Advances and Challenges in Super-Resolution

Sina Farsiu,[1] Dirk Robinson,[1] Michael Elad,[2] Peyman Milanfar[1]

[1] Electrical Engineering Department, University of California, Santa Cruz CA 95064

[2] Computer Science Department, The Technion–Israel Institute of Technology, Israel

Received 30 January 2004; accepted 15 March 2004

ABSTRACT: Super-Resolution reconstruction produces one or a set of high-resolution images from a sequence of low-resolution frames. This article reviews a variety of Super-Resolution methods proposed in the last 20 years, and provides some insight into, and a summary of, our recent contributions to the general Super-Resolution problem. In the process, a detailed study of several very important aspects of Super-Resolution, often ignored in the literature, is presented. Specifically, we discuss robustness, treatment of color, and dynamic operation modes. Novel methods for addressing these issues are accompanied by experimental results on simulated and real data. Finally, some future challenges in Super-Resolution are outlined and discussed. © 2004 Wiley Periodicals, Inc. Int J Imaging Syst Technol, 14, 47–57, 2004; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/ima.20007

Key words: Super-Resolution; demosaicing; inverse problem; dynamic Super-Resolution; image-reconstruction; robust estimation; robust regularization

I. INTRODUCTION

On the quest to achieve high resolution imaging systems, one quickly runs into the problem of diminishing returns. Specifically, the imaging chips and optical components necessary to capture very high-resolution images become prohibitively expensive, costing in the millions of dollars for scientific applications (Parulski et al., 1992). Super-resolution 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 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.

The basic idea behind Super-Resolution is the fusion of a sequence of low-resolution noisy blurred images to produce a higher-resolution image or sequence. Early works on Super-Resolution showed that the aliasing effects in the high-resolution fused image can be reduced (or even completely removed), if a relative sub-pixel motion exits between the undersampled input images (Huang and Tsai, 1984). However, contrary to the naive frequency domain description of this early work, we shall see that, in general, super-resolution is a computationally complex and numerically ill-posed problem. All this makes Super-Resolution one of the most appealing research areas for image processing researchers.

Although several articles have surveyed the different classical Super-Resolution methods and compared their performances (e.g., Borman and Stevenson, 1998; Kang and Chaudhuri, 2003), the intention of this article is to pinpoint the various difficulties inherent to the Super-Resolution problem for a variety of application settings often ignored in the past. We review many of the most recent and popular methods, and outline some of our recent work addressing these issues.

The organization of this article is as follows. In Section II we study Super-Resolution as an inverse problem and address related regularization issues. In Section III we analyze a general model for imaging systems applicable to various scenarios of Super-Resolution. In Section IV we describe three different application settings and our approaches to dealing with them. Specifically, we address the problem of robust Super-Resolution, the treatment of color images and mosaiced sources, and dynamic Super-Resolution. Finally, we conclude with a list of challenges to be addressed in future work on Super-Resolution.

II. SUPER-RESOLUTION AS AN INVERSE PROBLEM

Super-resolution algorithms attempt to extract the high-resolution image corrupted by the limitations of the optical imaging system. This type of problem is an example of an inverse problem, wherein the source of information (high-resolution image) is estimated from the observed data (low-resolution image or images). Solving an inverse problem in general requires first constructing a forward model. By far, the most common forward model for the problem of Super-Resolution is linear in form:

$$\underline{Y}(t) = M(t)\underline{X}(t) + \underline{V}(t), \tag{1}$$

where $\underline{Y}$ is the measured data (single or collection of images), M represents the imaging system, $\underline{X}$ is the unknown high-resolution image or images, $\underline{V}$ is the random noise inherent to any imaging

Correspondence to: S. Farsiu; e-mail: farsiu@ee.uese.edu

Grant sponsor: This work was supported in part by the National Science Foundation Grant CCR-9984246, 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.

© 2004 Wiley Periodicals, Inc.

system, and t represents the time of image acquisition. We use the underscore notation such as $\underline{X}$ to indicate a vector. In this formulation, the image is represented in vector form by scanning the 2D image in a raster or any other scanning format[1] to 1D.

Armed with a forward model, the practitioner of Super-Resolution must explicitly or implicitly [e.g. the POCS-based methods of Patti et al. (1997)] define a cost function to estimate $\underline{X}$ (for now we ignore the temporal aspect of Super-Resolution). This type of cost function assures a certain fidelity or closeness of the final solution to the measured data. Historically, the construction of such a cost function has been motivated from either an algebraic or a statistical perspective. Perhaps the cost function most common to both perspectives is the least-squares (LS) cost function, which minimizes the L_2 norm of the residual vector,

$$\hat{\underline{X}} = \underset{\underline{X}}{\operatorname{argmin}}\, J(\underline{X}) = \underset{\underline{X}}{\operatorname{argmin}}\, \|\underline{Y} - M\underline{X}\|_2^2. \tag{2}$$

For the case where the noise $\underline{V}$ is additive white, zero mean Gaussian, this approach has the interpretation of providing the maximum likelihood estimate of $\underline{X}$ (Elad and Feuer, 1997). We shall show in this paper that such a cost function is not necessarily adequate for Super-Resolution.

An inherent difficulty with inverse problems is the challenge of inverting the forward model without amplifying the effect of noise in the measured data. In the linear model, this results from the very high, possibly infinite, condition number for the model matrix M. Solving the inverse problem, as the name suggests, requires inverting the effects of the system matrix M. At best, this system matrix is ill conditioned, presenting the challenge of inverting the matrix in a numerically stable fashion (Golub and Loan, 1994). Furthermore, finding the minimizer of (2) would amplify the random noise $\underline{V}$ in the direction of the singular vectors (in the Super-Resolution case these are the high spatial frequencies), making the solution highly sensitive to measurement noise. In many real scenarios, the problem is worsened by the fact that the system matrix M is singular. For a singular model matrix M, there is an infinite space of solutions minimizing (2). Thus, for the problem of Super-Resolution, some form of regularization must be included in the cost function to stabilize the problem or constrain the space of solutions.

Traditionally, regularization has been described from both the algebraic and statistical perspectives. In both cases, regularization takes the form of constraints on the space of possible solutions often independent of the measured data. This is accomplished by way of Lagrangian type penalty terms as in

$$J(\underline{X}) = \|\underline{Y} - M\underline{X}\|_2^2 + \lambda\rho(\underline{X}). \tag{3}$$

The function $\rho(\underline{X})$ poses a penalty on the unknown $\underline{X}$ to direct it to a better formed solution. The coefficient λ dictates the strength with which this penalty is enforced. Generally speaking, choosing λ could be either done manually, using visual inspection, or automatically using methods like generalized cross-validation (Lukas, 1993; Nguyen et al., 2001a) L-curve (Hansen and O'Leary, 1993) and other techniques.

[1] Note that this conversion is semantic and bears no loss in the description of the relation between measurements and ideal signal.

Tikhonov regularization, of the form $\rho(\underline{X}) = \|T\underline{X}\|_2^2$, is a widely employed form of regularization, where T is a matrix capturing some aspect of the image such as its general smoothness. This form of regularization has been motivated from an analytic standpoint to justify certain mathematical properties of the estimated solution. For instance, a minimal energy regularization ($T = I$) easily leads to a provably unique and stable solution. Often, however, little attention is given to the effects of such simple regularization on the super-resolution results. For instance, the regularization often penalizes energy in the higher frequencies of the solution, opting for a smooth and hence blurry solution. From a statistical perspective, regularization is incorporated as *a priori* knowledge about the solution. Thus, using the maximum a-posteriori (MAP) estimator, a much richer class of regularization functions emerges, enabling us to capture the specifics of the particular application [e.g., Schultz and Stevenson (1996) captured the piecewise-constant property of natural images by modeling them as Huber-Markov random field data].

