Robust Shift and Add Approach to Super-Resolution

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

[a]Department of Electrical Engineering, University of California, Santa Cruz CA. 95064 USA.
[b] Department of Computer Science (SCCM), Stanford University, Stanford CA. 94305-9025 USA.

ABSTRACT

In the last two decades, many papers have been published, proposing a variety of methods for multi-frame resolution enhancement. These methods, which have a wide range of complexity, memory and time requirements, are usually very sensitive to their assumed model of data and noise, often limiting their utility. Different implementations of the non-iterative Shift and Add concept have been proposed as very fast and effective super-resolution algorithms. The paper of Elad & Hel-Or 2001 provided an adequate mathematical justification for the Shift and Add method for the simple case of an additive Gaussian noise model. In this paper we prove that additive Gaussian distribution is not a proper model for super-resolution noise. Specifically, we show that L_p norm minimization ($1 \leq p \leq 2$) results in a pixelwise weighted mean algorithm which requires the least possible amount of computation time and memory and produces a maximum likelihood solution. We also justify the use of a robust prior information term based on bilateral filter idea. Finally, for the underdetermined case, where the number of non-redundant low-resolution frames are less than square of the resolution enhancement factor, we propose a method for detection and removal of outlier pixels. Our experiments using commercial digital cameras show that our proposed super-resolution method provides significant improvements in both accuracy and efficiency.

Keywords: Super-Resolution, Robust Estimation, Robust Regularization, Shift and Add, Outlier Detection

1. INTRODUCTION

Theoretical and practical limitations usually constrain the achievable resolution of any imaging device. Super-resolution is the process of combining a sequence of low-resolution noisy blurred images to produce a higher resolution image or sequence.

Two common matrix notations are used to formulate the super-resolution problem in the pixel domain. The more popular notation[1–4] considers only camera lens/CCD blur and is defined as:

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

where $[r^2M^2 \times r^2M^2]$ matrix F_k is the geometric motion operator between the high-resolution frame $\underline{X}$ (of size $[r^2M^2 \times 1]$) and the k^{th} low-resolution 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 low-resolution frames.

Considering only atmosphere and motion blur an alternate matrix formulation of the super-resolution problem is presented as[5]

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

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). In this paper we use the model (1). Note that, under some assumptions which will be discussed in Section 3, blur and motion matrices commute and the general super-resolution formulation can be written as:

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

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

Defining $H_k = H_k^{cam} H_k^{atm}$ merges both models into a form similar to (1).

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. The performance of frequency domain approaches to super-resolution problem[6,7] significantly degrades with small deviations from the assumed translational motion model[8]. The iterative spatial domain super-resolution methods[1,3,9–11] are usually time consuming and their mathematical justification is only valid for the additive Gaussian noise model. Furthermore, regularization is either not implemented or it is limited to Tikhonov regularization. Considering outliers, Ref.[4] describes a very successful iterative robust super-resolution method, but lacks the proper mathematical justification. To reduce the computational cost, some papers[2,5,12,13] have broken the super-resolution problem to three separate steps: non-iterative image fusion (a.k.a. shift and add) step followed by interpolation and deblurring steps (usually iterative). The very important step of image fusion is mathematically defined and justified only for additive Gaussian noise[2].

In what follows in this paper, we first show the inadequacy of the Gaussian noise model for super-resolution problem (Section 2). Then following Ref.[2], we mathematically justify a non-iterative image fusion method for more general noise models (Section 3). In Section 4 we introduce a robust regularization term to help us interpolate and deblur the shift and add result. A non-iterative outlier detection method is introduced in Section 5. Finally simulation results and conclusive remarks are given in Sections 7 and 8, respectively.

