Preimage

Technological Advances Paving the Way for New-Age Drone Photogrammetry

Introduction

Photogrammetry is the technique of using a series of overlapping images to generate 3D digital representations of real-world environment, which can then be used for measurements, inspections, or advanced simulations. These days, such images are often captured using drones, DSLRs and smartphones. By using enough overlapping aerial images or terrestrial scans, specialized photogrammetry software can be used to generate photorealistic 3D representations of real-world environments.

Widely used industry-standard software generate these 3D models from images using algorithms (like Multi-view Stereopsis, Poisson surface reconstruction etc.) These algorithms are compute-heavy and can take multiple hours of processing even for moderate sized datasets, and often fail to generalize in presence of noise. In commercial settings, these challenges necessitate the need for cumbersome manual post-processing and higher oversight, thereby increasing project costs and delaying decision-making by users or businesses that use these outputs.

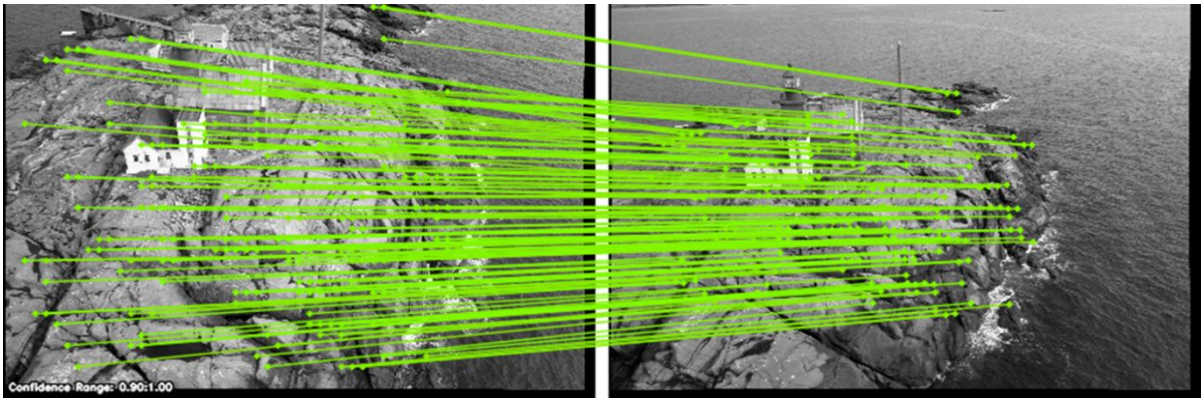
The recent advances in technology especially the improvements in AI algorithms (3D deep-learning), distributed cloud computing, embedded hardware, and fiber/5G connectivity can significantly improve the current process of creation of 3D models.

Preimage is one of the several companies trying to modernize the photogrammetry industry by combining the best technologies to alleviate the inefficiencies in the standard photogrammetry pipeline. We believe, a better, faster, and economical solution will unlock opportunities and spark innovations in industries beyond drone-based mapping like gaming, AR/VR, and 3D content creation.

