Direct Georeferencing of Ultrahigh-Resolution UAV Imagery

Darren Turner, *Member, IEEE*, Arko Lucieer, *Member, IEEE*, and Luke Wallace, *Member, IEEE*

***Abstract*—Micro-unmanned aerial vehicles often collect a large amount of images when mapping an area at an ultrahigh resolution. A direct georeferencing technique potentially eliminates the need for ground control points. In this paper, we developed a camera–global positioning system (GPS) module to allow the synchronization of camera exposure with the airframe's position as recorded by a GPS with 10–20-cm accuracy. Lever arm corrections were applied to the camera positions to account for the positional difference between the GPS antenna and the camera center. Image selection algorithms were implemented to eliminate blurry images and images with excessive overlap. This study compared three different software methods (Photoscan, Pix4D web service, and an in-house Bundler method). We evaluated each based on processing time, ease of use, and the spatial accuracy of the final mosaic produced. Photoscan showed the best performance as it was the fastest and the easiest to use and had the best spatial accuracy (average error of 0.11 m with a standard deviation of 0.02 m). This accuracy is limited by the accuracy of the differential GPS unit (10–20 cm) used to record camera position. Pix4D achieved a mean spatial error of 0.24 m with a standard deviation of 0.03 m, while the Bundler method had the worst mean spatial accuracy of 0.76 m with a standard deviation of 0.15 m. The lower performance of the Bundler method was due to its poor performance in estimating camera focal length, which, in turn, introduced large errors in the Z-axis for the translation equations.**

***Index Terms*—Remote sensing, unmanned aerial vehicles (UAVs).**

I. Introduction

In RECENT years, the use of civilian unmanned aerial vehicles (UAVs) as remote sensing platforms (also known as Unmanned Aircraft Systems) has been increasing, primarily due to improvements in the availability of accurate and miniature global positioning systems (GPSs) and inertial measurement units (IMUs), along with the availability of quality off-the-shelf consumer-grade digital cameras [1], [2]. In the realm of scientific research, the use of micro-UAVs, such as multirotor helicopters and small fixed-wing planes (typically below 5 kg in total weight) is becoming more commonplace. This is the type of UAV on which this study is based. The ability of UAVs to acquire useful data for environmental remote sensing applications has been successfully demonstrated by Dunford *et al.* [3] who looked at riparian forests and Rango *et al.* [4] who mapped rangelands in NM. In an agricultural context, the applicability of UAVs has been demonstrated by authors such as Johnson *et al.* [5] who collected imagery over vineyards and Lelong *et al.* [6] who mapped the health of wheat crops.

The temporal resolution of UAV imagery is superior to imagery collected by satellite and conventional aerial photography platforms which are restricted due to limitations in the availability of aircraft, weather, and satellite orbits [7]. Flying at low altitudes (50–120 m AGL), UAV imagery also provides superior spatial resolution (on the order of 1 cm/pixel) when compared with the resolution typically available from digital aerial photography platforms (25 cm/pixel) [8].

Despite its advantages, the footprint of a micro-UAV system collecting imagery at 1-cm/pixel spatial resolution is of limited extent. To map a large area (greater than 2 ha for example) many images need to be collected and processed to create a useable product, such as a georectified image mosaic. Micro-UAVs are typically used as a cost-effective alternative to full-scale airborne surveys, and it is therefore essential that the image processing routines are automated as much as possible such that image processing costs are also kept to a minimum while maintaining a suitable level of accuracy [9].

In recent years, there have been several methods developed to georectify and mosaic UAV imagery in a semiautomated fashion [7], [10]–[12]. Many of these techniques, however, rely upon ground control points (GCPs) or orientation measurements from an IMU to enable accurate georeferencing of the imagery. The collection of GCPs and the purchase of accurate miniature IMUs can be costly, not to mention they are often too heavy for micro-UAVs. Typically, low-cost UAVs only carry cost-effective sensors, and thus, the accuracy of position and orientation data is often limited [13].

Direct georeferencing is a technique that enables imagery to be georectified without the need for GCPs [14]. This has advantages in terms of cost and the expertise required for accurate collection of GCPs [15]. It also removes the need for human interaction to identify GCPs within the imagery, thus allowing a fully automatic processing chain to be implemented. Eugster and Nebiker [16] presented a method to directly georeference a video stream collected from a UAV. Their method used onboard low-accuracy IMU/GPS data and projected the frames onto a preexisting terrain model of the landscape. They reported a planimetric spatial accuracy of 0.7 m when flying at an altitude 50 m above ground level. Chiang *et al.* [17] also implemented a direct georeferencing technique and achieved a planimetric accuracy of around 5 m when flying at 300 m. They proposed that this sort of system and accuracy would be suitable for near-real-time disaster-relief operations. Pfeifer *et al.* [18]

Manuscript received December 20, 2012; revised March 19, 2013; accepted May 16, 2013. Date of publication July 3, 2013; date of current version March 3, 2014. This work was supported in part by the Winifred Violet Scott Trust and in part by the Australian Antarctic Division.

The authors are with the School of Geography and Environmental Studies, University of Tasmania, Hobart, TAS 7001, Australia (e-mail: Darren.Turner@utas.edu.au; Arko.Lucieer@utas.edu.au; Luke.Wallace@utas.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2013.2265295

Fig. 1. Oktokopter in flight fitted with Canon 550D.

developed a system that processed raw low-quality data from gyros, accelerometers, and magnetometers to estimate camera position with a calculated planimetric accuracy of around 1 m when flying at 25 m. Turner *et al.* [12] implemented a direct georeferencing technique that could achieve a planimetric accuracy of around 0.65–1.2 m when flying at 50 m. The limited accuracy of each of these techniques is primarily due to the low accuracy of the navigation-grade GPS units used to record camera (airframe) position at the time of image capture. A second contributing factor to the relatively low accuracies is the imprecise time synchronization between the camera acquisition and GPS receiver.