2. ERROR MODELLING

The Maximum Likelihood (ML) estimators of the high-resolution image developed in many previous works[1–3,14] are valid when the noise distribution follows the Gaussian model. Unfortunately, Gaussian noise assumption is not valid for many real world image sequences. To appreciate this claim we set up the following experiments. In these experiments according to the model in (3) a high-resolution $[256 \times 256]$ image was shifted, blurred, and downsampled to create 16 low-resolution images (of size $[64 \times 64]$). The effect of readout noise of CCD pixels, was simulated by adding Gaussian noise to these low-resolution frames achieving SNR [†] equal to 25dB. We considered three common sources of error (outliers) in super-resolution system:

1. Error in motion estimation.
2. Inconsistent pixels: effect of an object which is only present in a few low-resolution frames (e.g. the effects of a flying bird in a static scene).
3. Salt and Pepper noise.

In the first experiment, to simulate the effect of error in motion estimation, a bias equal to $\frac{1}{4}$ of a pixel was intentionally added to the known motion vector of only one of the low-resolution frames. In the second experiment, a $[10 \times 10]$ block of only one of the images was replaced by a block from another data sequence. And finally in the third experiment we added salt and pepper noise to approximately 1% of the pixels of only one of the low-resolution frames. We used the GLRT test (Appendix A) to compare the goodness of fit of Laplacian and Gaussian distributions for modelling the noise in these three sets of low-resolution images. The GLRT test results for these three experiments were 0.6084, 0.6272 and 0.6081, respectively. The test result for the original low-resolution images contaminated only with pure Gaussian noise was 0.7991. Based on the criterion in (27), the distribution of the noise with a test result smaller than 0.7602 is better modelled by the Laplacian model rather than the Gaussian model. Note that the outlier contamination in these tests was fairly small, and more outlier contamination (larger error in motion estimation, larger blocks of inconsistence pixels, and higher percentage of Salt and Pepper noise) results in even smaller GLRT test results.

3. DATA FUSION USING SHIFT AND ADD

In the previous section we showed that in many cases, Laplacian model is a better candidate for describing the super-resolution noise than the Gaussian model. The high-resolution ML estimate of a set of low-resolution

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

images contaminated with additive generalized Gaussian distribution (GGD) noise model[‡] can be calculated from the following L_p norm minimization criterion:

$$\widehat{\underline{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[\sum_{k=1}^{N} \|D_k H_k F_k \underline{X} - \underline{Y}_k\|_p^p \right] \qquad 1 \leq p \leq 2 \tag{4}$$

Considering translational motion and with reasonable assumptions such as common space-invariant PSF, and similar decimation factor for all low-resolution 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, when $1 \leq p \leq 2$[§].

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, (4) may be rewritten as:

$$\widehat{\underline{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[\sum_{k=1}^{N} \|D F_k H \underline{X} - \underline{Y}_k\|_p^p \right] \tag{5}$$

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

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

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 Ref.[2] 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 $\widehat{\underline{Z}}$, we substitute $H\underline{X}$ with $\underline{Z}$:

$$\widehat{\underline{Z}} = \underset{\underline{Z}}{\text{ArgMin}} \left[\sum_{k=1}^{N} \|D F_k \underline{Z} - \underline{Y}_k\|_p^p \right] \tag{6}$$

The gradient of the cost in (6) is:

$$\underline{G}_p = \frac{\partial}{\partial \underline{Z}} \left[\sum_{k=1}^{N} \|D F_k \underline{Z} - \underline{Y}_k\|_p^p \right] = \sum_{k=1}^{N} F_k^T D^T sign(D F_k \underline{Z} - \underline{Y}_k) \odot |D F_k \underline{Z} - \underline{Y}_k|^{p-1} \tag{7}$$

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

The vector $\widehat{\underline{Z}}$ which minimizes the criterion (6) will be the solution to $\underline{G}_p = \underline{0}$. There is a simple interpretation for the solution: The vector $\widehat{\underline{Z}}$ 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

$$\underline{G}_2 = \sum_{k=1}^{N} F_k^T D^T (D F_k \underline{Z}_n - \underline{Y}_k) \tag{8}$$

which is proved in Ref.[2] to be the pixelwise average of measurements. If $p = 1$ then the gradient term will be:

$$\underline{G}_1 = \sum_{k=1}^{N} F_k^T D^T sign(D F_k \widehat{\underline{Z}} - \underline{Y}_k) = \underline{0} \tag{9}$$

