# Fast and Robust Multi-Frame Super-Resolution

**Sina Farsiu**[*], **Dirk Robinson**[†], **Michael Elad**[‡], **Peyman Milanfar**[§]

**EDICS: 2-MLTF (Multi-Frame Image Restoration)**

**Abstract**

In the last two decades, many papers have been published, proposing a variety of methods for multi-frame resolution enhancement. These methods are usually very sensitive to their assumed model of data and noise, which limits their utility. This paper reviews some of these methods and addresses their shortcomings. We propose an alternate approach using $L_1$ norm minimization and robust regularization based on a bilateral prior to deal with different data and noise models. This computationally inexpensive method is robust to errors in motion and blur estimation, and results in images with sharp edges. Simulation results confirm the effectiveness of our method and demonstrate its superiority to other super-resolution methods.

## I. Introduction

Theoretical and practical limitations usually constrain the achievable resolution of any imaging device. A dynamic scene with continuous intensity distribution $X(x,y)$ is seen to be warped at the camera lens because of the relative motion between the scene and camera. The images are blurred both by atmospheric turbulence and camera lens by continuous point spread functions $H_{atm}(x,y)$ and $H_{cam}(x,y)$. Then they

[*]Corresponding author: Electrical Engineering Department, University of California Santa Cruz, Santa Cruz CA. 95064 USA. Email: farsiu@ee.ucsc.edu, Phone:(831)-459-4141, Fax: (831)-459-4829

[†]Electrical Engineering Department, University of California Santa Cruz, Santa Cruz CA. 95064 USA. Email: dirkr@ee.ucsc.edu, Phone:(831)-459-4141, Fax: (831)-459-4829

[‡] Department of Computer Science (SCCM), Stanford University, Stanford CA. 94305-9025 USA. Email: elad@sccm.stanford.edu, Phone:(650)-723-7685, Fax: (650)-723-2411

[§]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

This work was supported in part by the National Science Foundation Grants CCR-9984246 and CCR-9971010, 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 .

will be discretized at the CCD resulting in a digitized noisy frame $Y[m,n]$. We represent this forward model by the following equation:

$$Y[m,n] = [H_{cam}(x,y) * *F(H_{atm}(x,y) * *X(x,y))] \downarrow + V[m,n] \tag{1}$$

in which $**$ is the two dimensional convolution operator, $F$ is the warping operator, $\downarrow$ is the discretizing operator, $V[m,n]$ is the system noise and $Y[m,n]$ is the resulting discrete noisy and blurred image. Figure 1 illustrates this equation.

Super-resolution is the process of combining a sequence of low-resolution (LR) noisy blurred images to produce a higher resolution image or sequence. The multi-frame super-resolution problem was first addressed in [1], where they proposed a frequency domain approach, extended by others such as [2]. Although the frequency domain methods are intuitively simple and computationally cheap, they are extremely sensitive to model errors [3], limiting their use. Also by definition, only pure translational motion can be treated with such tools and even small deviations from translational motion significantly degrade performance.

Another popular class of methods, solves the problem of resolution enhancement in the special domain. Non-iterative spatial domain data fusion approaches were proposed in [4], [5] and [6]. The iterative back-projection method was developed in papers such as [7] and [8]. In [9], the authors suggested a method based on the multichannel sampling theorem. In [10], a hybrid method, combining the simplicity of maximum likelihood (ML) with proper prior information was suggested.

The spatial domain methods discussed so far are generally computationally expensive. The authors in [11] introduced a block circulant preconditioner for solving the Tikhonov regularized super-resolution problem formulated in [10], and addressed the calculation of regularization factor for the under-determined case by generalized cross-validation. Later, a very fast super-resolution algorithm for pure translational motion and common space invariant blur was developed in [5]. Another fast spatial domain method was recently suggested in [12], where LR images are registered with respect to a reference frame defining a nonuniformly spaced high-resolution (HR) grid. Then, an interpolation method called Delaunay triangulation is used for creating a noisy and blurred high-resolution image, which is subsequently deblurred. All of the above methods assumed the additive Gaussian noise model. Furthermore, regularization was either not implemented or it was limited to Tikhonov regularization. Considering outliers, [13] describes a very successful robust super-resolution method, but lacks the proper mathematical justification ( limitations of this robust method and its relation to our proposed method are discussed in Appendix II). Finally, [14] and [15] have considered quantization noise resulting from video compression and proposed iterative

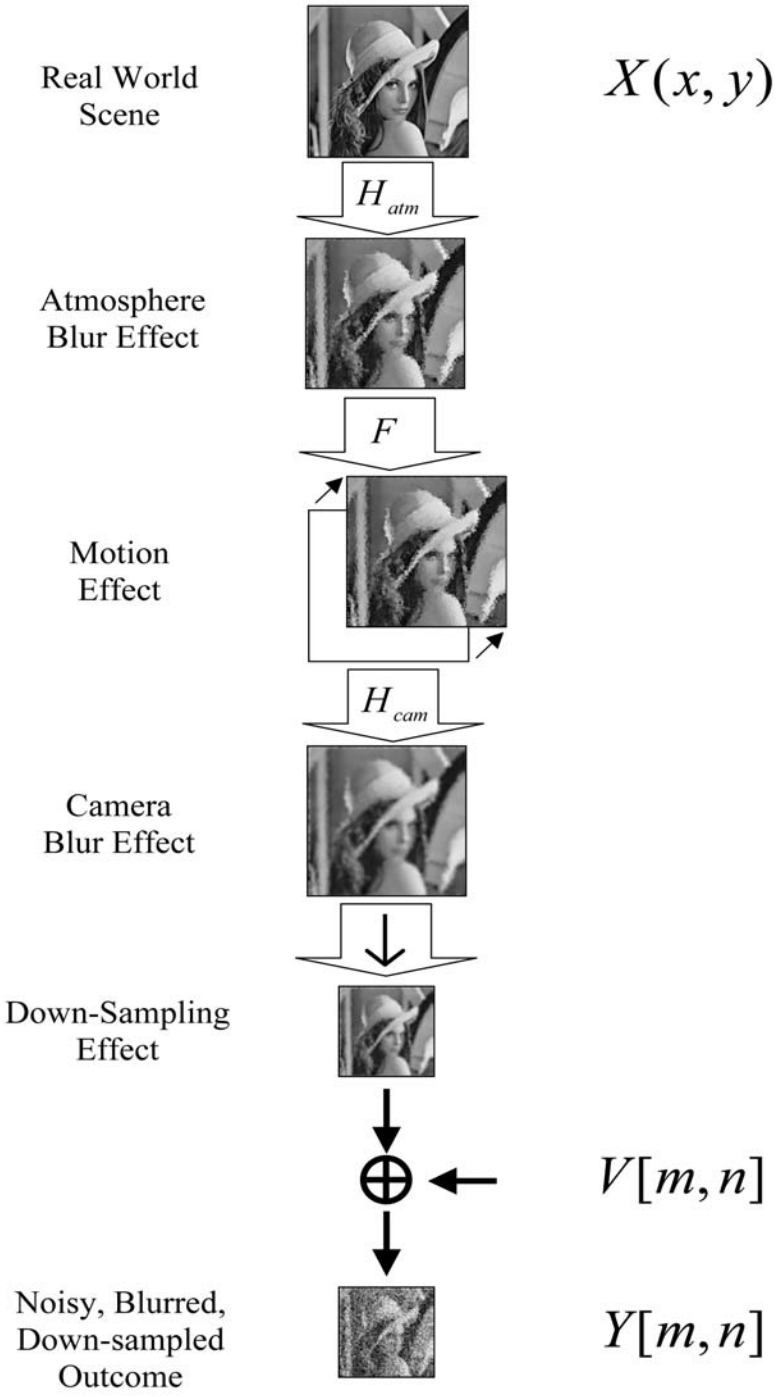



Fig. 1. Block diagram representation of (1), where $X(x,y)$ is the continuous intensity distribution of the scene, $V[m,n]$ is the additive noise, and $Y[m,n]$ is the resulting discrete low-quality image.

methods to reduce compression noise effects in the super-resolved outcome.

The two most common matrix notations used to formulate the general super-resolution model of (1) represent the problem in the pixel domain. The more popular notation used in [5], [11], and [13] considers only camera lens blur and is defined as:

$$\underline{Y}_k = D_k H_k^{cam} F_k \underline{X} + \underline{V}_k \qquad k = 1, \ldots, N \tag{2}$$

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

Considering only atmosphere and motion blur, [12] recently presented an alternate matrix formulation of (1) as

$$\underline{Y}_k = D_k F_k H_k^{atm} \underline{X} + \underline{V}_k \qquad k = 1, \ldots, N \tag{3}$$

In conventional imaging systems (such as video cameras), camera lens blur has more important effect than the atmospheric blur (which is very important for astronomical images). In this paper we use the model (2). Note that, under some assumptions which will be discussed in section II-B, blur and motion matrices commute and the general matrix super-resolution formulation from (1) can be rewritten as:

$$\underline{Y}_k = D_k H_k^{cam} F_k H_k^{atm} \underline{X} + \underline{V}_k = D_k H_k^{cam} H_k^{atm} F_k \underline{X} + \underline{V}_k \qquad k = 1, \ldots, N \tag{4}$$

Defining $H_k = H_k^{cam} H_k^{atm}$ merges both models into a form similar to (2).

In this paper, we propose a fast and robust super-resolution algorithm using the $L_1$ norm, both for the regularization and the data fusion terms. Whereas the former is responsible for edge preservation, the latter seeks robustness with respect to motion error, blur, outliers, and other kinds of errors not explicitly modelled in the fused images. We show that our method's performance is superior to what was proposed earlier in [5], [11], [13], etc. and has fast convergence. We also mathematically justify a non-iterative data fusion algorithm using a median operation and explain its superior performance.

This paper is organized as follows: Section II explains the main concepts of robust super-resolution; subsection II-B justifies using the $L_1$ norm to minimize the data error term; subsection II-C justifies using our proposed regularization term; subsection II-D combines the results of the two previous sections and explains our method. Section III proposes a faster implementation method (subsection III-A) and

provides some ideas for outlier detection and removal (subsection III-B). Simulations on both real and synthetic data sequences are presented in Section IV, and Section V concludes this paper.

