Chapter 1
Simultaneous demosaicing and resolution enhancement from under-sampled image sequences

SINA FARSIU
Duke University Eye Center
Durham, NC
Email: sina.farsiu@duke.edu

DIRK ROBINSON
Ricoh Innovations Inc.
Menlo Park, CA
Email: dirkr@rii.ricoh.com

MICHAEL ELAD
Computer Science Department
The Technion – Israel Institute of Technology, Haifa, Israel
Email: elad@cs.technion.ac.il

PEYMAN MILANFAR
Electrical Engineering Department,
University of California, Santa Cruz, CA.
Email: milanfar@ee.ucsc.edu

1.1   Challenge of Single-sensor Color Imaging

Single sensor color imaging provides a convenient, yet low cost solution for the problem of capturing images at multiple wavelengths. To capture color images, single-sensor imaging systems add a color filter array (CFA) on top of the photo-detector array. The CFA provides a simple mechanism for sampling of the different color channels in a multiplexed fashion. Figure 1.1 shows an example of the most popular CFA pattern known as the Bayer pattern [1] (design and properties of many CFA patterns are discussed in [2]). As apparent in Fig. 1.1, using a CFA trades off the spatial sampling resolution of the three sensor system, to achieve multispectral sampling. The single sensor approach assumes that due to spatial- and wavelength-correlations, the missing values could be interpolated reliably. Figure 1.2 shows a real image after such interpolation stage, demonstrating the various image quality issues encountered in single-sensor color imaging.

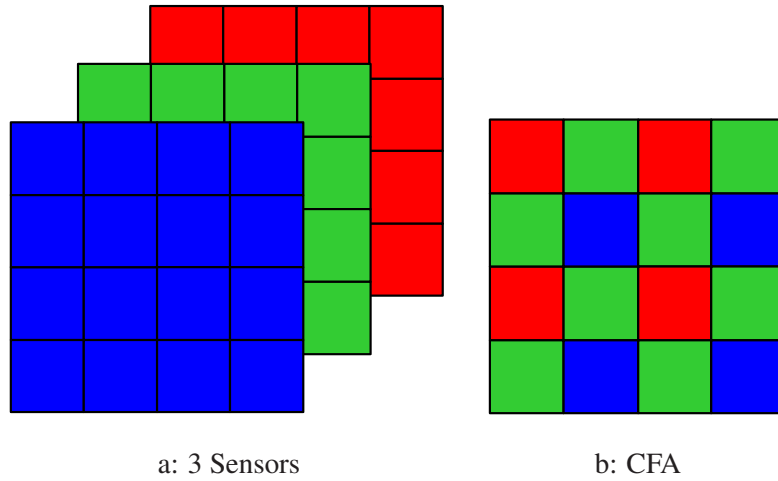a: 3 Sensors   b: CFA

Figure 1.1: Comparison of the more expensive 3CCD sensor arrays (a), with the cheaper CFA sensor (b) configurations.

These shortcomings motivated the development of image processing algorithms to improve image quality for single sensor imaging systems. Arguably, the most powerful image processing methods are the multiframe image processing approaches. The multiframe approaches combine the information from multiple images to produce higher quality images. Figure 1.3 provides a general picture of the multiframe imaging approach to mitigate the shortcomings of imaging systems.

The need to interpolate the missing values in CFA images calls for a model that ties the full color images in the sequence to the measurements obtained. The mathematical model of the forward imaging model for single-sensor imaging is

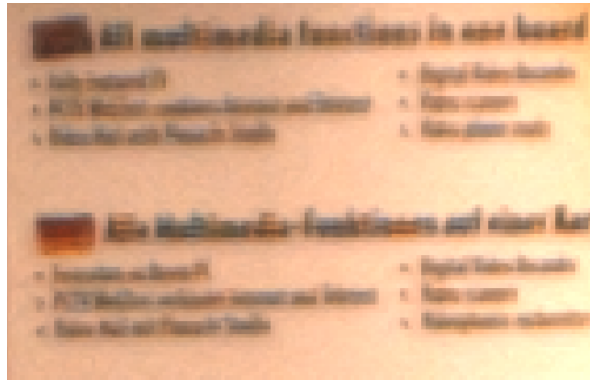$$\mathbf{y}(k) = \mathbf{ADHF}(k)\mathbf{x} + \mathbf{v}(k). \tag{1.1}$$

Figure 1.2: Example showing the typical image quality problems associated with single-sensor color imaging. The image is noisy with evident color artifacts and has limited overall resolution as evidenced by the difficulty in reading the text numbers.
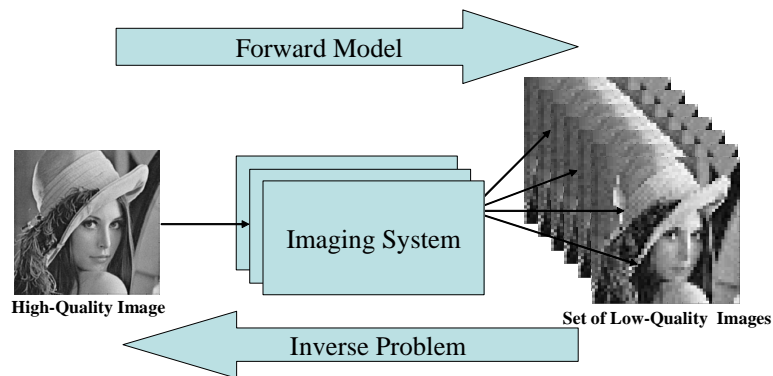

Figure 1.3: A block diagram representation of the multiframe image processing. The forward model is a mathematical description of the image degradation process. The multiframe inverse problem addresses the issue of retrieving (or estimating) the original scene from a *set* of low-quality images.

where $\mathbf{y}$ is a vector containing all of the captured pixel values for a single image. The vector $\mathbf{x}$ contains the pixel values of the unknown high resolution, noise free image. Conceptually, we can consider this vector to be comprised of the three unknown color channel images

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_R \\ \mathbf{x}_G \\ \mathbf{x}_B \end{pmatrix}. \tag{1.2}$$

The matrix $\mathbf{F}$ represents a motion operator capturing the image motion induced by the relative motion between the imaging system and the camera. The matrix $\mathbf{H}$ captures the blurring effects due to the camera optics, and $\mathbf{D}$ represents the downsampling effect due to limited resolution of the sensor array. These three operators ($\mathbf{F}$, $\mathbf{H}$, and $\mathbf{D}$) are applied to each of the color channels, either separately or jointly. In this work

we assume that their effect is separated to each color channel.

The matrix $\mathbf{A}$ represents the sampling effect due to the CFA. This again stands for an operator that operates on the three color channels, sampling each differently, based on the Bayer pattern. The vectors $\mathbf{v}$ represent the additive random noise inherent in the imaging process. For the multiframe imaging case, the variable k indexes the N captured frames. Notice that in our proposed model, only the motion operators are assumed to vary as a function of k, implying that the blur, downsampling, and CFA sampling, are all independent of time. An extension of this model could be proposed, where the various operators vary in time, and operate on the RGB in a mixed way.

Figure 1.4 depicts the image capture model for multiframe image processing.
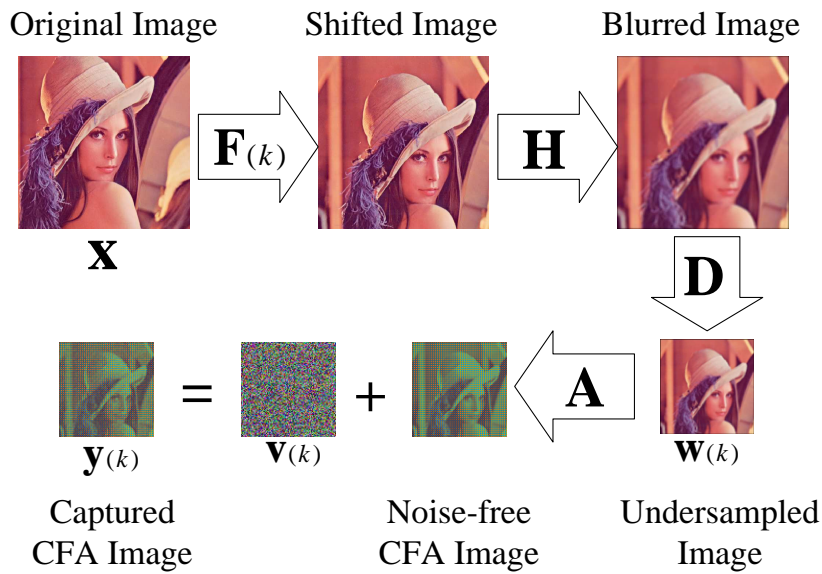

Figure 1.4: Block diagram representing the image formation model considered in this chapter, where $\mathbf{x}$ is the perfect intensity image of the scene, $\mathbf{v}$ is the additive noise, and $\mathbf{y}$ is the resulting color-filtered low-quality image. The operators $\mathbf{F}$, $\mathbf{H}$, $\mathbf{D}$, and $\mathbf{A}$ are representatives of the warping, blurring, down-sampling, and color-filtering processes, respectively.

The goal of multiframe processing is to estimate the high resolution image $\mathbf{x}$ from a collection of low resolution images $\{\mathbf{y}(k)\}$. In general, the problem of multiframe processing for single-sensor imaging systems is complicated. Earlier approaches to this problem sequentially solved single frame demosaicing, followed by multiframe resolution enhancement or super-resolution. The proposed approach, a multiframe demosaicing scheme, solves both problems in a joint fashion offering improved performance. In Sections 1.2 and 1.3 we describe both of these approaches in more detail. In Section 1.4, we describe a fast implementation of the joint approach enabling multiframe enhancement of single-sensor video sequences. Concluding remarks

of this chapter are given in Section 1.5.

1.2 Sequential Demosaicing and Super-Resolution

The classic approach for multiframe processing of single-sensor color images involves two stages of processing. First, single frame demosaicing interpolates the missing color pixel values in the single-sensor CFA imaging. This single frame demosaicing is applied to a collection of images independently, producing a set of full-color low resolution images. Second, these low resolution color images are processed using multi-frame resolution enhancement or *super-resolution* algorithms. The super-resolution algorithms are either applied to each color channel independently [3, 4], or to the luminance (grayscale) components of the color images [5]. This traditional two-step approach is illustrated in Figure 1.5.
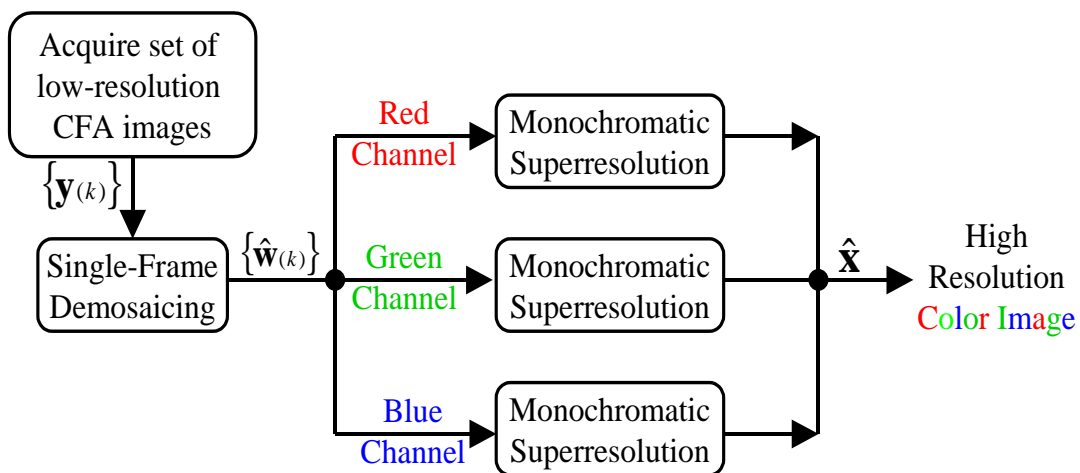

Figure 1.5: Block diagram representing the traditional two-step approach to the multi-frame reconstruction of color images. First, a demosaicing algorithm is applied to each of the CFA images independently. Second, a super-resolution algorithm is applied to each of the resulting color channel image sets independently.