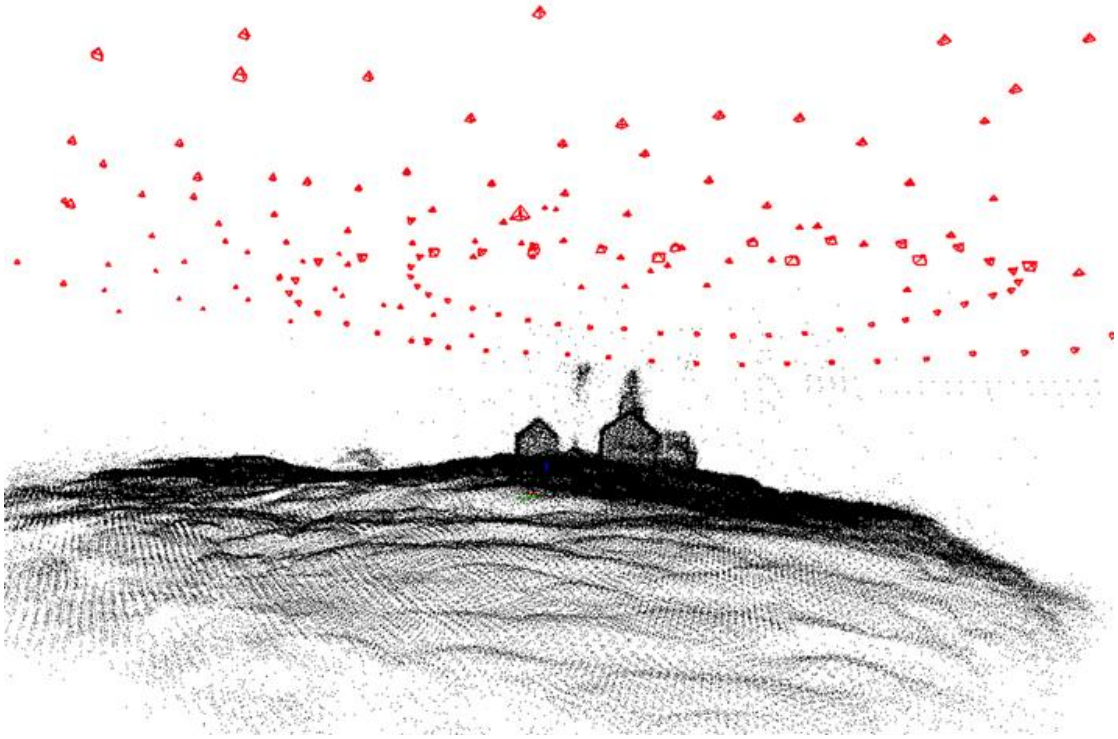
Classical Photogrammetry Practices

Photogrammetry as a field developed hand in hand with advances in geometric computer vision and multi view geometry. The classical pipeline usually involves extracting & matching key features in pairs of images, robustly estimating relative orientations between the pairs, triangulating an initial point cloud, and then solving a large bundle adjustment problem where the camera transformations, intrinsic parameters (focal lengths, lens distortions etc.) and the 3D point cloud are jointly optimized.

Feature matches between two views of the same scene



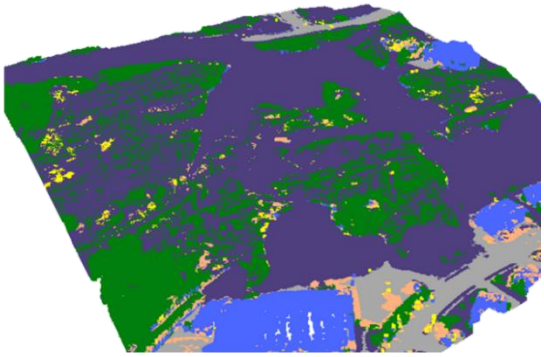
Output from Bundle Adjustment showing camera positions and 3D sparse cloud



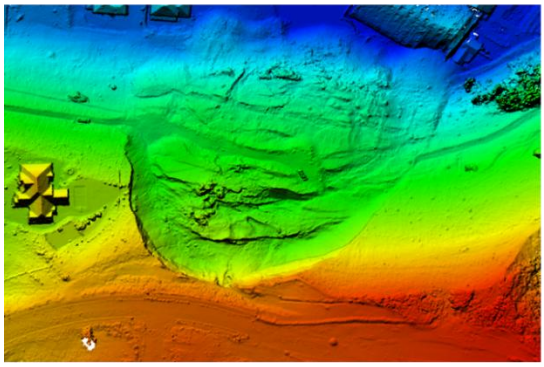
Once the camera transformations are established, a dense reconstruction of the scene can be produced using a technique known as Multi-View Stereo (MVS). This dense reconstruction is used as is, classified into ground, vegetation, building, cars etc., or is converted to other outputs.