Another area of difficulty is identifying an appropriate software package to process ultrahigh-resolution UAV imagery. Traditional aerial photography image processing packages can have difficulty with UAV imagery given the different acquisition geometry parameters. In many cases, image blocks acquired from UAV imagery are more similar to close-range terrestrial image blocks [12], [19]. Several authors [12], [20] have demonstrated that modern Computer Vision (CV) and Structure from Motion (SfM) algorithms are well suited to the processing of low-altitude (less than 300 m above ground level) UAV imagery. A discussion about the differences between traditional photogrammetry and low-altitude ultrahigh-resolution UAV imagery can be found in [12] along with an explanation as to why modern CV and SfM algorithms work well with UAV imagery.

In this paper, we describe a UAV image collection system and processing workflow for efficient direct georeferencing in a fully automated fashion with high accuracy. We compare the direct georeferencing accuracy for three different methods: 1) Photoscan [21], a commercial off-the-shelf software package; 2) a method developed in-house based on the Bundler SfM algorithms [12]; and 3) the Pix4D Web-based image processing service [22]. We also discuss the limitations and advantages of each technique, explaining why they achieve different levels of accuracy.

II. Methods

A. Platform

This study used a multirotor UAV made up of a Droidworx AD8 heavy lift airframe (www.droidworx.com.au) fitted with Mikrokopter electronics (www.mikrokopter.com) (see Fig. 1). The UAV has eight rotors (an OktoKopter) and can lift a payload of up to 2 kg for a flight duration of 5–10 min. The

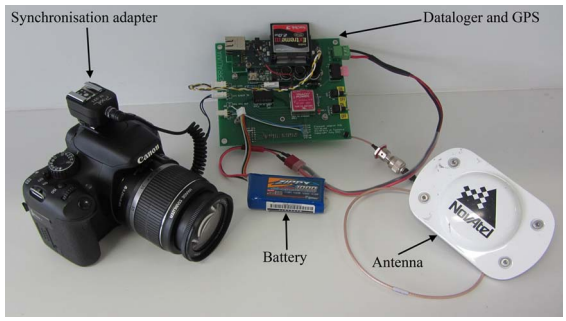

Fig. 2. Image capture and position determination system. Based on camera with synchronization adapter, GPS with data logging board, battery, and GPS antenna.

OktoKopter has a Photohigher AV130 stabilized camera mount (photohigher.co.nz) to which we have fitted a small format digital camera (Canon 550D DSLR, 15 megapixel, 5184 × 3456 pixels, with Canon EF-S 18–55 mm F/3.5–5.6 IS lens). The Mikrokopter onboard electronics consist of a pressure sensor, accelerometers, gyroscopes, and magnetometers for attitude determination along with a navigation-grade GPS (U-blox LEA6S) for position determination. The onboard electronics are used to maintain level flight, control the altitude of the UAV, fly the UAV through a series of predefined 3-D waypoints, and log system information during the flight.

B. GPS Logging System With Camera Sync

To accurately log airframe position information, a single-frequency carrier-phase differential GPS (DGPS) unit (Novatel Flexpak-G2L with an OEMV-2 board) was modified such that it could be mounted to the OktoKopter. The GPS board was removed from its housing to reduce weight, and a power management system was built. A single-board computer (Gumstix Verdex Pro XM4-BT with netCF and console expansion cards) was used to log the GPS data during the flight (see Fig. 2). Python scripts were implemented on the Gumstix computer that initialized the Novatel GPS, checked for sufficient satellite coverage, and configured the Novatel to log the required data streams: carrier-phase range data at a rate of 20 Hz, camera exposure events, and updates to the ephemeris as they occur. During the flight, a Leica 1200 base station also collected static observations over a known point at 20 Hz. Combining the two data streams allows postprocessing to be carried out with the WayPoint Grafnav software [23], calculating aircraft position every 0.05 s with an accuracy of 10–20 cm.

The camera positions are then synchronized with the images collected and written to the appropriate image EXIF header such that they can be read by other software in the later parts of the processing chain. To enable this synchronization, it is essential to accurately record the time of image capture. To achieve this, we fitted a flash sync unit to the camera's hotshoe adapter which was then connected to the GPS unit (see Fig. 2). When the camera shutter is opened, a pulse (normally requesting the flash to be fired) is sent to the GPS unit and subsequently time stamped in the GPS log file to a precision of 0.001 s. This allows the camera position to be interpolated between the 20-Hz postprocessed positions from the data recorded in the log file.

The accuracy of this system is dependent on the delay between the shutter release request and the flash trigger pulse.

The specifications of the Canon 550D state that the maximum (fastest) shutter speed allowable with flash is 1/200th s. This means the longest that the delay between shutter and flash pulse can be is 0.005 s. At a maximum airframe speed of approximately 5 m/s, we have a worst-case scenario movement of 2.5 cm during the 0.005-s delay. This is far below the measurement accuracy of the GPS system (10–20 cm). As this is the worst-case scenario with flying speeds typically only 2–3 m/s and the actual delay between shutter and pulse likely to be less than 0.005 s, the positional error caused by time synchronization delays are insignificant when compared with the accuracy of the GPS measurements.

Once camera position has been determined from the GPS data, we need to correct it to account for the lever arm between the camera center (the center of the CCD in this case) and the location of the GPS antenna. The location of the GPS antenna in relation to the camera is fixed; this means we are able to simply measure the distances between the two locations in pitch, roll, and yaw axes defined by the UAV's onboard the low-accuracy IMU. The approximate roll, pitch, and yaw of the OktoKopter at each exposure time was then read from the onboard log file (recorded at 1 Hz and time stamped with GPS time) and used to determine the lever-arm correction to be applied to the GPS position observation at each exposure station. As the IMU outputs are the result of the combination of measurements made by several sensors (gyros, accelerometers, and magnetometers) that are then Kalman filtered (with proprietary software), it is difficult to know the true accuracy of the IMU readings. We pessimistically assumed an error of $\pm 5^{\circ}$ and modeled the error. We found that the use of low-accuracy orientation estimates to correct for the lever-arm offset introduce errors of up to 4 cm. This is mainly due to the small physical offsets (x: 9 cm, y: 4 cm, z: 22 cm) between the camera and IMU; however, this correction still provides a significant improvement in the position estimate, which is reflected in the results (see Section III-C).

