Performance Evaluation of Optimization Methods for Super-resolution Mosaicking on UAS Surveillance Videos

Aldo Camargo*[a], Qiang He[b], K, Palaniappan[c]
[a]Research and Development at I+T, Lima 39, Lima, Peru ,
[b]Department of Mathematics, Computer and Information Sciences, Mississippi Valley State University, MVSU 7257
[c] Dept. of Computer Science, University of Missouri, Columbia, MO USA 65211

ABSTRACT

Unmanned Aircraft Systems (UAS) have been widely applied for reconnaissance and surveillance by exploiting the information collected from the digital imaging payload. However, the data analysis of UAS videos is frequently limited by motion blur; the frame-to-frame movement induced by aircraft roll, wind gusts, and less than ideal atmospheric conditions; and the noise inherent within the image sensors. Therefore, using super-resolution mosaicking of low-resolution UAS surveillance video frames, becomes a critical requirement for UAS video processing and important for further effective image understanding. In this paperwe develop a novel super-resolution framework which does not require the construction of sparse matrices. The proposed method applies image operators in the spatial domain and uses an iterated back-projection method to construct super-resolution mosaics from overlapping UAS surveillance video frames. The Steepest Descent method, Conjugate Gradient method and Levenberg-Marquardt algorithm are used to numerically solve the nonlinear optimization problem for estimating a super-resolution mosaic. A quantitative performance comparison in terms of computation time and visual quality of the super-resolution mosaic using the three numerical techniques is presented. The Levenberg-Marquardt algorithm provides a numerical solution for least squares curve fitting, which avoids the time-consuming computation of the inverse of the pseudo-Hessian matrix in regular singular value decomposition (SVD). The Levenberg-Marquardt method, interpolating between the Gauss–Newton algorithm (GNA) and the method of gradient descent, is efficient, robust, and easy to implement. The results obtained in our simulations shows a great improvement of the resolution of the low resolution mosaic of around 47 dB for synthetic images, and a considerable visual improvement in sharpness and visual details for real UAS surveillance frames. The convergence is rapid requiring typically ten iterations.

Keywords: Super-resolution, Conjugate Gradient Method, Steepest Descent Method, Levenberg Marquardt Algorithm, Ill-condition problems, Video Mosaicking

.

1. INTRODUCTION

Multi-frame super-resolution refers to the particular case where multiple images of a particular scene are available [2][3]. The idea is to use the low-resolution images containing motion, camera zoom, focus, and out of focus blur to recover extra data to reconstruct an image with a resolution above of limits of the camera. The super-resolved image should have more detail that any of the low resolution images. Mosaicking is the alignment or stitching of two or more images into a single composition representing a 3D scene [5][12]. Generally, the mosaics are used to create a map which is impossible to visualize with only one video frame.

Super-resolution mosaicking combines both methods, and it has number of applications when surveillance video from UAS or satellite is used. One clear application is the surveillance of certain areas even during night with the use of an infrared (IR) imaging system. The UAS can fly over areas of interest and generate super-resolved mosaics that can be analyzed at the ground control station. Other important applications involve the supervision of high voltage transmission lines, oil pipes, and the highway system. NASA also uses super-resolution mosaics to study the surface of Mars, the Moon, and other planets.

Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIII, edited by Gerald C. Holst, Keith A. Krapels,
Proc. of SPIE Vol. 8355, 83550Z · © 2012 SPIE · CCC code: 0277-786X/12/$18 · doi: 10.1117/12.919641

