Automatic Registration and Mosaicking for Airborne Multispectral Image Sequences

Qian Du, Nareenart Raksuntorn, Adnan Orduyilmaz, and Lori M. Bruce

Abstract
Airborne remote sensing has important applications in agriculture monitoring because of the flexibility of system deployment. The major obstacle in practical use is its high cost. To reduce the cost, a multispectral system can be assembled by using individual cameras onboard a small aerial platform, such as a miniature unmanned aerial vehicle (mini-UAV). In such a case, the cameras may have shifting and rotational misalignment, even after careful adjustment. Contiguous frames are captured as the platform flies. So multi-band registration within a single frame and frame-to-frame mosaicking are necessary to obtain a co-registered multispectral image for the entire monitoring area before any commercial product can be generated to support practical decision-making. In this paper, we present automatic algorithms to achieve this goal. These algorithms are particularly useful to the image scenes where no distinctive features are available. Both automatic and manual evaluations confirm the effectiveness of the developed algorithms in multi-sensor data fusion for overall flat terrain without distinctive features.

Introduction
Remote sensing has important applications in agriculture monitoring, where airborne remote sensing is of particular interest due to its flexibility. The major obstacle of airborne remote sensing in practical use is its high cost. To reduce the cost, a multispectral system can be assembled by using individual cameras onboard a small aerial platform. One example is the multispectral imaging system assembled by the United States Department of Agriculture (USDA) Kika de la Garza Subtropical Agricultural Research Center (Yang et al., 2002), which is composed of three Kodak MegaPlus 1.4i digital charge-coupled device (CCD) cameras with a NIR (845 to 857 nm) filter, a Red (625 to 635 nm) filter, and a Green (555 to 565 nm) filter, respectively, and a computer equipped with three image digitizing boards that have the capability of obtaining images with 1,024 pixels × 1,024 pixels. The cameras have a built-in analog-to-digital (A/D) converter that produces a digital output signal with 256 gray levels. A Cessna 206 aircraft was used to acquire imagery at an altitude of 760 m. Each frame has 480 pixels × 720 pixels with a 45° horizontal field of view (FOV). The deployment for data collection from these systems is very flexible. For instance, the mini-UAV can be hand-launched. Such an economic system is cost-effective in agricultural studies, since it can provide useful image data with greatly reduced data collection cost for farmers. The processed images (after registration and mosaicking) are also very useful for disaster assessment and relief operations. In these systems, image registration is needed to co-register the bands from individual cameras, particularly when the georeferencing capability is not equipped. In addition, frames with multispectral bands are taken contiguously as the aerial platform is flying. So image mosaicking is needed to combine adjacent frames to have a large overview of the monitoring field.

Many image registration tasks are accomplished manually, requiring expert knowledge of image analysts. During the last decades, image acquisition devices have developed rapidly. They capture a huge amount of images with great diversity. This invokes the development on automatic image registration. Reviews can be found in Brown (1992) and Moigne (2003), and Bentoutou et al. (2005). It is expected that automatic image registration can greatly reduce the long turn-around time, which currently is another bottleneck in practical applications of remote sensing in agriculture.

Image mosaicking is the act of combining two or more images with overlapping areas for an overview of a large image scene. The aim is to combine images with an undistorted and smooth transition area so that it appears to have been acquired from a single sensor. Radiometric normalization and blending processes can be employed for this purpose. However, registration is the key step in image mosaicking (Su et al., 2004; Du et al., 2001).

Automatic image registration has recently been widely studied. Some techniques can be found in Fonseca and Manjunath (1996), Moigne et al. (2002), Kennedy and Cohen (2003), Ingland and Giros (2004), Hong and Schowengerdt (2005), and Bentoutou et al. (2005). There exist iterative algorithms searching for a registered image that can minimize the difference from (or maximize the similarity to) the reference image without the need of control points (CPs), and a pyramid method may be used to decrease the computational
time (Thevanaz, et al., 1988; Chen, et al., 1990). However, in this research, we only focus on traditional, non-rigid image registration containing three major steps: control point identification, image transformation, and image resampling.

- Control Point Identification: There are two main methods for the CP detection: area-based and feature-based. In area-based methods, a small window of points in the reference image is statistically compared with windows of the same size in the sensed image. The comparison uses a similarity metric, which measures the similarity between two given windows. In feature-based algorithms, an image is represented in a compact form by a set of features. The features should be invariant to the scaling, rotation, and gray level modification. The common features are edges, regions, line endings, line intersections, or region centroids. Interest operators such as Moravec and Förstner operators can be employed for this purpose (Moravec, 1977 and 1979; Förstner and Gülch, 1987). In general, CP selection is necessary to keep the CPs with good quality from detected CP candidates.

