Computer vision–based orthorectification and georeferencing of aerial image sets

Mohammad Reza Faraji
Xiaojun Qi
Austin Jensen
Computer vision–based orthorectification and georeferencing of aerial image sets

Mohammad Reza Faraji,a,* Xiaojun Qi,a and Austin Jensenb

aUtah State University, Department of Computer Science, 4205 Old Main Hill, Logan, Utah 84322-4205, United States
bUtah State University, Utah Water Research Laboratory, 8200 Old Main Hill, Logan, Utah 84322-8200, United States

Abstract. Generating a georeferenced mosaic map from unmanned aerial vehicle (UAV) imagery is a challenging task. Direct and indirect georeferencing methods may fail to generate an accurate mosaic map due to the erroneous exterior orientation parameters stored in the inertial measurement unit (IMU), erroneous global positioning system (GPS) data, and difficulty in locating ground control points (GCPs) or having a sufficient number of GCPs. This paper presents a practical framework to orthorectify and georeference aerial images using the robust features-based matching method. The proposed georeferencing process is fully automatic and does not require any GCPs. It is also a near real-time process which can be used to determine whether aerial images taken by UAV cover the entire target area. We also extend this framework to use the inverse georeferencing process to update the IMU/GPS data which can be further used to calibrate the camera of the UAV, reduce IMU/GPS errors, and thus produce more accurate mosaic maps by employing any georeferencing method. Our experiments demonstrate the effectiveness of the proposed framework in producing comparable mosaic maps as commercial software Agisoft and the effectiveness of the extended framework in significantly reducing the errors in the IMU/GPS data. © 2016 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JRS.10.036027]

Keywords: aerial image sets; mosaic process; unmanned aerial vehicle; georeferencing; orthorectification; feature matching; M-estimator sample consensus.

Paper 16374 received May 16, 2016; accepted for publication Aug. 30, 2016; published online Sep. 22, 2016.

1 Introduction

Unmanned aerial vehicles (UAVs), also known as unmanned aircraft systems, as remote sensing platforms provide imagery which maps an area to high-resolution images. The employment of UAVs in civilian applications has increased in recent years due to the availability of accurate and miniature inertial measurement units (IMUs) and global positioning systems (GPSs), along with off-the-shelf high-quality consumer-grade digital cameras.1,2 Some examples for civilian applications include agriculture management,3,4 policing and firefighting,5 architectural recording of archaeological sites,6 earth observation,7,8 rangeland management,9 disaster management,10 and pipeline monitoring.11 Moreover, mostly in agricultural-related cases, periodic flights are scheduled to update maps and keep track of the main subjects such as specific plants, soil, and so on.

The wide varieties of applications have increased the need for robust and accurate orthorectification/georeferencing tools to orthorectify and georeference aerial images and produce a mosaic for an aerial image set. Generally, methods to generate a georeferenced mosaic from an aerial image set are classified into two categories: direct and indirect georeferencing methods. In the following, we briefly review these two categories of georeferencing methods.

Direct georeferencing methods use the respective IMU/GPS data such as pitch, roll, yaw, altitude, latitude, and longitude to georeference each aerial image.12-14 They are easy, fast,
straightforward, and automatic. They first perform a series of transformations to project four corners of each image to the earth coordinate and then create a three-dimensional surface and overlay images on the map using the terrestrial digital elevation model (DEM). These transformations are from the camera frame to the body frame, then the body frame to the local north, east, down (NED) coordinate system shown in the aircraft coordinate system [Fig. 1(a)], and finally, from the NED coordinate system to the earth-centered earth-fixed (ECEF) coordinate system [Fig. 1(b)]. However, these direct georeferencing methods can hardly produce an accurate mosaic map due to the possibly of an off calibrated camera, windy conditions, relatively large errors in exterior orientation (EO) parameters stored in the log, the delay between the camera exposure and logging the aerial data, and inaccurate computational data from the GPS. Figure 2 shows a sample mosaic map generated by the direct georeferencing method using the IMU/GPS data. This mosaic map contains the information from all the images in the aerial image set. However, the two skewed big circular grass areas and the disconnected railways and roads clearly show that each image in the aerial image set is positioned in the neighborhood of the right location of the produced mosaic map. This inaccurate georeferencing result mainly comes from the errors in EO parameters stored in the log. In other words, if the log stores the accurate EO parameters (i.e., pitch, roll, yaw, altitude, latitude, and longitude), the direct

---

Fig. 1 Illustration of two coordinate systems used in the transformation process of direct georeferencing methods. (a) The local NED coordinate system in the aircraft coordinate system and (b) The ECEF and the local NED coordinate systems.

Fig. 2 A mosaic map produced by the direct georeferencing method using the original IMU/GPS data.
georeferencing method is able to produce an accurate mosaic map. Some researchers presented various techniques to obtain more accurate EO parameters. For example, Laliberte et al. propose an iterative process to adjust the EO parameters using tie points and a reference image. Nevertheless, the erroneous EO parameters make it challenging to generate an accurate mosaic map for both the direct georeferencing methods and the indirect georeferencing methods.

Indirect georeferencing methods use ground control points (GCPs) or tie points to georeference images. Most GCPs are manually determined visually detectable and robust targets on the ground. The respective GCPs in the aerial images are identified manually or automatically by the use of computer vision-based tools. Finally, the actual locations of GCPs on the ground and their respective locations in aerial images are used to compute rotation and translation matrices, which can be further used to orthorectify and georeference each image. Indirect georeferencing methods normally outperform direct georeferencing methods due to the use of GCPs. However, setting up GCPs on the ground increases the implementation time and cost. Moreover, it might be impossible to set up GCPs when the area is not easily accessible or it might require a significant number of GCPs to cover the whole area. On the other hand, the GCPs-based indirect georeferencing methods do not normally produce an accurate mosaic map when there is difficulty in locating GCPs points or there are not a sufficient number of GCPs.

To address the shortcomings of GCPs-based indirect georeferencing methods, some researchers propose feature points-based indirect georeferencing methods. To this end, they first use computer vision-based techniques to quickly find reliable and robust feature points in images. They then employ threshold-based feature matching methods to find the best matched features. Two images are stitched together if they have enough matching points. Both images are also orthorectified based on their matched points. Other images in the image set are then stitched to the current stitched image to finally produce a mosaic. Representative features include Lowe’s scale-invariant feature transform (SIFT), features from accelerated segment test (FAST), binary robust independent elementary features (BRIEF), binary robust invariant scalable keypoints (BRISK), speeded up robust features (SURF), Harris & Stephens corner features, and Min Eigen (also known as Shi-Tomasi) features. SIFT generally produces the most reliable and robust features with the highest computational cost. FAST and BRIEF have lower computational cost, but they are not reliable and robust against image distortions and transformations. BRISK and SURF also produce fewer reliable features with lower computational cost. Harris & Stephens and Min Eigen are corner detectors, which mostly focus on corners and edges in an image. The random sample consensus (RANSAC) algorithm or its variant M-estimator sample consensus (MSAC) algorithm is usually used to remove the outliers.