Producing a high resolution image requires a smart combination of data processing and prior knowledge on the spatial and spectral (color) properties of images. The prior information about the images characterizes both the color and the spatial properties of images. In the following sections, we review the various approaches to single-frame demosaicing and super-resolution, including analysis of the different forms of prior information needed to solve these two problems.

1.2.1   Single-frame Demosaicing

Single-frame demosaicing algorithms, or just demosaicing algorithms, attempt to restore the missing color components eliminated by using a color filter array. Essentially, demosaicing algorithms try to estimate a low resolution full-color image $\mathbf{w}(k)$ defined as

$$\mathbf{w}(k) = \mathbf{DHF}(k)\mathbf{x}. \tag{1.3}$$

The full-color image $\mathbf{w}(k)$ has three color values for every pixel location, whereas the captured image $\mathbf{y}(k) = \mathbf{Aw}(k) + \mathbf{v}(k)$ contains only one color value per pixel location. The single-frame demosaicing algorithm is applied to each captured image $\mathbf{y}(k)$ independently.

Numerous demosaicing methods have been proposed over the years to solve this under-determined problem and in this section we review some of the more popular approaches. The simplest approach estimates the unknown pixel values using linear interpolation of the known color values in the neighboring pixel locations. Figure 1.6 illustrates the standard neighborhood interpolation problem of single-frame demosiacing.
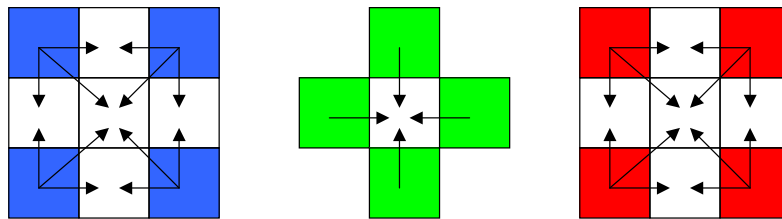
Figure 1.6: The diagram illustrates the basic interpolation problem of single-frame demosaicing. The demosaicing algorithm must estimate the missing color pixel value using the observed neighboring values.

The straightforward interpolation approach ignores some important information about the correlation between the color channel images and often produces serious color artifacts. For example, in [6] it was shown that the Red and Blue channels are related to the Green channel via the approximate equality of the ratios $\frac{Red}{Green}$ and $\frac{Blue}{Green}$. As observed in Fig. 1.6, the Green channel has twice the number of pixels, compared to the Red and Blue channels. Thus, using the correlations between the color channels just described, allows us to provide better interpolation of the Red and Blue channels using the more reliably interpolated Green channel. This approach forms the basis of the smooth-hue transition method first discussed in [6].

Note that this correlation between color channels does not hold across edges in an image. Consequently, although the smooth-hue transition algorithm works for smooth regions of the image, it does not work in the high-frequency (edge) areas. Considering this fact, gradient-based methods, first addressed in [7], do not perform interpolation across the edges of an image. This non-iterative method [7] 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 [8], 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. Both the smooth hue and edge-based methods were later combined in [9]. In this iterative method, the smooth hue interpolation is performed along the local image gradients computed in eight directions about a pixel of interest. A second stage using anisotropic inverse diffusion further enhances the quality of the reconstructed image. This two step approach of interpolation followed by an enhancement step has been used in many other settings. In [10], 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 filter is used as the enhancement step of the method described in [11], that takes advantage of a homogeneity assumption in the neighboring pixels.

Maximum A-Posteriori (MAP) methods form another important category of demosaicing methods. These MAP approaches apply global assumptions about the correlations between the color channels and the spatial correlation using a penalty function of the form

$$\Omega(\mathbf{w}) = J_0(\mathbf{w}, \mathbf{y}) + P(\mathbf{w}) \tag{1.4}$$

where $J_0()$ captures the correspondence between the estimated image $\mathbf{w}$ and the observed data $\mathbf{y}$ and $P()$ captures the quality of the proposed solution $\mathbf{w}$ based on the prior knowledge about the spatial and color correlations in the destination image. For example, a MAP algorithm with a smooth chrominance prior is discussed in [12]. The smooth chrominance prior is also used in [13], where the original image is first transformed to YIQ representation [1]. The chrominance interpolation is performed 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 [15], [16], [17], [18], [19], [20], [21] and [22]. We also note that a few post-demosaicing image enhancement techniques have been proposed (e.g. [23, 24]) to reduce the color artifacts introduced through the color-interpolation process. Almost all of the proposed demosaicing methods are based on one or more of these following assumptions:

1. In the measured image with the mosaic 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 4 Green pixels).

---

[1] YIQ is the standard color representation used in broadcast television (NTSC systems) [14].

2. The CFA pattern is most likely the Bayer pattern having the color sampling geometry shown in Fig. 1.1.

3. For each pixel, one and only one color band value is available.

4. The color pattern of available pixels does not change through the measured image.

5. The human eye is more sensitive to the details in the luminance component of the image than the details in chrominance component [13].

6. The human eye is more sensitive to chromatic changes in the low spatial frequency region than the luminance change [18].

7. Interpolation should be performed along and not across the edges.

8. Different color bands are correlated with each other.

9. Edges should align between color channels.

The accuracy of these single-frame demosaicing algorithms depends to a large degree on the resolution of the single-sensor imaging system. For example, the image Fig.1.7.a shows a high-resolution image captured by a 3-sensor camera. If we had captured this image instead using a single-sensor camera and then applied the single frame demosaicing algorithm of [9], we would obtain the image Fig. 1.7.d. This demosaiced image shows negligible color artifacts. Figure 1.7.b shows the same scene from a simulated 3-sensor camera which is undersampled by a factor of four. Fig. 1.7.e shows the single-frame demosaicing result, where the color artifacts in this image are much more evident than Fig. 1.7.d. In this case, inadequate sampling resolution and the subsequent aliasing produces severe color artifacts. Such severe aliasing happens in cheap commercial still or video digital cameras, with small number of sensor pixels [25].

Some imaging systems attempt to avoid these aliasing artifacts by adding optical low-pass filters to eliminate high spatial frequency image content prior to sampling. Such optical low pass filters not only increase system cost but lower image contrast and sharpness. These optical-low pass filters can remove some, but not all color artifacts from the demosaicing process. Figure 1.7.c shows a simulated low-resolution image which was presmoothed by a Gaussian low-pass digital filter simulating the effects of an optical low-pass filter prior to downsampling. The demosaiced version of this captured image is shown in Fig. 1.7.f. The demosaiced image contains fewer color artifacts than Fig. 1.7.e, however it has lost some high-frequency details.

The residual aliasing artifacts amplified by the demosaicing algorithm motivated the application of multiframe resolution enhancement algorithms. We describe these algorithms in the next section.

| a: Original | b: Down-sampled | c: Blurred and down-sampled |

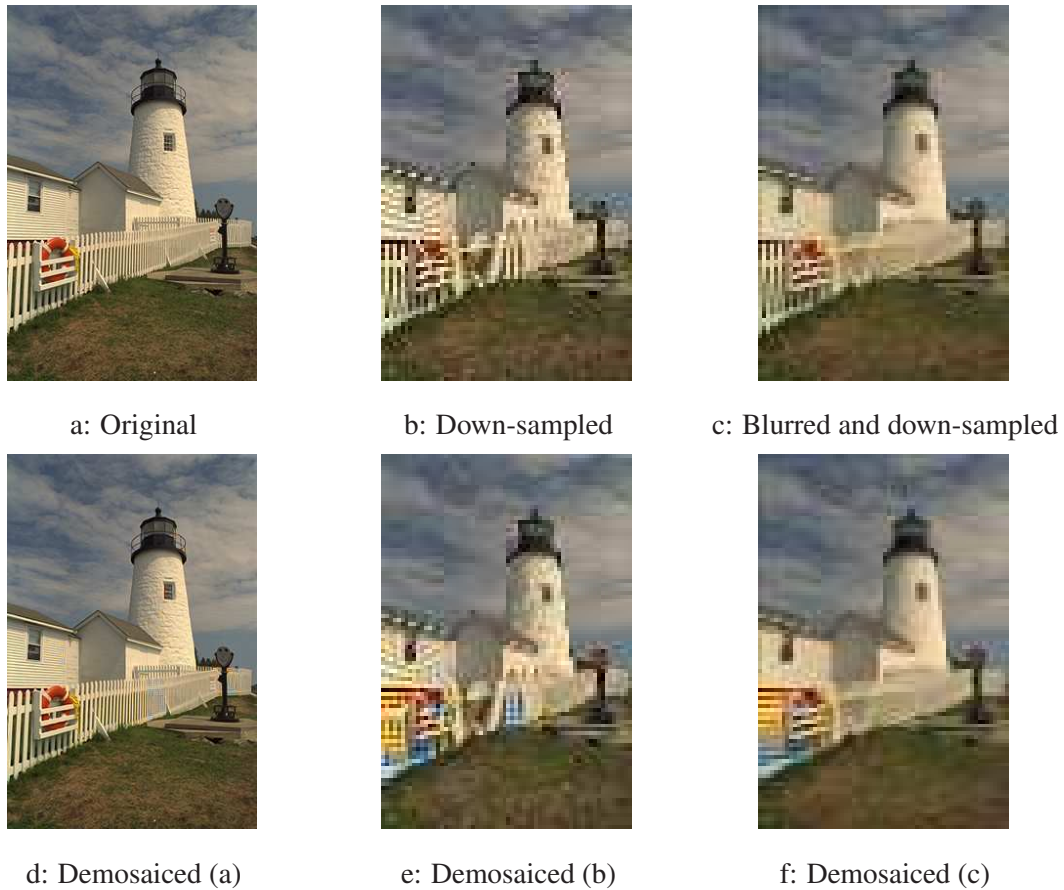| d: Demosaiced (a) | e: Demosaiced (b) | f: Demosaiced (c) |

Figure 1.7: A high-resolution image (a) captured by a 3-CCD sensor camera is down-sampled by a factor of four (b). In (c) the image in (a) is blurred by a Gaussian kernel before down-sampling by a factor of 4. The images in (a), (b), and (c) are color-filtered and then demosaiced by the method of [9]. The results are shown in (d), (e), (f), respectively (©[2007] IEEE).

1.2.2 Multiframe Super-Resolution

A relatively recent approach to dealing with aliasing in imaging systems is to combine multiple aliased images with some phase modulation between them to reconstruct a single high resolution image free of (or with significantly less) aliasing artifacts with improved SNR[2]. Such methods have been studied extensively for monochromatic imaging [26, 27, 28] and are commonly known as multiframe resolution enhancement (super-resolution) algorithms, which produce sharp images with less noise and higher *spatial resolution* than the capture images.