Super-resolution mosaicking has been studied by many researchers; Zomet and Peleg [6] use the overlapping area within a sequence of video frames to create a super-resolved mosaic. In this method, the SR reconstruction technique proposed in [7] is applied to a strip rather than a whole image. This means that the resolution of each strip is enhanced by the use of all the frames that contain that particular strip. The disadvantage is that this method is computationally expensive. Ready and Taylor [8] use a Kalman filter to compute the super-resolved mosaic. They add unobserved data to the mosaic using Dellaert's method. Basically, they construct a matrix that relates the observed pixels to estimated mixel values. This matrix is constructed using the homography matrix and the point spread function (PSF). The problem is that this matrix is extremely large, so they use a Kalman filter and diagonalization of the covariance matrix to reduce the amount of storage and computation required. The drawback of this algorithm is the use of the large matrix, and the best results with synthetic data obtain a PSNR of 31.6 dB. Simolic and Wiegand [9] use a method based on image warping. In this method, each pixel of each frame is mapped into the SR mosaic, and its gray level value is assigned to the corresponding pixel in the SR mosaic within a range of ± 0.2 pixel units. The drawback of this method is that it requires that the motion vectors and homography must be highly accurate, which is difficult when dealing with real surveillance video from UAS. Wang, Fevig, and Schultz [10] use the overlapped area within five consecutive frames from a video sequence. Then they use sparse matrices to model the relationship between the LR and SR frames, which is solved using maximum a posteriori estimation. To deal with the ill-posed problem of the super-resolution model, they use hybrid regularization. The drawback of this method is that it has to be used every five frames, which means that every five frames several sparse matrices has to be built. Therefore, this method does not seem to be appropriate to deal with a real video sequence which has thousand of frames. Pickering and Ye [1] proposed an interesting model for mosaicking and super-resolution of video sequences, using the Laplacian operator to find the regularization factor. The problem with the use of the Laplacian factor is that forces spatial smoothness. Therefore, noise and edge pixels are removed in the regularization process, eliminating sharp edges [11]. Arican and Frossard [17] use the Levenberg Marquardt algorithm to compute the SR of onmidirectional images. Chung [16] proposed different Gauss Newton methods to compute the SR of images, the disadvantage is that works only for small images.

Our method combines the ideas of most of these techniques, but it also inserts a different way to deal with the super-resolution mosaicking that does not require the construction of sparse matrices. Therefore, it is feasible to apply the algorithm to a relative large image sequence and obtain a video mosaic. Also, we use Huber regularization which preserves high frequency pixels, so sharp edges are preserved, and makes the super-resolution problem convex [4] that helps in the converge of our proposed algorithm.

2. MATHEMATICAL MODEL

2.1. Observation Model

Assuming that there are *K* frames of LR images available, the observation model can be represented as

$$\mathbf{y}_k = \mathbf{D}\mathbf{B}_k\mathbf{W}_k\mathbf{R}[\mathbf{x}]_k + \boldsymbol{\eta}_k = \mathbf{H}_k\mathbf{x} + \boldsymbol{\eta}_k \tag{1}$$

Here, y_k (k =1, 2, ..., K), x, and η_k represent the k^{th} LR image, the part of the real world depicted by the super-resolution mosaic, and the additive noise, respectively. The observation model in (1) introduces $R[x]_k$, which represents the reconstruction of the k^{th} warped SR image from the original high-resolution data x [1]. The geometric warp operator and the blur matrix between x and the k^{th} LR image, y_k are represented by W_k and B_k, respectively. To compute W_k we compute the homography by using the features obtained by SIFT and RANSAC. We used SIFT because its performance and accuracy [13]. The decimation operator is denoted by D. The estimation of the unknown SR mosaic image is not only based on the observed LR images, but also on many assumptions such as the blurring process and additive noise. The motion model is computed as a projective model using the homography between frames; the blur is considered only optical. The additive noise, η_k, is considered to be independent and identically distributed white Gaussian noise. Therefore, the problem of finding the maximum likelihood estimate of the SR mosaic image $\hat{x}$ can be formulated as

.

