Remote Sensing

Article

Assessing the Accuracy of Georeferenced Point Clouds Produced via Multi-View Stereopsis from Unmanned Aerial Vehicle (UAV) Imagery

Steve Harwin * and **Arko Lucieer**

School of Geography and Environmental Studies, University of Tasmania, Private Bag 76, Hobart, TAS 7001, Australia; E-Mail: arko.lucieer@utas.edu.au

* Author to whom correspondence should be addressed; E-Mail: Stephen.Harwin@utas.edu.au; Tel.: +61-3-6226-2703; Fax: +61-3-6226-7628.

Abstract: Sensor miniaturisation, improved battery technology and the availability of low-cost yet advanced Unmanned Aerial Vehicles (UAV) have provided new opportunities for environmental remote sensing. The UAV provides a platform for close-range aerial photography. Detailed imagery captured from micro-UAV can produce dense point clouds using multi-view stereopsis (MVS) techniques combining photogrammetry and computer vision. This study applies MVS techniques to imagery acquired from a multi-rotor micro-UAV of a natural coastal site in southeastern Tasmania, Australia. A very dense point cloud ($<$1–3 cm point spacing) is produced in an arbitrary coordinate system using full resolution imagery, whereas other studies usually downsample the original imagery. The point cloud is sparse in areas of complex vegetation and where surfaces have a homogeneous texture. Ground control points collected with Differential Global Positioning System (DGPS) are identified and used for georeferencing via a Helmert transformation. This study compared georeferenced point clouds to a Total Station survey in order to assess and quantify their geometric accuracy. The results indicate that a georeferenced point cloud accurate to 25–40 mm can be obtained from imagery acquired from ~50 m. UAV-based image capture provides the spatial and temporal resolution required to map and monitor natural landscapes. This paper assesses the accuracy of the generated point clouds based on field survey points. Based on our key findings we conclude that sub-decimetre terrain change (in this case coastal erosion) can be monitored.

1. Introduction

Remote sensing technology has improved a great deal in recent decades and the miniaturisation of sensors and positioning systems has paved the way for the use of Unmanned Aerial Vehicles (UAVs) for a wide range of environmental remote sensing applications [1,2]. The use of UAVs for non-military applications has only become possible in more recent times as these miniaturised systems have become affordable for research and commercial entities [3]. UAVs are now a viable alternative for collecting remote sensing data for a wide range of practical applications. The miniaturisation and commercialisation of sensors, positioning systems, and UAV hardware provide scientists with a means to overcome some of the limitations of satellite imagery and aerial photography, namely spatial and temporal resolution. The datasets produced by UAV remote sensing are at such high detail that characteristics of the landscape can be mapped that are simply not distinguishable at the lower resolutions generally obtainable via manned aircraft (~10–100 cm) and satellite systems (>50 cm). Furthermore, the ease of deployment and low running costs of these UAV systems allows for frequent missions providing very high spatial and temporal resolution datasets on-demand [1]. Recent advances in computer vision include multi-view stereopsis (MVS) techniques [4], which can derive 3D structure from overlapping photography taken from multiple angles. Recent studies [5–9] have successfully adopted MVS to derive dense point clouds from UAV photography. Creating an accurately georeferenced point cloud using these methods will be referred to as UAV-MVS as it combines photogrammetric and computer vision techniques to process the UAV data.

1.1. Structure from Motion - Photogrammetry Meets Computer Vision

The UAV-MVS process yields a 3D point cloud similar to that produced using active sensors such as LiDAR and interferometric RADAR and the point density of the cloud is a function of the image resolution and camera object separation. The 3D point cloud is a good data structure for storing complex surface structure and a digital surface model (DSM) can be generated to represent the captured surface. This complexity is not usually well represented in a digital elevation model (DEM) as these are commonly 2.5D datasets, *i.e.*, there is only one Z-value at each 2D coordinate (x, y) [10]. An advantage of UAV-borne sensors is the ability to acquire data from close range at multiple viewing angles (*i.e.*, nadir and oblique). A nadir view commonly used in photogrammetry results in more occlusion and detail can be missed. "The central theme of photogrammetry is accuracy" [11], and the techniques used in this field for deriving 3D coordinates are well-established and robust. Technological advances have improved the efficiency and automation of these accurate established techniques. Robotics and computer vision have also advanced significantly in recent decades. The achievement of human-level capability for information extraction from image data is the theme of this field [11].

The 3D reconstruction from imagery relies on the extraction of image correspondences. In recent years both fields have sought to improve automated image matching. Matched feature points in overlapping photography enable the derivation of 3D coordinates as point clouds. In computer vision this is done through a process known as Structure from Motion (SfM) that incorporates MVS techniques to derive camera position and orientation and 3D model coordinates. The success of MVS via the feature matching process is hindered by untextured surfaces, occlusions, illumination changes and acquisition geometry [12]. Of the recent advancements in this area, the Scale Invariant Feature Transform (SIFT) operator [13] has proven to be one of the most robust to large image variations [12,14]. A number of alternatives to SIFT exist such as Gradient Location and Orientation Histogram (GLOH) [15], Speeded Up Robust Features (SURF) [16], LDAHash [17] and Principal Component Analysis (PCA)SIFT [18]. However, they all aim to achieve essentially the same result.

Advances such as SIFT have allowed MVS 3D reconstruction systems to solve for the orientation of the camera and derive 3D positions of the feature surface points using bundle block adjustment techniques. As outlined in Triggs *et al.* [19], the theory and methods for bundle adjustment have been around for a long time. A number of software solutions exist that perform the bundle adjustment required to solve the camera parameters (including image orientation) and generate a 3D point cloud of a scene, including Bundler [20–22], Microsoft Photosynth [23], Agisoft PhotoScan [24] and PhotoModeler [25]. These tools are optimised for consumer-grade cameras with an uncalibrated focal length and close-range imagery acquired from different view angles. The density of the point clouds created is a function of the number of unambiguous point matches found. Generally, the density is quite sparse, which is adequate for the purpose of basic 3D modelling and tourism photo collection management. To increase the density it is necessary to revisit the images and use the knowledge of camera parameters to extract more points. Multi-view stereo techniques such as patch-based multi-view stereo (PMVS2) [26] and cluster multi-view stereo (CMVS) [27]) take the output from a standard bundle adjustment and perform a match, expand, filter approach to densify the original sparse point cloud [4,28]. This point cloud densification is usually done using the down-sampled imagery (<3 Megapixels) in order to reduce computing overhead.

In this paper we propose a modified workflow so that full-size images can be used in PMVS2 resulting in much denser and more accurate point clouds. Seitz *et al.* [29] compares over one hundred MVS algorithms [30] and this approach outperforms most other algorithms (although the objects were not natural landscapes). Strecha *et al.* [31] used LiDAR reference data to compare the Furukawa and Ponce [4] approach to the Strecha and Fransens [32] and Strecha *et al.* [33] approaches and their results favoured the Furukawa and Ponce [4] algorithm for completeness and relative accuracy. A number of alternative MVS approaches have been developed such as Semi-Global Matching (SGM) [34,35], Plane-sweep strategies [36], and the MVS pipeline developed by Vu *et al.* [37], some of which are now also freely available and may be evaluated in a future study. The PMVS2 software is open source; it integrates easily with Bundler, and creates a very dense and accurate point cloud. Whilst SfM and MVS were not designed for environmental monitoring and modelling nor intended for UAV imagery, these techniques are proving to be well suited to UAV data capture as they combine images from multiple angles and varying overlap. The low UAV flying height also improves feature definition as the technique can capture complex shapes allowing for the representation of features such as hollows and overhangs.

1.2. UAVs for 3D Reconstruction of Natural Landscapes

The use of UAVs for 3D reconstruction and point cloud generation via aerial imagery has been considered in the past, particularly in recent years [5–9,38–41]. These studies usually focused on assessing the accuracy of similar techniques, however, this manuscript presents the first attempt to quantify the accuracy of the whole UAV-MVS close-range data capture and georeferencing process applied to a natural landscape based on a comparison with Total Station survey data. Eisenbeiss and Sauerbier [38] examined the use of UAVs in archaeological applications. They employed a more traditional photogrammetric approach to obtaining 3D data (DSM and ortho-images) from UAV photography. Neitzel and Klonowski [5] compared a number of web services and software packages that "automatically generate 3D points from arbitrary image configurations" [5]. Whilst the accuracy assessment performed in Neitzel and Klonowski [5] provided some insight into the comparative accuracy of the successfully generated point clouds, they were not able to derive a general rule or prediction of accuracy due mainly to their uncertainty relating to the influence of topography on the point clouds produced. The images used were down-sampled from 12 Megapixels to 3 Megapixels and only PMVS2 and Photoscan produced point clouds dense enough (~90 and ~110 points per m^2 respectively) to see the ground control points (GCPs) across the entire study site (a relatively flat parking lot with few GCPs). Küng *et al.* [39] used Pix4D [42] to generate and compare georeferenced DEMs and orthmosiacs based on UAV GPS camera positions (geotags) and GCPs measured using DGPS and identified in the captured imagery. They flew at 130–900 m over non-natural sites and found that the accuracy of the geotagging was 2–8 m and the GCP method was accurate to 5–20 cm. The accuracy was strongly influenced by the resolution of the imagery and the texture and terrain in the scene [39]. Vallet *et al.* [40] compared georeferenced DTMs produced from LiDAR, Pix4D and NGATE (in SOCET SET [43]). The UAV flew at 100–150 m over a semi-natural scene containing 12 GCPs measured using static DGPS. The results suggest 10–15 cm accuracy is achievable when flying at 150 m. Rosnell *et al.* [44] looked at imaging conditions in different seasons and how the point cloud generation performed. They chose more natural sites but focused on a comparison between a 1 m resolution DEM resampled from a relatively sparse Photosynth point cloud (2–3 points per m^2) and a detailed terrain produced using NGATE. The photography was captured from an altitude of 110–130 m and it is unclear how the GCPs were found in the imagery. Hirschmüller [41] briefly discussed the use of Bundler and SGM with UAV imagery and provided a qualitative accuracy assessment. Dandois and Ellis [8] focused on vegetation structure mapping and chose to use GCPs from photography and DEMs resulting in poor georeferencing precision. They compared their tree height estimates from point clouds to LiDAR methods and found that DTMs produced using SfM techniques suffered from inaccuracy due to the complex canopy structure resulting in poor ground point extraction. The canopy surface was well represented and compared well to the LiDAR equivalent. Lucieer *et al.* [9] used the UAV-MVS technique to create point clouds of complex terrain with 1–2 cm point spacing. The 1 cm resolution DEMs generated were used to derive terrain derivatives such as topographic wetness index. Turner *et al.* [7] used the Bundler to create DSMs from point clouds with an estimated accuracy of ~10 cm. The derived transformations were then applied to the matched SIFT feature locations in each image to allow georectified image mosaics to be created.

