Comparison of digital image resolution enhancement techniques

Lisimachos P. Kondi[a], Dean A. Scribner[b] and Jonathon M. Schuler[b]

[a]Dept. of Electrical Engineering, SUNY at Buffalo, Buffalo, NY 14260
[b]Optical Sciences Division, U.S. Naval Research Laboratory, Washington, DC 20375

ABSTRACT

In this paper, we compare two resolution enhancement techniques. The first technique is based on Maximum A Posteriori (MAP) estimation while the second technique performs Temporal Accumulation of Registered Image Data (TARID) followed by Wiener filtering. Both techniques and described and the merits of each one are discussed. Experimental results are presented and conclusions are drawn.

Keywords: Resolution enhancement, superresolution, MAP estimation, image restoration

1. INTRODUCTION

In many imaging systems, the optical image of the lens is projected onto a Focal Plane Array (FPA) of photo-detectors. The job of the FPA is to photo-convert and sample the optical image to create a digital image sequence. However, in many cases, the array of detectors is not sufficiently dense to accurately sample the scene (meet the Nyquist criterion). In addition, the capturing process introduces noise. Thus, the video produced by the imaging system has lower resolution than what is theoretically achievable by the optic. Also, the point spread function of the lens and the effects of the finite size of the photo-detectors further degrade the acquired video frames. *Resolution enhancement* techniques combine multiple video frames to create a single composite image of higher resolution, less noise and less alias distortion than the original data.

Resolution enhancement using multiple frames is possible when there exists subpixel motion between the captured frames. Thus, each of the frames provides a unique look into the scene. An example scenario is the case of a camera that is mounted on an aircraft and is imaging objects in the far field. The vibrations of the aircraft will generally provide the necessary motion between the focal plane array and the scene, thus yielding frames with subpixel motion between them and minimal occlusion effects.

The rest of the paper is organized as follows. In section 2, some of the previous work on resolution enhancement is outlined. In section 3, a MAP-based resolution enhancement technique is discussed, while in section 4 the Temporal Accumulation of Registered Image Data (TARID) method is outlined. In section 5, experimental results are presented. Finally, in section 6, conclusions are drawn.

2. PREVIOUS WORK ON RESOLUTION ENHANCEMENT

The problem of video resolution enhancement is an active research area. A comprehensive review of the literature can be found in Ref. 1. We next outline a few of the approaches that have appeared in the literature. Among the earliest efforts in the field is the work by Tsai and Huang.[2] Their method operates in the frequency domain and capitalizes on the shifting property of the Fourier Transform, the aliasing relationship between the Continuous Fourier Transform (CFD) and the Discrete Fourier Transform (DFT), and the fact that the original scene is assumed to be bandlimited. The above properties are used to construct a system of equations relating the aliased DCT coefficients of the observed images to samples of the CFT of the unknown high resolution image. The system of equations is solved, yielding an estimate of the DFT coefficients of the original high resolution image, which can then be obtained using inverse DFT. This technique was further improved by Tekalp et. al. in Ref. 3 by taking into account a Linear Shift Invariant (LSI) blur Point Spread Function (PSF) and using a least

Correspondence Email: lkondi@eng.buffalo.edu

squares approach to solving the system of equations. The big advantage of the frequency domain methods is their low computational complexity. However, these methods are applicable only to global motion and a priori information about the high resolution image cannot be exploited.

Most of the other resolution enhancement techniques that have appeared in the literature operate in the spatial domain. The most computationally efficient techniques involve interpolation of non-uniformly spaced samples. This requires the computation of the optical flow between the acquired frames with sub-pixel accuracy. Then, the samples of the acquired low-resolution frames are combined in order to create a high-resolution frame. Interpolation techniques are used to estimate pixels in the high-resolution frame that did not correspond to pixels in one of the acquired frames. Finally, image restoration techniques are used to compensate for the blurring introduced by the imaging device. A method based on this idea is the Temporal Accumulation of Registered Image Data (TARID).[4,5] The TARID method is further described in section 4.

Another method that has appeared in the literature is the iterated backprojection method.[6] In this method, the estimate of the high resolution image is updated by backprojecting the error between motion-compensated, blurred and subsampled versions of the current estimate of the high resolution image and the observed low resolution images, using an appropriate backprojection operator.

