Digitizing the trails; the design of a low-cost wearable camera rig and development of a computer vision tree species detection model

Nicholas W. G. Trump

Mott MacDonald, Vancouver, British Columbia, Canada - nick.trump@mottmac.com

Keywords: Photogrammetry, Structure from Motion, wearable camera rig, 360° images, computer vision

Abstract

This paper presents a novel solution to a reality capture and 3D contextual mapping problem. A wearable camera rig was designed to capture photographic information of existing trails within North Vancouver, British Columbia, Canada. High-resolution video and 360° images were captured for 3D reconstruction models and the creation of an online virtual tour environment. Pathway width of the existing trails was extracted using a linear feature extraction tool to create detailed base map drawings for the client. A computer vision model was trained to detect tree species from images of tree bark. A summary of the stages of the project, lessons learnt from the use of photogrammetric (Structure from Motion) methods, suggested improvements to the computer vision model, and thoughts on the direction of reality capture methods within the infrastructure sector are given.

1. Introduction

1.1 Project background

This paper provides a case study overview of a digital solution developed for the eastern extension of the Spirit Trail located in North Vancouver, British Columbia, Canada. The client, The District of North Vancouver required assistance on collecting information on their existing trails to inform design decisions on route upgrades. The Spirit Trail is a 35km multiuse accessible greenway that crosses North Vancouver connecting communities and providing safe, comfortable, and sustainable travel. The active transportation trail provides equitable access to nature for people with varying abilities.

1.2 Objectives

To aid the design process of the extension of the Spirit Trail network, the client requested creation of base map imagery showing pathway position and tree location. The trees were to be split into different retention values and for this to be graphically indicated on the drawings. The base maps would allow the client to understand where the pathway width was insufficient (<4m) and how a potential widening might affect the trees of importance. For the client, promoting active trail whilst minimising impact and maintaining the sensitive natural environment of the forest was of upmost importance.

1.3 Available reality capture techniques

Whilst the problem might seem simple at first with pathway width being generally being relatively consistent, the greater problem was really mapping pathway orientation, direction, and location (alongside any width variations as well). Various approaches were considered and reviewed and excluded for varying reasons. Table 1 summarizes the different existing potential approaches and their disadvantages.

Technique	Reasons for exclusion
Walk and measure	Slow, inaccurate, and requires consistent bending over (increased back pain) by the site team. There would also be no way to capture path direction / orientation.
Use aerial imagery	The trails are mostly covered by the tree canopy.
Use open-source trail data	This data is inaccurate, and location / orientation is often incorrect.
LiDAR	2km of trail would be very slow and expensive. The equipment hire is expensive. Using terrestrial scanning in areas of vegetation can cause a lot of errors with scan registration.
Simultaneous Localization and Mapping (SLAM)	SLAM is very novel, the equipment is costly, and the use within areas of vegetation can cause a lot of errors with scan registration.
Drone flight (LiDAR / photogrammetry)	The equipment is expensive and requires trained personnel. No benefits over walking due to the need to be so low in elevation due to canopy cover.

Table 1. Evaluation of available solutions

As a result, it was decided to pursue a photogrammetric approach. Photogrammetry is the science of extracting dimensional information from 2D images (Bachiller & Upequi, 2022). 3D reconstruction is the approach of building 3D photogrammetric models from 2D images. Several techniques exist to do this, such as Structure from Motion (SfM) and Mutliview Stereo Reconstruction (MVS). Typically, SfM relies on multiple images from a single camera, whereas MVS relies on a stereo set of lenses (Bachiller & Upequi, 2022).

This paper focuses on the use of SfM to create 3D reconstructions of the trails to extract existing information.

With SfM relying on images the technology is passive (in contrast to an active laser system such as LiDAR), one key benefit here is therefore the low-cost nature of photogrammetric solutions. In recent years many others have turned to photogrammetric solutions for obtaining contextual information

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-2-W8-2024-439-2024

from constrained environments (Perfetti, et al., 2022) and (Liang, et al., 2014).

1.4 Research and development (R&D)

Alongside the reality capture work, the team obtained internal research funding to develop computer vision workflows. The objectives were to automatically detect and blur personal information (people / car number plates), and to identify different tree species adjacent to the trails.