1.3. Georeferenced Point Clouds and Reference Data

The point cloud generated by UAV-MVS is generally in an arbitrary reference frame and needs to be registered to a real-world coordinate system. This is achieved by identifying key features in the point cloud that can be matched to known real world coordinates. In natural environments GCPs that stand out are not often available. The solution is to distribute highly visible targets. Once the coordinates for feature points have been established and matched (manually or automatically) a 3D Helmert transformation (with seven parameters: three translations, three rotations and one scale) can be used to transform the point cloud from an arbitrary reference frame into a real-world coordinate reference frame. The georeferenced point clouds produced need to be compared to reference data. The use of a Total Station survey to accurately map a set of reference points around the study area is an accepted method of obtaining "ground truth". Walker and Willgoose [45] assessed the accuracy of their Total Station data using error propagation theory and found that uncertainty in position is ~1 cm and uncertainty in elevation is ~2 cm. Shrestha *et al.* [46] used traditional surveying techniques to acquire profiles to assess the accuracy of LiDAR; Töyrä *et al.* [47] used Total Station elevation data to assess LiDAR; and Farah *et al.* [48] used Total Station data to assess the accuracy of DEMs derived from GPS. In a number of these studies the Root Mean Squared Error (RMSE) for each dimension and the total RMSE have been used as accuracy metrics. There are other possible metrics such as the mean difference, standard deviation, correlation length, minimum/maximum difference and bias [45,49,50]. The RMSE is a recognised and relatively easily understood proxy for answering this question when the "ground truth" dataset is a set of distributed points rather than a continuous "truth" surface.

This study seeks to evaluate the accuracy of the UAV-MVS point cloud generated from imagery of a natural environment, namely a section of protected coastline. This accuracy will be assessed by comparing georeferenced point clouds to a Total Station and differential GPS (DGPS) survey. The site was chosen due to the fact that it is gradually eroding and this erosion may serve as an indicator for climate change. The erosion on this protected section of coastline is subtle and may not be visible via traditional aerial and satellite change detection techniques. We aim to use the UAV-MVS technique to generate dense and accurate 3D point clouds of this site and detect and quantify change over time. This investigation into the accuracy of UAV-MVS is the first step in a series of investigations into the application of these systems and processes to hyperspatial and hypertemporal earth observation and environmental monitoring using UAVs. To reliably quantify change we must first verify that the technique is sufficiently accurate to allow subtle (sub-decimetre) changes to be detected and measured. This accuracy assessment will serve to validate our GCP georeferencing process and quantify the uncertainty in the absolute position of the point cloud. We hypothesize that sub-decimetre change can be monitored using the UAV-MVS process.

2. Methods

2.1. Study Area

The site chosen for this study is a 100 m section of coast in a sheltered estuary in southeast Tasmania, Australia (Figure 1). The site was selected to evaluate the suitability of the UAV-MVS technique to

fine-scale change detection. The southern end of the site is a salt marsh and the remainder contains grasses along an erosion scarp with intermittent scrub bush (Figure 2).

Figure 1. Coastal monitoring site in an estuary in southeast Tasmania.
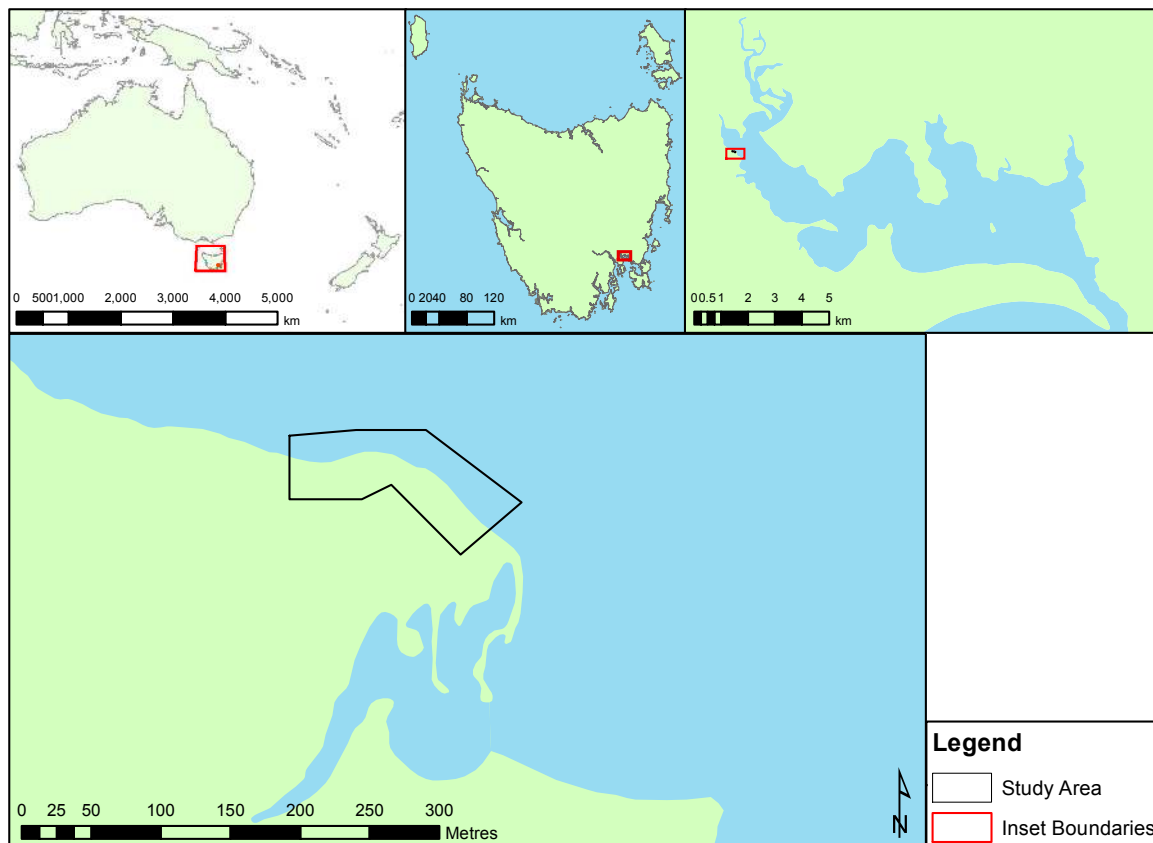
Figure 2. Images of the site (the first two are taken looking east, the third is taken looking west). The first image shows a ~2 m high erosion scarp and the second shows the much smaller 5–10 cm scarp. The third image shows that this section of coast is representative of the area.

2.2. Hardware

The TerraLuma UAV used for this study is based on the OktoKopter platform [51]. The OktoKopter is an electric, multi-rotor system with an approximate payload limit of 1 kg. When carrying a full payload the flight time is approximately 6 minutes, which is more than enough to capture UAV-MVS imagery for a ~1–2 ha area. The on-board GPS and navigation sensors provide 5–10 m positional accuracy and

the on-board computer is able to navigate the UAV to pre-defined GPS waypoints. The OktoKopter has a stabilised camera mount that can carry different sensors. To create UAV-MVS point clouds a standard digital camera can provide imagery with sufficient resolution. We have chosen the Canon 550D digital SLR camera as it has excellent image quality and a lightweight body. The focus of the lens is fixed to infinity, the ISO is set to 200, and the aperture is fixed to f3.5 resulting in a minimum shutter speed of $1/2000^{\text{th}}$ of a second. These settings reduce motion blur. The camera is triggered once per second (1 Hz) by the OktoKopters flight controller acquisition interval. This frequency provides a great deal of overlap (70%–95%) and redundant photography (over 300 photos per flight).

A Leica Viva real-time kinematic dual-frequency differential GPS (RTK DGPS) was used to survey the GCPs for UAV-MVS point cloud georeferencing. A Leica Total Station (TC407) was also used to survey the GCPs and create a reference dataset for accuracy assessment.

2.3. Data Collection

For accurate georeferencing of the UAV imagery accurate GCP coordinates are required. We distributed around 90 orange circular flat disks, ~10 cm in diameter, across the study site at a spacing of ~3–5 m. Initially traffic cones (witches hats) were used for GCPs, however the exact centre and height reference were difficult to establish when surveying the GCPs. These disks were our first attempt at ground control and this study was partially set up to assess if their small size was potentially reducing georeferencing accuracy. To evaluate an alternative 21 larger 22 cm pizza trays have been used. A hole was drilled in the centre of each tray. A 3 cm wide rim of was painted on each tray in colours designed to allow automated unique identification (since the datasets used for this study were captured the colour has been reconsidered and the trays now have an orange rim). For future studies we are considering custom made cones that may provide better centre point matching once point clouds have been extracted.

The larger trays were distributed along the two sides of the study area at intervals of ~6 m. Figure 3 shows the layout of the GCP trays and disks. We carried out both an RTK DGPS survey and an additional Total Station survey (with the prism mounted on a pole) to provide a reference dataset of GCP coordinates for all trays and disks. The orthometric height obtained from the Total Station survey was converted to an ellipsoid height by subtracting a geoid-ellispoid separation value (or N value) of 3.256 m (derived using AUSGeoid09 Geoid-Ellipsoid Separation Interpolation [52]). These GCPs were surveyed using RTK DGPS which were compared to Total Station coordinates to gauge the accuracy of the GCP survey technique. The UAV was deployed at a flying height of 30–50 m above ground level (AGL) capturing a photograph every second. The first flight captured nadir photography and the second flight captured oblique photography with the camera tilted to approximately 45°. The captured photos were screened and a subset of clear (*i.e.*, not blurred) photos of the area were selected for the UAV-MVS process.

Figure 3. Map of GCP layout. The trays are mainly along the edge of the study area and a number are placed toward the central portion. This distribution is considered favourable to accurate georeferencing. The smaller GCP disks are spread throughout the study area.
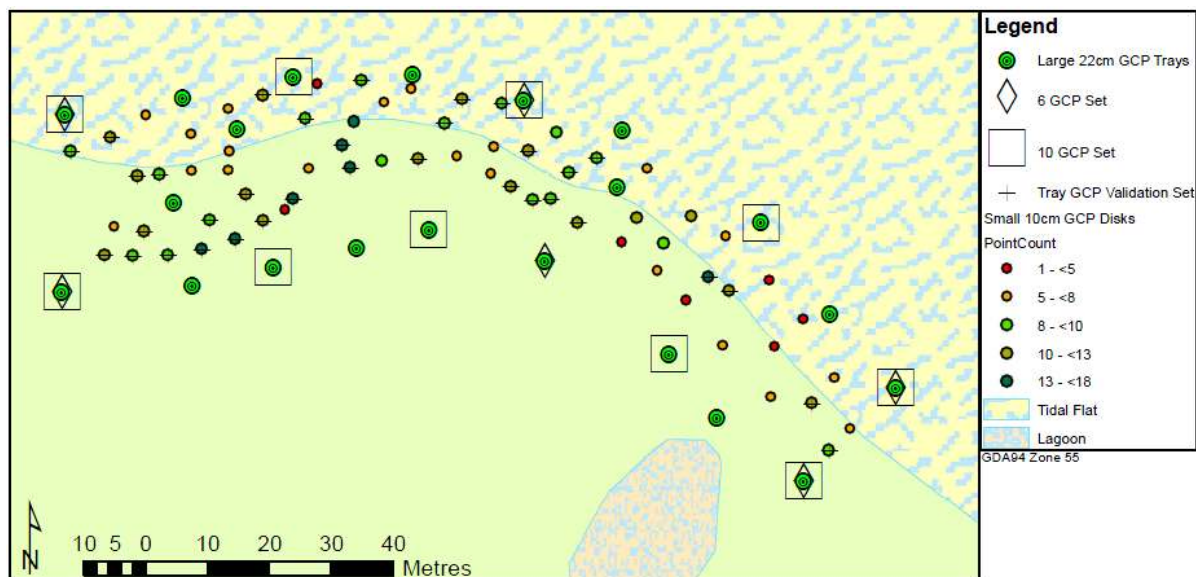
2.4. UAV-MVS

The first stage in the UAV-MVS process is feature extraction. Automated methods rely on features that can be distinguished, described, and matched in multiple views of a scene. This is done using the method described in Snavely *et al.* [21] and Snavely *et al.* [22] whereby a least squares bundle adjustment is performed based on the matching of SIFT features from down-sampled versions of the images. Lowe [13] describes the SIFT process as follows. A 128 element SIFT feature vector (or invariant descriptor vector) is created for each interest point in the image that is determined to be invariant to scale and orientation. The vector describes a chosen stable keypoint and is designed to reduce the effects of illumination and shape distortion. A database of these keypoints is then created and the matching process exhaustively compares each feature from a new image to all features in the database. Candidates are chosen based on Euclidean distance of their feature vectors using a nearest neighbour algorithm. A typical image can contain thousands of SIFT keypoints [13,53].