Another proposed method is the Projection Onto Convex Sets (POCS).[7,8] In these methods, the space of high resolution images is intersected with a set of convex constraint sets representing desirable image characteristics, such as positivity, bounded energy, fidelity to data, smoothness, etc.

A major class of resolution enhancement algorithms is based on stochastic techniques. Methods in this class include Maximum Likelihood (ML)[9] and Maximum A Posteriori (MAP) approaches.[10,11] We next describe our MAP-based implementation which is similar to that described in Ref. 10 and uses a Gaussian prior model.

3. MAP-BASED RESOLUTION ENHANCEMENT

In the following, we assume that all images are lexicographically ordered (i.e. rearranged into one-dimensional arrays). We assume that we observe p low resolution images of size $N_1 \times N_2$,
$\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_p$. Each $\mathbf{y}_i$ is a vector of length $M = N_1 \times N_2$. The full set of the p low resolution images is denoted by

$$\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_p^T] \tag{1}$$

where $\mathbf{y}$ is a vector of length pM.

We assume that the data $\mathbf{y}$ are observations of a high resolution image $\mathbf{z}$ of size $L_1 N_1 \times L_2 N_2$. The observation model is

$$\mathbf{y} = \mathbf{W_s z} + \mathbf{n} \tag{2}$$

where $\mathbf{W_s}$ performs blurring, motion compensation and subsampling. For the motion compensation, vector $\mathbf{s}$ is used. We assume global motion for the whole frame, i.e., every pixel in the frame has the same displacement. Thus, $\mathbf{s}$ contains two parameters for each frame. $\mathbf{n}$ is Additive White Gaussian Noise.

Our problem is to estimate the high resolution image $\mathbf{z}$ and the motion $\mathbf{s}$ given the observed data $\mathbf{y}$ and the observation model. The MAP estimator selects the $\mathbf{z}$ and $\mathbf{s}$ that maximize the conditional pdf of the quantities to be estimated ($\mathbf{z}$ and $\mathbf{s}$) given the observations ($\mathbf{y}$). Thus, the estimates can be computed as

$$\hat{\mathbf{z}}, \hat{\mathbf{s}} = \arg\max_{\mathbf{z},\mathbf{s}} Pr(\mathbf{z}, \mathbf{s}|\mathbf{y}). \tag{3}$$

Using Bayes rule, the above equation can be expressed as

$$\hat{\mathbf{z}}, \hat{\mathbf{s}} = \arg\max_{\mathbf{z},\mathbf{s}} \frac{Pr(\mathbf{y}|\mathbf{z}, \mathbf{s}) Pr(\mathbf{z}, \mathbf{s})}{Pr(\mathbf{y})}. \tag{4}$$

Since the denominator of the above equation is not a function of $\mathbf{z}$ or $\mathbf{s}$, it suffices to minimize the numerator. If we further assume that $\mathbf{z}$ and $\mathbf{s}$ are independent, the above equation can be written as

$$\hat{\mathbf{z}}, \hat{\mathbf{s}} = \arg\max_{\mathbf{z},\mathbf{s}} Pr(\mathbf{y}|\mathbf{z}, \mathbf{s}) Pr(\mathbf{z}) Pr(\mathbf{s}). \tag{5}$$

We assume that the image $\mathbf{z}$ has Gaussian distribution with pdf

$$Pr(\mathbf{z}) = \frac{1}{(2\pi)^{N/2}|C_z|^{1/2}} \exp\{-\frac{1}{2}\mathbf{z}^T C_z^{-1}\mathbf{z}\} \tag{6}$$

where C_z is the $N \times N$ covariance matrix of $\mathbf{z}$. In Ref. 10, the (i,j)th element of C_z^{-1} is defined by

$$C_{i,j}^{-1} = \frac{1}{\lambda}\sum_{r=1}^{N} d_{r,i} d_{r,j}. \tag{7}$$

where

$$d_{i,j} = \begin{cases} 1, & \text{for } i = j \\ -1/4, & \text{for } j : z_j \text{ is a cardinal neighbor of } z_i \end{cases} \tag{8}$$

This implies that the image is assumed to be smooth. The parameter λ controls our confidence in this prior model. The pdf $Pr(\mathbf{s})$ is not used in the minimization, i.e., an assumption is being made that all possible motion displacements are equally probable.