The standard model for such algorithms assumes that a single monochromatic low-resolution aliased image in a set of N images is defined by

$$\mathbf{m}(k) = \mathbf{DHF}(k)\mathbf{g} + \mathbf{v}(k), \quad k = 0 \ldots N - 1 \tag{1.5}$$

---

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

where $\mathbf{m}(k)$ is the kth captured monochrome image. In this case, $\mathbf{g}$ represents the unknown high resolution monochromatic image. In the context of the sequential multiframe demosaicing approach, the captured images $\mathbf{m}(k)$ are obtained by extracting one of the color channels from the demosaiced images. In other words, $\mathbf{m}(k)$ is either $\hat{\mathbf{w}}_R(k)$ or $\hat{\mathbf{w}}_G(k)$ or $\hat{\mathbf{w}}_B(k)$ and $\mathbf{g}$ is either $\mathbf{x}_R$, or $\mathbf{x}_G$, or $\mathbf{x}_B$. As before, $\mathbf{F}(k)$ is a warping operator capturing the relative motion between the kth frame and the anchor frame $k = 0$ and $\mathbf{v}(k)$ represents the additive noise in the image system. The matrix $\mathbf{H}$ represents the blurring associated with the optical point spread function and $\mathbf{D}$ represents the downsampling operator reflecting the limited spatial resolution associated with the image sensor. In effect, the downsampling operator is the cause of the aliasing artifacts observed in the final image. While we do not address the motion estimation in full detail in this chapter, Appendix .1 provides a brief discussion about this topic, with an emphasis on the need to estimate the cross-motion between the N frames jointly to avoid accumulated error.

Early studies of super-resolution showed that the aliasing artifacts in the low-resolution images enable the recovery of the high-resolution fused image, provided that a relative sub-pixel motion exists between the under-sampled input images [29]. Figure 1.8 shows an example of the resulting image after applying super-
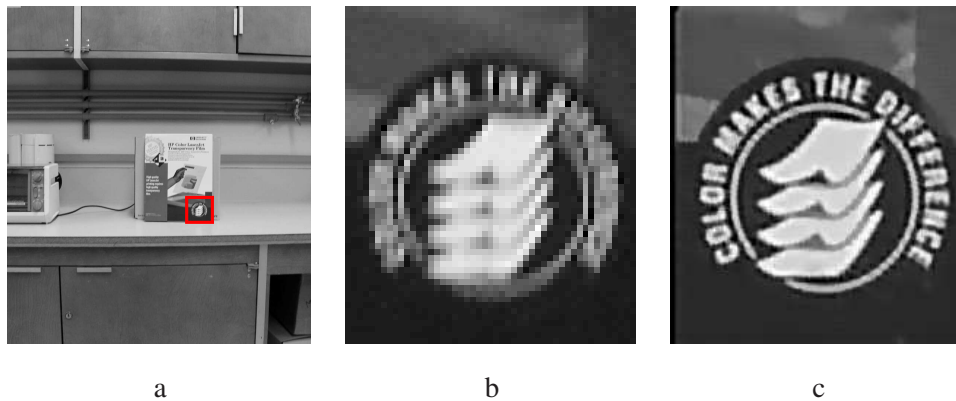

a   b   c

Figure 1.8: super-resolution experiment on real image data. A set of 26 low quality images were combined to produce a higher quality image. One captured image is shown in (a). The Red square section of (a) is zoomed in (b). Super-resolved image in (c) is the high quality output image.

resolution processing to a collection of images taken by a commercial digital still camera. The application of super-resolution is not restricted to optical imaging. An example is illustrated in Figure 1.9, where a collection of 17 images captured by a Siemens digital X-ray imaging system were fused to create the shown high resolution X-ray image. Figure 1.9.a shows one of the input images, a selected region of which is zoomed in Figure 1.9.b for a closer examination. A factor of three superresolved image of 1.9.a is zoomed in Figure 1.9.c, showing a significant reduction in aliasing artifacts.

The multi-frame super-resolution problem was first addressed in [29], where the authors proposed a
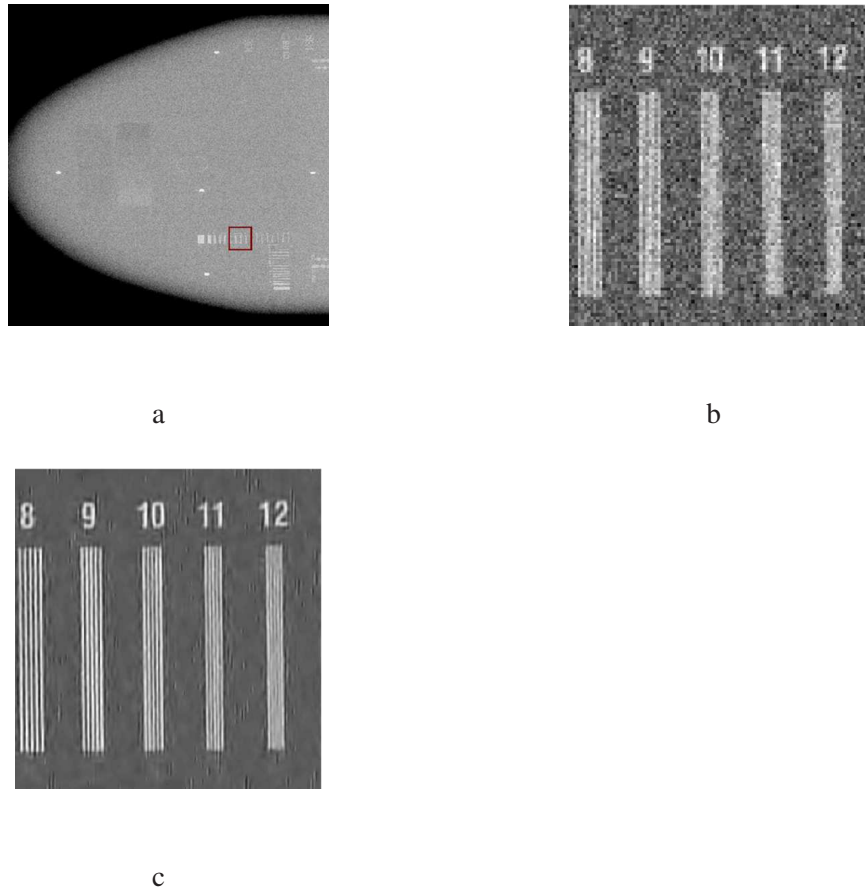
a

b

c

Figure 1.9: Super-resolution experiment on real image data. A set of 17 low quality images from a digital X-ray machine were combined to produce a higher quality image. One captured image is shown in (a). The Red square section of (a) is zoomed in (b). Super-resolved image in (c) is the high quality output image.

frequency domain approach, later extended by others [30]. Although the frequency domain methods are computationally cheap, they are very sensitive to noise and modeling errors [26]. Also, by design, these approaches handle only pure translational motion.

Another popular class of methods solves the problem of resolution enhancement in the spatial domain. The non-iterative spatial domain data fusion approaches are studied in [31], [32] and [33]. The iterative methods are based on construction of a cost function, which is minimized in an iterative function minimization approach. The cost function is of the form

$$\Omega(\mathbf{g}) = J_0(\mathbf{g}, \{\mathbf{m}(k)\}) + P(\mathbf{g}) \tag{1.6}$$

where $J_0(\mathbf{g}, \{\mathbf{m}(k)\})$ is the data penalty term that measures the closeness of the observed data $\{\mathbf{m}(k)\}$ to the predicted measurements using an estimate of the image $\mathbf{g}$. The $P(\mathbf{g})$ term represents the prior information about the unknown signal $\mathbf{g}$, as described before.

The most common data fidelity term is based on minimizing the L_2 difference between the observed data and the predicted data for an estimated image $\mathbf{g}$. The cost function looks like

$$J_0(\mathbf{g}, \{\mathbf{m}(k)\}) = \sum_{k=0}^{N-1} \|\mathbf{DHF}(k)\mathbf{g} - \mathbf{m}(k)\|_2^2. \qquad (1.7)$$

The solution to this penalty function is equivalent to the Least Squares estimate of the image $\mathbf{g}$, and is the optimal Maximum Likelihood (ML) solution [34] when noise follows the white Gaussian model, and when the motion vectors implicit in $\mathbf{F}(k)$ are assumed to be known exactly, and a-priori.

One distinct advantage of the spatial-domain approaches is the ability to incorporate complicated prior information into the estimation process. The most common form of prior information arises as a type of *regularization* to the commonly ill-conditioned or ill-posed estimation problem. This regularization or prior information takes the form of some assumptions about the spatial variability of the unknown high resolution image $\mathbf{g}$. Many of the early super-resolution algorithms, however, applied rudimentary regularization such as Tikhonov regularization [35], which describes the smoothness of the image $\mathbf{g}$. A Tikhonov regularization has the form

$$P(\mathbf{g}) = \lambda\|\Lambda\mathbf{g}\|_2^2 \qquad (1.8)$$

where Λ represents a spatial high-pass operator applied to the image $\mathbf{g}$. In other words, the estimated image will have a reduced high spatial frequency content, depending on the magnitude of the weighting parameter λ.

The early iterative algorithms used search techniques to minimize cost functions. The simplest approach to minimize a cost function of the form of (1.6) is to use the Steepest Descent (SD) algorithm. In SD, the derivative of (1.6) is computed with respect to the unknown parameter vector $\mathbf{g}$ to update the estimate. The estimate for the $n + 1$th iteration is updated according to

$$\widehat{\mathbf{g}}^{n+1} = \widehat{\mathbf{g}}^n - \beta\nabla_x\Omega(\widehat{\mathbf{g}}^n) \qquad (1.9)$$

where β is the SD step size. This approach is similar to the iterative back-projection method developed in [36] and [5].

These spatial domain methods discussed so far were often computationally expensive, especially when solved via explicit construction of the modeling matrices of Eq. 1.5. The authors of [37] introduced a block circulant preconditioner for solving the Tikhonov regularized super-resolution problem formulated in [35], and addressed the calculation of regularization factor for the under-determined case[3] by generalized cross-validation in [38].

---

[3]where the number of non-redundant low-resolution frames is smaller than the square of resolution enhancement factor. A resolution enhancement factor of r means that low-resolution images of dimension $Q_1 \times Q_2$ produce a high-resolution output of dimension $rQ_1 \times rQ_2$. Scalars Q_1 and Q_2 are the number of pixels in the vertical and horizontal axes of the low-resolution images, respectively.

Later, a very fast super-resolution algorithm for pure translational motion and common space invariant blur was developed in [32]. This fast approach was based on the observation that the deblurring can be separated from the merging process of the low-resolution images; and that merging of the images is obtained simply by Shift-and-Add operation. Using these, it was shown that the result is nevertheless ML optimal. Another interesting observation in this work is the fact that the matrices $\mathbf{H}$, Λ, $\mathbf{D}$, etc., can be applied directly in the image domain using standard operations such as convolution, masking, down-sampling, and shifting. Implementing the effects of these matrices as a sequence of operators on the images eliminates the need to explicitly construct the matrices. As we shall see momentarily, this property helps us develop fast and memory efficient algorithms.

A different fast spatial domain method was recently suggested in [39], where low-resolution images are registered with respect to a reference frame defining a nonuniformly spaced high-resolution grid. Then, an interpolation method called Delaunay triangulation is used for creating a noisy and blurry high-resolution image, which is subsequently deblurred. All of the above methods assumed the additive Gaussian noise model. Furthermore, regularization was either not implemented or it was limited to Tikhonov regularization. Another class of super-resolution techniques use the projection onto convex sets (POCS) formulation [40, 41, 42]. The POCS approach allows complicated forms of prior information such as set inclusions to be invoked.

Recently, a powerful class of super-resolution algorithms leverage sophisticated *learned*-based prior information. In these algorithms, the prior penalty functions $P(\mathbf{g})$ are derived from collections of training image samples [43, 44, **?**, **?**]. For example, in [44] an explicit relationship between low-resolution images of faces and their known high-resolution image is learned from a face database. This learned information is later used in reconstructing face images from low-resolution images. Due to the need for gathering a vast number of examples, often these methods are only effective when applied to very specific scenarios, such as faces or text.