C. Automated Image Selection

Imagery is collected at a rate of approximately one image every 1.5 s by the means of an automatic trigger that sends an alternating autofocus and shutter release command. A 5-min flight can yield around 200 images and cover 1–2 ha of the Earth's surface. To improve processing efficiency, it is necessary to remove some of these images from the data set. We have developed a series of algorithms that do this in an intelligent and automated fashion by applying a set of rules to the image set.

When examining the images captured during a flight, we find that there is generally a small amount ($< 5\%$) of the images that have an unacceptable amount of motion blur. Blurry images have the potential to negatively affect the SfM process. We collect imagery with a high level of redundancy (approximated 90% overlap), so it is possible to exclude the blurry images from the data set. To achieve this, we implemented an algorithm that calculates a blur metric for each image based on the method described by [24]. This method compares the magnitude of intensity variations between neighboring pixels within an original image, and this same image which has been deliberately blurred using a low-pass filter to evaluate blur annoyance in the form of a metric. By manually checking several test data sets, we discovered that a threshold of 0.35 for the blur metric worked well to exclude the blurry, and thus unusable, images.

The automatic camera trigger commences operation as soon as the OktoKopter's motors are started. This means that there is a series of images captured that do not need to be processed while the UAV is climbing to operational altitude. Similarly, once the main part of the flight is complete, the OktoKopter has to descend and land, during which time the camera continues to collect images until the motors are switched off. These unwanted images are removed from the data set based on the flying height at time of capture as recorded in the image EXIF data.

During a flight over an area of interest, the OktoKopter will move from one waypoint to the next, pausing briefly so that at least one image is collected at each waypoint. The waypoints are spaced such that the required overlap and sidelap between the images is achieved. However, as the camera has an automatic trigger, many images are captured during the transit between waypoints, giving a high level of image redundancy. This necessitates the need to remove some of these overlapping images to improve processing efficiency, which is achieved by reading the image location information from image EXIF headers. Through experimentation and comparison with other research, such as [25], we found that an image overlap of around 80% yields the best results for SfM-based image processing algorithms. Image subsetting is achieved by computing the distance between consecutive images, and if this distance is below a predefined threshold (i.e., the OktoKopter is relatively stationary), then this image is skipped and we move on to the next image until the spacing between exposures achieves the desired level of overlap (80%).

D. Image Processing

1) Photoscan: For this study, we compared the performance of the Photoscan Professional [21] image processing software with the Pix4D [22] Web-based image processing service and with an in-house technique we previously developed based on Bundler, as detailed in [12]. To maintain the efficiency and cost effectiveness of our micro-UAV system, it was essential that all the evaluated image processing workflows maximize automation capacity, thus reducing human intervention to enable the creation of the desired products such as orthomosaics and digital surface models (DSMs). Photoscan is a CV- and SfM-based software package that does not include some of the specific functionality that is available in professional photogrammetric software and is thus a more cost-effective solution for processing low-altitude UAV imagery.

After image selection, the Photoscan workflow starts with an image alignment stage that uses SfM techniques to reconstruct the 3-D scene geometry based on feature points [based on Scale Invariant Feature Transform (SIFT)] that have been detected within and matched across the images [26]. The image alignment step also estimates the camera positions and refines the seven camera calibration parameters (focal length, principal point, and three radial and two tangential distortion coefficients). A high-accuracy setting was used for the image alignment stage, and image pair preselection was based on ground control. Photoscan has the ability to select photo pairs for feature matching based on potential overlap given approximate photo locations. Photoscan also allows the input of custom

lens distortion parameters that can then be used to constrain the bundle adjustment.

The second processing stage is to build the scene geometry by applying a dense multiview stereo reconstruction to the aligned images [26]. Detailed surface geometry is built from images that have been undistorted using the camera calibration parameters obtained during the alignment stage of processing. Photoscan has a number of settings that the user needs to define for the geometry build phase. Through experimentation, we found the optimal settings for geometry build to be a high target quality, the object type was set to a height field, and the geometry was set to smooth. Once the scene geometry has been built, an orthophoto and a DSM of the area can be exported. All stages of the processing (except the initial selection of the images) can be run as a batch process, requiring no human interaction and, thus, is a fully automated process.

Georeferencing of the images and 3-D model is achieved by using onboard GPS coordinates attached to each image in the JPEG EXIF header. Generally, these onboard GPS coordinates have a low accuracy (2–5 m) due to the use of a navigation-grade GPS receiver, and therefore, more accurate GCPs are used to optimize the bundle adjustment. In this paper, GCPs are only used for accuracy assessment; instead, the coordinates of the camera stations based on high-accuracy GPS observations are used for georeferencing. Photoscan can use these coordinates in the bundle adjustment to produce accurately positioned orthomosaics and DSMs.

2) Bundler Method: We have previously described a technique to georectify and mosaic ultrahigh-resolution UAV imagery in [12]. This technique is based on the Bundler SfM software package developed by Snavely [27] to align the photos and estimate camera positions in a similar fashion to the first stage of the Photoscan process. In this paper, we developed an algorithm that uses the output of the Bundler process to georectify the individual images in a fully automated fashion. The final stage is to select the images to be merged into the final image mosaic.