Sample Outputs from Drone Photogrammetry Software

3D Mesh



Classified Point Cloud



Digital Elevation Model

Orthomosaic

For example, orthorectification & stitching of input images is performed in order to create a high-resolution orthophoto/orthomosaic which is a "top-down" view of a scene, useful for measuring distances. Screened Poisson Surface Reconstruction (SPSR) is typically performed on the MVS output to generate a watertight surface in the form of a mesh, which can then be imported into 3D CAD (Computer Aided Design) software like AutoCAD, game engines like Unreal Engine, or BIM (Building information modeling) software like Revit.

What's good about the existing software?

There is no dearth of photogrammetry software that do all of the above, both open-source (Meshroom, COLMAP, OpenMVG etc.) and commercial (Pix4D, Metashape, ContextCapture, RealityCapture etc.). All these software packages accomplish the eventual goal of reconstructing scenes using multiple camera images with survey-

grade quantitative accuracy, which can then be used to measure distances, areas, volumes.

Sparse model of central Rome using 21K photos produced by COLMAP's SfM pipeline.

Dense models of several landmarks produced by COLMAP's MVS pipeline.

Source: https://colmap.github.io

Most of these tools have been developed over a long period (>5 years) and hence are stable, support multiple file formats both for inputs & outputs, and have a helpful community of users. Moreover, most of these software applications are not limited to just aerial images and thus also work with images taken from DSLRs and smartphones.

There are also several specialized offerings for agricultural health monitoring and for specific kinds of sensors like multispectral cameras, some of which come bundled with many software packages.

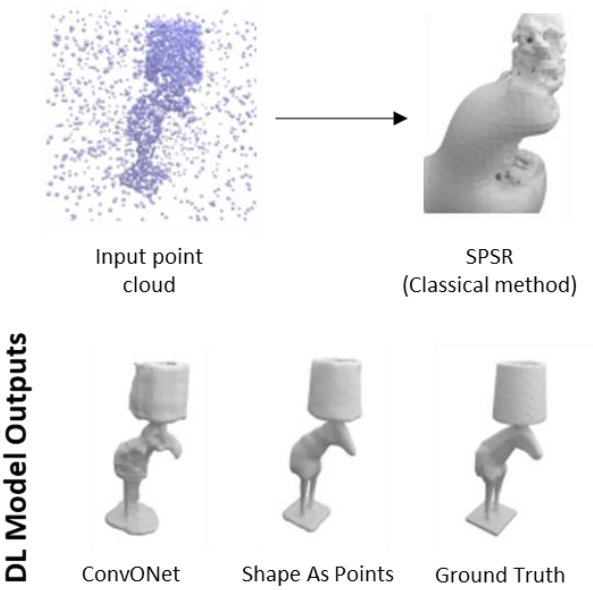
What needs to improve?

Even though the quantitative/measurement accuracy of reconstructions produced by most contemporary software is usually satisfactory, there are aspects of the pipeline that end up being pain points for many users. For example, the classification of the generated point clouds is often quite inaccurate, which in-turn also affects the quality of the interpolated terrain models (DTMs) produced.

Classical computer vision algorithms for point cloud segmentation usually employ a whole bunch of feature engineering, and thus are sensitive to outlier points and might not generalize to scenes they haven't been fine-tuned for.

There are also issues with the quality of the meshes produced by SPSR and/or derived variants, with the meshes sometimes containing jagged surfaces and inconsistent textures. Sometimes the reconstructed point cloud itself suffers from surface noise and artefacts. In commercial scenarios, these inaccuracies often necessitate manual post-processing of the outputs before they can be used. This is time-consuming and laborious.