Using the observation model of (2) and the pdf of (6), it is possible to derive $Pr(\mathbf{y}|\mathbf{z},\mathbf{s})$.[10] After some manipulations, our problem reduces to minimizing the cost function

$$L(\mathbf{z},\mathbf{s}) = \frac{1}{2\sigma_\eta^2}(\mathbf{y} - \mathbf{W_s z})^T(\mathbf{y} - \mathbf{W_s z}) + \frac{1}{2}\mathbf{z}^T C_z^{-1}\mathbf{z}. \tag{9}$$

where σ_η^2 is the variance of the additive white noise $\mathbf{n}$.

If we express the above equation using the individual elements of the matrices and vectors, we get

$$L(\mathbf{z},\mathbf{s}) = \frac{1}{2\sigma_\eta^2}\sum_{m=1}^{pM}(y_m - \sum_{r=1}^{N} w_{m,r}(\mathbf{s}) z_r)^2 + \frac{1}{2\lambda}\sum_{i=1}^{N}(\sum_{j=1}^{N} d_{i,j} z_j)^2. \tag{10}$$

The above cost function is the sum of two terms. The first term represents the fidelity of the desired $\mathbf{z}$ to the received data and the second term represents fidelity to the prior model for $\mathbf{z}$. For the selection of pdf in Ref. 10, the second term represents convolution of $\mathbf{z}$ with a high pass filter followed by summing the squares of the pixels. The parameter λ controls the balance between these two terms. A large λ would give larger weight to the fidelity to the received data while a small λ would give larger weight to the prior model. In Ref. 10, no analytic method for selecting λ is provided. However, it can be seen that our cost function is very similar to that of Ref. 12 which uses a regularization approach. The parameter λ is inversely proportional to the regularization parameter in Ref. 12. Therefore, a technique for the estimation of the regularization parameter[13] could be used to estimate λ.

The above cost function can be minimized using the coordinate-descent method. This iterative method starts with an initial estimate of $\mathbf{z}$ obtained using interpolation from a low resolution frame. Then, for a fixed $\mathbf{z}$, the cost function is minimized with respect to $\mathbf{s}$. Thus, the motion of each frame is estimated. Then, for fixed $\mathbf{s}$, a new estimate for $\mathbf{z}$ is obtained. This procedure continues until convergence is reached, i.e., $\mathbf{z}$ and $\mathbf{s}$ are updated in a cyclic fashion.

At iteration n, the vector $\hat{\mathbf{s}}_k^n$ denoting the global motion of frame k is found as

$$\hat{\mathbf{s}}_k^n = \arg\min_{\mathbf{s}_k}\{\sum_{m=1}^{M}(y_{k,m} - \sum_{r=1}^{N} w_{k,m,r}(\mathbf{s}_k)\hat{z}_r^n)^2\} \tag{11}$$

for $k = 1, 2, \ldots, p$. The above equation means that to estimate the motion, we take the current estimate of the high resolution image $\hat{\mathbf{z}}$, and, given the observation model, find the global motion that would minimize the mean square error between the blurred, motion compensated and subsampled high resolution frame and the corresponding low resolution (observed) frame.

In order to update the estimate of $\hat{\mathbf{z}}$, we use the following equation:

$$\hat{z}_k^{n+1} = \hat{z}_k^n - \varepsilon^n g_k(\hat{\mathbf{z}}^n, \hat{\mathbf{s}}^n) \tag{12}$$

where $g_k(\hat{\mathbf{z}}^n, \hat{\mathbf{s}}^n)$ is the gradient of the cost function and is given by

$$\begin{aligned} g_k(\mathbf{z},\mathbf{s}) &= \frac{\partial L(\mathbf{z},\mathbf{s})}{\partial z_k} \\ &= \frac{1}{\sigma_\eta^2}\sum_{m=1}^{pM} w_{m,k}(\mathbf{s})\left(\sum_{r=1}^{N} w_{m,r}(\mathbf{s}) z_r - y_m\right) + \\ &\quad \frac{1}{\lambda}\sum_{i=1}^{N} d_{i,k}\left(\sum_{j=1}^{N} d_{i,j} z_j\right). \end{aligned} \tag{13}$$

4. THE TARID METHOD

The Temporal Accumulation of Registered Image Data (TARID) method[4, 5] consists of three steps:

- Motion estimation.
- Reconstruction of a composite image lattice.
- Restoration filtering

In any resolution enhancement technique, it is necessary to estimate the motion between the low resolution frames. Any motion estimation algorithm can be used with TARID. In Ref. 5, a pyramidal algorithm that uses a block-based uniform shift estimation method to find a local motion field based on a coarse-to-fine stragegy.[14] Then, the coarse mesh of of shift vectors generated by the pyramidal algorithm is fitted to an affine parametric model that accounts for both geometric image transformations as well as simple optical distortion.