Unlike the traditional Tikhonov penalty terms, robust methods are capable of performing adaptive smoothing based on the local structure of the image. For instance, in Section IV.A we offer a penalty term capable of preserving the high-frequency edge structures commonly found in images. The edge-preserving property of this method has been extensively studied (Elad, 2002; Farsiu et al., 2004a; Rudin et al., 1992; Sochen et al., 1998).

In recent years there has also been a growing number of *learning*-based MAP methods, where the regularization-like penalty terms are derived from collections of training samples (Atkins et al., 1999; Baker and Kanade, 2002; Haber and Tenorio, 2003; Zhu and Muford, 1997). For example, in Baker and Kanade (2003) 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. Because of the need to gather a vast amount of examples, often these methods are effective when applied to very specific scenarios, such as faces or text.

Needless to say, the choice of regularization plays a vital role in the performance of any Super-Resolution algorithm.

III. ANALYSIS OF THE FORWARD MODEL

A. General Structure of the Linear Model. In this section, we focus on the construction of the model matrix M. Specifically, we explore the effects of various modeling assumptions relating to the computational efficiency and performance of Super-Resolution algorithms. Primarily, the three terms necessary to capture the image formation process are image motion, optical blur, and the sampling process. These three terms can be modelled as separate matrices by

$$M = DAHF, \tag{4}$$

where F represents the intensity conserving, geometric warp operation capturing image motion, H is the blurring operation due to the optical point spread function[2] (PSF), and D and A represent the effect of sampling by the image sensor. We use both D and A to

[2] A more general imaging model is defined as $M=DAHFH^{atm}$, where H^{atm} represents the effect of the atmosphere and motion blur (Lertrattanapanich and Bose, 2002). However, as in conventional imaging systems (such as video cameras), camera lens/CCD blur has more important effect than the atmospheric blur (which is very important for astronomical images), the effect of H^{atm} is usually ignored in the literature (Farsiu et al., 2004a).

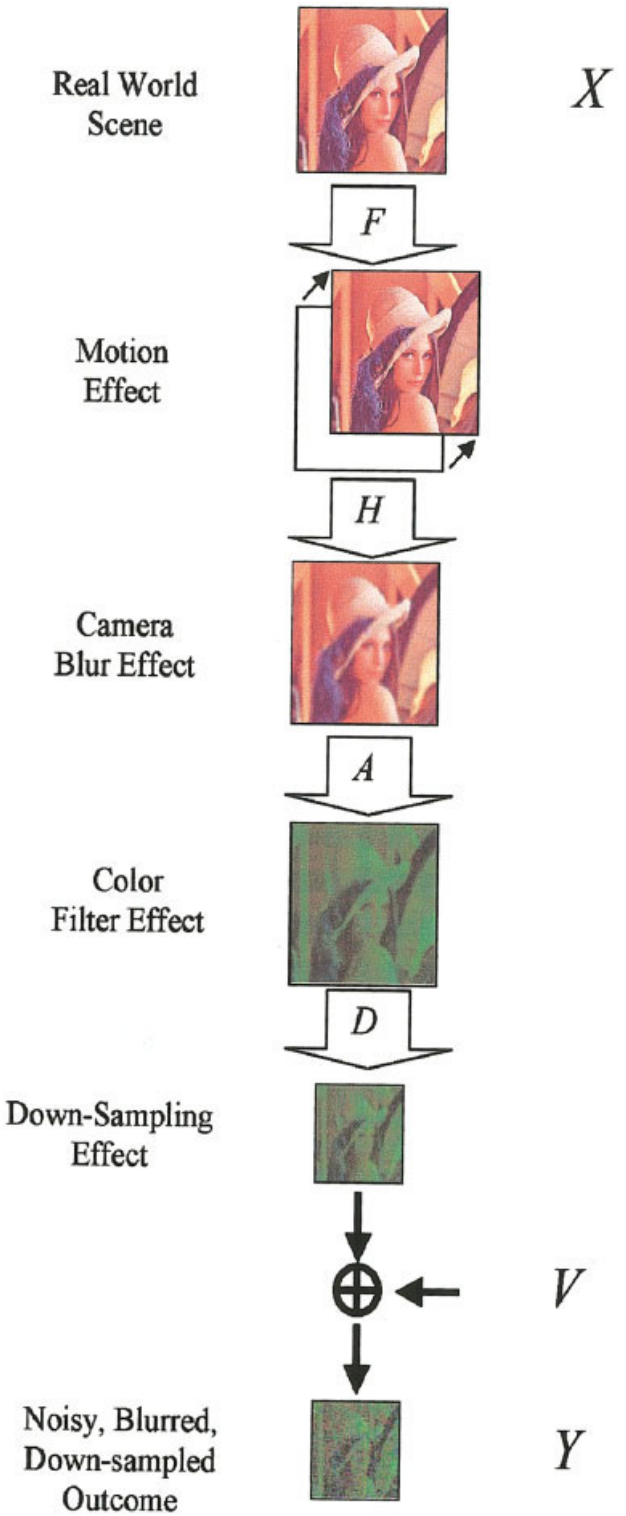

Figure 1. Block diagram representation of (4), where X is the original high-resolution color image, V is the additive noise, and Y is the resulting low-resolution blurred color filtered image.

distinguish between a generic down-sampling operation (or CCD decimation by a factor r) and the sampling operations specific to the color space (color filter effects). Although each of these components could in theory vary in time, for most situations, the down-sampling

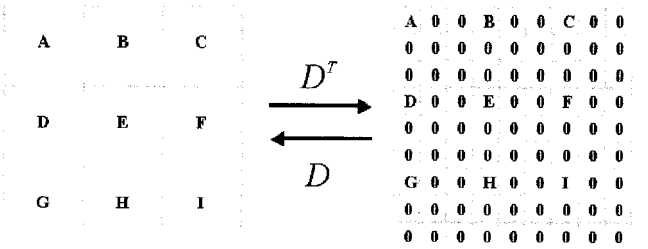

Figure 2. Effect of up-sampling D^T matrix on a 3 × 3 image and down-sampling matrix D on the corresponding 9 × 9 up-sampled image (resolution enhancement factor of 3). In this figure, to give a better intuition the image vectors are reshaped as matrices.

and blurring operations remain constant over time. Figure 1 illustrates the effect of each term in (4).

In ideal situations these modeling terms would capture the actual effects of the image formation process. In practice, however, the models used reflect a combination of computational and statistical limitations. For instance, it is common to assume simple parametric space-invariant blurring functions for the imaging system. This allows the practitioner to utilize efficient and stable algorithms for estimating an unknown blurring function. Or, the choice of resolution enhancement factor r often depends on the number of available low-resolution frames, the computational limitations (exponential in r), and the accuracy of motion estimates. Although this approach is reasonable, it must be understood that incorrect approximations can lead to significant reduction in overall performance.