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\{ \left\| \sum_{k=1}^{K} \mathbf{y}_k - \mathbf{D}\mathbf{B}_k \mathbf{W}_k \mathbf{R}[\mathbf{x}]_k \right\|_2^2 \right\} \tag{2}$$

In this case, $\| \|_2$ denotes the Euclidean norm. As the SR reconstruction is an ill-posed inverse problem, we need to add another term for regularization, which must contain prior information for the SR mosaicking. This regularization term helps to convert the ill-posed problem into a well-posed problem. We use Huber regularization:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\{ \left\| \sum_{k=1}^{K} \mathbf{y}_k - \mathbf{D}\mathbf{B}_k \mathbf{W}_k \mathbf{R}[\mathbf{x}]_k \right\|_2^2 + \lambda \sum_{g \in \mathbf{G}_x} \rho(g, \alpha) \right\} \tag{3}$$

The Huber function is defined as

$$\rho(x,\alpha) = \begin{cases} x^2, & \text{if} \quad |x| \le \alpha, \\ 2\alpha|x| - \alpha^2, & \text{otherwise.} \end{cases} \tag{4}$$

2.2. Super-resolution Mosaicking Using Steepest Descent

Based on the gradient descent algorithm for minimizing (3), the robust iterative update for $\hat{\mathbf{x}}$ can be expressed as

$$\hat{\mathbf{x}}^{(n+1)} = \hat{\mathbf{x}}^{(n)} + \alpha^{(n)} \left\{ \mathbf{R}^T \left[\mathbf{W}_k^T \mathbf{B}_k^T \mathbf{D}^T (\mathbf{y}_k - \mathbf{D}\mathbf{B}_k \mathbf{W}_k \mathbf{R}[\hat{\mathbf{x}}^{(n)}]_k \right]_{k=1}^{K} - \lambda^{(n)} \mathbf{G}^T \rho'(G\hat{\mathbf{x}}^{(n)}, \alpha) \right\} \tag{5}$$

where $\mathbf{G}$ is the gradient operator over the cliques [81,7], and $\lambda^{(n)}$, the regularization operator can be computed as

$$\lambda^{(n)} = \left(\frac{\sum_{k=1}^{K} \left\| \mathbf{y}_k - \mathbf{D}\mathbf{B}_k \mathbf{W}_k \mathbf{R}[\hat{\mathbf{x}}^{(n)}]_k \right\|_2}{K \sum_{g \in Gx} \rho(g, \alpha)} \right)^2 \cdot \ldots\ldots \quad \ldots\ldots \quad \ldots..(6)$$

Furthermore, the derivative of the Huber function is given as:

$$\rho'(x,\alpha) = \begin{cases} 2x, & \text{if} \quad |x| \le \alpha, \\ 2\alpha sign(x), & \text{otherwise.} \end{cases} \tag{7}$$

The gradient operator $\mathbf{G}$ has the advantage over the Total Variation (TV) prior; the function and its gradient with respect to $\hat{\mathbf{x}}^{(n)}$ are continuous as well as convex [19]. Therefore, the optimization problem can be solved easily using gradient-descent methods such as steepest descent and conjugate gradient methods.

The clique structure determines the spatial interactions. The spatial interactions are used with our proposed method, and its activity is computed using finite difference approximations to second-order directional derivatives (vertical, horizontal, and two diagonal directions) in each super-resolution mosaic $\hat{\mathbf{x}}^{(n)}$.

2.3. Super-resolution Mosaicking Using Conjugate Gradient

The solution of (3) can be estimated using conjugate gradient as

$$\hat{\mathbf{x}}^{(n+1)} = \hat{\mathbf{x}}^{(n)} + \beta^{(n)}\mathbf{p}^{(n)}, \tag{8}$$

where $\mathbf{p}^{(n)}$ is chosen to be conjugate to all previous search directions with respect to the Hessian matrix $\mathbf{H}$:

$$\mathbf{p}^{(n)} = \nabla f(\hat{\mathbf{x}}^{(n)}) + \left(\frac{\nabla f(\hat{\mathbf{x}}^{(n)})^T \nabla f(\hat{\mathbf{x}}^{(n)})}{\nabla f(\hat{\mathbf{x}}^{(n-1)})^T \nabla f(\hat{\mathbf{x}}^{(n-1)})} \right) \mathbf{p}^{(n-1)}. \tag{9}$$

The gradient vector, $\nabla f(\hat{\mathbf{x}}^{(n)})$,is given by the following expression:

$$\nabla f(\hat{\mathbf{x}}^{(n)}) = \mathbf{R}^T \left[\mathbf{W}_k^T \mathbf{B}_k^T \mathbf{D}^T (\mathbf{y}_k - \mathbf{D}\mathbf{B}_k \mathbf{W}_k \mathbf{R}[\hat{\mathbf{x}}^{(n)}]_k \right]_{k=1}^{K} - \lambda^{(n)} \mathbf{G}^T \rho'(G\hat{\mathbf{x}}^{(n)}, \alpha) \tag{10}$$

The gradient operator $\mathbf{G}$ is the same as in the steepest descent method.