Once the motion between the low resolution frames has been estimated, one frame is taken as a coordinate reference and the absolute motion is computed relative to this reference frame. Thus, each pixel in the low resolution image sequence is assigned a coordinate. Then, a high resolution image is created using a high-density image grid that spans the same range of coordinates as the reference image, but with higher sampling density than the reference image. A correspondence is found between each pixel of the low resolution images and a pixel in the high-density image grid. Thus, the grid is populated with the low resolution image pixels. Of course, it is likely that some grid locations will not be filled in this fashion. These locations are filled using interpolation from the neighboring pixel locations. Thus, a high resolution image is created.

The next and final step in the TARID algorithm is restoration filtering. This step is necessary in order to compensate for the pixel distortion and optical distortion present in the low-resolution images and and, consequently, in the high resolution composite image. In Ref. 5, a Wiener filter is used. Square pixel distortion and a Gaussian lens blur are assumed along with spectrally flat additive noise.

5. EXPERIMENTAL RESULTS

We next present experimental results using MAP estimation. The data used were from a Long Wave InfraRed (LWIR) camera and were supplied by the Naval Research Laboratory (NRL). Four low-resolution frames of size 256×256 were used. One of the frames is shown in Fig. 1. In Fig. 2, the result of the MAP-based resolution enhancement is presented. The enhanced image has a resolution of 512×512. For comparison purposes, Fig. 3 shows a bicubic interpolation of Fig. 1. We can see that the result of the MAP-based algorithm is of much better quality than the bicubic-interpolated image.

In order to be better able to evaluate the two methods presented in this paper, we applied both MAP-based image restoration and Wiener filtering on the same high resolution composite image. It can be seen that the MAP-restored image exhibits less pronounced restoration artifacts than the Wiener-filtered image.