2. Data capture

2.1 The wearable camera rig

The wearable camera rig that was designed includes different components that piece together to capture as much information as possible, whilst being low cost for the client and lightweight for the site team walking long distances. Figure 1 provides an overview of the camera rig.

Figure 1. Wearable camera rig

2.2 Problems

With any novel digital solution an important task is pre-planning and predicting problems that will arise with the data collection / processing. Here, the main expected issues were:

Twisting of the 3D model – Trails and other linear features pose several challenges to the use of photogrammetry. They are long and relatively narrow meaning any errors in 3D reconstruction / inaccuracy in GCP elevation could cause rotation around the longitudinal axis, i.e. twisting longitudinally. One way to arrest this issue is using GCPs at the edges of the trails to try and 'widen the base' as much as possible.

Surface mesh issues from vegetation – The trails are adjacent to vegetation, and movement of vegetation (due to wind) during data capture can create errors in mesh generation. Camera position / angle was therefore key in ensuring that most of the frames were filled with trail elements rather than the surrounding vegetation.

Similarity in trail images – Filling most of the frames with trail elements created a separate issue; similarity in every frame and an inability to discern one section of trail from the next. The reference number on the custom targets solved this by providing a chainage along a trail when encountering problems with 3D reconstruction. Additional coloured features from the survey targets helped improve this further.

Orientation / dimensions of trails – One drawback of SfM is the lack of scale and orientation provided by the 3D reconstruction (León-Vega & Rodríguez-Laitón, 2019), the approach used here to resolve this was using GCPs to both orientate / position the model in global coordinates but also provide scale for the trails. A Leica GS18 GNSS unit was used to obtain GCPs with real-time kinematic (RTK) positioning corrections to achieve centimetre-level accuracy.

2.3 Equipment costs

Table 2 provides an overview of the cost of the solution. At an approximate total of $3800 CAD this might initially seem high, but this is for the purchase cost (excluding the Leica GS18). For comparison the purchase of a terrestrial LiDAR scanner would be in the region of $80,000-100,000 and would also require substantially longer data capture and data processing times. The ongoing direct cost for each subsequent day on site is only $400, a negligible expense.

Component	Approx. cost (CAD $)	Comment
Insta 360 X4	$900	
iPhone 15 Pro	$1800	
Leica GS18	$400	Hire cost per day
Misc.	$700	Gimbal, SD card, additional batteries, attachments, etc
Total cost	**$3800**	

Table 2. Approximate costs for the wearable camera rig

The low-cost nature of SfM techniques (both direct and labour) is one of the main reasons for the rise in camera-based 3D reconstruction techniques (Kossieris, et al., 2017), (Cardaci, et al., 2019), (Granshaw, 2018), and (Santise, et al., 2017).

2.4 Data capture

Redundancy of data was a top priority, the client scope had to be met and all potential eventualities had to be planned and considered ahead of time. Two such examples are undertaking multiple walk-throughs of every trail section with a varying iPhone camera angle and walking the trails with the Insta360 capturing video and interval images. In total with these multiple walk-throughs 8km of trails were effectively mapped within a single day (4x 2km).

The camera settings used are noted in Table 3.

Insta 360 X4	72 MP 360° photos at 3 second interval 8K video at 30 FPS
iPhone 15 Pro	4K video at 30 FPS

Table 3. Camera settings

3. Data processing – 3D models

The workflow for the reality capture data is shown in Figure 2. Prior to the contract award several test runs were conducted using sample data collected. Identifying a suitable workflow from the site data to the client deliverables was critical for effective project delivery.

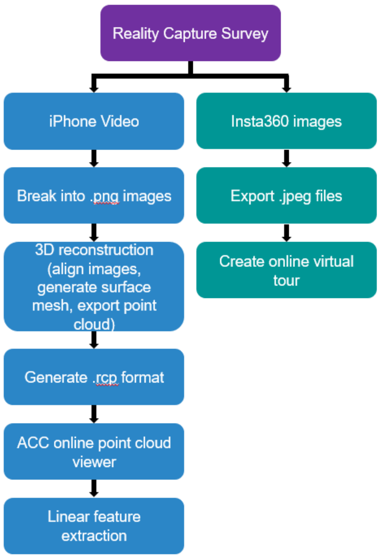

Figure 2. Reality capture data workflow

3.1 Image extraction

Still images were extracted from the video using a Python script. Overlap of images for SfM is a critical parameter. Ideal overlap varies by use-case but typically sits within the 60-85% range (Róg & Rzonca, 2021) and (Wang, et al., 2022). To attain this the team iterated the frame extraction based upon the surveyor's walking speed.