- Spatial Transformation: Once CPs are identified, the transformation parameters in a mapping function for registration can be determined. In order to define the mapping function, a priori information is needed about the types of deformations. If there is no a priori information available, mapping functions must be flexible so that they can be suitable to all possible combinations of degradations.

- Image Resampling: After spatial transformation, the registered image pixel coordinates are not integers anymore. The intensity of the pixels with integer-valued coordinates is computed by an appropriate interpolation technique.

Control point identification is the key step in image registration. Area-based methods do not use the salient points or distinctive features. Instead, intensity values of a window are compared using similarity metrics. In this type of methods, images should have similar or dependent intensity values. These methods generally can handle small rotation and translation misalignments. Feature-based methods are adopted when the features of objects are distinctive. These methods are relatively more powerful for the registration of different types of images with distortions.

Image registration and mosaicking techniques take advantage of the intensity similarity and/or distinctive features in two images. Our research is challenging because the existing techniques cannot be directly applied and can be even inapplicable in many cases. The first challenge in this multi-sensor image registration research comes from the fact that the band images, which are acquired from different spectral sensors and visually appear different, due to different solar reflectance characteristics. The second challenge is induced when most images are of agricultural areas with no distinctive features being present. In feature-based image registration algorithms, distinctive features are needed for the comparison of two images. Unfortunately, in our case, it is difficult to get these reliable features for registration.

In summary, we developed an automatic image registration algorithm to co-register different bands within a single frame and an image mosaicking algorithm to generate a mosaic using adjacent frames after within-frame multi-band co-registration. The major uniqueness of our work exists in the following three aspects:

1. An area-based approach is developed to identify reliable control points from different bands and scenes without prominent features.
2. The area-based method is extended to combine images with large rotational misalignment in image mosaicking (in addition to translation misalignment).
3. Thorough manual evaluations are conducted on the registration accuracy without using the control points that are involved in the registration process.

This paper is organized as follows. In the next section, an automatic image registration algorithm is introduced to co-register the bands within a single frame followed by an automatic image mosaicking algorithm for combining adjacent frames. The overall system diagram is then presented followed by experimental results and quantitative evaluation. Finally, conclusions are drawn in the last section.

Multi-sensor Image Registration

When an area-based method is applied to multi-sensor image registration for image scenes with large homogeneous areas, special approaches need to be taken to ensure the quality of CPs. In our research, the primary strategies include Region of Interest (ROI) selection before CP detection, and CP selection after potential CPs are detected.

Region of Interest Selection

Since the image scenes studied in this research contain large homogeneous regions, region of interest (ROI) selection is conducted before CP detection. It is to select the areas with relatively large grayscale variation. Then, the CP detection will be confined within the selected ROIs. This will greatly reduce the false alarms in the following CP identification step.

To select the distinctive areas, an image is divided into adjacent non-overlapping small blocks. Then, entropy is calculated for each block, which can be used to measure the local variation within the block. Figure 1b shows the entropy map of the image scene (the NIR band) in Figure 1a of size 480 pixels × 720 pixels, where the blocks with large entropies correspond to the areas with relatively high

![Figure 1. ROI Selection: (a) Original image, and (b) Entropy map.](image-url)
variation such as trees and buildings located in the left part of the image. A threshold η is set to choose those blocks with globally large entropies. This method is referred to as global ROI selection.

A potential problem is that the resulting ROIs may not be spatially well-distributed. As shown in Figure 1, the crop field in the right part of the image does not contain a block with a globally large entropy. This means no CPs can be detected from this area, and the CPs will not be widely spread. To overcome this problem, ROIs with locally large entropies should also be selected. Hence, first an image is divided into non-overlapping large patches; then, each patch is further divided into small blocks for entropy evaluation. After using the threshold η to select the blocks with globally large entropies, a patch that does not have any blocks being selected as ROI will choose the block with the local maximal entropy to be an ROI. This method is called local ROI selection. It ensures at least one ROI will be selected from each patch, which may result in CPs being more widely distributed. It should be noted that bad CPs will be removed during the following CP selection procedure as described in the next section.

Control Point Identification

**Control Point Detection**

To find similar areas in the sensed and reference images, a template window is selected at the ROI centers. Each window of the sensed image is compared to the corresponding window in the reference image. If the similarity is above a threshold of the sensed image is compared to the corresponding window in the right part of the image does not contain a block with a maximum much easier.

