FAST AND ROBUST SUPER-RESOLUTION

Sina Farsiu [1], *Dirk Robinson* [2], *Michael Elad* [3], *Peyman Milanfar* [4]

[1,2,4] Electrical Engineering Department, University of California Santa Cruz
Email: {farsiu, dirkr, milanfar}@ee.ucsc.edu
[3] Computer Science Department, SCCM Program, Stanford University
Email: elad@sccm.stanford.edu

1. ABSTRACT

In the last two decades, many papers have been published, proposing a variety methods of 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 a different implementation using L_1 norm minimization and robust regularization to deal with different data and noise models. This computationally inexpensive method is robust to errors in motion and blur estimation, and results in sharp edges. Simulation results confirm the effectiveness of our method and demonstrate its superiority to other robust super-resolution methods.

2. INTRODUCTION

Theoretical and practical limitations usually constrain the achievable resolution of any imaging device. Super-resolution techniques are employed to combine a sequence of low-resolution frames from a scene and produce a higher resolution picture or sequence. Many different iterative methods have been proposed in the last two decades (e.g. [1], [2] and [3]). In [4], [5] and [6] non-iterative methods have been addressed, assuming correct estimations of motion and blur are available. These methods have a wide range of complexity, memory and time requirements. None of the above methods addressed noise models other than Gaussian additive noise, and regularization was either not implemented or it was limited to Tikhonov regularization. Considering outliers, [7] has proposed a successful robust method, without proper mathematical justification.

In this paper, we will use the L_1 norm, both for the regularization and the error 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 in the fused images. Moreover, following [8], we use a robust regularizer, related to Total-Variation [9], called the bilateral filter, to get better performance. We will show that our method's performance is superior to what was proposed earlier in [7], with the same or better convergence speed.

This paper is organized as follows: Section 3 defines our model , 3.1 introduces formulation for the general case, and 3.2 justifies a very fast and effective robust super-resolution method for special models of data and noise. Simulations are presented in Section 4, and Section 5 concludes this paper.

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 CO-OPERATIVE AGREEMENT NO. AST - 9876783 .

3. ROBUST SUPER-RESOLUTION

Based on the number of available low-resolution frames, the accuracy of estimated motion or noise model, several data structures are possible. In this section, we propose L_1 norm minimization based solutions to deal with situations in which the confidence in motion estimation results is not high or noise model is not pure additive Gaussian. We will show that L_1 norm minimization is a superior method in dealing with outliers, and produces outstanding results even when there are no outliers present, and the noise is restricted to pure additive Gaussian.

We use the super-resolution notation of [2]:

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

where F_k is the geometric warp operator between the high-resolution frame X and k^{th} low-resolution frame Y_k which are rearranged in lexicographic order ($\mathbf{X}$ and $\mathbf{Y_k}$ present their matrix form). The camera's point spread function (PSF), is modelled by the blur matrix H_k, and D_k represents the decimation operator. V_k is the additive noise and N is the number of available low-resolution frames.

3.1. General Formulation

In this subsection, we explain an iterative solution for the super-resolution problem. As we desire to produce a high-resolution result which is similar to low-resolution frames, we need a cost function which compares the similarity of low and high resolution frames. We choose L_1 norm cost function instead of the common L_2 norm cost function, as including the effects of outliers in this comparison is not desirable. In the underdetermined super-resolution cases ($N < l^2$ in which N is the number of non-redundant low-resolution frames and l is the resolution enhancement factor), we need to add a regularization term for efficiently calculating missing data (i.e. interpolation). Regularization is a useful tool even in the square and overdetermined cases ($N = l^2$ and $N > l^2$ respectively) as generally super-resolution is an ill-conditioned problem. The following expression formulates our minimization criteria:

$$\widehat{\underline{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[\sum_{k=1}^{N} \|D_k H_k F_k \underline{X} - \underline{Y}_k\|_1 + \lambda \sum_{l=0}^{P} \sum_{m=0}^{P} \alpha^{m+l} \|\underline{X} - S_x^l S_y^m \underline{X}\|_1 \right] \quad (2)$$

λ is a scalar for properly weighting the first term (similarity cost) against the second term (regularization cost). S_x^l is the operator corresponding to shifting $\mathbf{X}$ by l pixels in horizontal direction and operator S_y^k shifts $\mathbf{X}$ by k pixels in vertical direction, presenting several scales of derivatives. Scalar weight α, $0 < \alpha < 1$, is applied to give a spatially decaying effect to the summation of the regularization term. The steepest descent solution of this minimization problem is:

$$\widehat{\underline{X}}_{n+1} = \widehat{\underline{X}}_n - \beta \left[\sum_{k=1}^{N} F_k^T H_k^T D_k^T sign(D_k H_k F_k \widehat{\underline{X}}_n - \underline{Y}_k) + \lambda \sum_{l=0}^{P} \sum_{m=0}^{P} \alpha^{m+l} [I - S_y^{-m} S_x^{-l}] sign(\widehat{\underline{X}}_n - S_x^l S_y^m \widehat{\underline{X}}_n) \right] \quad (3)$$

where β is the step size, 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. 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 $m = 0, l = 1$ 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 (2) results in:

$$\widehat{\underline{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[\sum_{k=1}^{N} \|D_k H_k F_k \underline{X} - \underline{Y}_k\|_1 + \lambda (\|Q_x \underline{X}\|_1 + \|Q_y \underline{X}\|_1) \right] \quad (4)$$

which is very close in spirit to the Total-Variation prior [10].