In our experience, the performance of motion estimation is of paramount importance to the performance of Super-Resolution. In fact, we offer the observation that difficulties in estimating motion represent the limiting factor in practical Super-Resolution. In reality, performance of motion estimation techniques is highly dependent on the complexity of actual motion. For instance, estimating the completely arbitrary motion encountered in real-world image scenes is an extremely difficult task with almost no guarantees of estimator performance. In practice, incorrect estimates of motion have disastrous implications on overall Super-Resolution performance (Farsiu et al., 2004a). In another aspect of our work, we have studied

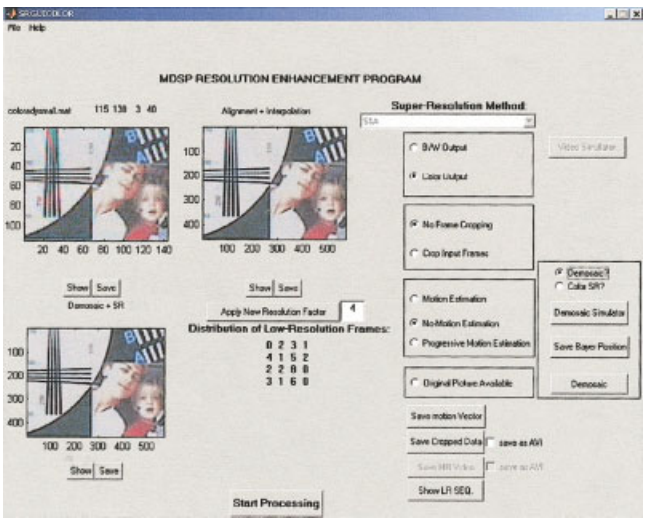

Figure 3. MDSP Resolution Enhancement Program screenshot.

fundamental performance limits for image registration (Robinson and Milanfar, 2004). We shall say more on this topic later.

B. Computational Aspects of Super-resolution.

A characteristic difficulty of the Super-Resolution problem is the dimensionality of the problem. This difficulty will be influenced both by the dimensionality of the unknown, $\underline{X}$, and the dimension of the measurement vector, $\underline{Y}$, and in both cases these numbers in the hundreds of thousands and beyond. The dimensionality of the problem demands high computational efficiency of any algorithm, if the algorithm is to be of practical utility. One such mechanism for simplifying the problem of Super-Resolution comes from a careful study of particular modeling scenarios. This theme plays a vital role in the work presented in this paper. This dimensionality problem is also the reason for the popularity of iterative solvers for the super-resolution problem in general.

For the case of quadratic penalty terms (LS) and Tikhonov regularization, the task of minimization is reduced to that of solving a very large linear system of equations. Many novel and powerful algebraic techniques have been proposed to minimize the complexity and maximize the performance for this class of routines. For example, Nguyen et al. (2001) propose efficient block circulant preconditioners to accelerate convergence of a conjugate gradient minimization algorithm. Although these methods are mathematically justifiable and numerically stable, they often belie a dependence on unrealistic assumptions such as perfect motion estimation. As we shall show, applying nonquadratic penalty terms offers much in the way of accuracy and at the same time realizing important speedups in minimization.

The speedup comes from the application of the matrix operators *F, H, D, A,* and their transposes directly as the corresponding image operations of shifting, blurring, and decimation (Zomet and Peleg, 2000; Farsiu et al., 2004a). For example, the operation of the decimation (down-sampling) matrix D and its transpose (up-sampling) matrix D^T is illustrated in Figure 2. Application of these operations in the image domain obviates the need to explicitly construct the matrices.

Throughout this article, we focus on the simplest of motion models, namely the translational model. The reasons for this are several. First, there exist efficient and accurate estimation algorithms with well studied performance limits (Robinson and Milanfar, 2004; Lin and Shun, 2004). Second, although simple, the model fairly well approximates the motion contained in image sequences where the scene is stationary and only the imaging system moves. Third, for sufficiently high frame rates most motion models can be (at least locally) approximated by the translational model. Finally, and most importantly, we believe that an in-depth study of this simple case allows much insight to be gained about the problems inherent to Super-Resolution.

One interesting implication of the translational motion model is the ability to greatly simplify the task of Super-Resolution. If the optical blur of the imaging system is translation invariant, then the order of the operations of the image shift and image blur are commutative (Elad and Hel-Or, 2001). By substituting $\underline{Z} = H\underline{X}$, the inverse problem may be separated into the much simpler sub-tasks of fusing the available images to estimate the blurry image $\underline{Z}$ followed by a deblurring/interpolation step estimating $\underline{X}$ from $\hat{\underline{Z}}$, the estimate of the blurred image. In Section IV, we make use of this property to explain and construct highly efficient Super-Resolution algorithms.

IV. RECENT WORK

In this section, we explore three specific Super-Resolution scenarios, each of which addresses a particular aspect of the general super-resolution challenge. Also, we highlight some of the important contributions we have made to each scenario. These scenarios have emerged from our effort to create a general Super-Resolution software tool capable of handling a wide variety of input image data. Figure 3 shows a sample screenshot of our Super-Resolution tool.[3] It is our hope that this work provides the foundation for future work addressing the more complete Super-Resolution problem.

A. Robust Super-resolution.

As indicated in Section III, often the parameters of the imaging system (such as motion and PSF) must be either assumed or estimated from the data to construct a forward model. When the terms in the model are assumed or estimated incorrectly, the data no longer match the model, leading to data outliers. Outliers, which are defined as data points with different distributional characteristics than the assumed model, will produce erroneous estimates when a nonrobust algorithm is applied. We have previously addressed (Farsiu et al., 2003a, 2004a) the problem of estimating a single high-resolution monochrome image $\underline{X}$ from a collection of low resolution images $\underline{Y}(t)$.

Drawing on the theory of robust statistics (Huber, 1981), we have developed a novel framework combining a robust data fidelity term and robust regularization term to build an efficient Super-Resolution framework exhibiting improved performance for real-world image sequences. It has been shown (Huber, 1981) that the LS type estimator of (2) is highly susceptible to the presence of outliers in the data, producing quite poor results. The lack of robustness is attributed to the use of the L_2 norm to measure data fidelity, which is only optimal for the case of Gaussian noise. A statistical study of the noise properties found in many real image sequences, however, suggests a heavy-tailed noise distribution such as Laplacian (Farsiu et al., 2003b). Instead of LS, we propose an alternate data fidelity term based on the L_1 norm, which has been shown to be very robust to data outliers. Also, we propose a novel regularization term called Bilateral-TV, which provides robust performance while preserving the edge content common to real image sequences. The proposed method is a generalization of the Total Variation principle of Rudin et al. (1992), which has been proposed as an edge-preserving regularization term.

Combining these two terms, we formulate our robust estimation framework as the following cost function[4]

$$J(\underline{X}) = \left[\sum_{t} \left\| D(t)H(t)F(t)\underline{X} - \underline{Y}(t) \right\|_1 + \lambda \underbrace{\sum_{l=-P}^{P} \sum_{m=0}^{P}}_{l+m \geq 0} \alpha^{|m|+|l|} \left\| \underline{X} - S_x^l S_y^m \underline{X} \right\|_1 \right], \quad (5)$$

[3] All the multiframe Super-Resolution methods (robust, color, demosaic, dynamic) discussed in this section plus many other Super-Resolution and motion estimation methods have been included in our software package. More information on this software tool is available at http://www.ee.ucsc.edu/~milanfar

[4] In this section we only consider the resolution enhancement problem for monochromatic images. Later, in Section IV.B, we extend this method for the case of color Super-Resolution.

where the first expression is relating the measurements to the desired image $\underline{X}$ through the model we described. S_x^l and S_y^m are the operators corresponding to shifting the image represented by $\underline{X}$ by l pixels in the horizontal direction and m pixels in the vertical direction, respectively. These act as derivatives across multiple scales. The scalar weight α, $0 < \alpha < 1$, is applied to give a spatially decaying effect to the summation of the regularization term. The shifting and differencing operations are very cheap to implement.

As mentioned in Section II, for the special case of translational motion and common space invariant blurring operation, where the blur and motion operators commute, we suggest a very efficient two-stage method for minimizing (5). The optimality of this method was extensively discussed in (Farsiu et al., 2004a). The first stage estimates the blurry high-resolution image $\underline{Z}$ from the collection of low resolution images as

$$\hat{\underline{Z}} = \underset{\underline{Z}}{\operatorname{argmin}} \left[\sum_t \left\| DF(t)\underline{Z} - \underline{Y}(t) \right\|_1 \right]. \tag{6}$$

We showed (Farsiu et al, 2004a) that for a given high-resolution pixel this cost function is minimized by performing a pixel-wise median of all the measurements after proper zero filling and motion compensation. We call this operation Median Shift-And-Add, which bears some similarity to the median-based algorithm proposed by Zomet et al. (2001).

