Leveraging cloud compute and open source software to generate 3D models from drone photography

Gregory Biegel

Amazon Web Services, 240 St. Georges Terrace, Perth, Western Australia 6000, Australia - gbiegel@amazon.com

Keywords: Cloud Computing, Photogrammetry, Digital Twin, Open Source Software

Abstract

Modern drone platforms provide the opportunity to capture large quantities of high quality, geotagged aerial imagery quickly and over large distances, at a relatively low cost. As organisations increasingly turn to digital twin technology to manage their operations, rich, photorealistic three dimensional (3D) models of their assets become increasingly important. This paper describes how cloud computing, and container based serverless compute in particular, can be used together with open source software to generate a 3D model, converted to 3D tiles for optimised transmission, from images of large outdoor assets captured by drone. This approach offers a flexible and cost effective way to processing drone imagery into 3D models whilst allowing you to maintain ownership and control of your data.

1. Introduction

The continued proliferation and advances in image capture with low-cost, consumer-grade, unmanned aerial vehicles (drones) and smart phone technology has created a range of opportunities for their application in fields such as agriculture, surveying, and remote structure inspection. High-resolution imagery captured by these aircraft and handsets can be processed using photogrammetry to create photorealistic three dimensional (3D) models that closely reflect reality and can be used to create immersive digital twins. One of the main challenges in creating 3D models from two dimensional (2D) images is the requirement for a large amount of computational power and storage for a relatively short period of time for image processing.

OpenDroneMap (ODM) is a popular open source software package that processes 2D images into 3D models, digital surface models (DSM), and orthophotos using photogrammetry (Toffanin, 2019). Despite its name, ODM can process any suitable images into a 3D model, although georeferenced aerial images obtained from a drone, combined with ground control points, are ideal.

Cloud computing offers on demand, virtually unlimited processing power and secure, flexible data storage and management without the need to invest in expensive and specialised hardware. Leveraging cloud computing allows users to only pay for any resources used, and the time they were used for, which is ideal for a process such as photogrammetry that requires large amounts of compute and storage temporarily whilst generating a 3D model.

This paper describes a data processing pipeline that leverages ODM running within the cloud to automate the processing of photogrammetry reconstruction tasks, allowing you to create a 3D model and convert it to a 3D tileset simply by uploading your images to cloud storage. Crucially, you will only pay for the storage and processing power used whilst the process is running. The generated tileset can then be used within digital twin applications to build immersive, photorealistic environments.

2. Background

Organisations are increasingly adopting aerial image capture using drones alongside terrestrial laser scanning and other reality capture technologies to gather data about large outdoor assets such as industrial plants and infrastructure including buildings, roads, and bridges. The benefits of this form of data collection include its speed, lower cost, flexibility, and relative safety allowing data to be captured significantly more frequently than using other methods. Whereas in the past aerial photography was a costly and infrequent option only available to a few, drones have almost commoditised aerial photography. In order to be useful, drone data needs to be stored, managed and processed in a way that enables insights to be drawn from it. Digital twins provide an effective way in which to present this data, combining a multitude of sensor and other data sources with recent imagery.

2.1 Digital twins

A digital twin is a virtual representation of a physical object or system that is designed to accurately reflect the physical entity it models. Modern digital twins commonly include rich three dimensional visualisations of the physical world onto which is layered additional data to provide an immersive experience that connects users with contextualised data. Until fairly recently, to understand a physical object, one would need to be in close proximity to the object and any investigation and analysis would need to occur on the object itself or a physical replica (Grieves and Vickers, 2017). A digital twin offers a high fidelity, virtual version of physical objects that can be used from the other side of the world by users to understand, model, and perform simulations on.

The scale at which digital twins are being deployed is increasing around the world. The Western Australian government recently announced plans to create a state-wide digital twin for Western Australia which is approximately 2.5 million square kilometers (the second largest subdivision of any country) (WA Government, 2024). Given their scale, city, state, or global scale digital twins benefit from the use of 3D models created from aerial photography using photogrammetry, to provide accurate

Product	License	Type
PIX4Dmapper	Commercial	Installable
PIX4Dcloud	Commercial	Managed service
Agisoft Metashape	Commercial	Installable
Agisoft Cloud	Commercial	Managed service
ArcGIS Drone2Map	Commercial	Installable
AliceVision Meshroom	Open Source	Installable
OpenDroneMap	Open Source	Installable

Table 1. Popular photogrammetry tools for generating 3D models from 2D images

representations of large geographical areas. This can be seen in the the photorealistic 3D tiles feature recently released on the Google Maps Platform that provides subscribers with a 3D mesh textured with high definition imagery for many parts of the globe (Google, 2024).