Another class of super-resolution algorithms focus on being robust to modeling errors and different noise. These methods implicitly [45] or explicitly [46, 47] take advantage of L_1 distance metrics when constructing the cost functions. Compared to (1.7), the methods of [46, 47] apply the L_1 metric as the data fidelity term

$$J_0(\mathbf{g}, \{\mathbf{m}(k)\}) = \sum_{k=0}^{N-1} \|\mathbf{DHF}(k)\mathbf{g} - \mathbf{m}(k)\|_1, \tag{1.10}$$

which is the optimal ML solution in presence of additive white Laplacian noise.

In [46], in the spirit of the total variation criterion [48], [49] and a related method called the bilateral filter [50], [51], the L_1 distance measure was also incorporated into a prior term called the Bilateral Total

Variation (BTV) regularization penalty function, defined as

$$P(\mathbf{g}) = \lambda \sum_{l=-L_{max}}^{L_{max}} \sum_{m=-M_{max}}^{M_{max}} \alpha^{|m|+|l|} \|\mathbf{g} - \mathbf{S}_x^l \mathbf{S}_y^m \mathbf{g}\|_1. \tag{1.11}$$

where S_x^l and S_y^m are the operators corresponding to shifting the image represented by $\mathbf{g}$ by l pixels in the horizontal and m pixels in the vertical directions, respectively. The parameters M_{max},L_{max} define the size of the corresponding bilateral filter kernel. The scalar weight α, $0 < \alpha \leq 1$, is applied to give a spatially decaying effect to the summation of the regularization terms.

Alternative robust super-resolution approaches were presented in [52, 53] based on normalized convolution and non-parametric kernel regression techniques. To achieve robustness with respect to errors in motion estimation, the recent work of [54] proposed a novel solution based on modifying camera hardware. Finally, in [41, 55, 56] the super-resolution framework has been applied to reduce the quantization noise resulting from video compression. More comprehensive surveys of the different monochromatic multi-frame super-resolution methods can be found in [26, 57, 58, 59].

1.2.3   Sequential Processing Example

Figure 1.10 shows an example of the sequential multiframe processing. In this example, ten CFA images were synthetically generated using the high resolution lighthouse image shown in Fig. 1.7.a. To produce the simulated CFA images, each colorband of the high resolution image was blurred by a 5×5 pixel Gaussian blurring kernel with unit standard deviation. Then, the ten images were taken from this blurry image by subsampling by a factor of four starting at randomly offset pixels. White Gaussian noise was added to each of the simulated low resolution images such that the image had an SNR of 30dB. Each of the simulated CFA images was processed by the single frame demosaicing algorithm described in [9].

Fig. 1.10.a shows an example low resolution image after single frame demosaicing. Then, the robust monochromatic super-resolution algorithm of [46] was applied to each of the three color channels independently. The resulting high resolution image is shown in Fig. 1.10.b, which shows improvement in resolution and noise reduction. The high resolution image still, however, contains some color artifacts and has oversmoothing in some regions. The weakness of these results motivated the use of multiframe demosaicing algorithms, an example of which is shown in Fig. 1.10.c. These techniques are described in details in the next section.

1.3   Multiframe Demosaicing

The sequential multiframe processing approach works reasonably well when the amount of aliasing artifacts is small. With significant color aliasing artifacts, however, the sequential approach fails to adequately correct

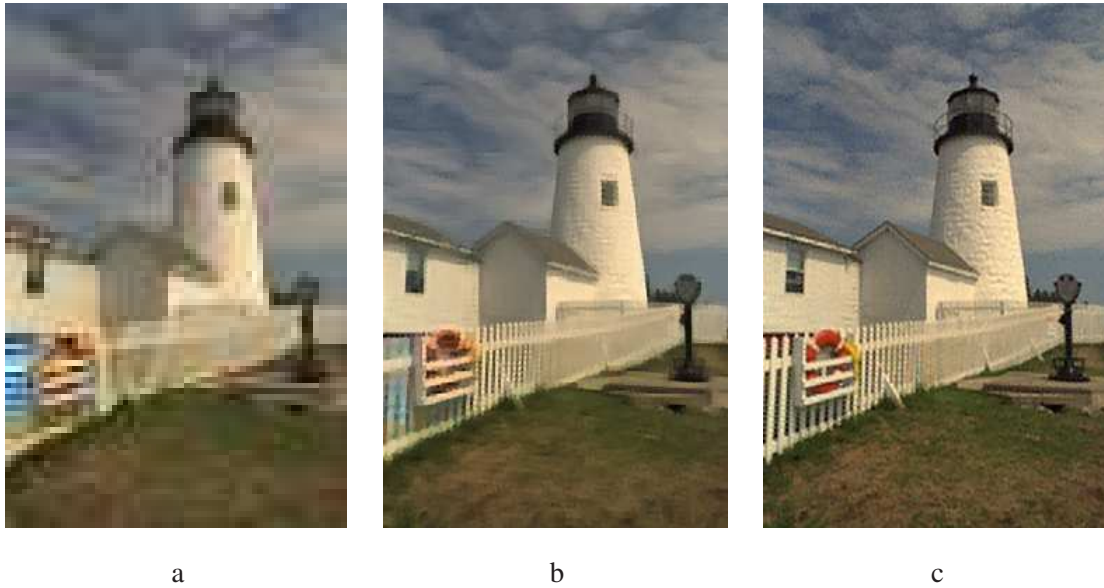a   b   c

Figure 1.10: The image (a) shows an example low resolution CFA image processed by the single frame demosaicing algorithm of [9]. The image (b) shows the resulting color image after applying the robust super-resolution algorithm [46] to each of the color channels independently completing the sequential multiframe processing algorithm, which shows improved resolution with significantly reduced artifacts. Some residual color artifacts, however, are apparent along the fence. Better reconstruction results are shown in (c), where the multiframe demosaicing method of Section 1.3, has fused the information of ten CFA frames (©[2007] IEEE).

all of the color artifacts. Basically, the failure is traced to the fact that all of the color prior information is applied at the individual (single) frame processing level during demosaicing. A more complicated, yet robust method performs the demosaicing and super-resolution *simultaneously*. This approach is known as *multiframe demosaicing* and has been developed relatively recently [60, 61, 62, 63, 64, 65, 66].

1.3.1 Modeling the Problem

The multiframe demosaicing approach tries to estimate the full-color high resolution image $\mathbf{x}$ using the the *set* of unprocessed CFA images $\mathbf{y}(k)$. In this case, the complete forward data model given by (1.1) is used to reconstruct the full-color high resolution image $\mathbf{x}$.

The generic MAP multiframe demosaicing approach estimates $\mathbf{x}$ by minimizing a cost function of the form

$$\Omega(\mathbf{x}) = J_1(\mathbf{x}, \{\mathbf{y}(k)\}) + P_1(\mathbf{x}) + P_2(\mathbf{x}) + P_3(\mathbf{x}). \tag{1.12}$$

This cost function is similar to the super-resolution cost function of (1.6) with the exception that now we have three prior information terms. These multiple terms include both the spatial and color prior information

constraints. In [64], the MAP-based penalty terms are described as:

1. $J_1()$: A penalty term to enforce similarities between the raw data and the high-resolution estimate (Data Fidelity Penalty Term).

2. $P_1()$: A penalty term to encourage sharp edges in the luminance component of the high-resolution image (Spatial Luminance Penalty Term).

3. $P_2()$: A penalty term to encourage smoothness in the chrominance component of the high-resolution image (Spatial Chrominance Penalty Term).

4. $P_3()$: A penalty term to encourage homogeneity of the edge location and orientation in different color bands (Inter-Color Dependencies Penalty Term).

These terms combine several forms of prior information, forming a powerful constraint when performing multiframe demosaicing. We shall describe these terms in details below:

- **Color Data Fidelity Penalty Term:** In the multiframe demosaicing case, the data fidelity term must include all three color channels. Considering the general motion and blur model of (1.1), a reasonable L_2-based multiframe data fidelity penalty term is given by:

$$J_1(\mathbf{x}, \{\mathbf{y}(k)\}) = \sum_{i=R,G,B} \sum_{k=0}^{N-1} \|\mathbf{ADHF}(k)\mathbf{x}_i - \mathbf{y}_i(k)\|_2^2. \qquad (1.13)$$

Essentially, this functional penalizes estimates of the high resolution color image that, when passed through the forward model of Eq. 1.1, differs from the observed image.

The data penalty function Eq. 1.13 works well when the model is accurate, e.g., the motion estimation produces accurate estimates of the forward imaging model, and the blur is the correct one. When these are not true, the forward model contains errors, and a more robust form of Eq. 1.13 could be proposed, based on the L_1 norm:

$$J_1(\mathbf{x}, \{\mathbf{y}(k)\}) = \sum_{i=R,G,B} \sum_{k=0}^{N-1} \|\mathbf{ADHF}(k)\mathbf{x}_i - \mathbf{y}_i(k)\|_1. \qquad (1.14)$$

This is the full-color generalization of the robust monochromatic super-resolution penalty term of (1.10). As before, this penalty term provides robustness at the expense of slight increase in computational complexity, and reduced denoising efficiency when applied to Gaussian noise.

- **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 [13]. It is important, therefore, that the edges in the luminance component of the reconstructed high-resolution image look sharp. As

explained in [46], the bilateral total variation (BTV) functional of (1.11) produces images with sharp edges. For the multiframe demosaicing application, we apply the BTV cost function to the reconstruction of the luminance component. The luminance image can be calculated as the weighted sum $\mathbf{x}_L = 0.299\mathbf{x}_R + 0.597\mathbf{x}_G + 0.114\mathbf{x}_B$ as explained in [14]. This luminance image is the Y component of the YIQ image format commonly found in video coding. The BTV luminance regularization term is then defined as

$$P_1(\mathbf{x}) = \lambda_1 \sum_{l=-L_{max}}^{L_{max}} \sum_{m=-M_{max}}^{M_{max}} \alpha^{|m|+|l|} \|\mathbf{x}_L - \mathbf{S}_x^l \mathbf{S}_y^m \mathbf{x}_L\|_1. \tag{1.15}$$

where the λ_1 defines the strength of the image prior.

- **Spatial Chrominance Penalty Term:** Spatial regularization is required also for the chrominance channels. However, since the human visual system is less sensitive to the resolution of these bands, we can apply a simpler regularization functional, based on the L_2 norm, in the form of

$$P_2(\mathbf{x}) = \lambda_2 \left(\|\Lambda \mathbf{x}_I\|_2^2 + \|\Lambda \mathbf{x}_Q\|_2^2 \right), \tag{1.16}$$

where the images $\mathbf{x}_I$ and $\mathbf{x}_Q$ are the I and Q portions of the YIQ color representation and λ_2 defines the strength of the image prior. As before, Λ is a spatial high-pass operator such as derivative, Laplacian, or even identity matrix. Such penalty term encourages the two chrominance components to vary smoothly over the high resolution image.

- **Inter-Color Dependencies Penalty Term:** This penalty term characterizes the correlation between the original three color channel images (RGB). This penalty function minimizes the mismatch between locations or orientations of edges across the color bands. Following [12], minimizing the vector product norm of any two adjacent color pixels forces different bands to have similar edge location and orientation. Based on the theoretical justifications of [67], the authors of [12] suggest a pixelwise inter-color dependencies cost function to be minimized. This term has the vector outer product norm of all pairs of neighboring.