[‡] $P_{GGD}(v, p, \alpha) = \frac{p}{2\alpha\Gamma(\frac{1}{p})} \exp^{-(\frac{|v|}{\alpha})^p}$, where $\Gamma(.)$ is the gamma function. Note that with $p = 1$ and $p = 2$, Laplacian and Gaussian noise models are the special cases of GGD, respectively.

[§] We did not consider the cases for which $p < 1$ as such assumption results in non-convex minimization, which is outside the scope of this paper.

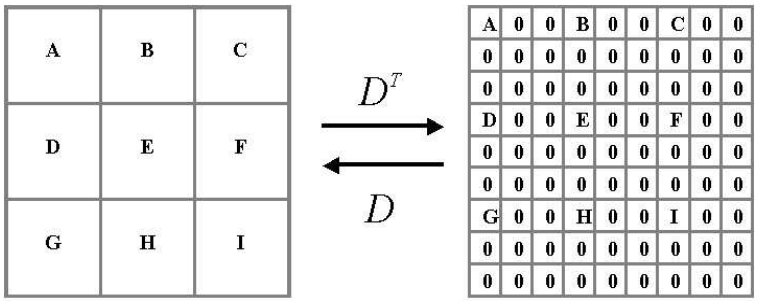


Figure 1. Effect of upsampling D^T matrix on a 3×3 image and downsampling matrix D on the corresponding 9×9 upsampled image (resolution enhancement factor of three). In this figure, to give better intuition the image vectors are reshaped as matrices.

We note that $F_k^T D^T$ copies the values from the low-resolution grid to the high-resolution grid after proper shifting and zero filling, and DF_k copies a selected set of pixels in high-resolution grid back on the low-resolution grid (Figure 1 illustrates the effect of upsampling and downsampling matrices D^T, and D). Neither of these two operations changes the pixel values. Therefore, each element of $\underline{G}_1$, which corresponds to one element in $\widehat{\underline{Z}}$, is the aggregate of the effects of all low-resolution 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 $\widehat{\underline{Z}}$ was larger than the corresponding contributing pixel from frame $\underline{Y}_k$.
3. Addition of -1, which means a pixel in $\widehat{\underline{Z}}$ 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 low-resolution 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 (7), 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. For the rest of this paper, we use the most robust member of this family of estimators (L_1 norm), and its median interpretation for estimating the high-resolution image.

4. REGULARIZATION

Note that for the under-determined cases ($N < r^2$) not all $\widehat{\underline{Z}}$ 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 Ref.[5], interpolation and deblurring are done simultaneously. As deblurring-interpolation is an ill-posed problem, it requires regularization which compensates the missing measurement information with some general prior information about the desirable high-resolution solution, and is usually implemented as a penalty factor. Based on the spirit of Total Variation (TV) criterion[15] and a related technique called the bilateral filter[16, 17] , we introduce our robust regularizer called Bilateral-TV, which is computationally cheap to implement, and preserves edges. The regularizing function looks like,

$$\Upsilon_{BTV}(X) = \sum_{l=0}^{P}\sum_{m=0}^{P} \alpha^{m+l} \|\underline{X} - S_x^l S_y^m \underline{X}\|_1 \tag{10}$$

where matrices (operators) S_x^l, and S_y^k shift $\mathbf{X}$ by l, and k pixels in horizontal and vertical directions respectively, presenting several scales of derivatives. The scalar weight α, $0 < \alpha < 1$, is applied to give a spatially decaying effect to the summation of the regularization 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 (10) results in:

$$\Upsilon_{BTV}(X) = \|Q_x \underline{X}\|_1 + \|Q_y \underline{X}\|_1 \tag{11}$$

which is suggested in Ref.[18] as a reliable and computationally efficient approximation to the Total-Variation prior.