As we see with Google's photorealistic 3D tiles at a planetary scale, with larger 3D visualisations there is a need for optimised transmission of data. Cesium introduced 3D Tiles in 2015 as an open specification for streaming massive heterogeneous 3D geospatial datasets and defines a spatial data structure and a set of tile formats designed for 3D and optimised for streaming and rendering (Patrick Cozzi, 2024). In 2019 3D Tiles was adopted as a community standard by the Open Geospatial Consortium (OGC) (Patrick Cozzi and Sean Lilley, 2023). Google's photorealistic 3D tiles are based on the OGC 3D Tile standard.

2.2 Processing images by photogrammetry

There are a number of options available to build 3D visualisations with photogrammetry for use in a digital twin, based on both the processing software used (commercially licensed vs. open source), the compute that you use to run the software (on premises vs. in the cloud), and whether you choose to use a fully managed service that provides compute, storage, and processing, often based on a subscription model. Table 1 shows some of the popular tools available today and whether they can be installed (locally or in the cloud), or whether they are a managed service.

The chosen approach must be balanced across economic considerations (the cost of acquiring, running, managing your own compute, storage, and processing software at sufficient scale) as well as considerations around data privacy, residency, and ownership within managed services.

Effective photogrammetry based on structure-from-motion (SFM) relies on certain characteristics of source images to create an effective 3D model, with possibly the most important being a high degree of overlap between images, as well as the horizon not being visible in images and the inclusion of both nadir and non-nadir images in the data set. Whilst a skilled drone pilot or photographer is able to manually capture images with such characteristics, more consistent results can be achieved via automated capture tools. For the purposes of this paper we will assume that a set of images suitable for processing via photogrammetry has already been captured.

2.3 Open source software for photogrammetry

ODM is a highly configurable structure-from-motion (SFM) photogrammetry tool that uses feature matching from multiple overlapping offset images to simultaneously solve camera pose and scene geometry (Westoby et al., 2012). ODM may be run on your own infrastructure or within the cloud, and also offers a subscription based managed service to process images.

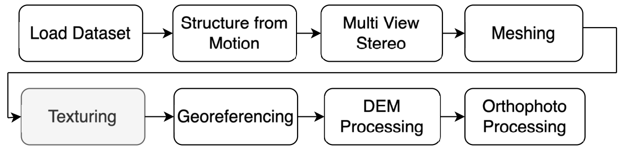

Figure 1. The ODM processing pipeline. (Toffanin, 2019) ©Piero Toffanin, reproduced with permission

ODM follows a series of steps internally when processing a dataset starting with structure from motion and ending with orthophoto processing as shown in Figure 1. It is the output of the texturing step (highlighted in Figure 1) that we use in our processing. ODM provides the **NodeODM** component which is a standard REST API interface to the ODM processing pipeline that can be accessed over HTTP. In our pipeline we deploy NodeODM and call the API from our orchestration function to manage the processing of a dataset.

The pipeline described in this work creates a 3D tileset from images, taking the textured polygonal mesh output from the ODM pipeline, being a Wavepoint OBJ file and associated texture files, and passing these to Cesium Ion (Cesium GS, 2024) via the Cesium Ion REST API where they are converted to a 3D tileset. Cesium Ion provides a community plan for personal use, and a range of paid plans for commercial use (Cesium GS, 2024). The ODM pipeline will also produce an orthophoto and digital elevation model (DEM) output.

2.4 Serverless computing in the cloud

Serverless computing is a relatively new form of compute offered by major public cloud providers whereby all server infrastructure is managed by the provider with customers bringing only application code to be executed. The application is often provided in the form of a container image which is a portable, lightweight collection of everything (code, dependencies, and runtime) needed to run it. ODM provides NodeODM as an image available via the publically accessible Docker Hub registry that can be provided to a container to run.