## II. Robust Super-Resolution

### A. Robust Estimation

Estimation of an unknown HR image is not exclusively based on the LR measurements. It is also based on many assumptions such as noise or motion models. These models are not supposed to be exactly true, as they are merely mathematically convenient formulations of some general prior information.

From many available estimators, which estimate a HR image from a set of noisy LR images, one may choose an estimation method which promises the optimal estimation of the HR frame, based on certain assumption on data and noise models. When the fundamental assumptions of data and noise models do not faithfully describe the measured data, the estimator performance degrades. Furthermore, existence of outliers, which are defined as data points with different distributional characteristics than the assumed model, will produce erroneous estimates. A method which promises optimality for a limited class of data and noise models may not be the most effective approach. Often, suboptimal estimation methods which are not as sensitive to modelling and data errors may produce better and more stable results (robustness).

To study the effect of outliers the concept of a breakdown point has been used to measure the robustness of an algorithm. The breakdown point is the smallest percentage of outlier contamination that may force the value of the estimate outside some range [16]. For instance, the breakdown point of the simple mean estimator is zero, meaning that one single outlier is sufficient to move the estimate outside any predicted bound. A robust estimator, such as the median estimator, may achieve a breakdown equal to 0.5, which is the highest value for breakdown points. This suggests that median estimation may not be affected by data sets in which outlier contaminated measurements form less that $50\%$ of all data points.

A popular family of estimators are the Maximum Likelihood type estimators (M-estimators)[17]. We rewrite the definition of these estimators in the super-resolution context as the following minimization problem:

$$\underline{\widehat{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[ \sum_{k=1}^{N} \rho(\underline{Y}_k, D_k H_k F_k \underline{X}) \right] \tag{5}$$

or by an implicit equation

$$\sum_{k} \Psi(\underline{Y}_k, D_k H_k F_k \underline{X}) = 0 \tag{6}$$

where $\rho$ is measuring the "distance" between the model and measurements, and $\Psi(\underline{Y}_k, D_k H_k F_k \underline{X}) = \frac{\partial}{\partial \underline{X}} \rho(\underline{Y}_k, D_k H_k F_k \underline{X})$. The maximum likelihood estimate of $\underline{X}$ for an assumed underlying family of expo-

nential densities $f(\underline{Y}_k, D_k H_k F_k \underline{X})$ can be achieved when $\Psi(\underline{Y}_k, D_k H_k F_k \underline{X}) = -\log f(\underline{Y}_k, D_k H_k F_k \underline{X})$.

To find maximum likelihood (ML) estimate of the HR image, many papers such as [2], [5], [11] adopt a data model such as (2) and model $\underline{V}_k$(additive noise) as white Gaussian noise. With this noise model, least squares approach will result in the maximum likelihood estimate [18]. The least squares formulation is achieved when $\rho$ is the $L_2$ norm of residual:

$$\underline{\widehat{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[ \sum_{k=1}^{N} \| D_k H_k F_k \underline{X} - \underline{Y}_k \|_2^2 \right] \tag{7}$$

For the special case of super-resolution, based on [5], we will show in the next section, that least square estimation has the interpretation of being a non-robust mean estimation. As a result, least squares based estimation of a HR image, from a data set contaminated with non-Gaussian outliers, produces an image with visually apparent errors.

To appreciate this claim and study the visual effects of different sources of outliers in a video sequence, we set up the following experiments. In these experiments, four LR images were used to reconstruct a higher resolution image with two times more pixels in vertical and horizontal directions (a resolution enhancement factor of two using the least-squares approach (7)). Figure 2(a) shows the original HR image and Figure 2(b) shows one of these LR images which has been acquired by shifting Figure 2(a) in vertical and horizontal directions and subsampling it by factor of two(pixel replication is used to match its size with other pictures).

In the first experiment one of the four LR images contained affine motion with respect to the other LR images. If the model assumes translational motion, this results in a very common source of error when super-resolution is applied to real data sequences, as the respective motion of camera and the scene are seldom pure translational. Figure 2(c) shows this outlier image. Figure 2(d) shows the effect of this error in the motion model (shadows around the Lena's hat) when the non robust least-squares approach [5] is used for reconstruction.

To study the effect of non-Gaussian noise models, in the second experiment all four LR images were contaminated with salt and pepper noise. Figure 2(e) shows one of these LR images, and Figure 2(f) is the outcome of the least-squares approach for reconstruction.

As the outlier effects are visible in the output results of least square based super-resolution methods, it seems essential to find an alternative estimator. This new estimator should have the essential properties of robustness to outliers, and a fast implementation.

a:Original HR Frame b:LR Frame

c:LR Frame with Salt and Pepper Outlier d:Least Squares Result

e:LR Frame with Zoom f:Least Squares Result

Fig. 2. Simulation results of outlier effects on super-resolved images. The original HR image in (a) was warped with translational motion and down-sampled resulting in four images such as (b). (c) One of four LR images acquired by adding salt and pepper noise to set of images in (b). (d) Reconstruction of images in (c) with Least Squares approach. (e) is an image acquired with downsampling and zoom(affine motion). (f) Reconstruction of these four LR images with least squares approach.

*B. Robust Data Fusion*

In subsection II-A, we discussed the shortcomings of least squares based HR image reconstruction. In this subsection, we study the family of $L_p$, $1 \leq p \leq 2$ norm estimators. We choose the most robust estimator of this family and show how implementation of this estimator requires minimum memory usage and is very fast.

The following expression formulates the $L_p$ minimization criterion:

$$\underline{\widehat{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[ \sum_{k=1}^{N} \|D_k H_k F_k \underline{X} - \underline{Y}_k\|_p^p \right] \tag{8}$$

Note that if $p = 2$ then (8) will be equal to (7).

Considering translational motion and with reasonable assumptions such as common space-invariant PSF, and similar decimation factor for all LR frames (i.e. $\forall k \quad H_k = H \quad \& \quad D_k = D$ which is true when all images are acquired with a unique camera), we calculate the gradient of the $L_p$ cost. We will show that $L_p$ norm minimization is equivalent to pixelwise weighted averaging of the registered frames. We calculate these weights for the special case of $L_1$ norm minimization and show that $L_1$ norm converges to median estimation which has the highest breakpoint value. Since $H$ and $F_k$ are block circulant matrices, they commute ($F_k H = H F_k$ and $F_k^T H^T = H^T F_k^T$). Therefore, (8) may be rewritten as:

$$\underline{\widehat{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[ \sum_{k=1}^{N} \|D F_k H \underline{X} - \underline{Y}_k\|_p^p \right] \tag{9}$$

We define $\underline{Z} = H\underline{X}$. So $\underline{Z}$ is the blurred version of the ideal HR image $\underline{X}$. Thus, we break our minimization problem in two separate steps:

1) Finding a blurred HR image from the LR measurements (we call this result $\underline{\widehat{Z}}$).
2) Estimating the deblurred image $\underline{\widehat{X}}$ from $\underline{\widehat{Z}}$

Note that anything in the null space of $H$ will not converge by the proposed scheme. However, if we choose an initialization that has no gradient energy in the null space, this will not pose a problem (see [5] for more details). As it turns out, the null space of $H$ corresponds to very high frequencies, which are not part of our desired solution. To find $\underline{\widehat{Z}}$, we substitute $H\underline{X}$ with $\underline{Z}$:

$$\underline{\widehat{Z}} = \underset{\underline{Z}}{\text{ArgMin}} \left[ \sum_{k=1}^{N} \|D F_k \underline{Z} - \underline{Y}_k\|_p^p \right] \tag{10}$$

The gradient of the cost in (10) is:

$$\underline{G}_p = \frac{\partial}{\partial \underline{Z}} \left[ \sum_{k=1}^{N} \|D F_k \underline{Z} - \underline{Y}_k\|_p^p \right] = \sum_{k=1}^{N} F_k^T D^T sign(D F_k \underline{Z} - \underline{Y}_k) \odot |D F_k \underline{Z} - \underline{Y}_k|^{p-1} \tag{11}$$

where operator $\odot$ is the element-by-element product of two vectors.

The vector $\underline{\widehat{Z}}$ which minimizes the criterion (10) will be the solution to $\underline{G}_p = \underline{0}$. There is a simple interpretation for the solution: The vector $\underline{\widehat{Z}}$ is the weighted mean of all measurements at a given pixel, after proper zero filling and motion compensation.

To appreciate this fact, let us consider two boundary values of $p$. If $p = 2$, then

$$\underline{G}_2 = \sum_{k=1}^{N} F_k^T D^T (D F_k \underline{Z}_n - \underline{Y}_k) \tag{12}$$

which is proved in [5] to be the pixelwise average of measurements. If $p = 1$ then the gradient term will be:

$$\underline{G}_1 = \sum_{k=1}^{N} F_k^T D^T sign(D F_k \underline{\widehat{Z}} - \underline{Y}_k) = \underline{0} \tag{13}$$

We note that $F_k^T D^T$ copies the values from the LR grid to the HR grid after proper shifting and zero filling, and $DF_k$ copies a selected set of pixels in HR grid back on the LR grid (Figure 3 illustrates the effect of upsampling and downsampling matrices $D^T$, and $D$). Neither of these two operations changes

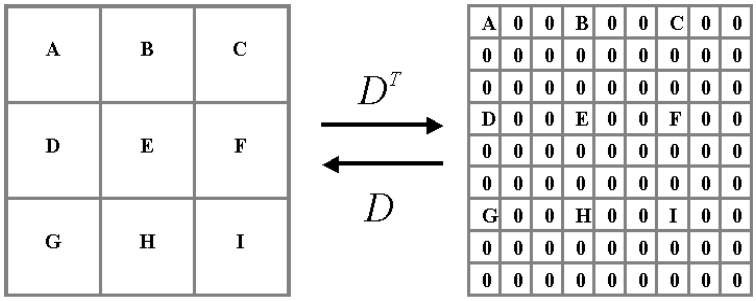



Fig. 3. Effect of upsampling $D^T$ matrix on a $3 \times 3$ image and downsampling matrix $D$ on the corresponding $9 \times 9$ upsampled image (resolution enhancement factor of three). In this figure, to give better intuition the image vectors are reshaped as matrices.

the pixel values. Therefore, each element of $\underline{G}_1$, which corresponds to one element in $\underline{\widehat{Z}}$, is the aggregate of the effects of all LR frames. The effect of each frame has one of the following three forms:

1) Addition of zero, which results from zero filling.
2) Addition of $+1$, which means a pixel in $\underline{\widehat{Z}}$ was larger than the corresponding contributing pixel from frame $\underline{Y}_k$.
3) Addition of $-1$, which means a pixel in $\underline{\widehat{Z}}$ was smaller than the corresponding contributing pixel from frame $\underline{Y}_k$.