The matching of these features across overlapping photography produces a sparse set of 3D coordinates of the surface features, the position and orientation of the camera, and radial distortion parameters for each photograph. These outputs from the bundle adjustment are based on the lower resolution images. The PMVS2 software can be used to "fill in" or "densify" the point cloud [4]. However, this is usually done using the down-sampled imagery rather than the original full resolution imagery, which potentially reduces the density and accuracy of the final point cloud.

Our UAV-MVS process improves the densification by utilising the full resolution imagery in the PMVS2 process. As portrayed in Figure 4, the process extracts SIFT features (in fact "SIFTFast" [54] features) from a reduced resolution dataset and performs the bundle adjustment to retrieve a sparse point cloud and camera parameters. We then transform the coordinates of the sparse point cloud and the camera coordinates to match their equivalent values for the full resolution imagery,

i.e., essentially scaling up the coordinate system. The radial distortion of the full resolution images is removed and these images are then processed with PMVS2 resulting in a dense set of 3D coordinates, including point normals. To evaluate the point derivation performance increase and assess the increase in computation time, PMVS2 was run on down-sampled imagery and full resolution imagery. The point cloud produced (see example point cloud from the full resolution imagery in Figure 5) is in an arbitrary reference frame and must be transformed into a real-world coordinate system via a Helmert transformation.

Figure 4. The UAV-MVS point cloud generation process. The key difference from the standard work flow is at Step 6 where the full resolution imagery is undistorted and provided to PMVS2 for point cloud densification.
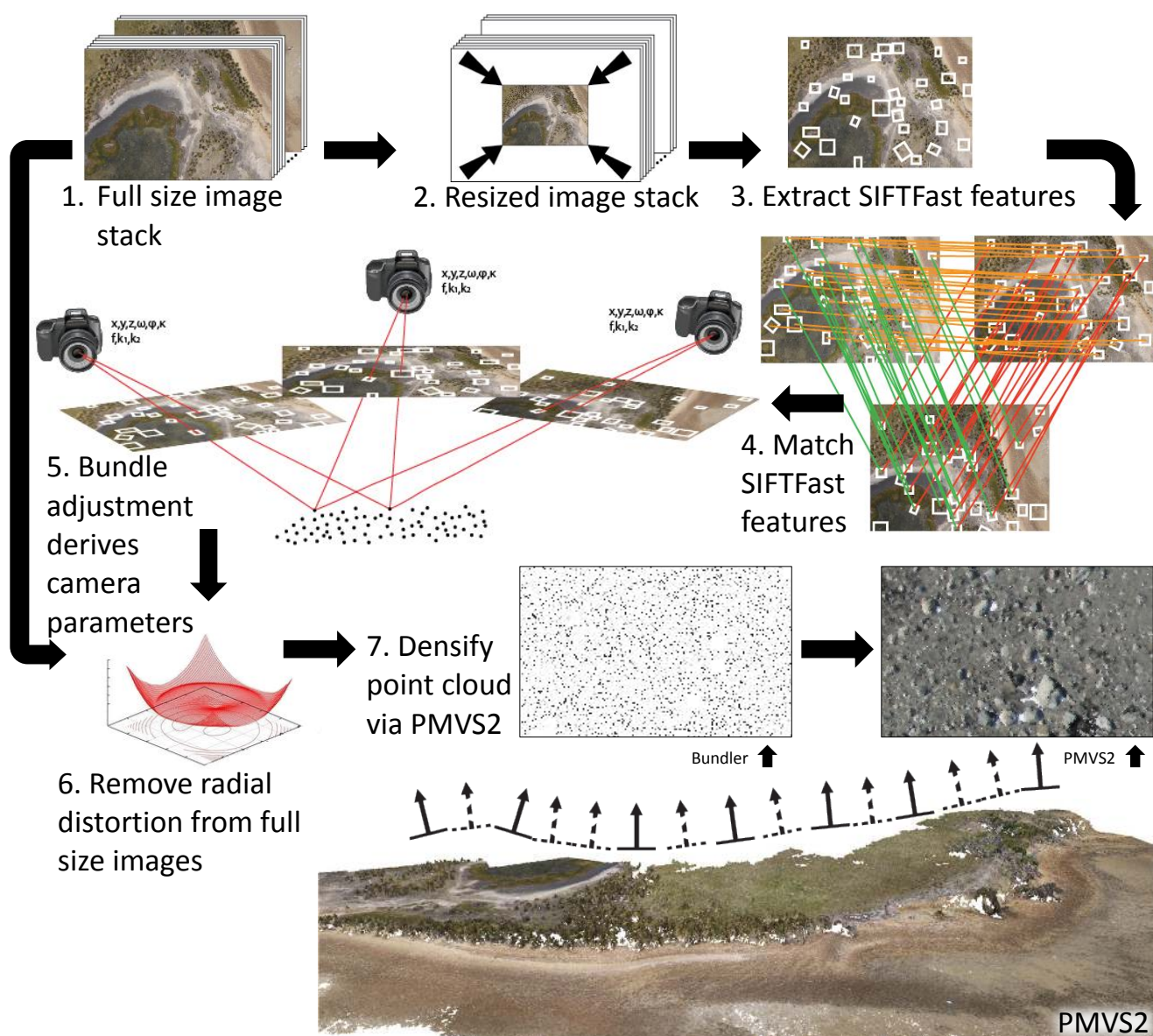
Figure 5. A dense UAV-MVS point cloud after PMVS2 processing with full resolution imagery. The majority of the surface is represented in the cloud at <1–3 cm point spacing. The patches with no points are either scrub bush or tussock grass. The erosion scarp is usually bare earth (see Figure 2) and is well represented in the cloud.
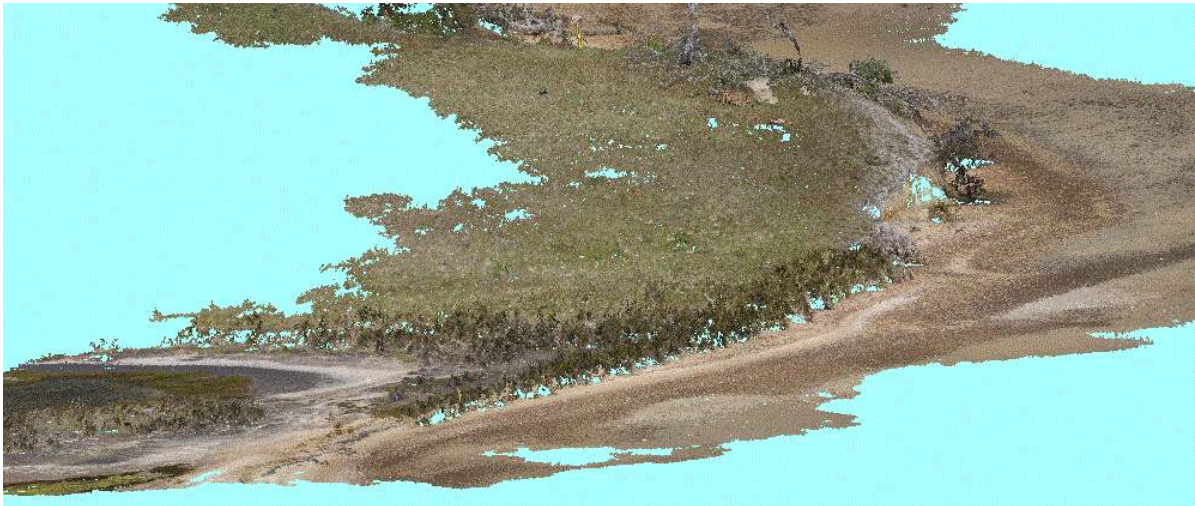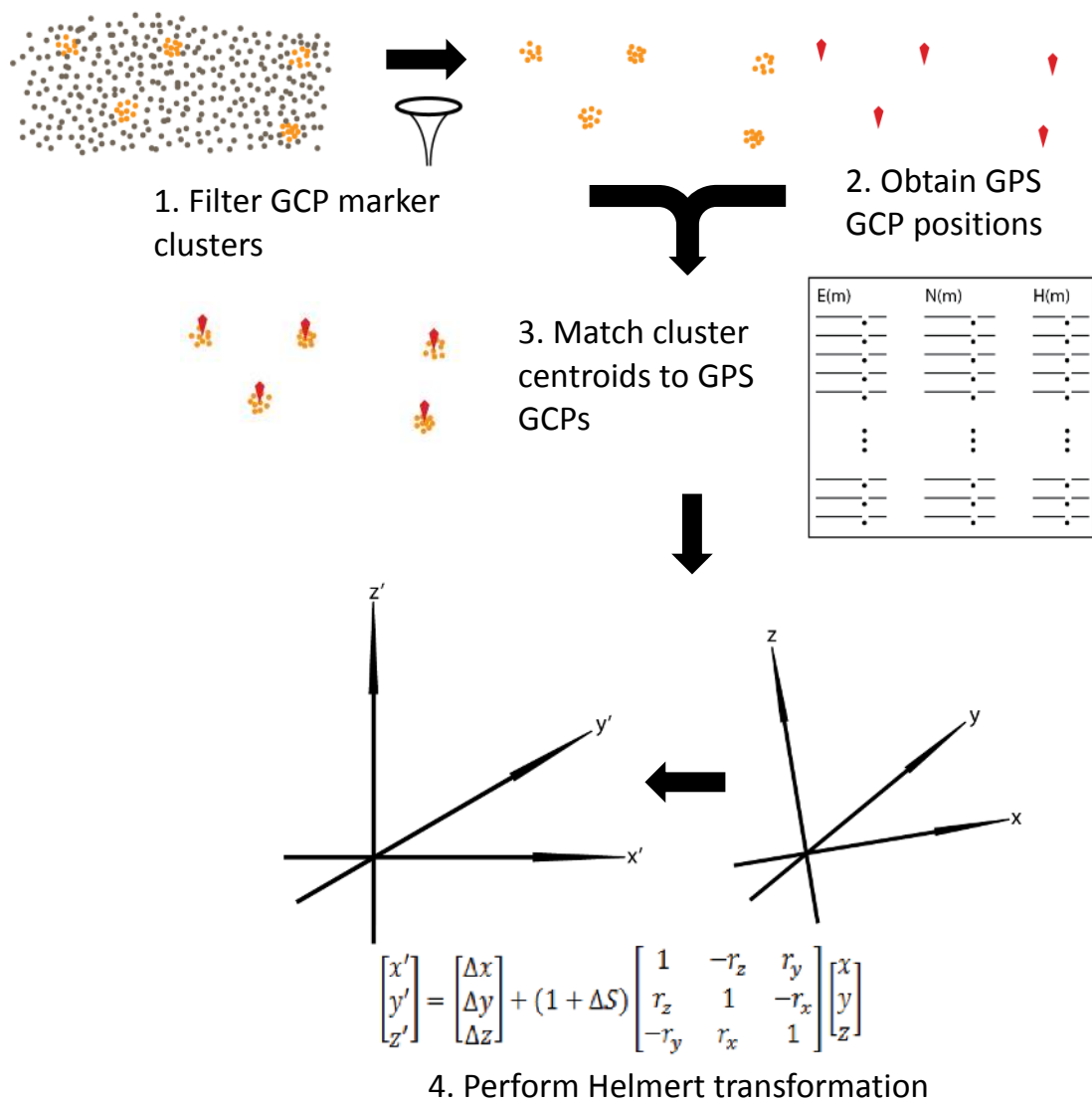