2.4. Super-resolution Mosaicking Using Levenverg Marquardt

This algorithm shares with gradient methods their ability to converge from an initial guess which may be outside of the region of convergence of other methods. Based on the Levenberg Marquardt method for minimizing (5.3), and defining $f(\mathbf{x})$ as

$$f(x) = \left\| \sum_{k=1}^{K} \left(y_k - \mathbf{D}\mathbf{B}_k \mathbf{W}_k \mathbf{R}[\mathbf{x}]_k \right) \right\|_2^2 + \lambda \sum_{g \in G_x} \rho(g, \alpha),$$
$$f(\mathbf{x} + \delta\mathbf{x}) \approx f(\mathbf{x}) + J\delta\mathbf{x}, \tag{11}$$

where $J(\mathbf{x})$ is given as the Jacobian matrix:

$$J = \partial \frac{f(\mathbf{x})}{\partial \mathbf{x}} = \partial \frac{\left\{ \left\| \sum_{k=1}^{K} \mathbf{y}_k - \mathbf{D}\mathbf{B}_k \mathbf{W}_k \mathbf{R}[\mathbf{x}]_k \right\|_2^2 + \lambda \sum_{g \in G_x} \rho(g, \alpha) \right\}}{\partial \mathbf{x}} \tag{12}$$

$$J = \mathbf{R}^T \left[\mathbf{W}_k^T \mathbf{B}_k^T \mathbf{D}^T (\mathbf{y}_k - \mathbf{D}\mathbf{B}_k \mathbf{W}_k \mathbf{R}[\hat{x}^{(n)}]_k \right]_{k=1}^{K} - \lambda^{(n)} G^T \rho'(G\hat{\mathbf{x}}^{(n)}, \alpha) \tag{13}$$

The Levenberg Marquardt method is iterative. Initiated at the starting point $\hat{\mathbf{x}}^{(0)}$, the method requires finding $\delta\mathbf{x}$ that minimizes

$$\left\| \hat{\mathbf{x}} - f(\hat{\mathbf{x}} + \delta\mathbf{x}) \right\| \approx \left\| \hat{\mathbf{x}} - f(\mathbf{x}) - J\delta\mathbf{x} \right\| = \left\| \varepsilon - J\delta\mathbf{x} \right\| \tag{14}$$

$\delta\mathbf{x}$ is found by solving a linear least squares problem [15][18]. The minimum is attained when $J\delta\mathbf{x} - \boldsymbol{\varepsilon}$ is orthogonal to the column space of J. This leads to:

$$J^T J\delta\mathbf{x} = J^T\boldsymbol{\varepsilon}$$
$$H^*\delta\mathbf{x} = J^T\boldsymbol{\varepsilon} \tag{15}$$

where H^* is called the pseudo Hessian, defined as $H^* = J^T J$ Levenberg Marquardt solves equation (15), adding a damping term to the diagonal elements of H^*. Therefore, the Levenberg Marquardt equation is

$$(H^* + cI)\delta\mathbf{x} = J^T\boldsymbol{\varepsilon}, \tag{16}$$

where $\delta\mathbf{x}$ is found by solving

$$\delta\mathbf{x} = \arg\min_{\delta\mathbf{x}} \left\| \left(H^* + cI\right)\delta\mathbf{x} - J^T\boldsymbol{\varepsilon} \right\|_2 . \tag{17}$$

