Towards generation of effective 3D surface models from UAV imagery using open source tools

P. S. Singh[1,*], Mayuri Sharma[2], Victor Saikhom[1], Dibyajyoti Chutia[1], Chirag Gupta[1], Avinash Chouhan[1] and P. L. N. Raju[1]

[1]North Eastern Space Applications Centre, Department of Space, Government of India, Umiam 793 103, India
[2]Department of Computer Science, Assam Don Bosco University, Guwahati 781 017, India

There has been increasing popularity in large scale mapping for deriving 3D surface and elevation models of earth and building structures. The techniques of computer vision comprising feature detections and matching and photogrammetry play an important role in deriving near accurate 3D reconstruction of scenes from 2D images. Since the images captured by the unmanned aerial vehicle (UAVs) are of high resolution, there is need for more sophisticated processing and analysis of the imagery to generate 3D models and other useful imagery products. The open source softwares are excellent tools for research and can be modified or changed to suit our model, as specific or combinations of algorithms behave differently based on the nature of UAV image scene to be processed. Though many algorithms are available for performing feature extractions from images, few studies have been carried out to identify suitable detector algorithms to be used based on the nature of image or scene that the UAV captures. An attempt has been made to understand and analyse the suitability of feature detection and descriptor algorithms for different scene types. This article also describes the popular technique called structure from motion process pipeline for sequential processing of UAV images with high overlapping, which involves the estimation of 3D point clouds from the keypoint correspondences. The relative accuracy of the 3D point cloud derived from our approach is comparable with similar output from other state-of-the-art UAV processing systems and is found to match with high precision.

Keywords: 3D reconstruction, open source, point clouds, remote sensing, structure from motion, unmanned aerial vehicle.

THE unmanned aerial system (UAS) is a drone with payloads such as cameras, sensors and detectors – both imaging and non-imaging. UAVs armed with remote sensing technology are increasingly being used to acquire high resolution spatial data about land cover resources, and provide the environment for processing and analysing remote sensing data. Further, they can be deployed in inaccessible or dangerous areas for mapping. The ability to acquire data from low flying altitude has enabled high resolution, large scale mapping even with consumer grade camera sensors. This has boosted remote sensing groups and has led to several applications in relevant areas. The imagery data product obtained from UAS image post-processing, can immensely help in many applications ranging from large scale building modelling to vegetation structure mapping which can in turn greatly benefit the local planning and development, in the North Eastern region of India owing to its limited road connectivity and physical infrastructure. The output of UAV image processing gives us valuable products such as orthophotos and elevation/surface models at various levels of details. However, there are challenges in processing huge volume of high resolution aerial imagery. The simpler and lighter platforms and sensors of UAV acquisition technologies do not necessarily translate into simpler processing systems. In contrast, more sophisticated processing is required for processing the imagery. The technique of photogrammetry and computer vision[1] has a big role to play for accurate and automatic processing of high resolution imagery captured using UAV. The recent advances in 3D scene reconstruction technique using structure from motion (SfM) approach[2,3] combined with the established rules of classic stereoscopic photogrammetric survey helps in deriving high resolution and precise three-dimensional texture models. Accurate feature detection and matching are critical steps in any UAV data processing pipeline involving 3D scene reconstruction. There are various computer algorithms for feature detectors, descriptors and matching for detecting and describing the different types of features in the image[4]. There is a need to compare the performance of these algorithms against images with different viewpoints, scales, illumination and image compressions[5]. There are popular feature extraction algorithms such as scale invariant feature transform[6] which extracts distinct features which are invariant to

*For correspondence. (e-mail: ss.puyam@nesac.gov.in)

scale and rotation from hundreds of high resolution aerial photographs. The image-based modelling approach using SfM approach is inexpensive and highly automated, but capable of producing highly accurate dense point clouds comparable to that of point cloud generated by a terrestrial laser scanning (TLS) surveys[7] and airborne LiDAR with horizontal and vertical precision in the centimetre range[8]. SfM is best suited for series of unstructured aerial images with high overlapping. It can estimate the 3D sparse point clouds along with positions and orientations of the camera[9]. The noisy sparse 3D point output from SfM is not sufficient for full 3D and high quality surface reconstructions and therefore the need to increase the number of 3D point clouds[10]. The multi view stereo algorithms are then applied to SfM outputs to generate 3D models with many details[11].

We have examined the efficacy of popular algorithms for feature detection using a few scenes of our study area. In this article, we describe an effective approach to UAV data processing using open source-based process pipelines to develop a 3D model of a landslide-affected area in Ribhoi, Meghalaya. The various stages of UAV data processing have been highlighted and described in detail with relevant techniques and algorithms used. Further, the processing output has been analysed for accuracy.