However, the feature points-based indirect georeferencing methods do not normally produce a decent mosaic map when there is considerable image distortion and/or there is difficulty in locating a sufficient number of matching points (a decent mosaic map is defined as a map where all constituent images are geometrically corrected to some extent so the user can visually validate these images as correctly orthorectified). The distortion in a few images may severely affect the orthorectification result since the mosaic at each intermediate step keeps the distortion to a certain level and the final mosaic may have aggregated distortion which makes the entire mosaic less useful. To address the aforementioned challenges associated with orthorectification of the aerial images, we propose a reference image-based orthorectification mosaicking approach to seamlessly incorporate the strength of direct and indirect georeferencing techniques for generating an orthorectified mosaic of decent quality without requiring manual GCPs. A good reference map (i.e., a reference image or an orthophoto, which we refer to as a reference map or sometimes simply as a map) should cover the same area flown by UAV and should have similar texture as the aerial images taken by UAV. Due to the need to update the map and keep track of the main subjects in most civilian applications, UAVs are usually periodically flown around the same investigated area and a good quality reference map is already available. In our process, we first normalize the reference map and the aerial images to have a proper and approximately equal scale. We then use the IMU/GPS data to quickly estimate the location of each image on the reference map. Next, we use feature and/or corner detectors to find important features in each aerial image and its estimated region in the reference map. Finally, we orthorectify and georeference each image by matching its robust and reliable features with the corresponding features in its estimated region in the reference map and finding the projective transformation...
matrix. In addition, we may use some points (e.g., corner points) from the obtained georeferenced and orthorectified images and their respective locations in the camera frame to update the IMU/GPS data, which can be further used to calibrate the camera or to reduce errors in the IMU/GPS data.

It should be noted that there are some commercial software such as EnsoMosaic, Agisoft PhotoScan, and Pix4d in the market to perform photogrammetric processing. For example, EnsoMosaic, which is completely based on photogrammetry, allows full operator control, intervention, and editing, in addition to automatic processing. However, it requires an in-house camera calibration file for input images to accomplish these tasks. Agisoft PhotoScan uses machine vision as well as photogrammetry techniques to generate the orthophoto. It includes four main steps such as aligning and matching photos, generating dense point cloud, generating DEM, and generating orthophoto. Some researchers also use different features-based methods together with one of these software to generate orthophotos. These software can usually produce good-quality mosaics. However, they are not completely automatic, since they need a trained expert to supervise the process to some extent. They require manually labeled GCPs to determine the scale of the images and further improve the alignment and georeference aligned images. More importantly, the process to produce a mosaic (i.e., orthophoto) using these software usually takes a few hours, which makes them unusable for real-time or near real-time applications.

Unlike indirect georeferencing methods and commercial software, we propose a fully automatic orthorectification and georeferencing framework, which is able to produce an orthorectified and georeferenced mosaicking image in near real-time. Our contributions are: (1) Presenting a practical framework to orthorectify and georeference aerial images using the robust feature-based matching method. (2) Automatically georeferencing aerial images without requiring manually labeled GCPs. (3) Automatically georeferencing aerial images in near real-time without generating DEM, which makes it useful for many real applications. (4) Using inverse georeferencing to update the IMU/GPS data, which may consequently be used to calibrate the camera and reduce errors. It should be mentioned that camera calibration, which is generally performed in the labs using grid patterns or in the field using manually located targets, is beyond the scope of this research.

The rest of this paper is organized as follows: Sec. 2 presents our UAV platform, sensors, and image acquisition. Section 3 presents the framework to orthorectify and georeference aerial images and generate mosaic maps. Section 4 explains the process to update the IMU/GPS data using orthorectified and georeferenced images. Section 5 illustrates experimental results for the proposed framework. Finally, Sec. 6 draws the conclusions and presents the directions for future work.

2 Unmanned Aerial Vehicle Platform, Sensors, and Image Acquisition

In this study, AggieAir™, a remote sensing platform developed at Utah State University (USU) for aerial imagery acquisition, is used to obtain our remote sensing data. AggieAir (Fig. 3) is an unmanned aerial system (UAS) designed to carry camera payloads to acquire aerial imagery for ecological and agricultural applications. The wings have a Clark Y airfoil built with EPP foam and wrapped with Kevlar. The fuselage is also constructed with Kevlar and holds the batteries, the payload (imaging system), and the avionics. The maximum takeoff weight of the aircraft is 11 lbs and its maximum flight time is 45 min. The aircraft is propelled using an electric brushless motor and can roll, pitch, and yaw by means of the ailerons, elevator, and rudder using electric servos. AggieAir also has flaps to help slow the aircraft down during landing. The motor and the servos are all controlled by the Paparazzi autopilot. Through the autopilot, the aircraft can be flown autonomously or manually. In the autonomous mode, Paparazzi controls the movement of the aircraft to follow a preprogrammed flight plan. In the manual mode, a human operator controls the aircraft through Paparazzi and the 2.4 GHz RC receiver. An inertial navigation system (INS) estimates the position and orientation of the aircraft and sends this data to the autopilot and the payload. The autopilot uses the inertial data to navigate the aircraft; the payload logs the inertial data every time a picture is exposed. This inertial data, included with every picture, is used after the flight for direct georeferencing.
The AggieAir system flies between 300 and 1000 m above ground. This translates to a ground resolution of 10 to 30 cm. With a 45-min flight time, the AggieAir system can cover between 1 and 3 square miles at each respective altitude. By the end of a 45-min flight, AggieAir will capture 500 pictures from each camera.

3 Proposed Mosaicking Framework

In this section, we explain the proposed framework for mosaicking aerial image sets when a reference map of the area covered by aerial image sets is available. We assume that the world file of the reference map is available. We illustrate each stage of the proposed mosaicking process by showing related results for three sample images in an aerial image set covering a farm in the state of Utah in the US, whose latitude is 39.2200 and longitude is −112.1100. The image set was captured on June 9, 2013, using the AggieAir system by the research team of the Utah Water Research Laboratory, who is responsible for collecting environmental data for scientific applications and natural resources management problems. Figure 4 shows three sample images from this image set, where the two images shown in Figs. 4(a) and 4(b) cover a portion of the same area.