Then,

$$\hat{\mathbf{x}}^{(n+1)} = \hat{\mathbf{x}}^{(n)} + \delta\mathbf{x}, \tag{18}$$

where c is the Levenberg Marquardt damping term that determines the behavior of the gradient in each iteration. If c is close to zero, then the algorithm behaves like a Gauss Newton (GN) method, but if $c \to \infty$, then the algorithm behaves like the steepest descent (SD) algorithm. The values of c during the iterative process are chosen in the following way: at the beginning of the iterations, c is set to a large value, so that the LM method uses the robustness of SD, and the initial guess of the solution to (3) can be chosen with less caution. It is necessary to save the value of the errors for each iteration, and do the comparison between two consecutive errors. In the case that $error_{(k)} < error_{(k-1)}$, c is decreased by a certain amount so that LM behaves like Gauss-Newton to take advantage of the speed up to convergence. Otherwise, c increases to a larger value, thus increasing the searching area, which means that LM behaves like SD. The $error_{(k)}$ is defined as

$$error_{(k)} = \frac{\left\| \hat{\mathbf{x}}_{k+1} - \hat{\mathbf{x}}_k \right\|_2}{\left\| \hat{\mathbf{x}}_k \right\|_2} \tag{19}$$

2. EXPERIMENTAL RESULTS

This section compares the results of the three proposed algorithms for super-resolution mosaicking. The comparison will be based on PSNR, time, and error for the synthetic data sets, and time and error for the real frames from UAS because there is no ground truth data available to compute the PSNR with real frames. We created synthetic LR frames from a single high resolution image. These LR frames where created using different translations (18 to 95 pixels), rotations (5^o to 10^o), and scales (1 to 1.5); then blurred them with a Gaussian Kernel. From the comparisons shown in Figures 1 and 2 and from Tables 1 and 2, all the methods improve the resolution of the LR mosaic, and all of them improve the color, details, and sharpness. But, when the image is black and white (IR images), the Levenberg Marquardt produces some artifacts since it solves linear squares equation that is close to be singular (17). The final error for the steepest descent and conjugate gradient algorithms decreases with every iteration, which means that they find the optimal solution in every iteration; but for the Levenberg Marquardt algorithm, this error can decrease or increase due to the use of the damping factor, c, which accelerates the search for the optimal solution.

