1. Introduction

In practice, surveillance video captured by a small Unmanned Aircraft System (UAS) digital imaging payload is almost always blurred and degraded because of limits of the imaging equipment and less than ideal atmospheric conditions. Small UAS vehicles typically have wingspans of less than four meters and payload carrying capacities of less than 50 kilograms, which results in a high vibration environment due to winds buffeting the aircraft and thus poorly stabilized video that is not necessarily pointed at a target of interest. Super-resolution image reconstruction can reconstruct a highly-resolved image of a scene from either a single image or a time series of low-resolution images based on image registration and fusion between different video frames [1, 6, 8, 18, 20, 27]. By fusing several subpixel-registered, low-resolution video frames, we can reconstruct a high-resolution panoramic image and thus improve imaging system performance. There are four primary applications for super-resolution image reconstruction:

1. *Automatic Target Recognition:* The interesting target is hard to identify and recognize under degraded videos and images. For a series of low-resolution images captured by a small UAS vehicle flown over an area under surveillance, we need to perform super-resolution to enhance image quality and automatically recognize targets of interest.

2. *Remote Sensing:* Remote sensing observes the Earth and helps monitor vegetation health, bodies of water, and climate change based on image data gathered by wireless equipments over time. We can gather additional information on a given area by increasing the spatial image resolution.

3. *Environmental Monitoring:* Related to remote sensing, environmental monitoring helps determine if an event is unusual or extreme, and to assist in the development of an appropriate experimental design for monitoring a region over time. With the
development of green industry, the related requirements become more and more important.

4. Medical Imaging: In medical imaging, several images of the same area may be blurred and/or degraded because of imaging acquisition limitations (e.g., human respiration during image acquisition). We can recover and improve the medical image quality through super-resolution techniques.

An Unmanned Aircraft System is an aircraft/ground station that can either be remote-controlled manually or is capable of flying autonomously under the guidance of pre-programmed GPS waypoint flight plans or more complex onboard intelligent systems. UAS aircrafts have recently been found a wide variety of military and civilian applications, particularly in intelligence, surveillance, and reconnaissance as well as remote sensing. Through surveillance videos captured by a UAS digital imaging payload over the same general area, we can improve the image quality of pictures around an area of interest. Super-resolution image reconstruction is capable of generating a high-resolution image from a sequence of low-resolution images based on image registration and fusion between different image frames, which is directly applicable to reconnaissance and surveillance videos captured by small UAS aircraft payloads.

Super-resolution image reconstruction can be realized from either a single image or from a time series of multiple video frames. In general, multiframe super-resolution image reconstruction is more useful and more accurate, since multiple frames can provide much more information for reconstruction than a single picture. Multiframe super-resolution image reconstruction algorithms can be divided into essentially two categories: super-resolution from the spatial domain [3, 5, 11, 14, 26, 31] and super-resolution from the frequency domain [27, 29], based on between-frame motion estimation from either the spatial or the frequency domains.

Frequency-domain super-resolution assumes that the between-frame motion is global in nature. Hence, we can register a sequence of images through phase differences in the frequency domain, in which the phase shift can be estimated by computing the correlation. The frequency-domain technique is effective in making use of low-frequency components to register a series of images containing aliasing artifacts. However, frequency-domain approaches are highly sensitive to motion errors. For spatial-domain super-resolution methods, between-frame image registration is computed from the feature correspondences in the spatial domain. The motion models can be global for the whole image or local for a set of corresponding feature vectors [2]. Zomet et al. [31] developed a robust super-resolution method. Their approach uses the median filter in the sequence of image gradients to iteratively update the super-resolution results. This method is robust to outliers, but computationally expensive. Keren et al. [14] developed an algorithm using a Taylor series expansion on the motion model extension, and then simplified the parameter computation. Irani et al. [11] applied local motion models in the spatial domain and computed multiple object motions by estimating the optical flow between frames.