Figure showing sensitivity of SPSR to noise



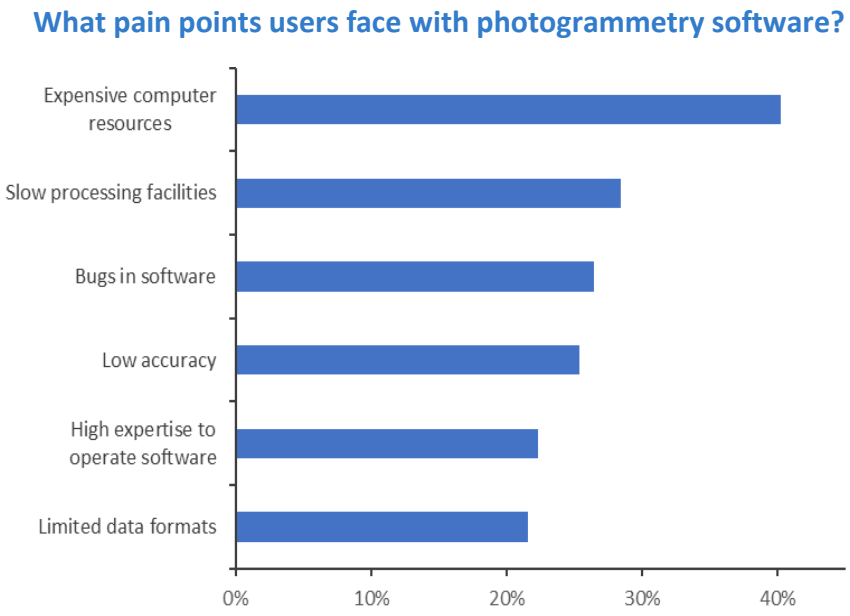
Source: Shape as Points: A Differentiable Poisson Solver

Photogrammetry is a high-compute process, and thus a common complaint about almost all the software is the processing speeds they can achieve. Much of classical 3D computer vision relies on heuristics to speed up the brute-force computations that are needed to run an end-to-end 3D reconstruction. But in case of large, interconnected scenes, the speed-up is seldom enough, and it's not uncommon for photogrammetry pipelines to take more than a day to process even moderate-sized scenes (~5000 images, 3-5 square kms).

Another thing of note is the computation hardware required for running the photogrammetry pipeline. Algorithms like MVS, SPSR, and Orthorectification are all compute-heavy, and sometimes may require copious amounts of RAM, CPU, and GPU processing power to run. Thus, it is very common for these processing pipelines to be run on machines with expensive 32+ core CPUs, workstation GPUs, and often 128+ gigabytes of RAM.

Even while running on machines with such specifications, processing speeds are slow, sometimes prohibitively so. The main reason being, the software itself is

designed to run on various machine configurations, and thus not optimized for specific hardware. There are also instances where the processing fails after hours, requiring the end-users to run it all over again.



Source: GIM International

A hidden problem with any on-premise deployment is the inertia associated with scaling. The high-end hardware setup recommended for running such software makes it prohibitively expensive for users to process multiple datasets simultaneously. This again increases the friction associated with undertaking projects that are time-critical, for example analyzing damage caused by a hurricane or measuring ore stockpile volume at a quarry, both of which need quick results.

There are several services that offer managed cloud-based offerings that claim to solve just this problem. For example, one can rent cloud instances from providers like GeoCloud that are custom engineered to run photogrammetry software without having to purchase them. But these are partner deployments that also run software from other vendors and hence are not tailored for specific software.

Also, some software like Pix4D and Bentley ContextCapture have cloud offerings, but these also suffer from the same bottlenecks as their on-premise counterparts since they aren't optimized for the cloud.

Technological Advances that can Reshape Photogrammetry

Although technological innovations are difficult to predict, there have been a few shifts in recent years that have a great potential to change the way photogrammetry pipelines are developed. We believe, improvements in AI, cloud computing, and hardware together will become pivotal in driving this change.