The second stage of deblurring/interpolating the image $\hat{\underline{Z}}$ is performed using an iterative minimization method. This step is both a deblurring and interpolation step because it is possible to have no measurements associated with some pixels in the image $\hat{\underline{Z}}$ defined on the high-resolution grid. The following expression formulates our minimization criterion for obtaining $\hat{\underline{X}}$ from $\hat{\underline{Z}}$:

$$\hat{\underline{X}} = \underset{\underline{X}}{\operatorname{argmin}} \left[\left\| B(H\underline{X} - \hat{\underline{Z}}) \right\|_1 + \lambda \sum_{\underbrace{l=-P}_{l+m\geq 0}}^{P} \sum_{m=0}^{P} \alpha^{|m|+|l|} \left\| \underline{X} - S_x^l S_y^m \underline{X} \right\|_1 \right]. \tag{7}$$

Again, we see that the first term encourages a robust fidelity to the fused image $\hat{\underline{Z}}$ and the second term represents the robust regularization term. Here, the matrix 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 $\hat{\underline{Z}}$. This weighting ensures that pixels of $\hat{\underline{Z}}$ that have more measurements are weighted higher than those that have little or no measurements.

As an example, Figure 4(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 global translational motion between each frame. Second, the alpaca statue was repositioned several times throughout the input frames [notice this relative motion in Figs. 4(a) and 4(b)]. The nonrobust L_2 norm reconstruction with Tikhonov regularization results in Figure 4(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 Figures 4(e) and 4(f). Here, we also see that the performance of the faster method shown in Figure 4(f) is almost as good as the optimal method shown in Figure 4(e).

B. Robust Multiframe Demosaicing and Color Super-Resolution. There is very little work addressing the problem of color Super-Resolution, and the most common solution involves applying monochrome Super-Resolution algorithms to each of the color channels independently (Tom and Katsaggelos, 2001). Another approach is transferring the problem to a different color space where chrominance layers are separated from luminance, and where Super-Resolution is applied to the luminance channel only (Irani and Peleg, 1991). In this section, we review the work of Farsiu et al. (2004), which details the problems inherent to color Super-Resolution and proposes a novel algorithm for producing a high-quality color image from a collection of low-resolution 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 5 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 (see, e.g., Alleysson et al., 2002; Hel-Or and Keren, 2002; Keren and Osadchy, 1999; Kimmel, 1999; Laroche and Prescott, 1994), yet almost none of them [but Zomet and Peleg's (2002) method] to date are directly applicable to the problem of multiframe 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 multiframe situation. To better understand the multiframe demosaicing problem, we offer an example for the simple case of translational motion. Figure 5 illustrates the pattern of sensor measurements in the high-resolution image grid. In such situations, the sampling pattern is quite arbitrary depending on the relative motion of the low-resolution images. This necessitates a different demosaicing algorithm than those designed for the original Bayer pattern.

The challenge of multiframe color Super-Resolution is much more difficult than that of monochrome imaging and should not be solved by applying monochrome methods for several reasons. First, the additional down-sampling (matrix A) 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 that 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 Super-Resolution would only amplify such artifacts and lead to suboptimal perfor-

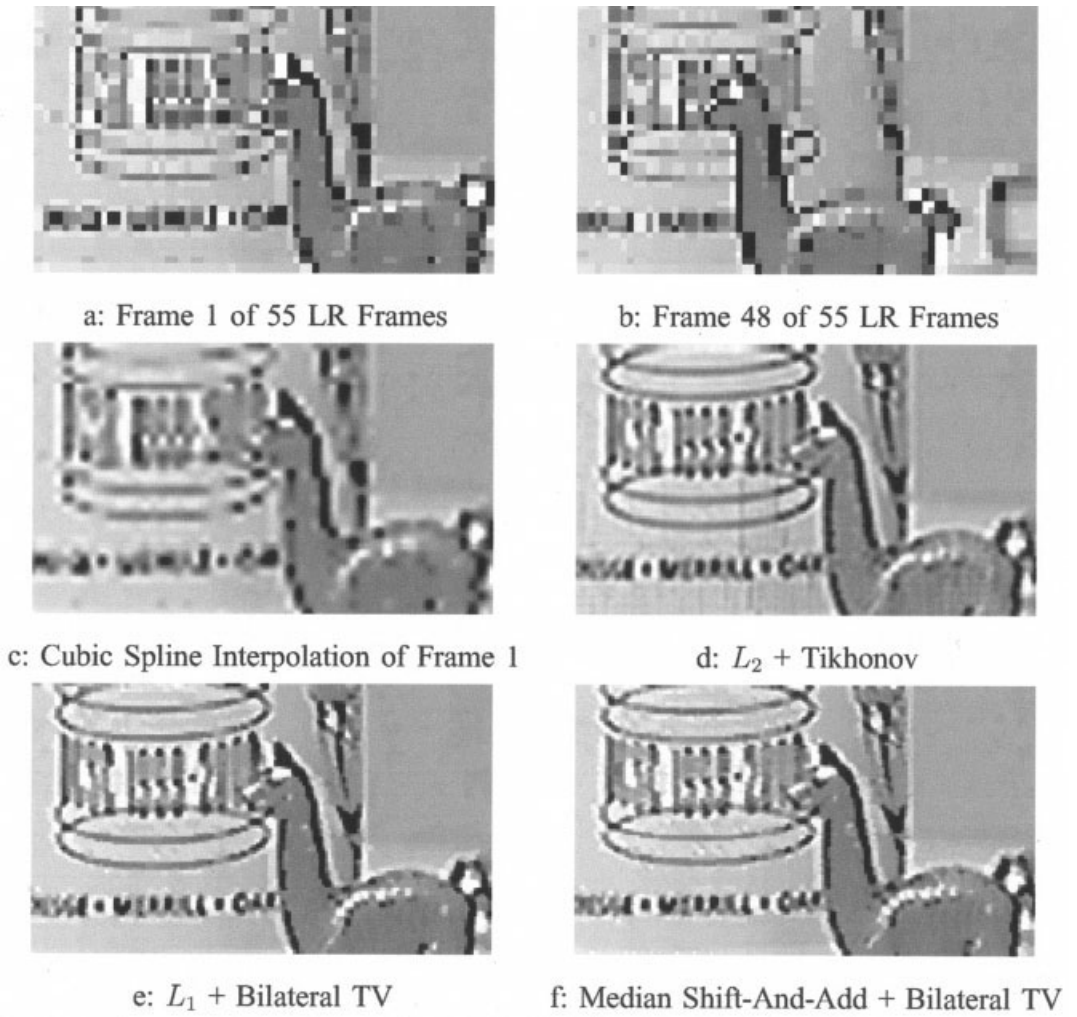

Figure 4. Results of different resolution enhancement methods applied to the alpaca sequence. Outlier effects are apparent in the nonrobust reconstruction method (d). However, the robust methods (e)–(f) were not affected by the outliers.

mance. Instead, all three channels must be estimated simultaneously to maximize the overall color Super-Resolution performance.