Our goal here is to develop an efficient (i.e., real-time or near-real-time) and robust super-resolution image reconstruction algorithm to recover high-resolution video captured from a low-resolution UAS digital imaging payload. Because of the time constraints on processing video data in near-real-time, optimal performance is not expected, although we still anticipate obtaining satisfactory visual results.
This paper proceeds as follows. Section 2 describes the basic modeling of super-resolution image reconstruction. Our proposed super-resolution algorithm is presented in Section 3, with experimental results presented in Section 4. We draw conclusions from this research in Section 5.

2. Modeling of Super-Resolution Image Reconstruction

Following the descriptions in [4, 7], we extend the images column-wise and represent them as column vectors. We then build the linear relationship between the original high-resolution image \( \hat{X} \) and each measured low-resolution image \( \hat{Y}_k \) through matrix representation. Given a sequence of low resolution images \( i_1, i_2, \ldots, i_n \) (where \( n \) is the number of images), the relationship between a low-resolved image \( \hat{Y}_k \) and the corresponding highly-resolved image \( \hat{X} \) can be formulated as a linear system,

\[
\hat{Y}_k = D_k C_k F_k \hat{X} + \hat{E}_k \text{, for } k = 1, \ldots, n
\]

where \( \hat{X} \) is the vector representation for the original highly-resolved image, \( \hat{Y}_k \) is the vector representation for each measured low-resolution image, \( \hat{E}_k \) is the Gaussian white noise vector for the measured low-resolution image \( i_k \), \( F_k \) is the geometric warping matrix, \( C_k \) is the blurring matrix, and \( D_k \) is the down-sampling matrix. Assume that the original highly-resolved image has a dimension of \( p \times p \), and every low-resolution image has a dimension of \( q \times q \). Therefore, \( \hat{X} \) is a \( p^2 \times 1 \) vector and \( \hat{Y}_k \) is a \( q^2 \times 1 \) vector. In general, \( q \ll p \), so equation (1) is an underdetermined linear system. If we group all \( n \) equations together, it is possible to generate an overdetermined linear system with \( nq^2 \geq p^2 \):

\[
\begin{bmatrix}
\hat{Y}_1 \\
\vdots \\
\hat{Y}_n
\end{bmatrix} = \begin{bmatrix} D_1 C_1 F_1 \\
\vdots \\
D_n C_n F_n \end{bmatrix} \hat{X} + \begin{bmatrix} \hat{E}_1 \\
\vdots \\
\hat{E}_n \end{bmatrix}
\]

Equivalently, we can express this system as

\[
Y = H \hat{X} + E,
\]

where

\[
Y = \begin{bmatrix}
\hat{Y}_1 \\
\vdots \\
\hat{Y}_n
\end{bmatrix}, \quad H = \begin{bmatrix} D_1 C_1 F_1 \\
\vdots \\
D_n C_n F_n \end{bmatrix}, \quad E = \begin{bmatrix} \hat{E}_1 \\
\vdots \\
\hat{E}_n \end{bmatrix}.
\]

In general, the solution to super-resolution reconstruction is an ill-posed inverse problem. The accurate analytic mathematical solution can not be reached. There are three practical estimation algorithms used to solve this (typically) ill-posed inverse problem [4], that is, (1) maximum likelihood (ML) estimation, (2) maximum a posteriori (MAP) estimation, and (3) projection onto convex sets (POCS).

