

# A Practical Approach to Super-Resolution

Sina Farsiu <sup>\*</sup>, Michael Elad <sup>†</sup>, Peyman Milanfar <sup>‡</sup>

## ABSTRACT

Theoretical and practical limitations usually constrain the achievable resolution of any imaging device. Super-Resolution (SR) methods are developed through the years to go beyond this limit by acquiring and fusing several low-resolution (LR) images of the same scene, producing a high-resolution (HR) image. The early works on SR, although occasionally mathematically optimal for particular models of data and noise, produced poor results when applied to real images. In this paper, we discuss two of the main issues related to designing a practical SR system, namely reconstruction accuracy and computational efficiency. Reconstruction accuracy refers to the problem of designing a robust SR method applicable to images from different imaging systems. We study a general framework for optimal reconstruction of images from grayscale, color, or color filtered (CFA) cameras. The performance of our proposed method is boosted by using powerful priors and is robust to both measurement (e.g. CCD read out noise) and system noise (e.g. motion estimation error). Noting that the motion estimation is often considered a bottleneck in terms of SR performance, we introduce the concept of “constrained motions” for enhancing the quality of super-resolved images. We show that using such constraints will enhance the quality of the motion estimation and therefore results in more accurate reconstruction of the HR images. We also justify some practical assumptions that greatly reduce the computational complexity and memory requirements of the proposed methods. We use efficient approximation of the Kalman Filter (KF) and adopt a dynamic point of view to the SR problem. Novel methods for addressing these issues are accompanied by experimental results on real data.

## 1. INTRODUCTION

On the path to designing HR imaging systems, one quickly runs into the problem of diminishing returns. Specifically, the imaging chips and optical components necessary to capture very HR images become prohibitively expensive, costing in the millions of dollars for scientific applications. Hence, there is a growing interest in the multi-frame image reconstruction algorithms that compensate for the shortcomings of the imaging systems. Such methods can achieve high-quality images using less expensive imaging chips and optical components by capturing multiple images and fusing them. The application of such algorithms will certainly continue to proliferate in any situation where high quality optical imaging systems cannot be incorporated or are too expensive to utilize.

SR is the term generally applied to the problem of transcending the limitations of optical imaging systems through the use of image processing algorithms, which presumably are relatively inexpensive to implement. The basic idea behind SR is the fusion of a sequence of LR noisy blurred images to produce a higher resolution image. The resulting HR image (or sequence) has more high-frequency content and less noise and blur effects than any of the LR input images. Early works on SR showed that it is the aliasing effects in the LR images that enable the recovery of the HR fused image(s), provided that a relative sub-pixel motion exists between the under-sampled input images<sup>1</sup>.

The experiment in Figure 1 shows an example of SR technology. In this experiment, a set of 26 images were captured by an OLYMPUS C-4000 camera. One of these images is shown in Figure 1(a). Unfortunately due to the limited number of pixels in the digital camera the details of these images are not clear, as shown in the zoomed image of Figure 1(b). SR helps us to reconstruct the details lost in the imaging process. The result of applying the SR algorithm described in Section 2 is shown in Figure 1(c), which is a high-quality image with 16 times more pixels than any LR frame (resolution enhancement factor of 4 in each direction)<sup>§</sup>.

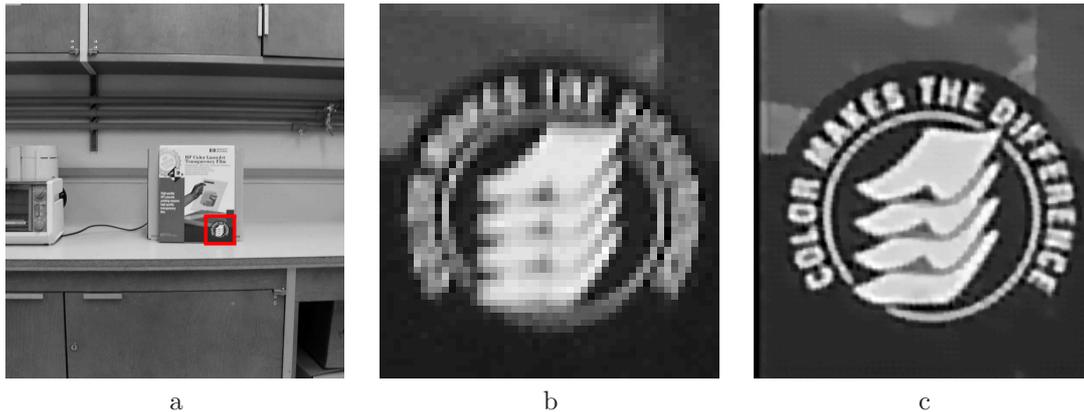
---

<sup>\*</sup> Corresponding author: Electrical Engineering Department, University of California Santa Cruz, Santa Cruz CA. 95064 USA. Email: farsiu@ee.ucsc.edu, Phone:(831)-459-1582, Fax: (831)-459-4829

<sup>†</sup> Computer Science Department, The Technion, Israel Institute of Technology, Israel. Email: elad@cs.technion.ac.il, Phone:972-4-829-4169, Fax: 972-4-829-4353

<sup>‡</sup>Electrical Engineering Department, University of California Santa Cruz, Santa Cruz CA. 95064 USA. Email: milanfar@ee.ucsc.edu, Phone:(831)-459-4929, Fax: (831)-459-4829

<sup>§</sup>This paper (with all color pictures) plus other related works containing more details about the ideas/experiments presented here are available at <http://www.ee.ucsc.edu/~milanfar/> .



**Figure 1.** SR experiment on real world data. A set of 26 low quality images were fused resulting in 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.

Overall, SR is a computationally complex<sup>2</sup> and numerically ill-posed problem<sup>3</sup>, which makes it one of the most appealing research areas in image processing. In the past two decades, a variety of SR methods have been proposed. These methods are usually very sensitive to their assumed model of data and noise, or computationally very expensive, which limits their utility. Recently, alternative approaches are proposed to overcome such shortcomings. For instance, the very recent work of Ref.[4] proposes a solution based on modifying camera hardware.

In contrast, in this paper we outline some of our recent work addressing these problems based on statistical signal processing techniques. We study a general Maximum A-Posteriori (MAP) framework for solving multi-frame image fusion problems. This framework not only helps us to propose a fast and robust solutions for the classic SR problem, but also guides us to an effective solution for the multi-frame demosaicing problem. Also, as the performance of motion estimation is of paramount importance to the performance of SR, we propose an efficient robust multi-frame motion estimation method.

Our proposed framework, in summary, is an efficient solution to the multi-frame imaging inverse problem that

1. defines a forward model describing all the components of the imaging channel (such as probability density function (PDF) of additive noise, blur point spread function (PSF), relative motion vectors,...).
2. adopts proper prior information to turn the ill-posed inverse problem to a well-posed problem (regularization)
3. applies a method for fusing the information from multiple images which is
  - (a) robust to inaccuracies in the forward model and the noise in the estimated data.
  - (b) computationally efficient.