Methodology

For construction of the textured model, we first import the set of 2D images. Then the detection of distinguished keypoints was initiated. In this step, the software also searched for matched points in all the images. For feature detection phase, we used a few of the popular algorithms for detecting features such as edges, corners or blobs. The SIFT[6] was used to build the feature descriptors for each feature or interest point detected. Computer vision techniques were used to accurately match the features. The matched keypoints were then used to reconstruct external orientation and the position of each camera scene gave the 3D sparse point clouds. The 3D points were joined to form triangular facets using surface reconstruction algorithms. Then, the texturing was applied to each facet with the colour information from the original raster giving the final full textured 3D surface model. The schematic block diagram describing the detailed methodology is shown in Figure 1.

Image acquisition

The ready-to-fly quadcopter was fitted with an optical sensor that can capture high-resolution geo-tagged aerial photos and high-definition videos for aerial surveillance and suitable for immediate deployment for near-real time assessment of landslide-affected areas. It can attain a maximum altitude of up to 2 km with a scanning radius of 5 km. With the currently available LiPo battery, it can fly for about 20 min and sufficiently cover an area of 1–1.5 sq km. Images of the target scene were taken from different positions (Figure 2). These images formed the dataset for further processing.

For capturing the images, we used T600 DJI inspire series of UAV. It has Zenmuse X3-FC350 (Figure 3) camera with focal length of 4 mm and an effective resolution of 12.4 mega pixels giving a maximum image size of 4000×3000.

It was flown at a height of 120 m to collect 90 images of the landslide-affected area located in Ribhoi, Meghalaya. The camera has a sensor width of 6.17 mm × 4.55 mm which translates the pixel size of the sensor to 0.0015 mm. Using the formula of GSD = (pixel size of the sensor × flight altitude)/focal length, we obtained high resolution images with GSD of 4.5 cm per pixel. While flying the UAV, we ensured 75% overlapping for the images. With this, we had more than 5 overlapping images for every scene in our area of study (Figure 4). This later helped in generating a large number of repeated feature points in multiple images which were useful in getting denser 3D point clouds.

Feature detection and matching

Initially, features were identified from images and the respective descriptors were constructed for every detected feature[12]. For every descriptor pair, the distance was calculated and marked as matched, if the distance fell below a threshold. Smaller the distance, the better the match. Based on the types of features to be detected, different combinations of detector-descriptor pairs were used to accurately detect the features on every image. The corner or interest point features were better detected by Harris, FAST and Shi-Tomasi. Edge features were well detected by Canny detectors. Among many feature detectors, SIFT and SURF were the most commonly used feature detection techniques[4].

These algorithms select the robust features which are localized, scale-invariant, distinct and repeatable in nature. SIFT descriptors, after identifying a set of keypoints, compute a descriptor vector or feature vector for every keypoint. These feature vectors were then used to align the images. We then established feature keypoints that corresponded to correct matches. The accuracy of the match was known by finding the ratio of the distances to the second nearest neighbour and the first nearest neighbour at some threshold. For it to be called a good match, the ratio has to be high.

Analysis of feature detectors: There have been few studies on effective selection of a particular feature detection algorithm and a feature matching algorithm, especially for 3D scene reconstruction involving 2D