**Spatial Transformation**

After the CPs are identified, the transformation parameters can be determined using their coordinates. According to the complexity and properties of the distortion, the mapping function type is selected. For instance, when a large number of CPs are available with some noise, the weight mean (WM) may be a good choice; when the images experience local geometric distortion, the piecewise linear (PL) is useful (Zagorchev and Goshtasby, 2006). In our research on agricultural monitoring, most imaged terrain has small variation such as flat crop fields. Therefore, we first consider two transformations: affine transform and eight-parameter projective model, both of which are useful to flat terrain. In particular, the non-linear projective model can accommodate all camera motions. In addition, we consider the thin-plate spline (TPS) to correct nonlinear geometric distortion from hilly areas with terrain changes that may come across; the TPS model is very flexible due to the fact that it is constructed from an affine transform and a weighted combination of nonlinear terms (i.e., radial basis functions). Its performance is relatively robust when the number of CPs is small (Zagorchev and Goshtasby, 2006).

**Affine Transform**

An affine transform is linear, which can be represented as:

\[ u = a_0 + a_1 x + a_2 y \]  
\[ v = b_0 + b_1 x + b_2 y \]  

where \( a_0 \) and \( b_0 \) are for shifting adjustment, and \( a_1, b_1, a_2, \) and \( b_2 \) are for rotational, scaling, and skew adjustments. Here, \((x, y)\) and \((u, v)\) are coordinates of CPs in the reference and sensed images, respectively. Theoretically, three independent CPs are necessary to solve the equations uniquely. In practice, the number of CPs is much larger for higher registration accuracy, and the least squares solution is used to estimate the parameters that can minimize the estimation error.

**Eight-parameter Projective model**

The affine model cannot capture camera pan and tilt. The eight-parameter projective (8PP) model gives the exact eight desired parameters to account for all the possible camera motions, including rotation, pan, and tilt. Note that going from the first-order (i.e., affine) to the second-order results in the 12-parameter biquadratic model. While an increase in the number of model parameters will result in a better fit, there is a tradeoff where the model begins to fit noise. Because the physical camera model yields the projective model with exactly eight parameters, when imaging over a
flat (planar) terrain, the 8-parametric model is studied. It is expressed as (Rao and Yip, 2001):

\[
\begin{bmatrix}
  u \\
  v
\end{bmatrix} = \frac{D \begin{bmatrix} x \\ y \end{bmatrix} + e}{f^T \begin{bmatrix} x \\ y \end{bmatrix} + 1}
\]

where \( D = \begin{bmatrix} p_1 & p_2 \\ p_3 & p_4 \end{bmatrix}, e = \begin{bmatrix} p_5 \\ p_6 \end{bmatrix}, \) and \( f = \begin{bmatrix} p_7 \\ p_8 \end{bmatrix} \) are the parameters to be estimated. Equation can be re-expressed as:

\[
\begin{align*}
  u &= p_1 x + p_2 y + p_3 - p_7 xu - p_8 yu \\
  v &= p_3 x + p_4 y + p_6 - p_7 xv - p_8 yv.
\end{align*}
\]

The matrix form for Equations 7 and 8 is:

\[
\mathbf{x} = \mathbf{Xp}
\]

where \( \mathbf{p} = (p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8)^T \),

\[
\mathbf{X} = \begin{bmatrix} x & y & 1 & 0 & 0 & 0 & -xu & -yu \\ 0 & 0 & 0 & x & y & 1 & -xv & -yv \end{bmatrix}, \text{ and } \mathbf{x} = (uv)^T
\]

(Mann and Picard, 1997). Then, the eight parameters can be estimated using a least squares solution when the number of CPs is larger than four.

**Thin-plate Spline**

Thin-plate spline (TPS) is the most widely used non-linear transformation function in image registration. It was first used by Goshtasby in the registration of remote sensing images and then by Bookstein in the registration of medical images.
images (Goshtasby, 1988; Bookstein, 1989; Rohr et al., 2001). TPS is found to be the most suitable approach when a small set of CPs is available (Zagorchev and Goshtasby, 2006). The pixel coordinates in the sensed image are related to those in the reference image using the following equations:

\[ u = a_0 + a_1 x + a_2 y + \sum_{i=1}^{N} c_i r_i^2 \ln r_i^2 \]  
\[ v = b_0 + b_1 x + b_2 y + \sum_{i=1}^{N} d_i r_i^2 \ln r_i^2 \]  

where \( r_i^2 = (x - x_i)^2 + (y - y_i)^2 \), \((x_i, y_i)\) is the \(i\)th CP in the reference image, and \(N\) is the number of CPs. Equation 10 or 11 contains \(N + 3\) unknowns. By substituting the coordinates of the \(N\) control points and their values in the reference image, \(N\) equations will be obtained. Three more equations are obtained using the following three constraints:

\[ \sum_{i=1}^{N} c_i = 0, \quad \sum_{i=1}^{N} d_i = 0 \]  

which ensure the spline has square-integrable second derivatives.