The georeferencing of the point cloud can be done in a number of ways. The simplest and least accurate method is direct georeferencing. This is done by geotagging the photography using the navigation-grade GPS on-board the UAV with approximate GPS locations of the time-synchronised camera at the moment of capture. These coordinates are then used to calculate the Helmert transformation parameters by matching the camera coordinates in the arbitrary reference frame to the corresponding GPS locations. The second method, which shall be referred to as "semi-automatic GCP georeferencing" (portrayed in Figure 6), analyses the colour attributes of the points in the point cloud and extracts the point subsets that match the colour of the orange GCP disks. This colour is based on a threshold collected from a selection of images of the disks (*i.e.*, disks are located in a random set of images and a colour picker is used to calculate an RGB average for the disks). The threshold is applied to the Euclidean distance for each point in RGB colour space to find points that match the disk colour. When all disk points are extracted, the reference points for the point clusters (an example of which is shown in Figure 7(a)) need to be determined to identify the centre coordinate for each disk. An alternative approach may be to use least squares template matching [55–57] or ellipse fitting [58] to determine corresponding GCP locations in multiple images and then compute 3D centre point coordinates in the arbitrary coordinate system based points in the cloud (found using cluster extraction) and their matched feature descriptor vectors (containing corresponding image coordinates). This has not been attempted here and is being considered for future studies. The automated extraction of GCP clusters has potential, particularly if GCP target design is improved further. The approach will therefore be used here to evaluate its feasibility and the resulting centre location determination accuracy.

A third method, which shall be referred to as "manual GCP georeferencing", produces the transformation parameters based on manually selected point clusters representing the large GCP trays (see Figure 7(b)). The Helmert transformation derived from the large GCP trays can be validated against

the cluster centres for the automatically extracted orange GCP disks. As with the automated approach the cluster centres are calculated and matched to the GCP positions.

Figure 6. The UAV-MVS georeferencing process. The filter in Step 1 can either be manual or automatic. The match in Step 3 could either be based on cluster centroid or cluster mean. In Step 4 a Helmert transformation is derived for transforming the point cloud or generated DSMs.

1. Filter GCP marker clusters

2. Obtain GPS GCP positions

3. Match cluster centroids to GPS GCPs

$$
\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} + (1 + \Delta S) \begin{bmatrix} 1 & -r_z & r_y \\ r_z & 1 & -r_x \\ -r_y & r_x & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}
$$

4. Perform Helmert transformation

Figure 7. GCP Clusters in the point cloud used for georeferencing by matching cluster centres to GCP locations. (**a**) A small ~10 cm orange GCP disk. The orange points can be extracted from the cloud by applying a colour threshold. These disks do not result in clusters with many points when flying at ~50 m, larger disks or cones are now considered more suitable unless flying lower or for terrestrial MVS; (**b**) A large 22 cm GCP tray. The GCP tray clusters were manually extracted from the point cloud due to their varying colour. Future studies will ensure these GCP trays (or cones) are designed and painted so that they result in dense clusters of many points and can be found automatically.

(**a**)   (**b**)
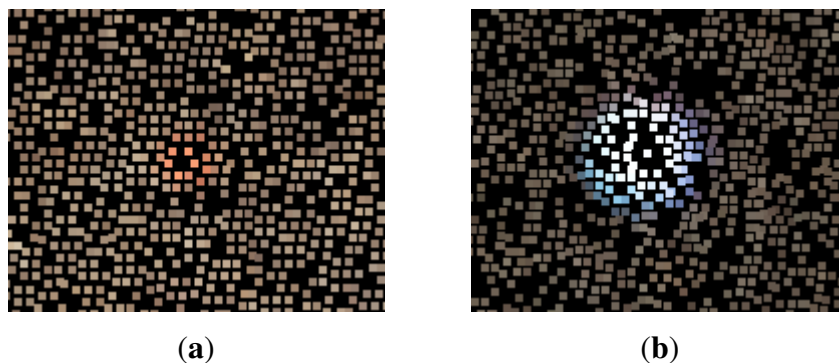
2.5. Accuracy Assessment

The accuracy of the GPS GCP survey impacts on the subsequent transformation, therefore the GPS survey is compared to the Total Station survey results. The initial assessment relates to the choice of mean or centroid cluster centre. To assess the effect of the cluster centre derivation method on the derived transformations, the 12 best centroid-based and 12 best mean-based transformation results are compared (those with a RMSE of less than 40 mm). Subsequently, an assessment of the layout and number of GCP clusters used to derive the Helmert transformation is conducted by evaluating the results from a number of scenarios (Scenarios 1, 2 and 3). In each scenario the transformed cluster centre locations of the validation disks are compared to the GCP reference coordinates (Total Station data). The validation set is a subset of GCPs not used to derive the transformation.

The first and second scenarios use a set of GCP clusters extracted manually from the large trays, *i.e.*, manual GCP georeferencing. All 21 GCP trays are used for the initial transformation derivation. To assess the effect of the number of GCPs on the accuracy of the transformation, ten and six GCP trays distributed across the area are used (see Figure 3). Ideally, the reference dataset would be a continuous coverage over the entire study area, unfortunately this is not available at sufficient accuracy and precision in the study area to allow us to compare with UAV-MVS point clouds. For validation a set of orange disk GCP clusters made up of eight or more points will be used to derive a set of cluster centres. This validation set (see Figure 3) will be transformed using each version of the Helmert transformation derived from the 21, 10, and 6 GCP tray sets respectively. The results will then be compared.

In the first scenario (Scenario 1), only Total Station coordinates for the GCP trays are used in the Helmert transformation and then its accuracy is assessed against the Total Station coordinates of the GCP disks. This provides a "best case" accuracy, even though the additional time required to undertake

a Total Station survey may not be viable for most cases. If required, the Total Station could use tripod mounted prisms instead of pole mounted prisms to further improve the accuracy of the GCP survey. The second scenario (Scenario 2) uses the RTK DGPS tray coordinates for manual GCP georeferencing and the transformed GCP disk cluster centres are compared to the Total Station GCP coordinates.

The third scenario (Scenario 3) assesses the accuracy of our semi-automatic georeferenced UAV-MVS technique. The small orange disk GCPs are automatically extracted from the point cloud and the cluster centres are used to derive a Helmert transformation by matching cluster centres to DGPS GCPs (*i.e.*, semi-automatic GCP georeferencing). The number of points per disk cluster and GCP disk layout are examined and six sets of disk GCPs are chosen to examine the effect of GCP density and distribution, and the impact of cluster point count on accuracy. The GCP disk layout and the effect of poor orange point cluster extraction (*i.e.*, a low number of points in the cluster) can then be evaluated. Similar to the first scenario, these sets are used to derive Helmert transformations which are applied to validation sets of GCP cluster centres, one validation set being automatically selected GCP disks and the other being manual extracted trays. Both validation sets are evaluated to assess whether the semi-automatic cluster extraction or manual cluster selection processes have a systematic influence on accuracy. After transformation the resulting cluster centre coordinates are compared. By changing the distribution and number of GCP disks used to derive the transformation, the optimal number of GCPs and the optimal GCP layout can be evaluated and the minimum number of points in a cluster required to achieve accurate georeferencing can be determined.

3. Results and Discussion

The data collection and processing methods described are the proposed technique for future change monitoring studies, hence there is a need for a clear understanding of the geometric accuracy of the UAV-MVS point clouds. Our georeferencing technique relies on accurate and sufficient ground control and RTK DGPS is the most time efficient means of surveying GCPs. The accuracy of the Total Station survey is within ±10–15 mm in both horizontal and vertical components with respect to fixed control. When these coordinates are compared to the RTK DGPS coordinates they are typically ±17 mm apart and always less than 26 mm horizontally and less than 40 mm vertically. These results correspond to the standard deviations reported by the GPS.

There were three UAV flights flown over the site on the 30th of November 2010, two flights for nadir photography and one flight for oblique photography. Almost 1000 photographs were taken and from this large set a subset of 105 photographs were chosen based on image clarity and content. These images were down-sampled (5,184 × 3,456 pixels ⇒ 2,000 × 1,333 pixels) and processed by Bundler. An initial point cloud containing approximately 230,000 points was extracted (including points for each of the 105 camera locations). The Bundler output was prepared for use with PMVS2 (including transforming the parameters to suit full resolution imagery). The full resolution images were radially undistorted using the calculated coefficients and PMVS2 was run to produce a dense point cloud. The resulting point cloud contained over seven million points. The processing time was 26 h 43 min 54 s (or 96,234 s) on a Dell PowerEdge R815 with four AMD Opteron processors (32 cores at 2.2 GHz), 256 GB of RAM, and 15K RPM SAS drives. The PMVS2 processing time was 11 h 34 min 3 s (or 41,643 s). The resulting point spacing was <1–3 cm. When PMVS2 was run on the down-sampled imagery the resulting point

clouds had only ~1.3 million points (or a ~5–15 cm point spacing) and the PMVS2 processing time was 1 h 33 min 15 s (or 5595 s). The use of full resolution imagery in PMVS2 results in 5 times more points in ~11 times the processing time.