We proposed (Farsiu et al., 2004), a computationally efficient method to fuse and demosaic a set of low-resolution 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 Super-Resolution, additional regularization penalty functions are required. To facilitate the explanation, we represent the high-resolution 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:

$$J(\underline{X}) = \sum_{i=R,G,B} \sum_{t=1}^{N} \|(D(t)A_iH(t)F(t)\underline{X}_i - \underline{Y}_i(t))\|_1,$$

where A_i and $\underline{Y}_i(t)$ are the red, green, or blue components of the color filter and the low-resolution frame, respectively. As in the previous section, the fast two-stage method for the case of constant, space-invariant blur and global translation is also applicable to the multiframe demosaicing method, leading to an initial Median Shift-And-Add operation on Bayer-filtered low-resolution data followed by a deblurring step. Thus, the first stage of the algorithm is the Median Shift-And-Add operation of producing a blurry high-resolution image $\hat{\underline{Z}}_{R,G,B}$ (e.g., the left side of the accolade in Fig. 5). In this case, however, the median operation is applied in a pixel-wise fashion to each of the color channels independently (for more details, see Farsiu et al., 2004).

2) Luminance Regularization: Here, we use a penalty term regularizing the luminance component of the high-resolution 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 (Hel-Or and Keren, 2002). Therefore, we apply the Bilateral-TV regularization to the luminance component to offer robust edge preservation. The luminance image can be calculated as the weighted sum $\underline{X}_L$

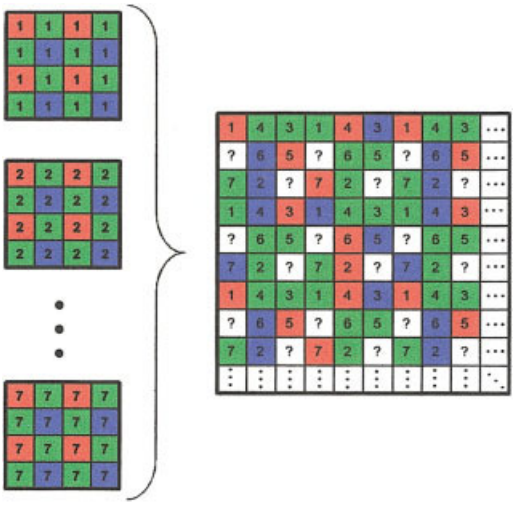

Figure 5. Fusion of 7 Bayer pattern low-resolution images with relative translational motion (the figures in the left side of the accolade) results in a high-resolution image ($\hat{Z}$) that does not follow a Bayer pattern (the figure in the right side of the accolade). The symbol "?" represents the high-resolution pixel values that were undetermined after the Shift-And-Add step (as a result of insufficient low-resolution frames).

$= 0.299\underline{X}_R + 0.597\underline{X}_G + 0.114\underline{X}_B$ as explained by Pratt (2001). The luminance regularization term is similar to (5) in Section IIIA:

$$J_1(\underline{X}) = \underbrace{\sum_{l=-P}^{P}\sum_{m=0}^{P}}_{l+m\geq 0}\alpha^{|m|+|l|}\left\|\underline{X}_L - S_x^l S_y^m \underline{X}_L\right\|_1. \tag{8}$$

3) Chrominance Regularization: This penalty term ensures the smoothness in the chrominance components of the high-resolution image. This removes many of the color artifacts objectionable to the human eye. Again, the two chrominance channels $\underline{X}_{C1}$ and $\underline{X}_{C2}$ can be calculated as the weighted combination of the RGB images using the weights $(-0.169, -0.331, 0.5)$ for $C1$ and $(0.5, -0.419, -0.081)$ for $C2$ (Pratt, 2001). 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:

$$J_2(\underline{X}) = \left\|\Lambda\underline{X}_{C1}\right\|_2^2 + \left\|\Lambda\underline{X}_{C2}\right\|_2^2, \tag{9}$$

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

4) Orientation Regularization: This term penalizes the nonhomogeneity 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

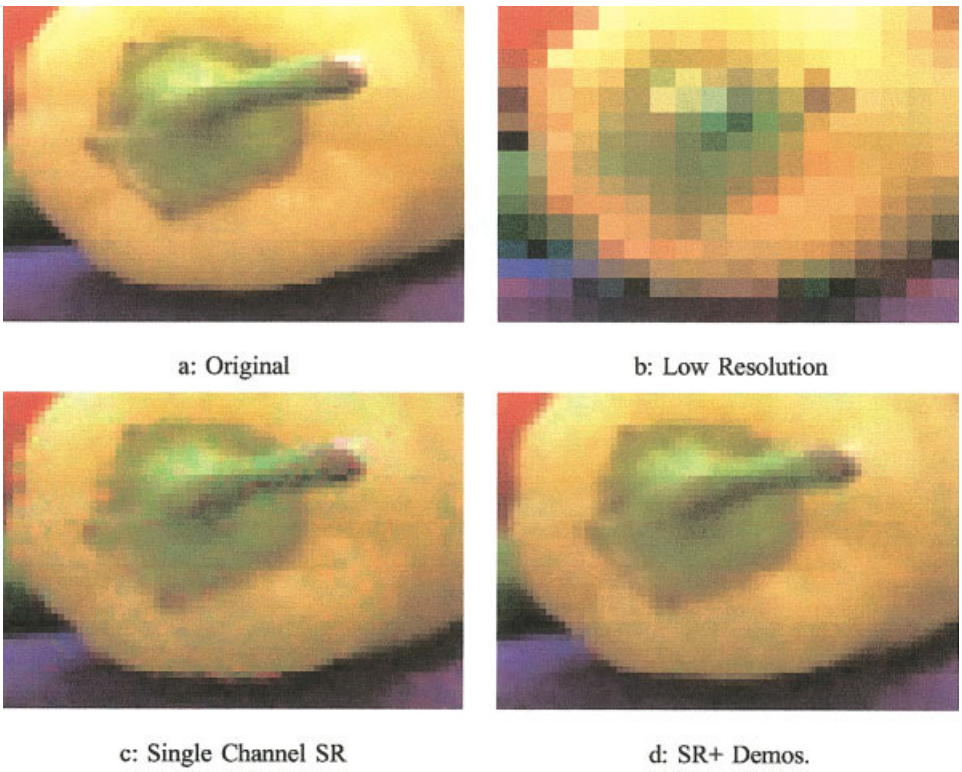

Figure 6. A high-resolution image (a) is passed through our model of camera to produce a set of low-resolution images. One of these low-resolution images, demosaiced by Laroche and Prescott's (1994) method, is shown in (b). The result of super-resolving each color band separately, considering only bilateral regularization, is shown in (c). And, finally, (d) is the result of applying the proposed method to this data set (factor of 4 resolution enhancement).

all color channels have the same edge orientation. That is, if a vertical (or horizontal) edge appears in the red band, a vertical (or horizontal) edge with similar orientation in the same location is likely to appear in the green and blue bands. Following Keren and Osadchy (1999), 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 by Keren and Osadchy (1999), our orientation penalty term is a differentiable cost function:

$$J_3(\underline{X}) = \underbrace{\sum_{l=-1}^{1}\sum_{m=0}^{1}}_{l+m\geq 0}(\|\underline{X}_G \odot S_x^l S_y^m \underline{X}_B - \underline{X}_B \odot S_x^l S_y^m \underline{X}_G\|_2^2 + \|\underline{X}_B \odot S_x^l S_y^m \underline{X}_R - \underline{X}_R \odot S_x^l S_y^m \underline{X}_B\|_2^2 + \|\underline{X}_R \odot S_x^l S_y^m \underline{X}_G - \underline{X}_G \odot S_x^l S_y^m \underline{X}_R|_2^2) \quad (10)$$

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:

$$\hat{\underline{X}} = \underset{\underline{X}}{\operatorname{argmin}}[J(\underline{X}) + \lambda_1 J_1(\underline{X}) + \lambda_2 J_2(\underline{X}) + \lambda_3 J_3(\underline{X})]. \quad (11)$$

We previously proposed (Farsiu et al., 2004) a method for applying a steepest descent algorithm to minimize this cost function. Interestingly, this cost function can also be applied to color images where an unknown demosaicing algorithm has already been applied prior to the Super-Resolution process.