3) Pix4D Web Service: The Pix4D service processes UAV imagery in a way similar to the already described software packages [28]; initially, it searches for and matches points using the SIFT algorithm [29]. The matches and approximate locations of the cameras (as supplied by the user, typically in the image EXIF header) are used in a bundle block adjustment to reconstruct the position and orientation of each camera [28]. The matched points can then have their 3-D coordinate calculated, which can be used to obtain a DSM which is, in turn, used to project every image pixel, thus allowing a georeferenced orthomosaic to be generated [28]. However, as it is a Web-based service, none of these steps are visible to the user who is only required to upload the images to be processed, await an initial processing evaluation accuracy report, select the products required, await final processing to complete, and download the results.

III. Results

A. Study Area Used for Testing

We tested our direct georeferencing system at Houston's Lettuce Farm in southeast Tasmania, Australia (see Fig. 3) as part of a project to test the feasibility of using micro-UAVs to map and monitor the health of lettuce crops. Lettuce is a high-value crop (worth ~AUD$25,000/ha) that could greatly benefit from a micro-UAV's ability to rapidly monitor the crop's health, which can then inform management decisions on irrigation, fertilization, and pest control. Our trial mapped a 0.8-ha section of harvest-ready lettuce plants; the site slopes gently from an elevation of 65 m in the southeast to 55 m in the northwest. Before the images were collected, 22 GCPs were distributed in the central area of the study area. For the GCPs, we used aluminum disks, 30 cm in diameter, with the outer rims painted with orange fluorescent paint. The center location of each disk was then measured using a survey-grade dual-frequency DGPS, with a typical accuracy of 2 cm in the horizontal and 4 cm in the vertical (relative to a local coordinated benchmark).

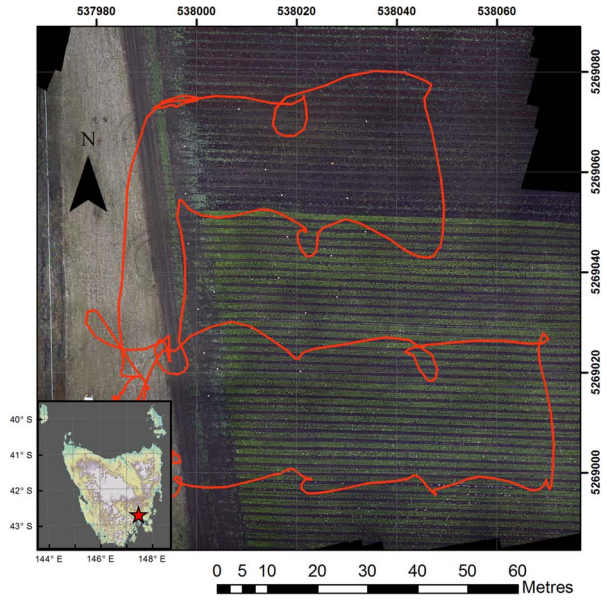

Fig. 3. Mosaic of study site with location map and UAV flight path overlaid (coordinate system: GDA94 UTM55S).

The flight was carried out on the 20th of April 2012 in windy (10–15 knots) and overcast conditions and yielded 297 photos. The wind affected the flight and navigation stability of the OktoKopter, resulting in the UAV circling around a waypoint before attaining it and moving on to the next (see Fig. 3). This was a good test for automated image selection routines to remove some of the redundant photos captured during the circling. Of the 297 images that were captured by the platform, 143 were selected for processing by the image selection algorithms (see Section II-C), i.e., blurry images, images captured during ascent/descent, and images collected too close together were removed.

B. Processing Time

Table I presents a comparison between the methods tested and the processing time required for the test data set. Photoscan is faster than the Bundler technique, as it has been developed to make use of the graphic processing units found on the high-end graphics card installed in the desktop PC used for processing. In comparison, the Bundler technique only uses a single-processor core at any one time and runs considerably slower. The Pix4D software seems slower than Photoscan at first glance; however, when we take note that a majority of the time taken by Pix4D

TABLE I
SUMMARY OF TOTAL PROCESSING TIME FOR 143 IMAGES WITH EACH METHOD

	Method		
	Photoscan	Bundler technique	Pix4D
Processing time	4.3 hours	41 hours	11 hours
Comments	Alignment 1.8 hr Geometry 1.8 hr Orthophoto 0.7 hr	Optimised technique reduced processing time to 16 hr	7 hours of total is time to upload image dataset
Hardware	Windows based, Quad core desktop PC (Intel i7 – 8 cores) with NVIDIA GTX590 graphics card and 32GB of RAM	Linux server with 256GB of RAM and 32 processor cores	Web-based cloud computing service

is during the image upload (7 h), we can see that the actual processing times are quite similar. The data upload was carried out over a very fast (shared 10 Gbit/s) connection, so there is nothing that can be done to improve this part of the process except perhaps reduce the quality of the images and thus reduce their file size and upload time. This may, however, have a negative impact on the quality of the final products. It should also be noted that if the user was in a remote area (e.g., carrying out fieldwork) with no, or limited, Internet access, it would be very difficult to make use of the Pix4D service.

To address the long time taken by the Bundler technique (due to its single-processor usage), some modifications to the processing chain were made. The first stage of the Bundler processing chain is to detect image features with the SIFT algorithm [29] and then to match the features among the images. Neither of these processes, as they stand, makes use of the multiple processor cores. Siftfast [30] is an implementation of the SIFT algorithm that use multiple cores and can also run SIFT on full resolution images, whereas the conventional SIFT algorithm used by the Bundler software requires image resolution be reduced to a maximum of 2000×1500 pixels. The matching stage can take a long time as it has to match the thousands of features (sometimes in the hundreds of thousands per image) found in each image with the features found in all the other images.

As we have prior knowledge of the camera positions, we can determine which images are likely to have a spatial overlap and, thus, likely to yield matches. There would be 10 153 image pairs to be processed if all images were matched exhaustively; however, by removing the pairs that are unlikely to have any overlap, this number was reduced to 1107 matches. These optimizations greatly speed up the feature detection and matching stages of the Bundler algorithm, reducing the total processing time by more than 50% in this example (see Table I); however, the final stage of processing is a bundle adjustment that still only uses a single-processor core and is therefore a processing speed constraint. Of the 15 h taken to process the data set with the optimized Bundler algorithm, 14 h was spent on the bundle adjustment (single-core) stage of processing.