A zero gradient state ($\underline{G}_1 = \underline{0}$) will be the result of adding an equal number of $-1$ and $+1$, which means each element of $\widehat{\underline{Z}}$ should be the median value of corresponding elements in the LR frames. $\widehat{\underline{X}}$, the final super-resolved picture, is calculated by deblurring $\widehat{\underline{Z}}$ .

So far we have shown that $p = 1$ results in pixelwise median and $p = 2$ results in pixelwise mean of all measurements after motion compensation. According to (11), if $1 < p < 2$ then both $sign(DF_k\underline{Z}_n - \underline{Y}_k)$ and $|DF_k\underline{Z}_n - \underline{Y}_k|^{p-1}$ terms appear in $\underline{G}_p$. Therefore, when the value of $p$ is near one, $\widehat{\underline{Z}}$ is a weighted mean of measurements, with much larger weights around the measurements near the median value, while when the value of $p$ is near two the weights will be distributed more uniformly.

In this subsection we studied $L_p, 1 \leq p \leq 2$ norm minimization family. As $p \longrightarrow 1$, this estimator takes the shape of median estimator, which has the highest breakpoint value, making it the most robust cost function. For the rest of this paper, we choose $L_1$ to minimize the measurement error (note that we left out the study of $L_p, 0 \leq p < 1$ norm minimization family as they are not convex functions).

In the square or underdetermined cases ($N = r^2$ and $N < r^2$ respectively), there is only one measurement available for each HR pixel and as median and mean operators for one or two measurements give the same result, $L_1$ and $L_2$ norm minimization will result in identical answers, also certain pixel locations will have no estimate at all. For these cases, it is essential for the estimator to have an extra term, called regularization term, to remove outliers. The next section discusses different regularization terms and introduces a robust and convenient regularization term.

### *C. Robust Regularization*

Super-resolution is an ill-posed problem [11], [19]. For the underdetermined cases, there exist an infinite number of solutions which satisfy (2). The solution for square and overdetermined cases is not stable, which means small amount of noise in measurements will result in large perturbations in the final solution. Therefore, considering regularization in super-resolution algorithm as a means for picking a stable solution is very useful, if not 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 is easy to implement.

A regularization term compensates the missing measurement information with some general prior information about the desirable HR solution, and is usually implemented as a penalty factor in the generalized minimization cost function (5):

$$\widehat{\underline{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[ \sum_{k=1}^{N} \rho(\underline{Y}_k, D_k H_k F_k \underline{X}) + \lambda \Upsilon(\underline{X}) \right] \tag{14}$$

where $\lambda$, the regularization parameter, is a scalar for properly weighting the first term (similarity cost) against the second term (regularization cost) and $\Upsilon$ is the regularization cost function.

One of the most widely referenced regularization cost functions is the Tikhonov cost function [10], [11]:

$$\Upsilon_T(\underline{X}) = \|\Gamma \underline{X}\|_2^2 \tag{15}$$

where $\Gamma$ is usually a high-pass operator such as derivative, Laplacian, or even identity matrix. The intuition behind this regularization method is to limit the total energy of the image (when $\Gamma$ is the identity matrix) or forcing spatial smoothness (for derivative or Laplacian choices of $\Gamma$). As the noisy and edge pixels both contain high-frequency energy, they will be removed in the regularization process and the resulting denoised image will not contain sharp edges.

Certain types of regularization cost functions work efficiently for some special types of images but are not suitable for general images (such as maximum entropy regularization which produce in sharp reconstructions of point objects, such as star fields in astronomical images [20]).

One of the most successful regularization methods for denoising and deblurring is the Total Variation (TV) method [21]. The total variation criterion penalizes the total amount of change in the image as measured by the $L_1$ norm of the gradient and is defined as:

$$\Upsilon_{TV}(\underline{X}) = \|\nabla \underline{X}\|_1$$

where $\nabla$ is the gradient operator. The most useful property of total variation criterion is that it tends to preserve edges in the reconstruction [20], [21], [22], as it does not severely penalize steep local gradients.

Based on the spirit of total variation criterion, and a related technique called the bilateral filter (Appendix I), we introduce our robust regularizer called Bilateral-TV, which is computationally cheap to implement, and preserves edges. The regularizing function looks like,

$$\Upsilon_{BTV}(X) = \sum_{l=0}^{P} \sum_{m=0}^{P} \alpha^{m+l} \|\underline{X} - S_x^l S_y^m \underline{X}\|_1 \tag{16}$$

where matrices (operators) $S_x^l$, and $S_y^k$ shift $\mathbf{X}$ by $l$, and $k$ pixels in horizontal and vertical directions repectively, presenting several scales of derivatives. The scalar weight $\alpha$, $0<\alpha<1$, is applied to give a spatially decaying effect to the summation of the regularization term.

It is easy to show that this regularization method is a generalization of other popular regularization methods. If we limit $m, l$ to the two cases of $m = 1$, $l = 0$ and $m = 0$, $l = 1$ with $\alpha = 1$, and define operators $Q_x$ and $Q_y$ as representatives of the first derivative ($Q_x = I - S_x$ and $Q_y = I - S_y$) then (16)

results in:

$$\Upsilon_{BTV}(X) = \|Q_x \underline{X}\|_1 + \|Q_y \underline{X}\|_1 \tag{17}$$

which is suggested in [23] as a reliable and computationally efficient approximation to the Total-Variation prior [21].

To compare the performance of Bilateral-TV ($P \geq 1$) to common TV prior ($P = 1$), we set up the following denoising experiment. We added Gaussian white noise of mean zero and variance 0.045 to the image in Figure 4(a) resulting in the noisy image of Figure 4(b). If $\underline{X}$ and $\underline{Y}$ represent the original and corrupted images then following (14), we minimized:

$$\underline{\widehat{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[ \|\underline{Y} - \underline{X}\|_2^2 + \lambda \Upsilon(\underline{X}) \right] \tag{18}$$

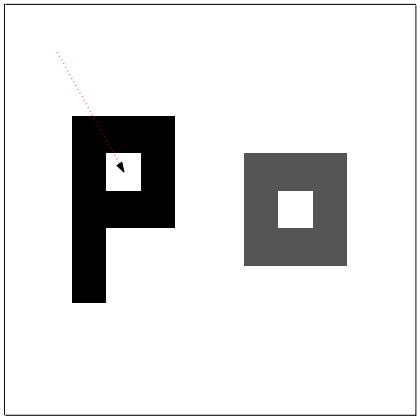
to reconstruct the noisy image. Tikhonov denoising resulted in Figure 4(c)[1], where $\Gamma$ in (15) was replaced by the Laplacian kernel:

$$\Gamma = \frac{1}{8} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{19}$$

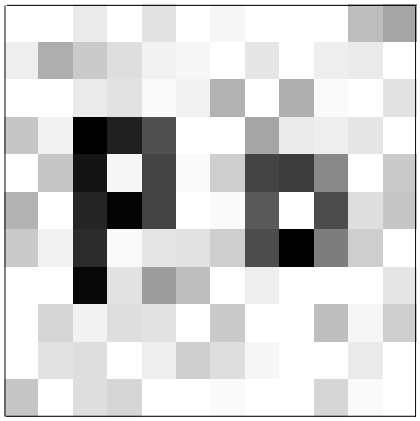
The result of using TV prior ($P = 1$) for denoising is shown in Figure 4(d). Figure 4(e) shows the result of applying Bilateral-TV prior ($P = 3$). Notice the effect of each reconstruction method on the pixel indicated by an arrow in Figure 4(a). As this pixel is surrounded by non-similar pixels, TV prior considers it as a heavily noisy pixel, and uses the value of immediate neighboring pixels to estimate its original value. On the other hand, Bilateral-TV considers a larger neighborhood. By bridging over immediate neighboring pixels, the value of similar pixels are also considered in graylevel estimation of this pixel, therefore the smoothing effect in Figure 4(e) is much less than Figure 4(d). Figure 4(f) compares the performance of TV and Bilateral-TV denosing methods in estimating graylevel value of the arrow indicated pixel. Unlike Bilateral-TV regularization, increasing the number of iterations in Tikhonov and TV regularization will result in more undesired smoothing.