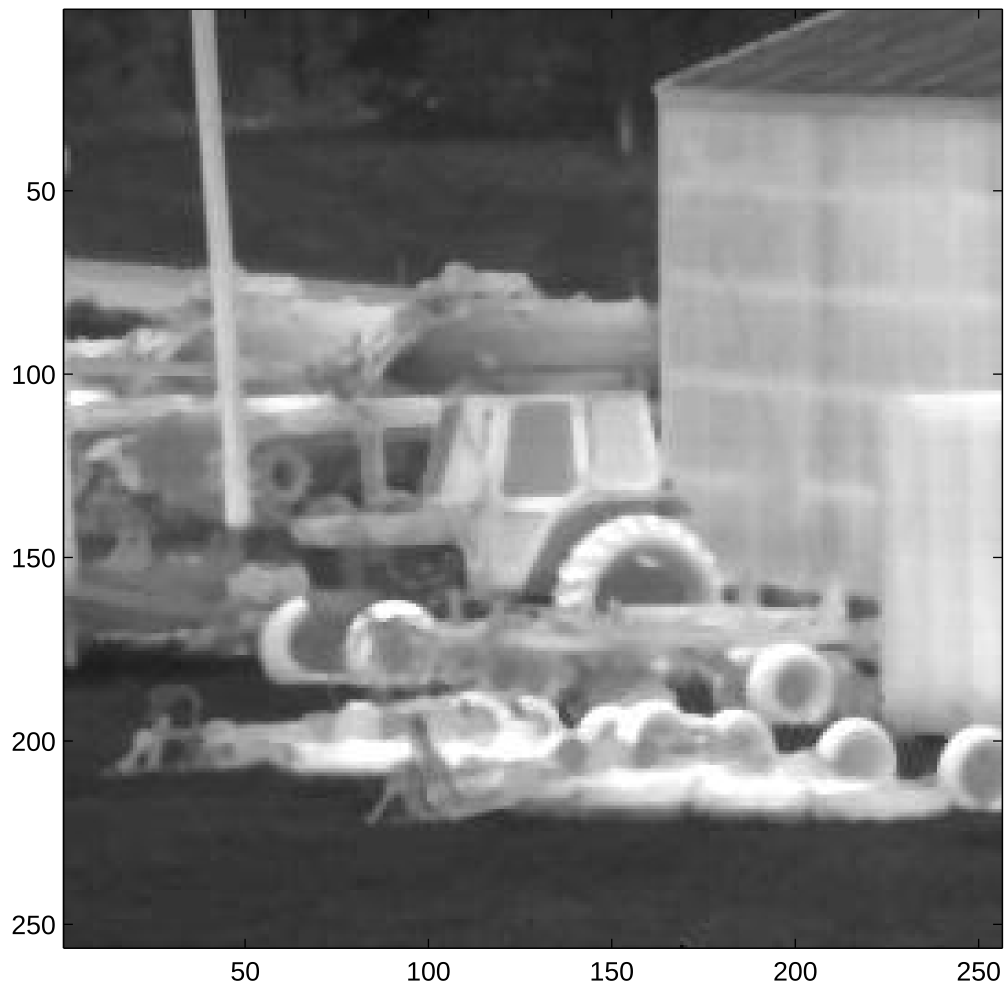

Figure 1: Original 256×256 frame.

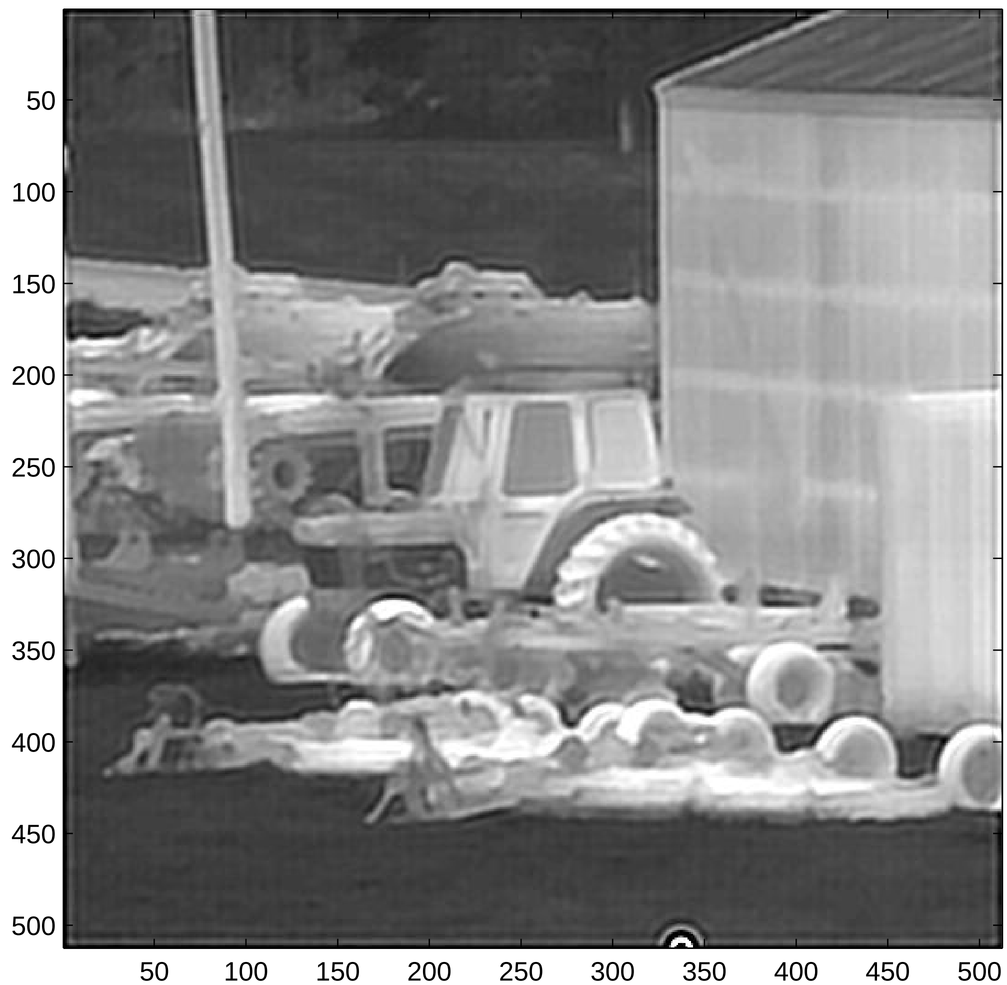

Figure 2: Result of the MAP resolution enhancement algorithm.