Equations 10 through 14 can be represented in matrix forms as:

\[ L = \begin{bmatrix} K & P \\ P^T & 0 \end{bmatrix} \]  

where

\[ K = \begin{bmatrix} 0 & k_{12} & \ldots & k_{1N} \\ k_{12} & 0 & \ldots & k_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ k_{1N} & k_{2N} & \ldots & 0 \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_N & y_N \end{bmatrix} \]  

with \(k_{ij} = r_{ij}^2 \ln r_{ij}^2\) and \(r_{ij}^2 = (x_i - x_j)^2 + (y_i - y_j)^2\). The unknown parameters are represented as \(c = (c_1, \ldots, c_N, a_0, a_1, a_2)^T\), and \(d = (d_1, \ldots, d_N, b_0, b_1, b_2)^T\). In addition, the coordinates in the sensed image are included in \(u = (u_1, \ldots, u_N, 0, 0)^T\), and \(v = (v_1, \ldots, v_N, 0, 0)^T\). They are related as:

\[ u = Lc \quad \text{and} \quad v = Ld. \]  

Since \(L\) is a square matrix with full-rank, the unknown parameters in Equations 10 and 11 can be easily solved by calculating \(L^{-1}\).

Note that the TPS models in Equations 15 and 16 include the original affine model in Equations 4 and 5 and the weighted combination of nonlinear radial functions. If the nonlinear distortion is small, then the weights in \(c\) and \(d\) are very small values, which is reduced to the original affine model.

**Multi-frame Mosaicking**

Image mosaicking includes an image registration process. Postprocessing such as radiometric normalization is not required in this case because images to be mosaicked are taken from the same sensor during the same time period. In the registration process for image mosaicking, the slight differences from multi-band image registration include:

1. The images to be registered for mosaicking, which have been co-registered within each frame, are taken in the same band in consecutive frames.
2. Since the NIR bands have higher contrast, the reference image is selected as the NIR band. After the NIR bands are mosaicked, the Red and Green bands can be mosaicked accordingly.
3. In the multi-band image registration, there was a slight translation, so the searching area for potential CPs can be selected as three or four times of the window size to save time. However, in image mosaicking, the searching area is the entire overlapped area, so the resultant computation cost is higher.
4. In the multi-band image registration, there is a small rotation, which can be handled using an affine transform. However, in image mosaicking, the images have larger rotation as the aerial platform flying. So, the rotation angle should be detected with a preprocessing step before CP detection. Otherwise, the area-based similarity comparison is not applicable.
5. As mentioned earlier, a non-linear mapping is preferred because the orientation of the focal planes to the Earth’s surface is changed from frame to frame.

Considering the differences explained above, the mosaicking algorithm has an additional step: rotation angle detection. There may be a large rotation between the consecutive frames. The window content at the same location is changing at each degree, which leads to the difficulties in similarity comparison.
The rotation should be estimated and adjusted before the comparison of windows. With pre-introducing the rotation, the comparison of windows is still conducted using CC or MI. An example of the resulting rotation angle versus CC or MI are given in Figure 5. The rotation angle corresponding to the maximum CC or MI is selected for each window. The rotation angle with maximum occurrence frequency is considered as the actual rotation angle between two frames. From Figure 5, we can see that an MI curve has a sharper peak, and a CC curve is smoother. So, the rotation angle can be more efficiently detected using MI.

After the rotation angle is detected, the NIR image is first rotated back using this angle. The remaining steps are the same as in the registration algorithm described in the previous sections. Then, the Green and Red bands are mosaicked accordingly.

Since the images are taken within a very close amount of time using the same sensors, the radiometric difference between the overlap areas in the reference and sensed images is assumed to be very small. In the transition areas close to the image boundaries, the intensity values are weighted on average to create a smooth and compact mosaic.

**Overall Data Processing System Diagram**

To summarize, the overall system diagram is shown in Figure 6. After multi-band co-registration within a frame and frame-to-frame mosaicking, commercial products are ready to be generated, such as normalized difference vegetation index (NDVI) and color infrared (CIR) composite, which are very useful to agricultural and forestry studies.

---

Figure 5. Rotation detection using (a) CC, and (b) MI for the rotation angle $\theta = 7^\circ$.

Figure 6. System diagram (the methods in the parentheses are to be compared and those highlighted are the better choices to be confirmed).
Experiments and Evaluation
The images collected at four sites (Lake Columbus, Stennis Space Center, Greenwood, and Oswalt in Mississippi) in 2005 by the AOSI's mini-UAV which was used in the experiments. In each site, six panels were placed for the purpose of evaluation. However, most of the frames did not contain these panels.

