (d), and (f) are the zoomed images of the Figures (a), (c), and (e), respectively. This paper (with all color pictures) is available at *http://www.ee.ucsc.edu/∼milanfar/publications.htm*