The colour matching parameters for orange GCP disks were determined and 67 GCP disk clusters were extracted. The cloud was manually processed to extract 21 GCP tray clusters. Figure 3 shows the layout of the GCP trays and disks.

3.1. Cluster Centres—Centroid or Mean?

The initial question relates to the choice of cluster centre calculation, *i.e.*, the choice between centroid and mean. If we consider the 24 GCP disk cluster set transformations with a total RMSE of less than 4 mm and analyse the mean RMSE for the "centroid" derived results versus the "mean" derived results (as portrayed in Table 1) there is evidence to favour the mean over the centroid if the overall RMSE (*i.e.*, ENH_{RMSE} or combined Easting, Northing and Height Root Mean Squared Error) is used as the main accuracy metric. However, there is only a 1.1 mm difference. The other accuracy metrics shown are Easting RMSE (E_{RMSE}); Northing RMSE (N_{RMSE}); Height RMSE (H_{RMSE}); and combined Easting and Northing RMSE (EN_{RMSE}).

Table 1. RMSE errors (in millimetres) for Means vs. Centroids. Height is the least accurate dimension. The Easting and Northing error or horizontal position error is higher for the mean based transformations. This is important for GCP matching and georeferencing accuracy, therefore the centroid based transformation is the favoured method for determining cluster centre.

|	E_{RMSE}	N_{RMSE}	H_{RMSE}	EN_{RMSE}	ENH_{RMSE}
Centroid based transformations	15.2	14.4	53.1	14.8	34.4
Mean based transformations	18.0	15.4	49.0	16.7	33.5

These cluster points are filtered based on colour and proximity. If the filter has identified more coloured points on one side of a disk than the other, then the mean will be biased to one side. The centroid, on the other hand, is based on the bounding box of all pixels in a cluster, which is less influenced by the distribution of points within the bounding box. Both methods result in a poor centre calculation when points are only found on one side of a disk and not the other, so perhaps a measure of shape would help highlight good GCP cluster candidates in future studies. As discussed, template matching and ellipse fitting may be alternatives worth considering. The centroid option results in a better EN_{RMSE} and less favourable H_{RMSE} with a 4 mm difference, which impacts on the overall accuracy (*i.e.*, ENH_{RMSE}). The disks are flat and usually placed so that they are reasonably level, therefore the variation in height across the disk should be much less than the variation in horizontal position. The control is captured using DGPS and the predicted accuracy for height measurements is usually ~4 cm, which is an order of magnitude more than the cluster point height difference (~4 mm) seen between the two cluster centre options in that dimension. Based on these considerations the centroid of the clusters

will be used to define cluster centre, as it is more robust to poor cluster point distribution and it results in a more accurate horizontal position of the disk centres.

3.2. Automated GCP Disk Cluster Extraction Performance

Figure 8 provides a histogram of frequency distribution cluster point counts along with the mean, median and standard deviation of those counts. These results indicate that the majority of clusters contain between five and thirteen points, with eight being the average. More than half the clusters contain more than eight points. The scenarios discussed below will compare the effect of using only clusters with more than eight points versus allowing clusters with six or more points to be used.

Figure 8. A histogram of the number of automatically extracted points per cluster representing each of the orange disks. The mean is 8.5 points per cluster, the median is 8 and the standard deviation is 3.5.
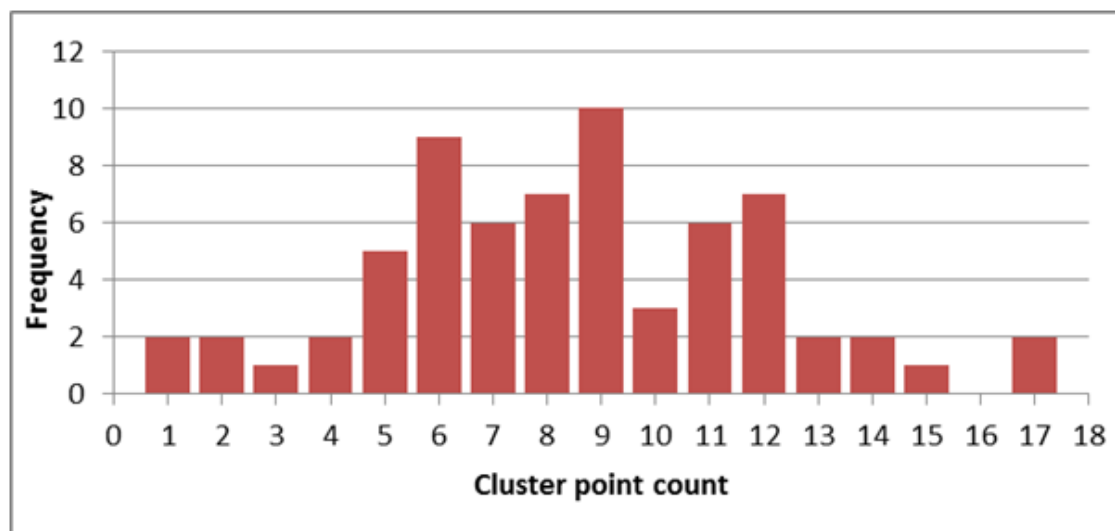
To estimate the accuracy of the georeferenced point clouds and to evaluate the effect on accuracy of GCP layout for Scenarios 1 and 2, the Helmert transformations are compared using the RMSE derived from the comparison of the reference Total Station dataset to the 34 transformed GCP disk cluster centres (*i.e.*, those with eight or more points in a cluster, see GCP tray validation set in Figure 3).

3.3. Scenario 1 and 2

Scenario 1 tests the accuracy of the georeferenced point cloud based on the manually selected GCP tray clusters Helmert transformation (Table 2) and a Total Station GCP survey. Scenario 2 uses the manually selected GCP tray clusters Helmert transformation (Table 3) and a DGPS GCP survey for the accuracy assessment. The comparative accuracy of the three transformation outcomes for the two scenarios is summarised in Tables 4 and 5. The distribution and orientation of these errors were visualised in 3D in Eonfusion [59], allowing the visual assessment of the X, Y, and Z components of the error. Two example views are shown in Figure 9 for the residuals for the GCP disks transformed using the tray centroid transformation for all 21 trays (Figure 9(a)) and for 6 trays (Figure 9(b)).

Table 2. Scenario 1 Helmert transformation results (translation parameters are in metres, rotation parameters are in degrees and accuracies are in millimetres). Only Total Station coordinates for the GCP trays are used in this Scenario, its accuracy is assessed against the Total Station coordinates of the GCP disks.

Description	Tx	$\pm$	Ty	$\pm$	Tz	$\pm$	Rx	$\pm$	Ry	$\pm$	Rz	$\pm$	Scale	$+/-$
All trays	536, 154.565	61.1	5, 262, 637.035	98.2	30.916	68.6	-6.216	1.1	-18.8783	2.5	-32.9718	0.9	9.4409	8.2
10 trays	536, 154.522	108.6	5, 262, 636.977	169.9	30.837	118.6	34.6250	1.9	9.4528	4.3	-73.8128	1.4	9.4383	13.4
6 trays	536, 154.401	154.2	5, 262, 636.794	244.2	30.6975	165.2	3.2108	2.5	-3.1168	6.2	-48.6806	1.9	9.4352	17.8

Table 3. Scenario 2 Helmert transformation results (translation parameters are in metres, rotation parameters are in degrees and accuracies are in millimetres). This scenario uses the RTK DGPS tray coordinates for manual GCP georeferencing and compares the transformed GCP disk cluster centres to the Total Station GCP coordinates.

| Description	Tx	$\pm$	Ty	$\pm$	Tz	$\pm$	Rx	$\pm$	Ry	$\pm$	Rz	$\pm$	Scale	$\pm$
All trays	536, 154.554	60.7	5, 262, 637.027	97.6	30.947	68.1	-56.4816	1.1	-31.4445	2.5	-32.9719	0.9	9.4415	8.1
10 trays	536, 154.511	108.1	5, 262, 636.970	169.1	30.870	118.1	-40.7732	1.8	3.1694	4.3	-42.3968	1.4	9.4389	13.4
6 trays	536, 154.392	152.8	5, 262, 636.792	242.0	30.732	163.7	3.2107	2.5	-3.1168	6.1	-48.6806	1.9	9.4358	17.6

Table 4. Scenario 1 result for manually selected tray transformation validation against Total Station GCP disks (accuracies in millimetres). Total Station coordinates for the GCP trays are assessed against the Total Station coordinates of the GCP disks.

| Description	GCP Count	Test Count	E_{RMSE}	N_{RMSE}	H_{RMSE}	EN_{RMSE}	ENH_{RMSE}
All trays	21	34	28.1	18.7	49.2	23.4	34.4
10 trays	10	34	67.5	43.8	102.9	55.6	75.4
6 trays	6	34	143.0	97.0	171.0	120.0	140.4

Table 5. Scenario 2 result for manually selected tray transformation validation against DGPS GCP disks (accuracies in millimetres). In this scenario RTK DGPS tray coordinates are used to transform GCP disk cluster centres. These are assessed against the Total Station GCP coordinates.

Description	GCP Count	Test Count	E_{RMSE}	N_{RMSE}	H_{RMSE}	EN_{RMSE}	ENH_{RMSE}
All trays	21	34	36.8	19.6	21.0	28.2	27.0
10 trays	10	34	76.9	43.8	73.6	60.3	66.5
6 trays	6	34	153.2	97.5	143.7	125.3	133.7