Two sample orthorectified and georeferenced mosaic images are also shown in Fig. 5. The image shown in Fig. 5(a) is the mosaic image produced by experts using the Agisoft PhotoScan software from the aerial image set captured on June 9, 2013. Another mosaic image shown in Fig. 5(b) is produced by the same software with the aid of experts from the aerial image set captured on June 17, 2013. Both aerial image sets were obtained by flying the same UAV over the same farm of Utah using the same settings. It clearly shows the significant field changes at the same farm over a short period of time (e.g., 8 days). Since the two sample mosaic images...
are of high quality, we can use either of them as the reference map to orthorectify and georeference the aerial images of the same area captured at the same or different times. To facilitate discussion, we use the mosaic image shown in Fig. 5(a) as the reference map to illustrate the important intermediate results of orthorectifying and georeferencing the three sample images shown in Fig. 4 and the final mosaic map generated from the entire aerial image set using the proposed framework.

The proposed mosaicking framework seamlessly incorporates the strength of direct and indirect georeferencing techniques for generating an orthorectified mosaic map of decent quality without requiring manual GCPs. It uses a good quality map covering the same area flown by UAV and having a similar texture as the aerial images taken by UAV as the reference. Specifically, it first makes the reference map and the aerial images have a proper and approximately equal scale. It then uses the IMU/GPS data to quickly estimate the location of each aerial image on the reference map. Feature and/or corner detectors are then employed to find important features in each aerial image and its estimated region in the reference map. Finally, it orthorectifies and georeferences each image by matching its robust and reliable features with the corresponding features in its estimated region in the reference map and finding the projective transformation matrix. The georeferenced images can be further used to compute the position and orientation of the camera and consequently to update the IMU/GPS data. Figure 6 shows the framework of the proposed mosaicking process together with the IMU/GPS update process.

The five important steps of the proposed reference map-based orthorectification mosaicking method are summarized as follows:

Step 1: Normalize the reference map to have a similar scale as the aerial images taken by UAVs.
Step 2: Utilize the IMU/GPS information of each aerial image to quickly estimate its corresponding pixel location in the reference map.
Step 3: Find important feature points for each aerial image and its estimated region in the reference map using the MinEigen corner feature detector.
Step 4: Identify the robust and reliable feature points in each aerial image that match with the feature points in its estimated region using the nearest neighbor ratio matching method and the MSAC method.
Step 5: Apply the projective transformation on the matched feature points in each pair of an aerial image and its estimated region to orthorectify the aerial image.

Fig. 5 High quality orthorectified and georeferenced mosaic images for the farming area produced by experts using aerial images captured on (a) June 9, 2013, and (b) June 17, 2013.
In the following, we explain each of these five steps in detail.

Step 1: **Normalize the reference map to have a similar scale as the aerial images taken by UAVs.** This step aims to resize aerial images and the reference map to have a proper size and approximately equal scale so the later matching process likely finds more robust and reliable matching features. We use the world file of the reference map and the IMU/GPS data associated with the aerial image sets to perform the normalization operation. The world file contains the universal transverse mercator (UTM) coordinates (i.e., UTM easting and UTM northing) of the upper-left corner of the reference map and also the scale of the reference map. The UTM coordinates for the lower-right corner of the map can be simply computed using the coordinates of the upper-left corner, the size, and the scale of the map. The IMU/GPS data contains the actual coordinates of the UAV when it takes the image and logs the data. These coordinates can be considered as an estimation of the location (i.e., latitude and longitude) of the center pixel of the aerial image. These latitude and longitude coordinates are then converted to their respective UTM easting and UTM northing coordinates. We also have the scale for each aerial image, which can be considered as the same for all aerial images, assuming the UAV flies at a fixed altitude when taking images. The original size of the three sample images is $2736 \times 3648$ pixels and their scales are $-0.15$ and $0.15$ in the east and north directions. The UTM coordinates of the center pixel, the size, and the scale of each aerial image can finally be used to compute the UTM coordinates of the upper-left and lower-right corners of each image.

To normalize the reference map and aerial images to a similar scale, we first decide the new size of the reference map by considering the complexity of each mosaicking step, the desired accuracy, and the estimated run-time acceptable for the users. We then resize the reference map and all the aerial images accordingly. For example, we may set the new size for the reference map as $1/6$ of the original size (i.e., $1/6$ in each direction) to reduce the computational time by at least six fold. This resizes the reference map shown in Fig. 5(a) with the original size of $13527 \times 11073$ pixels to the new size of $2255 \times 1846$ pixels. We then compute the new scale for the reference map using the new pixel size and the coordinates of its upper-left and lower-right corners. Specifically, the east scale of the reference map can be computed by dividing the difference of the UTM easting coordinates between the lower-right and upper-left corners by the width of the map. Similarly, the north scale of the map can be computed by dividing the difference of the UTM northing coordinates between the lower-right and upper-left corners by the height of the map. Following these steps, the new east and north scales for the reference map shown in Fig. 5(a) are $-0.8998$, and $0.8998$, respectively. Since aerial images need to have the same scale as the map, we compute the new pixel size for each image using the new scale of the map and the UTM coordinates of the upper-left and lower-right corners of the image. Finally, we resize the images to the newly computed normalized size. For the three sample images, the new size is $457 \times 609$ pixels. As a result, the reference map and the images have a proper size and equal scale.