Figure 6 illustrates the performance of the proposed method with respect to other methods. Figure 6(a) shows an image acquired with a high-resolution 3-CCD camera. A set of 10 low-resolution color filtered images was constructed following the forward imaging model to simulate the effect of imaging with a low-resolution single CCD Bayer-CFA camera. Figure 6(b) shows one of these images demosaiced by the method of Laroche and Prescott (1994), which is employed in Kodak DCS-200 digital cameras (Ramanath et al., 2002). In Figure 6(c) the method of Farsiu et al. (2004a) is used to fuse these images and increase the resolution by a factor of 4 in each color band, independently. The color artifacts are still apparent in this result. The result of applying our method on this sequence is shown in Figure 6(d), where color artifacts are significantly reduced.

As mentioned earlier, that this method may also be applied to a set of color low-resolution frames previously demosaiced to enhance their spatial resolution while reducing color artifacts. Figure 7 offers an example of this application on a real data sequence courtesy of Adyoron Intelligent Systems Ltd., Tel Aviv, Israel. The available color images were previously demosaiced using an unknown algorithm. Clearly, the color artifacts are reduced using our method.

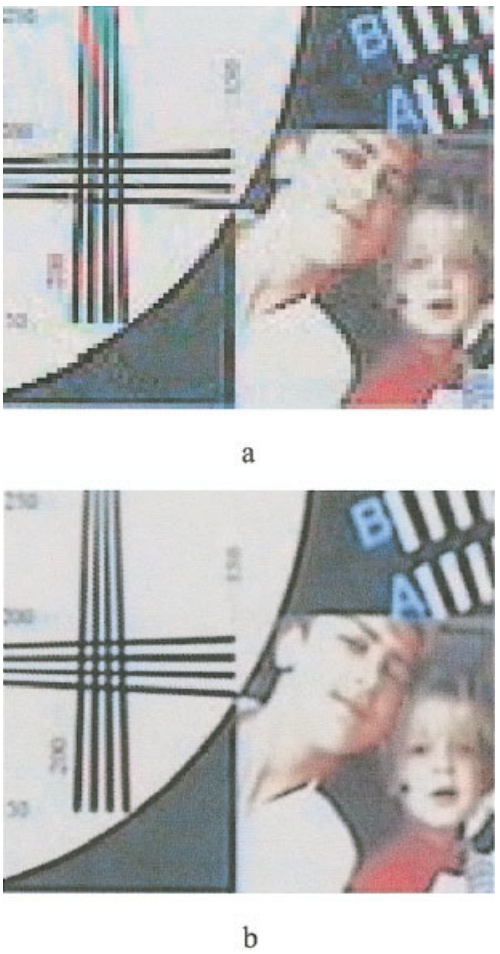

Figure 7. Multi-frame color Super-Resolution implemented on a real-world data sequence. (a) shows one of the input low-resolution images and (b) is the result of implementing the proposed method which has increased the spatial resolution by a factor of 4, removed the compression artifacts, and also reduced the color artifacts.

C. Dynamic Super-Resolution. In this section we address the computational challenges inherent to dynamic Super-Resolution. By dynamic Super-Resolution, we refer to the situation in which a *sequence* of high-resolution images are estimated from a sequence of low-resolution frames. Although it may appear that this problem is a simple extension of the static Super-Resolution situation, the memory and computational requirements for the dynamic case are so taxing as to preclude its application without highly efficient algorithms. We review the method introduced previously (Farsiu et al., 2004b), which offers an extremely efficient recursive algorithm for dynamic Super-Resolution. Although such a recursive solution for Super-Resolution has been addressed before (Elad and Feuer, 1999), we now show the speedups applicable for the case of translational motion and common space-invariant blur. This simplified model empowers us to use the two step algorithm that was described in Section IV.A for solving the dynamic case.

According to (1), we set up the forward model of the dynamic Super-Resolution problem as

$$Y(t) = DH(t)F(t)\underline{X}(t) + \underline{V}(t). \quad (12)$$

An efficient and intuitive approach of acquiring the high-resolution image is using weighted least square optimization (Elad and Feuer, 1999):

$$\hat{\underline{X}}(t) = \underset{\underline{X}}{\operatorname{argmin}}\left[\sum_{\tau=0}^{N-1}\gamma^{\tau}\|DHF^{\tau}(t-\tau)\underline{X}(t) - \underline{Y}(t-\tau)\|_2^2\right], \quad (13)$$

where γ is a parameter between 0 and 1. The weighting γ^{τ} places more emphasis on recent image data than on previous data. Note that

in order to consider the varying reliability of measurements gathered at each location, the weighting can also be applied on a pixel-by-pixel basis [(Farsiu et al., 2004b)]. We use the L_2 norm to follow Elad and Feuer's (1999) model (a robust data fusion term using L_1 norm minimization is part of our ongoing work).

As before, we first consider the estimation of the unknown blurry high-resolution image $\underline{Z}(t)$ before considering the task of deblurring. For this formulation, we have shown (Farsiu et al., 2004) that the update of the blurry high-resolution estimate is given by the recursive equation (14) below. Note that only those pixels in the high-resolution image that have new measurements from $\underline{Y}(t)$ are updated, and all other pixels are left unaltered. The pixels that satisfy this criterion (indexed by m) are updated according to

$$[\underline{Z}(t)]_m = \frac{1}{1+\beta_m(t)}[D^T\underline{Y}(t)]_m + \frac{\beta_m(t)}{1+\beta_m(t)}[F(t)\underline{Z}(t-1)]_m. \quad (14)$$

The adaptive weighting is given by the recursive equation

$$\beta_m(t) = [1+\beta_m(t-t')]\gamma^{t-t'}, \quad (15)$$

where t' represents the most recent time from time t in which a low-resolution pixel measurement was used to update pixel m. This type of weighting encourages a larger *forgetting* factor when the high-resolution pixels have not been updated recently.

Such a recursive solution shows that there is no need to keep any previous low-resolution frames (except the most recent one) in memory. Only the high-resolution image estimate $\hat{\underline{Z}}(t)$ at any given time and a same size weighting image containing the updated β values of corresponding pixels, need to be stored in memory, leading to a very memory-efficient algorithm. Furthermore, the update operation is simply shifting the previous estimate $\hat{\underline{Z}}(t-1)$ and updating the proper pixels using (14). Note that a Kalman filtering approach provides another recursive solution that offers a more mathematically justifiable estimate of the fused image $\hat{\underline{Z}}(t)$. This additional approach is studied in Farsiu et al. (2004b).

At this point, we have an efficient recursive estimation algorithm producing estimates of the blurry high-resolution images sequence $\hat{\underline{Z}}(t)$. From these frames, the sequence $\hat{\underline{X}}(t)$ must be estimated. Note that the first few frames will not have estimates for every pixel in $\hat{\underline{Z}}(t)$, necessitating a further joint interpolation and deblurring step. To perform robust deblurring and interpolation, we utilize a similar cost function as (7) for every time t:

$$\hat{\underline{X}}(t) = \underset{\underline{X}(t)}{\operatorname{argmin}}\left[\left\|B(H\underline{X}(t) - \hat{\underline{Z}}(t))\right\|_2^2 + \lambda \underbrace{\sum_{l=-P}^{P}\sum_{m=0}^{P}}_{l+m\geq 0}\alpha^{|m|+|l|}\left\|\underline{X} - S_x^l S_y^m \underline{X}\right\|_1\right]. \quad (16)$$

Here, the matrix B is a diagonal matrix whose values are chosen relative to both the number of measurements that contributed to make each element of $\hat{\underline{Z}}(t)$ and their time lag with respect to the current estimate. This is the primary distinction between (16) and (7).