In what follows in this paper, we study several important multi-frame image fusion/reconstruction problems under this general framework that helps us provide practical fast and robust solutions. In Section 2, we study the “multi-frame SR” problem for the gray-scale images. In Section 3, we focus on color images and search for an efficient method for removing color artifacts in digital images. We propose a fast and robust hybrid method of SR and demosaicing, based on a MAP estimation technique by minimizing a multi-term cost function. In Section 4, we propose a dynamic SR method for producing a sequence of HR images, based on a very fast and memory efficient approximation of the Kalman Filter, applicable to both grayscale and color(filtered) images. In Section 5, we address the problem of estimating the relative motion between the frames of a video sequence. We exploit the global motion consistency conditions and propose a MAP based robust multi-frame motion estimation method, which is resilient against the effects of outliers and noise. Finally, Section 6 concludes this paper.

## 2. ROBUST MULTI-FRAME SUPER-RESOLUTION OF GRAYSCALE IMAGES

As we discussed in the introduction section, theoretical and practical limitations usually constrain the achievable resolution of any imaging device. In this section, we focus on the incoherent grayscale imaging systems and propose an effective multi-frame SR method that helps improve the quality of the captured images.

We use a popular matrix notation to formulate the general SR model<sup>5-7</sup>:

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

where  $F(k)$  is the geometric motion operator between the HR frame  $\underline{X}$  and  $k^{th}$  LR frame  $\underline{Y}(k)$  which are rearranged in lexicographic order.  $\underline{X}$  and  $\underline{Y}(k)$  represent the matrix form of the HR and  $k^{th}$  LR frames, respectively, where  $\underline{X}$  has  $r$  times more pixels than  $\underline{Y}(k)$  in each direction. The camera's point spread function (PSF), is modeled by the blur matrix  $H(k)$ , and  $D(k)$  represents the decimation operator.  $\underline{V}(k)$  is the additive noise and  $N$  is the number of available LR frames.

### 2.1. General Formulation

In this subsection, we explain an iterative robust solution for the SR problem. As we desire to produce an HR result which is similar to LR frames, we need a cost function which computes the similarity between LR and HR frames. As explained in Ref.[8],  $L_1$  norm used to define such error term results in robust reconstruction of the HR image in the presence of uncertainties such as motion error. Considering the general model of (1), the (grayscale) data fidelity penalty term is defined as:

$$J_0(\underline{X}) = \sum_{k=1}^N \|D(k)H(k)F(k)\underline{X} - \underline{Y}(k)\|_1. \quad (2)$$

Note that  $L_1$  norm minimization is the Maximum Likelihood (ML) estimate of data in the presence of Laplacian noise. The statistical analysis and experiments presented in Ref.[9] justifies modeling the SR noise in the presence of different sources of outliers as Laplacian probability density function (PDF) rather than Gaussian PDF.

As mentioned in Section 1, SR is an ill-posed problem<sup>6,10</sup>. For the under-determined cases<sup>¶</sup>, there exist an infinite number of solutions which minimize (2) equally well. The solution for square and over-determined <sup>||</sup> cases is not stable, which means that small amounts of noise in measurements results in large perturbations in the final solution. Therefore, considering regularization in SR algorithm as a means for picking a stable solution is indeed necessary. Also, regularization can help the algorithm to remove artifacts from the final answer and improve the rate of convergence. Of the many possible regularization terms, we desire one which results in HR images with sharp edges and one that is easy to implement.

In Ref.[8], based on the spirit of total variation criterion<sup>11,12</sup>, and a related technique called the bilateral filter<sup>13,14</sup>, we introduce our robust regularizer called Bilateral Total Variation (BTV), which is computationally cheap to implement, and preserves edges. The BTV regularizing function looks like,

$$J_1(\underline{X}) = \sum_{l,m=-P}^P \alpha^{|m|+|l|} \|\underline{X} - S_x^l S_y^m \underline{X}\|_1, \quad (3)$$

where  $S_x^l$  and  $S_y^m$  are the operators corresponding to shifting the image represented by  $\underline{X}$  by  $l$  pixels in horizontal direction and  $m$  pixels in vertical direction, respectively. This cost function in effect computes derivatives across multiple scales of resolution (as determined by the parameter “ $P$ ”). The scalar weight  $\alpha$ ,  $0 < \alpha < 1$ , is applied to give a spatially decaying effect to the summation of the regularization terms. In Ref.[8], several experiments are presented justifying the superior edge preserving property of the BTV prior with respect to other popular techniques such as Tikhonov regularization<sup>15</sup>.

<sup>¶</sup>where the number of non-redundant LR frames is smaller than the square of resolution enhancement factor ( $r^2$ ).

<sup>||</sup>where the number of non-redundant LR frames is equal or larger than the square of resolution enhancement factor ( $r^2$ ), respectively.

Combining the ideas presented thus far, we propose a robust solution of the SR problem as follows:

$$\hat{\underline{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[ \sum_{k=1}^N \|D(k)H(k)F(k)\underline{X} - \underline{Y}(k)\|_1 + \lambda \sum_{l,m=-P}^P \alpha^{|m|+|l|} \|\underline{X} - S_x^l S_y^m \underline{X}\|_1 \right]. \quad (4)$$

We use steepest descent to find the solution to this minimization problem:

$$\begin{aligned} \hat{\underline{X}}^{n+1} = \hat{\underline{X}}^n - \beta & \left\{ \sum_{k=1}^N F^T(k)H^T(k)D^T(k)\text{sign}(D(k)H(k)F(k)\hat{\underline{X}}^n - \underline{Y}(k)) \right. \\ & \left. + \lambda \sum_{l,m=-P}^P \alpha^{|m|+|l|} [I - S_y^{-m} S_x^{-l}] \text{sign}(\hat{\underline{X}}^n - S_x^l S_y^m \hat{\underline{X}}^n) \right\}, \end{aligned} \quad (5)$$

where  $\beta$  is a scalar defining the step size in the direction of the gradient.  $S_x^{-l}$  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 as  $S_x^l$  and  $S_y^m$ . The matrices  $F$ ,  $H$ ,  $D$ ,  $S$  and their transposes can be exactly interpreted as direct image operators such as shift, blur, and decimation<sup>8,16</sup>. Noting and implementing the effects of these matrices as a sequence of operators spares us from explicitly constructing them as matrices. This property and the fact that only five to twenty iterations are required for convergence helps our method to be implemented in a fast and memory efficient way.