**Fig. 6** Framework of the proposed mosaicking process together with the IMU/GPS update process.
Step 2: Utilize the IMU/GPS information of each aerial image to quickly estimate its corresponding pixel location in the reference map. This step aims to quickly estimate the location of each aerial image in the reference map so the later matching process focuses on matching features in the aerial image and features in the estimated region of the reference map. Due to the delay between the camera exposure and logging the UAV data and inaccurate computational data from the GPS, it is impossible to accurately compute the location of each aerial image in the reference map. To quickly estimate the location of each aerial image, we directly compute the UTM coordinate of the center of the image in the reference map and use this UTM coordinate as the center of the estimated region. The size of the estimated region is the same as the resized aerial image obtained in Step 1. It should be mentioned that we also evaluated several neighboring estimated regions by shifting the center of the estimated region along the eight directions (east, north-east, north, north-west, west, south-west, south, and south-east) by 5 pixels, 10 pixels, 15 pixels, and 20 pixels, respectively. The experimental results showed the estimated region, whose center corresponds to the UTM coordinate of the center of the image, achieves a comparable performance among all the 32 evaluated neighboring estimated regions. Figure 7 shows the three sample aerial images and their corresponding estimated regions in the reference map in three rows. For example, Figs. 7(a) and 7(b) present a pair of an aerial image and its estimated region in the reference map. Figures 7(c) and 7(d) present another pair of an aerial image and its estimated region in the reference map. A small portion of the circular green grass area and the railway show up in both pairs of the aerial image and its estimated region. However, the oval shaped parking lot and the narrow road leading to the parking lot in Fig. 7(c) do not show up in its estimated region. Similarly, the irregular shaped muddy area on the left side of Fig. 7(e) does not show up in its estimated region as shown in Fig. 7(f). These clearly indicate that the estimated regions are not accurate since they contain some of the important objects in the paired aerial images while missing some other important objects. However, a sufficient number of feature points can be located in the common important objects presented in each pair for later processing. Therefore, instead of searching for the best matched region in the neighborhood of the estimated location, we suggest using the quickly estimated region.

Step 3: Find important feature points for each aerial image and its estimated region in the reference map using the MinEigen corner feature detector. This step aims to find a sufficient number of important feature points in each pair of an aerial image and its respective estimated region so the later matching process is able to find enough matching feature points in the pair to compute the projective transformation matrix. We choose the MinEigen corner feature detector over the SIFT descriptor for the following two reasons: (1) It has significantly less computational cost than the SIFT descriptor; and (2) it is able to achieve comparable performance as the SIFT descriptor when the image is of decent quality. In other words, the MinEigen corner detector is capable of quickly finding a sufficient number of important feature points in an image of decent quality. The identified feature points for the three pairs of aerial images and their corresponding estimated regions are also shown in blue dots in Fig. 7. The number of feature points detected by MinEigen and SIFT descriptors on six images shown in Fig. 7 and their corresponding computational times are summarized in Table 1. It clearly shows that the MinEigen corner feature detector can identify a sufficient number of feature points in each image in a significantly less computational time when compared with the SIFT detector.

Step 4: Identify the robust and reliable feature points in each aerial image that match with the feature points in its estimated region using the nearest neighbor ratio matching method and the MSAC method. This step aims to find the robust and reliable matching feature points in each pair of an aerial image and its respective estimated region. We use the nearest neighbor ratio matching method to match identified feature points in each pair of images. For each feature point in an aerial image, the nearest neighbor ratio matching method first finds potential matching feature points in its estimated region. If the distance (e.g., difference) between the feature vector of the investigated feature point in an aerial image and the feature vector of a feature point in its estimated region is less than a specified threshold, the two feature points are considered as potentially matched. It then ranks the matched feature points based on the distance. The smaller the distance, the more similar the two feature points. The best matched feature point in the estimated region is identified if it further passes the ratio test. Otherwise, no matched feature point in the estimated region is found for the investigated feature point in the aerial image. Specifically, the nearest neighbor ratio matching method computes the ratio of two distances: one distance is the difference
between the feature vector of the investigated feature point and the feature vector of the best matched feature point and the other distance is the difference between the feature vector of the investigated feature point and the feature vector of the second best matched feature point. If this ratio is less than a predefined threshold, it indicates that the best matched feature point is much more similar to the feature point in the aerial image than the second best matched feature point and, therefore, should be kept. Otherwise, it indicates that the top two matched features points have a similarity to the feature point in the aerial image, therefore, there is no best matched feature point in the estimated region. Finally, we use the MSAC method, an improved version of the RANSAC method, to remove outliers (i.e., ambiguous pairs of feature points) by evaluating the quality of the consensus set formed by the pairs of the matched feature points. Different from the RANSAC method, which does not consider a score for each inlier and considers a constant penalty for each outlier, the MSAC method uses the redescending M-estimators to score inliers on how well they fit the data and still gives outliers a constant penalty. This significantly benefits all robust estimations without

Fig. 7 Three sample images shown in Fig. 4 [(a), (c), and (e)] and their estimated regions in the reference map [(b), (d), and (f)] together with their MinEigen corner feature points shown as blue dots. Note: images shown in (a), (c), and (e) are displayed and processed as they are captured by cameras.
any additional computational cost. Figure 8 shows the three sample aerial images and their corresponding estimated regions together with the final valid matching feature points, where the yellow line connects a robust and reliable feature point in the aerial image with its matching feature point in its estimated region after removing the outliers. The number of valid matching feature points is also listed for each pair of an aerial image and its estimated region in the reference map. It clearly shows that only a small portion of the feature points shown in Fig. 7 are considered as the robust and reliable feature points and thus are used for matching. Specifically, 281, 238, and 118 robust and reliable pairs of feature points are, respectively, found for three pairs of sample aerial images and their estimated regions shown in Fig. 7 after applying the nearest neighbor ratio matching method and the MSAC method on the feature points identified by the MinEigen corner feature detector. On the contrary, 462, 441, and 247 robust and reliable pairs of feature points are, respectively, found for the same three pairs after applying the same method on the features points identified by the SIFT feature detector. This verifies that the SIFT feature detector is able to locate more reliable and robust matching points than the MinEigen corner detector with a significantly higher computational cost. However, we suggest using the MinEigen corner detector in the proposed framework due to its effectiveness in finding enough matching points in an image of reasonable resolution with a lower computational cost.

Step 5: Apply the projective transformation on the matched feature points in each pair of an aerial image and its estimated region to orthorectify the aerial image. This step aims to find the optimal projective transformation matrix which is able to transform each robust and reliable feature point in an aerial image to its corresponding matching feature point in its estimated region in the reference map. We use the projective geometry on all pairs of the matching feature points to estimate rotation and translation matrices. The projective geometry requires at least four pairs of matching feature points and supports translation, rotation, isotropic and nonisotropic scaling, and tilting. The obtained projective transformation has advantages over the affine transformation by not preserving parallelism, length, and angle (i.e., tilting), which makes it suitable for image orthorectification.