To compare the performance of our regularization method to Tikhonov regularization method, we set up another experiment. We corrupted an image by blurring it with a Gaussian blur kernel followed by adding Gaussian additive noise. We reconstructed the image using Tikhonov and our proposed regularization terms (this can be assumed as a super-resolution problem with resolution factor of one). If $\underline{X}$ and $\underline{Y}$

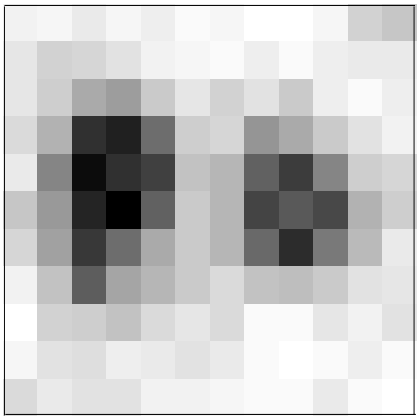
[1]The criteria for parameter selection in this example (and other examples discussed in this paper) was to choose parameters which produce visually most appealing results. Therefore to ensure fairness, each experiment was repeated several times with different parameters and the best result of each experiment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



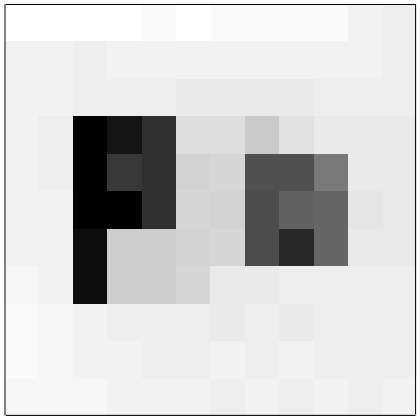
a: Original



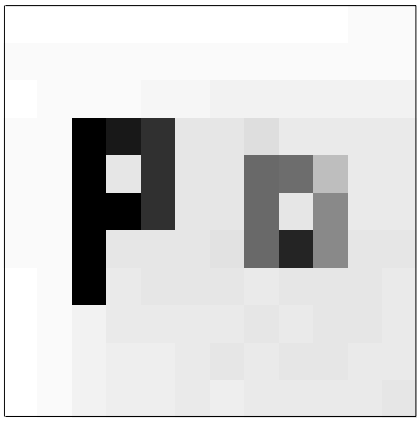
b: Noisy



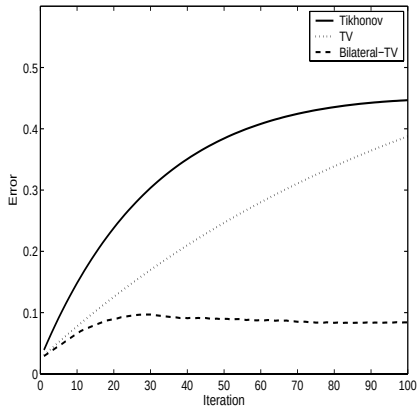
c:Reconstruction using Tikhonov



d:Reconstruction using TV



e:Reconstruction using Bilateral-TV




f:Error in gray-level value estimation of pixel indicated by arrow

Fig. 4. a-e: Simulation results of denoising using different regularization methods. f: Error in gray-level value estimation of the pixel indicated by arrow in (a) versus the iterations number in Tikhonov (solid line), TV (dotted line), and Bilateral TV (broken line) denoising.

represent the original and corrupted images and $H$ represents the matrix form of the blur kernel then following (14), we minimized:

$$\underline{\widehat{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[ \|\underline{Y} - H\underline{X}\|_2^2 + \lambda \Upsilon(X) \right] \tag{20}$$

to reconstruct the blurred noisy image.

Figure 5 shows the results of our experiment. Figure 5(a) shows the original image($\underline{X}$). Figure 5(b) is the corrupted $\underline{Y} = H\underline{X} + \underline{V}$, where $V$ is the additive noise. Figure 5(c) is the result of reconstruction with Tikhonov regularization, where $\Gamma$ in (15) was replaced by the Laplacian kernel (19). Figure 5(d) shows the result of applying our regularization criterion (16) with the following parameters: $\alpha = .7$ and $P = 2$. The best Mean Square Error (MSE) achieved by Tikhonov regularization was 313 versus 215 for the proposed regularization. The edge preserving and the superiority of the bilateral prior is apparent in this example.

### *D. Robust Super-Resolution Implementation*

In this subsection, based on the material that was developed in sections II-B and II-C, a solution for the robust super-resolution problem will be proposed. Recall that

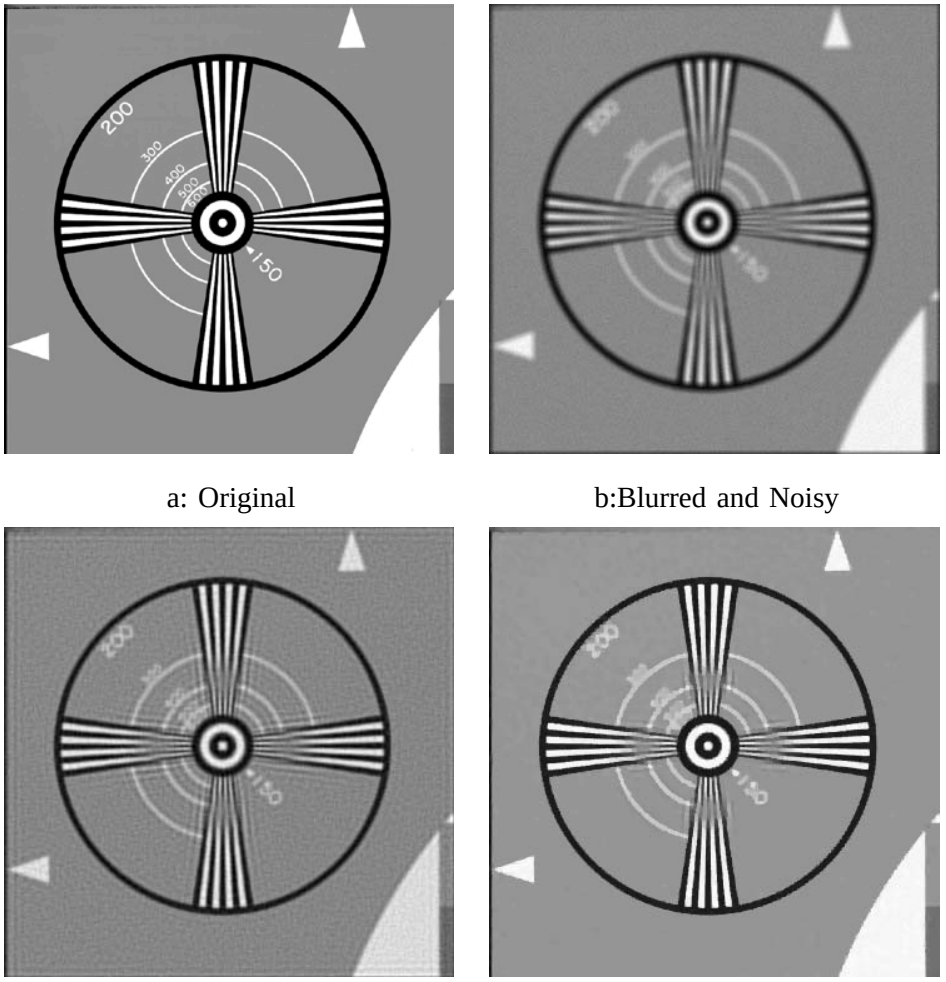
$$\begin{aligned} \underline{\widehat{X}} \;=\; \underset{\underline{X}}{\text{ArgMin}} \Bigg[ & \sum_{k=1}^{N} \|D_k H_k F_k \underline{X} - \underline{Y}_k\|_1 + \\ & \lambda \sum_{l=0}^{P} \sum_{m=0}^{P} \alpha^{m+l} \|\underline{X} - S_x^l S_y^m \underline{X}\|_1 \Bigg] \end{aligned} \tag{21}$$

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

$$\begin{aligned} \underline{\widehat{X}}_{n+1} = \underline{\widehat{X}}_n - \beta \Bigg\{ & \sum_{k=1}^{N} F_k^T H_k^T D_k^T sign(D_k H_k F_k \underline{\widehat{X}}_n - \underline{Y}_k) \\ & + \lambda \sum_{l=0}^{P} \sum_{m=0}^{P} \alpha^{m+l} [I - S_y^{-m} S_x^{-l}] sign(\underline{\widehat{X}}_n - S_x^l S_y^m \underline{\widehat{X}}_n) \Bigg\} \end{aligned} \tag{22}$$

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 of $S_x^l$ and $S_y^m$.

Simulation results in Section IV will show the strength of the proposed algorithm. The matrices $F$, $H$, $D$, $S$ and their transposes can be exactly interpreted as direct image operators such as shift, blur, and decimation. Noting and implementing the effects of these matrices as a sequence of operators spares us from explicitly constructing them as matrices. This property helps our method to be implemented in an



a: Original b:Blurred and Noisy

c:Best Tikhonov Regularization d:Proposed Regularization

Fig. 5. Simulation results of deblurring using different regularization methods.

extremely fast and memory efficient way.

Figure 6 is the block diagram representation of (22). There, each LR measurement $Y_k$ will be compared to the warped, blurred and decimated current estimate of HR frame $X_n$. Block $G_k$ represents the gradient back projection operator that compares the $k^{th}$ LR image to the estimate of the HR image in the $n^{th}$ steepest descent iteration. Block $R_{l,m}$ represents the gradient of regularization term, where the HR estimate in the $n^{th}$ steepest descent iteration is compared to its shifted version ($l$ pixel shift in horizontal and $m$ pixel shift in vertical directions).

Details of the blocks $G_k$ and $R_{m,l}$ are defined in Figures 7(a) and 7(b). Block $T(PSF)$ in Figure 7(a) replaces the matrix $H_k^T$ with a simple convolution. Function $T$ flips the columns of PSF kernel in the

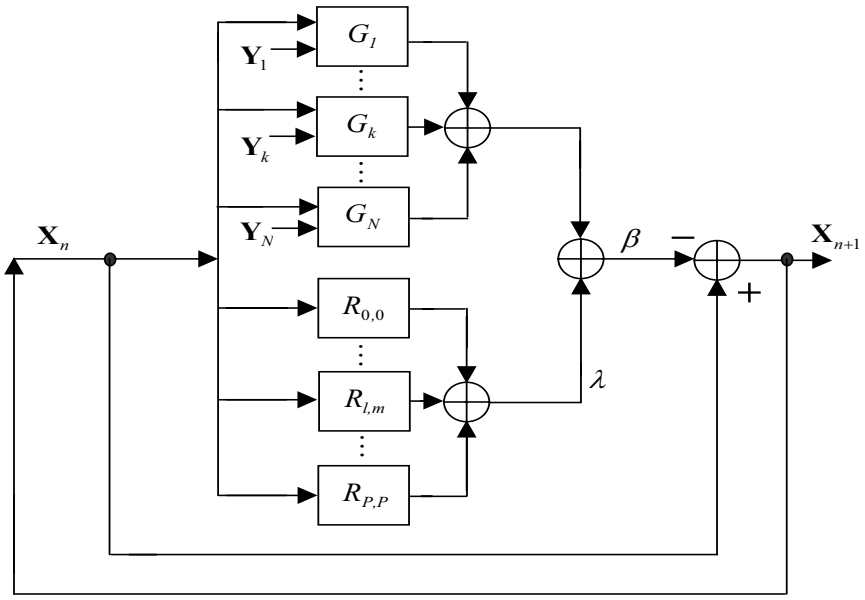



Fig. 6. Block diagram representation of (22), blocks $G_k$ and $R_{m,l}$ are defined in Figure 7.