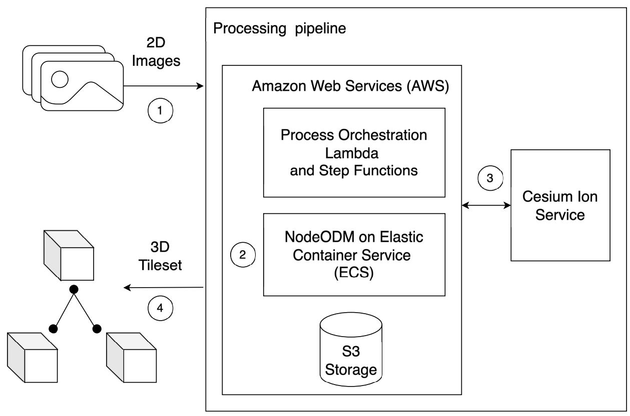

Figure 2. Cloud based processing pipeline to create a 3D tileset from 2D photographs

CPU Value	Memory	Operating System
2048 (2 vCPU)	4 - 16GB	Linux, Windows
4096 (4 vCPU)	8 - 30GB	Linux, Windows
8192 (8 vCPU)	16 - 60GB	Linux
16384 (16 vCPU)	32 - 120GB	Linux

Table 2. Available Fargate task sizes (from 2 vCPU upwards (Amazon Web Services, 2024a))

3. Processing pipeline

The processing pipeline leverages NodeODM to provide an interface to process georeferenced images into a textured mesh in a completely serverless approach. The pipeline uses Amazon S3 (Amazon Web Services, 2024b) to store image data and Amazon Elastic Container Service (ECS) (Amazon Web Services, 2024a) provides a container to run the NodeODM image to process images into a 3D model. The Cesium Ion service (Cesium GS, 2024) is then used to create a tileset from the textured mesh generated by ODM. AWS Lambda (Amazon Web Services, 2024e) and AWS Step Functions (Amazon Web Services, 2024f) are used to orchestrate the entire process. All components of the pipeline are defined as Infrastructure as Code (IaC) using AWS CloudFormation (Amazon Web Services, 2024c) making the pipeline simple to deploy, configure, and manage. The pipeline, shown in Figure 2, takes the following steps to process a set of images

1. A zip file containing JPG images is uploaded to a landing S3 bucket
2. The upload initiates a new compute task in ECS running NodeODM and creates a processing job within the executing task. The processing task creates a textured model in Wavefront OBJ format. When processing completes, the compute task is stopped so that no more cost is incurred
3. The textured model is uploaded to Cesium Ion and the resulting tileset is saved in an S3 bucket
4. The tileset is downloaded and can be used in a digital twin application and have data overlaid within it

Our processing pipeline uses ECS to run a Docker container image of the NodeODM software that is downloaded from Docker Hub. ECS offers two launch options to run the NodeODM image within the cloud and a further option to incorporate your own infrastructure into a cluster.

3.1 ECS Fargate launch type

AWS Fargate (Amazon Web Services, 2024d) is a serverless way to run the workload where you do not need to manage any of the underlying infrastructure, simply specifying the desired CPU and memory for your compute task. Table 2 illustrates task size options for Fargate (from 2 vCPU upwards being considered the minimum for processing by ODM). A Fargate task can be dynamically started when there is data available to process, and stopped once processing is complete. Using Fargate to run ODM is considered the option with the least complexity but consequently does not offer much control or configurability for processing larger, more complex datasets. Our pipeline currently uses Fargate to provide a high-level, serverless compute abstraction and has proven to be adequate in the tests we have run. More complex processing jobs can be completed using the same pipeline, and simply replacing the compute task with either EC2 or specialised on-premises hardware.

3.2 ECS EC2 launch type

If you require more control over the CPU, GPU, and memory available for photogrammetry processing (as may be the case for large or complex data sets), you can manage your own Elastic Compute Cloud (EC2) servers within the Elastic Container Service. Managing your own servers also provides you with access to instances that have NVIDIA GPUs such as **G4dn** instances. Studies have shown that adding a GPU to the ODM processing system can offer improvement of between 15% and 45% (Gbagir et al., 2023). Whilst managing your own EC2 servers provides more flexibility than Fargate, there is the additional complexity of having to manage the servers (for example ensuring that they are patched appropriately).