Image Registration
Both the global and local ROI selection methods were used for comparison purposes. The CPs resulting from global ROI selection is referred to as CP Pool 1, and those resulting from local ROI selection is referred to as CP Pool 2. In the similarity comparison, the window size was chosen as 51 pixels \times 51 pixels. If the window size is too small, the variation contained in an image block may not be distinctive enough to be accurately measured as similar or dissimilar to that in another window. If the window size is too large, then the computational time is greatly increased. It is found that a window size such as 51 \times 51 can achieve a good trade-off in these experiments. When applying the MI as the similarity metric, 128 bins were used to estimate a histogram, which can also achieve a good trade-off in feature extraction and computational complexity. For each dataset collected from the four sites, approximately 40 to 70 CPs were identified and used for registration after CP detection and selection.

An example of the original and registered three-band images and generated CIR and NDVI composites using CP Pool 2 are given in Figure 7 and Plate 1. The images were acquired from Oswalt. An NIR band was used as a reference image because it had higher contrast, and Green and Red

Figure 7. Registration for an Oswalt image (from left to right: NIR, Red, and Green). (a) Original bands, (b) Registered bands using the CP Pool 2 and affine transform, (c) Registered bands using the CP Pool 2 and thin-plate spline, and (d) Registered bands using the CP Pool 2 and eight-parameter model.
bands were registered to the NIR band individually. In the CIR and NDVI images, the effect from registration can be easily recognized. As the alignment accuracy increases, registered images have more compact and clearer CIR and NDVI products, while the mis-registration causes blurry CIR images and some artificial artifacts in NDVI images. For instance, the blue-purple color around the panels and trees were not correct in the CIR composite from the original image, but after registration, the colors became appropriate; some areas such as those around panels and the road with no vegetation had high gray levels in the NDVI generated from the original image. After registration, the gray levels were adjusted.

There was no difference that can be visually perceived between using CP Pool 1 and CP Pool 2, and between using affine transform, 8PP, and TPS. The subtle difference needs to be evaluated with quantitative comparison as described later in this section.

Plate 1. CIR and NDVI images from original and registered images in Figure 7. (From left to right: CIR color composite, NDVI image): (a) Original bands, (b) Registered bands using the CP Pool 2 and affine transform, (c) Registered bands using the CP Pool 2 and thin-plate spline, and (d) Registered bands using the CP Pool 2 and eight-parameter model.
Image Mosaicking
An example for image mosaicking is shown in Plate 2, where ten consecutive frames were mosaicked. The NIR bands were mosaicked first, and then the Red and Green bands were mosaicked using the parameters found for the NIR bands. This is because the NIR bands have higher contrast for image scenes with large areas of vegetation and soil. The CIR images generated after ten frames were mosaicked as shown in Plate 2, where local ROI selection (CP Pool 2) and MI were employed for CP detection. From the mosaic, we can see that the UAV flew around a country road; the left side of this road is primarily bare soil, and the right side of this road is a crop field. This gives us a synoptic view of the entire monitoring area.

The difference between using affine, 8PP, and TPS cannot be visually perceived. The quantitative evaluation is described in the next subsection.

Evaluation of the Results
Eight datasets containing panels were used for the evaluation of registration: three image sets were taken from Oswalt, two were from Greenwood, and another three were taken from Stennis Space Center. For mosaicking, there were three sets for evaluation, and both of them were from Oswalt.

Automatic Evaluation
In the automatic evaluation, the registration accuracy about CPs is measured in mean squared error (MSE). Table 1 lists the average error when using the CC and MI with affine transform. Once again, it confirms that MI can yield smaller registration error than CC in the registration of CPs. Table 2 lists the average error when using the MI but with different sets of CP pools and transforms. Here, the labels such as “Affine 1” means the use of affine transform and CP Pool 1, etc. We can see that there was no big difference on the registration of those CP pixels from the use of CP Pool 1 and CP Pool 2 since the evaluation is about those pixels acting as CPs only. TPS can yield smaller residual error at CPs than the affine transform in all the cases, while 8PP performs better than affine in most cases. Table 3 shows the registration error in mosaicking, where the improvement from 8PP and TPS is more significant. When evaluated with the CPs, TPS did the best among the three transforms for both registration and mosaicking. This may be because TPS employed the largest number of parameters to fit the correspondence between each pair of CPs.

Since the image transformations were designed to minimize the residual error at the CPs, the result from such an automatic evaluation cannot be well-generalized to other regions that do not contain CPs. In particular, when the number of CPs is small, the evaluation result may be biased. So, manual evaluation was also conducted as described in the following subsection.