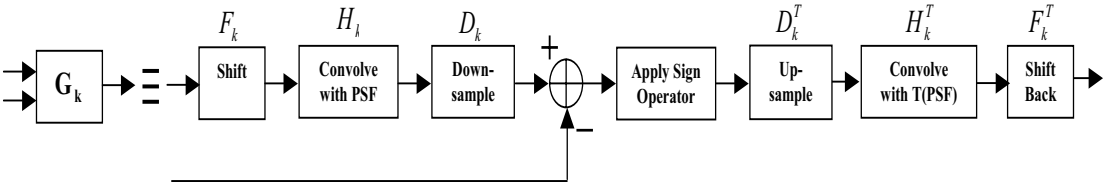
left-right direction (that is, about the vertical axis), and then flips the rows of PSF kernel in the up-down direction (that is, about the horizontal axis)[2]. The $D_k^T$ up-sampling block in Figure 7(a) can be easily implemented by filling $r-1$ zeros both in vertical and horizontal directions around each pixel (Figure 3). And finally the $F_k^T$ shift-back block in Figure 7(a), is implemented by inverting the translational motion in the reverse direction. Note that even for the more general affine motion model a similar inverting property (though more complicated) is still valid.

Parallel processing potential of this method, which significantly increases the overall speed of implementation, can be easily interpreted from Figure 6 (the computation of each $G_k$ or $R_{l,m}$ blocks may be assigned to a separate processor).

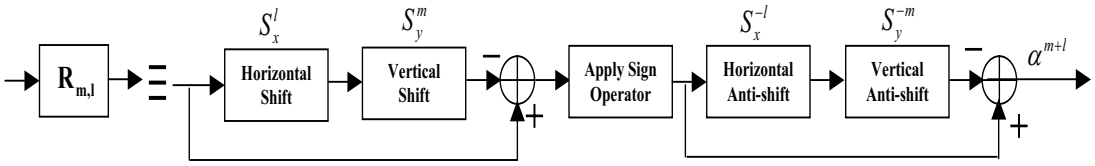
Our robust super-resolution approach also has an advantage in the computational aspects over other methods including the one proposed in [13]. In our method, an inherently robust cost function has been proposed, for which a number of computationally efficient numerical minimization methods[3] are applicable. On the contrary, [13] uses steepest descent method to minimize the non-robust $L_2$ norm cost function, and robustness is achieved by modifying the steepest descent method, where median

[2]If the PSF kernel has even dimensions, one extra row or column of zeros will be added to it, to make it odd size (zero columns and rows have no effect in convolution process).

[3]Such as Conjugate Gradient (CG), Preconditioned Conjugate Gradient (PCG), Jacobi, and many others.




a:Block diagram representation of similarity cost derivative($G_k$)




b:Block diagram representation of regularization cost derivative ($R_{m,l}$)

Fig. 7. Extended Block diagram representation of $G_k$ and $R_{m,l}$ blocks in Figure 6.

operator is used in place of summation operator in computing the gradient term of (12). Implementing the same scheme of substituting summation operator with median operator in computationally more efficient methods such as conjugate gradient is not a straightforward task and besides it is no longer guaranteed that the modified steepest descent and conjugate gradient minimization converge to the same answer.

As an example, Figure 8(a) and Figure 8(b) show the result of implementing the proposed method on the same image sets that was used to generate Figures 2(d), and Figure 2(f) respectively. The outlier effects have been reduced significantly (more detailed examples are presented in section IV).

In the next section we propose an alternate method to achieve further improvements in computational efficiency.

## III. Further Extensions and Improvements

### *A. Fast Robust Super-Resolution Formulation*

In Section II-D, we proposed an iterative robust super-resolution method based on equation (21). Although implementation of (21) will result in one of the fastest ever proposed super-resolution methods, for real-time image sequence processing, faster methods are always desirable. In this subsection, based on the interpretation of (13) that was offered in Section II-B, we simplify (21) to achieve a faster method. In this method, resolution enhancement is broken into two consecutive steps:

a b

Fig. 8. Reconstruction of the outlier contaminated images in Figure 2 using equation (22). (a) is the robust reconstruction of the same images that was used to produce Figure 2(d) and (b) is the robust reconstruction of the same images that was used to produce Figure 2(f).

1) Non-iterative data fusion.
2) Iterative deblurring-interpolation

As we described in section II-B, registration followed by the median operation (what we call median shift and add) results in $\underline{\widehat{Z}} = H\underline{\widehat{X}}$.

The goal of the deblurring-interpolation step is finding the deblurred HR frame $\underline{\widehat{X}}$. Note that for the under-determined cases not all $\underline{\widehat{Z}}$ pixel values can be defined in the data fusion step, and their values should be defined in a separate interpolation step. In this paper unlike [12], interpolation and deblurring are done simultaneously.

The following expression formulates our minimization criterion for obtaining $\underline{\widehat{X}}$ from $\underline{\widehat{Z}}$.

$$\underline{\widehat{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[ \|A(H\underline{X} - \underline{\widehat{Z}})\|_1 + \lambda' \sum_{l=0}^{P} \sum_{m=0}^{P} \alpha^{m+l} \|\underline{X} - S_x^l S_y^m \underline{X}\|_1 \right] \tag{23}$$

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

As $A$ is a diagonal matrix, $A^T = A$, and the corresponding steepest descent solution of minimization

problem (23) can be expressed as:

$$\underline{\widehat{X}}_{n+1}=\underline{\widehat{X}}_{n}-\beta\Bigg\{H^{T}A^{T}sign(AH\underline{\widehat{X}}_{n}-A\underline{\widehat{Z}})$$

$$+\lambda^{'}\sum_{l=0}^{P}\sum_{m=0}^{P}\alpha^{m+l}[I-S_{y}^{-m}S_{x}^{-l}]sign(\underline{\widehat{X}}_{n}-S_{x}^{l}S_{y}^{m}\underline{\widehat{X}}_{n})\Bigg\} \tag{24}$$

Decimation and warping matrices ($D$ and $F$) and summation of measurements are not present anymore, which makes the implementation of (24) much faster than (22). Note that physical construction of matrix $A$ is not necessary as it can be implemented as a mask matrix with the size equal to image $\mathbf{X}$.

### *B. Bilateral Non Iterative Artifact Removal*

In Section II-D we argued that the combination of $L_1$ error minimization and robust regularization efficiently removes outliers and results in sharp edges for any kind of data set (over or under determined cases). In Section III-A we presented a faster two step method. Although the non-iterative data fusion step works well for over-determined systems, it fails to remove outliers in the under-determined case, as it ignores the correlation of neighboring pixels. The outliers may be removed by the regularization term in the deblurring-interpolation step, but this needs a relatively large regularization factor $\lambda^{'}$, which may result in blurring of the final answer. In this section, we add a non-iterative outlier removal step, after data fusion and, before deblurring-interpolation step using the bilateral filter.

Our proposed method essentially calculates the correlation of different measurements (pixels from different frames) with each other and removes the inconsistent data. The computed correlation is based on the bilateral idea, so the high-frequency (edge-information) data will be differentiated from outliers. We assign a weight to each pixel in the measurements based on its bilateral correlation with corresponding pixels in the data-fused image. After computing these weights, pixels with very small weights will be removed from the data set. As pixels containing high-frequency information receive higher weights than the ones located in the low-frequency areas, it is reasonable to compute and compare the penalty weights for blocks of pixels rather than for single pixels.

The penalty weight for pixel $(i,j)$ in $k^{th}$ LR frame is calculated as:

$$w_{k}(i,j)=\sum_{m=-q}^{q}\sum_{n=-q}^{q}e^{-\frac{1}{2}(\frac{|\mathbf{Y}_{k}(i,j)-\widehat{\mathbf{Z}}(i+m,j+n)|}{\sigma_{r}})}\times e^{-\frac{1}{2}\frac{m^{2}+n^{2}}{\sigma_{d}^{2}}} \tag{25}$$

where $\widehat{\mathbf{Z}}(i,j)$ is the gray level value of pixel $(i,j)$ in the data fused image discussed in III-A, $\mathbf{Y}_{k}(i,j)$ is the gray level value of pixel $(i,j)$ in the $k^{th}$ LR frame, $q$ controls the bilateral kernel size, $\sigma_{r}$ and $\sigma_{d}$

control corresponding photometric and distance spread.

If each LR frame is divided into $R$ blocks then the overall weight of block $r$ of the $k$th frame is:

$$W_{k,r} = \sum_{i,j \in \Omega_{k,r}} w_k(i,j) \tag{26}$$

where $\Omega_{k,r}$ defines the pixels in the $r$th block of the $k$th LR frame.

In the next step, mean and variance of all corresponding blocks are computed and the blocks with weights smaller than $mean(W_{k,r}) - n \times var(W_{k,r})$ will be removed from the input data set (constant $n$ will control the number of rejected blocks). Finally, the value of $\underline{\widehat{Z}}$ for the blocks which were labelled as outliers will be recalculated. Then (24) may be used to iteratively calculate $\underline{\widehat{X}}$.

## IV. Experiments

In this section we compare the performance of the different resolution enhancement algorithms proposed in this paper to existing resolution enhancement methods. The first example is a controlled simulated experiment. In this experiment we create a sequence of LR frames by using one HR image (Figure 9(a)), downsampling it by factor of 4 in each axis and choosing all 16 possible starting points. Then to simulate the effect of camera PSF, each of the 16 shifted HR images was convolved with a symmetric Gaussian lowpass filter of size $4 \times 4$ with standard deviation equal to one. After subsampling these HR blurred images by the factor of four, we added Gaussian noise to the resulting LR frames to achieve SNR[4] equal to 18dB. One of these LR frames is presented in Figure 9(b). To simulate the errors in motion estimation, a bias equal to one pixel shift was intentionally added to the known motion vectors of three LR frames.