Following geometry transformation techniques, the projective transformation $H$ transfers a set of two-dimensional points $P_i = (x_i, y_i, 1)^T$, $i = 1, \ldots, n$ and $n \geq 4$, to their corresponding points $P’_i = (x’_i, y’_i, 1)^T$ by the equation $P’_i = HP_i$. This equation can be expressed as $P’_i \times HP_i = 0$, where “$\times$” denotes the cross product. Let $h^T_j$ denote the $j$'th row in the matrix $H$, where $T$ indicates the transpose operation, then we have

$$HP_i = \begin{bmatrix} h^T_1 P_i \\ h^T_2 P_i \\ h^T_3 P_i \end{bmatrix}. \quad (1)$$

As a result, the cross product in $P’_i \times HP_i = 0$ is computed by

Table 1 The number of feature points detected by MinEigen and SIFT and their corresponding computational time in seconds shown inside the parentheses together with the number of valid matching pairs after applying the matching and MSAC method for three sample images and their respective estimated regions in Fig. 7.

<table>
<thead>
<tr>
<th>Figure/Method</th>
<th>MinEigen #Features(time)</th>
<th>Matched pairs</th>
<th>SIFT #Features(time)</th>
<th>Matched pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 7(a)</td>
<td>5340 (.3207)</td>
<td>281</td>
<td>1259 (3.1508)</td>
<td>462</td>
</tr>
<tr>
<td>Fig. 7(b)</td>
<td>5465 (.2947)</td>
<td>1169</td>
<td>1169 (2.9112)</td>
<td></td>
</tr>
<tr>
<td>Fig. 7(c)</td>
<td>3597 (.2765)</td>
<td>238</td>
<td>1251 (3.2594)</td>
<td>441</td>
</tr>
<tr>
<td>Fig. 7(d)</td>
<td>4814 (.2688)</td>
<td>1294</td>
<td>1294 (3.3543)</td>
<td></td>
</tr>
<tr>
<td>Fig. 7(e)</td>
<td>4191 (.3343)</td>
<td>118</td>
<td>916 (2.3909)</td>
<td>247</td>
</tr>
<tr>
<td>Fig. 7(f)</td>
<td>4163 (.2776)</td>
<td>807</td>
<td>807 (2.1622)</td>
<td></td>
</tr>
</tbody>
</table>


Fig. 8 Illustration of valid matching pairs of robust and reliable MinEigen feature points in the same three aerial images shown in Fig. 4 (left) and their estimated regions (right) after removing outliers using the nearest neighbor ratio matching method and the MSAC method. The number of valid matching robust and reliable MinEigen feature points in each pair of three sample aerial images and their estimated regions is (a) 281; (b) 238; and (c) 118.

Since $h^T_{ij}P_i = P^T_{ij}h^j \ (j = 1, \ldots, 3)$, Eq. (2) can be written as follows

$$P^i_j \times HP_i = \begin{bmatrix} y_i'h^{3T}P_i - h^{2T}P_i \\ h^{1T}P_i - x_i'h^{3T}P_i \\ x_i'h^{2T}x_i - y_i'h^{1T}P_i \end{bmatrix} = 0.$$  \tag{2}$$

Since $h^T_{ij}P_i = P^T_{ij}h^j \ (j = 1, \ldots, 3)$, Eq. (2) can be written as follows

$$\begin{bmatrix} 0^T \\ P^T_{ij} \\ -y_i'^T P_i \end{bmatrix} \begin{bmatrix} h^1 \\ h^2 \\ h^3 \end{bmatrix} = 0.$$  \tag{3}$$
Eq. (3) has the form of $A_i h = 0$, where $A_i$ is a $3 \times 9$ matrix and $h$ is a vector with nine elements from the matrix $H$

$$h = \begin{bmatrix} h^1 \\ h^2 \\ h^3 \end{bmatrix}, \quad H = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix},$$

with $h_i$ being the $i$'th element of $h$ [i.e., $h^1 = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}$, $h^2 = \begin{bmatrix} h_4 \\ h_5 \\ h_6 \end{bmatrix}$, and $h^3 = \begin{bmatrix} h_7 \\ h_8 \\ h_9 \end{bmatrix}$]. The first two rows in $A_i$ are linearly independent and the third row can be obtained from the sum of $x_i'$ times the first row and $y_i'$ times the second row. So, the third equation can be omitted and Eq. (3) is simplified as follows:

$$\begin{bmatrix} 0^T & -P_i^T & y_i'P_i^T & h^1 \\ P_i^T & 0^T & -x_i'P_i^T & h^2 \\ & & & h^3 \end{bmatrix} = 0.$$  

Eq. (5) can be written as $E_i h = 0$, where $E_i$ is the $2 \times 9$ matrix. As a result, if we concatenate $E_i$ for each of $n$ feature points ($1 \leq i \leq n$), we have a single $2n \times 9$ matrix $E$, satisfying the following condition

$$Eh = \begin{bmatrix} 0 & 0 & 0 & -x_1 & -y_1 & -1 & y_1'x_1 & y_1'y_1 & y_1' \\
 x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1'x_1 & x_1'y_1 & x_1' \\
 . & . & . & . & . & . & . & . & . \\
 x_i & y_i & 1 & 0 & 0 & 0 & -x_i'x_i & x_i'y_i & x_i' \\
 . & . & . & . & . & . & . & . & . \\
 0 & 0 & 0 & -x_n & -y_n & -1 & y_n'x_n & y_n'y_n & y_n' \\
 x_n & y_n & 1 & 0 & 0 & 0 & -x_n'x_n & x_n'y_n & x_n' \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ h_9 \end{bmatrix} = 0.$$  

A solution for $Eh = 0$ can be obtained by minimizing $\|Eh\|$ subject to $\|h\| = 1$. Suppose that the singular value decomposition of the matrix $E$ is represented by $USV^T$, where $U$ and $V$ are the unitary matrices and $S$ is the diagonal matrix with nonnegative real values in the descending order, minimizing $\|Eh\|$ is equivalent to minimizing $\|USV^T h\|$ subject to $\|h\| = 1$. Since $\|USV^T h\| = \|SV^T h\|$ and $\|h\| = \|V^T h\|$,

setting $d = \|V^T h\|$ transfers the problem to minimizing $\|Sd\|$ subject to $\|d\| = 1$, which results in $d = (0, 0, \ldots, 0.1)^T$. $h$ can then be computed by $\|Vd\|$. In other words, the last column of $V$ (i.e., nine values in the last column) gives the solution $h$.

Using all pairs of matching feature points between an aerial image and its estimated region, the nine transformation coefficients are estimated. All pixels of an aerial image are then transformed using the projective transformation matrix to their orthorectified locations. In this way, an aerial image is orthorectified. Figures 9(a) and 9(b) show the three sample aerial images shown in Fig. 4 and their corresponding orthorectified images, respectively.