To compare the performance of Bilateral-TV ($P \geq 1$) to common TV prior ($P = 1$) and Tikhonov prior[19] we set up the following denoising experiment. We added Gaussian white noise of mean zero and variance 0.045 to the image in Figure 2(a) resulting in the noisy image of Figure 2(b). If $\underline{X}$ and $\underline{Y}$ represent the original and corrupted images then we minimized:

$$\widehat{\underline{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[\|\underline{Y} - \underline{X}\|_2^2 + \lambda \Upsilon(\underline{X}) \right] \tag{12}$$

to reconstruct the noisy image, where λ, the regularization parameter, is a scalar for properly weighting the first term (similarity cost) against the second term (regularization cost). Tikhonov denoising resulted in Figure 2(c) ¶ The result of using TV prior ($P = 1$) for denoising is shown in Figure 2(d). Figure 2(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 2(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 2(e) is much less than Figure 2(d). Figure 2(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.

5. DEBLURRING-INTERPOLATION

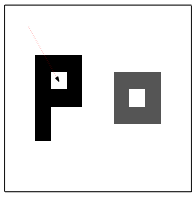
Based on the material that was developed in Sections 3 and 4, the following expression formulates our minimization criterion for obtaining $\widehat{\underline{X}}$ from $\widehat{\underline{Z}}$.

$$\widehat{\underline{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[\|A(H\underline{X} - \widehat{\underline{Z}})\|_1 + \lambda \sum_{l=0}^{P} \sum_{m=0}^{P} \alpha^{m+l} \|\underline{X} - S_x^l S_y^m \underline{X}\|_1 \right] \tag{14}$$

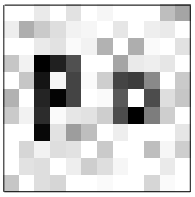
where matrix A is a diagonal matrix with diagonal values equal to the square root of the number of measurements that contributed to make each element of $\widehat{\underline{Z}}$ (in the square case ($N = r^2$), A is the identity matrix). So, the undefined pixels of $\widehat{\underline{Z}}$ have no effect on the high-resolution estimate $\widehat{\underline{X}}$. On the other hand, those pixels of $\widehat{\underline{Z}}$ which have been produced from numerous measurements, have a stronger effect in the estimation of the high-resolution frame $\widehat{\underline{X}}$.

¶ $\Upsilon_T(\underline{X}) = \|\Gamma \underline{X}\|_2^2$ where where Γ is usually a high-pass operator such as derivative, Laplacian, or even identity matrix. For this example Γ was replaced by the Laplacian kernel:

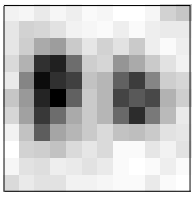
$$\Gamma = \frac{1}{8} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{13}$$



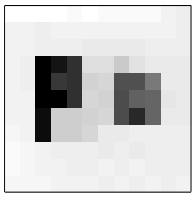
a: Original



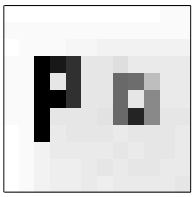
b: Noisy



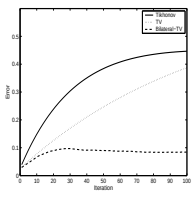
c:Reconstruction using Tikhonov



d:Reconstruction using TV



e:Reconstruction using Bilateral-TV



f:Estimation Error plot

Figure 2. 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.