With some modifications to what was proposed in [12], our inter-color dependencies penalty term is a differentiable cost function

$$P_3(\mathbf{x}) = \lambda_3 \sum_{l,m=-1}^{1} \left[\|\mathbf{x}_G \odot \mathbf{S}_x^l \mathbf{S}_y^m \mathbf{x}_B - \mathbf{x}_B \odot \mathbf{S}_x^l \mathbf{S}_y^m \mathbf{x}_G\|_2^2 + \right.$$
$$\left. \|\mathbf{x}_B \odot \mathbf{S}_x^l \mathbf{S}_y^m \mathbf{x}_R - \mathbf{x}_R \odot \mathbf{S}_x^l \mathbf{S}_y^m \mathbf{x}_B\|_2^2 + \|\mathbf{x}_R \odot \mathbf{S}_x^l \mathbf{S}_y^m \mathbf{x}_G - \mathbf{x}_G \odot \mathbf{S}_x^l \mathbf{S}_y^m \mathbf{x}_R\|_2^2 \right], \tag{1.17}$$

where $\odot$ is the element by element multiplication operator. The λ_3 defines the strength of the image prior.

As with the monochromatic super-resolution case, we minimize the generic full-color cost function (1.12) using a type of SD. In each step, the derivative will be computed with respect to one color channel assuming the other color channels are fixed. Thus, for the $n + 1$ iteration, the algorithm updates the ith color channel according to

$$\widehat{\mathbf{x}}_i^{n+1} = \widehat{\mathbf{x}}_i^n - \beta \nabla_{x_i} \Omega(\mathbf{x}^n) \qquad i = R, G, B \tag{1.18}$$

where the scalar β is the steepest descent step size. This minimization approach converges relatively quickly for most image sequences, providing satisfactory results after 10-15 such iterations.

1.3.2   Multiframe Demosaicing Examples

In this section, we present some visual results of the multiframe demosaicing algorithm described in the previous section. The first example demonstrates the advantages of the multiframe demosaicing algorithm when dealing with raw CFA images taken directly from an image sensor. In this example, we acquired 31 uncompressed raw CFA images from a 2 megapixel CMOS sensor. The (unknown) camera point spread function (PSF) was modeled as a tapered 5×5 disk PSF. In all examples, to compute the unknown motion matrices, the raw CFA images were first demosaiced by the single-frame method of [7]. Then, the luminance component of these images was registered in a pairwise fashion using the motion estimation algorithm[4] described in [69].

Fig. 1.11.a shows a single low-resolution image after applying the single-frame demosaicing method of [7]. The single-frame demosaicing method exhibits the typical problems of low resolution, color aliasing, and sensor noise. Fig. 1.11.b shows the resulting image after applying the multiframe demosaicing algorithm directly to the set of low-resolution CFA images to increase the spatial resolution by a factor of three.

In the second example, we repeated the above experiment using a different camera, capturing 26 color filtered images. Fig. 1.12.a shows a single low-resolution image after applying the more complex single-frame demosaicing method of [9] and Fig. 1.12.b shows the resulting image after applying the multiframe demosaicing algorithm directly to the set of low-resolution CFA images to increase the spatial resolution by a factor of three. Both above examples (Fig. 1.11.b and Fig. 1.12.b) show improved resolution and almost no color artifacts.

In the third example, we apply the multiframe demosaicing algorithm to a set of color images which have already undergone single-frame demosaicing and compression. Fig. 1.13.a shows an example low resolution image which has undergone an unknown demosaicing algorithm followed by compression. Forty such images were combined using the multiframe demosaicing algorithm to produce the image on the right.

---

[4]Application of an accurate multiframe registration technique (Appendix.1) for directly registering the CFA data is described in [68].

a   b

Figure 1.11: Multi-frame color super-resolution applied to a real data sequence. (a) shows one of the 31 low-resolution input images (of size $[141 \times 147 \times 3]$) demosaiced by the single-frame method of [7]. (b) shows the resulting image after multiframe demosaicing using the robust data penalty term of (1.14). The resulting image shows improved resolution and reduced color aliasing artifacts (Ⓒ[2007] IEEE).

This resulting image demonstrates the capability of the multiframe demosaicing approach to handle images later in the imaging pipeline after standard compression. While the results would presumably be better if the multiframe algorithm had direct access to the raw CFA images, the final image shows significant improvement over the input image.

1.4   Fast and Dynamic Multiframe Demosaicing

In this section, first we describe the conditions under which the multiframe demosaicing algorithm described in the previous section may be implemented very efficiently (Section 1.4.1). Using this fast formulation, in Section 1.4.2 we extend the multiframe demosaicing algorithm to *dynamic* multiframe processing. Dynamic multiframe demosaicing differs from basic multiframe demosaicing in that the final result is a high resolution image sequence or video. This is also known as video-to-video super-resolution, which has been addressed in the literature for the grayscale [70, 71, 72] and color (demosaiced) [73] sequences.

1.4.1   Fast Multiframe Demosaicing

While the iterative method described in the previous section is fast enough for many practical applications, speeding up the implementation is desirable when handling large amounts of data. In this section, we
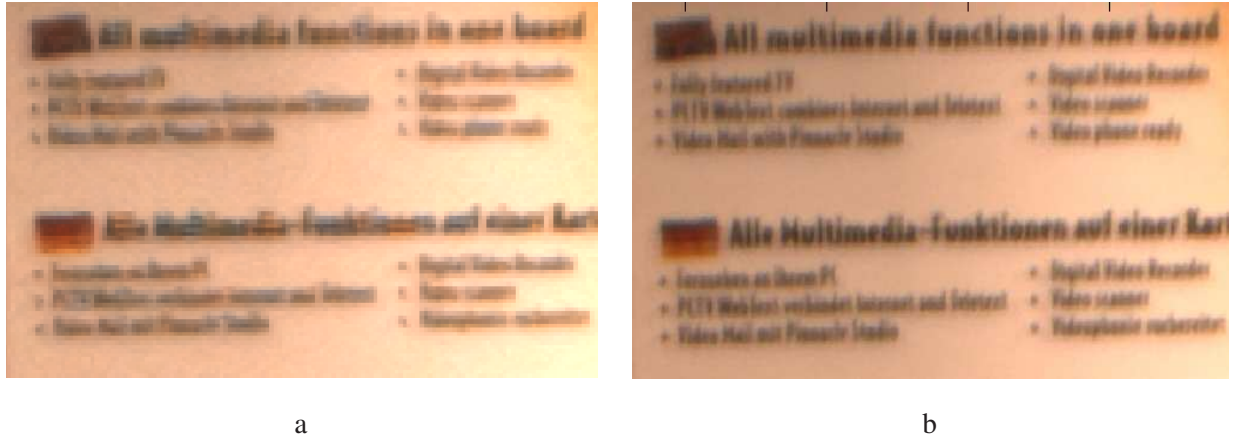
Figure 1.12: Multi-frame color super-resolution applied to a real data sequence. (a) shows one of the 26 low-resolution input images demosaiced by the single-frame method of [9] . (b) shows the resulting image after multiframe demosaicing using the robust data penalty term of (1.14). The resulting image shows improved resolution and reduced color aliasing artifacts.

describe a fast, two-stage approach to multiframe demosaicing, in the spirit of the method developed in [74]. The two stage approach works only when two conditions are satisfied. First, the common system PSF is spatially invariant. Second, the motion between the low-resolution frames (at least locally) is translational in nature. When these two conditions are satisfied, the shifting operator $\mathbf{F}(k)$ and the blurring operator $\mathbf{H}$ commute. We then define $\mathbf{z} = \mathbf{H}\mathbf{x}$ as the unknown high-resolution blurry image. In the fast approach, multiframe demosaicing is divided into the following two steps:

1. Non-iterative data fusion (Shift-and-Add – see below) provides an estimate of $\mathbf{z}$ from the captured data. The model for this step is

$$\mathbf{y}(k) = \mathbf{A}\mathbf{D}\mathbf{F}(k)\mathbf{z} + \mathbf{v}(k) \tag{1.19}$$

2. Iterative deblurring and interpolation provides an estimate of $\mathbf{x}$ from the estimate $\hat{\mathbf{z}}$.

The two-step approach is fast since the first stage is non-iterative. After this stage, we no longer need to store the set of capture images $\{\mathbf{y}(k)\}$. In essence, the estimate $\hat{\mathbf{z}}$ is a sufficient statistic with which we can estimate the high resolution image $\mathbf{x}$.

Data Fusion Step

The data fusion step solves the estimation problem for the model (1.19) in a non-iterative fashion for each color channel independently by using an analytic minima of a simple data penalty cost function [46]. If the
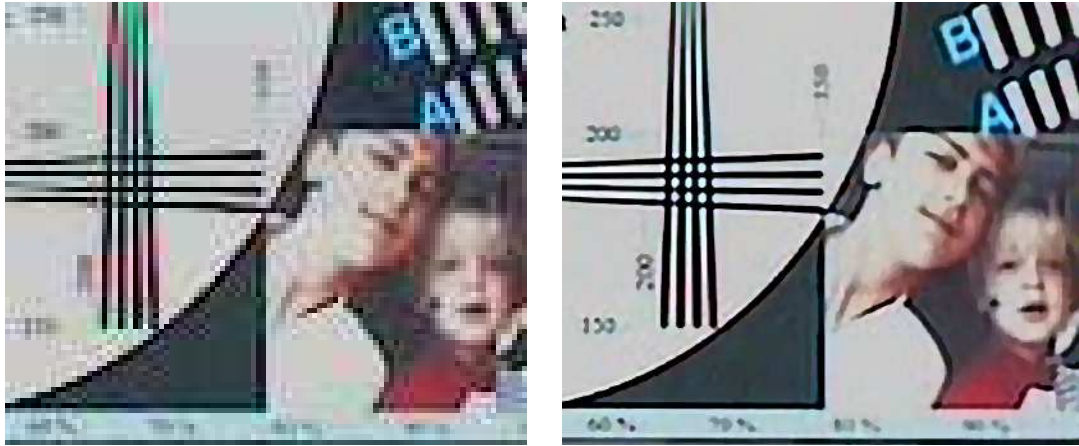
Figure 1.13: (a) shows an example of one of the 40 captured images which are demosaiced and compressed in an unknown fashion by the imaging system. These images were used to reconstruct the higher resolution image in (b) exploiting the multiframe demosaicing algorithm (©[2007] IEEE).

penalty function is based on the L_2 norm penalty function

$$\sum_k \|\mathbf{y}_i(k) - \mathbf{DF}(k)\mathbf{z}_i\|_2^2, \quad i = R, G, B \tag{1.20}$$

the fusion step is called the Shift-and-Add algorithm. In this algorithm, the input images are upsampled by zero-padding, shifted by the inverse of the translation amount, and averaged over the set of N images. Figure 1.14 shows a visual example of the Shift-and-Add process for one color channel. This is equivalent
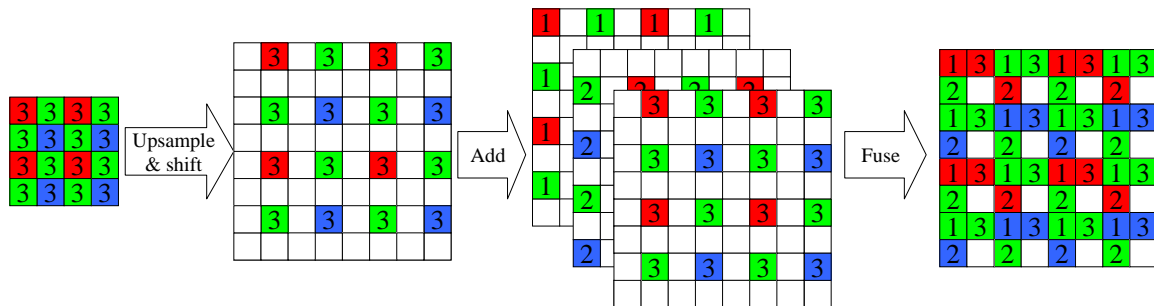

Figure 1.14: Diagram showing an example of the Shift-and-Add process. The input color channel image is upsampled by the resolution enhancement factor $r = 2$, and shifted according to the inverse of the image translation. This shifted image is added and averaged with the other $N - 1$ (in this case 2) upsampled and shifted low resolution images.

to the ML estimate of the image $\mathbf{z}$. As shown in [46], if the robust L_1 data fidelity

$$\sum_k \|\mathbf{y}_i(k) - \mathbf{DF}(k)\mathbf{z}_i\|_1 \quad i = R, G, B \tag{1.21}$$

term is used, then a pixel-wise median operation over the set of frames implements the robust penalty function of Eq. 1.14. This is called the Robust Shift-and-Add algorithm, the result of which is a single image containing estimates of the blurry image **z**.