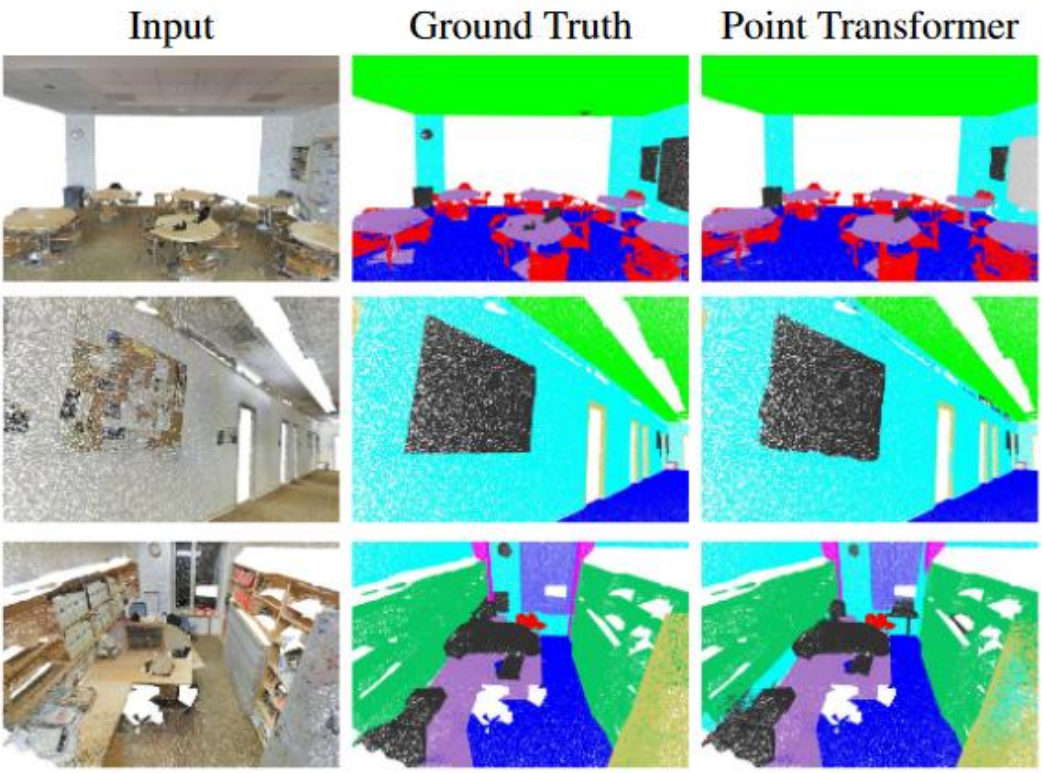
Recent advances in 3D computer vision

Recent works show that a plethora of 3D computer vision problems, from the likes of multi-view reconstruction to camera/object pose estimation and localization are better tackled with the use of deep CNNs (convolutional neural networks). Traditional methods rely on either heuristic filters or work by enforcing handcrafted photo consistency metrics which would need to be applied across all interconnected images requiring hours and even days of processing.

Deep learning methods sidestep these problems by learning filters that are specific to the task and scene at hand. Deep learning approaches provide better accuracy and generalizability at a fraction of the compute overhead when compared with traditional approaches.

Obtaining a dense set of matching feature descriptors between two images is a cornerstone of a variety of 2D and 3D computer vision tasks, including Structure from Motion (SfM), Simultaneous Localization and Mapping (SLAM) and Visual localization. With the use of self and cross attention layers in Vision transformers, it is now possible to obtain a dense set of matches conditioned on the two images being matched even on low texture areas, incidentally, an area where traditional feature detectors struggle to obtain matches.

Semantic segmentation on S3DIS dataset



Source: Point Transformer (2020)

Self-supervised vision transformers were found to contain a greater amount of explicit information about the semantic segmentation of a scene than their CNN counterparts, giving way to semi-supervised training strategies that pretrain on large quantities of data, with no supervision, and can then be fine-tuned on the target domain data. Such models have exhibited far greater semantic understanding of scenes and objects on an area than any traditional approach.