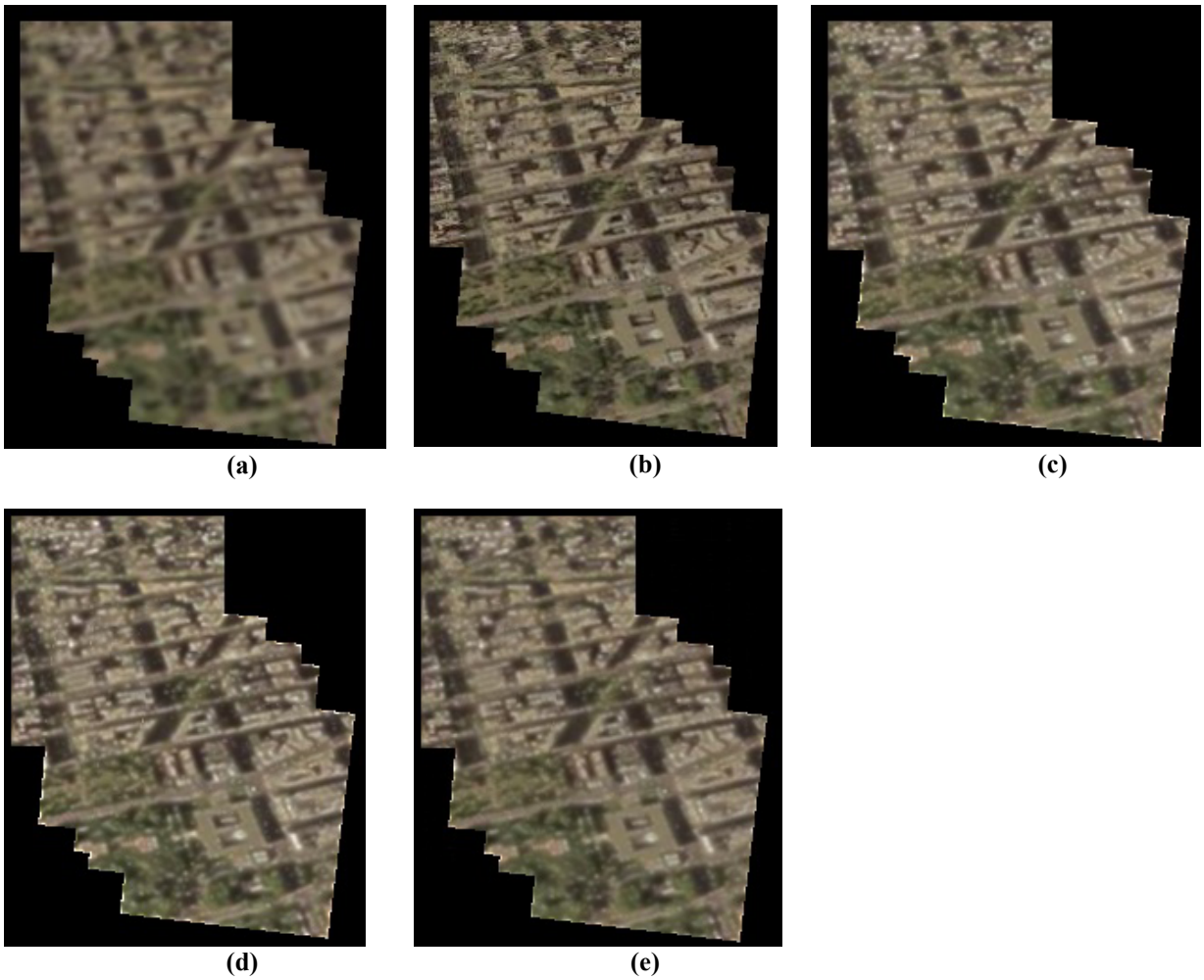

Figure 1. Comparison of the three proposed algorithms: steepest descent, conjugate gradient, and Levenberg Marquardt. These images belong to the first set of synthetic frames created. (a) LR mosaic. (b) Ground truth HR mosaic. (c) SR mosaic using steepest descent. (d) SR mosaic using conjugate gradient. (e) SR mosaic using Levenberg Marquardt.

Table 1. Comparison of the three proposed algorithms to compute super-resolution mosaics for the first set of synthetic color frames.

Algorithm	PSNR (dB)	Final error $\left\|\hat{\mathbf{x}}_{k+1} - \hat{\mathbf{x}}_k\right\|_2 \Big/ \left\|\mathbf{x}_k\right\|_2$	Total Processing Time on CPU (sec)
Super-resolution using steepest descent algorithm.	43.86	0.006391	4.625
Super-resolution using conjugate gradient algorithm.	43.98	0.004381	5.047
Super-resolution using Levenberg Marquardt algorithm.	43.77	0.002833	5.422