## 2.2. Fast Robust SR Formulation

In the previous subsection, we proposed an iterative robust SR method based on equation (5). Although implementation of (5) is very fast, for real-time image sequence processing, faster methods are always desirable. In this subsection, based on the analysis that was offered in Ref.[8], we simplify (4) to achieve a faster method. In this method, resolution enhancement is broken into two consecutive steps:

1. Non-iterative data fusion (what we call median Shift-and-Add), and
2. Iterative deblurring-interpolation.

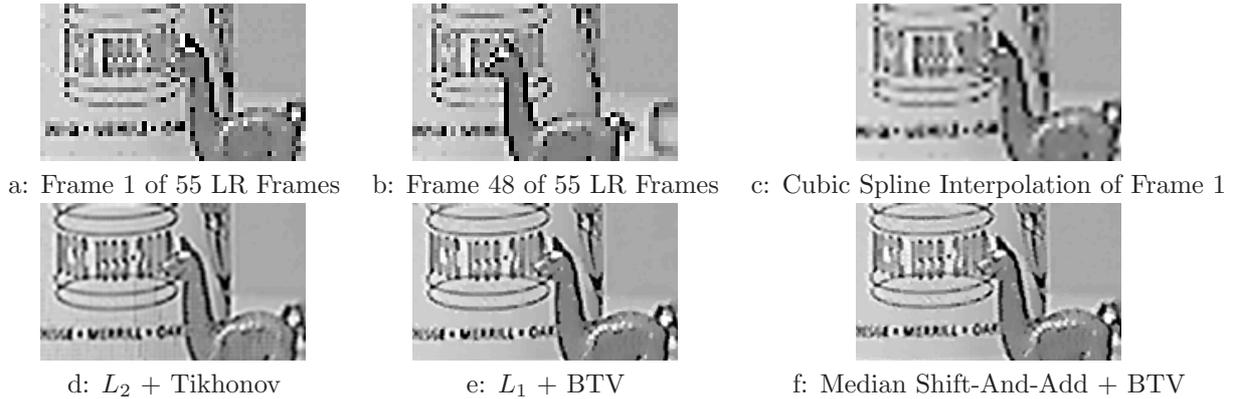
The Shift-and-Add step constructs a blurry (possibly with some missing pixels) HR image by performing a pixel-wise median of all measurements after proper (factor of  $r$ ) upsampling and motion compensation. In doing so, we refer to the penalty term defined in (2), while removing the blur effect (i.e.,  $H = I$ ), and disregarding the regularization. Different applications of the median operator for fusing LR images are also suggested in Ref.[7, 17, 18].

The goal of the deblurring-interpolation step is finding the deblurred HR frame. Note that in some cases not all pixel values of the Shift-and-Add image can be defined in the data fusion step, and their values should be defined in a separate interpolation step. The following expression formulates our minimization criterion for obtaining  $\hat{\underline{X}}$  from  $\hat{\underline{Z}}$  (the Shift-and-Add image).

$$\hat{\underline{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[ \|\Phi(H\underline{X} - \hat{\underline{Z}})\|_1 + \lambda' \sum_{l,m=-P}^P \alpha^{|m|+|l|} \|\underline{X} - S_x^l S_y^m \underline{X}\|_1 \right], \quad (6)$$

where the confidence matrix  $\Phi$  is a diagonal matrix with diagonal values equal to the square root of the number of measurements that contributed to make each element of  $\hat{\underline{Z}}$ . So, the undefined pixels of  $\hat{\underline{Z}}$  have no effect on the HR estimate  $\hat{\underline{X}}$ . On the other hand, those pixels of  $\hat{\underline{Z}}$  which have been produced from numerous measurements, have a stronger effect in the estimation of the HR frame  $\hat{\underline{X}}$ .

As an example, Figure 2(a) shows one of 55 images captured with a commercial web camera. In this sequence, two separate sources of motion were present. First, randomly shaking the camera introduced approximately



**Figure 2.** Results of different resolution enhancement methods applied to the Alpaca sequence. Outlier effects are apparent in the non-robust reconstruction method (d). However, the robust methods (e)-(f) were not affected by the outliers.

global translational motion between each frame. Second, the Alpaca statue was repositioned several times throughout the input frames (notice this relative motion in Figure 2(a) and Figure 2(b)). The non-robust  $L_2$  norm reconstruction (applying  $L_2$  norm in place of  $L_1$  norm in (2)) with Tikhonov regularization results in Figure 2(d) where the shadow-like artifacts to left of the Alpaca due to the Alpaca motion are apparent. The robust estimation methods, however, reveal the ability of the algorithm to remove such artifacts from the image as shown in Figure 2(e) and 2(f). Here, we also see that the performance of the faster method shown in Figure 2(f) is almost as good as the optimal method shown in Figure 2(e). More details about this experiment is given in Ref.[8].

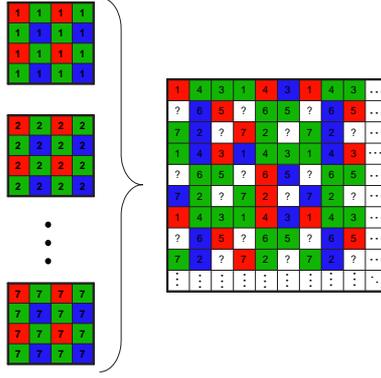
### 3. ROBUST MULTI-FRAME DEMOSAICING AND COLOR-SUPER-RESOLUTION

There is very little work addressing the problem of color SR and the most common solution involves applying monochrome SR algorithms to each of the color channels independently<sup>19</sup>. Another approach is transferring the problem to a different color space where chrominance layers are separated from luminance, and where SR is applied to the luminance channel only<sup>20</sup>. In this section, we review the work of Ref.[21] which details the problems inherent to color SR and proposes a novel algorithm for producing a high quality color image from a collection of LR color (filtered) images.

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

Numerous single-frame demosaicing methods have been proposed through the years<sup>22-26</sup>, yet almost none of them (but the method in Ref.[27, 28]) to date are directly applicable to the problem of multi-frame color demosaicing. In fact, the geometry of the single-frame and multi-frame demosaicing problems are fundamentally different, making it impossible to simply cross apply traditional demosaicing algorithms to the multi-frame situation. To better understand the multi-frame demosaicing problem, we offer an example for the simple case of translational motion. Figure 3 illustrates the pattern of sensor measurements in the HR image grid. In such situations, the sampling pattern is quite arbitrary depending on the relative motion of the LR images. This necessitates a different demosaicing algorithm than those designed for the original Bayer pattern.

The challenge of multi-frame color SR is much more difficult than that of monochrome imaging and should not be solved by applying monochrome methods independently on each color channel for several reasons. First, the



**Figure 3.** Fusion of 7 Bayer pattern LR images with relative translational motion (the figures in the left side of the accolade) results in an HR image ( $\widehat{Z}$ ) that does not follow Bayer pattern (the figure in the right side of the accolade). The symbol “?” represents the HR pixel values that were undetermined after the Shift-And-Add step (as a result of insufficient LR frames).

additional down-sampling of each color channel due to the color filter array makes the independent reconstruction of each channel much harder. For many situations, the information contained in a single color channel is insufficient to solve such a highly ill-posed problem and therefore acceptable performance is virtually impossible to achieve. Second, there are natural correspondences between the color channels which should be leveraged during the reconstruction process. Third, the human visual system is very sensitive to certain artifacts in color images which can only be avoided by processing all of the color channels together. Merely applying a simple demosaicing algorithm prior to SR would only amplify such artifacts and lead to sub optimal performance. Instead, all three channels must be estimated simultaneously to maximize the overall color SR performance.