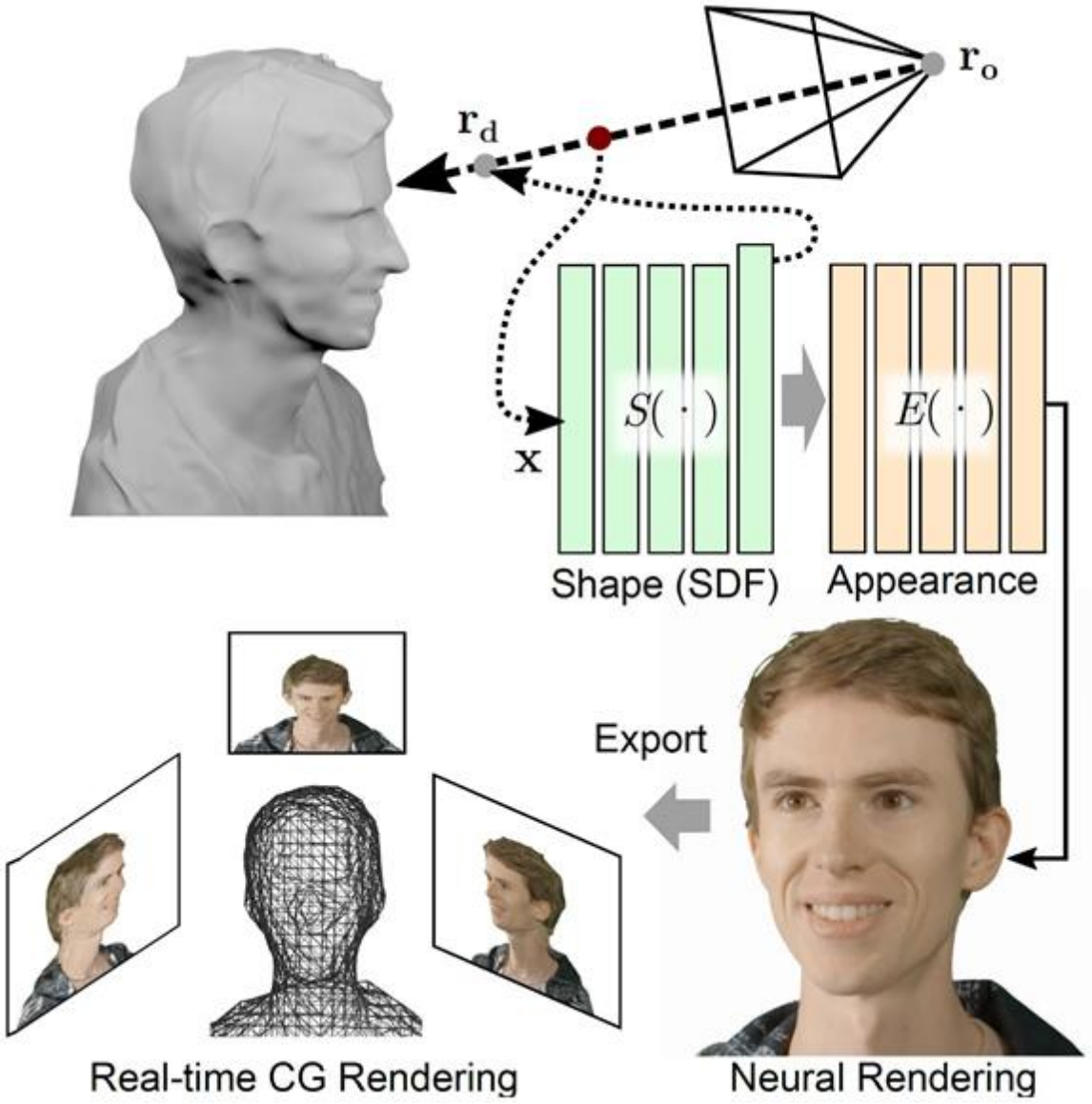
Earlier 3D deep learning approaches, as a consequence of using 3D CNNs, relied on voxel grids as an ordered representation of three-dimensional data. Voxel grids are an unnecessarily voluminous data structure whose memory footprint makes them unwieldy to use. But, by employing novel network architectures like PointNet, neural networks can now handle immense amounts of 3D data in the form of unordered point clouds at a much lower memory footprint.