Different from these three approaches, Zomet et al. [31] developed a robust super-resolution method. The approach uses a median filter in the sequence of image gradients to iteratively
update the super-resolution results. From equation (1), the total error for super-resolution reconstruction in the $L_2$-norm can be represented as

$$L_2(\tilde{X}) = \frac{1}{2} \sum_{i=1}^{n} \| \tilde{y}_i - D_i C_i F_i \tilde{X} \|^2.$$  

(4)

Differentiating $L_2(\tilde{X})$ with respect to $\tilde{X}$, we have the gradient $\nabla L_2(\tilde{X})$ of $L_2(\tilde{X})$ as the sum of derivatives over the low-resolution input images:

$$\nabla L_2(\tilde{X}) = \sum_{i=1}^{n} F_i^T C_i^T D_i^T \left( D_i C_i F_i \tilde{X} - \tilde{y}_i \right)$$

(5)

We can then implement an iterative gradient-based optimization technique to reach the minimum value of $L_2(\tilde{X})$, such that

$$\tilde{X}^{i+1} = \tilde{X}^i + \lambda \nabla L_2(\tilde{X}),$$

where $\lambda$ is a scalar that defines the step size of each iteration in the direction of the gradient $\nabla L_2(\tilde{X})$.

Instead of a summation of gradients over the input images, Zomet [31] calculated $n$ times the scaled pixel-wise median of the gradient sequence in $\nabla L_2(\tilde{X})$. That is, we have:

$$\tilde{X}^{i+1} = \tilde{X}^i + \lambda \cdot n \cdot \text{median} \left[ F_1^T C_1^T D_1^T \left( D_1 C_1 F_1 \tilde{X} - \tilde{y}_1 \right), \ldots, F_n^T C_n^T D_n^T \left( D_n C_n F_n \tilde{X} - \tilde{y}_n \right) \right],$$  

(7)

where $t$ is the iteration step number. It is well-known that the median filter is robust to outliers. Additionally, the median can agree well with the mean value under a sufficient number of samples for a symmetric distribution. Through the median operation in equation (7), we supposedly have a robust super-resolution solution. However, we need to execute many computations to implement this technique. We not only need to compute the gradient map for every input image, but we also need to implement a large number of comparisons to compute the median. Hence, this is not truly an efficient super-resolution approach.

3. Efficient and Robust Super-Resolution Image Reconstruction

In order to improve the efficiency of super-resolution, we do not compute the median over the gradient sequence for every iteration. We have developed an efficient and robust super-resolution algorithm for application to small UAS surveillance video that is based on a coarse-to-fine strategy. The coarse step builds a coarsely super-resolved image sequence from the original video data by piece-wise registration and bicubic interpolation between every additional frame and a fixed reference frame. If we calculate pixel-wise medians in the coarsely super-resolved image sequence, we can reconstruct a refined super-resolved image. This is the fine step for our super-resolution image reconstruction algorithm. The advantage of our algorithm is that there are no iterations within our implementation, which is unlike traditional approaches based on highly-computational iterative algorithms [15]. Thus, our algorithm is very efficient, and it provides an acceptable level of visual performance.

3.1 Up-sampling process between additional frame and the reference frame

Without loss of generality, we assume that $i_k$ is the reference frame. For every additional frame $i_k$ ($1 < k \leq n$) in the video sequence, we transform it into the coordinate system of the reference frame through image registration. Thus, we can create a warped image...
\[ i_{u} = \text{Interpolation}(i_{w}, i_{r}, \text{factor}), \]

where \text{factor} is the up-sampling scale.

3.2 Motion estimation
As required in multiframe super-resolution approaches, the most important step is image registration between the reference frame and any additional frames. Here, we apply subpixel motion estimation [14, 23] to estimate between-frame motion. If the between-frame motion is represented primarily by translation and rotation (i.e., the affine model), then the Keren motion estimation method [14] provides a good performance. Generally, the motion between aerial images observed from an aircraft or a satellite can be well approximated by this model. Mathematically, the Keren motion model is represented as

\[
\begin{pmatrix}
  x' \\
  y'
\end{pmatrix} = s \begin{pmatrix}
  \cos(\theta) & -\sin(\theta) \\
  \sin(\theta) & \cos(\theta)
\end{pmatrix} \begin{pmatrix}
  x \\
  y
\end{pmatrix} + \begin{pmatrix}
  a \\
  b
\end{pmatrix},
\]

where \( \theta \) is the rotation angle, and \( a \) and \( b \) are translations along directions \( x \) and \( y \), respectively. In this expression, \( s \) is the scaling factor, and \( x' \) and \( y' \) are registered coordinates of \( x \) and \( y \) in the reference coordinate system.