To improve the speed of the entire algorithm, we propose using the shifted version of the previous high-resolution estimate $F(t)\hat{\underline{X}}(t-1)$ as the initial guess for $\hat{\underline{X}}(t)$. For most applications, this allows the iterative deblurring algorithm to converge in only a few steps.

Figure 8 shows an example of the dynamic Super-Resolution algorithm for a couple of frames of a 300-frame video sequence. The deblurred images (c) and (f) show the benefits achieved by only a few iterations of deblurring with the proper initial guess.

V. SUMMARY AND FURTHER CHALLENGES

In Section IV we presented only a few methods and insights for specific scenarios of Super-Resolution. Many questions still persist in developing a generic Super-Resolution algorithm capable of producing high-quality results on general image sequences. In this section, we outline a few areas of research in Super-Resolution that remain open. The types of questions to be addressed fall into mainly two categories. The first concerns analysis of the performance limits associated with Super-Resolution. The second is that of Super-Resolution system level design and understanding.

A thorough study of Super-Resolution performance limits will have a great effect on the practical and theoretical activities of the image reconstruction community. In deriving such performance limits, one gains insight into the difficulties inherent to super-resolution. One example of recent work addressing the limitations of optical systems is given by Sharam and Milanfar (2004), where the objective is to study how far beyond the classical Rayleigh resolution limit one can reach at a given signal to noise ratio. Another recent study (Baker and Kanade, 2002), shows that, for a large enough resolution enhancement factor, any smoothness prior will result in reconstructions with very little high-frequency content. Lin and Shum (2004), for the case of translational motion, studied limits based on a numerical perturbation model of reconstruction-based algorithms. However, the question of an optimal resolution factor (r) for an arbitrary set of images is still wide open. Also, the role of regularization has never been studied as part of the analysis is proposed. Given that it is the regularization that enables the reconstruction in practice, any future contribution of worth on this matter must take it into effect.

Systematic study of the performance limits of Super-Resolution would reveal the true information bottlenecks, hopefully motivating focused research to address these issues. Furthermore, analysis of this sort could possibly provide understanding of the fundamental limits to the Super-Resolution imaging, thereby helping practitioners to find the correct balance between expensive optical imaging system and image reconstruction algorithms. Such analysis may also be phrased as general guidelines when developing practical super-resolution systems.

In building a practical Super-Resolution system, many important challenges lay ahead. For instance, in many of the optimization routines used in this and other articles, the task of tuning the necessary parameters is often left up to the user. Parameters such as regularization weighting λ can play an important role in the performance of the Super-Resolution algorithms. Although the cross-validation method can be used to determine the parameter values for the nonrobust Super-Resolution method (Nguyen et al., 2001a), a computationally efficient way of implementing such method for the robust Super-Resolution case has not yet been addressed.

Although some work has addressed the joint task of motion estimation and Super-Resolution (Hardie et al., 1997; Schultz et al., 1998; Tom and Katsaggelos, 2001), the problems related to this still remain largely open. Another open challenge is that of blind super-resolution wherein the unknown parameters of the imaging system's PSF must be estimated from the measured data. Many single-frame

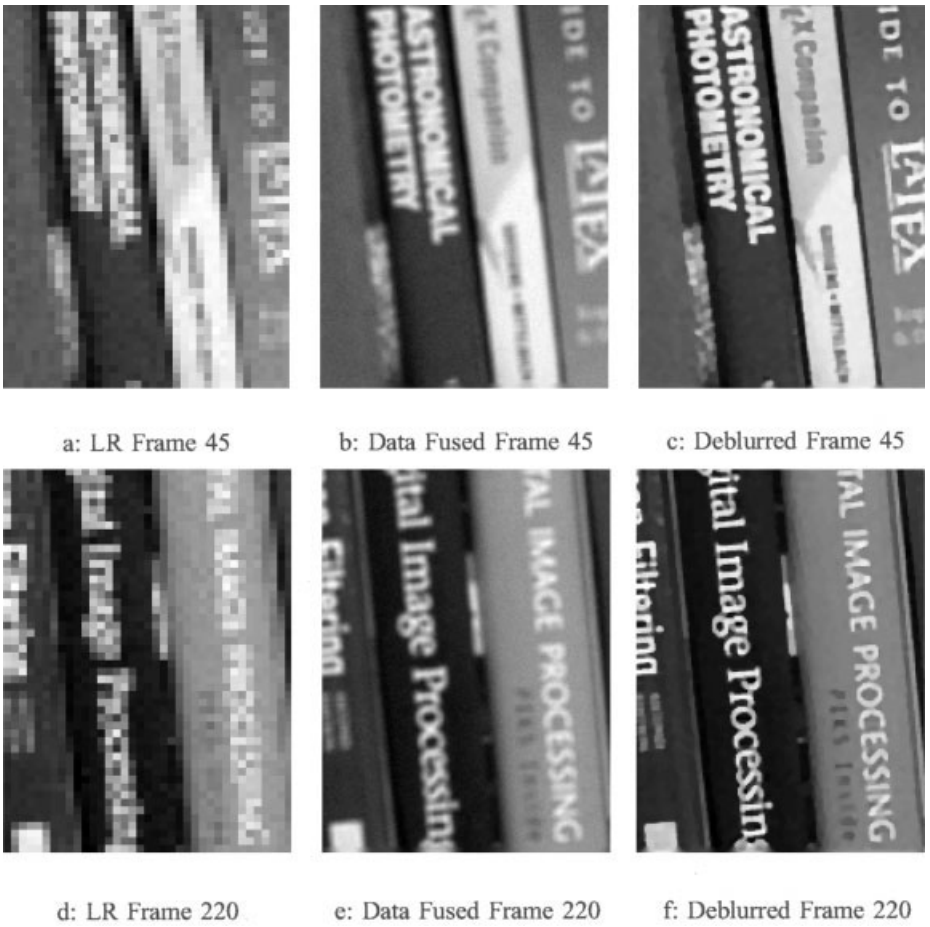

Figure 8. A set of low-resolution frames are used to produce a set of high-resolution frames. Two low-resolution frames in this sequence are shown in (a) and (d). The result of image fusion for these low-resolution frames are shown in (b) and (e). The result of deblurring these images after two iterations of steepest descent is shown in (c) and (f).

blind deconvolution algorithms have been suggested in the last 30 years (Kondur and Hatzinakos, 1996), and recently (Nguyen et al., 2001a) incorporated a single parameter blur identification algorithm in their Super-Resolution method, but there remains a need for more research to provide a Super-Resolution method along with a more general blur estimation algorithm from aliased images. Also, recently the challenge of simultaneous resolution enhancement in time as well as space has received growing attention (Robertson and Stevenson 2001; Shechtman et al., 2002).

Finally, it is the case that the low-resolution images are often, if not always, available in compressed format. Although a few articles have addressed resolution enhancement of DCT-based compressed video sequences (Segall et al., 2001; Altunbasak et al., 2002), the more recent advent and utilization of wavelet-based compression methods requires novel adaptive Super-Resolution methods. Adding features such as robustness, memory and computation efficiency, color consideration, and automatic selection of parameters in super-resolution methods will be the ultimate goal for the Super-Resolution researchers and practitioners in the future.

REFERENCES

D. Alleysson, S. Süsstrunk, J. Hrault, 2002. Color demosaicing by estimating luminance and opponent chromatic signals in the fourier domain. In Proc. IS&T/SID 10th Color Imaging Conf Nov, 2002, pp. 331–336.

Y. Altunbasak, A. Patti, R. Mersereau. 2002. Super-resolution still and video reconstruction from mpeg-coded video. IEEE Trans Circuits Syst Video Technol. 12:217–226.