In Ref.[21], we propose a computationally efficient method to fuse and demosaic a set of LR color frames (which may have been color filtered by any CFA) resulting in a color image with higher spatial resolution and reduced color artifacts. To address the challenges specific to color SR, additional regularization penalty functions are required. To facilitate the explanation, we represent the HR color channels as  $\underline{X}_G$ ,  $\underline{X}_B$ , and  $\underline{X}_R$ . The final cost function consists of the following terms:

1. **Data Fidelity:** Again, we choose a data fidelity penalty term using the  $L_1$  norm to add robustness:

$$J_0(\underline{X}) = \sum_{i=R,G,B} \sum_{k=1}^N \|D_i(k)H(k)F(k)\underline{X}_i - \underline{Y}_i(k)\|_1 \quad (7)$$

where  $\underline{Y}_i(t)$  is the red, green or blue component of the color (filtered) LR frame, and  $D_i$  represents the down-sampling effects of the CCD and color filter arrays on red, green, or blue color bands. As in the previous section, the fast two stage method for the case of common, space-invariant blur and global translation is also applicable to the multi-frame demosaicing method, leading to an initial median Shift-And-Add operation on Bayer-filtered LR data followed by a deblurring step. Thus, the first stage of the algorithm is the median Shift-And-Add operation of producing a blurry HR image  $\widehat{Z}_{R,G,B}$  (e.g. the left side of the accolade in Figure 3). In this case, however, the median operation is applied in a pixel-wise fashion to each of the color channels independently. See Ref.[21] for more details.

2. **Luminance Regularization:** Here, we use a penalty term regularizing the luminance component of the HR image instead of each color channel separately. This is because the human eye is more sensitive to the details in the luminance component of an image than the details in the chrominance components<sup>25</sup>. Therefore, we apply the BTV regularization to the luminance component to offer robust edge preservation.

The luminance regularization term is similar to (3)

$$J_1(\underline{X}) = \sum_{l,m=-P}^P \alpha^{|m|+|l|} \|\underline{X}_L - S_x^l S_y^m \underline{X}_L\|_1, \quad (8)$$

where the luminance image can be calculated as the weighted sum of the RGB components<sup>29</sup>  $\underline{X}_L = 0.299\underline{X}_R + 0.597\underline{X}_G + 0.114\underline{X}_B$ .

3. **Chrominance Regularization:** This penalty term ensures the smoothness in the chrominance components of the HR image. This removes many of the color artifacts objectionable to the human eye. Again, the two chrominance channels  $\underline{X}_{C_1}$  and  $\underline{X}_{C_2}$  can be calculated as the weighted combination of the RGB images using the weights<sup>29</sup>  $(-0.169,-0.331,0.5)$  for  $C_1$  and  $(0.5,-0.419,-0.081)$  for  $C_2$ . As the human eye is less sensitive to the chrominance channel resolution, it can be smoothed more aggressively. Therefore, the following regularization is an appropriate method for smoothing the chrominance term:

$$J_2(\underline{X}) = \|\Lambda \underline{X}_{C_1}\|_2^2 + \|\Lambda \underline{X}_{C_2}\|_2^2, \quad (9)$$

where  $\Lambda$  is the matrix realization of a high-pass operator such as the Laplacian filter.

4. **Orientation Regularization:** This term penalizes the non-homogeneity of the edge orientation across the color channels. Although different bands may have larger or smaller gradient magnitudes at a particular edge, it is reasonable to assume that all color channels have the same edge orientation. That is, if a vertical (or horizontal) edge appears in the red band, an edge with similar orientation is likely to appear in the green and blue bands. Following Ref.[24, 30], minimizing the vector product norm of any two adjacent color pixels forces different bands to have similar edge orientation. With some modifications to what was proposed in Ref.[24], our orientation penalty term is a differentiable cost function:

$$J_3(\underline{X}) = \sum_{l,m=-1}^1 [\|\underline{X}_G \odot S_x^l S_y^m \underline{X}_B - \underline{X}_B \odot S_x^l S_y^m \underline{X}_G\|_2^2 + \|\underline{X}_B \odot S_x^l S_y^m \underline{X}_R - \underline{X}_R \odot S_x^l S_y^m \underline{X}_B\|_2^2 + \|\underline{X}_R \odot S_x^l S_y^m \underline{X}_G - \underline{X}_G \odot S_x^l S_y^m \underline{X}_R\|_2^2] \quad (10)$$

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

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

$$\hat{\underline{X}} = \underset{\underline{X}}{\text{ArgMin}} [J_0(\underline{X}) + \lambda_1 J_1(\underline{X}) + \lambda_2 J_2(\underline{X}) + \lambda_3 J_3(\underline{X})] \quad (11)$$

In Ref.[21], a method for applying a steepest descent algorithm to minimize this cost function is proposed. Interestingly, this cost function can also be applied to full color images where an unknown demosaicing algorithm has already been applied prior to the SR process.

Figure 4 illustrates the performance of the proposed method. We used 31 uncompressed, raw CFA images from a video camera (based on Zoran 2MP CMOS Sensors using the coach chipset). We applied the sophisticated demosaicing method of Ref.[23] to demosaic each of these LR frames, individually. Figure 4(a) shows one of these images. To increase the spatial resolution by a factor of three, we applied the proposed multi-frame color SR method on the demosaiced<sup>23</sup> LR sequence. Figure 4(b) shows such HR color SR image. Finally, we directly applied the proposed multi-frame demosaicing method on the raw CFA data to increase the spatial resolution by the same factor of three. Figure 4(c) shows this HR image, where color artifacts are significantly reduced and the reconstructed image is sharper than the result of other methods. More details about this experiment are given in Ref.[21].



**Figure 4.** Multi-frame color SR implemented on a real data sequence. (a) shows one of the LR demosaiced<sup>23</sup> input images. (b) is the result of applying our method on a sequence of already demosaiced<sup>23</sup> LR images, and (c) is the result of applying our method on the original un-demosaiced raw LR images.

#### 4. DYNAMIC SUPER-RESOLUTION

In the previous sections, our approach to the SR problem was to fuse a set of LR images and reconstruct a single HR image. We refer to this as the *static* SR method, since it does not exploit the temporal evolution of the process. In this section, we briefly review the work of Ref.[31] in which we consider SR applied on an image sequence, producing a sequence of HR images; a process known as *dynamic* SR. The natural approach, as most existing works so far suggest, is to apply the static SR on a set of images with the  $t$ -th frame as a reference\*\*, produce the SR output, and repeat this process all over again per each temporal point. However, the memory and computational requirements for the static process are so taxing as to preclude its direct application to the dynamic case.

In this section, similar to Section 2.2, we limit our model to the case of translational motion for several reasons. First, such a motion model allows for an extremely fast and memory efficient dynamic SR algorithm. Second, while simple, the model fairly well approximates the motion contained in many image sequences, where the scene is stationary and only the camera moves in approximately linear fashion. Third, for sufficiently high frame rates most motion models can be (at least locally) approximated by the translational model. Finally, we believe that an in-depth study of this simple case yields much insight into the more general cases of motion in dynamic SR.