Manual Evaluation
Before the flights, six panels with the size of 3 meters × 3 meters were put in the sites. The first panel is white, the last is black, and the others have the gray level tones. The corners of these panels are distinctive, so these corners can be used for evaluation. The samples of the image scenes

<table>
<thead>
<tr>
<th>Image</th>
<th>MI</th>
<th>CC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.50</td>
<td>0.79</td>
</tr>
<tr>
<td>2</td>
<td>0.50</td>
<td>0.98</td>
</tr>
<tr>
<td>3</td>
<td>0.40</td>
<td>0.96</td>
</tr>
<tr>
<td>4</td>
<td>0.28</td>
<td>0.91</td>
</tr>
<tr>
<td>5</td>
<td>0.49</td>
<td>0.59</td>
</tr>
<tr>
<td>6</td>
<td>0.56</td>
<td>0.61</td>
</tr>
<tr>
<td>7</td>
<td>0.60</td>
<td>0.54</td>
</tr>
<tr>
<td>8</td>
<td>0.49</td>
<td>0.52</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Image</th>
<th>Affine 1</th>
<th>TPS 1</th>
<th>8PP 1</th>
<th>Affine 2</th>
<th>TPS 2</th>
<th>8PP 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.97</td>
<td>0.62</td>
<td>0.45</td>
<td>1.08</td>
<td>0.51</td>
<td>0.34</td>
</tr>
<tr>
<td>2</td>
<td>0.57</td>
<td>0.40</td>
<td>0.51</td>
<td>0.65</td>
<td>0.33</td>
<td>0.27</td>
</tr>
<tr>
<td>3</td>
<td>0.53</td>
<td>0.34</td>
<td>0.52</td>
<td>0.76</td>
<td>0.51</td>
<td>0.44</td>
</tr>
<tr>
<td>4</td>
<td>1.43</td>
<td>0.64</td>
<td>0.44</td>
<td>0.40</td>
<td>0.36</td>
<td>0.44</td>
</tr>
<tr>
<td>5</td>
<td>0.51</td>
<td>0.23</td>
<td>0.45</td>
<td>0.52</td>
<td>0.32</td>
<td>0.37</td>
</tr>
<tr>
<td>6</td>
<td>0.72</td>
<td>0.39</td>
<td>0.51</td>
<td>0.67</td>
<td>0.37</td>
<td>0.37</td>
</tr>
<tr>
<td>7</td>
<td>0.33</td>
<td>0.28</td>
<td>0.53</td>
<td>0.38</td>
<td>0.31</td>
<td>0.59</td>
</tr>
<tr>
<td>8</td>
<td>0.36</td>
<td>0.23</td>
<td>0.48</td>
<td>0.40</td>
<td>0.25</td>
<td>0.21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Image</th>
<th>Affine 1</th>
<th>TPS 1</th>
<th>8PP 1</th>
<th>Affine 2</th>
<th>TPS 2</th>
<th>8PP 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.06</td>
<td>0.61</td>
<td>0.68</td>
<td>1.17</td>
<td>0.64</td>
<td>1.18</td>
</tr>
<tr>
<td>2</td>
<td>1.03</td>
<td>0.59</td>
<td>0.77</td>
<td>0.91</td>
<td>0.43</td>
<td>0.99</td>
</tr>
<tr>
<td>3</td>
<td>1.20</td>
<td>1.16</td>
<td>0.90</td>
<td>0.98</td>
<td>0.69</td>
<td>0.78</td>
</tr>
</tbody>
</table>
with panels are shown in Figure 8. In registration and mosaicking, the corner pixels were intentionally not used as CPs even if they may be identified as CPs. Six users participated in the manual evaluation. They selected the corners of panels by clicking the mouse on the pixels that they think are the corners, whose coordinates were recorded and compared with the actual coordinates.

First, the performance of CC and MI using affine transform was compared. As an example, the selected corner locations for NIR, Red, and Green bands are illustrated in Plate 3, where the blue points represent the pixel locations of the corners in NIR band, red point represents Red band, and green points represent Green band. Typically, one cluster corresponds to one corner. Before registration, the corners were misaligned with slight rotation and large shifting, so the red, green, and blue clusters for the same corner were far away from each other. After registration, the red, green, and blue clusters for the same corner were very close. Ideally, the corner coordinates should be the same for the bands after registration. The quantified misalignment in MSE between the NIR and Green bands in the eight datasets is given in Table 4. Here, $\mu$ represents the mean of the misalignment for the corners determined by the six users, $\text{std}$ represents the standard deviation, and the unit is in pixels. We can see that the misalignment was fairly high before registration, and it was very small after registration. Comparing the results from MI with CC, the accuracy using MI generally was higher than CC.