C.B. Atkins, C.A. Bouman, J.P. Allebach. 1999. Tree-based resolution synthesis. In IS&T Conf on Image Processing, Image Quality, Image Capture Systems, 1999, pp. 405–410.

S. Baker, T. Kanade. 2002. Limits on super-resolution and how to break them. IEEE Trans Pattern Anal Machine Intelli 24:1167–1183.

S. Borman, RL. Stevenson. 1998. Super-resolution from image sequences — a review. In Proc 1998 Midwest Symp Circuits and Systems, Vol. 5, Apr

M. Elad. 2002. On the bilateral filter and ways to improve it. IEEE Trans Image Process 11:1141–1151.

M. Elad, A. Feuer. 1997. Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images IEEE Trans Image Process 6:1646–1658.

M. Elad, A. Feuer. 1999. Super-resolution reconstruction of image sequences. IEEE Trans Pattern Anal Machine Intell 21:817–834.

M. Elad, Y. Hel-Or. 2001. A fast super-resolution reconstruction algorithm for pure translational motion and common space invariant blur. IEEE Trans Image Process 10:1187–1193.

S. Farsiu, M. Elad, P. Milanfar. 2004. Multi-frame demosaicing and super-resolution from under-sampled color images. Proc 2004 IS&T/SPIE 16th Annual Sympo on Electronic Imaging, Jan., 2004, pp. 222–233.

S. Farsiu, D. Robinson, M. Elad, P. Milanfar. 2003a. Fast and robust Super-Resolution. Proc 2003 IEEE Int Conf on Image Process 2003, pp. 291–294.

S. Farsiu, D. Robinson, M. Elad, P. Milanfar. 2003b. Robust shift and add approach to super-resolution. Proc. 2003 SPIE Conf on Applic Digital Signal and Image Process, 2003, pp. 121–130.

S. Farsiu, D. Robinson, M. Elad, P. Milanfar. 2004a. Fast and robust multi-frame super-resolution, to appear in IEEE Trans Image Processing, October, 2004.

S. Farsiu, D. Robinson, M. Elad, P. Milanfar. 2004b. Dynamic demosaicing and color Super-Resolution video sequences, to appear in the Proc SPIE Conf on Image Reconstruction from Incomplete Data, 2004.

G. Golub, C.V. Loan. 1996. Matrix computations, 3rd ed. The Johns Hopkins University Press: London, 1996.

Haber E, Tenorio L. 2003. Learning regularization functionals-a supervised training approach. Inverse Problems 19:611–626.

P.C. Hansen, D.P. O'Leary. 1993. The use of the *L*-curve in the regularization of ill-posed problems. SIAM J Sci Comput 14:1487–1503.

R. Hardie, K.J. Barnard, EE. Armstrong. 1997. Joint map registration and high-resolution image estimation using a sequence of undersampled images IEEE Trans Image Process 6:1621–1633.

Y. Hel-Or, Keren D. 2002. Demosaicing of color images using steerable wavelets. Tech. Report HPL-2002-206R1 20020830, HP Labs Israel, 2002.

T.S. Huang, R.Y. Tsai, 1984. Multi-frame image restoration and registration. Adv Comput Vision Image Process 1: 317–339.

P.J. Huber. 1981. Robust Statistics. Wiley: New York, 1981.

M. Irani, S. Peleg. 1991. Improving resolution by image registration. CVGIP: Graph. Models Image Process 53:231–239.

M.G. Kang, S. Chaudhuri. 2003. Super-resolution image reconstruction. IEEE Signal Process Mag 20:21–36.

D. Keren, M. Osadchy. 1999. Restoring subsampled color images. Machine Vision and Appli 11:197–202.

R. Kimmel. 1999. Demosaicing: Image reconstruction from color ccd samples. IEEE Trans Image Process 8:1221–1228.

D. Kondur, D. Hatzinakos. 1996. Blind image deconvolution. IEEE Signal Process Mag 13:43–64.

C. Laroche, M. Prescott. 1994. Apparatus and method for adaptive for adaptively interpolating a full color image utilizing chrominance gradients. United States Patent (1994), 5,373,322.

S. Lertrattanapanich, N.K. Bose. 2002. High resolution image formation from low resolution frames using Delaunay triangulation. IEEE Trans Image Process 11:1427–1441.

Z.C. Lin, H.Y. Shum. 2004. Fundamental limits of reconstruction-based superresolution algorithms under local translation. IEEE Trans Pattern Anal Machine Intell 26:83–97.

M.A. Lukas. 1993. Asymptotic optimality of generalized cross-validation for choosing the regularization parameter. Numerische Mathematik 66:41–66.

N. Nguyen, P. Milanfar, G. Golub. 2001a. Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement. IEEE Trans Image Process 10:1299–1308.

N. Nguyen, P. Milanfar, G.H. Golub. 2001b. A computationally efficient image superresolution algorithm. IEEE Trans Image Process 10:573–583.

K.A. Parulski, L.J. D'Luna, B.L. Benamati, P.R. Shelley. 1992. High performance digital color video camera. J Electron Imaging 1:35–45.

A. Patti, M. Sezan, A.M. Tekalp. 1997. Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time. IEEE Trans Image Process 6:1326–1333.

W.K. Pratt, 2001. Digital image processing, 3rd ed. Wiley, New York,

R. Ramanath, W. Snyder, G. Bilbro, W. Sander. 2002. Demosaicking methods for the Bayer color arrays. J Electron Imaging 11:306–315.

M.A. Robertson, R.I. Stevenson. 2001. Temporal resolution enhancement in compressed video sequences. EURASIP J Appl Signal Process 230–238.

D. Robinson, P. Milanfar, 2003. Fundamental performance limits in image registration, to appear in IEEE Trans Image Processing, Sept., 2004.

L. Rudin, S. Osher, E. Fatemi. 1992. Nonlinear total variation based noise removal algorithms. Physica D 60:259–268.

R.R. Schultz, L. Meng, R. Stevenson. 1998. Subpixl motion estimation for Super-Resolution image sequence enhancement. J Visual Communi Image Repres 9:38–50.

R.R. Schultz, R.L. Stevenson. 1996. Extraction of high-resolution frames from video sequences. IEEE Trans Image Process 5:996–1011.

C.A. Segall, R. Molina, A. Katsaggelos, J. Mateos. 2001. Bayesian high-resolution reconstruction of low-resolution compressed video. In: Proc IEEE Int Conf. Image Process. Oct 2001, Vol 2, pp. 25–28.

M. Shahram, P. Milanfar. 2003. Imaging below the diffraction limit: A statistical analysis. IEEE Trans Image Processing 5:677–689.

E. Shechtman, Y. Caspi, M. Irani. 2002. Increasing space-time resolution in video. In: Proc. European Conf on Computer Vision (ECCV), May 2002, pp. 331–336.

N. Sochen, R. Kimmel, R. Malladi. 1998. A general framework for low level vision. IEEE Trans Image Process 7:310–318.

B.C. Tom, A. Katsaggelos. 2001. Resolution enhancement of monochrome and color video using motion compensation. IEEE Trans Image Process 10:278–287.

S.C. Zhu, D. Mumford. 1997. Prior learning and Gibbs reaction-diffusion IEEE Trans Pattern Analy Machine Intelli 19:1236–1250.

A. Zomet, S. Peleg. 2000. Efficient super-resolution and applications to mosaics. In Proc Int Conf on Pattern Recognition (ICPR), Sep. 2000, pp. 579–583.

A. Zomet, S. Peleg. 2002. Multi-sensor super resolution. In: Proc IEEE Workshop on Applications of Computer Vision, Dec 2002, pp. 27–31.

A. Zomet, A. Rav-Acha, S. Peleg. 2001. Robust super resolution, In: Proc. Int Conf Computer Vision and Pattern Recognition (CVPR), Dec 2001, Vol. 1, pp. 645–650.