3.3 ECS Anywhere

ECS provides a third option that allows you to bring existing on-premises infrastructure into an ECS cluster. This option provides flexibility to allow you to incorporate existing infrastructure investment into the same processing pipeline, giving you the ability to consume cloud compute (via Fargate or EC2 launch types) together with your own hardware within the same ECS cluster in the processing pipeline. Following this option you are not only responsible for maintaining the server software (e.g operating system), but also managing the physical device itself (e.g. power, cooling).

3.4 Orchestration

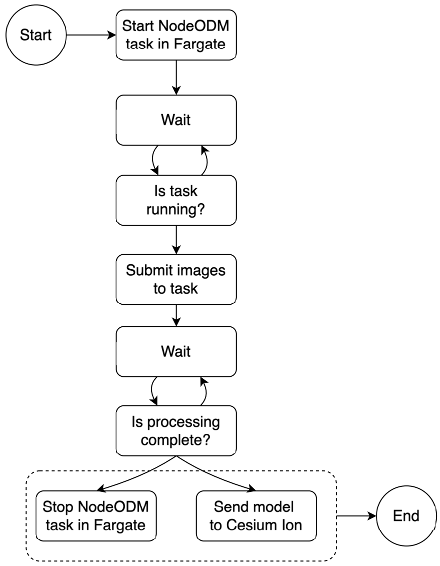

Figure 3. Step Function state machine to orchestrate 3D model processing and tileset creation

Due to the serverless nature of the processing pipeline and the approach to minimise compute costs by only running a compute task when there is data to be processed, we use the workflow capabilities provided by a step function to orchestrate the tasks involved in the processing pipeline. This includes starting the compute task, periodically checking to see if it has completed, transferring the data to Cesium Ion for tiling, and stopping the

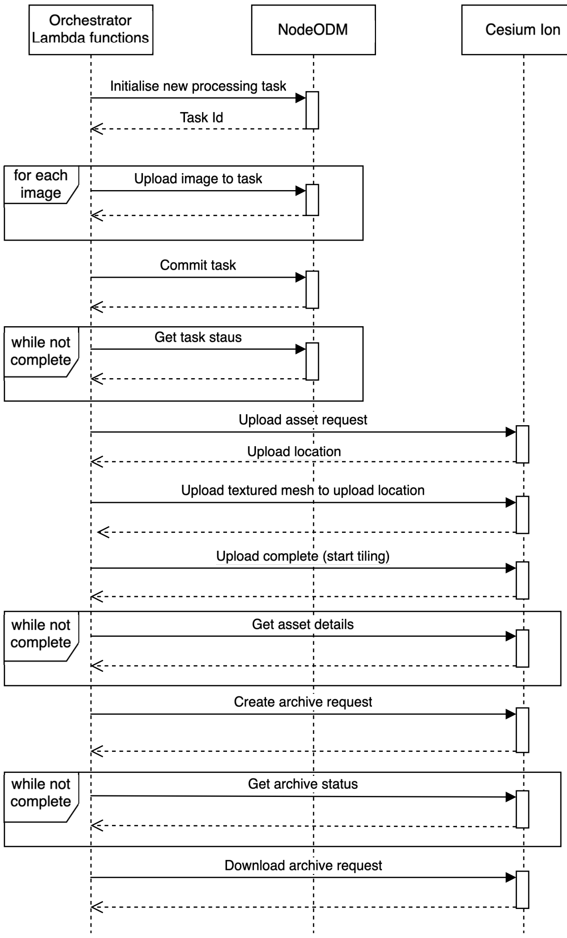

Figure 4. Interaction diagram showing orchestration of calls to NodeODM and Cesium Ion

compute task. The orchestration step function used in the processing pipeline is shown in Figure 3.

The step function uses lambda functions to invoke both the NodeODM REST API (running within ECS) to create a 3D model, and the Cesium Ion REST API (in the Cesium managed service) to create the 3D tileset. The sequence of calls to these APIs is illustrated in Figure 4.

4. Results

An example dataset of 149 images of an abandoned power station in Fremantle, Western Australia was captured with a DJI Phantom 4 consumer drone. We utilised the flight planning capabilities of DroneDeploy to plan the capture, which took around 9 minutes to cover the 6 acres of ground as shown in Figure 5.