As apparent in the fused image in Fig. 1.14, the color sampling pattern after Shift-and-Add can be quite arbitrary depending on the relative motion of the low-resolution images. For some pixels we may not have any value, while for some we might have multiple measurements of one, two, or even all three color bands. Because this sampling pattern is arbitrary, we rely on the iterative deblurring and interpolation algorithm to restore the missing information. As an aside, Fig. 1.14 partly explains the inferiority of the sequential strategy of first demosaicing the low resolution images followed by super-resolution. Image motion between captured frames will often provide data in some of the pixel locations lacking color data in a single CFA image. Interpolating the missing pixels for each individual frame prematurely estimates the missing pixel data thereby limiting the final performance of multiframe resolution enhancement.

Deblurring and Interpolation Step

The second step uses the estimate of the blurry high resolution image $\hat{\mathbf{z}}$ to estimate the high resolution image **x**. This step relies on a cost function very similar to that of Eq. 1.12 with the exception that the data penalty term is replaced by

$$\Omega(\mathbf{x}, \hat{\mathbf{z}}) = \sum_{i=R,G,B} \|\Phi_i \left(\mathbf{H}\mathbf{x}_i - \hat{\mathbf{z}}_i\right)\|_2^2 + P_1(\mathbf{x}) + P_2(\mathbf{x}) + P_3(\mathbf{x}) \tag{1.22}$$

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 $\hat{\mathbf{z}}_i$. In this way, the data fidelity term applies no weight to pixel locations which have no observed data. On the other hand, those pixels which represent numerous measurements, have a stronger influence on the data fidelity penalty function. Here, we observe that the data fidelity term no longer requires summation over the set of captured frames, saving a considerable amount of processing time as well as memory.

1.4.2   Dynamic Multiframe Demosaicing

Armed with the fast implementation of the multiframe demosaicing algorithm, we turn to apply this algorithm to video sequences. This is called *dynamic* multiframe demosaicing that produces an image *sequence* or video with higher resolution. A naive approach to this problem is to apply the static algorithm on a set of images while changing the reference frame. However, the memory and computational requirements for the static process are so taxing as to preclude its direct application to the dynamic case.

In contrast, we present a dynamic algorithm which takes advantage of the fact that the resulting high resolution image for the previous time frame $t - 1$ helps predict the solution for the current time frame

t. In this section, we replace the generic frame index k with t to indicate the temporal ordering of the low resolution frames. A simple forward model capturing the temporal relationship of the low-resolution images is

$$\mathbf{x}(t) = \mathbf{F}(t)\mathbf{x}(t-1) + \mathbf{u}(t), \tag{1.23}$$

and

$$\mathbf{y}(t) = \mathbf{ADH}(t)\mathbf{x}(t) + \mathbf{v}(t). \tag{1.24}$$

In other words, the current frame is a shifted version of the previous frame with some additive noise with covariance $\mathbf{C}_u(t)$.

The equations given above describe a system in its *state-space* form, where the state is the desired ideal image. Thus, a Kalman-filter (KF) [75] formulation can be employed to recursively compute the optimal estimates $(\mathbf{x}(t), t \in \{1, ..., N\})$ from the measurements $(\mathbf{y}(t), t \in \{1, ..., N\})$, assuming that $\mathbf{D}, \mathbf{H}$, $\mathbf{F}(t)$, $\mathbf{C}_u(t)$, and $\mathbf{C}_v(t)$ (covariance of $\mathbf{v}(t)$) are known [34, 71, 70]. In the following, we study the application of the causal Kalman filter, where the estimates are obtained through on-line processing of an incoming sequence[5].

To implement the dynamic multiframe demosaicing algorithm in a fast and efficient manner, we rewrite the state space model Eqs. 1.23 and 1.24 in terms of $\mathbf{z}$ ($\mathbf{z}(t) = \mathbf{Hx}(t)$) as

$$\mathbf{z}(t) = \mathbf{F}(t)\mathbf{z}(t-1) + \mathbf{e}(t), \tag{1.25}$$

and

$$\mathbf{y}(t) = \mathbf{ADz}(t) + \mathbf{v}(t). \tag{1.26}$$

Note that the first of the two equations is obtained by left multiplication of both sides of Eq. 1.23 by $\mathbf{H}$ and using the fact that it commutes with $\mathbf{F}(t)$. Thus, the perturbation vector $\mathbf{e}(t)$ is a colored version of $\mathbf{u}(t)$, leading to $\mathbf{C}_e = \mathbf{HC}_u\mathbf{H}^T$ as its covariance matrix.

The following defines the forward Kalman propagation and update equations [34], that account for a causal (on-line) process. We assume that at time $t-1$ we already have the mean-covariance pair, $(\hat{\mathbf{z}}(t-1), \hat{\Pi}(t-1))$, and those should be updated to account for the information obtained at time t. We start with the covariance matrix update based on Equation (1.25),

$$\tilde{\Pi}(t) = \mathbf{F}(t)\hat{\Pi}(t-1)\mathbf{F}^T(t) + \mathbf{C}_{e(t)}, \tag{1.27}$$

where $\tilde{\Pi}(t)$ is the propagated covariance matrix (initial estimate of the covariance matrix at time t). The KF gain matrix is given by

$$\mathbf{K}(t) = \tilde{\Pi}(t)(\mathbf{AD})^T[\mathbf{C}_v(t) + \mathbf{AD}\tilde{\Pi}(t)\mathbf{D}^T]^{-1}. \tag{1.28}$$

---

[5] A closely related non-causal processing mode, where every high-resolution reconstructed image is derived as an optimal estimate incorporating information from all the frames in the sequence is studied in [73].

Based on $\mathbf{K}(t)$, the updated state vector mean is computed by

$$\hat{\mathbf{z}}(t) = \mathbf{F}(t)\hat{\mathbf{z}}(t-1) + \mathbf{K}(t)[\mathbf{y}(t) - \mathbf{ADF}(t)\hat{\mathbf{z}}(t-1)]. \tag{1.29}$$

The final stage requires the update of the covariance matrix, based on Equation 1.26,

$$\hat{\Pi}(t) = \text{Cov}\,(\hat{\mathbf{z}}(t)) = [\mathbf{I} - \mathbf{K}(t)\mathbf{AD}]\tilde{\Pi}(t). \tag{1.30}$$

More on the meaning of these equations and how they are derived can be found in [34, 76].

While in general the above equations require the propagation of intolerably large matrices in time, if we refer to $\mathbf{C}_e(t)$ as a diagonal matrix, then $\tilde{\Pi}(t)$, and $\hat{\Pi}(t)$ are diagonal matrices. Following [71, 70], if we choose a matrix $\sigma_e^2\mathbf{I} \geq \mathbf{C}_e(t)$, it implies that $\sigma_e^2\mathbf{I} - \mathbf{C}_e(t)$ is a positive semi-definite matrix, and there is always a finite σ_e that satisfies this requirement. Replacing $\mathbf{C}_e(t)$ with this majorizing diagonal matrix, the new state-space system in Equations (1.25) and (1.26) simply assumes a stronger innovation process. The effect on the KF is to rely less on the temporal relation in (1.25) and more on the measurements in (1.26). Since $\mathbf{C}_e(t)$ is diagonal, Eqs. 1.27, 1.28, 1.29, and (1.30) can be implemented on an extremely fast pixel-by-pixel basis. We refer the interested reader to [73].

The output of the temporal Kalman filter equations Eqs. 1.27, 1.28, 1.29, and (1.30) is an estimate of the blurred high resolution video sequence $\hat{\mathbf{z}}(t)$. We refer to this as the dynamic Shift-and-Add sequence. At this point, we apply the iterative deblurring and interpolation step described earlier. The iterative deblurring of the blurry image $\hat{\mathbf{z}}(t)$ for time t is accomplished by minimizing the cost function

$$\Omega(\mathbf{x}(t)) = J_1'(\mathbf{x}(t), \hat{\mathbf{z}}(t)) + P_1(\mathbf{x}(t)) + P_2(\mathbf{x}(t)) + P_3(\mathbf{x}(t)) \tag{1.31}$$

using the forward shifted version of the previous estimate $\mathbf{F}(t)\hat{\mathbf{x}}(t-1)$ as the initial guess. By relying on the shifted high resolution estimate from the previous frame $\hat{\mathbf{x}}(t-1)$ as an initial guess, very few iterations are required for subsequent frames.

Figure 1.15 shows the block diagram describing the fast dynamic multiframe demosaicing algorithm. The dynamic multiframe demosaicing algorithm as presented describes only the forward causal processing model. This can also be extended to consider both forward filtering and smoothing as described in [64, 77].

1.4.3 Example of Dynamic Multiframe Demosaicing

W present an example demonstrating the dynamic multiframe demosaicing algorithm. As before, we used 74 uncompressed, raw CFA images from a video camera (based on Zoran 2MP CMOS Sensors). The upper and lower images in the left column of Fig. 1.16 show the low-resolution frames (frames number 1, and 69, respectively) demosaiced by the method in [7]. The central images show the resulting color images after the recursive dynamic Shift-and-Add processing [73]. Some of the resolution has been restored at this
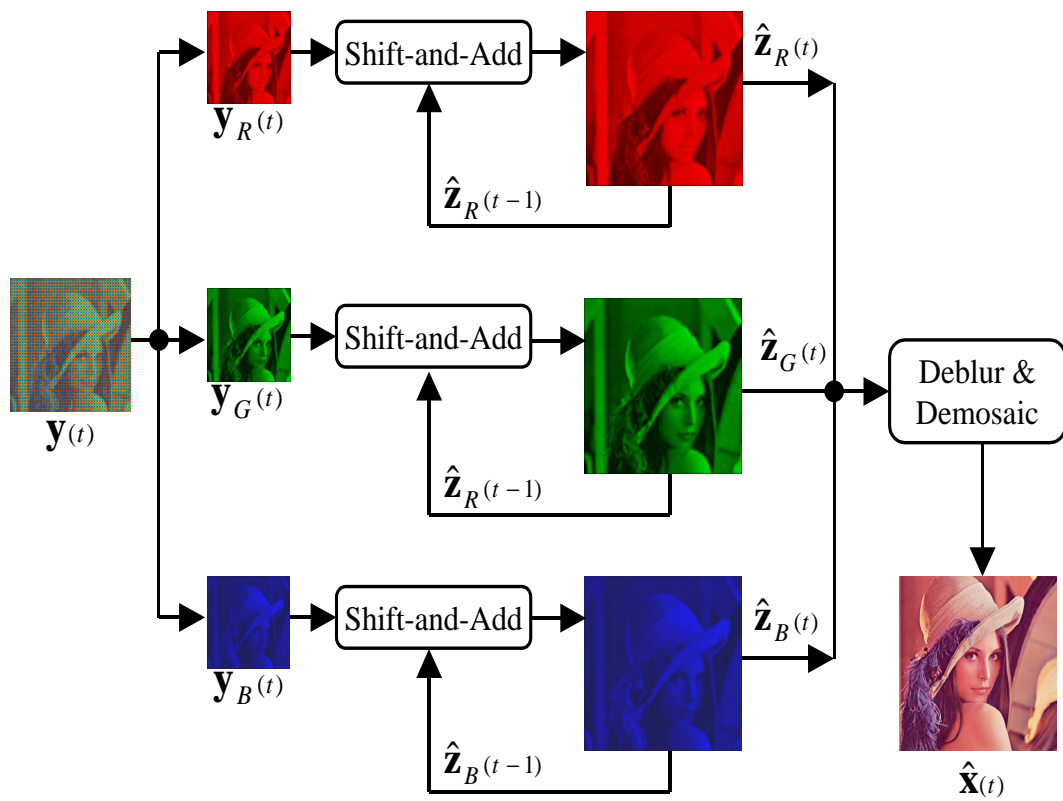
Figure 1.15: Block diagram representation of the fast dynamic multiframe demosaicing process.