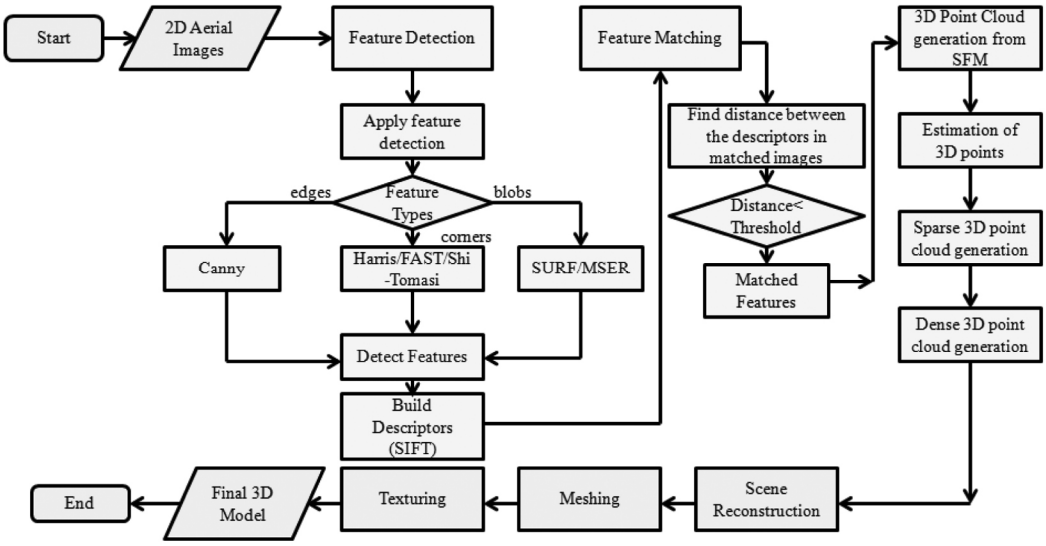

Figure 1. Methodology.

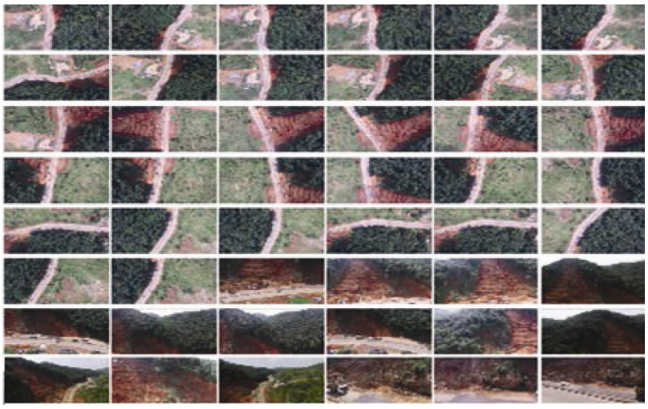

Figure 2. 2D images of landslide affected area in Ribhoi, Meghalaya.

Figure 3. The Zenmuse X3-FC350.

images from UAVs. We did a comparative analysis of popular feature detector algorithms and certain feature matching techniques for chosen scenes. Performance analysis was carried out which suggested the suitability of the algorithm for our applications. We randomly chose three popular feature detection and matching algorithms from VLFeat and OpenCV. The Harris Corner, FAST and Minimum Eigen were compared on the basis of (a) maximum number of robust feature keypoints detected, (b) average computation time, (c) even distribution of keypoints. As observed from Figure 5, MinEigen was able to detect keypoints (in green colour) with uniform distribution in the scene. Among popular feature detectors such as Harris, MinEigen and FAST, the MinEigen was able to collect a larger number of keypoints from the image that contains mostly tree-clad areas. The keypoints are represented by green dots (Figures 5 and 6). Further, MinEigen was able to collect large number of keypoints from the scene when compared to Harris and FAST (Figure 6). The MinEigen performed well for matching the detected features at the scene. At threshold 60, it gave maximum numbers of correct matches when compared to FAST, SURF and SIFT (Figure 7).

Suitability of feature detectors for different scene types captured by our UAV: Although there are numerous feature detectors and descriptors available, there is a need to analyse them properly and measure their performance in terms of the number of robust features detected, the even distribution of keypoints and repeatability of features detected[13]. Specific detectors/descriptors or a combination of them which work well on a scene behave differently on different types of scenes. To understand and analyse this properly, we took three different scene types – Scene-I: Natural scene with tree-clad areas, scene-II: Buildings and artificial structures, scene-III: Mix scene with natural and artificial structures (Figure 8).

Even distribution of detected features is important and helps in uniform 3D reconstruction. In scene-I, the detectors such as SIFT and DoG were able to detect uniform and evenly distributed number of keypoints (Figure 9).

In scene-II, the popular feature detector and descriptor SIFT, failed to detect large number frames when compared to DoG detector, whereas MSER was able to detect

only boundaries of the object and not the interested keypoints (Figure 10).

In scene-III, both SIFT and DoG detected uniform keyframes. The scale invariant Harris Laplace was best suited for detecting corners as well as highly textured keypoints. Further, Harris Laplace was able to detect regions missed by both SIFT and DoG (Figure 11). Figure 12 depicts the number of feature points detected by various detectors for different scene types.

Figure 4. Number of overlapped images for our study area.

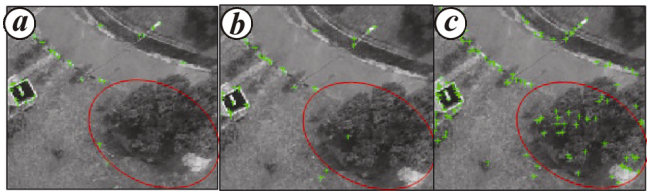

Figure 5. Distribution of keypoints. ***a***, Harris corner; ***b***, FAST; ***c***, MinEigen.