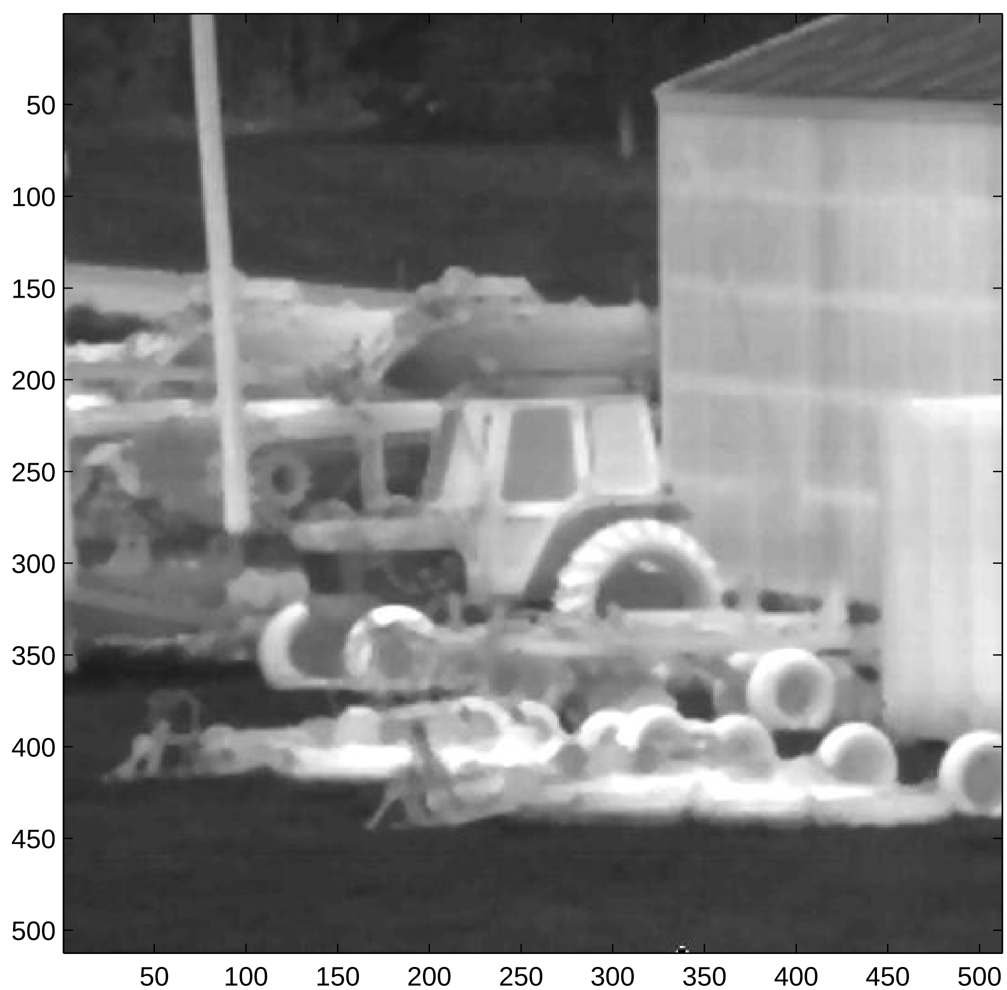

Figure 3: Bicubic interpolation of the low resolution image.