As Pix4D is a cloud-based web service, processing times may vary depending on the current demand on the system. However, with each of the methods, they are all really "set and forget" systems, where the user will initiate processing and return to review the results the next day.

TABLE II
SUMMARY OF SPATIAL ERRORS FOR EACH METHOD

Software	Mean absolute Easting error (m)	Mean absolute Northing error (m)	Mean absolute Total error (m)	Standard deviation of mean error (m)
Photoscan	0.0299	0.108	**0.1149**	0.0214
Pix4D web service	0.2371	0.0598	**0.2471**	0.0277
Bundler method	0.7381	0.1372	**0.7574**	0.1481

To assess the benefit of reducing the number of images to be processed by the use of the automated image selection algorithm, the entire set of images was also processed. The Photoscan processing time increased to 8.6 h, a doubling in processing time from a 63% increase in number of images. The Bundler method processing time increased to 115.5 h, a 180% increase. A test with the full image set was not carried out with the Pix4D service as its major overhead is the data upload time required for the cloud computing. The increase in image processing time would not be significant in comparison to this increase in upload time (estimated at 10 h).

C. Spatial Accuracy

The primary objective of this study is to produce spatially accurate georeferenced mosaics of the area covered by a UAV flight. To assess the accuracy of each method, we measured the offset between the mosaic and recorded locations of the GCPs (see Section III-A for the description of GCPs). A summary of the errors for each method is presented in Table II, and a typical example of the error for each method is illustrated in Fig. 4.

As discussed earlier, not all images were used for processing. It may be the case that using fewer images reduces the accuracy of the final image mosaic. To test this, we also ran all the images (excluding the ascent/descent images which were manually removed) through both the Photoscan and the Bundler software. The accuracy results for Photoscan were worse, with a mean absolute error of 0.2037 m (standard deviation of 0.0285 m). The Bundler method was slightly better with a mean absolute error of 0.68 m (standard deviation of 0.16 m). However, the slight improvement achieved by the Bundler method was at a great expense in terms of processing time (115.5 h versus 16 h previously; see Section III-B).

The spatial accuracy achieved by these methods is proportional to the absolute accuracy of the GPS measurements of camera position at time of exposure. The onboard flight computer logs navigation-grade GPS positions during the flight. Postflight, it is possible to use this data to geotag the images with their position from the navigation log. Verhoeven *et al.* [15] suggest that this type of GPS data can be used to georeference a Photoscan model, but they did not recommend it due to the inaccuracies in the GPS measurements. We processed the data set using only these low-quality GPS positions and found that both Photoscan and the Bundler method could only achieve a mean absolute error of 2.3 m (with standard deviation of

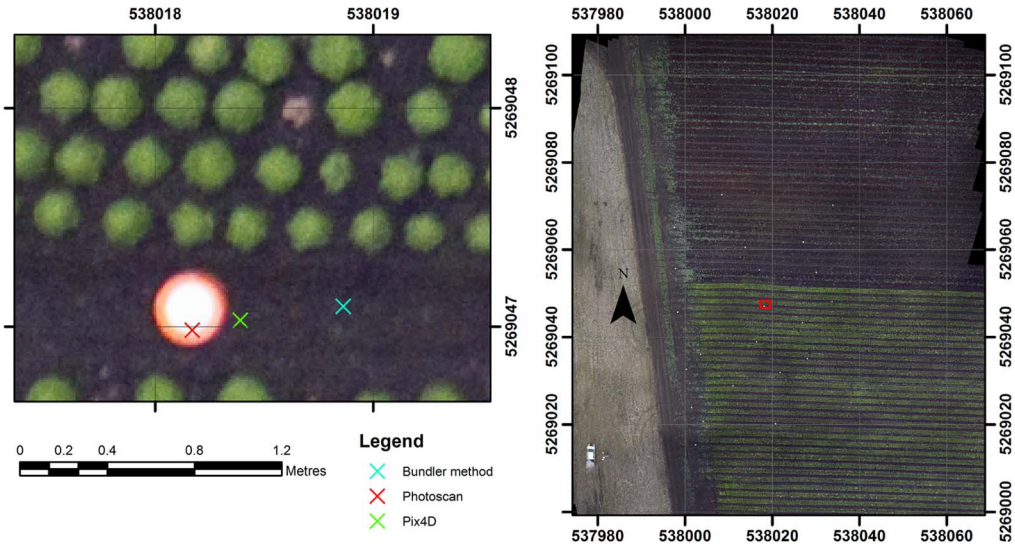

Fig. 4. Close up of one GCP showing spatial error for each method (coordinate system: GDA94 UTM55S).

0.22 m). This experiment validates that the quality of the GPS position data drives the spatial accuracy of the image mosaics and that a direct georeferencing technique based on high-accuracy GPS observations with precise camera synchronization is a valid endeavor.

It is also important to note the improvement to the results that the application of lever-arm corrections yielded. To quantify this, we ran the same set of photos through Photoscan but we did not apply the lever-arm correction to the camera positions. The mean error of this method was 0.19 m with a standard deviation of 0.024 m. This is a significant reduction in the accuracy compared with the results that included lever-arm correction (around 42%). The lever-arm correction is therefore an important part of the image processing workflow.

D. Helmert Transformations

The Bundler method performs its bundle adjustment in an arbitrary coordinate system. We therefore rely on a seven-parameter Helmert transformation (three translations, three rotations, and one scale parameter) to transform the output into a real-world coordinate system [e.g., a projected Universal Transverse Mercator (UTM) easting and northing, and height]. The Helmert transformation parameters and their associated uncertainties are estimated using a least squares approach (see [12] for discussion on Helmert transformation in this context). It is also possible to have Photoscan run in an arbitrary coordinate system by not supplying it with camera positions before processing commences, allowing us to compare how well each method was able to estimate the camera positions with no initial position input to the bundle adjustment. Table III shows that the formal errors for the Photoscan transformation are much smaller than the Bundler method, indicating that the Photoscan software has estimated the camera positions more accurately than the Bundler method.

TABLE III
HELMERT TRANSFORMATION PARAMETERS WITH FORMAL ERRORS (1 SIGMA) FROM LEAST SQUARES SOLUTION

	Translations			Scale	Rotations		
Software	X (m)	Y (m)	Z (m)	Factor	X (°)	Y (°)	Z (°)
Bundler	537994.538 ± 0.13	5269004.837 ± 0.12	92.807 ± 0.16	23.31 ± 0.05	169.81 ± 0.20	187.60 ± 0.18	269.55 ± 0.13
Photoscan	537999.43 ± 0.02	5269075.448 ± 0.02	92.016 ± 0.03	4.49 ± 0.01	357.56 ± 0.03	187.98 ± 0.03	282.86 ± 0.02

IV. DISCUSSION

When considering processing time and spatial accuracy as metrics, it is clear that Photoscan is the best performer of the software methods tested in this study. Photoscan's spatial accuracy is comparable with the accuracy of the DGPS measurements used in the processing chain to georeference the imagery. The processing time was the fastest of the methods, although Pix4D would be comparable without the delay caused by image upload.

The spatial accuracy obtained by Pix4D is good and would be sufficient for most purposes. As Pix4D is a Web-based system, we do not have intimate knowledge of the computing resources and processing algorithms used, and thus, it is difficult to comment as to why the accuracy achieved is lower than that obtained by Photoscan. It could be hypothesized that the Pix4D processing chain has been developed around the user also supplying GCPs and has not been optimized for direct georeferencing.

The Bundler method did not perform well in either spatial accuracy or processing time, despite optimizations that were applied to the process. The difference in the formal errors of the respective Helmert transformations (see Table III) indicates where the greatest source of error may be found. The residuals of the Helmert transform for each of the x, y, and z translations indicate that the z translation residuals for the Bundler Helmert transform are very high in relation to all the other formal errors (the average z residual for Bundler was 1.27 m while only 0.07 m for Photoscan). Further analysis reveals that the errors in the z translation are related to the focal length that has been estimated by the Bundler method (see Fig. 5).

A similar problem with Bundler was found by Rosnell and Honkavaara [25] who discovered that it was hard to get consistent camera calibration parameters when using Microsoft

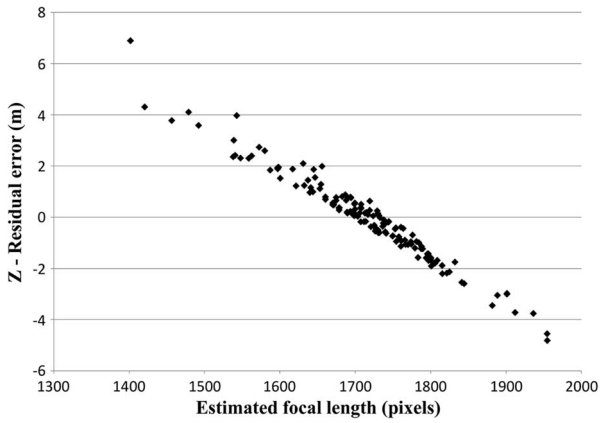

Fig. 5. Estimated focal length and z translation residual errors or bundler processing method.

Photosynth (which is the Web-based implementation of Bundler). In particular, Rosnell and Honkavaara [25] found the focal lengths estimated by Photosynth (Bundler) to be highly variable and suggested that this was because Photosynth was designed to process each image as if it were from a different camera. Thus, Photosynth is more suited to phototourism [31] and not ideal for applications using a single camera with a fixed focal length such as a UAV conducting an aerial mapping survey.

In comparison to Bundler, Photoscan estimates seven camera calibration parameters (see Section II-D1) as opposed to the three parameters estimated by Bundler (focal length and two radial distortion parameters). It would appear that Photoscan performs better than Bundler in this scenario as it is expecting images from the same camera and is thus better able to model the camera parameters. Bundler has some user-definable parameters to constrain or fix the focal estimates for the cameras. We carried out experiments to see if fixing or constraining the focal length could improve results; however, these tests were unsuccessful, generally resulting in the algorithm falling over or the resulting point cloud showing significant artefacts.

Although the Bundler method has the poorest spatial accuracy, there is one advantage to this method in that it is possible to georeference each of the individual images. Photoscan is also able to georeference individual images. The output of the Pix4D software is a mosaic of all the images. There may be scenarios in which rectified individual images may be of use. It should also be noted that each of the three methods also produces a DSM as an additional product with the same spatial accuracy as the mosaics.

The Bundler method suffers from poor processing speed due the lack of parallelization of the processing algorithms. Other implementations of the Bundler software are freely available and use multiple CPU cores to speed up processing times, namely, SfM Toolkit [32] and VisualSfM [33]. Trials with both of these programs revealed that they did not achieve any better results than the original Bundler; in fact, the results were usually poorer. This may be due to the difference in implementation of the parallel SIFT and bundle adjustment algorithms and the optimal settings of processing parameters. For this study, we therefore decided to use the original Bundler implementation that we used in [12].

There is also the option to process the data with professional photogrammetric software packages such as Leica Photogrammetry Suite (LPS) [34] or SOCET SET [35]. Rosnell and Honkavaara [25] processed low-altitude UAV imagery with SOCET SET but discovered that it was not able to deal with the strong perspective distortions in the multiview photography. Our tests with LPS revealed the same issues; it would seem that traditional aerial photography software is not suited to processing low-altitude aerial photography that has significant perspective distortions and very high overlap [25]. Verhoeven *et al.* [15] highlighted some other limitations of LPS and similar packages, in that they require significant photogrammetric skills to operate, calibrated cameras, and high-resolution DSMs.

V. Conclusion

Our study has demonstrated that it is possible to integrate an accurate direct georeferencing system onto a micro-UAV. The developed hardware components combined with current software based on CV algorithms allows the generation of accurate directly georeferenced ultrahigh-resolution orthophotos. The elimination of the need for GCPs simplifies the data collection process, making the use of UAVs more cost effective and time efficient. The absolute spatial accuracy of the mosaics created in this study ($\sim$0.11 m) was limited by the accuracy of the onboard DGPS measurement. The integration of a dual-frequency (L1/L2) onboard GPS unit that should be able to measure the camera position with a higher accuracy (on the order of 2–5 cm) will be the subject of further research.

Acknowledgment

The authors would like to thank Houston's Lettuce Farm for providing access to their property, Dr. C. Watson for the valuable suggestions on the focal length analysis, and S. Harwin for his assistance in collecting field data.

References

[1] S. Nebiker, A. Annena, M. Scherrerb, and D. Oeschc, "A light-weight multispectral sensor for micro UAV—Opportunities for very high resolution airborne remote sensing," *Int. Arch. Photogram. Rem. Sens. Spatial Inform. Sci.*, vol. 37, pp. 1193–1198, 2008.

[2] G. Zhou, V. Ambrosia, A. J. Gasiewski, and G. Bland, "Foreword to the special issue on unmanned airborne vehicle (UAV) sensing systems for earth observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 687–689, Mar. 2009.

[3] R. Dunford, K. Michel, M. Gagnage, H. Piégay, and M. L. Trémelo, "Potential and constraints of unmanned aerial vehicle technology for the characterization of Mediterranean riparian forest," *Int. J. Remote Sens.*, vol. 30, no. 19, pp. 4915–4935, Sep. 2009.

[4] A. Rango, A. Laliberte, J. E. Herrick, C. Winters, K. Havstad, C. Steele, and D. Browning, "Unmanned aerial vehicle-based remote sensing for rangeland assessment, monitoring, and management," *J. Appl. Remote Sens.*, vol. 3, no. 1, pp. 033542-1–033542-15, Aug. 2009.

[5] L. F. Johnson, S. R. Herwitz, S. E. Dunagan, B. M. Lobitz, D. Sullivan, and R. Slye, "Collection of ultra high spatial and spectral resolution image data over California vineyards with a small UAV," in *Int. Symp. Remote Sens. Env.*, Honolulu, HI, USA, 2003, p. 3.

[6] C. C. D. Lelong, P. Burger, G. Jubelin, B. Roux, S. Labbe, and F. Baret, "Assessment of unmanned aerial vehicles imagery for quantitative monitoring of wheat crop in small plots," *Sensors*, vol. 8, no. 5, pp. 3557–3585, May 2008.

[7] J. A. J. Berni, P. J. Zarco-Tejada, L. Suarez, and E. Fereres, "Thermal and narrowband multispectral remote sensing for vegetation monitoring from