3.3 Proposed algorithm for efficient and robust super-resolution
Our algorithm for efficient and robust super-resolution image reconstruction consists of the following steps:

1. Choose frame \( i_{r} \) as the reference frame.
2. For every additional frame \( i_{k} \):
   - Estimate the motion between the additional frame \( i_{k} \) and the reference frame \( i_{r} \).
   - Register additional frame \( i_{k} \) to the reference frame \( i_{r} \) using the \( i_{k} = \text{Regis}(i_{r}, i_{k}) \) operator.
   - Create the coarsely-resolved image \( i_{k} = \text{Interpolation}(i_{k}w, i_{r}, \text{factor}) \) through bicubic interpolation between the registered frame \( i_{k}w \) and the reference frame \( i_{r} \).
3. Compute the median of the coarsely resolved up-sampled image sequence \( [i_{1}w, \ldots, i_{n}w] \) as the updated super-resolved image.
4. Enhance the super-resolved image if necessary by sharpening edges, increasing contrast, etc.

4. Experimental Results
The proposed efficient and robust super-resolution image reconstruction algorithm was tested on two sets of real video data captured by an experimental small UAS operated by
Lockheed Martin Corporation flying a custom-built electro-optical (EO) and uncooled thermal infrared (IR) imager. The time series of images are extracted from videos with low-resolution 60 x 80. In comparison with five well-known super-resolution algorithms in real UAS video tests, namely the robust super-resolution algorithm [31], the bicubic interpolation, the iterated back projection algorithm [10], the projection onto convex sets (POCS) [24], and the Papoulis-Gerchberg algorithm [8, 19], our proposed algorithm gave both good efficiency and robustness as well as acceptable visual performance. For low-resolution 60 x 80 pixel frames with five frames in every image sequence, super-resolution image reconstruction with up-sampling factors of 2 and 4 can be implemented very efficiently (approximately in real-time). Our algorithm was developed using MATLAB 7.4.0. We implemented our algorithm on a Dell 8250 workstation with a Pentium 4 CPU running at 3.06GHz with 1.0GB of RAM. If we ported the algorithm into the C programming language, the algorithm would execute much more quickly.

Test data taken from small UAS aircraft are highly susceptible to vibrations and sensor pointing movements. As a result, the related video data are blurred and the interesting targets are hard to be identified and recognized. The experimental results for the first data set are given in Figures 1, 2, and 3. The experimental results for the second data set are provided in Figures 4, 5, and 6.

Fig. 1. Test Set #1 low-resolution uncooled thermal infrared (IR) image sequence captured by a small UAS digital imaging payload. Five typical frames are shown in (a), (b), (c), (d), and (e), with a frame size of 60 x 80 pixels.

Fig. 2. Test Set #1 super-resolved images, factor 2 (reduced to 80% of original size for display). Results were computed as follows: (a) Robust super-resolution [31]. (b) Bicubic interpolation. (c) Iterated back projection [10]. (d) Projection onto convex sets (POCS) [24]. (e) Papoulis-Gerchberg algorithm [8, 19]. (f) Proposed method.
Super-Resolution Reconstruction by Image Fusion and Application to Surveillance Videos Captured by Small Unmanned Aircraft Systems