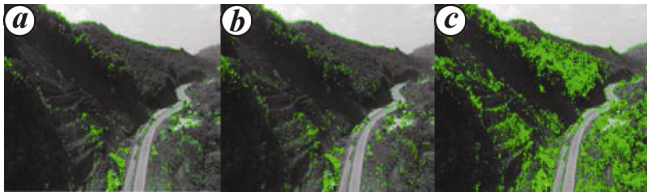

Figure 6. Robust detection of keypoints. ***a***, Harris: 1970; ***b***, FAST: 2784; ***c***, MinEigen: 22433.

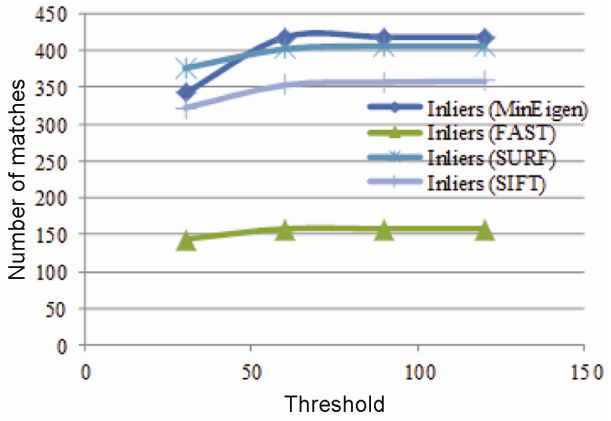

Figure 7. Keypoints matching.

The DoG and Multiscale Hessian were able to detect maximum number of keypoints in all scenes when compared to other detectors.

Once we chose the right feature detector for the scene, the next step was to describe each feature detected and establish tracks or feature correspondence for a pair of features (Figures 13 and 14).

High repeatability of features indicates how efficiently the regions are marked in the scene by the corresponding detector[3]. We observed the changes in repeatability by considering the changes in image viewpoints. The viewpoint-angle changes were set at 30° and 40° for all detectors. The detector with high repeatability and high number of correspondences was the preferred detector. The repeatability was calculated for every reference image-test image pair. At a viewpoint angle of 40°, the repeatability was found to be the highest for all the

Figure 8. Three different scenes captured by our UAV. ***a***, Scene-I; ***b***, Scene-II; ***c***, Scene-III.

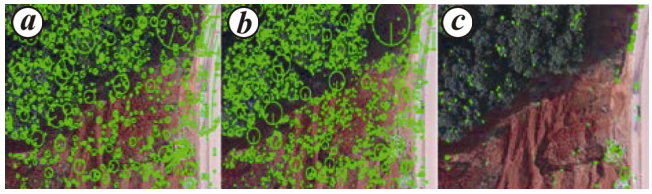

Figure 9. Distribution of the feature frames in scene I. ***a***, SIFT; ***b***, DoG; ***c***, Harris Laplacian.

Figure 10. Distribution of the feature frames in scene II. ***a***, SIFT; ***b***, DoG; ***c***, MSER.

Figure 11. Distribution of the feature frames in scene III. ***a***, SIFT; ***b***, DoG; ***c***, Harris Laplacian.

feature detectors. For scene-II, Harris Laplace was found to have better repeatability with 65% when viewpoint angle was at 40°. In short, our experimental analysis (Figures 11 and 15) confirmed that Harris Laplace must be used for precise detection of features for scene-III.

Point cloud generation

There are general techniques for recovering 3D shape from one or two 2D images which are defined under Shape from X, where X represents the specific 3D recovery technique, viz. stereo, motion, shading, texture, contours, etc. The structure from motion was suitable for estimating 3D structures from 2D image sequences captured through UAV imagery from multiple viewpoints. The coordinate points of the objects in the scene to be reconstructed were created. Here, an incremental approach for estimating the camera pose was adopted. To initiate reconstruction, the camera parameters for a set of images were first analysed. The third image was then taken and the number of tracks was observed for those camera poses estimated in the first step.