We use the general linear dynamic forward model for the SR problem as in Ref.[32, 33]. By substituting  $\underline{Z}(t) = H\underline{X}(t)$ , we represent our simplified state-space model<sup>34</sup> describing how the blurry super-resolved images are related to each other through time as

$$\underline{Z}(t) = F(t)\underline{Z}(t-1) + \underline{W}(t), \quad (12)$$

and

$$\underline{Y}(t) = D\underline{Z}(t) + \underline{V}(t). \quad (13)$$

where the so-called system noise  $\underline{W}(t)$ , is assumed to be additive zero mean Gaussian with  $C_w(t)$  as its covariance matrix. Note that the closer the overlapping regions of  $\underline{Z}(t)$  and the motion compensated  $\underline{Z}(t-1)$  are, the smaller  $C_w(t)$  becomes. Therefore,  $C_w(t)$  in a sense reflects the accuracy of the motion estimation process (as  $\underline{W}(t)$  also captures unmodeled motions), and for overlapped regions it is directly related to the motion estimation covariance matrix. The noise vector  $\underline{V}(t)$  is assumed to be additive, zero mean, white Gaussian noise. Thus, its covariance matrix is  $C_v(t) = \sigma_v^2 I$ . We further assume that  $\underline{W}(t)$  and  $\underline{V}(t)$  are independent of each other.

\*\*To emphasis on the temporal order of the frames, unlike other sections of this paper, we use the term “ $t$ ” to indicate the frame numbers.

With this alternative definition of the state of the dynamic system, the solution of the inverse problem at hand decomposes, without loss of optimality, into the much simpler sub-tasks of fusing the available images to compute the estimated blurry image  $\hat{\underline{Z}}(t)$  (dynamic Shift-and-Add), followed by a deblurring/interpolation step, estimating  $\hat{\underline{X}}(t)$  from  $\hat{\underline{Z}}(t)$ . The dynamic Shift-and-Add step treats the three color bands separately. For instance, only the red band values in the input frames,  $\underline{Y}(t)$ , contribute to the reconstruction of the red band values in  $\hat{\underline{Z}}(t)$ . Figure 5 shows an overall block diagram of the dynamic SR process for mosaiced images (the feedback loops are eliminated to simplify the diagram).

We next study the application of KF to estimate  $\underline{Z}(t)$ . In general, the application of KF requires the update of the state-vector's covariance matrix per each temporal point, and this update requires an inversion of the state-vector's covariance matrix, implying a prohibitive amount of computations and memory.

Fast and memory efficient alternative ways are to be found, and such methods were first proposed in the context of the dynamic SR in Ref.[32, 33]. In this paper, we show that significant further speedups are achieved for the case of translational motion and common space-invariant blur.

The following defines the forward Kalman propagation and update equations<sup>34</sup>, that account for a causal (on-line) process. We assume that at time  $t - 1$  we already have the mean-covariance pair,  $(\hat{\underline{Z}}(t - 1), \hat{\Pi}(t - 1))$ , and those should be updated to account for the information obtained at time  $t$ . We start with the covariance matrix update based on Equation (12),

$$\tilde{\Pi}(t) = F(t)\hat{\Pi}(t - 1)F^T(t) + C_w(t). \quad (14)$$

The KF gain matrix is given by

$$K(t) = \tilde{\Pi}(t)D^T[C_v(t) + D\tilde{\Pi}(t)D^T]^{-1}. \quad (15)$$

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

$$\hat{\underline{Z}}(t) = F(t)\hat{\underline{Z}}(t - 1) + K(t)[\underline{Y}(t) - DF(t)\hat{\underline{Z}}(t - 1)]. \quad (16)$$

The final stage requires the update of the covariance matrix, based on Equation (13),

$$\hat{\Pi}(t) = \text{Cov}\left(\hat{\underline{Z}}(t)\right) = [\mathbf{I} - K(t)D]\tilde{\Pi}(t). \quad (17)$$

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

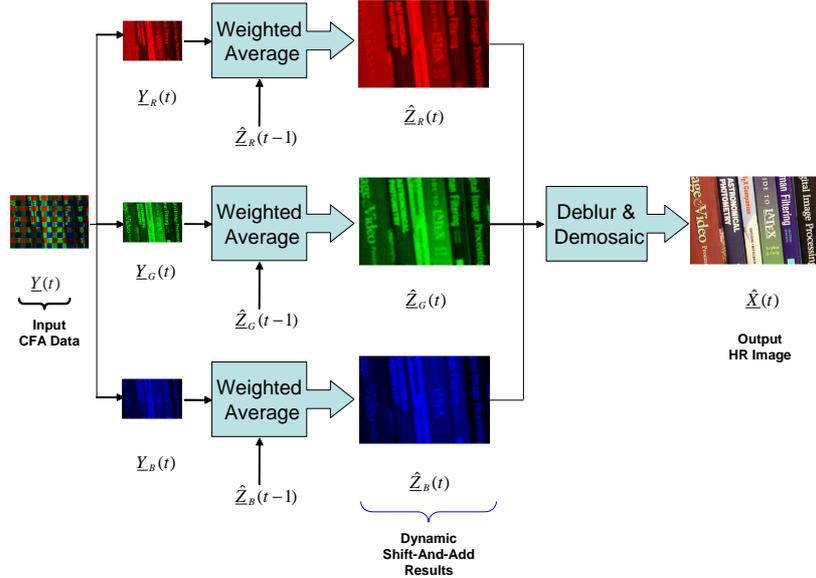
While in general the above equations require the propagation of intolerably large matrices in time, if we refer to  $C_w(t)$  as a diagonal matrix, then  $\tilde{\Pi}(t)$ , and  $\hat{\Pi}(t)$  are diagonal matrices as well. We verified such property in Ref.[31] which is a key assumption in transferring the general KF into a simple and fast pixel-by-pixel procedure. This algorithm is described in details in Ref.[31], suitable for causal, real-time processing, as it estimates the HR frames from the previously seen LR frames. However, often times SR is performed off-line, and therefore a more accurate estimate of an HR frame at a given time is possible by using both previous and future LR frames. Based on the fixed-interval smoothing method of Rauch-Tung-Striebel<sup>35</sup>, we have introduced such algorithm in Ref.[31].

Any of the above procedures result in a blurry estimate of the HR image sequence ( $\underline{X}(t)$ ), which we refer hereafter to as the dynamic Shift-and-Add sequence ( $\hat{\underline{Z}}(t)$ ). Each image in this sequence is later deblurred (and possibly interpolated) to result in the HR final image. As explained in details in Ref.[31], the deblurring-interpolation process for the gray scale images is very similar to the method explained in Section 2.2:

$$\hat{\underline{X}}(t) = \underset{\underline{X}(t)}{\text{ArgMin}} \left[ \|\Phi(t)(H\underline{X}(t) - \hat{\underline{Z}}(t))\|_2^2 + \lambda \sum_{l,m=-P}^P \alpha^{|m|+|l|} \|\underline{X}(t) - S_x^l S_y^m \underline{X}(t)\|_1 \right]. \quad (18)$$

The deblurring-interpolation step for the color case is also very similar to the one explained in Section 3:

$$\hat{\underline{X}}(t) = \underset{\underline{X}(t)}{\text{ArgMin}} [J_0(\underline{X}(t)) + \lambda' J_1(\underline{X}(t)) + \lambda'' J_2(\underline{X}(t)) + \lambda''' J_3(\underline{X}(t))] . \quad (19)$$