Figure 1.16: The images in the left hand column show frames #1 (top) and #69 (bottom) of the 74 frame low-resolution CFA image sequence. The second column shows the results after the recursive dynamic Shift-and-Add processing increasing their resolution by a factor of three in each direction. The right column shows the images after iteratively deblurring the dynamic Shift-and-Add images. We observe that the dynamic Shift-and-Add performs much of the work, while the subsequent deblurring restores the final details to the images.

point. The images in the right column show the iteratively deblurred versions of the Shift-and-Add images. Overall, the final image sequences shows significant improvement over the original sequence.

1.5   Conclusion

In this Chapter, we discussed two important inverse problems in imaging; namely super-resolution, and demosaicing. More significantly, we presented a unified approach for simultaneously solving these two interrelated problems. We illustrated how using the robust L_1 norm for the data fidelity term makes the proposed methods robust to errors in both measurements and in modeling. Furthermore, we used the bilateral regularization of the luminance term to achieve sharp reconstruction of edges, meanwhile the chrominance and inter-color dependencies cost functions were tuned to remove color artifacts from the final estimate.

While the implementation of the resulting algorithms may appear complex at first blush, from a practical point of view, all matrix-vector operations in the proposed methods are implementable as simple image operators, making the methods practically feasible and possibly useful for implementation on specialized or general purpose hardware within commercially available cameras. Furthermore, 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.

As we have pointed out, accurate subpixel motion estimation is an essential part of any image fusion process such as the proposed simultaneous multi-frame super-resolution and demosaicing. While multiscale methods based upon the optical flow or phase correlation principles are adequate, they are not specifically tailored or appropriate for estimating motions between color-filtered images. In fact, to the best of our knowledge, there is no significant body of literature that addresses the problem of estimating subpixel motion between Bayer filtered images. This is a significant gap in the literature which still needs to be filled, and which could create improved versions of the algorithms suggested in this chapter. More broadly, a new paradigm along the lines of [25] is needed in which accurate subpixel motion estimation is performed jointly inside and along with the overall super-resolution /demosaicing problem, instead of what has traditionally been a preprocessing step.

1.6   Acknowledgments

Bibliography

[1] B. Bayer, "Color imaging array." US Patent 3971065, 1976.

[2] R. Lukac and K. Plataniotis, "Color filter arrays: design and performance analysis," *IEEE Trans. Image Processing*, vol. 51, pp. 1260–1267, Nov. 2005.

[3] N. R. Shah and A. Zakhor, "Resolution enhancement of color video sequences," *IEEE Trans. Image Processing*, vol. 8, pp. 879–885, June 1999.

[4] B. C. Tom and A. Katsaggelos, "Resolution enhancement of monochrome and color video using motion compensation," *IEEE Trans. Image Processing*, vol. 10, pp. 278–287, Feb. 2001.

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

[6] D. R. Cok, "Signal processing method and apparatus for sampled image signals." United States Patent 4,630,307, 1987.

[7] C. Laroche and M. Prescott, "Apparatus and method for adaptively interpolating a full color image utilizing chrominance gradients." United States Patent 5,373,322, 1994.

[8] J. Hamilton and J. Adams, "Adaptive color plan interpolation in single sensor color electronic camera." United States Patent 5,629,734, 1997.

[9] R. Kimmel, "Demosaicing: Image reconstruction from color CCD samples," *IEEE Trans. Image Processing*, vol. 8, pp. 1221–1228, Sept. 1999.

[10] L. Chang and Y.-P. Tan, "Color filter array demosaicking: new method and performance measures," *IEEE Trans. Image Processing*, vol. 12, pp. 1194–1210, Oct. 2002.

[11] K. Hirakawa and T. Parks, "Adaptive homogeneity-directed demosaicing algorithm," in *Proc. of the IEEE Int. Conf. on Image Processing*, vol. 3, pp. 669–672, Sept. 2003.

[12] D. Keren and M. Osadchy, "Restoring subsampled color images," *Machine Vision and applications*, vol. 11, no. 4, pp. 197–202, 1999.

[13] Y. Hel-Or and D. Keren, "Demosaicing of color images using steerable wavelets," Tech. Rep. HPL-2002-206R1 20020830, HP Labs Israel, 2002.

[14] W. K. Pratt, *Digital image processing*. New York: John Wiley & Sons, INC., 3rd ed., 2001.

[15] D. Taubman, "Generalized Wiener reconstruction of images from colour sensor data using a scale invariant prior," in *Proc. of the IEEE Int. Conf. on Image Processing*, vol. 3, pp. 801–804, Sept. 2000.

[16] D. D. Muresan and T. W. Parks, "Optimal recovery demosaicing," in *IASTED Signal and Image Processing*, Aug. 2002.

[17] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Color plane interpolation using alternating projections," *IEEE Trans. Image Processing*, vol. 11, pp. 997–1013, Sep. 2002.

[18] S. C. Pei and I. K. Tam, "Effective color interpolation in CCD color filter arrays using signal correlation," *IEEE Trans. Image Processing*, vol. 13, pp. 503–513, June 2003.

[19] D. Alleysson, S. Süsstrunk, and J. Hérault, "Color demosaicing by estimating luminance and opponent chromatic signals in the Fourier domain," in *Proc. of the IS&T/SID 10th Color Imaging Conf.*, pp. 331–336, Nov. 2002.

[20] R. Ramanath and W. Snyder, "Adaptive demosaicking," *Journal of Electronic Imaging*, vol. 12, pp. 633–642, Oct. 2003.

[21] L. Zhang and X. Wu, "Color demosaicing via directional linear minimum mean square-error estimation," *IEEE Trans. Image Processing*, vol. 14, pp. 2167–2178, Dec. 2005.

[22] R. Lukac, K. Plataniotis, D. Hatzinakos, and M. Aleksic, "A new CFA interpolation framework," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 86, pp. 1559–1579, July 2006.

[23] R. Lukac, K. Martin, and K. Plataniotis, "Colour-difference based demosaicked image postprocessing," *Electronics Letters*, vol. 39, pp. 1805–1806, Dec. 2003.

[24] R. Lukac, K. Martin, and K. Plataniotis, "Demosaicked image postprocessing using local color ratios," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 914–920, June 2004.

[25] D. Robinson, S. Farsiu, and P. Milanfar, "Optimal registration of aliased images using variable projection with applications to superresolution," *Accepted for publication in The Computer Journal*, vol. Special Issue on Super-Resolution in Imaging and Video, 2007.

[26] S. Borman and R. L. Stevenson, "Super-resolution from image sequences - a review," in *Proc. of the 1998 Midwest Symposium on Circuits and Systems*, vol. 5, Apr. 1998.

[27] S. Park, M. Park, and M. G. Kang, "Super-resolution image reconstruction, a technical overview," *IEEE Signal Processing Magazine*, vol. 20, pp. 21–36, May 2003.

[28] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology*, vol. 14, pp. 47–57, Oct. 2004.

[29] T. S. Huang and R. Y. Tsai, "Multi-frame image restoration and registration," *Advances in computer vision and Image Processing*, vol. 1, pp. 317–339, 1984.

[30] N. K. Bose, H. C. Kim, and H. M. Valenzuela, "Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframes," in *Proc. of the IEEE Int. Conf. Acoustics, Speech, and Signal Processing(ICASSP)*, vol. 5, pp. 269–272, Apr. 1993.

[31] L. Teodosio and W. Bender, "Salient video stills: Content and context preserved," in *Proc. of the First ACM Int. Conf. on Multimedia*, vol. 10, pp. 39–46, Aug. 1993.

[32] M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space invariant blur," *IEEE Trans. Image Processing*, vol. 10, pp. 1187–1193, Aug. 2001.

[33] M. C. Chiang and T. E. Boulte, "Efficient super-resolution via image warping," *Image and Vision Computing*, vol. 18, pp. 761–771, July 2000.

[34] S. M. Kay, *Fundamentals of statistical signal processing:estimation theory*, vol. I. Upper Saddle River, New Jersey: Prentice-Hall, 1993.

[35] M. Elad and A. Feuer, "Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images," *IEEE Trans. Image Processing*, vol. 6, pp. 1646–1658, Dec. 1997.

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

[37] N. Nguyen, P. Milanfar, and G. H. Golub, "A computationally efficient image superresolution algorithm," *IEEE Trans. Image Processing*, vol. 10, pp. 573–583, Apr. 2001.

[38] M. Ng and N. Bose, "Mathematical analysis of super-resolution methodology," *IEEE Signal Processing Mag.*, vol. 20, pp. 62–74, May 2003.

[39] S. Lertrattanapanich and N. K. Bose, "High resolution image formation from low resolution frames using Delaunay triangulation," *IEEE Trans. Image Processing*, vol. 11, pp. 1427–1441, Dec. 2002.

[40] A. Patti, M. Sezan, and A. M. Tekalp;, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Trans. Image Processing*, vol. 6, pp. 1326–1333, Mar. 1997.

[41] Y. Altunbasak, A. Patti, and R. Mersereau, "Super-resolution still and video reconstruction from MPEG-coded video," *IEEE Trans. Circuits And Syst. Video Technol.*, vol. 12, pp. 217–226, Apr. 2002.

[42] O. Haik and Y. Yitzhaky, "Super-resolution reconstruction of a video captured by a vibrated tdi camera," *Journal of Electronic Imaging*, vol. 15, pp. 023006–1–12, Apr-Jun 2006.

[43] C. B. Atkins, C. A. Bouman, and J. P. Allebach, "Tree-based resolution synthesis," in *Proc. of the IS&T Conf. on Image Processing, Image Quality, Image Capture Systems*, pp. 405–410, 1999.

[44] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1167–1183, Sept. 2002.

[45] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super resolution," in *Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 645–650, Dec. 2001.

[46] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super-resolution," *IEEE Transactions on Image Processing*, vol. 13, pp. 1327–1344, Oct. 2004.

[47] Z. A. Ivanovski, L. Panovski, and L. J. Karam, "Robust super-resolution based on pixel-level selectivity," *In Proc. of the 2006 SPIE Conf. on Visual Communications and Image Processing*, pp. 607707–1–8, Jan. 2006.

[48] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, Nov. 1992.

[49] T. F. Chan, S. Osher, and J. Shen, "The digital TV filter and nonlinear denoising," *IEEE Trans. Image Processing*, vol. 10, pp. 231–241, Feb. 2001.

[50] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. of IEEE Int. Conf. on Computer Vision*, pp. 836–846, Jan. 1998.

[51] M. Elad, "On the bilateral filter and ways to improve it," *IEEE Trans. Image Processing*, vol. 11, pp. 1141–1151, Oct. 2002.