The normalized direct linear transform (DLT) under RANSAC[14] was used to find the extrinsic and intrinsic parameters of the new image. The 3D point was added if it was seen by at-least one more camera and if it helped in optimization of the estimates. This process was repeated for the rest of the images. The OpenSFM library build on top of OpenCV was used for estimation of camera posed and reconstruction of 3D scenes from multiple images. The sparse point cloud is shown in Figure 16. The proposed method was able to generate a fair number of points covering most of the area. Figure 17 shows few selected scenes and respective 3D points generated from 2D images having more than 5 overlapped images. We also observed that the usual SfM approach method was unable to generate enough points from areas with dense tree covers and was able to generate enough points for other areas. The multi-view stereo (MVS)[15] matching generated dense point clouds with relatively good accuracy (better than

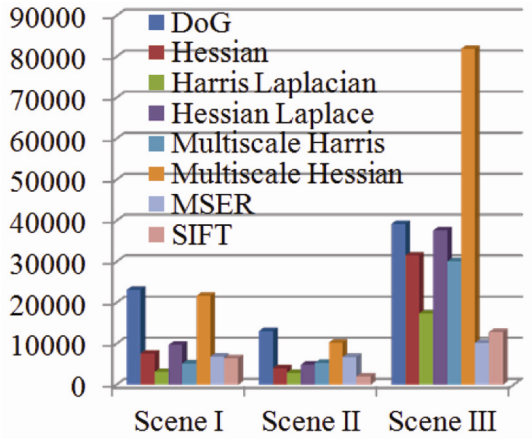

Figure 12. Number of feature keypoints detected by various detectors for different types of scenes.

Figure 13. Marking of feature descriptors of each features detected in scene II.

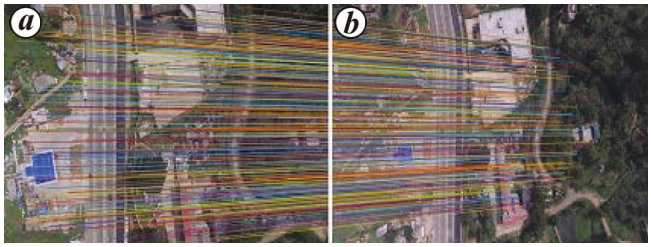

Figure 14. Matching of feature descriptors in scene II.

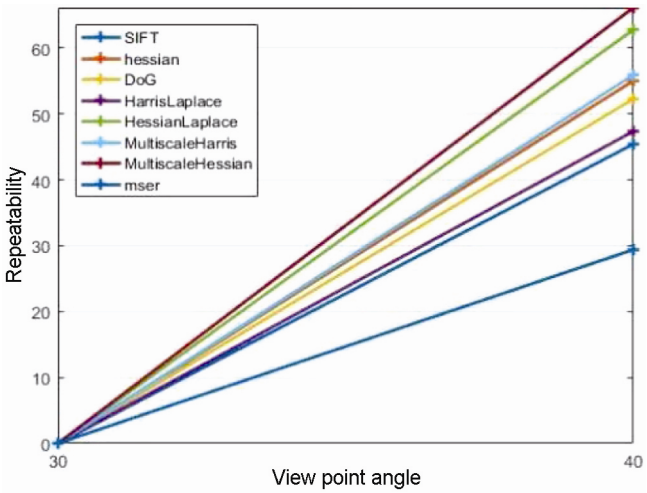

Figure 15. Matching of feature descriptors in scene III.

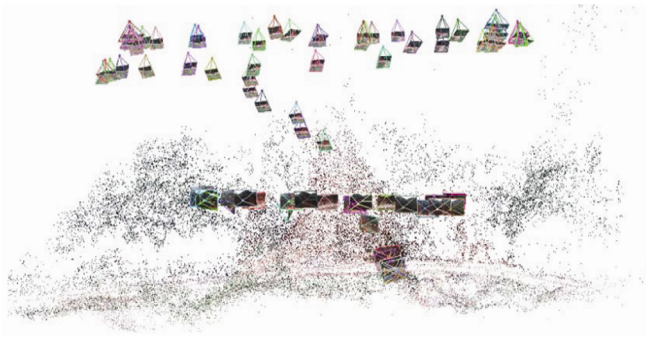

Figure 16. Sparse 3D point clouds.

1 mm for a 20 cm wide object) even from low resolution images[16]. To further accelerate point cloud generation specially for large number of images, a software having combined clustering views for MVS[17] and patch-based MVS software package was used for computing dense set of 3D points cloud from the set of 2D images. The patch was further extended to get more by segmenting the patches and expanding denser patches[18]. There were also techniques for obtaining denser patches through expansion of initial matches by leveraging nearby pixels[19,20]. The comparisons of point clouds generated by different approaches are shown in Figure 18.