**Figure 5.** Block diagram representation of the overall dynamic SR process for color filtered images. The feedback loops are omitted to simplify the diagram. Note that  $\hat{\underline{Z}}_{i \in \{R, G, B\}}(t)$  represents the dynamic Shift-and-Add estimates of the Red, Green and Blue bands.

where the first term is replaced by

$$J_0(\underline{X}(t)) = \sum_{i=R, G, B} \|\Phi_i(t) (H\hat{\underline{X}}_i(t) - \hat{\underline{Z}}_i(t))\|_2^2, \quad (20)$$

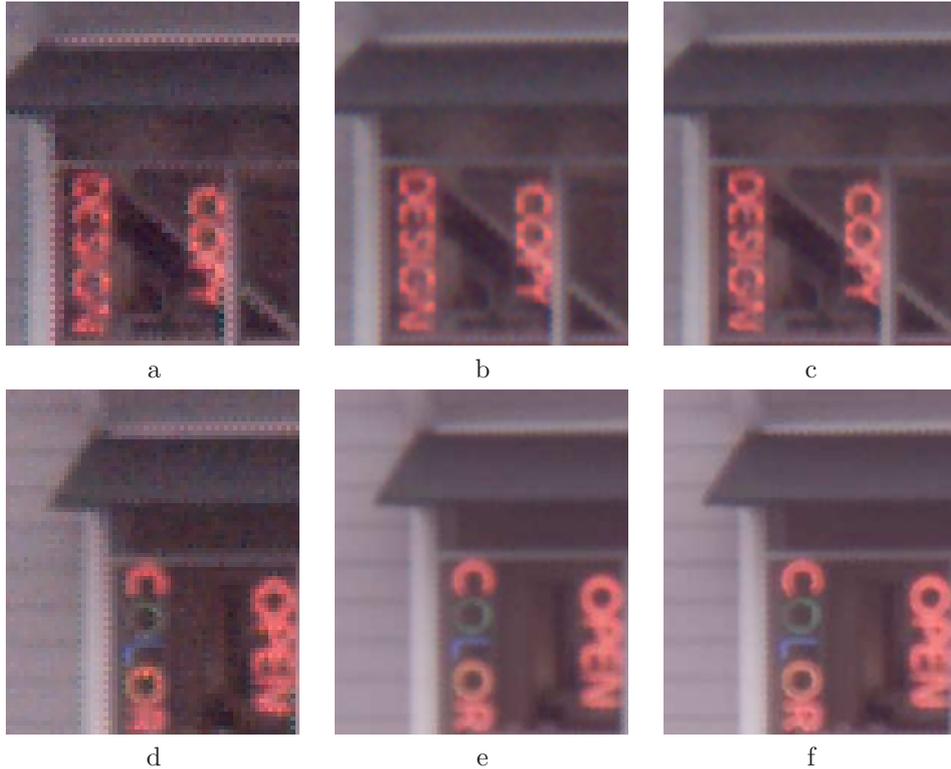
and all other terms are similar to the ones formulated in Section 3. Note that  $F(t)\hat{\underline{X}}(t-1)$  is a suitable candidate to initialize  $\hat{\underline{X}}^0(t)$ , since it follows the KF prediction of the state-vector updates and therefore only few iterations are required in this minimization task.

Figure 6 shows an example of the proposed dynamic SR algorithm for a couple of frames of a 74 uncompressed, raw CFA sequence from a video camera (based on Zoran 2MP CMOS Sensors). We applied the method of Ref.[22] to demosaic each of these LR frames, individually. Figure 6(a) and Figure 6(d) show frames #1 and #69 of this sequence, respectively. To increase the spatial resolution by a factor of three in each axis, we applied the proposed dynamic SR method on the raw CFA data. Figure 6(b) shows the smoothing Shift-and-Add result for the frame #1 (Figure 6(e) shows the corresponding image for the frame #69). These frames were further deblurred-demosaiced and the results are shown in Figures 6(c) and Figures 6(f) for the frames #1 and #69, respectively. More details about this experiment are given in Ref.[31].

## 5. CONSTRAINED, ROBUST MULTI-FRAME MOTION ESTIMATION

So far in this paper, we assumed that the relative motion vectors are known (estimated in a separate process). We acknowledged the fact that oftentimes the estimated motion vectors do not have the subpixel accuracy required by the classic SR algorithms and thus, robust SR algorithms which are relatively less sensitive to the errors in motion estimation are necessary. Indeed, more accurate estimates of motion vectors improves the performance of any SR method (including the robust ones).

Numerous image registration techniques have been developed throughout the years<sup>36</sup>. These methods are mainly developed to estimate the relative motion between *a pair* of frames. For the cases where several images are to be registered with respect to each other (e.g. SR applications), two simple strategies are commonly used.



**Figure 6.** A sequence of 74 real-world LR uncompressed color filtered frames (a & d show frames #1 and #69, respectively) are recursively fused, increasing their resolution by the factor of three in each direction (b & e). They were further deblurred and demosaiced, resulting in images with much higher-quality than the input LR frames (c & f).

The first is to register all frames with respect to a single reference frame<sup>8</sup> (often referenced to as the *anchored* method). The other popular strategy is the *progressive* registration method<sup>17,31</sup>, where images in a sequence are registered in pairs, with one image in each pair acting as the reference frame.

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 between any pair of frames to lie in a space whose geometry and structure is conveniently described.

In this section, we briefly study such priors and propose an optimal method for exploiting them which results in very accurate estimates of the relative motion in a sequence. The details of this procedure are described in Ref.[37]. Our proposed method is inspired by the Bundle Adjustment<sup>38</sup> (BA) technique which is a general, yet computationally expensive method for producing a jointly optimal 3D structure and viewing parameters. It is important to note that BA is not intended for motion estimation in 2-D images. On another front, our method is different to the recent multi-frame registration methods<sup>39-41</sup>, as it considers the robustness issue.

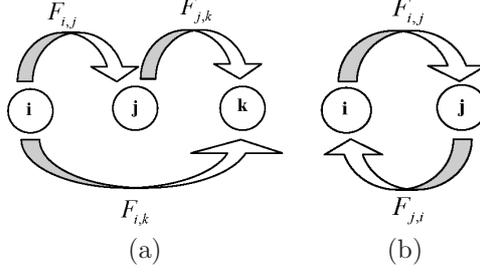
### 5.1. Constrained Motion Estimation Formulation

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

$$\underline{Y}(i) = \mathbf{F}_{i,j}\{\underline{Y}(j)\},$$

where  $\underline{Y}_i$  and  $\underline{Y}_j$  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 are inherent intuitive relationships between these pair-wise registration operators. In particular, the first condition dictates that the



**Figure 7.** The consistent flow properties: (a) Jacobi Identity and (b) Skew Anti-Symmetry.

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 7(a), taking any triplet of frames  $i$ ,  $j$ , and  $k$ , we have the first motion consistency condition as:

$$\forall i, j, k \in \{1, \dots, N\}, \quad \mathbf{F}_{i,k} = \mathbf{F}_{i,j} \circ \mathbf{F}_{j,k}. \quad (21)$$

The second rather obvious (but hardly ever used) consistency condition states that the composition of the operator mapping frame  $i$  to  $j$  with the operator mapping frame  $j$  to  $i$  should yield the identity operator. This is illustrated in Figure 7(b). Put another way,

$$\forall i, j \in \{1, \dots, N\}, \quad \mathbf{F}_{j,i} = \mathbf{F}_{i,j}^{-1}. \quad (22)$$

These natural conditions define an algebraic group structure (a Lie algebra) in which the operators reside. Therefore, any estimation of motion between frames of an image sequence could take these conditions into account. In particular, the optimal motion estimation strategy can be described as an estimation problem over a group structure, which has been studied before in other contexts<sup>42</sup>.