Lockheed Martin Corporation flying a custom-built electro-optical (EO) and uncooled thermal infrared (IR) imager. The time series of images are extracted from videos with low-resolution 60 x 80. In comparison with five well-known super-resolution algorithms in real UAS video tests, namely the robust super-resolution algorithm [31], the bicubic interpolation, the iterated back projection algorithm [10], the projection onto convex sets (POCS) [24], and the Papoulis-Gerchberg algorithm [8, 19], our proposed algorithm gave both good efficiency and robustness as well as acceptable visual performance. For low-resolution 60 x 80 pixel frames with five frames in every image sequence, super-resolution image reconstruction with up-sampling factors of 2 and 4 can be implemented very efficiently (approximately in real-time). Our algorithm was developed using MATLAB 7.4.0. We implemented our algorithm on a Dell 8250 workstation with a Pentium 4 CPU running at 3.06GHz with 1.0GB of RAM. If we ported the algorithm into the C programming language, the algorithm would execute much more quickly.

Test data taken from small UAS aircraft are highly susceptible to vibrations and sensor pointing movements. As a result, the related video data are blurred and the interesting targets are hard to be identified and recognized. The experimental results for the first data set are given in Figures 1, 2, and 3. The experimental results for the second data set are provided in Figures 4, 5, and 6.

Fig. 1. Test Set #1 low-resolution uncooled thermal infrared (IR) image sequence captured by a small UAS digital imaging payload. Five typical frames are shown in (a), (b), (c), (d), and (e), with a frame size of 60 x 80 pixels.

Fig. 2. Test Set #1 super-resolved images, factor 2 (reduced to 80% of original size for display). Results were computed as follows: (a) Robust super-resolution [31]. (b) Bicubic interpolation. (c) Iterated back projection [10]. (d) Projection onto convex sets (POCS) [24]. (e) Papoulis-Gerchberg algorithm [8, 19]. (f) Proposed method.

Fig. 3. Test Set #1 super-resolved images, factor 4 (reduced to 60% of original size for display). Results were computed as follows: (a) Robust super-resolution [31]. (b) Bicubic interpolation. (c) Iterated back projection [10]. (d) Projection onto convex sets (POCS) [24]. (e) Papoulis-Gerchberg algorithm [8, 19]. (f) Proposed method.

Fig. 4. Test Set #2 low-resolution uncooled thermal infrared (IR) image sequence captured by a small UAS digital imaging payload. Five typical frames are shown in (a), (b), (c), (d), and (e), with a frame size of 60 x 80 pixels.
Fig. 5. Test Set #2 super-resolved images, factor 2 (reduced to 80% of original size for display). Results were computed as follows: (a) Robust super-resolution [31]. (b) Bicubic interpolation. (c) Iterated back projection [10]. (d) Projection onto convex sets (POCS) [24]. (e) Papoulis-Gerchberg algorithm [8, 19]. (f) Proposed method.

Fig. 6. Test Set #2 super-resolved images, factor 4 (reduced to 60% of original size for display). Results were computed as follows: (a) Robust super-resolution [31]. (b) Bicubic interpolation. (c) Iterated back projection [10]. (d) Projection onto convex sets (POCS) [24]. (e) Papoulis-Gerchberg algorithm [8, 19]. (f) Proposed method.
Fig. 5. Test Set #2 super-resolved images, factor 2 (reduced to 80% of original size for display). Results were computed as follows: (a) Robust super-resolution [31]. (b) Bicubic interpolation. (c) Iterated back projection [10]. (d) Projection onto convex sets (POCS) [24]. (e) Papoulis-Gerchberg algorithm [8, 19]. (f) Proposed method.

Fig. 6. Test Set #2 super-resolved images, factor 4 (reduced to 60% of original size for display). Results were computed as follows: (a) Robust super-resolution [31]. (b) Bicubic interpolation. (c) Iterated back projection [10]. (d) Projection onto convex sets (POCS) [24]. (e) Papoulis-Gerchberg algorithm [8, 19]. (f) Proposed method.