Simulation results in Section 4 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. Simulating the effects of these matrices as a sequence of operators spares us from explicitly constructing them. This property helps our method to be implemented in an extremely fast and memory efficient way.

3.2. Special Case: No Regularization

In this section we propose a faster version of the method described in 3.1 for dealing with square or overdetermined cases when no regularization is considered in the data fusion process. We show that in the square case, the performance of L_1 and L_2 norm minimizations (without regularization) are exactly the same, and in the most realistic situations of overdetermined case, L_1 norm minimization has better performance than L_2 norm minimization.

When the additive noise is pure Gaussian, the ML optimum solution results from the following minimization formula [2]:

$$\widehat{\underline{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[\sum_{k=1}^{N} \|D_k H_k F_k \underline{X} - \underline{Y}_k\|_2^2 \right] \quad (5)$$

Considering translational motion and with reasonable assumptions such as space-invariant PSF, and similar decimation factor for all low-resolution frames (i.e. $\forall k \quad H_k = H \quad \& \quad D_k = D$), [5] showed that (5) can be solved in two separate steps, speeding up the implementation:

1. Pixelwise averaging the low-resolution frames after proper zero filling and warping (shift and add step).
2. Deblurring the resulting shifted and added picture (deconvolution step).

In this section, considering the same assumptions as [5], we will justify a similar two step implementation for L_1 norm minimization and we will also show that in special cases L_1 and L_2 norms converge to the same answer. Given

$$\widehat{\underline{X}} = \underset{\underline{X}}{\text{ArgMin}} \left[\sum_{k=1}^{N} \|D H F_k \underline{X} - \underline{Y}_k\|_1 \right] \quad (6)$$

the steepest descent solution will be:

$$\widehat{\underline{X}}_{n+1} = \widehat{\underline{X}}_n - \beta \left[\sum_{k=1}^{N} F_k^T H^T D^T sign(D H F_k \widehat{\underline{X}}_n - \underline{Y}_k) \right]$$

H and F_k are block circulant matrices which commute ($F_k H = H F_k$ and $F_k^T H^T = H^T F_k^T$), therefore:

$$\widehat{\underline{X}}_{n+1} = \widehat{\underline{X}}_n - \beta \left[H^T \sum_{k=1}^{N} F_k^T D^T sign(D F_k H \widehat{\underline{X}}_n - \underline{Y}_k) \right]$$

Multiplying both sides by H we obtain:

$$H\underline{\widehat{X}}_{n+1} = H\underline{\widehat{X}}_n - \beta \left[HH^T \sum_{k=1}^{N} F_k^T D^T sign(DF_k H \underline{\widehat{X}}_n - \underline{Y}_k) \right]$$

Defining $\underline{Z}_n = H\underline{\widehat{X}}_n$:

$$\underline{Z}_{n+1} = \underline{Z}_n - \beta \left[HH^T \sum_{k=1}^{N} F_k^T D^T sign(DF_k \underline{Z}_n - \underline{Y}_k) \right]$$

As HH^T is a positive semi-definite matrix it has no effect on the steady state solution $\underline{Z}_\infty$, which is reached when the gradient term $\underline{G} = \sum_{k=1}^{N} F_k^T D^T sign(DF_k \underline{Z}_n - \underline{Y}_k)$ tends to zero. There is a simple interpretation for the steady state solution. $\underline{Z}_\infty$ is nothing but the pixelwise median of all measurements after proper zero filling and motion compensation. To appreciate this fact, we note that $F_k^T D^T$ copies the values of low resolution grid to high resolution grid after proper shifting and zero filling, and DF_k copies a selected set of pixels in high resolution grid to low-resolution grid. None of these two operators changes the pixel values. Therefore, each element of $\underline{G}$, which corresponds to one element in $\underline{Z}_\infty$, is the aggregate of the effect of each low-resolution frame. Each effect 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 $\underline{Z}_\infty$ was larger than the corresponding pixel in frame $\underline{Y}_k$.

3. Addition of -1, which means a pixel in $\underline{Z}_\infty$ was smaller than the corresponding pixel in frame $\underline{Y}_k$.

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

In the square case, 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.

It is interesting to compare L_1 and L_2 norm performances for overdetermined cases where multiple measurements for each high resolution pixel is available. In an unrealistic situation in which the only noise in the system is pure additive Gaussian and motion estimation is very accurate, L_2 norm performance is slightly better than L_1 norm as averaging effectively reduces the noise variance. As noise variance increases, the accuracy of motion estimation decreases, resulting in the appearance of systematic motion artifacts in the L_2 norm minimization results. But as the median operator is robust to outliers, these motion artifacts do not have a significant effect on L_1 norm minimization results.