The above properties describe what is known as the Jacobi condition, and the skew anti-symmetry relations<sup>40</sup>. For some practical motion models (e.g. constant motion or the affine model), the relevant operators could be further simplified. For example, in the case of translational (constant) motion, the above conditions can be described by simple linear equations relating the (single) motion vectors between the frames:

$$\forall i, j, k \in \{1, \dots, N\}, \quad \underline{\delta}_{i,k} = \underline{\delta}_{i,j} + \underline{\delta}_{j,k}, \quad (23)$$

where  $\underline{\delta}_{i,j}$  is the motion vector field between the frames  $i$  and  $j$ . Note that  $\underline{\delta}_{i,i} = \underline{0}$ , and therefore the skew anti-symmetry condition is represented by (23), when  $k = i$ .

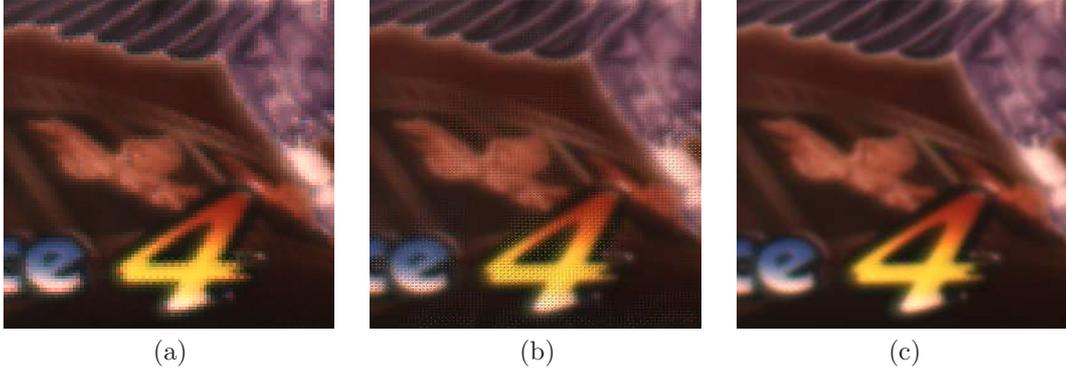
In a general setting, the optimal solution to the multi-frame registration problem can be obtained by

$$\hat{\underline{\delta}} = \underset{\underline{\delta}}{\text{ArgMin}} \rho(\underline{Y}, \underline{\delta}) \text{ such that } \Upsilon(\underline{\delta}) = 0, \quad (24)$$

where  $\rho$  represents a motion-related cost function (e.g. penalizing deviation from brightness constraint, or a phase-based penalty), and  $\Upsilon$  captures the constraints discussed above.

Special cases of this general formulation for different motion models are discussed in Ref.[37]. To get a feeling for this general formulation, we address the translational motion case, with  $\rho$  representing the Optical Flow model:

$$\rho(\underline{Y}, \underline{\delta}) = \sum_{\substack{i,j=1 \\ i \neq j}}^N \|\underline{Y}_i^{(x)} \delta_{i,j}^{(x)} + \underline{Y}_i^{(y)} \delta_{i,j}^{(y)} + \underline{Y}_{i,j}^{(t)}\|_1, \quad (25)$$



**Figure 8.** Experimental registration results for a real sequence. (a) One input LR frame after demosaicing. (b) Single Reference HR registration. (c) Optimal HR registration.

where  $\underline{Y}_i^{(x)}$  and  $\underline{Y}_i^{(y)}$  are the spatial derivatives (in  $x$  and  $y$  directions) of the  $i^{th}$  frame, and  $\underline{Y}_{i,j}^{(t)}$  is the temporal derivative (e.g., the difference between frames  $i$  and  $j$ ). Here the motion vector field  $\underline{\delta}_{i,j}$  is spatially constant, and it can be represented by the scalar components  $\delta_{i,j}^{(x)}$  and  $\delta_{i,j}^{(y)}$  in  $x$  and  $y$  axes, respectively, and for  $1 \leq i, j \leq N$ .

Using this, the translational consistency condition as in Equation (23) is then formulated as

$$\Upsilon(\underline{\delta}) : C\underline{\delta} = 0, \quad (26)$$

where the unknown motion vector  $\underline{\delta}$  has all the  $2N(N-1)$  entries  $\delta_{i,j}^{(x)}$  and  $\delta_{i,j}^{(y)}$  stacked to a vector. The constraint matrix  $C$  is of size  $[2(N-1)^2 \times 2N(N-1)]$ . Each row in  $C$  has only two or three non-zero ( $\pm 1$ ) elements representing the skew anti-symmetry and Jacobi identity conditions in (23), respectively. Non-linear programming (e.g. “fmincon” function in MATLAB) can be used to minimize this cost functions. To reduce computational complexity and increase robustness in Ref.[37] we proposed to optimize a modified cost function which includes a term representing the “soft” version of the constraints as:

$$\hat{\underline{\delta}} = \underset{\underline{\delta}}{\text{ArgMin}} \{ \rho(\underline{Y}, \underline{\delta}) + \lambda_m \Upsilon(\underline{\delta}) \}, \quad (27)$$

where  $\lambda_m$  represents the strength of the regularizing term.

A real experiment was conducted aligning 27 color-filtered LR (LR) images. One of these LR frames after demosaicing<sup>22</sup> is shown in Fig.8(a). We used the method described in Ref.[43] to compute the motion vectors in an *anchored* fashion. Then, the method of Ref.[21] was used to construct a HR image, by registering these images on a finer grid (resolution enhancement factor of three in  $x$  and  $y$  directions). Figure 8(b) shows the HR reconstruction using this method with clear mis-registration errors. Finally, the result of applying the proposed multi-frame registration method is shown in Fig.8(c), with almost no visible mis-registration error. More details about this experiment are given in Ref.[37].

## 6. CONCLUSION

In this paper, we presented SR algorithms to enhance the quality of a set of noisy, blurred, and possibly color (filtered) images to produce one (or a set) of monochromatic or color HR images with less noise, aliasing, and blur effects. We also presented a very accurate solution for the multi-frame motion estimation problem, which is often considered as a bottleneck to SR performance<sup>44</sup>. Our methods were based on a fast and robust MAP framework. We showed that the two main problems that usually restrict the real world (practical) application of classical SR methods, namely computational complexity-memory requirements, and outlier effects can be overcome using the presented methods.