3.2 Structure from Motion

Structure from Motion (SfM) is a photogrammetric technique to reconstruct 3D models from 2D images (Elkhrachy, 1988). Using computer vision algorithms, distinct features (tie points) are identified and mapped between the series of images, camera position is then estimated from matching these features across different images. This generates a sparse 3D point cloud from the tie points. A surface mesh can then be generated and colorized / textured using RGB data from the images.

There are many software packages available for SfM, some of the popular options include Metashape, Reality Capture, 3DF Zephyr, Bentley Context Capture, Pix4D, and ReCap Photo. The available advanced features and therefore required user experience varies dramatically across the software packages. For example, ReCap Photo provides a very easy solution for SfM but lacks any real option for advanced user input and is not suited for any 3D reconstruction beyond very simplistic 3D models of small static objects.

Decimation in the context of surface meshes is the process of reducing the total number of triangles in the mesh whilst trying to preserve accurate model topography (Schroeder, et al., 1997). Decimation provides an opportunity to dramatically reduce file size whilst often having minimal effect on the quality or accuracy of the end data. Care must be given however to ensure over-decimation, where model geometry is markedly affected, does not occur. The SfM process generated a dense poly mesh for each section of the trail. These high-poly meshes generated approximately 29 million triangles for a 100m section of the trail. To reduce file sizes and improve usability of the models a mesh decimation process was completed. There are several mesh decimation techniques but one of the most common is the edge collapse method, the fundamental idea of this method is to merge two or more vertices into a single vertex. (Kobbelt, et al., 1998). Adjusting the position of the remaining vertex can mostly resolve any topology errors from the decimation.

The high-poly mesh was simplified from 29m to 100k triangles. The textures from the high-poly mesh were then reprojected back onto the low-poly mesh to recover details and provide a sharp but lightweight mesh with sufficient accuracy (Figure 3).