The result of implementing the non-iterative resolution enhancement method described in [5] is shown in Figure 9(c). It is not surprising to see the motion error artifacts in the HR frame as the HR image is the result of zero-filling, shifting and adding the LR measurements. Deblurring this result with Wiener method (Figure 9(d)) does not remove these artifacts, of course. For reference, Figure 9(e) shows the result of applying an iterative method based on minimizing $L_2$ norm, both for the residual and the regularization terms. The following equation describes this minimization criterion:

$$\underline{\widehat{X}} = \text{ArgMin} \left[ \sum_{k=1}^{N} \|D_k H_k F_k \underline{X} - \underline{Y}_k\|_2^2 + \lambda \|\Gamma \underline{X}\|_2^2 \right] \tag{27}$$

in which $\Gamma$ is defined in (19) and regularization factor $\lambda$ was chosen to be 0.4. As $L_2$ norm is not robust to motion error, motion artifacts are still visible in the result. Note that the relatively high regularization

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

factor which was chosen to reduce the motion artifact has resulted in a blurry image.

The robust super-resolution method which was proposed in [13] resulted in Figure 9(f). Figure 10(a) was resulted by simply adding the regularization term defined in (27) to the proposed method of [13] which is far better than the $L_2$ approach, yet exhibiting some artifacts. Figure 10(b) shows the implementation of the proposed method described in section II-D. The selected parameters for this method were as follows: $\lambda = 0.005$, $p = 2$, $\beta = 110$, $\alpha = 0.6$. Figure 10(c) shows the implementation of the proposed method described in section III-B. The selected parameters for this method were as follows: $n = 1$, $p = 2$ and $\sigma_d^2 = \sigma_r^2 = 1.5$. $\Omega$ was chosen to be the size of each LR frame. After implementing this method all three frames with erroneous motion vectors were detected and removed from data set. Finally, Figure 10(d) shows the iteratively deblurred version of Figure 10(c) using the method discussed in Section III-A. The selected parameters for this method were as follows: $\lambda = 0.1$, $p = 2$, $\beta = .3$, $\alpha = 0.8$. Comparing Figure 10(b) and 10(d) to other methods, we notice not only our method has removed the outliers more efficiently, but also it has resulted in sharper edges without any ringing effects.

Our second example is a real infrared camera image sequences with no known outliers; courtesy of B. Yasuda and the FLIR research group in the Sensors Technology Branch, Wright Laboratory, WPAFB, OH. We used eight LR frames in our reconstruction to get resolution enhancement factor of four (Figure 11(a) shows one of the input LR images)[5]. Figure 11(b) shows the cubic spline interpolation of Figure 11(a) by factor of four. The (unknown) camera PSF was assumed to be a $4 \times 4$ Gaussian kernel with standard deviation equal to one. We used the method described in [24] to computed the motion vectors. $L_2$ norm reconstruction with Tikhonov regularization (27) result is shown in Figure 11(c) where $\Gamma$ is defined in (19) and regularization factor $\lambda$ was chosen to be 0.1. Figure 11(d) shows the implementation of (22) with the following parameters $\lambda = 0.006$, $p = 2$, $\beta = 81$, and $\alpha = 0.5$. Although modelling noise in these frames as additive Gaussian is a reasonable assumption, our method achieved a better result than the best $L_2$ norm minimization.

The final experiment is a real compressed sequence of 20 images from a commercial video camera; courtesy of Adyoron Intelligent Systems LTD., Tel Aviv, Israel. Figure 12(a) is one of these LR images and Figure 12(b) is the cubic spline interpolation of this image by factor of 3. We rotated five frames of this sequence (rotation from $20^\circ$ to $60^\circ$), creating a sequence of images with relative affine motion. The (unknown) camera PSF was assumed to be a $5 \times 5$ Gaussian kernel with standard deviation equal to 2. We used the method described in [24] to computed the motion vectors with translational motion

[5]Note that this is an underdetermined scenario.

assumption. The error in motion modelling results in apparent shadows in $L_2$ norm reconstruction with Tikhonov regularization (Figure 12(c)) where $\Gamma$ is defined in (19) and regularization factor $\lambda$ was chosen to be 0.5. These shadows are removed in Figure 12(d), where the method described in Section II-D (22) was used for reconstruction with the following parameters $\lambda = 0.003$, $p = 2$, $\beta = 50$, and $\alpha = 0.7$.

## V. Conclusions and Future Work

In this paper, we presented an algorithm to enhance the quality of a set of noisy blurred images and produce a HR image with less noise and blur effects. We presented a robust super-resolution method based on the use of $L_1$ norm both in the regularization and the measurement terms of our penalty function. We showed that our method removes outliers efficiently, resulting in images with sharp edges. Even for images in which the noise followed the Gaussian model, $L_1$ norm minimization results were as good as $L_2$ norm minimization method results, which encourages using $L_1$ norm minimization for any data set. The proposed method was fast and easy to implement. We also proposed and mathematically justified a very fast method based on pixelwise shift and add method and related it to $L_1$ norm minimization when relative motion is pure translational, and PSF and decimation factor is common and space-invariant in all LR images. Note that the mathematical derivation of the proposed shift and add method was independent of the constraint over decimation factor, but we included it as this constraint distinguishes super-resolution problem from the more general problem of multi-scale image fusion.

Success of our algorithm in removing motion estimation related outliers provides room for future research. For instance the rejected frames (pixels) can be detected and fed back to refine the motion estimation. The motion estimation algorithm may consider other motion models or use more iterations to recalculate the relative motion of these frames (pixels). Also, presumably motion estimation with respect to super-resolved image will be more accurate.

Other important extension for our algorithm include incorporation of blur identification algorithms in the super-resolution method. Although many single-frame blind deconvolution algorithms have been suggested in the last 30 years [25], and recently [26] incorporated a single parameter blur identification algorithm in their super-resolution method, still there is need for more research to provide a super-resolution method along with a more general blur estimation algorithm.

Few papers have addressed resolution enhancement of compressed video sequences [14] and [15]. Compression artifacts resulting from quantization of DCT coefficients can dramatically decrease the performance of super-resolution system. The results of Section III-A may be used to design a very fast none iterative method for reducing the compression artifacts in the super-resolved images.

One of the most apparent effects of DCT based compression methods such as *MPEG* for video and *JPEG* for still images is the blocking artifact. The quantization noise variance of each pixel in a block is space variant. For a block located in a low-frequency content area, pixels near boundaries contain more quantization noise than the interior pixels . On the other hand, for the blocks located in the high-frequency area, pixels near boundaries contain less quantization noise than the interior pixels [27].

This space-variant noise property of the blocks may be exploited to reduce the quantization noise. Because of the presence of motion in video sequences, pixel locations in the blocks change from one frame to the other. So two corresponding pixels from two different frames may be located on and off the boundaries of the blocks in which they are located. Based on the discussion that was presented in the previous paragraph, it is easy to determine which pixel has less quantization noise. It is reasonable to assign a higher weight to those pixels which suffer less from quantization noise in the data fusion step which was explained in III-A. The relative magnitude of the weight assigned because of quantization and the weight that was explained in III-A will depend on the compression ratio.

## APPENDIX I
## BILATERAL FILTER

The idea of the bilateral filter was first proposed in [28] as a very effective one pass filter for denoising purposes while keeping sharp edges. Unlike conventional filters such as Gaussian low-pass filter, the bilateral filter defines the closeness of two pixels not only based on geometric distance but also based on photometric distance. Considering 1-D case (for simplifying the notations), the result of applying bilateral filter for the $k$th sample in the estimated signal $\widehat{X}$ is:

$$\widehat{X}[k] = \frac{\sum_{-N}^{N} W[k,n]Y[k-n]}{\sum_{-N}^{N} W[k,n]} \tag{28}$$

where $\underline{Y} = \underline{X} + \underline{V}$ is the noisy image (vector). The weight $W[k,n] = W_S[k,n]W_P[k,n]$ considers both photometric and spatial difference of sample $k$ in noisy vector $\underline{Y}$ from its neighbors to define the value of sample $k$ in the estimated vector $\underline{\widehat{X}}$ . The spatial and photometric difference weights were arbitrary defined in [28] as

$$W_S[k,n] = exp\left\{-\frac{n^2}{2\sigma_S^2}\right\}$$

$$W_P[k,n] = exp\left\{-\frac{[Y[k]-Y[k-n]]^2}{2\sigma_R^2}\right\} \tag{29}$$

where parameters $\sigma_S^2$ and $\sigma_R^2$ control the strength of spatial and photometric property of the filter respectively.

In [29] it was proved that such filter is a single iteration of the following weighted least squares minimization

$$\begin{aligned}\widehat{\underline{X}} &= \underset{\underline{X}}{\text{ArgMin}} \left[ \|\underline{X} - \underline{Y}\|_2^2 + \lambda \sum_{n=1}^{N} \|\underline{X} - S^n \underline{X}\|_{W_n}^2 \right] \\ &= \underset{\underline{X}}{\text{ArgMin}} \left[ [\underline{X} - \underline{Y}]^T [\underline{X} - \underline{Y}] + \lambda \sum_{n=1}^{N} [\underline{X} - S^n \underline{X}]^T \mathbf{W_n} [\underline{X} - S^n \underline{X}] \right] \end{aligned} \tag{30}$$

with Jacobi method, where $S^n$ implies a shift right of $n$ samples. [29] also showed that using more iterations will enhance the performance of this filter.

Note that if we define the $(i,i)$th element of the diagonal weight matrix $\mathbf{W_n}$ as

$$\mathbf{W_n}(i,i) = \frac{\alpha^n}{|\underline{X}(i) - S^n \underline{X}(i)|} \qquad 0 < \alpha < 1$$

that is weighting the estimate with respect to both photometric distance $|\underline{X}(i) - S^n \underline{X}(i)|$ and geometric distance $\alpha^n$ (which is the main property of a bilateral filter), then (30) will become

$$\widehat{\underline{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[ \|\underline{X} - \underline{Y}\|_2^2 + \lambda \sum_{n=1}^{N} \alpha^n \|\underline{X} - S^n \underline{X}\|_1 \right] \tag{31}$$

which is the 1-D version of the Bilateral-TV criterion in (16).