We should note that the outliers are not limited to motion estimation artifacts. In many situations, it is desirable to remove an object which is only present in a few low-resolution frames (e.g. removing the effects of a flying bird from a static scene). For these type of outliers, L_1 norm minimization is clearly more adequate than L_2 norm minimization.

4. SIMULATIONS

In this section we compare the performance of the proposed algorithm to other methods. We present a simulated example, in which the original frame was blurred and downsampled by a factor of three. The motion vectors for two of the nine low-resolution frames were intentionally computed erroneously.

Figure 1 is one of the input low-resolution frames. The result of implementing the method described in [5] is shown in Figure 2. The robust super-resolution method which was proposed in [7] resulted in Figure 3. Figure 4 shows the implementation of the proposed method described in section 3.1. Comparing Figure 4 to Figures 2 and 3, we notice not only our method has removed the outliers more efficiently than other methods, but also it has resulted in sharper edges without any ringing effects.

5. CONCLUSION

In this paper we presented a robust super-resolution method based on L_1 norm used both in the regularization and the measurement term in our penalty function. We showed that our method removes outliers efficiently, resulting in images with sharp edges. The proposed method was fast and easy to implement. We also proposed a very fast method for the square and overdetermined cases and mathematically justified a pixelwise shift and add (median) method and related it to L_1 norm minimization.

6. REFERENCES

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

[2] M. Elad and A. Feuer, "Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images," *IEEE Transactions*

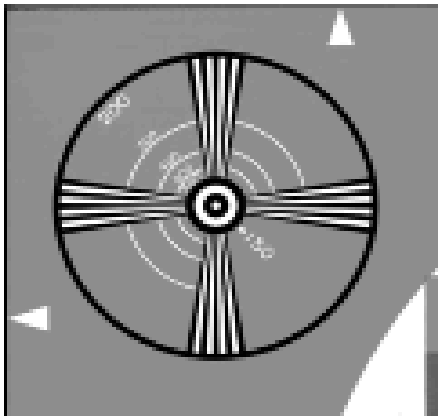

Fig. 1. Original low-resolution picture

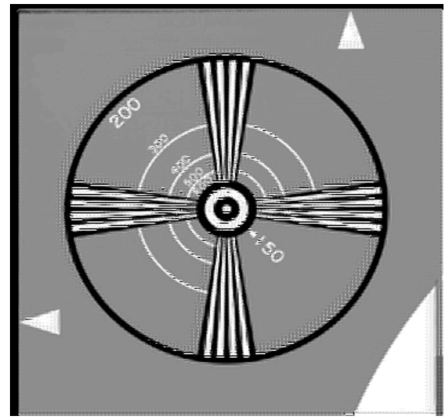

Fig. 3. Zomet method [7] reconstruction

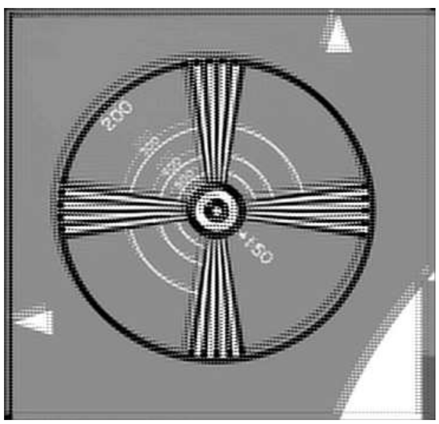

Fig. 2. Deblurred Shift and Add [5] reconstruction

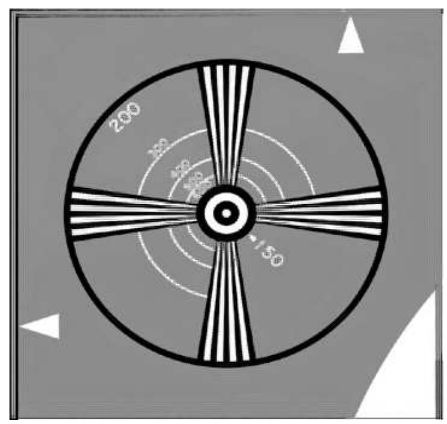

Fig. 4. Proposed method reconstruction

on Image Processing, vol. 6, no. 12, pp. 1646–1658, Dec. 1997.

[3] N. Nguyen, P. Milanfar, and G. Golub, “A computationally efficient image superresolution algorithm,” *IEEE Transactions on Image Processing*, vol. 10, no. 4, pp. 573–583, Apr. 2001.

[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 Transactions on Image Processing*, vol. 10, no. 8, pp. 1187–1193, Aug. 2001.

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

[7] 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.

[8] M. Elad, “On the bilateral filter and ways to improve it,” *IEEE Transactions on Image Processing*, vol. 11, no. 10, pp. 1141–1151, Oct. 2002.

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

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