Relative accuracy analysis of SfM based 3D point cloud: An attempt was made to compare the two-point cloud regenerated by SfM-based approach and Pix4D output using Cloud Compare[21]. Once the two point clouds were aligned and registered properly, the Euclidean distances were calculated with each pair of point clouds and the distance or shift of each point was observed. We observed near perfect registration of the point in the entire scene except for minor deviations in some of the areas. Figure 19 shows the point cloud differentiation map in cm. The SfM based point cloud captured almost all the important features in the study area except for the tree-clad areas where it generated less number of points when compared to Pix4D cloud.

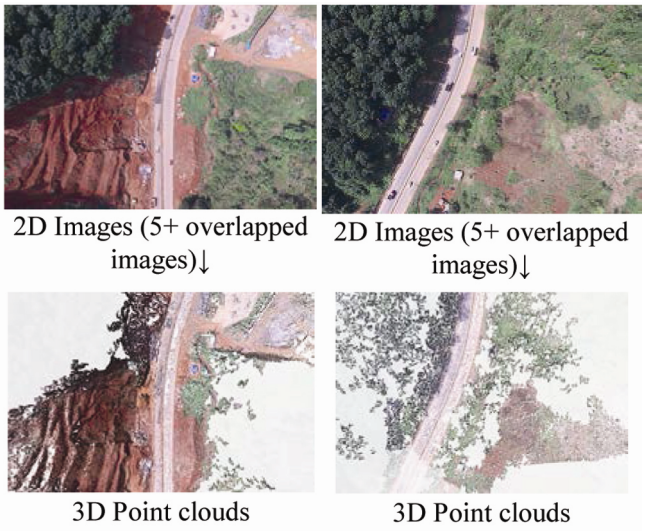

Figure 17. Point clouds generated from sequance of overlapped 2D images.

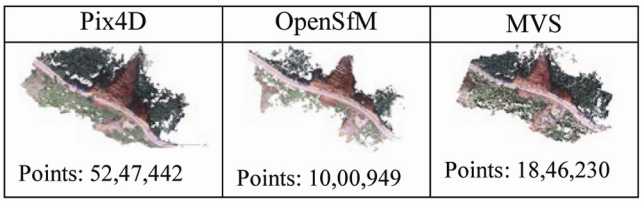

Figure 18. Comparision of generation of 3D point clouds.

The OpenSfM and MVS[22] approaches generated total vertices of 1000949 and 1846230 respectively, while the Pix4D gave 5247442 vertices covering all areas in the scene. However, all the objects with defined shape such as man-made structures were precisely captured by SfM approach. Further, the MVS approach was also able to capture the complete scene including tree-clad areas. We compared this point cloud with Pix4D cloud. We observed near perfect registration of the two-point clouds with centimetre level precision represented by blue colour except for deviations for few 3D points as shown in green and yellow point clouds (Figure 19).

Meshing and texturing

Once we obtained the dense point clouds, the surface reconstruction was carried out using Poisson surface reconstruction algorithm with greater detail[23]. Here, the points were connected together to obtain a mesh-like structure. The mesh surface reconstruction was performed through Poisson surface reconstruction algorithm by adjusting the octre-depth at 12 to get finer level of mesh structures (Figure 20). The colour from the original

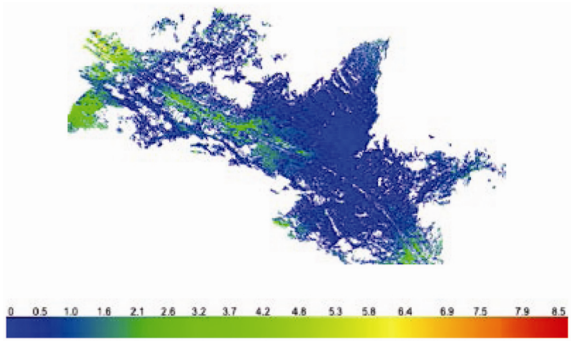

Figure 19. Comparison of point cloud accuracy.

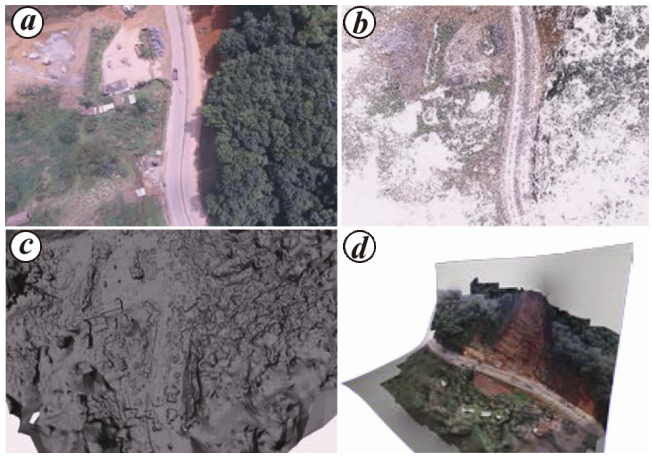

Figure 20. Images depicting different stages of converting 2D images to 3D textured models. ***a***, 2D images; ***b***, Point clouds; ***c***, Shaded mesh; ***d***, 3D textured models.