![Figure 8. Panels located in image scenes: (a) Stennis Space Center, (b) Greenwood, and (c) Oswalt.](image)

![Plate 3. An example of manual evaluation result for registration using MI/CC and affine transform: (a) Corners before registration, (b) Corners after registration: CC, and (c) Corners after registration: MI.](image)
The performance of 8PP and TPS seems to be comparable. In this case, geometric distortion contained in adjacent frames. The performance due to the fact that they can handle the non-linear component feature details are present in image scenes. The correlation coefficient (CC) and mutual information (MI) are the two metrics used for similarity comparison. Since the images are about vegetation areas and farms, the texture features in different areas are similar to each other. Without any preprocessing, the similarity measures may lead to large registration errors. So, the region of interest (ROI) determination is important. Because the registration is applied to different bands in a single frame but with small misalignment, the ROI is selected by the calculation of the entropy of non-overlapping blocks, and then control point detection is spatially confined within the ROIs. Control point identification is another key step where CC or MI is used for control point detection, followed by control point selection to eliminate outliers to ensure the quality of final control points. In image mosaicking, pre-introducing rotation is another key step, which makes the area-based method feasible when the rotational misalignment cannot be ignored. For image transformation, a non-linear mapping function such as eight-parameter projective model (8PP) or thin-plate spline (TPS) is needed to deal with the geometric difference between image and frame to frame.

Both automatic and manual evaluations are conducted. They confirm the effectiveness of the developed algorithms. In general, the misregistration is less than one pixel at control points, and approximately one pixel at other locations. In particular, MI is demonstrated as a better similarity metric, which can produce more control points with good quality.

<table>
<thead>
<tr>
<th>Table 4. Manual Evaluation for the Misalignment Between NIR-Green Band Using Affine Transform</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Image</strong></td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
</tbody>
</table>

In the following comparison, only MI was used as similarity metric. Plate 4 shows an example of the selected corner coordinates with affine transform, 8PP, and TPS using CP Pools 1 and 2 in registration. Table 5 lists the registration errors between NIR and Green bands. Using the CP Pool 2 generally is better than using the Pool 1 because of evener distribution. But, TPS did not necessarily provide better performance than the affine transform in registration. This may be because there is no significant geometric difference present among the three bands within a single frame. 8PP outperformed affine and TPS in most cases. Using the CP Pool 2 is also better than using the CP Pool 1.

**Conclusions**

In this paper, we developed automatic algorithms for registration and mosaicking for multispectral image sequences taken by an economic sensor system onboard a mini-UAV. After the image integration into the ground control station with the capability of real-time image recording, the near real-time decision-making support is achievable with the final commercial products, such as CIR and NDVI images, for agricultural, forestry, and environmental studies.

The specialties of the acquired images in our research include: most image scenes are of agricultural crop fields without distinctive spatial features, the multi-bands in a single frame are taken simultaneously with small shifting and rotational misalignment, and every two adjacent frames taken at similar altitudes have overlapping areas and relatively large rotational misalignment. Algorithms are developed based on these image characteristics. The area-based method is employed, which is applicable when no prominent feature details are present in image scenes. The correlation coefficient (CC) and mutual information (MI) are the two metrics used for similarity comparison. Since the images are about vegetation areas and farms, the texture features in different areas are similar to each other. Without any preprocessing, the similarity measures may lead to large registration errors. So, the region of interest (ROI) determination is important. Because the registration is applied to different bands in a single frame but with small misalignment, the ROI is selected by the calculation of the entropy of non-overlapping blocks, and then control point detection is spatially confined within the ROIs. Control point identification is another key step where CC or MI is used for control point detection, followed by control point selection to eliminate outliers to ensure the quality of final control points. In image mosaicking, pre-introducing rotation is another key step, which makes the area-based method feasible when the rotational misalignment cannot be ignored. For image transformation, a non-linear mapping function such as eight-parameter projective model (8PP) or thin-plate spline (TPS) is needed to deal with the geometric difference between image and frame to frame.

Both automatic and manual evaluations are conducted. They confirm the effectiveness of the developed algorithms. In general, the misregistration is less than one pixel at control points, and approximately one pixel at other locations. In particular, MI is demonstrated as a better similarity metric, which can produce more control points with good quality.