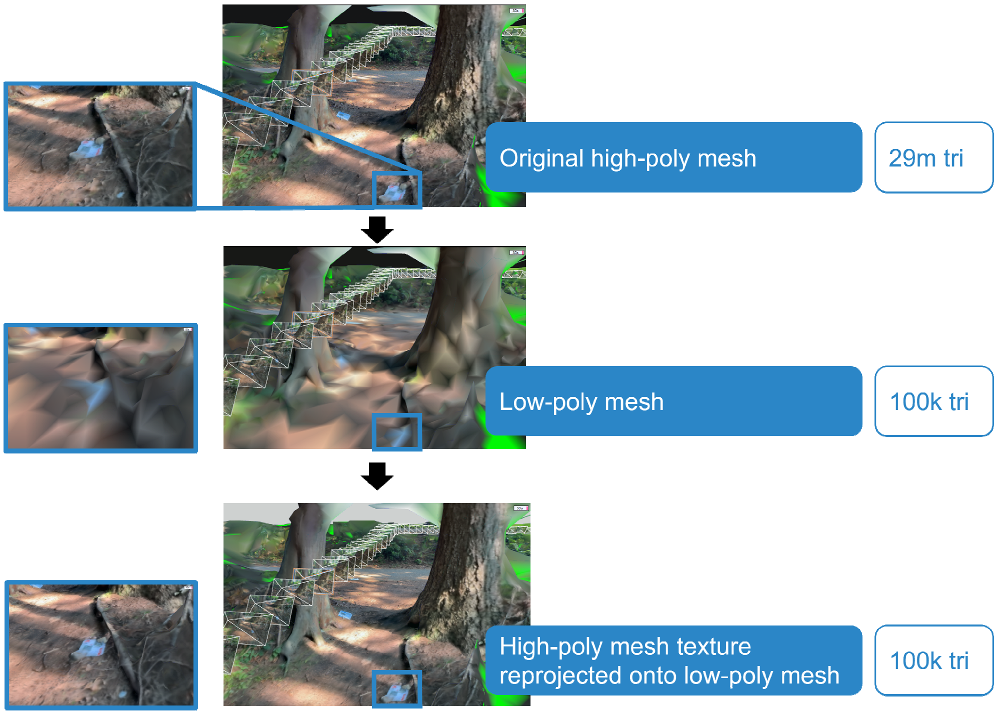

Figure 3. Visual synthesis of the SfM decimation process

3.3 Validation

A critical piece of any novel digital solution is checking and validation of the end results. A few methods were used, firstly, the point clouds were georeferenced so could be overlaid with existing digital elevation models of the trails to check for any misalignment in 3D reconstruction. Next, the GPS position of the GCPs was known and due to surveying a great number, there was a lot of redundancy and therefore opportunity for validation, Figure 4. Finally, the arborist tree data (with GPS) provided an opportunity for ensuring the geolocation of the trees adjacent to the trails was correct.

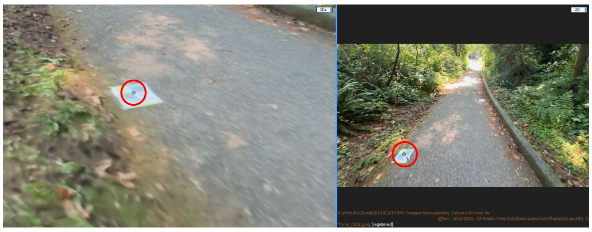

Figure 4. Validation of GCP on mesh (left) and input 2D image (right) with correct GCP location (red circle)

4. Data delivery

4.1 Point cloud models

The meshes were then exported as a dense point cloud with global coordinates in an .xyz format and imported and combined together in Autodesk ReCap. With the point clouds in an .rcp format they could be synced to Autodesk Construction Cloud (ACC). ACC was used to host and share the point clouds with the client. Accessible via an online link and with no hardware / software requirements, ACC gave the client an easy way to navigate the 3D models at their own pace and investigate areas of interest to them, rather than relying on prescribed screenshots / viewpoints from us, Figure 5. Additional measurement and annotation features were also available.

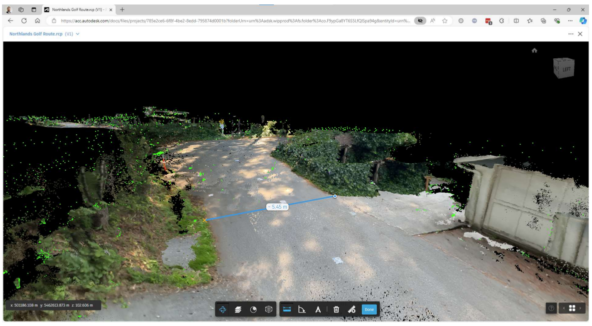

Figure 5. Point cloud model hosted on ACC

4.2 Preparation of base maps

Using the linear feature extraction tool within the ACC point cloud viewer, the existing pathway extents could be extracted. Whilst an automated process, in areas where the pathway was mostly comprised of soils / woodchips as opposed to asphalt / gravel, the feature extraction tool often needed some manual guidance to keep it on track.

Extracted to a .dwg file, the pathway extents were imported to ArcGIS Pro to combine with the other contextual data from the available open data sets.

Arborist tree inventories were imported as .shp files and trees layered by their ‘retention value’, an importance value (High to Low) based on size / age / species of tree. The tree trunk & root zones were illustrated as a series of circles and colours applied based on retention layers. An example of one of the base maps can be seen within Figure 7. Background context was provided using aerial imagery, contours (to show gradients of paths), and lot lines.

4.3 Virtual tour environment

The process of creating the virtual tour was a simple one but provided a very accessible environment. Being familiar (akin to Google Street View) and easy to navigate, meant this was a very valuable tool. Insta 360 cameras use the native .insp format, that can be easily exported as an equirectangular .jpeg (Figure 6).

Figure 6. Equirectangular 360° image

On the online virtual tour creator, the images could be linked together, maps created, an overlay interface added, and hotspots used to label key features. The end product (Figure 8) was a simple way for the client to “walk” the remote trails from the safety of their desktop.