When the flight was complete, the captured images were transferred from the drone and fed through the processing pipeline, resulting in the textured model shown in Figure 6. The 3D tileset generated by the subsequent processing step in Cesium Ion (and spatially located manually post processing), is shown rendered in a simple web application in Figure 7.

Figure 8 shows how the 3D scene rendered in the web application can be enhanced with geospatially located annotations, demonstrating how you can begin to build a rich, immersive digital twin that combines data and visualisations.

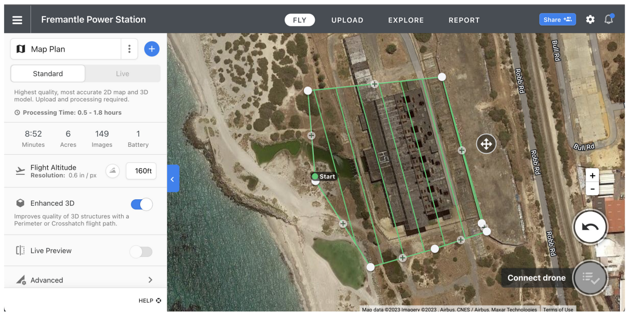

Figure 5. Flight plan within DroneDeploy for image capture

Figure 6. Output from OpenDroneMap processing of aerial images of Fremantle power station

We tested processing of the image set with several combinations of vCPU and memory configuration within Fargate which are the only configuration options available. For the purposes of the test, we only measure the time taken by ODM running on Fargate to generate the 3D model, since this is the majority of time taken in the complete pipeline, as well as the only part of the process that can be configured with vCPU and memory parameters.

4.1 Test with 2 vCPU, 16GB RAM

ODM recommends a minimum of 16GB of RAM be used for image processing, and in the first test 16GB of RAM was paired with 2 vCPUs in Fargate. As expected, this configuration resulted in the slowest processing the dataset, taking 149 minutes. The CPU and memory consumption of the processing task is shown in Figure 9 and as expected, both CPU usage and memory consumption are the highest for the longest time of the three tests.

4.2 Test with 8 vCPU, 60GB RAM

The next test paired 8 vCPUs with 60GB of RAM in Fargate, and as expected the processing completed significantly quicker, in 65 minutes. The CPU and memory consumption of the processing task is shown in Figure 10 and as can be seen CPU utilisation still spiked in this test, but for less time than the test with 2 vCPUs. Maximum memory consumption was also lower in this test.

4.3 Test with 16 vCPU, 120GB RAM

The final test configured the Fargate task with the maximum vCPU and memory resources possible and as expected, completed in the shortest time of the three tests, at 49 minutes.