To georeference an aerial image on the mosaic map which has the same dimension as the reference map, we can simply add the actual location of the upper-left corner of its estimated region in the reference map to every matching feature point in the region. In other words, the pixel location of $(p, q)$ of every matching feature point in the estimated region can be added by the pixel coordinate of the upper-left corner of the estimated region to obtain its location in the reference map. Using all pairs of matching feature points (i.e., robust and reliable feature points in an aerial image and the georeferenced matching feature points in the reference map), the nine transformation coefficients in the projective transformation matrix can be estimated. All pixels of
an aerial image are then transformed using the projective transformation matrix to their ortho-
rectified and georeferenced locations in the mosaic map. Figure 9(c) shows the three ortho-
rectified and georeferenced images of the three sample images in the mosaic map.

It should be noted that we project aerial images to their corresponding orthorectified and
georeferenced images using the projective transformation and finally mosaic them together
without generating DEM. Our mosaicking process removes the effects of image tilt and
makes the scale of the image uniform so that the location of each new rectified pixel represents
its true location. The new rectified features or pixels can now be used for different analyses
including measuring true distances. In other words, the aerial images are orthorectified and
georeferenced using the feature matching and the projective transformation method together
with a reference map.

4 Update the Inertial Measurement Unit/Global Positioning System Data

The orthorectified and georeferenced images can be used to compute the position and orientation
of the camera and consequently to find errors in the IMU/GPS data. The method is also called the
inverse georeferencing. Finding errors and calibrating the camera result in more accurate
IMU/GPS data, which obviously benefit both direct and indirect georeferencing methods to
achieve better performance.

We employ the general procrustes analysis (GPA) method to update the IMU/GPS data. The GPA method is a statistical algorithm for point-fitting. It computes a linear transformation of
a set of \( n \) points (e.g., the set \( A \)) to best conform them to another set of \( n \) points (e.g., the set \( B \)).
The transformation includes rotation, translation, reflection, and scaling and is computed by
minimizing the sum of squared errors between two sets of \( n \) points (i.e., the set \( A \) and the
set \( B \)). The GPA method first calculates the respective centroid of \( A \) and \( B \). It then computes
\( d_A^k(i) \) and \( d_B^k(i) \) (\( i = 1, 2, \ldots, n \)), the distances from each point \( i \) in \( A \) and \( B \) to its corresponding
centroid, respectively. Next, the GPA method solves the least squares problem by minimizing
\( \sum_{i=1}^{n} ||d_A^k(i) - Rd_B^k(i)||^2 \), where \( R \) is the transformation matrix to be found. Finally, the translation
vector \( T \) is simply computed by \( T = p_A - Rp_B \), where \( p_A \) and \( p_B \) are a pair of matching

---

**Fig. 9** Illustration of orthorectified and georeferenced images for the three sample images, which
are transformed by the projective transformation matrix. (a) Three sample aerial images shown in
Fig. 4; (b) their respective orthorectified images; and (c) their respective orthorectified and geore-
ferenced images.
points from sets $A$ and $B$. In the proposed framework, we compute the linear transformation from feature points (e.g., four corner points) in the camera frame (i.e., the set $A$) [Fig. 1(a)] to their respective matching feature points in the orthorectified and georeferenced image (i.e., the set $B$). The obtained translation vector directly gives updated UTM northing and easting and altitude. The obtained rotation matrix $R$ is a $3 \times 3$ matrix as follows:

$$
R = \begin{pmatrix}
  r_{11} & r_{12} & r_{13} \\
r_{21} & r_{22} & r_{23} \\
r_{31} & r_{32} & r_{33}
\end{pmatrix}.
$$

The matrix $R$ is used to update roll, pitch, and yaw using the following equation:

$$
\text{roll} = a \times \tan^2(r_{23}, r_{33}), \quad \text{pitch} = \tan\left(\frac{r_{13}}{r_{33}}\right), \quad \text{yaw} = \tan\left(\frac{r_{12}}{r_{11}}\right).
$$

Figure 10 shows the three sample aerial images shown in Fig. 4 and compares their corresponding georeferenced images produced by applying the direct georeferencing method on the original and updated IMU/GPS data, respectively. It clearly shows that the direct georeferencing method is able to produce more accurate georeferencing results using the updated IMU/GPS data. In other words, the IMU/GPS data updated by the GPA method are more accurate.

## 5 Experimental Results

We perform three experiments to demonstrate the effectiveness of the proposed framework. In the first experiment, we use either of the two high-quality mosaic maps shown in Fig. 5 as the reference map to automatically generate the mosaic maps for two aerial image sets of the same area, which were taken on June 9, 2013, and June 17, 2013, respectively. In the second experiment, we use the generated mosaic maps of the two aerial image sets to automatically update their IMU/GPS data and apply the direct georeferencing method on the updated IMU/GPS data to automatically generate their corresponding mosaic maps. In the third experiment, we use an appropriate 2014 national agriculture imagery program (NAIP) image of the same farm area of the state of Utah as the reference map to automatically generate mosaic maps for the two aerial image sets.

![Fig. 10](image)

**Fig. 10** Comparison of the direct georeferencing results on three sample images using original and updated IMU/GPS data. (a) Three sample aerial images; (b) their georeferenced images produced using original IMU/GPS data; and (c) their georeferenced images produced using updated IMU/GPS data. All georeferenced images are inferior to the georeferenced images (shown in Fig. 5) produced by Agisoft OrthoPhoto software through visual inspection, where the two red circles in (b) highlight the locations where the original IMU/GPS data fail and the red circle in (c) indicates the updated IMU/GPS data are not perfectly accurate.
5.1 The First Experiment

The aim of this experiment is to demonstrate the performance of the proposed framework and evaluate the influence of high-quality reference maps of the same area with the same or different textures as the aerial image set on the mosaicking results. To this end, we first use the high-quality reference map of the same area with the same texture as the aerial image set to generate the mosaic maps. Specifically, we use the reference map of the same date as the aerial image set to generate the mosaic maps for two aerial image sets. In other words, we use the reference map on June 9, 2013, to generate the mosaic map for the aerial image set taken on the same date. We use the reference map on June 17, 2013, to generate the mosaic map for the aerial image set taken on the same date. The two mosaic maps are shown in Figs. 11(a) and 12(a), respectively. Next, we use the reference map of the different date as the aerial image set to generate the mosaic maps for two aerial image sets. In other words, we use the reference map on June 17, 2013, to generate the mosaic map for the aerial image set taken on a different date, June 9, 2013. We use the reference map on June 9, 2013, to generate the mosaic map for the aerial image set taken on a different date, June 17, 2013. The two mosaic maps are shown in Figs. 11(b) and 12(b), respectively. The mosaicking results in these figures confirm that the proposed framework is able to produce the best mosaic maps using the high-quality reference maps of the same date and the comparable mosaic maps using the high-quality reference maps of the different date since the generated mosaic maps are identical to the mosaic maps produced by the Agisoft PhotoScan software with the aid of experts.