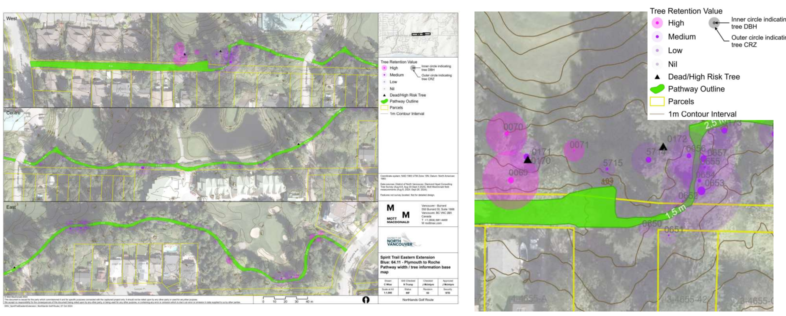

Figure 7. Example base map and excerpt

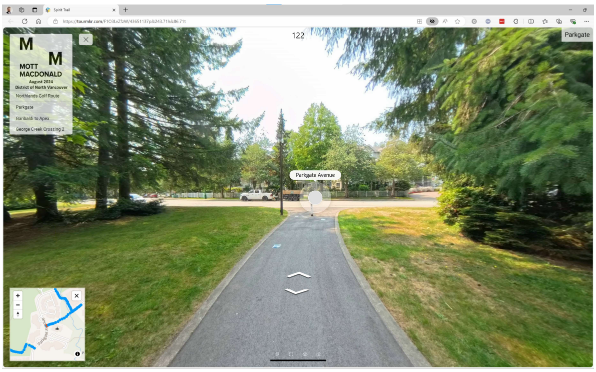

Figure 8. Virtual tour environment

5. Machine learning

5.1 Tree species detection model

Computer vision is a subset of deep learning focused on identification and extraction of information from images (Davies, 2012). Computer vision applications are very broad, for example on this project alone three applications were actually used; from the identification of features for the SfM 3D reconstruction (e.g. SIFT), to the instance segmentation algorithm for the blurring tool, to the object detection / classification completed for the tree species detection model.

A YOLO v10 model was trained with custom classes for object detection of tree species from images of the tree bark alone. Five main species of trees local to Vancouver were used in the training dataset: red alder, douglas fir, big-leaf maple, western hemlock, and western red cedar. In total 70 images per class were labelled resulting in 350 images total.

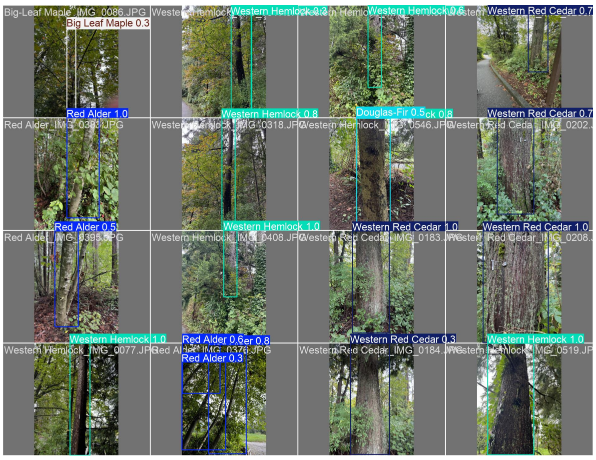

Figure 9. Tree species detection results

Figure 9 illustrates the performance of our model in identifying and classifying the tree species. The detection outcomes are displayed with bounding boxes around each detected tree bark, where the species classification and confidence score are indicated. The results show that the model correctly identified most of the trees within the input images. Confidence interval varied but generally was high suggesting strong model assurance in those classifications.

Figure 10 and Figure 11 show the F1-Confidence curve and Precision-Recall (PR) curve respectively.

F1-Confidence Curve – The F1 score is a widely used metric for classification algorithms (Hand, et al., 2021). The score illustrates the relationship between confidence levels and the F1-score for different tree species classes. Higher confidence indicates that the model is more certain of its prediction. The results indicate a strong result for a complex task like this, this curve indicates that the model maintains a reasonable balance between precision (avoiding false positives) and recall (capturing true positives).