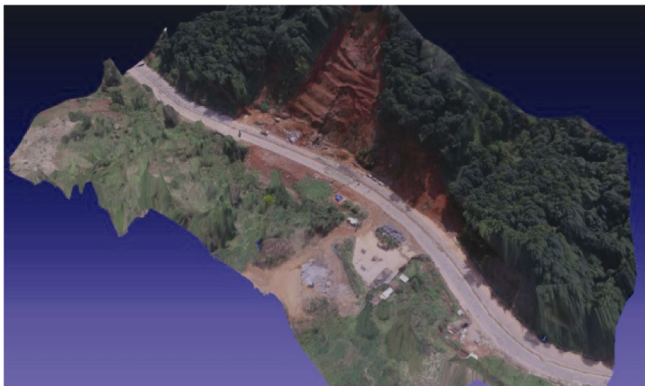

Figure 21. Full 3D textured model.

image was then incorporated into the model to get a textured output (Figure 21). The texturing of the 3D surface[24,25] mesh was very important for generating an aesthetically pleasing and realistic 3D model for visualization. The result of the texturing depends on the resolution of original input images and quality of mesh geometry. Further, the method adopted must be able to handle the texturing process even for large models. Then, texturing was done through a multi-view approach. Therefore a combined approach with SfM and MVS techniques having large number of input images and mesh geometry consisting of thousands of triangles in a reasonable time[26] was considered an effective technique for efficient texturing process.

The non-manifold edges or bad geometry were then removed. Finally, we projected the active raster colours from the original images to the mesh that we reconstructed. Meshlab was used externally for editing, cleaning and creating a mesh surface from point cloud file. The CMPMVS[27] was used to generate the 3D mesh and textured model of our study area.

Conclusions

This article has outlined the use of computer vision algorithms and SfM approach for effective processing of UAV derived imagery to generate precise and high resolution 3D surface models. It also described the detailed UAV data process workflow where open source tools are used for processing high resolution data to generate valuable and good quality data products. The high resolution 3D surface and terrain models have been increasingly used as an important input for further analysis to understand various landform processes. The proper understanding of the process deriving 3D surface models involving sophisticated analysis and detailed technical understanding of the entire process pipeline is utmost essential for precise 3D scene reconstruction from 2D aerial images. The ability to accurately detect and match the feature keypoints will result in deriving better and precise 3D surface models. To understand this, an in-depth study was done to identify suitable detector algorithms to be used based on the scene type. To better understand this, we selected two scenes with different textures and viewpoints and picked up three common detectors, viz. Harris, Fast and MinEigen and performed keypoint detections on these scenes. Out of the three chosen detectors, the MinEigen was able to detect 22,433 number of keypoints and the keypoints which were well distributed. Further, MinEigen was able to find maximum number of correct matches when compared to other two. Thus, the MinEigen detector was chosen for the scene. We also performed several experiments to understand the suitability of feature detectors and their response on three different types of scenes captured by UAV. As observed in the experiments, some detectors performed well in specific types of scenes with high repeatability when compared to others. We further studied the different approaches for generating dense 3D point clouds. The MVS method generated denser and well-distributed point clouds covering the entire region of the study area. The 3D dense point clouds generated using these feature correspondences were comparable with the Pix4D 3D point cloud with centimetre level precision.