Figure 9. Eonfusion screen captures of 3D residuals for the validation GCP set (red arrows of residuals for each GCP are scaled by a factor of 20). The underlying surface model is derived from the UAV-MVS point clouds (the two holes in the foreground are due to dead scrub bushes resulting in no points). The view angle is from the west looking down on the site. (**a**) The 21 tray set (*i.e.*, All trays). The largest horizontal residuals of ~25 cm occur at either end of the study area (vertically the largest residuals are as high as ~40 cm) whilst the majority of the residuals are ~14 cm. The smallest residuals occur on the beach; (**b**) The 6 tray set. The largest residuals of ~−31 cm occur in the central portion of the study area near the steep scarp whilst the majority of the residuals are ~−14 cm. Again, the smallest residuals occur on the beach.
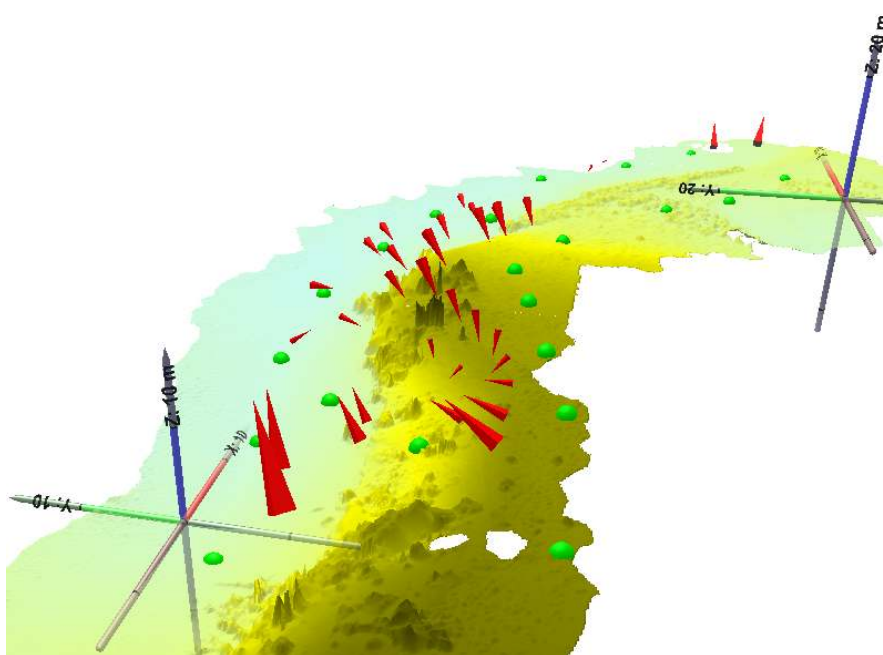
(**a**)
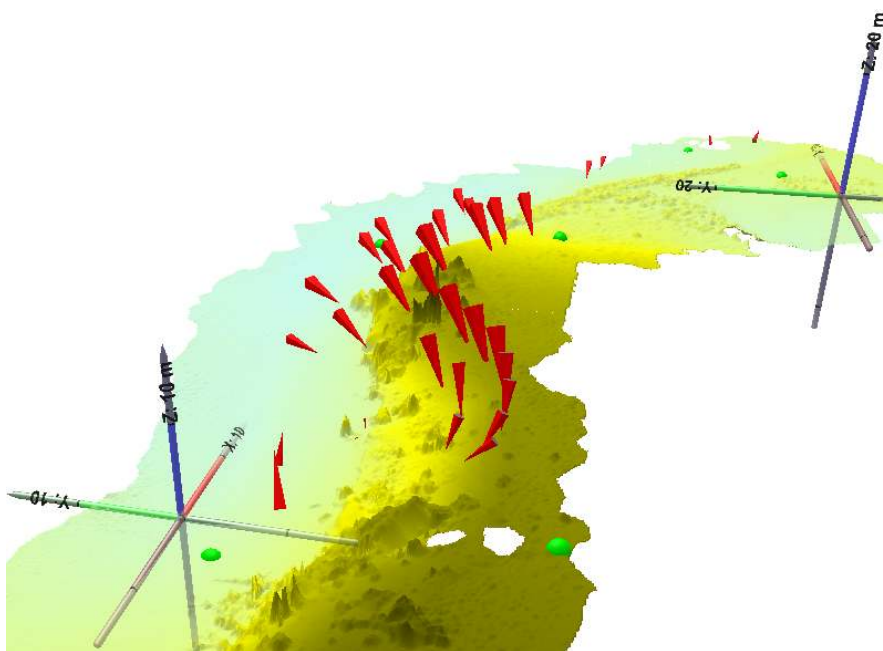
(**b**)

The higher accuracy Total Station survey of the GCP trays was expected to result in a more accurate transformation. However, the GPS survey surprisingly showed a slightly higher accuracy (7 mm difference in ENH_{RMSE}). The EN_{RMSE} is lower in all three GPS-based transformations (approximately 0.5 mm more accurate). The H_{RMSE} is driving the overall accuracy down, similar to what occurred in the cluster centre centroid versus mean comparison. The error estimates for each of the DGPS GCP derived Helmert transformation parameters (Table 3) are slightly better than the error estimates for each of the Total Station GCP derived Helmert transformation parameters (Table 2). The differences are small, however, as can be seen in the 3D residual portrayals (Figure 9), these slight differences and the often major differences in the parameter values can affect the transformation results by millimetres. Figure 9(a,b) shows that removing the majority of the GCPs from the transformation has a significant impact on the error in the central portion of the transformed point cloud. This region coincides with the portion of the site with most topographic relief. In both scenarios, the number of GCPs used has a major impact on the accuracy. The size of the error doubles in each case, from <35 mm to >75 mm in scenario 1 and <30 mm to >65 mm in scenario 2; and finally to ~140 mm and ~130 mm respectively when only 6 GCPs are used.

3.4. Scenario 3

The question that arises from the previous scenarios relates to an optimal GCP distribution and number of GCPs. Scenario 3 was developed to evaluate GCP layout and the success of automated orange disk cluster extraction. For this scenario, a number of GCP disk subsets were used to derive transformations via semi-automated georeferencing and the results compared to two validation sets, *i.e.*, the GCP tray dataset and the set of the GCP disks that were not used to derive the transformation and that had a cluster point count of eight or more.

Figure 10 portrays the chosen GCP sets and the number of points in the clusters. Table 6 provides the derived Helmert transformation results, this set of transformations was applied to the two validation sets. Table 7 compares the validation sets of the transformed GCP disk cluster centres to the corresponding Total Station coordinates of the validation GCPs. Similarly, Table 8 compares the transformed centres of the manually selected tray clusters to the reference data validation GCPs. Figure 11 compares the RMSE of the two validation scenarios. The resulting transformed validation sets show that the automatically extracted disk clusters provide a better georeferencing accuracy, the maximum ENH_{RMSE} is approximately <5 mm in all sets except set (b) (Figure 11); this effect is similar to the results seen in the other scenarios. The choice of cluster extraction method (manual or semi-automatic) has a systematic impact on accuracy. The impact of cluster density and distribution can therefore be evaluated by examining either validation set result.

The four remaining GCP sets test the effect of fewer GCPs where set (c) and set (e) contain a cluster with six points whereas sets (d) and (f) also have an additional four GCPs in the central portion of the study area. In some cases the removal of the six point cluster improves accuracy (Table 7) whereas in others it reduces accuracy (Table 8). The disk validation set shows a more accurate result, particularly in the horizontal dimension. The height dimension is the major contributor to the overall error. Set (f) using disk validation is by far the most accurate of these three options in the horizontal dimensions (EN_{RMSE} of 1 mm) and the H_{RMSE} is 59 mm which is similar to other H_{RMSE} values for the other

four sets. Removing disks with relatively few points (<8) might improve the overall accuracy, however, this reduction will result in fewer available GCP clusters to contribute to the transformation, which could ultimately lead to a poorer fit of the transformation model. Due to this potential impact, and due to the less than definitive results, it may be better to allow these six point clusters to remain in the transformation derivation. In addition, the shape of the cluster may need to be measured to help rank the clusters and discard those that are not circular enough in shape. The size and colour of GCP targets is important. The ~10 cm disks often result in GCP disk clusters of fewer than eight points. This is influenced by both the disk size and by the height of surrounding vegetation and other occluding surfaces. The accuracy of the cluster centre calculation is therefore affected. The larger 22 cm trays with a higher percentage of painted surface area might provide more accurate cluster representations in the generated point cloud.

Figure 10. GCP disk layouts, (**a**) Dense GCP coverage; (**b**) Very sparse GCP coverage; (**c**) GCPs along edge (≥6 cluster points); (**d**) GCPs along edge (≥8 cluster points); (**e**) GCPs along edge and within (≥6 cluster points); (**f**) GCPs along edge and within (≥8 cluster points). The disk distribution suffers when GCPs are removed due to low point counts.
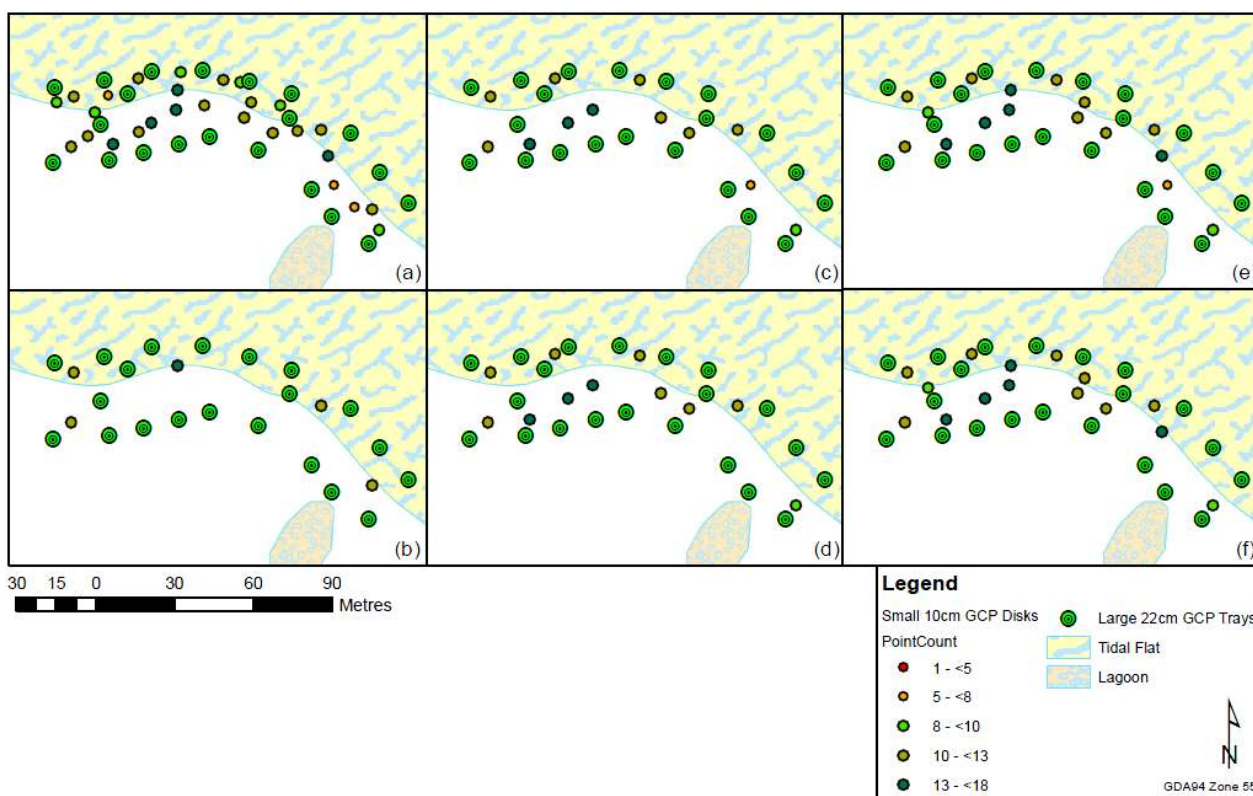
Table 6. Scenario 3 Helmert transformation results (translation parameters are in metres, rotation parameters are in degrees and accuracies are in millimetres). In this scenario, the small orange disk GCPs are automatically extracted from the point cloud and the cluster centres are used to derive a Helmert transformation by matching cluster centres to DGPS GCPs.

| Description	Tx	±	Ty	±	Tz	±	Rx	±	Ry	±	Rz	±	Scale	±
Dense GCP coverage	536, 154.462	39.3	5, 262, 636.876	73.4	30.905	46.4	−6.2140	0.8	−18.8730	2.0	−58.1048	0.6	9.4474	5.9
Very sparse GCP coverage	536, 154.393	94.9	5, 262, 636.718	193.5	30.812	104.6	0.0695	1.8	−0.0215	5.4	−45.5388	1.4	9.4445	13.1
GCPs along edge (≥6 cluster points)	536, 154.484	64.5	5, 262, 636.881	117.0	30.935	73.9	−15.6391	1.3	9.4484	3.3	−36.1137	1.0	9.4451	9.7
GCPs along edge (≥8 cluster points)	536, 154.483	68.4	5, 262, 636.875	125.5	30.941	79.1	0.0689	1.4	−0.0236	3.5	−39.2554	1.2	9.4465	11.0
GCPs along edge and within (≥6 cluster points)	536, 154.468	50.9	5, 262, 636.866	96.1	30.928	59.4	12.6356	1.0	−6.3064	2.7	−26.6889	0.8	9.4479	7.7
GCPs along edge and within (≥8 cluster points)	536, 154.466	53.0	5, 262, 636.860	101.7	30.934	62.1	−12.4972	1.1	−6.3063	2.8	−58.1050	0.9	9.4495	8.5

Table 7. Result for automatically extracted GCP disk cluster transformation (based on subsets of GCP disks) validated against GCP disks (accuracies in millimetres), see Figure 10 for mapped distributions.