5.2 Accuracy Assessment

Previously, we have six GCPs available for the imagery of the target farm flown by the UAV. The six GCPs were marked using 1m × 1m polyvinyl chloride targets with an iron cross pattern, and their accurate positions were recorded using an real-time kinematic GPS system. However, they are not sufficient enough to assess the accuracy. Therefore, we manually locate a few features visible on all mosaics and reference maps and use the location information of these visible features to assess the accuracy of generated mosaics. To be consistent in evaluating the accuracy, we will not consider the six GCPs in our calculations and will only focus on visible features. We then

![Fig. 11 The mosaic maps generated from aerial images captured on June 9, 2013, using a reference map of a high-quality image (a) on the same date and (b) on a different date, June 17, 2013.](http://remotesensing.spiedigitallibrary.org/)
use the ArcMap 10.3.1 software to pull out the location of each visible point, which may produce its own inherent errors. To this end, we import the Bing map as our base map in ArcMap and pull out the UTM coordinates of 18 points (which are visible on both the reference maps and the generated mosaics) from the Bing map using the ArcMap software. Figure 13 shows the locations of the 18 points on the Bing map for the investigated area. Each visible point in ArcMap is not a specific pixel to be extracted for each reference map or mosaic and there are errors in selecting an exact point in ArcMap and errors in recording the UTM coordinates. To address these issues, we select the location of each visible point five times in a row and consider their average as the ground truth UTM coordinate of the point.

We assess the accuracy of the generated mosaics using horizontal accuracy assessment (HAA). HAA is one of the statistical and testing methodologies implemented by the national standard for spatial data accuracy to estimate the positional accuracy of points on maps and in digital geospatial data, with respect to georeferenced ground positions of higher accuracy. We consider the location of each point on the Bing map as the highest accuracy and compare the accuracy of each point in our mosaics with the corresponding point in the Bing map. We also compute the horizontal accuracy for the mosaics produced using the Agisoft PhotoScan software, since they are used as reference maps in our mosaicking process. Tables 2 and 3, respectively, summarize the accuracy of three mosaics (i.e., the mosaic produced by Agisoft PhotoScan software and the two mosaics generated from aerial image sets captured on June 9 of 2013, and June 17 of 2013) in terms of root mean square errors (RMSE) with respect to the Bing map in ArcMap. In both tables, the first row presents the accuracy of the mosaic produced by Agisoft PhotoScan software. The second row presents the accuracy of the mosaic map generated by the proposed method, which uses the reference map of the same date produced by Agisoft PhotoScan. The third row presents the accuracy of the mosaic map generated by the proposed method, which uses the reference map of a different date produced by Agisoft PhotoScan. All RMSE errors are given in the scale 1:200. As expected, the reference maps produced by Agisoft PhotoScan software achieve the best accuracy (i.e., smallest errors) and our mosaics achieve comparable accuracy by exclusively using the feature matching techniques. However, as we mentioned, it usually takes a few hours for Agisoft PhotoScan software and similar software to produce a mosaic whereas it takes the proposed method twenty minutes to produce a
Fig. 13 Illustration of 18 manually chosen visible points on the Bing map. The symbol “+” shows the exact locations of the points. The UTM coordinates are pulled out using the ArcMap software. The UTM coordinates for the first point are shown in the rounded rectangular.

Table 2  Horizontal accuracy in the scale 1:200 for mosaics produced from the image set captured on June 9, 2013, with respect to the Bing map in ArcMap.

<table>
<thead>
<tr>
<th>Method</th>
<th>Scale: 1:200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agisoft PhotoScan</td>
<td>1.39</td>
</tr>
<tr>
<td>The proposed method (using the reference image from June 9)</td>
<td>1.50</td>
</tr>
<tr>
<td>The proposed method (using the reference image from June 17)</td>
<td>2.08</td>
</tr>
</tbody>
</table>

Table 3  Horizontal accuracy in the scale 1:200 for mosaics produced from the image set captured on June 17, 2013, with respect to the Bing map in ArcMap.

<table>
<thead>
<tr>
<th>Method</th>
<th>Scale: 1:200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agisoft PhotoScan</td>
<td>1.70</td>
</tr>
<tr>
<td>The proposed method (using the reference image from June 17)</td>
<td>1.99</td>
</tr>
<tr>
<td>The proposed method (using the reference image from June 9)</td>
<td>2.24</td>
</tr>
</tbody>
</table>
comparable mosaic. This real-time or near real-time orthorectification and georeferencing makes
the proposed method appealing to a lot of applications.

We also compare the horizontal accuracy of our mosaics with the Agisoft reference maps
since these reference maps are used to produce our mosaics. Tables 4 and 5, respectively, com-
pare the accuracy of the mosaics generated from aerial image sets captured on June 9 of 2013 and
June 17 of 2013 with the Agisoft reference maps from June 9 of 2013 and June 17 of 2013. It
clearly demonstrates that using the Agisoft orthophoto of the same date as the reference map is
able to produce a mosaic map of higher accuracy than using the Agisoft orthophoto of a different
date as the reference map.

5.3 The Second Experiment

The aim of this experiment is to demonstrate the performance of the proposed IMU/GPS updat-
ing method and evaluate the influence of IMU/GPS data on the mosaicking results. To this end,
we use the mosaic maps generated by the reference map of a different date to update the IMU/
GPS data. Specifically, we use the mosaic map shown in Fig. 11(b) to update the IMU/GPS data
of the aerial image set captured on June 9, 2013, and use the mosaic map shown in Fig. 12(b) to
update the IMU/GPS data of the aerial image set captured on June 17, 2013. Figures 14(a) and
14(b) show the mosaic maps for the aerial image set captured on June 9, 2013, generated by
applying the direct georeferencing method on the original and updated IMU/GPS data, respec-
tively. Figures 15(a) and 15(b) show the mosaic maps for the aerial image set captured on June
17, 2013, generated by applying the direct georeferencing method on the original and updated
IMU/GPS data, respectively. These figures clearly demonstrate that the mosaic maps generated
by using the updated IMU/GPS data have a much better georeferencing quality than the mosaic
maps generated by using the original IMU/GPS data. They also indicate the noisy nature of the
IMU/GPS data and the necessity to update these data to further improve any georeferencing
method.