**Table 5. Manual Evaluation for the Misalignment Between NIR-Green Band Before and After Registration**

<table>
<thead>
<tr>
<th>Image</th>
<th>BeforeReg.</th>
<th>Affine 1</th>
<th>TPS 1</th>
<th>8PP 1</th>
<th>Affine 2</th>
<th>TPS 2</th>
<th>8PP 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16.69</td>
<td>0.46</td>
<td>1.20</td>
<td>0.41</td>
<td>1.13</td>
<td>0.46</td>
<td>0.36</td>
</tr>
<tr>
<td>2</td>
<td>15.93</td>
<td>0.35</td>
<td>1.19</td>
<td>0.43</td>
<td>1.43</td>
<td>1.16</td>
<td>0.36</td>
</tr>
<tr>
<td>3</td>
<td>15.02</td>
<td>0.33</td>
<td>1.16</td>
<td>0.57</td>
<td>1.03</td>
<td>0.31</td>
<td>0.73</td>
</tr>
<tr>
<td>4</td>
<td>15.60</td>
<td>0.67</td>
<td>1.18</td>
<td>0.47</td>
<td>1.20</td>
<td>0.64</td>
<td>0.72</td>
</tr>
<tr>
<td>5</td>
<td>12.93</td>
<td>0.59</td>
<td>0.57</td>
<td>0.21</td>
<td>1.26</td>
<td>0.69</td>
<td>0.55</td>
</tr>
<tr>
<td>6</td>
<td>10.22</td>
<td>0.59</td>
<td>1.77</td>
<td>0.82</td>
<td>1.56</td>
<td>1.43</td>
<td>1.30</td>
</tr>
<tr>
<td>7</td>
<td>8.91</td>
<td>0.90</td>
<td>1.02</td>
<td>0.73</td>
<td>1.02</td>
<td>1.10</td>
<td>1.56</td>
</tr>
<tr>
<td>8</td>
<td>9.44</td>
<td>0.43</td>
<td>1.18</td>
<td>0.42</td>
<td>1.20</td>
<td>0.36</td>
<td>0.88</td>
</tr>
</tbody>
</table>

**Table 6. Manual Evaluation for the Misalignment Before and After Mosaicking**

<table>
<thead>
<tr>
<th>Pair</th>
<th>BeforeReg.</th>
<th>Affine 1</th>
<th>TPS 1</th>
<th>8PP 1</th>
<th>Affine 2</th>
<th>TPS 2</th>
<th>8PP 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>167.74</td>
<td>5.08</td>
<td>4.90</td>
<td>2.24</td>
<td>1.85</td>
<td>0.87</td>
<td>1.08</td>
</tr>
<tr>
<td>2</td>
<td>104.66</td>
<td>1.40</td>
<td>1.21</td>
<td>0.24</td>
<td>1.83</td>
<td>1.55</td>
<td>2.94</td>
</tr>
<tr>
<td>3</td>
<td>109.29</td>
<td>4.08</td>
<td>2.13</td>
<td>0.91</td>
<td>2.08</td>
<td>0.83</td>
<td>1.28</td>
</tr>
</tbody>
</table>
Plate 4. An example of manual evaluation result for registration using CP Pool 1 or CP Pool 2 and affine/TPS/8PP transforms: (a) Corners before registration, (b) Corners after registration (AR), Affine 1, (c) Corners AR, Affine 2, (d) Corners AR, TPS 1, (e) Corners AR, TPS 2, (f) Corners AR, 8PP1, and (g) Corners AR, 8PP2.

thereby achieving higher registration accuracy. The effect from global and local ROI selection is also evaluated. The locally selected ROI can provide more widely spread control points, so the following registration and mosaicking performance can be improved. Another comparison is conducted on different transforms. We discovered that the non-linear mapping such as 8PP and TPS are more capable of correcting the geometric differences due to imprecise camera installation and between-frame misalignment due to the movement of the aerial platform. For the flat terrain in the experiment, the 8PP provides overall the best result. Interestingly, TPS is still useful for flat terrain, and it yields the best results in CPs which may be because it includes more parameters to approximate these corresponding relationships.
The automated algorithms were tested on the multispectral image sequences taken by the mini-UAV system that experiences much vibration, and image registration and mosaicking are completely relied on the software due to the lack of other hardware or global positioning system (GPS) assistance. So, we believe they can be useful for similar small airborne platforms and other larger UAVs with or without hardware stabilization system when monitoring an overall flat terrain lack of distinctive features. Its performance for more complex terrain and the largest between-frame rotation that can be tolerated need to be further investigated.

Acknowledgments
This research was supported by National Aeronautics and Space Administration (NASA) through Stennis Space Center and National Geospatial-Intelligence Agency (NGA). The authors would like to thank the UAV team led by Mr. Skip Wright at AOSI for collecting the data used in the experiments.

References