Tables 1, 2, 3, and 4 show the CPU running times in seconds for five established super-resolution algorithms and our proposed algorithm with up-sampling factors of 2 and 4. Here, the robust super-resolution algorithm is abbreviated as RobustSR, the bicubic interpolation algorithm is abbreviated as Interp, the iterated back projection algorithm is abbreviated as IBP, the projection onto convex sets algorithm is abbreviated as POCS, the Papoulis-Gerchberg algorithm is abbreviated as PG, and the proposed efficient super-resolution algorithm is abbreviated as MedianESR. From these tables, we can see that bicubic interpolation gives the fastest computation time, but its visual performance is rather poor. The robust super-resolution algorithm using the longest running time is computationally expensive, while the proposed algorithm is comparatively efficient and presents good visual performance. In experiments, all of these super-resolution algorithms were implemented using the same estimated motion parameters.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>RobustSR</th>
<th>Interp</th>
<th>IBP</th>
<th>POCS</th>
<th>PG</th>
<th>MedianESR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU Time (s)</td>
<td>9.7657</td>
<td>3.6574</td>
<td>5.5575</td>
<td>2.1997</td>
<td>0.3713</td>
<td>5.2387</td>
</tr>
<tr>
<td>Table 1. CPU running time for Test Set #1 with scale factor 2.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>RobustSR</th>
<th>Interp</th>
<th>IBP</th>
<th>POCS</th>
<th>PG</th>
<th>MedianESR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU Time (s)</td>
<td>17.7110</td>
<td>2.5735</td>
<td>146.7134</td>
<td>11.8985</td>
<td>16.7603</td>
<td>6.3339</td>
</tr>
<tr>
<td>Table 2. CPU running time for Test Set #1 with scale factor 4.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>RobustSR</th>
<th>Interp</th>
<th>IBP</th>
<th>POCS</th>
<th>PG</th>
<th>MedianESR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU Time (s)</td>
<td>8.2377</td>
<td>2.8793</td>
<td>9.6826</td>
<td>1.7034</td>
<td>0.5003</td>
<td>5.2687</td>
</tr>
<tr>
<td>Table 3. CPU running time for Test Set #2 with scale factor 2.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>RobustSR</th>
<th>Interp</th>
<th>IBP</th>
<th>POCS</th>
<th>PG</th>
<th>MedianESR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU Time (s)</td>
<td>25.4105</td>
<td>2.7463</td>
<td>18.3672</td>
<td>11.0448</td>
<td>22.1578</td>
<td>8.2099</td>
</tr>
<tr>
<td>Table 4. CPU running time for Test Set #2 with scale factor 4.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5. Summary

We have presented an efficient and robust super-resolution restoration method by computing the median on a coarsely-resolved up-sampled image sequence. In comparison with other established super-resolution image reconstruction approaches, our algorithm is not only efficient with respect to the number of computations required, but it also has an acceptable level of visual performance. This algorithm should provide a movement in the right direction with respect to real-time super-resolution image reconstruction. In future research, we plan to try other motion models such as planar homography and multi-model motion in order to determine whether or not we can achieve better performance. In addition, we will explore to incorporate the natural image characteristics to set up the criterion of super-resolution algorithms such that the super-resolved images provide high visual performance under natural image properties.

6. References


This book aims to explore the latest practices and research works in the area of sensor fusion. The book intends to provide a collection of novel ideas, theories, and solutions related to the research areas in the field of sensor fusion. This book is a unique, comprehensive, and up-to-date resource for sensor fusion systems designers. This book is appropriate for use as an upper division undergraduate or graduate level textbook. It should also be of interest to researchers, who need to process and interpret the sensor data in most scientific and engineering fields. The initial chapters in this book provide a general overview of sensor fusion. The later chapters focus mostly on the applications of sensor fusion. Much of this work has been published in refereed journals and conference proceedings and these papers have been modified and edited for content and style. With contributions from the world’s leading fusion researchers and academicians, this book has 22 chapters covering the fundamental theory and cutting-edge developments that are driving this field.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following: