Article

An Automated Technique for Generating Georectified Mosaics from Ultra-High Resolution Unmanned Aerial Vehicle (UAV) Imagery, Based on Structure from Motion (SfM) Point Clouds

Darren Turner *, Arko Lucieer and Christopher Watson

School of Geography and Environmental Studies, University of Tasmania, Hobart, TAS 7001, Australia; E-Mails: Arko.Lucieer@utas.edu.au (A.L.); Christopher.Watson@utas.edu.au (C.W.)

* Author to whom correspondence should be addressed; E-Mail: Darren.Turner@utas.edu.au; Tel.: +61-3-6226-2212; Fax: +61-3-6226-2989.

Abstract: Unmanned Aerial Vehicles (UAVs) are an exciting new remote sensing tool capable of acquiring high resolution spatial data. Remote sensing with UAVs has the potential to provide imagery at an unprecedented spatial and temporal resolution. The small footprint of UAV imagery, however, makes it necessary to develop automated techniques to geometrically rectify and mosaic the imagery such that larger areas can be monitored. In this paper, we present a technique for geometric correction and mosaicking of UAV photography using feature matching and Structure from Motion (SfM) photogrammetric techniques. Images are processed to create three dimensional point clouds, initially in an arbitrary model space. The point clouds are transformed into a real-world coordinate system using either a direct georeferencing technique that uses estimated camera positions or via a Ground Control Point (GCP) technique that uses automatically identified GCPs within the point cloud. The point cloud is then used to generate a Digital Terrain Model (DTM) required for rectification of the images. Subsequent georeferenced images are then joined together to form a mosaic of the study area. The absolute spatial accuracy of the direct technique was found to be 65–120 cm whilst the GCP technique achieves an accuracy of approximately 10–15 cm.

1. Introduction

Historically, Unmanned Aerial Vehicles (UAVs) have primarily been used for military applications. More recently, the use of UAVs in the civilian domain as remote sensing tools presents new and exciting opportunities. Improvements in the availability of accurate and miniature Global Positioning Systems (GPS) and Inertial Measurement Units (IMUs), along with the availability of quality off-the-shelf consumer grade digital cameras and other miniature sensors have resulted in an increased use of civilian UAVs [1]. The highest spatial resolution data available from conventional platforms, such as satellites and manned aircraft, is typically in the range of 20–50 cm/pixel. UAVs are capable of flying much lower and hence can collect imagery at a much higher resolution [2,3], often at a sub-decimetre resolution, even as detailed as 1 cm/pixel. The temporal resolution of conventional systems is limited by the availability of aircraft platforms and orbit characteristics of satellites. For the purpose of monitoring highly dynamic vegetation, satellite sensors are often limited due to unfavourable re-visit times [4].

Many studies have successfully used UAVs to map and monitor areas of vegetation that are of an agricultural and/or an environmental interest, see for example [5–8]. Johnson *et al.* [6] used a small fixed wing UAV to collect imagery over a commercial vineyard in California. The imagery had a spatial resolution of 20 cm/pixel and was processed to segment the scenes into vegetation and soil areas and to subsequently calculate percentage vegetation cover. Monitoring of small plots within wheat crops in southwest France [7] is another example of UAVs assisting with agricultural processes. Lelong *et al.* [7] used a modified digital camera to collect imagery in four bands, red, green, blue and near-infrared to enable the calculation of vegetation indices such as the Normalized Difference Vegetation Index (NDVI).

In an environmental monitoring context Rango *et al.* [8] deployed a fixed wing UAV in the rangelands of southern New Mexico, acquiring imagery with at a 5–6 cm/pixel resolution. Laliberte [9] also collected imagery of the New Mexico rangelands, but also used a six band multispectral camera to capture high resolution data in the near infrared. Imagery of such high spatial resolution can provide a lot of information, such as detailed area of vegetation and bare soil coverage, composition by functional or structural group, spatial distribution of plants, inter canopy gaps and in some cases, vegetation type [10]. In another study, Dunford *et al.* [5] used a paraglider type UAV to acquire imagery with a spatial resolution of 6–21 cm/pixel over 179 ha of riparian forest in France. An object-based classification approach was then found to be the most accurate classifier for the detection of dead wood within the forested area [5].

Despite significant evidence highlighting the value of UAVs in the fields of precision agriculture and environmental monitoring, the collection of ultra-high resolution UAV imagery presents a number of challenges. Due to the relatively low flying height (e.g., 50–120 m) of micro-UAVs (<5 kg), the images have a small footprint (e.g., 50×40 m when flying at 50 m above ground level with a typical camera and lens configuration). This necessitates the capture of a large number of images to achieve the spatial coverage required for many applications. For example, a single flight covering approximately 2 ha can yield around 150–200 images. To maximise the potential of the UAV technology for environmental and agricultural applications, it is essential that an automated, efficient, and accurate technique be developed to rectify and mosaic the large volume of images generated.

There are fundamental differences between imagery collected by a UAV flying at low altitude compared to that collected by a traditional aerial platform flying at higher altitudes. UAV imagery is often collected in a haphazard manner (*i.e.*, flight lines with variable overlap and cross-over points); it has large rotational and angular variations between images [11]; the altitude of the platform is low in relation to the height variation within the scene, causing large perspective distortions [11]; and the exterior orientation (EO) parameters are either unknown or, if measured, they are likely to be inaccurate. UAV imagery often has high variability in illumination, occlusions and variations in resolution [12], which are characteristics more typical of those usually presented in close-range photogrammetry applications [13]. Hence, UAV photography has characteristics of both traditional aerial photography and terrestrial photography, and there are opportunities to use image processing algorithms that are applicable to both types of imagery, as suggested by Barazzetti *et al.* [12].

Recently there have been advances in the realm of Computer Vision (CV), resulting in new algorithms for processing terrestrial photography. Examples are the powerful Scale Invariant Feature Transform (SIFT) [14] feature detector, and the Structure from Motion (SfM) algorithms that make use of SIFT features to create 3D models from a series of overlapping photos [15]. SIFT is a region detector, rather than an interest point extractor that would typically be used by traditional photogrammetric software [16]. As a region detector it has been demonstrated that SIFT is applicable to UAV imagery due to its robustness against changes in rotation, scale, and translation between images [16].

The standard approach in modern photogrammetry is to employ a Bundle Block Adjustment (BBA) to solve for the exterior orientation of each photograph and, if required and provided the geometry of the block of photographs allows it, to solve for additional parameters such as the interior orientation (IO). An introduction to the BBA is provided by e.g., Wolf and Dewitt [17]. Most commonly, metric mapping cameras are used for aerial photography for which the IO parameters are known. UAV imagery is typically collected with consumer grade cameras for which IO parameters are neither known nor stable. Measured values for EO parameters, typically captured at relatively low accuracy in the case of UAV photography, can be included in the BBA, and provide approximate measurements for the bundle adjustment [18].

Increasingly, in the case of traditional aerial photogrammetry, the position and orientation of the camera can be derived from GPS and IMU data with sufficient accuracy to allow direct georeferencing without the need for Ground Control Points (GCPs). Often if ground control is available it is primarily used to ensure a reliable transformation from the GPS based coordinate system into the required map coordinate system. This is not the case for UAV photography because of the lower accuracy of the GPS/IMU data and because of the very large scale of the imagery and map products.

Tie/pass points are required to complete a BBA and are typically automatically generated in the case of traditional aerial photography by an interest point extractor algorithm. For UAV imagery, a SIFT algorithm can be used and has the potential to generate a large number of features that can be used as tie/pass points, supplying more redundant observations for a BBA and thus improving the accuracy of the results [11].

Table 1 clearly demonstrates that with UAV imagery, the IO and EO parameters are often not well known, making the use of a traditional BBA problematic or, at least, more similar to terrestrial or close-range photogrammetry. Attempts have been made to overcome these limitations by developing techniques to specifically work with UAV imagery. Berni *et al.* [4] used onboard IMU and GPS data

to estimate the camera's approximate EO parameters which were then imported into traditional photogrammetric software along with calibrated images to create a mosaic. The images collected had a high level of overlap, allowing only the central part of the images to be used to avoid the extremities where view angle caused perspective distortions [4]. A minimum number of GCPs were then manually measured and an aerotriangulation performed. Berni *et al.* [4] were then able to use an existing Digital Terrain Model (DTM) to generate an orthomosaic, however, no overall spatial accuracy for this method was reported.

Table 1. Comparison of Bundle Block Adjustment variables.

Variables	Traditional Aerial Photography	UAV Imagery
IO parameters—Camera calibration e.g., focal length, principle point, lens distortion parameters	Often known as metric, calibrated, cameras are used	Not usually known and often unstable because consumer grade cameras are used
EO parameters—Camera position and orientation	Often measured by high accuracy onboard GPS/IMU	Either unknown or inaccurate due to limited accuracy of navigation grade GPS and miniature MEMs IMU
GCPs—3D ground control	Manual identification of natural or artificial targets in the imagery and surveyed in situ for accurate 3D coordinates	Manual identification of natural or artificial targets identified in the imagery and surveyed in situ for accurate 3D coordinates
Tie/Pass points—2D image points	Manually identified or automatically generated by interest point extractor algorithm	Manually identified or automatically generated by region detector such as SIFT
Object points—3D points	The coordinates of tie and pass points are computed as part of the BBA. The coordinates of terrain points are computed using image matching techniques (usually a hybrid of area and feature based) to identify conjugate points in two or more images, and then by intersection based on co-linearity condition equations.	The coordinates of all SIFT features are computed as part of the BBA (bundler software). A denser point cloud of terrain points is calculated using patch-based multi-view stereo (PMVS) techniques from three or more images.

Laliberte *et al.* [19] developed a method that relied on an existing underlying orthorectified photo and DTM. They initially estimated camera EO parameters from onboard sensors and then iteratively matched each individual image with the existing orthophoto to improve the accuracy of the EO parameters and provide GCPs based on matched features between images. After many iterations of this process, photogrammetric software used the EO parameters and GCPs to orthorectify the images and generate a seamless mosaic. Laliberte *et al.* [19] identified that their methodology has a number of limitations: it requires pre-existing orthophotos that can quickly become out of date, the 10 m DEMs used for orthorectification were not detailed enough compared to the resolution of the UAV imagery, it suffered from problems finding accurate EO parameters, and achieved variable accuracy of the automatically generated tie points. The overall accuracy of the method was reported to have an RMS error of 0.48 m, (corresponding to ~10 pixels), however, it was acknowledged that the method had

only been tested over relatively flat terrain and algorithm performance in areas with higher vertical variability had not been confirmed [19].

Bryson *el al.* [31] presented a georectification and mosaicking technique that used onboard IMU/GPS data to initially estimate camera pose and then image features were matched across the image dataset. A bundle adjustment then used the initial camera pose estimates and the matched features to refine the camera poses; subsequently the images are then rectified and mosaicked using these poses. The method described by Bryson *et al.* [31] is similar to the method that we propose in that it uses similar processes (e.g., bundle adjustment, feature matching). However, there are significant differences in the platform used (rotary wing *versus* fixed wing) and the resolution of the imagery collected. Also, in this study we do not use onboard IMU data; we can automatically identify GCPs, and we integrate the use of multiview stereopsis algorithms into the solution.

These techniques performed well but many are based on traditional photogrammetric software designed to process imagery collected from conventional platforms. Some of these techniques have some key disadvantages: they use existing underlying DTMs and base orthophotos, they rely on complex workflows to estimate camera EO parameters, and, in some cases, require human intervention to identify GCPs.

In this study, we describe a methodology for geometric image correction that uses new CV and SfM algorithms that are more applicable to UAV photography. The technique is fully automated and can directly georeference and rectify the imagery with only low accuracy camera positions, resulting in UAV image mosaics in real-world coordinates. Alternatively, GCPs can be automatically identified to improve the spatial accuracy of the final product. The automation and simplicity of our technique is ideally suited to UAV operations that generate large image data sets that require rectification and mosaicking prior to subsequent analysis.

2. Methodology

2.1. UAV Platform and Photo Acquisition

The UAV platform used in this study is a multi-rotor OktoKopter (Figure 1). This platform is purpose designed for aerial photography [20] and has a stabilised camera mount, to which we have fitted a small format digital camera (Canon 550D 15 Megapixel, 5,184 × 3,456 pixels, DSLR, with Canon EF-S 18–55 mm F/3.5–5.6 IS lens). Image resolution (ground pixel size) at a typical flying height above terrain of 50 m is approximately 1 cm/pixel. The OktoKopter has a payload limit of approximately one kilogram and with a full payload has a flight duration of around 5–6 min. A single flight conducted at 50 m above ground level (AGL) can cover an area of around 4–5 ha, producing approximately 200–300 images under a standard operating configuration. Larger areas are covered with multiple flights, or by increasing the flying height and lowering the spatial resolution.

The Oktokopter has an onboard navigation system based on a navigation grade GPS receiver (U-blox LEA6S) and a small Microelectromechanical System (MEMS)-based IMU (Mikrokopter Flight Controller ME V2.0) enabling it to fly autonomously through a pre-defined set of waypoints. As part of this study, we developed flight planning software that calculates the spacing and layout of waypoints to optimise the image acquisition over a region of interest at a nominated image scale

(Figure 2). Imagery is acquired at the maximum rate allowed by the camera (approximately 1 Hz), providing ample image overlap in addition to redundancy to account for occasional outlier acquisitions (excessive tilt or poor exposure).

Figure 1. Oktokopter fitted with Canon 550D.

Figure 2. Software to plan flight over Antarctic moss bed.
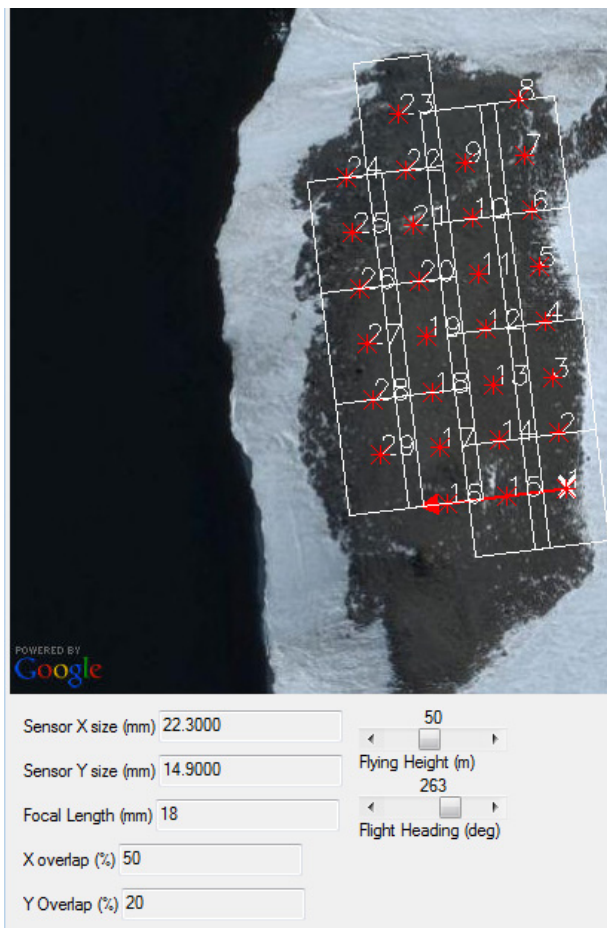
2.2. Block adjustment and Point Cloud Generation

The automated mosaicking technique encompasses a number of stages. The first step requires the manual elimination of any images outside the study region or of limited quality. This qualitative process is the only manual intervention required in the processing chain.

The ideal processing strategy for the imagery would be traditional photogrammetric software that uses GPS/IMU data for bundle adjustment and thus provide significant redundancy in block and photo invariant parameters. Our micro-UAV platform used for this study, however, does not carry a GPS receiver and IMU sensor that can collect data with sufficient accuracy for these techniques to work. In addition, a consumer grade digital camera is used, which means that IO calibration parameters are neither known nor stable. To overcome these problems we have applied bundle adjustment software (Bundler, [15]) specifically designed to enable automated 3D reconstruction of a scene captured by cameras with unknown internal parameters [21]. Dandois and Ellis [22] demonstrated that it has become relatively straightforward to use newly developed CV and SfM algorithms to generate 3D geometry from sets of overlapping digital photographs collected from UAV platforms.

The Bundler software [15] uses SfM algorithms to compute the camera geometry and to generate a sparse 3D point cloud for the area of interest. The SfM framework initially uses the SIFT algorithm [14] to detect and describe local features within each image. SIFT feature descriptors are invariant to scale, orientation, affine distortion and partial illumination changes [23] and can be matched across multiple images. Using the conjugate (matched) image points as input, a bundle block adjustment is applied to compute the exterior orientation (position and orientation) of each camera exposure station. In addition, the bundle adjustment computes the interior orientation parameters (focal length and two radial distortion parameters) of each image, although if required these parameters can be implicitly defined and fixed for all images. The bundle adjustment output includes 3D coordinates for a sparse point cloud of SIFT features in an arbitrary coordinate system which we denote (p_x, p_y, p_z). The Bundler software package is fully automated, requiring only images and a few optional user definable parameters as input.

2.3. 3D Point Cloud Transformation Using Direct Technique

A seven parameter Helmert transformation (three translations, three rotations and one scale parameter) can be used to describe the relationship between the point cloud coordinate system (model space) and a real-world (object space) coordinate system (e.g., a projected Universal Transverse Mercator (UTM) easting and northing, and height). We initially use the computed (bundle adjustment) and measured (GPS) values of the exposure station coordinates to solve for the Helmert transformation parameters. This approach, which does not rely on GCPs in the imagery, is often referred to as direct georeferencing [24], and is useful when working in unsafe or inaccessible areas where GCPs cannot be physically measured on the ground.

The GPS coordinates of the exposure station are determined using the OktoKopter's on-board GPS receiver, with pre-flight synchronisation of the camera's internal clock with GPS time so that during post-flight data analysis the GPS position at the moment of exposure can be written to the EXIF header information for each image. The height measurements from navigation-grade GPS receivers are

relatively poor, hence we use height measurements provided by the OktoKopter's barometric altimeter, which is estimated to be accurate to 1 m when used over short time scales as per a typical UAV flight.

The 3D coordinates of the exposure stations are extracted from the Bundler output and denoted p_x, p_y, p_z. Image EXIF header information is read to extract the matching GPS location. The GPS latitude, longitude and altitude (relative to the WGS84 datum) are subsequently converted into the UTM projected coordinate system, resulting in easting, northing, and ellipsoidal height coordinates. Transformation to an orthometric height system is also possible through the use of a local geoid model if required. Corresponding exposure station coordinates from the bundle adjustment and the GPS are then matched (see Table 2 for an example) to provide a list of point pairs used to compute the parameters of a Helmert transformation. The number of point pairs available is equal to the number of images used by Bundler to generate the point cloud, this number will depend on how large an area is being mapped, but for a single flight there can be as many as 200 point pairs.

Table 2. Sample point pairs list.

Real World Coordinate System			Bundler Coordinate System		
Easting	**Northing**	**Height**	p_x	p_y	p_z
481,495.15	2,638,913.85	39.81	5.2142	−14.3954	−0.7744
481,494.54	2,638,915.10	40.03	5.1918	−14.0937	−0.9143
481,494.53	2,638,918.55	40.90	5.0252	−13.6941	−0.8905
481,494.10	2,638,919.18	40.80	5.2283	−13.3615	−0.9766
481,495.25	2,638,920.18	40.41	5.2167	−13.1875	−0.8768
⋮	⋮	⋮	⋮	⋮	⋮

Errors in the measured GPS coordinates, the Bundler derived exposure station coordinates, and the lever arm between the camera and the GPS antenna contribute to uncertainty in the derived transformation parameters. The camera and GPS antenna share a common vertical axis to within a few centimetres and a vertical offset of approximately 25 cm. The solution for our system, however, is dominated by GPS errors that limit the absolute accuracy to 5–15 m when using a pseudorange only navigation-grade GPS receiver [25] The absolute accuracy of our derived point cloud is limited primarily by the navigation grade GPS, but we find that the translation parameters typically have low formal errors (often <±40 cm) indicating that the relative position of the GPS points, and thus the transformation model, has comparatively high precision.

2.4. 3D Point Cloud Transformation Using GCP Technique

If GCPs are established prior to photography, then the real-world coordinates of these GCPs can be used to derive the parameters of the Helmert transformation, rather than rely on GPS data from the UAV. Accurate GCP coordinates can potentially improve the solution of the Helmert transformation and therefore result in a higher accuracy of the final point cloud and image features. For this purpose, we use circular metal targets (12 cm diameter) painted with fluorescent orange paint distributed across the region to be mapped. The coordinates of these GCPs are measured using a survey grade dual frequency differential GPS, with a typical accuracy of 2 cm in the horizontal and 4 cm in the vertical (relative to a local coordinated benchmark).

The point cloud generated by the Bundler software is relatively sparse and insufficient to reliably identify the GCPs. A novel multi-view stereopsis algorithm [26] can be applied to the output from the Bundler software to densify the sparse point cloud. This algorithm is implemented in the Patch-based Multiview Stereo (PMVS2) software. A detailed description of the algorithm can be found in Furukawa and Ponce [26] and Lucieer *et al.* [27]. The resulting PMVS2 point cloud has extremely dense point spacing, typically around 1–2 cm, with each of the orange GCP targets generating multiple 3D points [27].

The coordinates of the points in the PMVS2 point cloud are still in the bundler coordinate system (p_x, p_y, p_z) but can be transformed into real-world coordinates with the Helmert transformation parameters determined from the direct georeferencing approach (Section 2.3). Transforming the point cloud into the same coordinate system as the GCPs enables automatic matching of the orange discs in the point cloud with their corresponding real-world GPS coordinates. A simple RGB threshold is applied as the point cloud is transformed to filter out the orange pixels. The locations of the orange points are recorded both in the original coordinate system (p_x, p_y, p_z) and the transformed real-world coordinate system: easting, northing and height.

In most cases, multiple orange points are returned for each target, using a search radius of 60 cm we group these points based on their spatial distribution. The centroid of each group of points is determined and used as the penultimate bundler based coordinate of the GCP. These centroid coordinates are then matched against the *in situ* field survey coordinates via a simple separation criterion, *i.e.*, identifying point pairs that are no more than 2 m apart thereby eliminating misidentified orange points as their location will typically not be close to a GPS coordinate.

As the original Bundler coordinates (p_x, p_y, p_z) were also recorded for the orange disc points we can now replace the calculated centroids with (p_x, p_y, p_z) and derive a new list of point pairs (similar to Table 2). If all GCPs were successfully identified and matched to their corresponding GPS coordinate we will typically have up to 60 point pairs from which to calculate a new set of Helmert transformation parameters that have an improved accuracy and precision compared to the direct georeferencing technique. This improvement is due to the fact that the Helmert transformation parameters are now based on higher accuracy GCPs based on survey-grade GPS measurements rather than the on-board navigation-grade GPS coordinates. The new solution also has an improved precision which can be seen in the reduced formal errors of transformation parameters (e.g., errors reduced from ~40 cm to ~5 cm in translation parameters).

2.5. Rectification of the Images

The locations of the matched image features used to derive the point cloud are extracted directly from the bundle adjustment output. For every image we extract the image coordinates of each of these features (*Image$_x$*, *Image$_y$*) and their corresponding 3D bundler coordinates (p_x, p_y, p_z). The previously derived Helmert transformation parameters are then applied to the (p_x, p_y, p_z) coordinates to transform them into the real world coordinate system (easting, northing, height). We generate a table of corresponding image coordinates (2D) and real-world coordinates (planimetric only) for every feature in the dataset (Table 3).

Table 3. Example list of GCPs for an image.

Easting	Northing	$Image_x$	$Image_y$
481,497.41	2,638,927.16	4,412.62	149.99
481,497.74	2,638,926.96	4,446.58	207.21
481,497.59	2,638,927.83	4,502.50	70.89
481,497.68	2,638,926.44	4,391.27	273.48
⋮	⋮	⋮	⋮

The large amount of image overlap in our datasets allows us to only transform the central part of the images to avoid distortions at the extremities, in a similar manner to Berni *et al.* [4]. The number of matched features and thus GCPs for each image is typically quite large (2,000–10,000) and a Delaunay triangulation uses these GCPs to rectify each individual image. The density of the GCPs gives us the equivalent of a high resolution Digital Terrain Model (DTM) of the area of the image allowing the triangulation to produce accurate results.

2.6. Mosaicking

The final stage of the process is to join the images into a single mosaic that covers the study area. Colour balancing can initially be used to remove differences in exposure and lighting conditions between the images resulting in an improvement of the visual integrity of the final mosaic. Colour balancing can be performed with standard image/photo processing packages, or within mosaicking software. However, to maintain visual integrity of the imagery, we chose not to use any colour balancing or seam blending, allowing the final product to be quantitatively assessed without bias.

As all images are now rectified and georeferenced, it is a straightforward process to mosaic them with a georeferenced mosaicking algorithm, which is for example available in ENVI [28]. As there is a large amount of overlap between the images in the dataset, only about one third of the images are required to create a mosaic of the study area. Selection of the images at this time is a simple manual process that involves adding images to the mosaic until sufficient coverage is achieved. This is a processing step that could be easily automated and this will be the subject of further research.

3. Results and Discussion

3.1. Study Area and Dataset

To illustrate the effectiveness of our georeferencing and mosaicking technique we present a case study of UAV remote sensing in Antarctica. The Windmill Islands region near Casey (Australia's largest station) has the most extensive and well-developed vegetation in Eastern Antarctica (Figure 3 site map). Mosses are the most dominant plants in Antarctica. These mosses preserve a record of past climate along their shoots, which make them a valuable proxy for climate change at remote sites. Climate change is now recognised as occurring in the high latitudes rendering Antarctica one of the most significant baseline environments for the study of global climate change. Temperature, UV-B, and changes in water availability have been identified as the three key factors that will change in the Antarctic regions with

climate change. Despite this, there have been few long-term studies of the response of Antarctic vegetation to climate [29,30].

Figure 3. (**a**) The continent of Antarctica with an arrow in Eastern Antarctica indicating the location of the Windmill Islands (**b**) The locations of the Robinson Ridge and Red Shed study sites in the Windmill Islands
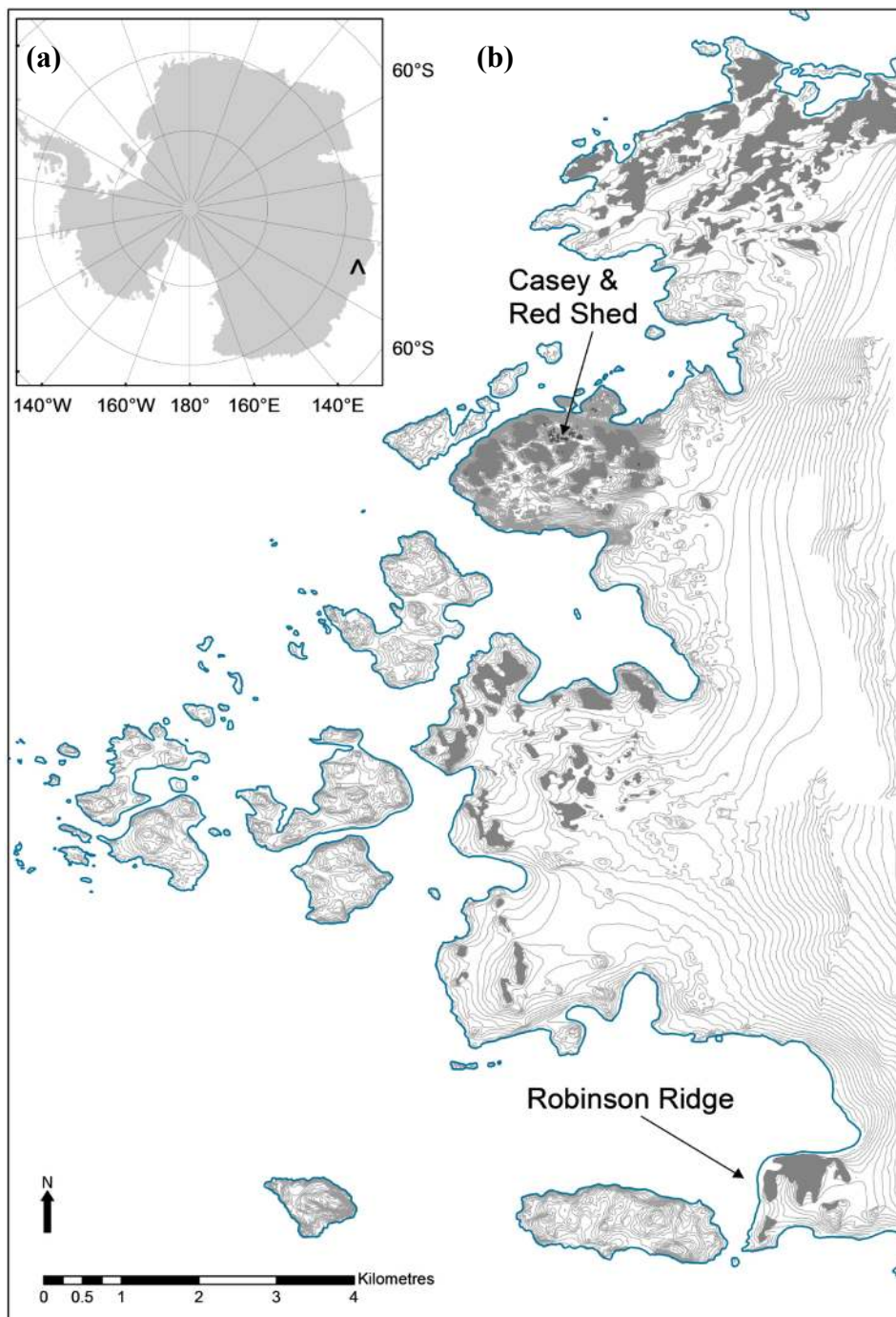
The spatial scale of the moss beds (tens of m^2) makes satellite imagery (even recent very high resolution imagery of 0.5 m resolution) unsuitable for mapping their extent in sufficient detail. Due to logistical constraints aerial photography is impractical and also does not provide the required spatial resolution. Recent developments in the use of UAVs provide exciting new opportunities for ultra-high

resolution mapping and monitoring of this unique Antarctic environment. The aim of this case study is to create ultra-high resolution and geometrically accurate image mosaics of two field sites near Casey: Robinson Ridge and Red Shed. Robinson Ridge is approximately 10 km south of Casey. These mosses grow near small melt streams on a ~100 m high ridge on the coastline (Figure 4 photograph). Two hundred photographs were selected from a UAV flight on 25 February 2011 flying at approximately 50 m AGL. The Red Shed site is a small bowl-shaped catchment fed by a snow melt lake right behind the main accommodation building at Casey. The mosses are concentrated around few main drainage channels. Sixty nine photos were selected from a UAV flight on 20 February 2011 at 50 m AGL.

Figure 4. Moss bed area at the Robinson Ridge site with a variety of healthy moss (green), stressed moss (red/brown), and dead moss (black). Small orange discs (~10 cm diameter) and trays (~30 cm diameter) used as GCPs for geometric correction and validation are visible in the photograph.

3.2. Helmert Transformation Parameters

For both datasets, Helmert transformation parameters were calculated initially via the direct technique (see Section 2.3) and then with the use of the GCP-based technique (see Section 2.4). A summary of the Helmert parameters obtained can be found in Table 4, in which the formal errors of each parameter are listed. These errors, which are the mean residual error from the least squares solution, represent the precision of the Helmert transform. It can be seen that the precision of the GCP technique for both datasets is better than the direct technique due to the error in the onboard GPS position. The precision of the direct technique for the Robinson's Ridge Helmert parameters (around 16–17 cm for the translations) is better than for the Red Shed site (around 36–45 cm for the translations). This difference in precision is most likely explained by a number of factors such as differences in flying height and the presence of outliers in the dataset used to calculate the Helmert transform. Another contributing factor is that there is more variation in the flying height for the Robinson Ridge dataset, improving the geometry of the solution and thus the precision of the transformation parameters. However, it is

important to note that the precision of the Helmert transform is not a good indication of the geometric accuracy of the resulting point clouds and image mosaics. It is thus necessary to measure the absolute spatial accuracy of the final mosaic using GCPs (see Section 3.4).

For each dataset the images were rectified (see Section 2.5) using both the direct and the GCP techniques. The number and density of the points used for the triangulation for individual images is very high (Figure 5). The homogenous areas such as the snow have poor point coverage, this is due to the poor performance of the SIFT algorithm over smooth surfaces. This study was not concerned with the snow areas so this limitation does not affect the results.

Figure 5. Example of 2888 of control points (shown in red—extracted from the process described in Section 2.5) on a single photograph.
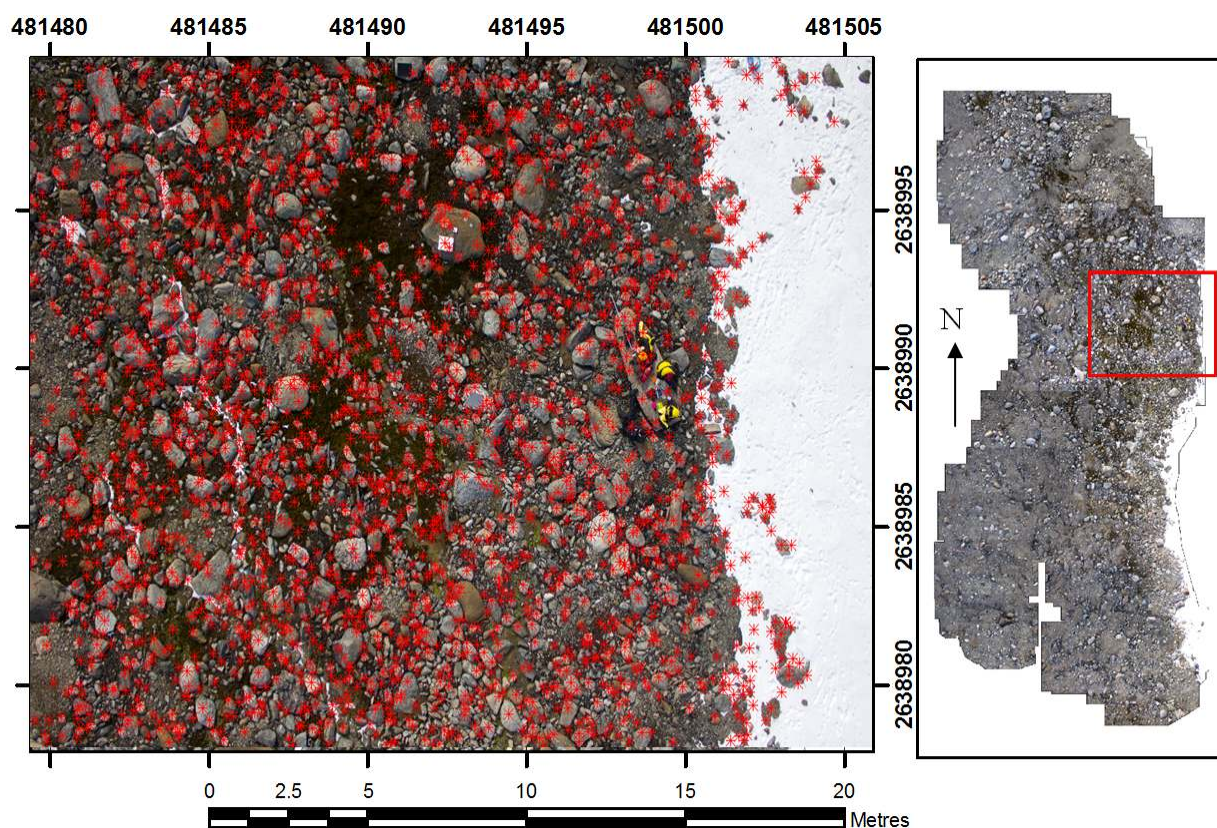

Table 4. Helmert transformation parameters with formal errors (1 sigma) from least squares solution.

Dataset	Method	Calculated Helmert Transform Parameters						
		Translation X (m)	Translation Y (m)	Translation Z (m)	Scale Factor	Rotation X (°)	Rotation Y (°)	Rotation Z (°)
Robinson's ridge	200 camera locations (Direct)	4,814,747.58 ± 0.160	2,638,997.85 ± 0.160	39.06 ± 0.167	12.658 ± 0.046	0.615 ± 0.286	1.204 ± 0.702	9.977 ± 0.207
Robinson's ridge	25 GCPs	481,472.54 ± 0.066	2,638,997.77 ± 0.039	40.30 ± 0.038	12.774 ± 0.009	0.994 ± 0.05	3.158 ± 0.113	9.810 ± 0.043
Red shed	69 camera locations (Direct)	478,776.001 ± 0.371	2,648,411.55 ± 0.368	63.31 ± 0.457	13.840 ± 0.068	2.945 ± 0.04	−10.277 ± 0.407	249.122 ± 0.286
Red shed	19 GCPs	478,777.397 ± 0.042	2,648,409.88 ± 0.059	54.23 ± 0.074	13.736 ± 0.008	−186.2325 ± 0.04	187.737 ± 0.057	−290.3135 ± 0.034

3.3. Mosaics

To show how well our technique works, we do not apply any colour balancing or blend the seam lines in the production of our mosaics. A detailed section of a typical seam within the mosaic for the Robinson's Ridge site can be seen in Figure 6. Here the seam line running across the middle of the image is barely visible indicating a high level of accuracy in the image alignment. A qualitative analysis of the visual integrity of the mosaic reveals that there are no obvious distortions around the seam lines and the colour balancing between the neighbouring images is even, despite only relying on automated colour balancing on the camera at the time of exposure.

Figure 6. Detailed section of a typical region within the image mosaic of Robinson's Ridge highlighting accuracy of seam lines (yellow box highlights seam line).
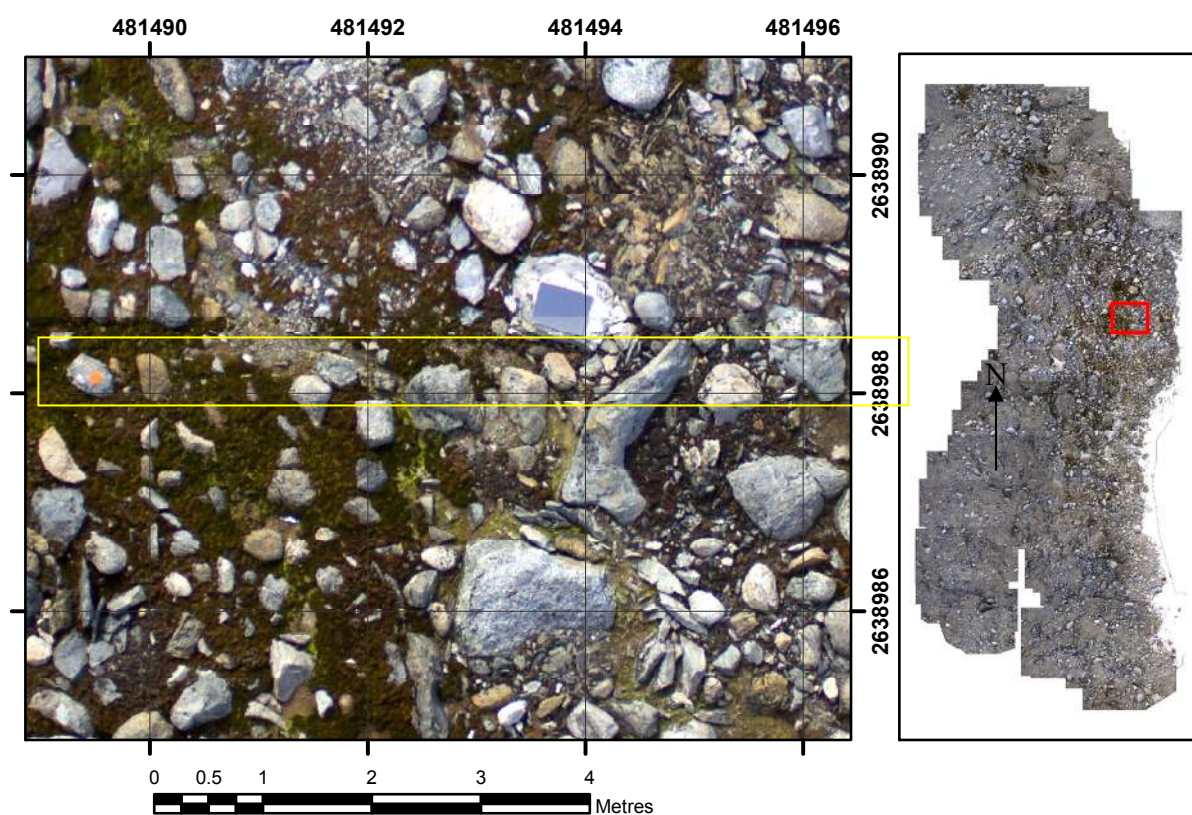

3.4. Spatial Accuracy

To quantify the absolute spatial accuracy for each image mosaic we measured the distance between the orange GCP markers in the image and their corresponding GPS coordinates. A summary of the absolute accuracy of each mosaic is presented in Table 5. These accuracy values highlight that the GCP technique has a superior spatial accuracy to the direct georeferencing technique, which is also visualised in Figure 7. The systematic nature of the errors from the direct georeferencing technique are also shown in Figure 7, this is typical of navigation grade GPS data collected over short time periods, where errors based on atmospheric and orbits effects are typically highly temporally correlated. The absolute spatial accuracy achieved with the GCP technique of around 10–15 cm is considered to be very good, especially when you bear in mind it is being compared to differential GPS measurements

that have an error of ±2–4 cm themselves. The absolute accuracy of the direct georeferencing technique is similar for both field sites.

Table 5. Summary of mosaics and their spatial accuracy.

Dataset	Method	Area (ha)	Number of Check Points	Topographic Variation (m)	Mean Absolute Easting Error (m)	Mean Absolute Northing Error (m)	Mean Absolute Total Error (m)	Standard Deviation of Mean Error (m)
Robinson's Ridge	200 camera locations	0.5	43	4–24	1.076	0.571	1.247	0.184
Robinson's Ridge	25 GCPs	0.5	44	4–24	0.087	0.103	0.129	0.061
Red Shed	69 camera locations	1.1	61	13–19	0.449	0.447	0.665	0.459
Red Shed	20 GCPs	1.1	63	13–19	0.086	0.042	0.103	0.064

Figure 7. Detailed section of an image mosaic of Robinson's Ridge showing typical spatial errors of direct and GCP techniques in relation to the actual GCPs (the small orange discs).
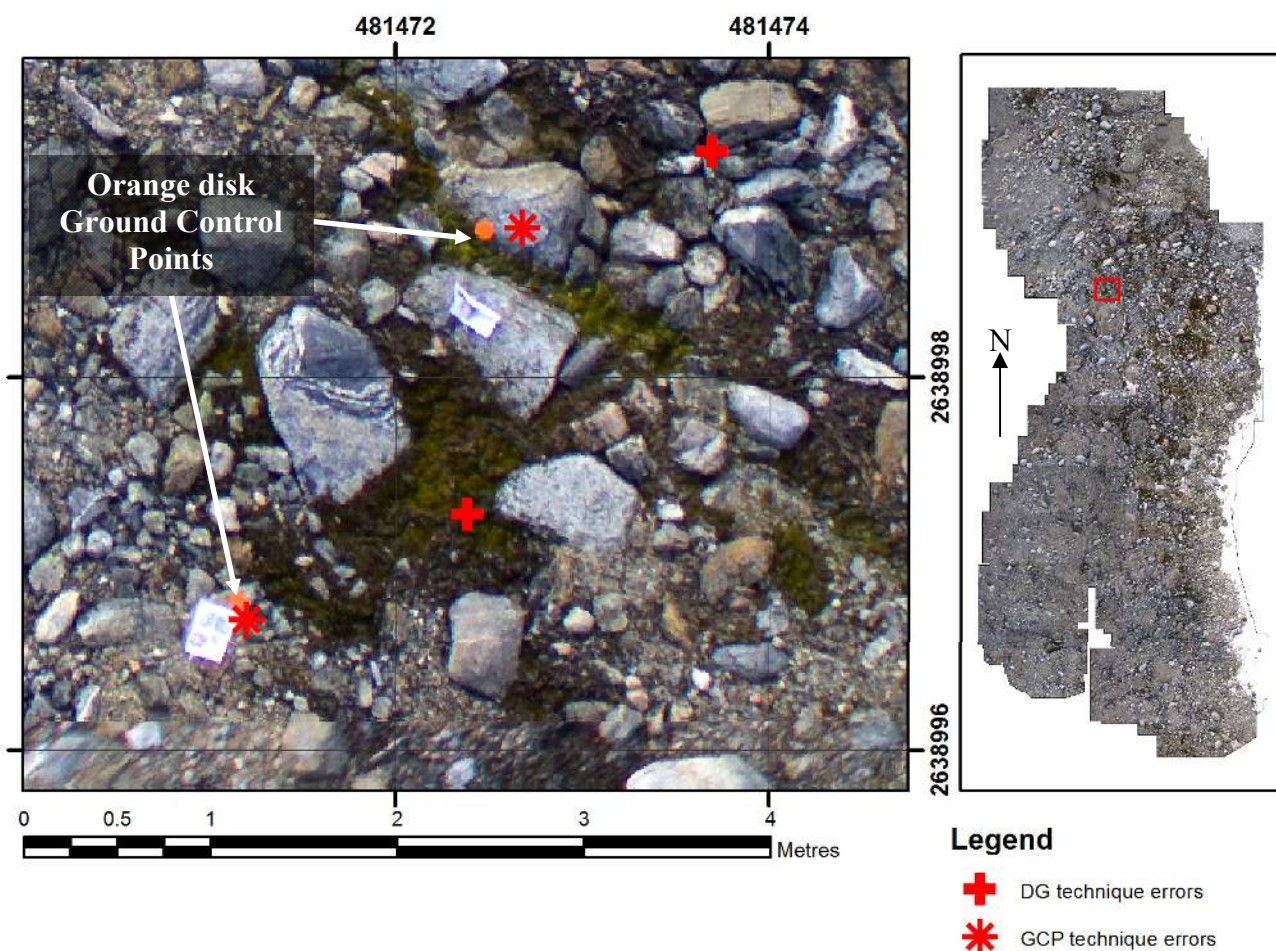
When we compare the absolute accuracy of the GCP technique between the two study sites, we see an insignificant difference between the two (0.10 ± 0.06 m *versus* 0.13 ± 0.06 m). Topographic influences (there is a much larger variation in height in the Robinson's Ridge dataset compared to the Red Shed dataset—Table 5) will drive some of the differences observed in accuracy between the datasets.

Further investigation of the absolute spatial error in the Robinson's ridge dataset, the largest errors are at the GCPs at the extremities of the mosaic, e.g., in the south at the top of the hill and in the north at the bottom of the hill. If we do not incorporate these points in the accuracy assessment the mean total error is reduced to 0.112 m (with a standard deviation of ±0.042) which is comparable to the Red Shed dataset. Another source of error in the mosaics may be the fact that for the sake of efficiency and automation, we used a dense triangulation to rectify the images rather than a rigorous orthorectification that would typically be undertaken in a traditional aerial photogrammetric treatment of such imagery.

An experiment was undertaken to test the robustness of the GCP technique against a reduction in the number of GCPs available. From the 20 GCPs that were automatically detected in the Red Shed dataset, 10 were randomly selected such that they were evenly distributed throughout the study area. New Helmert transform parameters were then calculated based on only these 10 GCPs followed by image rectification and mosaicking. An accuracy assessment based on 63 GCPs gave a mean spatial error of 0.108 m (with a standard deviation of ± 0.063) which is similar to the error when using all 20 GCPs for the Helmert transformation.

With the direct technique, a significant portion of this error is explained by the lack of precision in the measurements that it uses to generate the Helmert transform parameters. These camera location measurements were collected by a navigation-grade (single frequency) on-board GPS unit with no differential corrections and inaccurate time synchronisation between GPS and camera. The Robinson's Ridge dataset Helmert transform parameters had a higher precision than the Red Shed parameters but this was not reflected in the absolute spatial accuracy that was achieved for the two areas.

The absolute spatial accuracy of mosaics produced by our system is comparable to or exceeds the results achieved by others such as Laliberte *et al.* [19] and Berni *et al.* [4]. However, our technique is fully automated, requiring no user intervention and is thus very time-efficient.

4. Conclusions

Unmanned Aerial Vehicles (UAVs) are increasingly used for environmental remote sensing applications. A large number of UAV aerial photographs are required to cover even relatively small study areas. The characteristics of UAV-based aerial photography has necessitated the development of new geometric image correction and mosaicking techniques. Our approach applies modern Computer Vision (CV) algorithms to ultra-high resolution UAV imagery so that 3D point clouds can be generated and subsequently used to georeference the imagery. The combination of a micro-UAV platform with our novel image processing techniques provides an inexpensive, automated, and accurate system for producing ultra-high resolution mosaics of a study area that by far exceeds the resolutions typically available from conventional platforms.

Imagery of moss beds in Antarctica was used to validate the spatial accuracy of our technique which can directly georeference the imagery or use Ground Control Points (GCPs) if they are available. Two datasets (one containing 200 images, the other 69 images) were processed with both techniques producing four mosaics. The directly georeferenced mosaics had a spatial accuracy of 65–120 cm with whilst the GCP technique achieved a spatial accuracy of 10–15 cm.

The primary source of error for the direct georeferencing technique is the fact that it relies on an inaccurate navigation grade GPS to record the camera position. A significant portion of this error could

be removed with the addition of an on-board single/dual frequency carrier phase differential GPS. This could potentially eliminate the need for GCPs and greatly improve the efficiency of field surveys. A further improvement to the spatial accuracy of the mosaics could be achieved by applying a rigorous orthorectification rather than a Delaunay triangulation which is currently used. Investigation into the potential of these two improvements will be the subject of further research.

The technique presented in this study is applicable to other UAV surveys conducted over alternate surface types and terrains. The automated nature of our technique allows a large collection of ultra-high resolution UAV images to be quickly and efficiently transformed into a usable product for a range of subsequent analyses.

Acknowledgments

References

1. Nebiker, S.; Annena, A.; Scherrerb, M.; Oeschc, D. A light-weight multispectral sensor for micro UAV—Opportunities for very high resolution airborne remote sensing. *Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci.* **2008**, *37*, Part 1, 1193–1198.
2. Hunt, E.R.J.; Hively, W.D.; Fujikawa, S.; Linden, D.; Daughtry, C.S.; McCarty, G. Acquisition of nir-green-blue digital photographs from unmanned aircraft for crop monitoring. *Remote Sens.* **2010**, *2*, 290–305.
3. Scaioni, M.; Barazzetti, L.; Brumana, R.; Cuca, B.; Fassi, F.; Prandi, F. Rc-Heli and Structure & Motion Techniques for the 3-D Reconstruction of a Milan Dome Spire. In *Proceedings of the 3rd ISPRS International Workshop 3D-ARCH 2009: "3D Virtual Reconstruction and Visualization of Complex Architectures"*, Trento, Italy, 25–28 February 2009; p. 8.
4. Berni, J.A.J.; Zarco-Tejada, P.J.; Suarez, L.; Fereres, E. Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 722–738.
5. Dunford, R.; Michel, K.; Gagnage, M.; Piégay, H.; Trémelo, M.L. Potential and constraints of unmanned aerial vehicle technology for the characterization of mediterranean riparian forest. *Int. J. Remote Sens.* **2009**, *30*, 4915–4935.
6. Johnson, L.F.; Herwitz, S.R.; Dunagan, S.E.; Lobitz, B.M.; Sullivan, D.; Slye, R. Collection of Ultra High Spatial and Spectral Resolution Image Data over California Vineyards with a Small UAV. In *Proceedings of the International Symposium on Remote Sensing of Environment*, Honolulu, HI, USA, 10–14 November 2003; p. 3.
7. Lelong, C.C.D.; Burger, P.; Jubelin, G.; Roux, B.; Labbe, S.; Baret, F. Assessment of unmanned aerial vehicles imagery for quantitative monitoring of wheat crop in small plots. *Sensors* **2008**, *8*, 3557–3585.

8. Rango, A.; Laliberte, A.; Herrick, J.E.; Winters, C.; Havstad, K.; Steele, C.; Browning, D. Unmanned aerial vehicle-based remote sensing for rangeland assessment, monitoring, and management. *J. Appl. Remote Sens.* **2009**, *3*, 1–15.

9. Laliberte, A.S.; Goforth, M.A.; Steele, C.M.; Rango, A. Multispectral remote sensing from unmanned aircraft: Image processing workflows and applications for rangeland environments. *Remote Sens.* **2011**, *3*, 2529–2551.

10. Rango, A.; Laliberte, A.; Steele, C.; Herrick, J.E.; Bestelmeyer, B.; Schmugge, T.; Roanhorse, A.; Jenkins, V. Using unmanned aerial vehicles for rangelands: Current applications and future potentials. *Environ. Pract.* **2006**, *8*, 159–168.

11. Zhang, Y.; Xiong, J.; Hao, L. Photogrammetric processing of low-altitude images acquired by unpiloted aerial vehicles. *Photogramm. Rec.* **2011**, *26*, 190–211.

12. Barazzetti, L.; Remondino, F.; Scaioni, M. Automation in 3D reconstruction: Results on different kinds of close-range blocks. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2010**, *38*, Part 5, 55–61.

13. Luhmann, T.; Robson, S.; Kyle, S.; Harley, I. *Close Range Photogrammetry*; Whittles Publishing: Caithness, UK, 2006; p. 510.

14. Lowe, D. *Sift Keypoint Detector*; Available online: www.cs.ubc.ca/~lowe/keypoints/ (accessed on 14 April 2011).

15. Snavely, N. *Bundler: Structure from Motion (SFM) for Unordered Image Collections*; Available online: phototour.cs.washington.edu/bundler/ (accessed on 13 January 2011).

16. Lingua, A.; Marenchino, D.; Nex, F. Performance analysis of the sift operator for automatic feature extraction and matching in photogrammetric applications. *Sensors* **2009**, *9*, 3745–3766.

17. Wolf, P.R.; Dewitt, B.A. *Elements of Photogrammetry with Applications in GIS*, 3rd ed.; McGraw-Hill: New York, NY, USA, 2000.

18. Barazzetti, L.; Remondino, F.; Scaioni, M.; Brumana, R. Fully automatic UAV image-based sensor orientation. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2010**, *38*, Part 5, p. 6.

19. Laliberte, A.S.; Winters, C.; Rango, A. A procedure for Orthorectification of Sub-Decimeter RESOLUTION Imagery Obtained with an Unmanned Aerial Vehicle (UAV). In *Proceedings of the ASPRS 2008 Annual Conference*, Portland, OR, USA, 28 April – 2 May 2008; p. 9.

20. Mikrokopter Mikrokopter wiki. Available online: www.mikrokopter.com (accessed on 17 January 2011).

21. Snavely, N.; Seitz, S.M.; Szeliski, R. Modeling the world from internet photo collections. *Int. J. Comput. Vis.* **2008**, *80*, 189–210.

22. Dandois, J.P.; Ellis, E.C. Remote sensing of vegetation structure using computer vision. *Remote Sens.* **2010**, *2*, 1157–1176.

23. Lowe, D.G. Object Recognition from Local Scale-Invariant Features. In *Proceedings of the International Conference on Computer Vision*, Corfu, Greece, 21–22 September 1999.

24. Nagai, M.; Shibasaki, R.; Manandhar, D.; Zhao, H. Development of Digital Surface Model and Feature Extraction by Integrating Laser Scanner and CCD Sensor with IMU. In *Proceedings of the ISPRS Congress, Geo-Imagery Bridging Continents*, Istanbul, Turkey, 12–23 July 2004.

25. US DoD. *Global Positioning System Standard Positioning Service Performance Standard*, 4th ed.; Defence, D.O., Ed.; US Government: Washington, DC, USA 2008.

26. Furukawa, Y.; Ponce, J. Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. Pattern Anal.* **2009**, *32*, 1362–1376.

27. Lucieer, A.; Robinson, S.; Turner, D. Unmanned Aerial Vehicle (UAV) Remote Sensing for Hyperspatial Terrain Mapping of Antarctic Moss Beds Based on Structure from Motion (SFM) Point Clouds. In *Proceedings of the 34th International Symposium for Remote Sensing of the Environment (ISRSE)*, Sydney, Australia, 10–15 April 2011.

28. ITTVIS. *ENVI Software—Image Processing and Analysis Solutions*; Available online: http://www.ittvis.com/envi (accessed on 8 May 2012).

29. Convey, P.; Bindschadler, R.; di Prisco, G.; Fahrbach, E.; Gutt, J.; Hodgson, D.; Mayewski, P. Antarctic climate change and the environment. *Antarct. Sci.* **2009**, *21*, 541–563.

30. Robinson, S.A.; Wasley, J.; Tobin, A.K. Living on the edge–plants and global change in continental and maritime antarctica. *Glob. Chang. Biol.* **2003**, *9*, 1681–1717.

31. Bryson, M.; Reid, A.; Ramos, F.; Sukkarieh, S. Airborne vision-based mapping and classification of large farmland environments. *J. Field Robot.* **2010**, *27*, 632–655.