5.4 The Third Experiment

The aim of this experiment is to demonstrate the performance of the proposed framework and
evaluate the influence of decent publicly available reference maps of the same area with different
textures as the aerial image set on the mosaicking results. For most georeferencing tasks, a high-
quality reference map may not be available. However, we normally have access to imagery from
different sources such as NAIP and Google Earth. In this experiment, we use an appropriate
image from the NAIP imagery as the reference map when a high-quality reference map of
the same area flown by UAVs is unavailable. The NAIP imagery is administered by the United
State Department of Agriculture’s Farm Service Agency through the Aerial Photography Field

Table 4  Horizontal accuracy in the scale 1:200 for mosaics produced from the image set captured
on June 9, 2013, with respect to the Agisoft reference map from the same date in ArcMap.

<table>
<thead>
<tr>
<th>Method</th>
<th>Scale: 1:200</th>
</tr>
</thead>
<tbody>
<tr>
<td>The proposed method (using the reference image from June 9)</td>
<td>0.97</td>
</tr>
<tr>
<td>The proposed method (using the reference image from June 17)</td>
<td>2.03</td>
</tr>
</tbody>
</table>

Table 5  Horizontal accuracy in the scale 1:200 for mosaics produced from the image set captured
on June 17, 2013, with respect to the Agisoft reference map from the same date in ArcMap.

<table>
<thead>
<tr>
<th>Method</th>
<th>Scale: 1:200</th>
</tr>
</thead>
<tbody>
<tr>
<td>The proposed method (using the reference image from June 17)</td>
<td>1.59</td>
</tr>
<tr>
<td>The proposed method (using the reference image from June 9)</td>
<td>2.16</td>
</tr>
</tbody>
</table>
Fig. 14 The mosaic maps generated from the aerial image set captured on June 9, 2013, by applying the direct georeferencing method on two kinds of data: (a) its original IMU/GPS data and (b) its updated IMU/GPS data.

Fig. 15 The mosaic maps generated from the aerial image set captured on June 17, 2013, by applying the direct georeferencing method on two kinds of data: (a) its original IMU/GPS data and (b) its updated IMU/GPS data.
Office in Utah, USA. NAIP acquires the aerial imagery during the agricultural growing seasons and makes digital ortho-photography available to governmental agencies and the public. Figures 16(a)–16(f) show sample images from the same farm at the state of Utah for six different years. They clearly show the texture of the farm has been changed over years, especially when comparing to the mosaic images shown in Figs. 5(a) and 5(b).

After evaluating the quality of the mosaic maps generated from different NAIP maps, we use the most recent 2014 NAIP image shown in Fig. 16(f) as the reference map to generate mosaic maps for the two aerial image sets of the same farming area, which were taken on June 9, 2013, and June 17, 2013. Since the NAIP reference map has a lower resolution than the high-quality reference map used in the first experiment, a few modifications are suggested to be incorporated into the proposed mosaicking framework. First, the SIFT feature detector should be employed to find more feature points due to its robustness in locating feature points in images with different resolutions. Second, the average filtering should be applied to each aerial image to smooth the image and blur its details since the reference map contains fewer details than the aerial images. This blurring operation is necessary since it increases the chances for the matching method to find more reliable and robust matching feature points. Figures 17(a) and 17(b) show the two mosaic maps generated by applying the proposed mosaicking method with the two aforementioned modifications on the aerial image sets taken on June 9, 2013, and June 17, 2013, respectively. It is clear that these two mosaic maps, which are produced by using the low-quality NAIP image as the reference map, are not as good as the mosaic maps shown in Figs. 11 and 12, which are produced by using the high-quality images as the reference maps. Furthermore, information from some aerial images are missing since there are no matching feature points or insufficient matching feature points for finding the projective transformation matrix. However, these two mosaic maps still provide important information for researchers.

![Fig. 16 NAIP images for the same farm at the state of Utah for six different years: (a) 1990; (b) 2004; (c) 2006; (d) 2009; (e) 2011; and (f) 2014.](image_url)
6 Conclusions and Future Work

We present a practical five-step framework to orthorectify and georeference aerial images using the robust features-based matching method. The proposed framework is fully automatic and does not require manually labeled GCPs to georeference aerial images. Unlike commercial software such as Agisoft PhotoScan and EnsoMosaic, which take a few hours to produce a mosaic with the aid of the experts, the proposed framework is able to produce a comparable mosaic image in near real-time as long as there is an available reference image with similar or slightly different texture as the aerial images. The framework further employs inverse georeferencing to update the IMU/GPS data, which can be used to calibrate the camera of the UAV, reduce IMU/GPS errors, and thus produce more accurate mosaic maps by employing any georeferencing method. Our experiments show that the proposed framework is capable of producing comparable mosaic maps as commercial software when the reference image has similar or slightly different texture as the aerial images. They also verify that the effectiveness of the extended framework in significantly reducing the errors in the IMU/GPS data.

In the future, we will focus on decreasing the computational time to propose a real-time framework. We will further improve the mosaic results when a few NAIP, Google Earth, or other imagery sources are used as reference images, which have quite different textures as the aerial images. Finally, we will investigate the feasibility to propose a framework that can produce accurate mosaic results when there is no imagery available to be used as a reference image.

Acknowledgments

Funding for this research was provided by the Utah Water Research Laboratory, USU, Logan, Utah.

References


Mohammad Reza Faraji received his BS degree in applied mathematics from Zanjan University in 2005 and his MS degree in systems engineering from Amirkabir University of Technology in 2009. He received his PhD in 2015 in computer science at Utah State University. He is currently a postdoctoral researcher with the Computer Science Department at Utah State University. His research interests include pattern recognition, computer vision, and fuzzy systems.

Xiaojun Qi received her PhD in computer science from Louisiana State University in 2001. In 2002, she joined the Department of Computer Science at Utah State University as an assistant professor. In 2008, she received tenure and was promoted to associate professor. In 2015, she was promoted to professor. She has been a senior IEEE member since 2010. Her research interests include image retrieval, watermarking and information hiding, pattern recognition, and computer vision.
Austin Jensen received his PhD in electrical engineering from Utah State University in 2014. He played a major part in creating the UAS program and developing a low-cost autonomous UAS remote sensing platform, AggieAir. This included creating and running a service center at the UWRL, which flies AggieAir regularly to capture multispectral aerial imagery, and participating in research areas including autopilot navigation for radio localization, processing and calibrating thermal imagery, and aircraft state estimation.