[52] T. Q. Pham, L. J. van Vliet, and K. Schutte, "Robust fusion of irregularly sampled data using adaptive normalized convolution," *EURASIP Journal on Applied Signal Processing*, p. Article ID 83268, 2006.

[53] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Transactions on Image Processing*, vol. 16, pp. 349–366, Feb. 2007.

[54] M. Ben-Ezra, A. Zomet, and S. Nayar, "Video super-resolution using controlled subpixel detector shifts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 977–987, June 2004.

[55] C. A. Segall, R. Molina, A. Katsaggelos, and J. Mateos, "Bayesian high-resolution reconstruction of low-resolution compressed video," in *Proc. of IEEE Int. Conf. on Image Processing*, vol. 2, pp. 25–28, Oct. 2001.

[56] C. Segall, A. Katsaggelos, R. Molina, and J. Mateos, "Bayesian resolution enhancement of compressed video," *IEEE Trans. Image Processing*, vol. 13, pp. 898–911, July 2004.

[57] S. Borman, *Topics in Multiframe Superresolution Restoration.* PhD thesis, University of Notre Dame, Notre Dame, IN, May 2004.

[58] M. G. Kang and S. Chaudhuri, "Super-resolution image reconstruction," *IEEE Signal Processing Magazine*, vol. 20, pp. 21–36, May 2003.

[59] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology*, vol. 14, pp. 47–57, Aug. 2004.

[60] S. Farsiu, M. Elad, and P. Milanfar, "Multi-frame demosaicing and super-resolution from under-sampled color images," *Proc. of the 2004 IS&T/SPIE 16th Annual Symposium on Electronic Imaging*, vol. 5299, pp. 222–233, Jan. 2004.

[61] T. Gotoh and M. Okutomi, "Direct super-resolution and registration using raw CFA images," in *Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 600–607, July 2004.

[62] R. Lukac and K. Plataniotis, "Fast video demosaicking solution for mobile phone imaging applications," *IEEE Trans. Consumer Electronics*, vol. 51, pp. 675–681, May 2005.

[63] R. Sasahara, H. Hasegawa, I. Yamada, and K. Sakaniwa, "A color super-resolution with multiple non-smooth constraints by hybrid steepest descent method," in *in the Proc. of International Conference on Image Processing (ICIP)*, pp. 857–860, 2005.

[64] S. Farsiu, M. Elad, and P. Milanfar, "Multiframe demosaicing and super-resolution of color images," *IEEE Transactions on Image Processing*, vol. 15, pp. 141–159, January 2006.

[65] R. Lukac and K. Plataniotis, "Adaptive spatiotemporal video demosaicking using bidirectional multi-stage spectral filters," *IEEE Transactions on Consumer Electronics*, vol. 52, pp. 651–654, May 2006.

[66] P. Vandewalle, K. Krichane, D. Alleysson, and S. Süsstrunk, "Joint demosaicing and super-resolution imaging from a set of unregistered aliased images," in *IS&T/SPIE Electronic Imaging: Digital Photography III*, vol. 6502, 2007.

[67] D. Keren and A. Gotlib, "Denoising color images using regularization and correlation terms," *Journal of Visual Communication and Image Representation*, vol. 9, pp. 352–365, Dec. 1998.

[68] S. Farsiu, M. Elad, and P. Milanfar, "Constrained, globally optimal, multi-frame motion estimation," in *Proc. of the 2005 IEEE Workshop on Statistical Signal Processing*, pp. 1396 – 1401, July 2005.

[69] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierachical model-based motion estimation," *Proc. of the European Conf. on Computer Vision*, pp. 237–252, May 1992.

[70] M. Elad and A. Feuer, "Super-resolution reconstruction of image sequences," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, pp. 817–834, Sept. 1999.

[71] M. Elad and A. Feuer, "Superresolution restoration of an image sequence: adaptive filtering approach," *IEEE Trans. Image Processing*, vol. 8, pp. 387–395, Mar. 1999.

[72] C. Newland, D. Gray, and D. Gibbins, "Modified kalman filtering for image super-resolution,," in *In Proc. of the 2006 Conf. on Image and Vision Computing (IVCNZ06)*, pp. 79–84, Nov. 2006.

[73] S. Farsiu, M. Elad, and P. Milanfar, "Video-to-video dynamic super-resolution for grayscale and color sequences," *EURASIP Journal on Applied Signal Processing*, vol. 2006, p. Article ID 61859, 2006.

[74] M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space invariant blur," *IEEE Transactions on Image Processing*, vol. 10, pp. 1186–1193, Aug. 2001.

[75] B. Anderson and J. Moore, *Optimal filtering*. 1979.

[76] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York: Academic Press, 1970.

[77] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Dynamic demosaicing and color superresolution of video sequences," in *Proc. SPIE's Conf. on Image Reconstruction from Incomplete Data III*, pp. 169–178, Denver, CO. Aug. 2004.

[78] M. Robinson, S. Farsiu, J. Lo, P. Milanfar, and C. Toth, "Efficient registration of aliased x-ray images," *Proc. of the 41th Asilomar Conf. on Signals, Systems, and Computers*, pp. 215 – 219, Oct. 2007.

[79] S. Farsiu, M. Elad, and P. Milanfar, "A practical approach to super-resolution," in *Proc. of the SPIE: Visual Communications and Image Processing*, vol. 6077, pp. 24–38, Jan. 2006.

[80] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, pp. 325–376, Dec. 1992.

[81] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. of DARPA Image Understanding Workshop*, pp. 121–130, 1981.

[82] W. Zhao and H. Sawhney, "Is super-resolution with optical flow feasible?," in *ECCV*, vol. 1, pp. 599–613, 2002.

[83] M. Alkhanhal, D. Turaga, and T. Chen, "Correlation based search algorithms for motion estimation," in *Picture Coding Symposium*, Apr. 1999.

[84] H. Sawhney, S. Hsu, and R. Kumar, "Robust video mosaicing through topology inference and local to global alignment," in *ECCV*, vol. 2, pp. 103–119, 1998.

[85] V. Govindu, "Combining two-view constraints for motion estimation," in *Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 218–225, July 2001.

[86] V. Govindu, "Lie-algebraic averaging for globally consistent motion estimation," in *Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 684–691, July 2004.

[87] Y. Sheikh, Y. Zhai, and M. Shah, "An accumulative framework for the alignment of an image sequence," in *ACCV*, Jan. 2004.

[88] P. Vandewalle, S. Susstrunk, and M. Vetterli, "A frequency domain approach to registration of aliased images with application to super-sesolution," *EURASIP Journal on Applied Signal Processing*, p. Article ID 71459, 2006.

[89] H. Foroosh, J. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Transactions on Image Processing*, vol. 11, pp. 188–200, Mar. 2002.

[90] D. Robinson and P. Milanfar, "Statistical performance analysis of super-resolution," *IEEE Transactions on Image Processing*, vol. 15, June 2006.

[91] N. Woods, N. Galatsanos, and A. Katsaggelos, "Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images.," *IEEE Trans. Image Processing*, vol. 15, pp. 201–213, Jan 2006.

[92] J. Chung, E. Haber, and J. Nagy, "Numerical methods for coupled super-resolution.," *Inverse Problems*, vol. 22, pp. 1261 – 1272, June 2006.
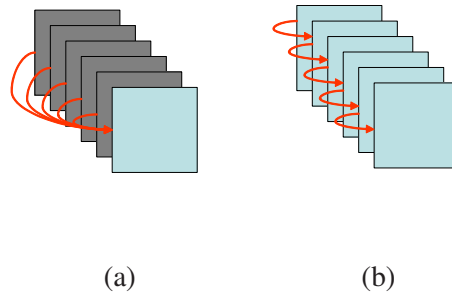
<center>(a)   (b)</center>

Figure 17: Common strategies used for registering frames of a video sequence. (a) Fixed reference ("anchored") estimation. (b) Pairwise ("progressive") estimation.

.1   Appendix: Motion Estimation

The first step in solving the multiframe demosaicing problem is estimating the motion between the collection of CFA images. Numerous image registration techniques have been developed throughout the years [80]. Of these, optical flow [81] [82], and correlation-based methods [83] are among the most popular.

While the detailed description of these techniques are out of the scope of this paper, we note that these methods are mainly developed to estimate the relative motion between *a pair* of frames. For cases where several images are to be registered with respect to each other (e.g. super-resolution applications), two simple strategies are commonly used. The first is to register all frames with respect to a single reference frame [46]. This may be called the *anchoring* approach, as illustrated in Figure .1(a). The choice of a reference or anchor frame is rather arbitrary, and can have a severe effect on the overall accuracy of the resulting estimates. This caveat aside, overall, this strategy is effective in cases where the camera motion is small and random (e.g. small vibrations of a gazing camera).

The other popular strategy is the *progressive* registration method [31] [77], where images in the sequence are registered in pairs, with one image in each pair acting as the reference frame. For instance, taking a causal view with increasing index denoting time, the ith frame of the sequence is registered with respect to the $(i+1)^{\text{th}}$ frame and the $(i+1)^{\text{th}}$ frame is registered with respect to the $(i+2)^{\text{th}}$ frame, and so on, as illustrated in Figure .1(b). The motion between an arbitrary pair of frames is computed as the combined motion of the above incremental estimates. This method works best when the camera motion is smooth. However, in this method, the registration error between two "nearby" frames is accumulated and propagated when such values are used to compute motion between "far away" frames. Neither of the above approaches take advantage of the important prior information available for the multi-frame motion estimation problem. This prior information constrains the estimated motion vector fields as defined in the following.

To begin, let us define $\mathbf{F_{i,j}}$ as the operator which maps (registers) frames indexed i and j as follows:

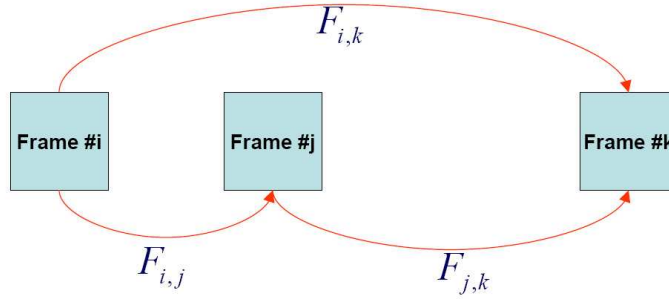$$\mathbf{y}_i = \mathbf{F_{i,j}}\{\mathbf{y}_j\},$$

Figure 18: The consistent motion property dictates that the motion between any pair of frames must be the composition of the motion between two other pairs of frames.

where $\mathbf{y}_i$ and $\mathbf{y}_j$ are the lexicographic reordered vector representations of frames i and j. Now given a sequence of N frames, precisely $N(N-1)$ such operators can be considered. Regardless of considerations related to noise, sampling, and the finite dimensions of the data, there is an inherent intuitive relationship between these pair-wise registration operators. This condition dictates that the operator describing the motion between any pair of frames must be the composition of the operators between two other pairs of frames. More specifically, as illustrated in Figure .1, taking any triplet of frames i, j, and k, we have the motion consistency condition as:

$$\forall i,j,k \in \{1,...,N\}, \quad \mathbf{F_{i,k}} = \mathbf{F_{i,j}} \circ \mathbf{F_{j,k}}. \tag{32}$$

Using these types of constraints provides a means for reliable estimation of the sub-pixel motions and hence the warping operators. While a detailed discussion will be out of the scope of this paper, we shall note that the constrained motion property has been successfully implemented in different guises and for various applications such as mosaicing [84], motion estimation [85, 86, 87], and super-resolution/demosaicing [82, 68, 88, 89, 90, 91, 92, 25].

Index