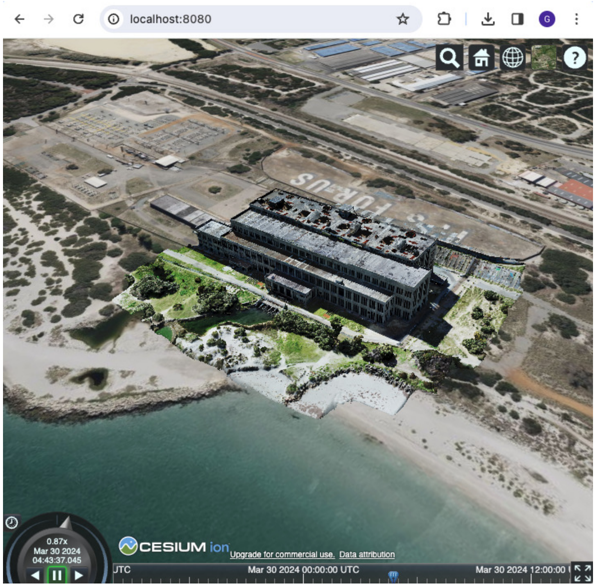

Figure 7. 3D tile output of processing pipeline running in a web application

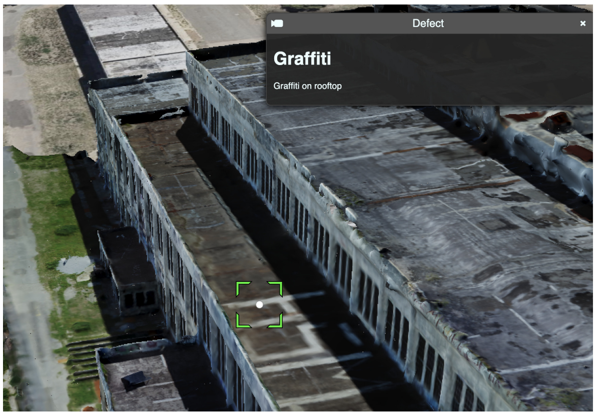

Figure 8. Markup annotation on the 3D scene

Again as expected, the CPU and memory consumption as illustrated in Figure 11 shows shorter spikes in CPU utilisation and lower memory consumption.

4.4 Indicative compute cost

Our processing pipeline consists of a number of components including storage, compute, and orchestration components (step functions and lambda functions). The majority of the cost of running the pipeline lies in the compute function within Fargate. Fargate is priced on an hourly basis per vCPU and GB memory, dependent on architecture, operating system, and region. Every Fargate task has 20GB of ephemeral storage included at no extra cost. Our experiments were run on x86/Linux in the northern Virginia (us-east-1) AWS region with the pricing for this mix illustrated in Table 3. Fargate tasks are charged from the time the container image is obtained, until the task terminates, rounded up to the nearest second, with a minimum charge of 1 minute (Amazon Web Services, 2024d). Table 4 illustrates the cost of processing the sample data set with Fargate for the 3 experimental configurations, and shows that the dataset can be processed for less than 1USD in all cases.

Figure 9. Fargate task configured with 2 vCPU and 16GB memory

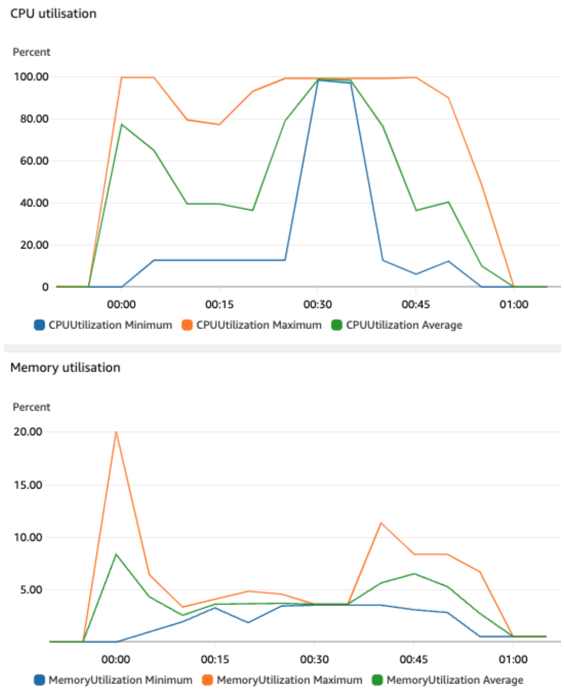

Figure 10. Fargate task configured with 8 vCPU and 60GB memory

4.5 Future work

The pipeline as presented provides a good starting point that can be further improved to process larger and more complex data sets efficiently.