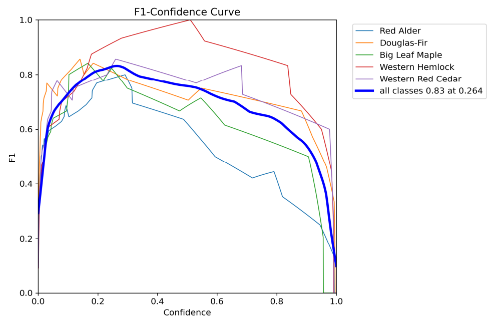

Figure 10. F1-Confidence curve

Precision-Recall curve – This curve measures the model's ability to capture true positives, ranging from 0 to 1. A higher recall indicates that the model correctly identifies a greater proportion of actual positives versus how many of the model's positive predictions are correct, a higher precision indicates fewer false positives.

The PR curve for all classes, has an overall mean average precision of 0.893 at an Intersection over Union threshold of 0.5. This value indicates strong overall performance. The scores for all species are approximately 0.80 or higher suggesting that the model can reliably identify these species with high confidence and accuracy.

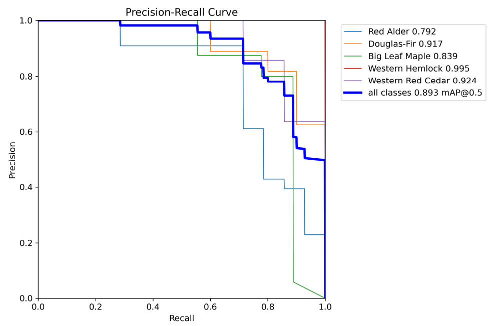

Figure 11. Precision-Recall curve

5.2 Anonymizing

A computer vision model was used to anonymise and safeguard personal data inadvertently collected during data capture. It ensures the protection of sensitive information while maintaining data integrity for analysis.

6. Reflection on approach taken

6.1 Applications in infrastructure

Beyond trail mapping, terrestrial photogrammetry has widespread applications all across infrastructure projects. Terrestrial photogrammetry provides a low-cost means of producing accurate contextual models from early-stage site visits that can be used for many downstream uses from concept design workflows, desk-based inspections, enhanced drawing production, and 3D visualization renders.

6.2 Wearable camera rig

iPhone video angle – The SfM 3D reconstruction generally provided accurate results, there were areas of the trails that the reconstruction was incorrect. This occurred due to user error with an incorrect camera angle chosen, future iterations of the rig may include a chest mount with a fixed angle.

GCP survey process – The process of using GCPs surveyed by an RTK GNSS was successful but did slow down the site survey process with needing to place and survey printed paper targets. This also would not be feasible in wet weather conditions without protected targets. There has been study by others in terrestrial GNSS-assisted photogrammetry (Morelli, et al., 2022). Additional iterations of the camera rig will investigate kinetic options for obtaining georeferencing.

Virtual tour – The virtual tour provided a lot of value as a low-cost addition to the camera rig. For example, looking at Figure 7 it can easily be seen that any widening of the 1.5m bottleneck of the existing trail would impact several large medium value trees immediately adjacent to the trail. Accessing the virtual tour (Figure 12) provides additional context to the client. When looking at this you can much more clearly see the challenges at this location than from looking at the base map alone.

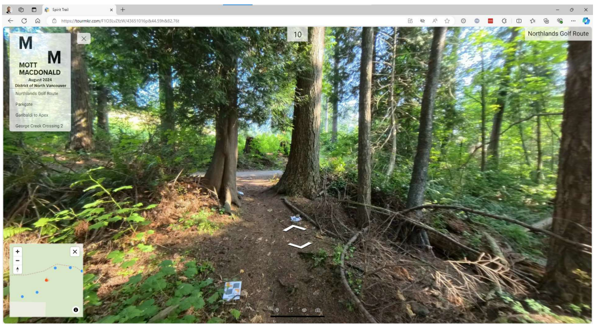

Figure 12. Virtual tour at bottleneck location

6.3 Computer vision results & next steps

Whilst the computer vision model has shown promising results the underlying machine learning model is in its early stages and needs refinement in several areas. The training data used was relatively limited due to time constraints and should be expanded with a much larger training data set of images with varying lighting / weather conditions to improve the accuracy, recall, and confidence interval of the tree species detection model.