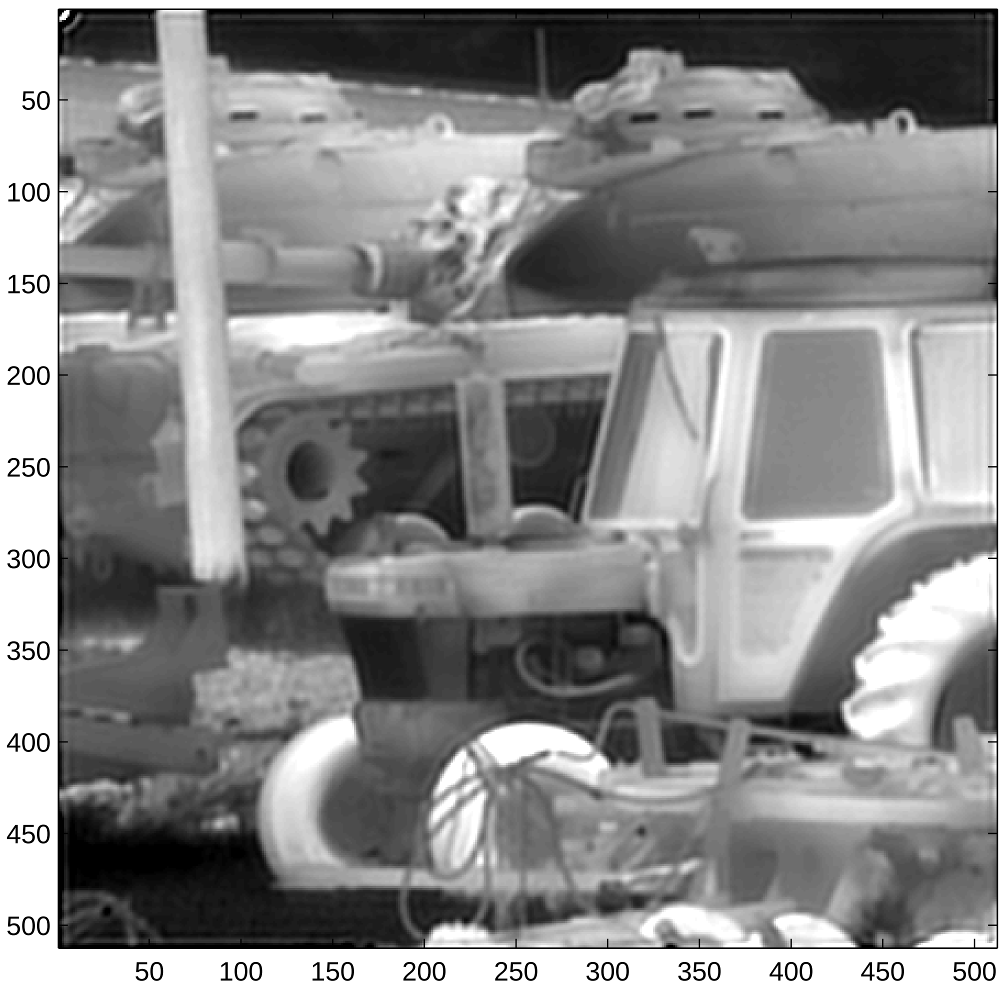

Figure 4: MAP restoration of the high-resolution composite image.