Description	Map	GCP Count	Test Count	E_{RMSE}	N_{RMSE}	H_{RMSE}	EN_{RMSE}	ENH_{RMSE}
Dense GCP coverage	a	27	13	15.2	3.0	40.0	9.1	24.8
Very sparse GCP coverage	b	5	31	87.9	77.6	38.7	82.7	71.3
GCPs along edge (≥6 cluster points)	c	12	24	15.5	1.3	63.1	8.4	37.5
GCPs along edge (≥8 cluster points)	d	11	24	9.6	1.7	61.7	5.7	36.1
GCPs along edge and within (≥6 cluster points)	e	16	21	6.6	2.8	59.9	4.7	34.8
GCPs along edge and within (≥8 cluster points)	f	15	21	0.7	1.3	59.1	1.0	34.1

Table 8. Result for manually extracted GCP tray cluster transformation (based on subsets of GCP disks) validated against manually extracted GCP trays (accuracies in millimetres), see Figure 10 for mapped distributions.

Description	Map	GCP Count	Test Count	E_{RMSE}	N_{RMSE}	H_{RMSE}	EN_{RMSE}	ENH_{RMSE}
Dense GCP coverage	a	27	21	8.1	22.6	41.0	15.4	27.5
Very sparse GCP coverage	b	5	21	64.8	47.7	44.0	56.3	53.0
GCPs along edge ($\geq$6 cluster points)	c	12	21	6.3	25.4	62.9	15.9	39.3
GCPs along edge ($\geq$8 cluster points)	d	11	21	13.8	28.9	61.0	21.3	39.8
GCPs along edge and within ($\geq$6 cluster points)	e	16	21	17.0	22.8	59.7	19.9	38.2
GCPs along edge and within ($\geq$8 cluster points)	f	15	21	24.6	24.5	58.5	24.5	39.3

Figure 11. Comparison of RMSE for each of the automatically extracted GCP disk cluster transformations assessed against remaining GCP disks (blue) and GCP trays (red). Set (**a**) (27 GCPs) performs the best due to the distribution and density of control. Set (**b**) (5 GCPs) performed poorly as expected. The remaining sets show mixed results, the differences between sets (**c**) and (**d**) and sets (**e**) and (**f**) are not definitive. This may suggest the number of GCPs is more important than avoiding clusters with only six or seven points.
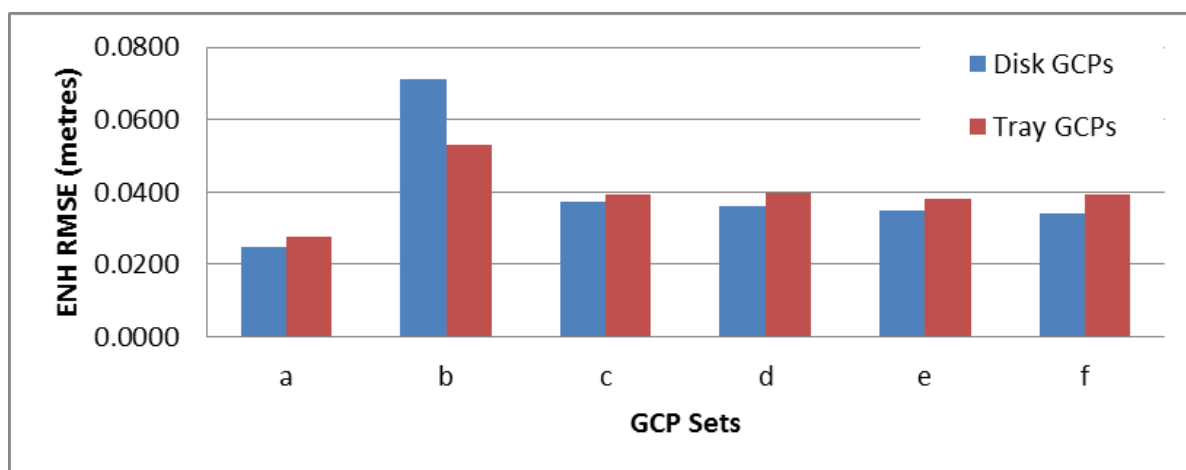
3.5. GCP Distribution

The georeferencing accuracy is strongly influenced by GCP distribution and to a lesser degree by the cluster centre to GCP match. Based on this assessment the best distribution of GCPs is evenly distributed throughout the focus area with a spacing of one fifth to one tenth the UAV flying height (AGL). The terrain variation is important and GCPs should be closer together in steeper terrain. The GCP targets should be clearly visible at the chosen flying height, camera resolution and focal length (>10 cm in diameter for a 40–50 m flying height with the Canon 550D), and they should be visibly different in colour to the surrounding landscape.

3.6. Applications and Limitations

SfM was developed mainly for 3D reconstruction of buildings and other objects from overlapping photography. Examples include modelling tourist destinations captured by hundreds of people who made their photos available on community Internet sites and modelling from photographs and video footage for applications such as architecture, archaeology, robotics and computer graphics. UAV-MVS point clouds have a great deal of potential due to their high point density. This results in an extremely detailed record of the surface at the time of data capture. A major limitation of the process is that the point clouds generated by the UAV-MVS do not represent areas in the landscape where vegetation is dense and complex (such as dead or dry bush with many overlapping branches) and when the surface has a homogeneous texture (e.g., water or a tin roof). These features do not provide the visible attributes needed for algorithms such as SIFT [13]. Techniques are emerging that may overcome these problems [60,61].

Natural environments present a range of complexities, including variable vegetation cover, strong topographic relief and variability in texture. Future studies will need to assess the impact of these complexities on the accuracy of the generated point clouds as landscape snapshots. Unlike LiDAR, the technique is not well suited to penetrating vegetation and, therefore, in vegetated areas it may not produce an accurate DEM when applying ground filtering algorithms [8,12]. In applications where the ground is not the focus, the point clouds can provide a very detailed picture of the surface/terrain. The technique is well suited to canopy monitoring, particularly when combined with LiDAR derived DEMs. Furthermore, in areas where vegetation is sparse such as along the coast, on mine sites and on farm land, the technique offers affordable hyperspatial and hypertemporal data.

4. Conclusions

This study presented an assessment of the accuracy and applicability of point clouds derived by multi-view stereopsis (MVS) based on Unmanned Aerial Vehicle (UAV) photography for natural landscape mapping and monitoring. The UAV-MVS technique generates dense point clouds (1–3 cm point spacing) of natural environments using Structure from Motion (SfM) techniques to process imagery captured from a micro-UAV and georeferences the derived point cloud using Differential Global Positioning System (DGPS) surveys of ground control points (GCPs). In general, the use of UAV-MVS for 3D surface reconstruction and monitoring of natural landscapes has a lot of potential. There have been previous studies that have looked at assessing the accuracy of similar techniques. However, this is the first attempt to quantify the accuracy of the whole data capture and georeferencing process applied to a natural landscape. We developed new additions to existing SfM workflows that allow for full resolution imagery to be used instead of down-sampled imagery, resulting in denser point clouds (~80% increase in point density for an 87% increase in processing time based on 12 Mega-pixel versus 3 Mega-pixel imagery). We present a case study of UAV-MVS point clouds for a natural coastal area in southeastern Tasmania, Australia. Accurate and dense 3D point clouds are required to quantify the impact of erosion events on the coastline. The main objective of this study was to test the geometric accuracy of the point clouds based on Real-Time Kinematic (RTK) DGPS and Total Station surveys of GCPs. We found that, when flying at 40–50 m, an accuracy of 2.5–4 cm can be achieved provided sufficient, clearly visible

GCPs are distributed evenly throughout the study area, and the flight planning ensures a high degree of overlap (70%–95%) between images. The accuracy obtained by UAV-MVS when properly controlled is, in fact, within the magnitude of accuracy achievable by DGPS. In this study the distribution and number of GCP disks used to derive the transformation was varied to assess the optimal GCP layout, the number of GCPs, and the best methods for automated GCP extraction. The use of RTK DGPS to survey the ground control compared favourably to the Total Station survey results. The estimated accuracy of the Total Station data is ~1 cm in position and ~2 cm in elevation compared to DGPS accuracy of ~2.5 cm and ~4 cm in position and elevation respectively. Semi-automatic GCP point cluster extraction where clusters have greater than six points can allow a cluster centroid to be calculated. When GCP targets are well placed, large (>10 cm in diameter) and visibly different in colour to the surrounding landscape, this cluster extraction will be more successful. Future studies will investigate improving GCP design and matching. Semi-automatic cluster extraction enables georeferencing to sufficient accuracy such that sub-decimetre terrain change can be detected and monitored. Assessing the accuracy of these point clouds was an essential first step towards proving the viability of the UAV-MVS technique for fine-scale landform change monitoring. In particular, coastal erosion monitoring requires sub-decimetre dense and accurate 3D point clouds. Fine scale change mapping cannot be achieved to sufficient spatial and temporal resolution with traditional airborne surveys and satellite sensors. The study site used in this paper will be monitored in the future to assess whether subtle coastal erosion in a sheltered estuary can be used as a climate change indicator. The MVS technique used fails to find sufficient features for matching in areas of complex vegetation and where surfaces have a homogeneous texture, as these result in gaps or sparse areas in the point cloud. The technique does not penetrate dense vegetation and the resulting point cloud contains very few ground points beneath vegetation. Despite these limitations, the techniques have great potential in a wide range of application areas beyond coastal monitoring, including mining, agriculture and habitat mapping, and this accuracy assessment will serve to solidify the viability of the process.

Acknowledgements

References

1. Laliberte, A.; Herrick, J.; Rango, A.; Winters, C. Acquisition, orthorectification, and object-based classification of unmanned aerial vehicle (UAV) imagery for rangeland monitoring. *Photogramm. Eng. Remote Sensing* **2010**, *76*, 661–672.