1. Bhandari, B., Oli, U., Panta, N. and Pudasaini, U., Generation of High Resolution DSM Using UAV Images, FIG Working Week Sofia, Bulgaria, 17–21 May 2015.
2. Alshawabkeh, Y., Haala, N. and Fritsch, D., 2*d*–3*d* feature extraction and registration of real world scenes, isprs Commission V Symposium Image Engineering and Vision Metrology, IAPRS Volume XXXVI, Part 5, Dresden 25–27 September 2006.
3. Changchang, W., Towards Linear-time Incremental Structure from Motion, 2013 International Conference on 3D Vision.
4. Hassaballah, M. *et al.*, Image feature detectors and descriptors. Studies in Computational Intelligence 630, Springer International Publishing, Switzerland, 2016.
5. Mikolajczyk, K. *et al.*, A comparison of affine region detectors. *Int. J. Comput. Vision*, 2006; doi:10.1007/s11263-005-3848-x.
6. Lowe, D. G., Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.*, 2004. **60**(2), 91–110.
7. Mancini, F. *et al.*, Using unmanned aerial vehicles (UAV) for high-resolution reconstruction of topography: the structure from motion approach on coastal environments. *Remote Sensing*, 2013, **5**, 6880–6898; doi:10.3390/rs5126880.
8. Mark, A. *et al.*, Topographic structure from motion: a new development in photogrammetric measurement. *Earth Surf. Proc. Landforms*, 2013, **38**(4), 421–430.
9. Westoby, M. J. *et al.*, Structure-from motion' photogrammetry: a low-cost, effective tool forgeoscience applications. *Geomorphology*, 2012, **179**, 300–314.
10. Mike, R. James and Robson, S., Straightforward reconstruction of 3D surfaces and topography with a camera: Accuracy and geoscience application. *J. Geophys. Res.*, 2012, **117**, F03017.
11. Furukawa, Y. and Hernández. C., Multi-view stereo: a tutorial. *Found. Trends in Comput. Graph. Vision*, 2013, **9**(1–2), 1–148; doi:10.1561/0600000052.
12. Hassaballah, M. *et al.*, Image Features Detection, Description and Matching, Volume 630 of the series Studies in Computational Intelligence, pp. 11–45.
13. Mikolajczyk, K. and Schmid, C., A performance evaluation of local descriptors. *CVPR*, 2003.

14. Snavely, A., Seitz, S. and Szeliski, Building Rome in a day. International Conference on Computer Vision, 2009.
15. Furukawa, Y. *et al.*, Towards Internet-scale Multi-view Stereo, *CVPR*, 2010.
16. Furukawa, Y. and Ponce, J., Accurate, dense and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2010, **32**(8), 1362–1376.
17. OpenSFM, https://github.com/mapillary/OpenSfM.
18. Shao, Z., A multi-view dense point cloud generation algorithm based on low-altitude remote sensing images. *Remote Sensing*, *MDPI*, 2016, **8**, 381; doi:10.3390/rs8050381.
19. Habbecke, M. and Kobbelt, L., Iterative multi-view plane fitting. Proceedings of the 11th Fall Workshop Vision, Modeling, and Visualization, 2006.
20. Kushal, A. and Ponce, J., A novel approach to modeling 3D objects from stereo views and recognizing them in photographs. Proceedings of the European Conference Computer Vision, 2006, vol. 2, pp. 563–574.
21. Maiellaro, N., Zonno, M. and Lavalle, P., Laser scanner and camera-equipped uav architectural surveys, *Int. Arch. Photogramm. Remote Sensing Spatial Inf. Sci.*, 2015, **XL-5/W4**, 381–386.
22. Seitz, S. M., Curless, B., Diebel, J., Scharstein, D. and Szeliski, R., A comparison and evaluation of multi-view stereo reconstruction algorithms. *CVPR*, **1**, 2006.
23. Kazhdan, M., Bolitho, M. and Hoppe, H., Poisson Surface Reconstruction, Eurographics Symposium on Geometry Processing, 2006.
24. Waechter, M., Moehrle, N. and Goesele, M., Let there be Color! Large-scale texturing of 3D Reconstructions, European Conference on Computer Vision (ECCV 2014), Zürich, Switzerland, 6–12 September 2014.
25. TexRecon – 3D Reconstruction Texturing. http://www.gcc.tu-darmstadt.de/home/proj/texrecon/
26. Waechter, M., Moehrle, N. and Goesele, M., Let There Be Color! Large-Scale Texturing of 3D Reconstructions, Volume 8693 of the series Lecture Notes in Computer Science, Computer Vision – ECCV 2014, 2014, pp. 836–850.
27. Jancosek, M. and Pajdla, T., Multi-View Reconstruction Preserving Weakly-Supported Surfaces, CVPR 2011 – IEEE Conference on Computer Vision and Pattern Recognition, 2011.

ACKNOWLEDGEMENTS. We thank the North Eastern Space Applications Centre, Department of Space, Government of India, Umiam, Meghalaya, India, for providing the necessary UAV images required for the study. We also thank Mr P. L. N. Raju, the Director of the Centre for his support and encouragement.

Received 13 April 2017; accepted 3 August 2017

doi: 10.18520/cs/v114/i02/314-321