4.5.1 Compute task One of the major benefits of the processing pipeline as presented, is the ability to easily replace

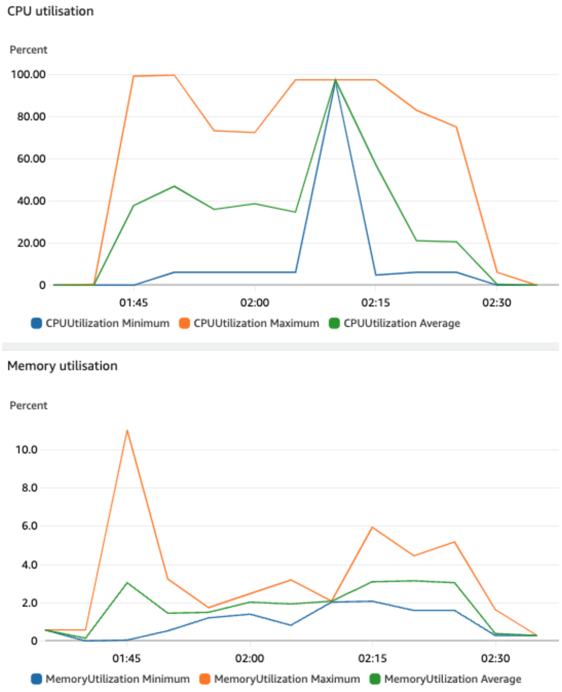

Figure 11. Fargate task configured with 16 vCPU and 120GB memory

Dimension	Price (USD)
per vCPU hour	$0.04048
per GB hour	$0.004445

Table 3. Fargate pricing for x86/Linux in us-east-1

the type of compute that executes the ODM image processing. Whilst Fargate provides the convenience of serverless, managed compute, it’s utility is limited for larger, more complex processing jobs that require GPU acceleration. ECS allows Fargate to be switched out for an EC2 based task, providing far more granular control over CPU, GPU, storage, and memory.

4.5.2 Automatic georeferencing The pipeline does not currently pass georeference data to Cesium Ion, even though ODM does perform georeferencing during processing. In the current pipeline, the generated tileset can be manually georeferenced via the Cesium Ion web application, but this is not precise.

5. Conclusions

ODM is a powerful and configurable software package that can produce high quality 3D models from 2D images. The benefits of executing ODM serverlessly within the cloud includes only paying for resources that you use and providing the ability to have fine grained control over the number and type of resources that you apply to the processing task, including CPU, GPU, and memory. We have created a pipeline that can seamlessly process drone imagery into a 3D tileset and shown how the output can form the basis of a visualisation within a digital twin. The pipeline is cost-effective, with the ability to dynamically trade processing time off against cost if required and to utilise existing on-premises hardware for processing if required.

vCPU	Memory	Elapsed (mins)	Cost (USD)
2	16	149	$0.38
8	60	65	$0.64
16	120	49	$0.97

Table 4. Pricing of three experimental processing jobs in Fargate based on configuration and elapsed time

6. Disclaimer

The views and opinions expressed in this publication are those of the author alone and do not relate to his employment with Amazon Web Services.

References

Amazon Web Services, 2024a. Amazon Elastic Container Service. https://aws.amazon.com/ecs/ (26 March 2024).

Amazon Web Services, 2024b. Amazon S3. https://aws.amazon.com/s3/ (26 March 2024).

Amazon Web Services, 2024c. AWS CloudFormation. https://aws.amazon.com/cloudformation/ (26 March 2024).

Amazon Web Services, 2024d. AWS Fargate. https://aws.amazon.com/fargate/ (26 March 2024).

Amazon Web Services, 2024e. AWS Lambda. https://aws.amazon.com/lambda/ (26 March 2024).

Amazon Web Services, 2024f. AWS Step Functions. https://aws.amazon.com/step-functions/ (26 March 2024).

Cesium GS, 2024. Cesium Ion. https://cesium.com/platform/cesium-ion/ (26 March 2024).

Gbagir, A.-M. G., Ek, K., Colpaert, A., 2023. OpenDroneMap: multi-platform performance analysis. *Geographies*, 3(3), 446–458.

Google, 2024. Photorealistic 3D Tiles. https://mapsplatform.google.com/maps-products/map-tiles/ (26 March 2024).

Grieves, M., Vickers, J., 2017. *Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems*. Springer International Publishing, Cham, 85–113.

Patrick Cozzi, 2024. The open specification for 3D data. https://cesium.com/why-cesium/3d-tiles/ (26 March 2024).

Patrick Cozzi and Sean Lilley, 2023. 3D Tiles Specification. https://docs.ogc.org/cs/22-025r4/22-025r4.html (26 March 2024).

Toffanin, P., 2019. *OpenDroneMap: The Missing Guide*. first edn, UAV4GEO.

WA Government, 2024. Digital Twin technology to elevate State’s data capabilities. https://www.wa.gov.au/government/media-statements/Cook-Labor-Government/Digital-Twin-technology-to-elevate-State’s-data-capabilities–20240301 (26 March 2024).

Westoby, M., Brasington, J., Glasser, N., Hambrey, M., Reynolds, J., 2012. ‘Structure-from-Motion’ photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology*, 179, 300-314. https://doi.org/10.1016/j.geomorph.2012.08.021.