## APPENDIX II
## THE LIMITATIONS OF ZOMET METHOD

A robust super-resolution method is recently proposed by Zomet et al. [13], where robustness is achieved by modifying the gradient of the $L_2$ norm cost function (7):

$$\underline{G}_2 = \sum_{k=1}^{N} \underline{B}_k = \sum_{k=1}^{N} F_k^T H_k^T D_k^T (D_k H_k F_k \underline{X} - \underline{Y}_k) = \sum_{k=1}^{N} F_k^T H_k^T D_k^T \underline{U}_k \tag{32}$$

in which $\underline{B}_k$ is the gradient resulted from frame $k$ and $\underline{U}_k$ represents the residual vector.They substituted (32) with the following approximation:

$$\widehat{\underline{G}}_2 = \text{MED}\{\underline{B}_k\}_{k=1}^N = \text{MED}\{F_k^T H_k^T D_k^T \underline{U}_k\}_{k=1}^N \tag{33}$$

where MED is a pixelwise median operator. Then steepest descent minimization was used to calculate $\widehat{\underline{X}}$:

$$\widehat{\underline{X}}_{n+1} = \widehat{\underline{X}}_n + \lambda'' \widehat{\underline{G}}_2 \tag{34}$$

where $\lambda''$ is the step size in the direction of gradient.

We show that in many different scenarios the approximated gradient (33) is zero in all iterations, which means estimated HR frame of the $n^{th}$ iteration ($X_n$) is the same as the initial guess ($X_0$). To appreciate

this fact, lets start with a square case in which blurring effect is negligible (i.e. $H_k$ is an identity matrix resulting in $B_k = F_k^T D_k^T \underline{U}_k$). A quick consultation with Figure 3 suggests that only one of every $r^2$ elements in $D_k^T \underline{U}_k$ has a non-zero value. Moreover, recall that $F_k^T$ just registers vector $D_k^T \underline{U}_k$ with respect to the estimated relative motion without changing its value. According to (33), $\widehat{\underline{G}}(i)$ (the $i^{th}$ element of the gradient vector) is equal to MED$\{\underline{B}_k(i)\}_{k=1}^N$. As $N-1$ elements in $\{\underline{B}_k(i)\}_{k=1}^N$ have zero value, their median will also be zero. Therefore every element of the approximated gradient vector will be zero. Even for a more general case in which the effect of blur matrix is not negligible ($H_k$ is a matrix form of a $m \times n$ blur kernel), the same approach may be employed to show that unless $(m \times n > \frac{r^2}{2})$, the gradient will remain zero for all iterations[6].

This limitation can be removed by modifying the MED operator in (33). This modified median operator does not consider those elements of $\underline{B}_k(i)$ which are the result of zero filling. It is interesting to note that such assumption will result in estimating the HR frame as the median of registered LR frames after zero filling, which is the exact interpretation of using $L_1$ norm minimization discussed in Section II-B.

## REFERENCES


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

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

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

[4] L. Teodosio and W. Bender, "Salient video stills: Content and context preserved," in *Proceedings of First ACM International Conference on Multimedia*, vol. 10, Aug. 1993, pp. 39–46.

[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*, vol. 10, no. 8, pp. 1187–1193, Aug. 2001.

[6] M. C. Chiang and T. E. Boulte, "Efficient super-resolution via image warping," *Image and Vision Computing*, vol. 18, no. 10, pp. 761–771, July 2000.

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

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

[9] H. Ur and D. Gross, "Improved resolution from sub-pixel shifted pictures," *CVGIP:Graph. Models Image Processing*, vol. 54, no. 181–186, Mar. 1992.


[6]This condition is also valid for underdetermined and even some overdetermined cases.

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

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

[12] S. Lertrattanapanich and N. K. Bose, "High resolution image formation from low resolution frames using delaunay triangulation," *IEEE Trans. Image Processing*, vol. 11, no. 12, pp. 1427–1441, Dec. 2002.

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

[14] Y. Altunbasak, A. Patti, and R. Mersereau, "Super-resolution still and video reconstruction from mpeg-coded video," *IEEE Trans. Circuits And Syst. Video Technol.*, vol. 12, no. 4, pp. 217–226, Apr. 2002.

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

[16] G. C. Calafiore, "Outliers robustness in multivariate orthogonal regression," *IEEE Trans. Syst. Man and Cybernetics*, vol. 30, no. 6, pp. 674–679, Nov. 2000.

[17] P. J. Huber, *Robust Statitics*. New York: Wiley, 1981.

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

[19] A. M. Tekalp, *Digital video processing*. Prentice-Hall, 1995.

[20] A. Bovik, *Handbook of image and video processing*. New Jersey: Academic Press Limited, 2000.

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

[22] T. F. Chan, S. Osher, and J. Shen, "The digital TV filter and nonlinear denoising," *IEEE Trans. Image Processing*, vol. 10, no. 2, pp. 231–241, Feb. 2001.

[23] Y. Li and F. Santosa, "A computational algorithm for minimizing total variation in image restoration," *IEEE Trans. Image Processing*, vol. 5, no. 6, pp. 987–995, June 1996.

[24] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierachical model-based motion estimation," *Proceedings European Conference on Computer Vision*, pp. 237–252, 1992.

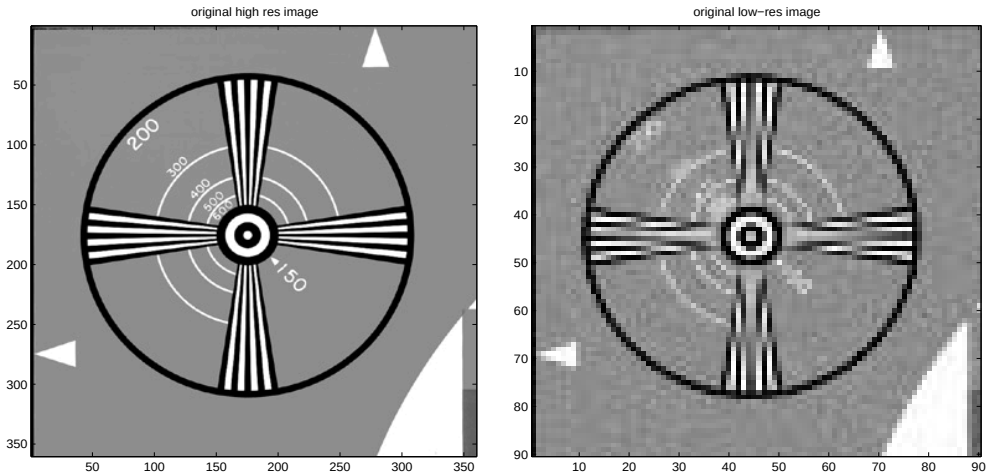
[25] D. Kondur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Mag.*, vol. 13, pp. 43–64, May 1996.

[26] N. Nguyen, P. Milanfar, and G. H. Golub, "Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement," *IEEE Trans. Image Processing*, vol. 10, no. 9, pp. 1299–1308, Sept. 2001.

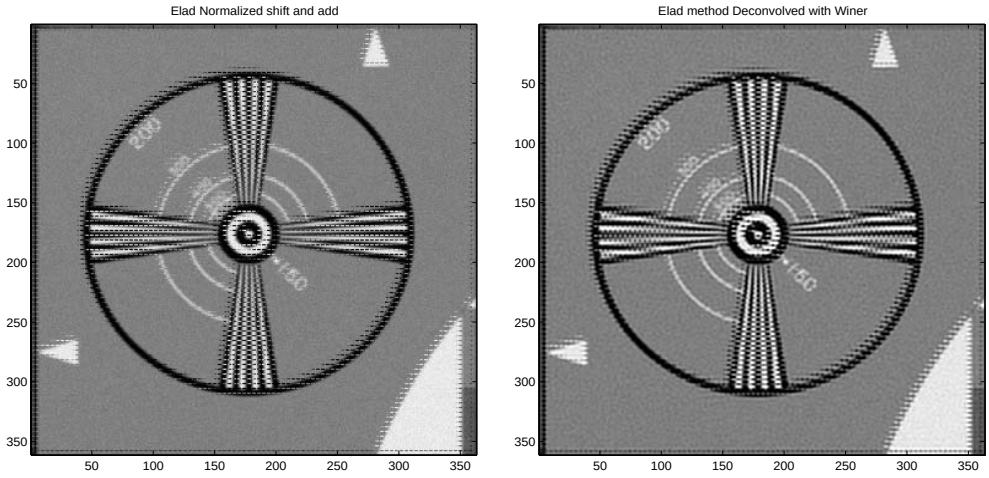
[27] M. Robertson and R. Stevenson, "DCT quantization noise in compressed images," in *IEEE Int. Conf. on Image Processing*, vol. 1, Oct. 2001, pp. 185–1888.

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

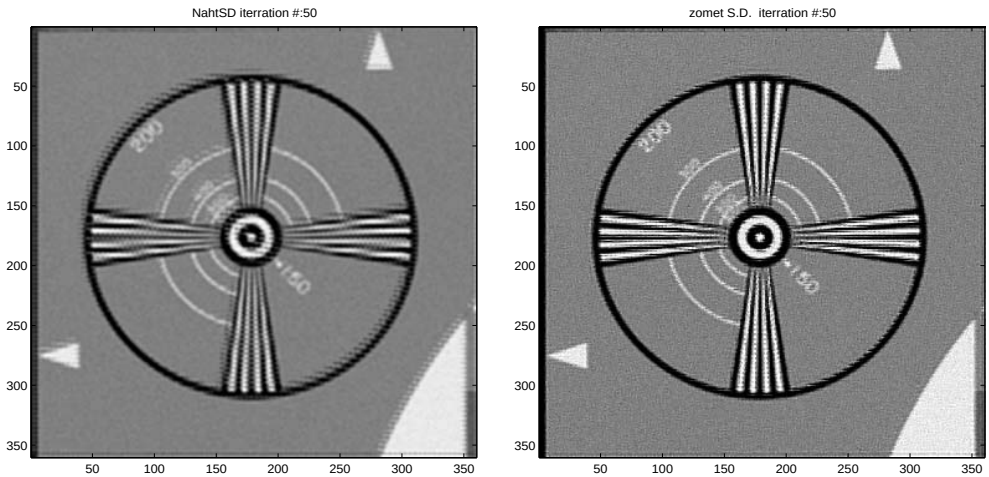
[29] M. Elad, "On the bilateral filter and ways to improve it," *IEEE Trans. Image Processing*, vol. 11, no. 10, pp. 1141–1151, Oct. 2002.