## ACKNOWLEDGMENTS

This work was supported in part by the US Air Force Grant F49620-03-1-0387, 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. M. Elad's work was supported in part by Jewish Communities of Germany Research Fund. We thank Prof. Ron Kimmel of the Technion-Israel Institute of Technology for providing us with the code that implements the method in Ref.[23]. We would like to thank Lior Zimet and Erez Galil from Zoran Corp. for providing the camera used to produce the raw CFA images of experiments in Figure 4 and Figure 6. We would also like to thank Eyal Gordon from the Technion-Israel Institute of Technology for helping us capture the raw CFA images used in the Figure 8 experiment.

## REFERENCES

1. 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.
2. M. Ng and N. Bose, "Mathematical analysis of super-resolution methodology," *IEEE Signal Processing Mag.* **20**, pp. 62–74, May 2003.
3. S. Borman, *Topics in Multiframe Superresolution Restoration*. PhD thesis, University of Notre Dame, Notre Dame, IN, May 2004.
4. M. Ben-Ezra, A. Zomet, and S. Nayar, "Video super-resolution using controlled subpixel detector shifts," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**, pp. 977–987, June 2004.
5. 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* **10**, pp. 1187–1193, Aug. 2001.
6. N. Nguyen, P. Milanfar, and G. H. Golub, "A computationally efficient image superresolution algorithm," *IEEE Trans. Image Processing* **10**, pp. 573–583, Apr. 2001.
7. A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super resolution," in *In Proc. of the Int. Conf. on Computer Vision and Patern Recognition (CVPR)*, **1**, pp. 645–650, Dec. 2001.
8. S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super-resolution," *IEEE Trans. Image Processing* **13**, pp. 1327–1344, Oct. 2004.
9. S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Robust shift and add approach to super-resolution," *In Proc. of the 2003 SPIE Conf. on Applications of Digital Signal and Image Processing*, pp. 121–130, Aug. 2003.
10. A. M. Tekalp, *Digital video processing*. Prentice-Hall, 1995.
11. L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D* **60**, pp. 259–268, Nov. 1992.
12. T. F. Chan, S. Osher, and J. Shen, "The digital TV filter and nonlinear denoising," *IEEE Trans. Image Processing* **10**, pp. 231–241, Feb. 2001.
13. C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *In Proc. of IEEE Int. Conf. on Computer Vision*, pp. 836–846, Jan. 1998.
14. M. Elad, "On the bilateral filter and ways to improve it," *IEEE Trans. Image Processing* **11**, pp. 1141–1151, Oct. 2002.
15. 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.
16. A. Zomet and S. Peleg, "Efficient super-resolution and applications to mosaics," in *In Proc. of the Int. Conf. on Pattern Recognition (ICPR)*, pp. 579–583, Sept. 2000.
17. L. Teodosio and W. Bender, "Salient video stills: Content and context preserved," in *In Proc. of the First ACM Int. Conf. on Multimedia*, **10**, pp. 39–46, Aug. 1993.
18. M. C. Chiang and T. E. Boult, "Efficient super-resolution via image warping," *Image and Vision Computing* **18**, pp. 761–771, July 2000.
19. 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.

20. M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP:Graph. Models Image Process* **53**, pp. 231–239, 1991.
21. S. Farsiu, M. Elad, and P. Milanfar, "Multi-frame demosaicing and super-resolution of color images," to appear in *IEEE Trans. Image Processing* **15**, Jan. 2005.
22. 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.
23. R. Kimmel, "Demosaicing: Image reconstruction from color ccd samples," *IEEE Trans. Image Processing* **8**, pp. 1221–1228, Sept. 1999.
24. D. Keren and M. Osadchy, "Restoring subsampled color images," *Machine Vision and applications* **11**(4), pp. 197–202, 1999.
25. Y. Hel-Or and D. Keren, "Demosaicing of color images using steerable wavelets," Tech. Rep. HPL-2002-206R1 20020830, HP Labs Israel, 2002.
26. D. Alleysson, S. Süsstrunk, and J. Hraut, "Color demosaicing by estimating luminance and opponent chromatic signals in the Fourier domain," in *In Proc. of the IS&T/SID 10th Color Imaging Conf.*, pp. 331–336, Nov. 2002.
27. A. Zomet and S. Peleg, "Multi-sensor super resolution," in *In Proc. of the IEEE Workshop on Applications of Computer Vision*, pp. 27–31, December 2002.
28. T. Gotoh and M. Okutomi, "Direct super-resolution and registration using raw CFA images," in *In Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, **2**, pp. 600–607, July 2004.
29. W. K. Pratt, *Digital image processing*. New York: John Wiley & Sons, INC., 3rd ed., 2001.
30. D. Keren and A. Gotlib, "Denoising color images using regularization and correlation terms," *Journal of Visual Communication and Image Representation* **9**, pp. 352–365, Dec. 1998.
31. S. Farsiu, M. Elad, and P. Milanfar, "Video-to-video dynamic superresolution for grayscale and color sequences," To appear in *EURASIP Journal of Applied Signal Processing*, 2005.
32. M. Elad and A. Feuer, "Superresolution restoration of an image sequence: adaptive filtering approach," *IEEE Trans. Image Processing* **8**, pp. 387–395, Mar. 1999.
33. M. Elad and A. Feuer, "Super-resolution reconstruction of image sequences," *IEEE Trans. Pattern Analysis and Machine Intelligence* **21**, pp. 817–834, Sept. 1999.
34. S. M. Kay, *Fundamentals of statistical signal processing:estimation theory*, vol. I, Englewood Cliffs, New Jersey: Prentice-Hall, 1993.
35. H. Rauch, F. Tung, and C. Striebel, "Maximum likelihood estimates of dynamic linear systems," *AIAA Journal* **3**, pp. 1445–1450, Aug. 1965.
36. L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys* **24**, pp. 325–376, December 1992.
37. S. Farsiu, M. Elad, and P. Milanfar, "Constrained, globally optimal, multi-frame motion estimation," in *To appear in the Proc. of the 2005 IEEE Workshop on Statistical Signal Processing*, July 2005.
38. B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustment – a modern synthesis," in *Vision Algorithms: Theory and Practice, Lecture Notes in Computer Science* **1883**, pp. 298–372, 2000.
39. H. Sawhney, S. Hsu, and R. Kumar, "Robust video mosaicing through topology inference and local to global alignment," in *ECCV*, **2**, pp. 103–119, 1998.
40. V. Govindu, "Lie-algebraic averaging for globally consistent motion estimation," in *In Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, **1**, pp. 684–691, July 2004.
41. Y. Sheikh, Y. Zhai, and M. Shah, "An accumulative framework for the alignment of an image sequence," in *ACCV*, Jan. 2004.
42. S. Marcus and A. Willsky, "Algebraic structure and finite dimensional nonlinear estimation," *SIAM J. Math. Anal.*, pp. 312–327, Apr. 1978.
43. J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," *In Proc. of the European Conf. on Computer Vision*, pp. 237–252, May 1992.
44. D. Robinson and P. Milanfar, "Statistical performance analysis of super-resolution," To Appear in *IEEE Transactions on Image Processing*, 2005.