an unmanned aerial vehicle," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 722–738, Mar. 2009.
[8] E. R. J. Hunt, Jr., W. D. Hively, S. Fujikawa, D. Linden, C. S. Daughtry, and G. McCarty, "Acquisition of NIR-green-blue digital photographs from unmanned aircraft for crop monitoring," *Remote Sens.*, vol. 2, no. 1, pp. 290–305, Jan. 2010.
[9] M. Sauerbier, E. Siegrist, H. Eisenbeiss, and N. Demir, "The practical application of UAV-based photogrammetry under economic aspects," in *Proc. The Int. Arch. Photogram. Remote Sens. Spatial Inf. Sci.*, Zurich, Switzerland, 2011, vol. XXXVIII-1/C22 UAV-g, pp. 1–6.
[10] A. S. Laliberte, J. E. Herrick, A. Rango, and C. Winters, "Acquisition, orthorectification, and object-based classification of unmanned aerial vehicle (UAV) imagery for rangeland monitoring," *Photogramm. Eng. Rem. S.*, vol. 76, pp. 661–672, Jun. 2010.
[11] M. Bryson, A. Reid, F. Ramos, and S. Sukkarieh, "Airborne vision-based mapping and classification of large farmland environments," *J. Field Robot.*, vol. 27, no. 5, pp. 632–655, Sep./Oct. 2010.
[12] D. Turner, A. Lucieer, and C. Watson, "An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (UAV) imagery, based on structure from motion (SfM) point clouds," *Remote Sens.*, vol. 4, no. 5, pp. 1392–1410, May 2012.
[13] M. Bláha, H. Eisenbeiss, D. Grimm, and P. Limpach, "Direct georeferencing of UAVs," in *Proc. UAV-g, Conf. Unmanned Aerial Veh. Geomatics*, Zurich, Switzerland, 2011, pp. 1–6.
[14] M. Nagai, R. Shibasaki, D. Manandhar, and H. Zhao, "Development of digital surface model and feature extraction by integrating laser scanner and CCD sensor with IMU," in *Proc. ISPRS Congr., Geo-Imagery Bridging Continents*, Istanbul, Turkey, 2004, p. 5.
[15] G. Verhoeven, M. Doneus, C. Briese, and F. Vermeulen, "Mapping by matching: A computer vision-based approach to fast and accurate georeferencing of archaeological aerial photographs," *J. Archaeol. Sci.*, vol. 39, no. 7, pp. 2060–2070, Jul. 2012.
[16] H. Eugster and S. Nebiker, "Geo-registration of video sequences capruted from mini UAVs—Approaches and accuracy assessment," in *Proc. 5th Int. Symp. Mobile Mapping Technol. Approaches Accuracy Assessment*, Padua, Italy, 2007, p. 8.
[17] K.-W. Chiang, M.-L. Tsai, and C.-H. Chu, "The development of an UAV borne direct georeferenced photogrammetric platform for ground control point free applications," *Sensors*, vol. 12, no. 7, pp. 9161–9180, Jul. 2012.
[18] N. Pfeifer, P. Glira, and C. Briese, "Direct georeferencing with on board navigation components of light weight UAV platforms," in *Proc. XXII ISPRS Congress Int. Arch. Photogram., Remote Sens. Spatial Inform. Sci.*, Melbourne, Australia, 2012, vol. XXXIX-B7, pp. 487–492.
[19] F. Remondino, L. Barazzetti, F. Nex, M. Scaioni, and D. Sarazzi, "UAV photogrammetry for mapping and 3D modeling—Current status and future perspectives," in *Proc. Int. Arch. Photogram., Remote Sens. Spatial Inform. Sci.*, Zurich, Switzerland, 2011, vol. XXXVIII-1/C22 UAV-g, p. 6.
[20] L. Barazzetti, F. Remondino, and M. Scaioni, "Automation in 3D reconstruction: Results on different kinds of close-range blocks," in *Proc. Int. Arch. Photogram., Remote Sens. Spatial Inform. Sci., Part 5 Commission V Symp.*, Newcastle upon Tyne, U.K., 2010, vol. XXXVIII, pp. 55–61.
[21] Agisoft, Agisoft Photoscan Professional, Jul. 2012. [Online]. Available: http://www.agisoft.ru/
[22] Pix4D, Next-Generation Aerial Image Processing Software, Jul. 2012. [Online]. Available: http://pix4d.com
[23] Novatel, GrafNav GNSS Post-Processing Software, Apr. 2012. [Online]. Available: http://www.novatel.com/products/software/waypoint-products/post-processingsoftware/grafnav
[24] F. Crete, T. Dolmiere, P. Ladret, and M. Nicolas, "The blur effect: Perception and estimation with a new no-reference perceptual blur metric," in *Proc. SPIE*, San Jose, CA, USA, 2007, p. 64 920I.
[25] T. Rosnell and E. Honkavaara, "Point cloud generation from aerial image data acquired by a quadrocopter type micro unmanned aerial vehicle and a digital still camera," *Sensors*, vol. 12, no. 1, pp. 453–480, Jan. 2012.
[26] G. Verhoeven, "Taking computer vision aloft—Archaeological three-dimensional reconstructions from aerial photographs with photoscan," *Archaeol. Prospect.*, vol. 18, no. 1, pp. 67–73, Jan./Mar. 2011.
[27] N. Snavely, Bundler: Structure from Motion (SfM) for Unordered Image Collections, Jul. 2010. [Online]. Available: http://phototour.cs.washington.edu/bundler
[28] O. Kung, C. Strecha, A. Beyeler, J.-C. Zufferey, D. Floreano, P. Fua, and F. Gervaix, "The accuracy of automatic photogrammetric techniques on ultra-light UAV imagery," presented at the Conference Unmanned Aerial Vehicle Geomatics, Zurich, Switzerland, 2011, XXXVIII-1/C22, Paper UAV-g, XXXVIII-1/C22.
[29] D. Lowe, SIFT Keypoint Detector 2005. [Online]. Available: http://www.cs.ubc.ca/~lowe/keypoints
[30] Libsift, Fast SIFT image features library 2012. [Online]. Available: http://libsift.sourceforge.net
[31] N. Snavely, S. M. Seitz, and R. Szeliski, "Modeling the world from internet photo collections," *Int. J. Comput. Vis.*, vol. 80, no. 2, pp. 189–210, Nov. 2008.
[32] H. Astre, SFM toolkit. 2012. [Online]. Available: http://www.visual-experiments.com/demos/sfmtoolkit
[33] C. Wu, VisualSFM: A Visual Structure From Motion System 2012. [Online]. Available: http://homes.cs.washington.edu/~ccwu/vsfm
[34] Intergraph, LPS 2012. [Online]. Available: http://geospatial.intergraph.com/products/LPS/LPS/Details.aspx
[35] BAE_SYSTEMS, SOCET SET 2012. [Online]. Available: http://www.socetgxp.com/content/products/socet-set

Darren Turner (M'13) received the B.S. degree in computer science and geography (with honors) from the University of Tasmania, Hobart, Australia, in 1994, where he is currently working toward the Ph.D. degree, researching the development of micro-UAVs and their associated image processing requirements for a range of remote sensing applications.

He has 25 years of experience with UAVs, holds instructor ratings for fixed- and rotary-wing radio-controlled aircraft, and holds a private pilot license.

Arko Lucieer (M'13) received the M.Sc. degree in physical geography with a specialization in hyperspectral remote sensing from Utrecht University, Utrecht, The Netherlands, in 2000, and the Ph.D. degree from the International Institute for Geo-Information Science and Earth Observation and Utrecht University in 2004. His Ph.D. research focused on image segmentation and visualization of uncertainty in satellite imagery.

He is a Senior Lecturer in Remote Sensing and GIS with the School of Geography and Environmental Studies, University of Tasmania, Hobart, Australia. Since 2004, he has been a Lecturer and Senior Lecturer (after 2010) with the University of Tasmania. The focus of his research is on texture measures, image segmentation and fuzzy classification, GEOBIA, and change detection techniques. In 2010, he founded a research group focusing on the use of unmanned aircraft systems and small sensors for environmental remote sensing and aerial surveys.

Luke Wallace (M'13) received the Bachelor of Surveying and Spatial Sciences degree (with honors) from the University of Tasmania, Hobart, Australia, in 2005, where he is currently working toward the Ph.D. degree in the School of Geography and Environmental Studies.

His current research interests focus on the use of Bayesian estimation for the determination of UAV position and orientation using vision- and MEMS-based devices for direct georeferencing of imagery and LiDAR data with application to forestry.