a: Original HR Frame b: LR Frame

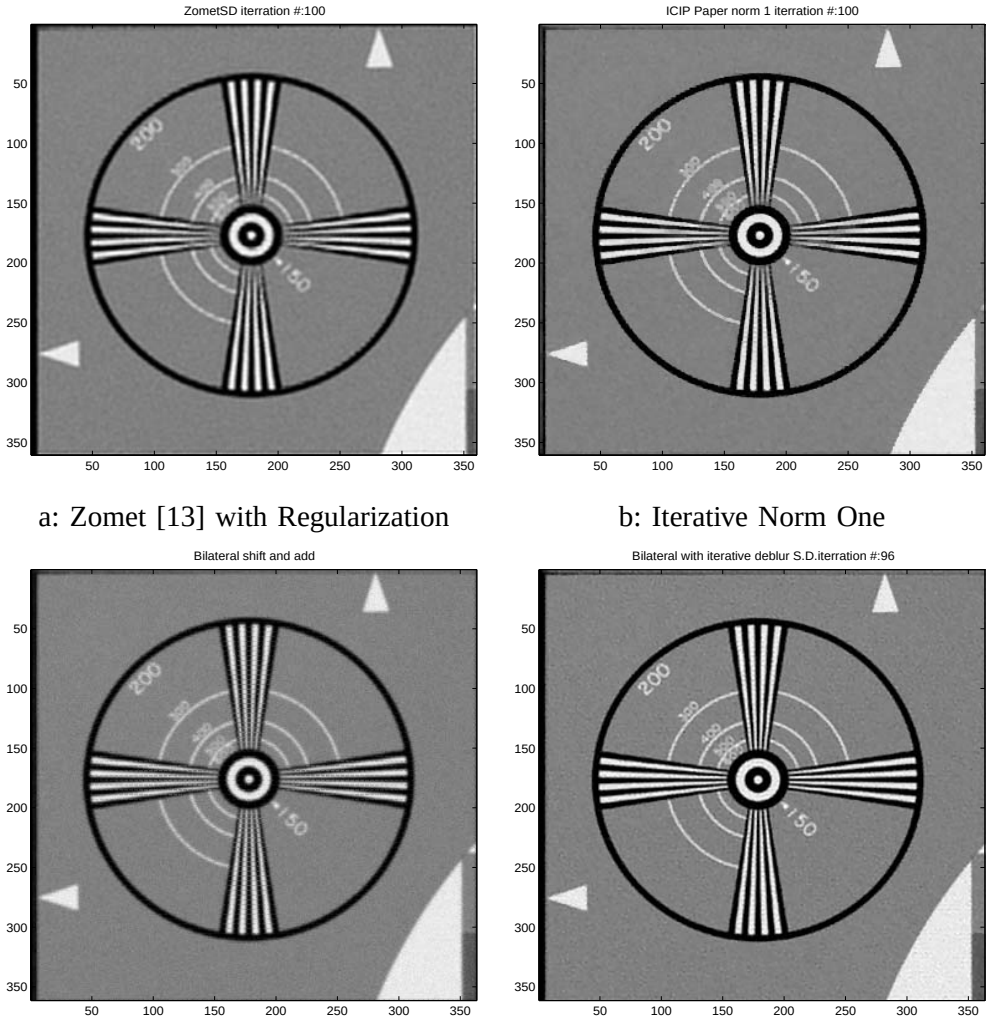



c: Shift and Add Result [5] d: Deconvolved Shift and Add [5]




e: Norm 2 Reconstruction f: Zomet Method [13]

Fig. 9. Simulation results of different resolution enhancement methods are applied to the Figure (a).



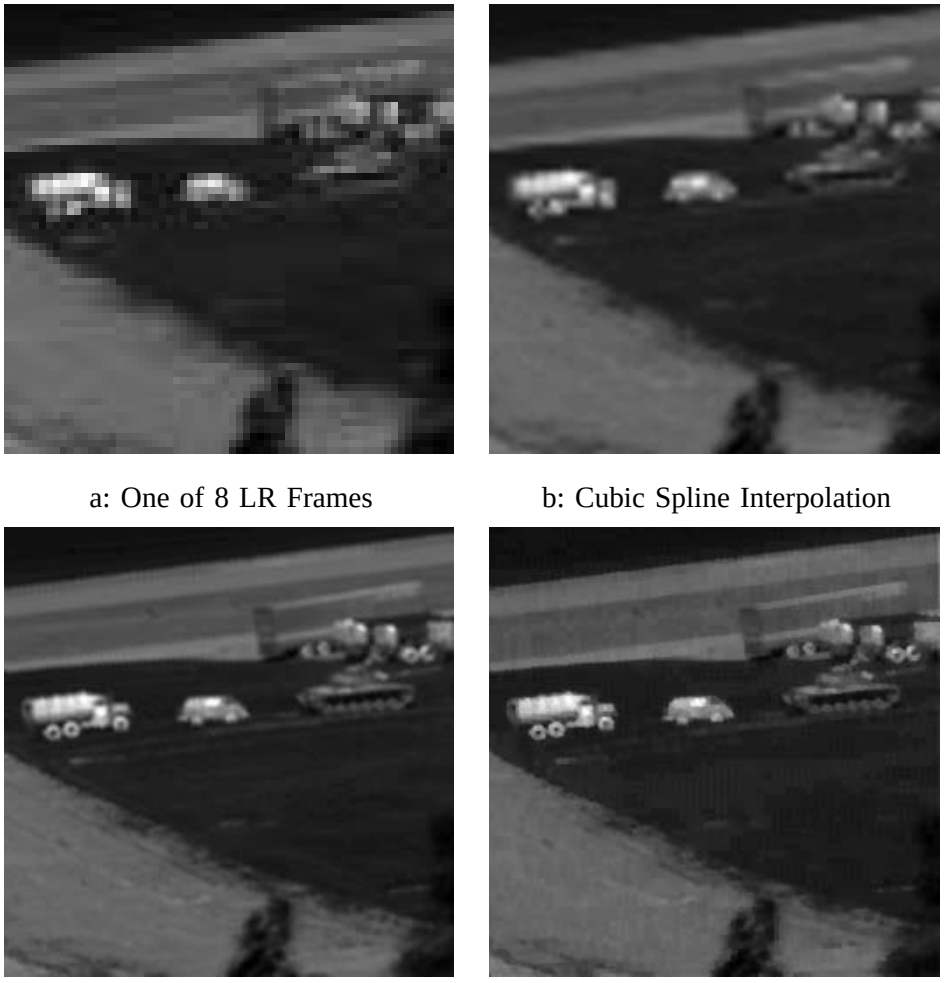

a: Zomet [13] with Regularization

b: Iterative Norm One

c: Bilateral Reconstruction

d: Iterative Deconvolved Bilateral

Fig. 10. Simulation results of different resolution enhancement methods are applied to the Figure (a).



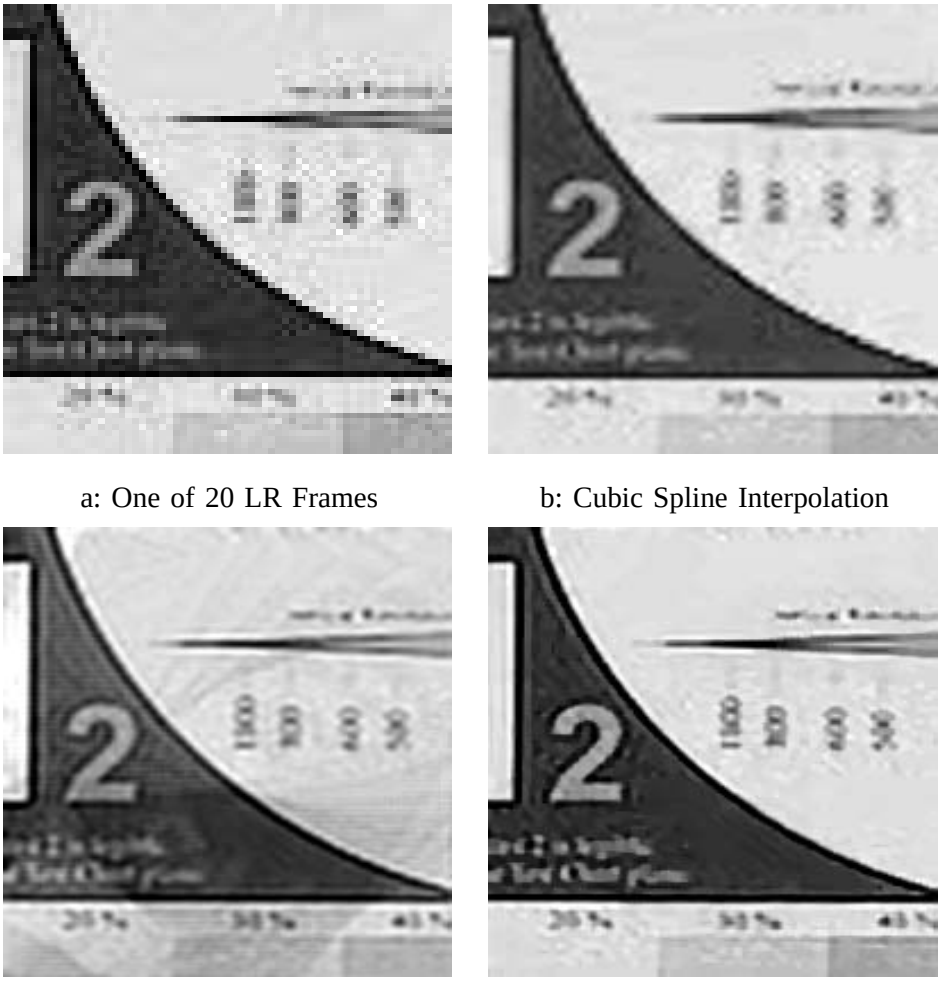
a: One of 8 LR Frames

b: Cubic Spline Interpolation

c: Iterative Norm Two

d: Iterative Norm One

Fig. 11. Results of different resolution enhancement methods applied to Tank sequence.



a: One of 20 LR Frames

b: Cubic Spline Interpolation

c: Iterative Norm Two

d: Iterative Norm One

Fig. 12. Results of different resolution enhancement methods applied to ADYORON test sequence.