Hierarchical feature extraction from 3D data using PointNets resulted in the models achieving better classification and segmentation of scenes. Neural scene representations strive to encode a 3D scene in a neural network by learning an implicit surface function, which can then be queried to retrieve the surface at

potentially infinite resolution at a fraction of the memory overhead compared to traditional approaches.

A new paradigm in this field is the advent of differentiable rendering. 3D Scenes represented as neural networks can also be rendered as images in a 'differentiable' fashion, which essentially means that 3D gradients of objects can propagate through the rendering, allowing networks to optimize 3D scenes while working on 2D representations. Works like NeRF (Neural Radiance Fields) and its derivatives produce such differentiable, photorealistic renders of 3D scenes from new view directions given some input images taken from other positions.

Overview of a pipeline that optimizes 3D geometry & appearance using 2D neural renders



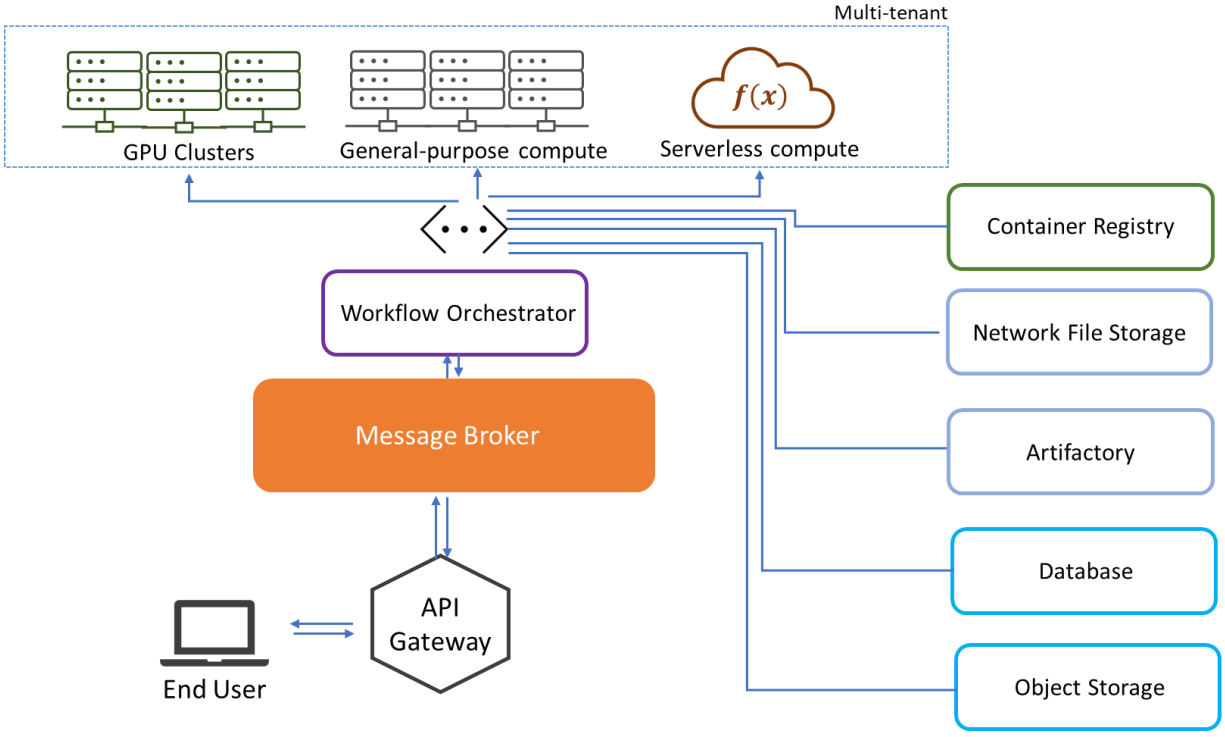
Source: Neural Lumigraph Rendering (2021)