7. Conclusion

This paper presents a case study on a novel approach to a reality capture problem through the use of a custom-designed wearable camera rig utilizing Structure from Motion 3D reconstruction and 360° virtual tours to map out existing trails within North Vancouver, British Columbia, Canada.

The custom computer vision object detection model developed is presented herein with a summary of the results showing a strong performance in its ability to detect tree species from images of tree bark alone. Recommendations for improving the accuracy and confidence interval are given.

Photogrammetric techniques are becoming more common within infrastructure projects and this project provides a case study of the ability of accurately using these techniques to provide low-cost methods of contextual mapping within constrained environments.

Acknowledgements

The author would like to acknowledge the following. Firstly, Shane Devine and Nitara Heywood from the District of North Vancouver for their support in the innovative approach. Next, Chris Lin & Sihan Bo (Mott MacDonald) for their efforts with the computer vision models, Jay Mezher (MM) for his support in preparing this paper, and finally Jamie McIntyre (MM) for his support and guidance in pursuing such a novel approach to a problem.

References

Bachiller, C. P. & Upequi, E., 2022. Low cost 3D reconstruction of cave paintings for the conservation of Colombian historical memory: case study Indigenous rock art of the sacred place “Piedras de Dunjo”. *ISPRS,* XLVIII-2/W1, 251-256.

Cardaci, A., Versaci, A. & Azzola, P., 2019. 3D low-cost acquisition for the knowledge of cultural heritage. *ISPRS,* XLII-2/W17, 93-100.

Davies, E. R., 2012. *Computer and Machine Vision.* Fourth ed. s.l.:Academic Press.

Elkhrachy, I., 1988. 3D Structure from 2D Dimensional Images Using Structure from Motion Algorithms. *Sustainability,* 14.

Granshaw, S., 2018. Structure from Motion: origins and originality. *The Photogrammetric Record,* 33(161), 6-10.

Hand, D. J., Christen, P. & Kirielle, N., 2021. F*: an interpretable transformation of the F-measure. *Machine Learning,* 110, 451-456.

Kobbelt, L., Campagna, S. & Seidel, H.-P., 1998. A General Framework for Mesh Decimation. *Proceedings of Graphics Interface,* 43-50.

Kossieris, S., Kourounioti, O., Agrafiotis, P. & Georgopoulos, A., 2017. Developing a low-cost system for 3D data acquisition. *ISPRS,* XLII-2/W8, 119-126.

León-Vega, H. A. & Rodríguez-Laitón, M. I., 2019. Fisheye lens image capture analysis for indoor 3D reconstruction. *ISPRS,* XLII-2/W17, 179-186.

Liang, X. et al., 2014. The Use of a Hand-Held Camera for Individual Tree 3D. *MDPI Remote Sensing,* 6(7), 6587-6603.

Morelli, L., Menna, F., Vitti, A. & Remondino, F., 2022. Action cams and low-cost multi-frequency antennas. *ISPRS,* XLVIII-2/W1, 171-176.

Perfetti, L., Teruggi, S., Achille, C. & Fassi, F., 2022. Rapid and low-cost photogrammetric survey of hazardous sites, from measurements to VR dissemination. *ISPRS,* XLVIII-2/W1, 207-214.

Róg, M. & Rzonca, A., 2021. The Impact of Photo Overlap, the Number of Control Points and the Method of Camera Calibration on the Accuracy of 3D Model Reconstruction. *Geomatics and Environmental Engineering,* 15, 67-87.

Santise, M. et al., 2017. Preliminary tests of a new low-cost photogrammetric system. *ISPRS,* XLII-2/W8, 229-236.

Schroeder, W. J., Zarge, J. A. & Lorensen, W. E., 1997. Decimation of Triangle Meshes. *ACM SIGGRAPH Computer Graphics,* 26(2), 65-70.

Wang, F., Zou, Y., Castillo, E. d. R. & Lim, J. B. P., 2022. Optimal UAV Image Overlap for Photogrammetric. *IOP Conf. Series: Earth and Environmental Science,* 1101, 1-10.