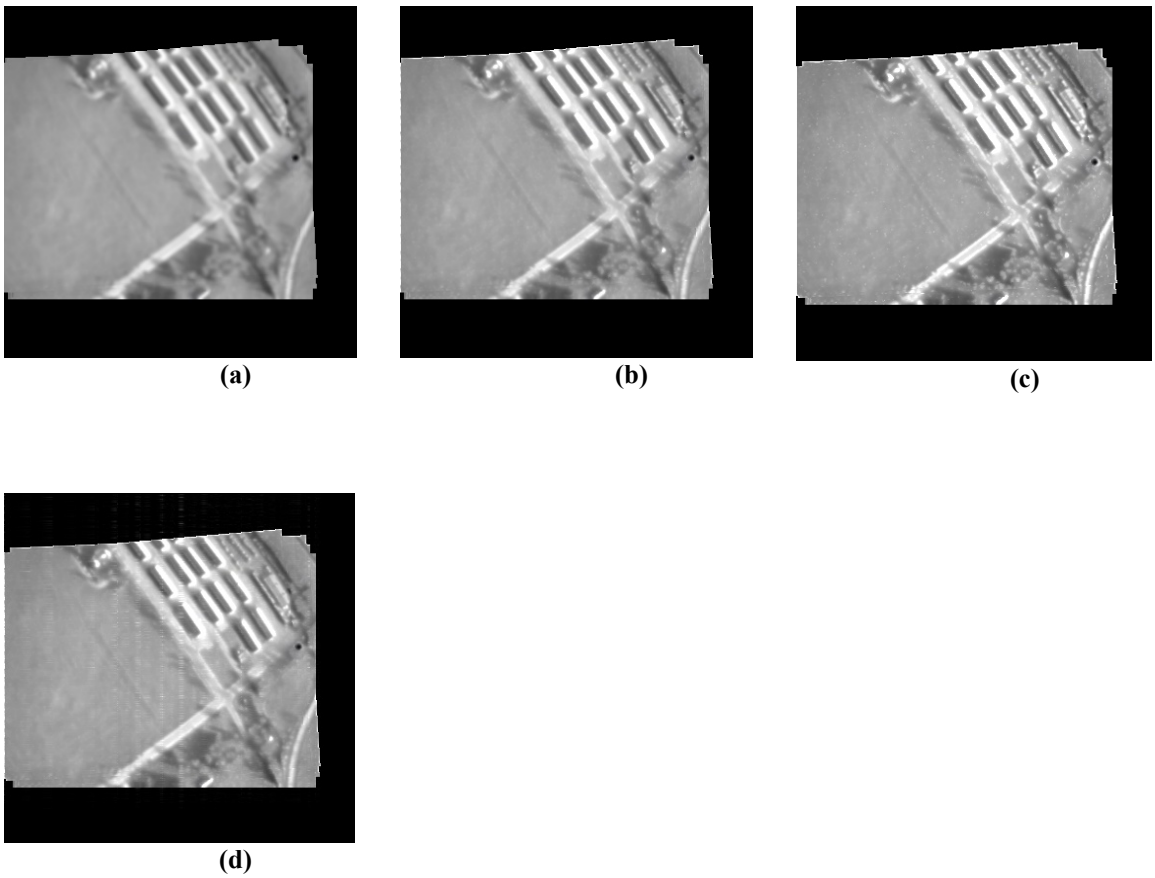

Figure 2. Comparison of the three proposed algorithms: steepest descent, conjugate gradient, and Levenberg Marquardt. These images belong to the first set of real IR video frames captured from UAS. (a) LR mosaic. (b) SR mosaic using steepest descent. (c) SR mosaic using conjugate gradient. (d) SR mosaic using Levenberg Marquardt.

Table 2. Comparison of the three proposed algorithms to compute super-resolution mosaics for the set of real video IR frames captured by UAS.

Algorithm	Final error $\left\|\hat{\mathbf{x}}_{k+1} - \hat{\mathbf{x}}_k\right\|_2 \Big/ \left\|\mathbf{x}_k\right\|_2$	Total Processing Time on CPU (sec)
Super-resolution using steepest descent algorithm.	0.065014	10.844
Super-resolution using conjugate gradient algorithm.	0.097590	11.907
Super-resolution using Levenberg Marquardt algorithm.	0.068155	11.750

3. CONCLUSIONS

The three optimization models improve the resolution of the mosaic images. According with the results using synthetic and real frames, the Conjugate Gradient produces the best results, the Levenberg Marquardt second last the Steepest Descent.

Acknowledgments

This research was supported in part by the Defense Experimental Program to Stimulate Competitive Research (DEPSCoR) program, Army Research Office grant number 50441-CI-DPS, Computing and Information Sciences Division, "Real-Time Super-Resolution ATR of UAV-Based Reconnaissance and Surveillance Imagery," (PI, Richard R. Schultz, Principal Investigator,). This research was also supported in part by Joint Unmanned Aircraft Systems Center of Excellence contract number FA4861-06-C-C006, "Unmanned Aerial System Remote Sensing and Avoidance System and Advanced Payload Analysis and Investigation," as well as the North Dakota Department of Commerce grant, "UND Center of Excellence for UAV and Simulation Applications." Additionally, the authors would like to acknowledge the contributions of the Unmanned Aircraft Systems Engineering (UASE) Laboratory team at the University of North Dakota, This research was also supported by Fincyt (Perú) under the SuperRIVAM project.