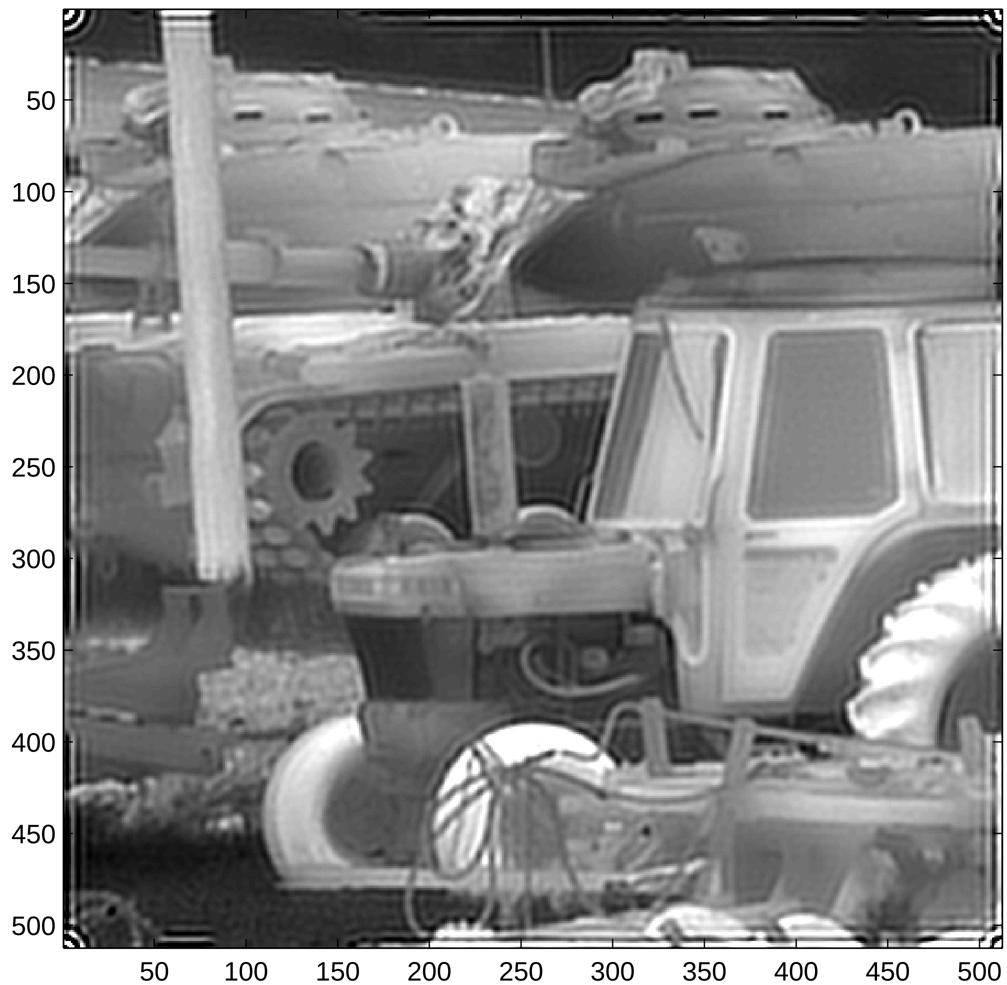

Figure 5: Wiener filtering result of the high-resolution composite image.

6. CONCLUSIONS

We have presented two different digital image resolution enhancement techniques. The MAP-based approach would be expected to yield an improvement in performance at the expense of computational complexity. As part of our future work in this area, we will conduct extensive evaluations in order to gauge the merits of different resolution enhancement algorithms. We also intend to determine the most appropriate prior pdf for the high resolution image.

REFERENCES

1. S. Borman and R. L. Stevenson, "Super-Resolution from Image Sequences - A Review," in *Proceedings of the 1998 Midwest Symposium on Circuits and Systems*, (Notre Dame, IN), 1998.
2. R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," in *Advances in Computer Vision and Image Processing*, R. Y. Tsai and T. S. Huang, eds., **1**, pp. 317–339, JAI Press Inc., 1984.
3. A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, "High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, **III**, pp. 169–172, (San Francisco, CA), 1992.
4. J. M. Schuler, D. A. Scribner, and M. R. Kruer, "Alias reduction and resolution enhancement by a temporal accumulation of registered data from focal plane array sensors," in *SPIE Visual Information Processing*, pp. 94–102, (Orlando, FL), 2000.
5. J. M. Schuler, J. G. Howard, P. R. Warren, and D. A. Scribner, "Resolution enhancement through TARID processing," (San Jose, CA), 2002.
6. M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion and transparency," *Journal of Visual Communications and Image Representation* **4**, pp. 324–335, Dec. 1993.
7. A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero apperture time," *IEEE Transactions on Image Processing* **6**, pp. 1064–1076, Aug. 1997.
8. P. E. Eren, M. I. Sezan, and A. M. Tekalp, "Robust, object-based high-resolution image reconstruction from low-resolution video," *IEEE Transactions on Image Processing* **6**, pp. 1446–1451, Oct. 1997.
9. B. C. Tom and A. K. Katsaggelos, "Reconstruction of a high resolution image from multiple degraded mis-registered low resolution images," in *Proceedings of the Conference on Visual Communications and Image P rocessing*, **2308**, pp. 971–981, (Chicago, IL), Sept. 1994.
10. R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Transactions on Image Processing* **6**, pp. 1621–1633, Dec. 1997.
11. R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Transactions on Image Processing* **5**, pp. 996–1011, June 1996.
12. B. C. Tom and A. K. Katsaggelos, "Resolution enhancement of monochrome and color video using motion compensation," *IEEE Transactions on Image Processing* **10**, pp. 278–287, Feb. 2001.
13. M. G. Kang and A. K. Katsaggelos, "Simultaneous multichannel image restoration and estimation of the regularization parameters," *IEEE Transactions on Image Processing* **6**, pp. 774–778, May 1997.
14. S. Rakshit and C. H. Anderson, "Computation of optical flow using basis functions," *IEEE Transactions on Image Processing* **6**, pp. 1246–1254, Sept. 1997.