As A is a diagonal matrix, $A^T = A$, and the corresponding steepest descent solution of minimization problem (14) can be expressed as:

$$\underline{\widehat{X}}_{n+1}=\underline{\widehat{X}}_n-\beta\left\{H^TA^Tsign(AH\underline{\widehat{X}}_n-A\underline{\widehat{Z}})\right.$$
$$\left.+\lambda\sum_{l=0}^{P}\sum_{m=0}^{P}\alpha^{m+l}[I-S_y^{-m}S_x^{-l}]sign(\underline{\widehat{X}}_n-S_x^lS_y^m\underline{\widehat{X}}_n)\right\} \quad (15)$$

where β is a scalar defining the step size in the direction of the gradient. Physical construction of matrices A, H, and S is not necessary as they can be implemented as direct image operators such as masking, blurring, and shifting operators, respectively. Note that decimation and warping matrices (D and F) and summation of measurements are not present in (15), which makes the implementation of our method much faster than the iterative methods suggested in Ref. [1, 3, 4].

6. BILATERAL NON ITERATIVE ARTIFACT REMOVAL

The non-iterative data fusion step presented in Section 3 works well for over-determined cases (where $N \gg r^2$). In the square or underdetermined cases, there is only one measurement available for each high-resolution 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 and fail to remove outliers. The outliers may be removed by the regularization term in the deblurring-interpolation step, but this needs a relatively large regularization factor λ, 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} low-resolution frame is calculated as:

$$w_k(i,j) = \sum_{m=-q}^{q} \sum_{n=-q}^{q} e^{-\frac{1}{2}\left(\frac{|\mathbf{Y}_k(i,j) - \widehat{\mathbf{Z}}(i'+m,j'+n)|}{\sigma_r}\right)} \times e^{-\frac{1}{2}\frac{m^2+n^2}{\sigma_d^2}} \tag{16}$$

where $\mathbf{Y}_k(i,j)$ is the gray level value of pixel (i,j) in the k^{th} low-resolution frame, (i',j') is the coordinates of pixel (i,j) mapped to the high-resolution grid, $\widehat{\mathbf{Z}}(i',j')$ is the gray level value of pixel (i',j') in the data fused image discussed in Section 5, q controls the bilateral kernel size, σ_r and σ_d control corresponding photometric and distance spread. Note that the values of $\widehat{\mathbf{Z}}(i'+m,j'+n)$ which are not defined in the shift and add step should not be considered in (16).

If each low-resolution frame is divided into R blocks then the overall weight of block r of the kth frame is:

$$W_{k,r} = \sum_{i,j \in \Omega_{k,r}} w_k(i,j) \tag{17}$$

where $\Omega_{k,r}$ defines the pixels in the rth block of the kth low-resolution frame.

In the next step, the median and variance of all corresponding blocks are computed and the blocks with weights smaller than $\text{MEDIAN}(W_{k,r}) - \tau \times \text{VAR}(W_{k,r})$ will be removed from the input data set (constant τ 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 (15) may be used to iteratively calculate $\underline{\widehat{X}}$.

7. EXPERIMENTS

In this section we compare the performance of the resolution enhancement algorithm proposed in this paper to the robust resolution enhancement method proposed in Ref.[4]. We used an Olympus C-4000 digital camera to capture 26 low-resolution images from a scene. One of these low-resolution images is shown in Figure 3. Figure 3(b) shows the cubic spline interpolation of Figure 3(a) by factor of four in each axis. The (unknown) camera PSF was assumed to be a 6×6 Gaussian kernel with standard deviation equal to two. We used the method described in Ref. [20] to compute the motion vectors. The robust super-resolution method which was proposed in Ref.[4] resulted in Figure 3(c)[‖]. The proposed method of this paper resulted in Figure 3(d)[**]. The run time of our method on a Pentium-III desktop computer is 18 seconds versus 48 seconds for the method of Ref.[4]. Note that increment of the number of low-resolution images has no significant effect on the runtime of our method as the time consuming deblurring-interpolation step is independent of the number of input images, but the run time of Ref.[4] has a direct relation with the number of input images [††].

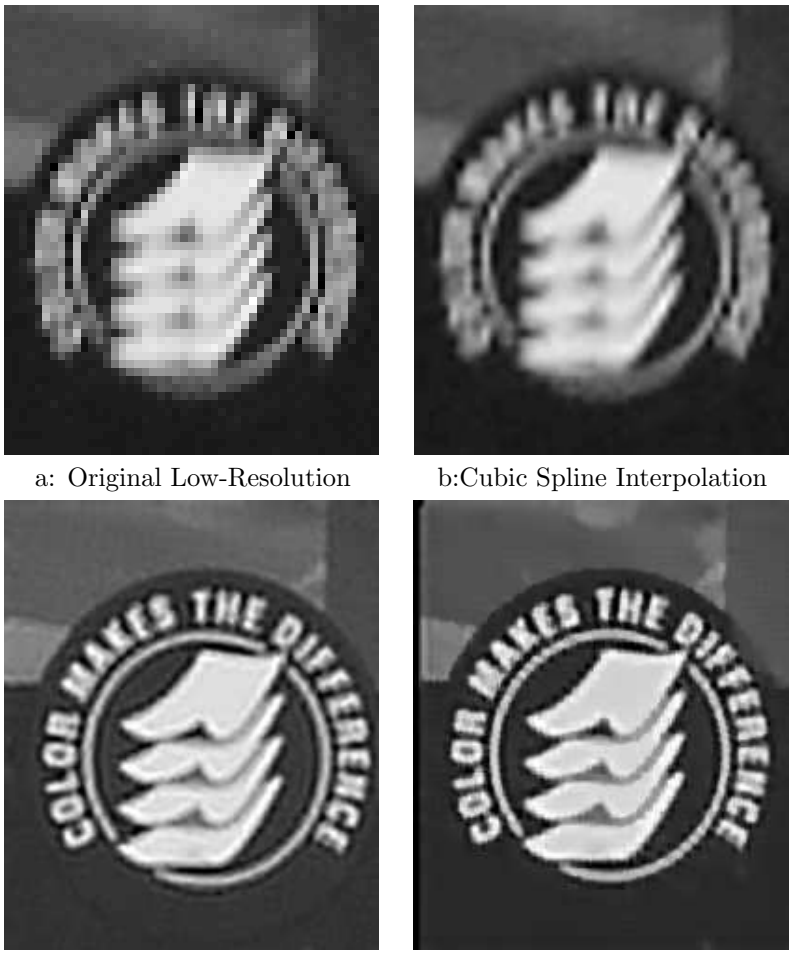
8. CONCLUSIONS AND FUTURE WORK

In this paper, we presented an algorithm to enhance the quality of a set of noisy blurred images and produce a high-resolution 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. The proposed method is fast and easy to implement as we have 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 low-resolution 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. We showed that our method removes outliers efficiently, resulting in images with sharp

[‖]To get the best results from this method we added a Tikhonov regularization term to the original method described in Ref.[4]. For this experiment the best result was attained in 10 iterations of steepest descent, using a step size equal to 20. The regularization factor (λ) was equal to 0.1.

[**]For this experiment the best result was attained in 30 iterations of steepest descent with the following parameters: $\lambda = .1$, $P = 2$, $\alpha = 0.7$, $q = 2$, $\sigma_r = \sigma_d = 2$, and $\tau = 1$. Ω was chosen to be the size of each low-resolution frame.

[††]Limitations of the method proposed in Ref.[4] plus more detailed experiments are presented in Ref.[21] available in: "http://www.soe.ucsc.edu/ milanfar/publications.htm".



a: Original Low-Resolution b:Cubic Spline Interpolation

c:Robust Ref.[4] d:Proposed Method

Figure 3. Experiment results for a set of real world images. Reconstruction of (c)[4] took about 48 seconds versus the proposed method which took about 18 seconds.

edges. The proposed method is suitable for real time super-resolution implementation as it is computationally inexpensive and memory efficient. The time consuming deblurring-interpolation step can be computed with parallel processors, which significantly increases the overall speed of implementation.

APPENDIX A. NOISE MODELLING BASED ON GLRT TEST

In this appendix we explain our approach of deciding which statistical model better describes the probability density function (PDF) of the noise. Gaussian and Laplacian distributions, the two major candidates for modelling the noise PDF, are defined as:

$$P_G(\underline{V}) = \frac{1}{(2\pi\sigma_G^2)^{N/2}} \exp(-\frac{\sum_{i=1}^{N}(\underline{V}(i) - m_G)^2}{2\sigma_G^2}) \tag{18}$$

$$P_L(\underline{V}) = \frac{1}{(2\sigma_L)^N} \exp(-\frac{\sum_{i=1}^{N}|\underline{V}(i) - m_L|}{\sigma_L}) \tag{19}$$

where $\underline{V}(i)$ is the i^{th} element of the noise vector $\underline{V}$ (of size $[1 \times N]$) and σ_G, m_G are the unknown parameters of the Gaussian PDF (P_G) and σ_L, m_L are the unknown parameters of the of the Laplacian PDF (P_L) which are

estimated from data. Noting logarithm is a monotonic function and

$$\ln P_L(\underline{V}) = -N\ln 2 - N\ln\sigma_L - \frac{\sum_{i=1}^{N}|\underline{V}(i)-m_L|}{\sigma_L} \tag{20}$$

then the ML estimates of σ_L and m_L are calculated as

$$\widehat{\sigma}_L, \widehat{m}_L = \underset{\sigma_L, m_L}{\text{ArgMax}}(P_L(\underline{V})) = \underset{\sigma_L, m_L}{\text{ArgMax}}(\ln P_L(\underline{V})) \tag{21}$$

so

$$\frac{\partial \ln P_L(\underline{V})}{\partial m_L} = \sum_{i=1}^{N}|\underline{V}(i)-m_L| = 0 \Longrightarrow \widehat{m}_L = \text{MEDIAN}(\underline{V}) \tag{22}$$

and

$$\frac{\partial \ln P_L(\underline{V})}{\partial \sigma_L} = -\frac{N}{\sigma_L} + \frac{\sum_{i=1}^{N}|\underline{V}(i)-m_L|}{\sigma_L^2} = 0 \Longrightarrow \widehat{\sigma}_L = \frac{\sum_{i=1}^{N}|\underline{V}(i)-\widehat{m}_L|}{N} \tag{23}$$

The same scheme can be used to estimate the Gaussian model parameters as:

$$\widehat{m}_G = \text{MEAN}(\underline{V}) \quad \text{and} \quad \widehat{\delta}_G = \sqrt{\frac{\sum_{i=1}^{N}(\underline{V}(i)-\widehat{m}_G)^2}{N}} \tag{24}$$

We use the generalized likelihood ratio test (GLRT)[22] to decide between the two hypothesis about the noise model:

$$\frac{P_G(\underline{V};\widehat{\sigma}_G,\widehat{m}_G)}{P_L(\underline{V};\widehat{\sigma}_L,\widehat{m})_L} > \gamma \tag{25}$$

where γ is the decision threshold, that is if the ratio in (25) is larger than γ then P_G is a more accurate PDF model for $\underline{V}$ than P_L and vice versa (γ was chosen equal to 1 as it mimicks a test which minimizes the probability of error and does not a priori favor either hypothesis). So:

$$\frac{\frac{1}{(2\pi\widehat{\sigma}_G^2)^{N/2}}\exp(-\frac{\sum_{i=1}^{N}(\underline{V}(i)-\widehat{m}_G)^2}{2\widehat{\sigma}_G^2})}{\frac{1}{(2\widehat{\sigma}_L)^N}\exp(-\frac{\sum_{i=1}^{N}|\underline{V}(i)-\widehat{m}_L|}{\widehat{\sigma}_L})} > 1 \tag{26}$$

Substituting $\widehat{m}_G$, $\widehat{\sigma}_L$, $\widehat{\sigma}_G$, and $\widehat{\sigma}_L$ with their corresponding estimates from (22), (23), and (24) and simplifying results in:

$$\frac{\widehat{\sigma}_L}{\widehat{\sigma}_G} > (\frac{\pi}{2e})^{\frac{1}{2}} \simeq 0.7602 \tag{27}$$

So if (27) is valid for a certain vector $\underline{V}$ then Gaussian model is a better estimate of PDF than Laplacian model and vice versa.

ACKNOWLEDGMENTS

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. Sina Farsiu would like to thank Morteza Shahram, for his very useful suggestions about the GLRT test.

REFERENCES

1. M. Elad and A. Feuer, "Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images," *IEEE Trans. Image Processing* **6**, pp. 1646–1658, Dec. 1997.
2. M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space invariant blur," *IEEE Trans. Image Processing* **10**, pp. 1187–1193, Aug. 2001.
3. N. Nguyen, P. Milanfar, and G. H. Golub, "A computationally efficient image superresolution algorithm," *IEEE Trans. Image Processing* **10**, pp. 573–583, Apr. 2001.
4. A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super resolution," in *Proceedings of the Int. Conf. on Computer Vision and Patern Recognition (CVPR)*, **1**, pp. 645–650, Dec. 2001.
5. S. Lertrattanapanich and N. K. Bose, "High resolution image formation from low resolution frames using delaunay triangulation," *IEEE Trans. Image Processing* **11**, pp. 1427–1441, Dec. 2002.
6. T. S. Huang and R. Y. Tsai, "Multi-frame image restoration and registration," *Advances in computer vision and Image Processing* **1**, pp. 317–339, 1984.
7. N. K. Bose, H. C. Kim, and H. M. Valenzuela, "Recurcive 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)*, **5**, pp. 269–272, Apr. 1993.
8. S. Borman and R. L. Stevenson, "Super-resolution from image sequences - a review," in *Proceedings of the 1998 Midwest Symposium on Circuits and Systems*, **5**, Apr. 1998.
9. S. Peleg, D. Karen, and L. Schweitzer, "Improving image resolution using subpixel motion," *CVGIP:Graph. Models Image Processing* **54**, pp. 181–186, March 1992.
10. M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP:Graph. Models Image Process* **53**, pp. 231–239, 1991.
11. H. Ur and D. Gross, "Improved resolution from sub-pixel shifted pictures," *CVGIP:Graph. Models Image Processing* **54**, Mar. 1992.
12. L. Teodosio and W. Bender, "Salient video stills: Content and context preserved," in *Proceedings of First ACM International Conference on Multimedia*, **10**, pp. 39–46, Aug. 1993.
13. M. C. Chiang and T. E. Boulte, "Efficient super-resolution via image warping," *Image and Vision Computing* **18**, pp. 761–771, July 2000.
14. D. Capel and A. Zisserman, "Super-resolution enhancement of text image sequences," in *Proceedings of International Conference on Pattern Recognition*, pp. 600–605, 2000.
15. L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D* **60**, pp. 259–268, Nov. 1992.
16. C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of IEEE Int. Conf. on Computer Vision*, pp. 836–846, Jan. 1998.
17. M. Elad, "On the bilateral filter and ways to improve it," *IEEE Trans. Image Processing* **11**, pp. 1141–1151, Oct. 2002.
18. Y. Li and F. Santosa, "A computational algorithm for minimizing total variation in image restoration," *IEEE Trans. Image Processing* **5**, pp. 987–995, June 1996.
19. A. Bovik, *Handbook of image and video processing.* Academic Press Limited, New Jersey, 2000.
20. J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierachical model-based motion estimation," *Proceedings European Conference on Computer Vision* , pp. 237–252, 1992.
21. S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super-resolution," *submitted to IEEE Trans. Image Processing* , June 2003.
22. S. M. Kay, *Fundamentals of statistical signal processing:detection theory*, vol. II, Prentice-Hall, New Jersey, 1998.