2. Coulter, L.; Lippitt, C.; Stow, D.; McCreight, R. Near Real-Time Change Detection for Border Monitoring. In *Proceedings of the ASPRS Annual Conference*, Milwaukee, WI, USA, 1–5 May 2011; pp. 9–17.

3. Chao, H.; Cao, Y.; Chen, Y. Autopilots for small unmanned aerial vehicles: A survey. *Int. J. Control Autom. Syst.* **2010**, *8*, 36–44.

4. Furukawa, Y.; Ponce, J. Accurate, Dense, and Robust Multi-View Stereopsis. In *Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, 18–23 June 2007; Volume 1, pp. 1–8.

5. Neitzel, F.; Klonowski, J. Mobile 3D mapping with a low-cost UAV system. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2011**, *38*, 1–6.

6. Rosnell, T.; Honkavaara, E. Point cloud generation from aerial image data acquired by a quadrocopter type micro unmanned aerial vehicle and a digital still camera. *Sensors* **2012**, *12*, 453–480.

7. Turner, D.; Lucieer, A.; Watson, C. Development of an Unmanned Aerial Vehicle (UAV) for Hyper Resolution Vineyard Mapping Based on Visible, Multispectral, and Thermal Imagery. In *Proceedings of the 34th International Symposium on Remote Sensing of Environment (ISRSE34)*, Sydney, Australia, 11–15 April 2011.

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

9. 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 on Remote Sensing of Environment (ISRSE34)*, Sydney, Australia, 11–15 April 2011.

10. Monserrat, O.; Crosetto, M. Deformation measurement using terrestrial laser scanning data and least squares 3D surface matching. *ISPRS J. Photogramm.* **2008**, *63*, 142–154.

11. Hartley, R.; Mundy, J. The relationship between photogrammetry and computer vision. *Proc. SPIE* **1993**, *14*, 92–105.

12. Remondino, F.; El-Hakim, S. Image-based 3D modelling: A review. *Photogramm. Rec.* **2006**, *21*, 269–291.

13. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.

14. Juan, L.; Gwun, O. A comparison of SIFT, PCA-SIFT and SURF. *Int. J. Image Process.* **2009**, *3*, 143–152.

15. Mikolajczyk, K.; Schmid, C. Scale & affine invariant interest point detectors. *Int. J. Comput. Vis.* **2004**, *60*, 63–86.

16. Bay, H.; Tuytelaars, T.; Van Gool, L. Surf: Speeded up Robust Features. In *Proceedings of 9th European Conference on Computer Vision*, Graz, Austria, 7–13 May 2006; pp. 404–417.

17. Strecha, C.; Bronstein, A.; Bronstein, M.; Fua, P. LDAHash: Improved matching with smaller descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 66–78.

18. Ke, Y.; Sukthankar, R. PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 27 June–2 July 2004; pp. 506–513.

19. Triggs, B.; McLauchlan, P.; Hartley, R.; Fitzgibbon, A. Bundle Adjustment—A Modern Synthesis. In *Proceedings of the ICCV '99 Vision Algorithms: Theory and Practice*, Corfu, Greece, 20–21 September 1999; pp. 153–177.

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

21. Snavely, N.; Seitz, S.M.; Szeliski, R. Photo tourism: Exploring photo collections in 3D. *ACM Trans. Graph.* **2006**, *25*, 835–846.

22. Snavely, N.; Seitz, S.M.; Szeliski, R. Modeling the world from internet photo collections. *Int. J. Comput. Vis.* **2007**, *80*, 189–210.

23. Microsoft Corporation. *Photosynth*; Available online: http://photosynth.net/ (accessed on 25 October 2010).

24. AgiSoft. *Agisoft PhotoScan*; Available online: http://www.agisoft.ru/products/photoscan/ (accessed on 1 November 2010).

25. Eos Systems Inc. *PhotoModeler*; Available online: http://www.photomodeler.com/products/photomodeler.htm (accessed on 1 October 2011).

26. Furukawa, Y.; Ponce, J. *Patch-Based Multi-View Stereo Software*; Available online: http://grail.cs.washington.edu/software/pmvs/ (accessed on 1 June 2010).

27. Furukawa, Y. *Clustering Views for Multi-View Stereo (CMVS)*; Available online: http://grail.cs.washington.edu/software/cmvs/ (accessed on 3 December 2010).

28. Furukawa, Y.; Ponce, J. Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1362–1376.

29. Seitz, S.; Curless, B.; Diebel, J. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, NY, USA, 17–22 June 2006; Volume 1, pp. 519–528.

30. Seitz, S.M.; Curless, B.; Diebel, J.; Scharstein, D.; Szeliski, R. *Multi-View Stereo*; Available online: http://vision.middlebury.edu/mview/ (accessed on 12 April 2012).

31. Strecha, C.; von Hansen, C.; Gool, L.V.; Fua, P.; Thoennessen, U. On Benchmarking Camera Calibration and Multi-View Stereo for High Resolution Imagery. In *Proceedins of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, USA, 24–26 June 2008.

32. Strecha, C.; Fransens, R. Wide-Baseline Stereo from Multiple Views: A Probabilistic Account. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 27 June–2 July 2004; pp. I:552–I:559.

33. Strecha, C.; Fransens, R.; van Gool, L. Combined Depth and Outlier Estimation in Multi-View Stereo. In *Proceedins of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, New York, NY, USA, 17–22 June 2006; Volume 2, pp. 2394–2401.

34. Hirschmuller, H. Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. In *Proceedins of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, 20–26 June 2005; Volume 2, pp. 807–814.

35. Hirschmüller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341.

36. Baillard, C.; Zisserman, A. A plane-sweep strategy for the 3D reconstruction of buildings from multiple images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2000**, *33*, 56–62.

37. Vu, H.H.; Labatut, P.; Pons, J.P.; Keriven, R. High accuracy and visibility-consistent dense multi-view stereo. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 889–901.

38. Eisenbeiss, H.; Sauerbier, M. Investigation of uav systems and flight modes for photogrammetric applications. *Photogramm. Rec.* **2011**, *26*, 400–421.

39. Küng, O.; Strecha, C.; Beyeler, A.; Zufferey, J.C.; Floreano, D.; Fua, P.; Gervaix, F. The Accuracy of Automatic Photogrammetric Techniques on Ultra-Light UAV Imagery. In *Proceedings of the International Conference on Unmanned Aerial Vehicle in Geomatics (UAV-g)*, Zurich, Switzerland, 14–16 September 2011; In *IAPRS*; 2011; Volume 38.

40. Vallet, J.; Panissod, F.; Strecha, C. Photogrammtric Performance of an Ultralightweight Swinglet UAV. In *Proceedings of the International Conference on Unmanned Aerial Vehicle in Geomatics (UAV-g)*, Zurich, Switzerland, 14–16 September 2011; In *IAPRS*; 2011; Volume 38.

41. Hirschmüller, H. Semi-Global Matching Motivation, Developments and Applications. In *Proceedins of the Invited Paper at the 54th Photogrammetric Week*, Stuttgart, Germany, 5–11 September 2011; pp. 173–184.

42. Pix4D. Available online: http://pix4d.com/ (accessed on 12 April 2012).

43. BAE Systems. *Next-Generation Automatic Terrain Extraction (NGATE)*; Available online: http://www.socetgxp.com/docs/products/modules/ss_ngate.pdf (accessed on 11 December 2011).

44. Rosnell, T.; Honkavaara, E.; Nurminen, K. On geometric processing of multi-temporal image data collected by light UAV systems. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2011**, *38*, 1–6.

45. Walker, J.P.; Willgoose, G.R. A comparative study of australian cartometric and photogrammetric digital elevation model accuracy. *Photogramm. Eng. Remote Sensing* **2006**, *72*, 771–779.

46. Shrestha, R.L.; Carter, W.E.; Lee, M.; Finer, P.; Sartori, M. Airborne laser swath mapping: ALSM. *Civil Eng.* **1999**, *59*, 83–94.

47. Töyrä, J.; Pietroniro, A.; Hopkinson, C.; Kalbfleisch, W. Assessment of airborne scanning laser altimetry (lidar) in a deltaic wetland environment. *Can. J. Remote Sens.* **2003**, *29*, 718–728.

48. Farah, A.; Talaat, A.; Farrag, F. Accuracy assessment of digital elevation models using GPS. *Artif. Satellites* **2008**, *43*, 151–161.

49. Hodgson, M.; Bresnahan, P. Accuracy of airborne lidar-derived elevation: Empirical assessment and error budget. *Photogramm. Eng. Remote Sensing* **2004**, *70*, 331–339.

50. Vaaja, M.; Hyyppä, J.; Kukko, A.; Kaartinen, H.; Hyyppä, H.; Alho, P. Mapping topography changes and elevation accuracies using a mobile laser scanner. *Remote Sens.* **2011**, *3*, 587–600.

51. Mikrokopter. Available online: http://www.mikrokopter.de/ucwiki/en/MikroKopter (accessed on 28 May 2012)

52. Geoscience Australia. Available online: http://www.ga.gov.au/geodesy/ausgeoid/nvalcomp.jsp (accessed on 11 December 2010).

53. Farenzena, M.; Fusiello, A.; Gherardi, R. Structure-and-Motion Pipeline on a Hierarchical Cluster Tree. In *Proceedings of the IEEE Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference*, Kyoto, Japan, 27 September–4 October 2010; pp. 1489–1496.

54. Libsift Sourceforge Project. Available online: http://libsift.sourceforge.net/ (accessed on 10 September 2010).

55. Zhang, Y.; Zhang, Z.; Zhang, J.; Wu, J. 3D building modelling with digital map, lidar data and video image sequences. *Photogramm. Rec.* **2005**, *20*, 285–302.

56. Zhang, Z.; Wu, J.; Zhang, Y.; Zhang, J. Multi-view 3D city model generation with image sequences. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2003**, *34*, 351–356.

57. Kocaman, S.; Zhang, L.; Gruen, A.; Poli, D. 3D City Modeling from High-Resolution Satellite Images. In *Proceedings of ISPRS Workshop on Topographic Mapping from Space*, Ankara, Turkey, 14–16 February 2006; Volume XXXVI, Part 1/W41.

58. Hanley, H. Geopositioning accuracy of IKONOS imagery: Indications from two dimensional transformations. *Photogramm. Rec.* **2001**, *17*, 317–329.

59. Myriax. Available online: http://www.eonfusion.com (accessed on 2 October 2009).

60. Lu, F.; Ji, X.; Dai, Q.; Er, G. Multi-View Stereo Reconstruction with High Dynamic Range Texture. In *Proceedings of the Computer Vision—ACCV 2010*, Queenstown, New Zealand, 8–12 November 2011; Springer: Berlin, Heidelberg, Germany, 2011; Volume 6493/2011, pp. 412–425.

61. Mičušík, B.; Košecká, J. Multi-view superpixel stereo in urban environments. *Int. J. Comput. Vis.* **2010**, *89*, 106–119.