REFERENCES

[1] Mark Pickering, Getian Ye, Michael Frater and Jhon Arnold. "A Transform-Domain Approach to Super-Resolution Mosaicing of Compressed Images". 4th AIP International Conference and the 1st Congress of the IPIA. Journal of Physics: Conference Series 124 (2008) 012039.

[2] Richard R Shultz, Li Meng, and Robert L. Stevenson. "Subpixel motion estimation for multiframe resolution enhancement". Visual Communication and Image Processing 1997, pp 1317-1328.

[3] Lynsey C. Pickup. "Machine Learning in Multi-frame Image Super-resolution". Ph.D. Dissertation, University of Oxford, 2007.

[4]Sean Borman. "Topic in Multiframe Superresolution Restoration". Ph.D. Dissertation, University of Notre Dame, Notre Dame, Indiana, 2004.

[5] David Peter Capel "Image Mosaicing and Super-resolution". University of Oxford, Ph. D. Dissertation. University of Oxford, 2001.

[6] Assaf Zomet, and Shmuel Peleg. "Efficient Super-resolution and Applications to Mosaics". Proc of International Conference of Pattern Recognition, Sept 2000.

[7] M. Irani and S. Peleg, "Improving resolution by image registration", Graph. Models Image Process., vol. 53.. pp 231-239. March 1991.

[8] Bryce B. Ready, Clark N. Taylor and Randal W. Beard. "A Kalman-filter Based Method for Creation of Super-resolved Mosaicks". Robotics and Automation, 2006. UCRA 2006.

[9] Aljoscha Smolic and Thomas Wiegand. "High-resolution video mosaicing"

[10] Yi Wang, Ronald Fevig and Richard R. Schultz. "Super-resolution Mosaicking of UAV Surveillance Video" , ICIP 2008, pp 345-348, 2008.

[11] Sina Farsiu, Dirk Robinson, Michael Elad, Peyman Milanfar. "Fast and Robust Multi-Frame Super-resolution," IEEE Transaction on Image Processing,Vol. 13, No. 10. (2004), pp. 1327-1344.

[12] M. Brown and D.G. Lowe, "Recognising Panoramas," ICCV, Proceedings of the Ninth IEEE International Conference on Computer Vision – Volume 2, pp. 1218, 2003.

[13] Krystian Mikolajczyk and Cordelia Schmid. "A Performance Evaluation of Local Descriptors," IEEE Transactions of Pattern Analysis and Machine Intelligence. Vol.27, No. 10, 2005.

[14] Antonio Plaza, Javier Plaza and Hugo Vegas. "Improving the Performance of Hyperspectral Image and Signal Processing Algorithms Using Parallel, Distributed and Specialized Hardware-Based Systems," Journal Signal Processing Systems, DOI 10.1007/s11265-010-0453-1

[15] D.W. Marquardt. An Algorithm for the Least-Squares Estimation of Nonlinear Parameters. SIAM Journal of Applied Mathematics, 11(2):431–441, Jun.1963

[16] Julianne Chung and James G. Nagy. “Nonlinear Least Squares and Super Resolution”. Journal of Physics: Conference Series 124 (2008) 012019
[17] Zafer Arican and Pascal Frossard. “Joint Registration and Super-resolution with Omnidirectional Images”. IEEE Transactions on Image Processing. 2009.
[18] W.H. Press, S.A. Teukolsky, A.W.T. Vetterling, and B.P. Flannery. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, New York, 1992.
[19] Richard R.Schultz. “Multichannel Stochastic Image Models: Theory, Applications and Implementations.” PhD dissertation. University of Notre Dame, Indiana.