The number of institutions (both universities and companies) involved in 3D deep learning research has grown significantly and along with it the amount of freely available 3D datasets that can be used for training (e.g. GL3D, MegaDepth). Moreover, availability of free 3D graphic design software and frameworks like Unreal Engine, Unity, and Blender have made generating task-specific 3D data easier than ever, with the visual quality of their rendering engines getting better with each release. Increase in easily available 3D sensor data and pipelines for customized data generation allows for greater fine-tuning of deep learning architectures, making way for more generalizable and robust solutions to 3D vision problems.

Democratization of high-performance distributed computing

With the advent of cloud computing platforms like AWS, high-performance compute services are now available to everyone with internet access. Any AI application needs tons of such computing resources both for training the models & deploying them. This need for compute power only intensifies when one attempts to build 3D computer vision AI applications, as the data increases in size and complexity, taking the form of huge point clouds, meshes etc. with billions of points. Distributed, specialized clusters in the cloud can handle smaller chunks of such data and effectively reduce processing time.

Schematic of an example HPC architecture



With deep learning frameworks like Pytorch Lightning, training strategies utilizing distributed data and model parallelism can be envisioned on top of such clusters. For even faster inference times, customized hardware can be built to accelerate the neural nets being deployed by harnessing programmable FPGA instances like Amazon's EC2 F1 or ASICs like Google's Tensor Processing Units (TPUs).

Modern CPUs and GPUs have great support for parallel computing, with specialized libraries like TBB, CUDA, Triton, OpenMP etc. abstracting the SIMD function calls that are natively supported by these devices. Such libraries provide great speedup for repetitive tasks in a few lines of code and are indispensable for HPC applications.

An added advantage of running HPC in the cloud is prior knowledge of the hardware specifications of machines running the applications, allowing the aforementioned libraries to be optimized even further for the specific hardware configuration. For example, the number of Streaming Multiprocessors (SMs) in an NVIDIA GPU can be optimized for in a CUDA kernel call.

Improvement in hardware and embedded systems

Among all the rapid developments on the software front, hardware systems for data collection and processing on-board the drone have steadily improved over years. Photogrammetry processes are set to duly benefit from quality sensor outputs, AI on powerful on-board computers and super-fast communication modules.

On the sensors front, the general availability of reliable, low-cost, and lightweight LiDAR sensors can be game-changing. LiDAR sensors offer high accuracy and can penetrate vegetation canopies, yielding reconstructions that would not have been visible to the naked eye. More and more drones are using global shutter cameras, that capture the image in a synchronous fashion, a property that aids better photogrammetry reconstructions. Multispectral & hyperspectral cameras offer great insights on crop and vegetation health and are also useful for mineral discovery.

Low-power computers have become quite powerful over recent years. For example, the NVIDIA Jetson series of boards are powerful enough (up to 200TOPS) to run full autonomy stacks on-board with less than 40W of power consumed. Even more promising is the advent of custom-designed FPGAs for specific operations, a great example being the Navion chip designed by MIT researchers in 2018, which can run real-time stereo SLAM at under 50mW of power consumption.

AGX Orin outperforms its predecessors by a wide margin

	AGX Orin	AGX Xavier	Jetson Nano
CPU	12x Cortex-A78AE @ 2.0GHz	8x Carmel @ 2.26GHz	4x Cortex-A57 @ 1.43GHz
GPU	Ampere, 2048 Cores @ 1000MHz	Volta, 512 Cores @ 1377MHz	Maxwell, 128 Cores @ 920MHz
Accelerators	2x NVDLA v2.0	2x NVDLA	N/A
Memory	32GB LPDDR5, 256-bit bus (204 GB/sec)	16GB LPDDR4X, 256-bit bus (137 GB/sec)	4GB LPDDR4, 64-bit bus (25.6 GB/sec)
Storage	64GB eMMC 5.1	32GB eMMC	16GB eMMC
AI Perf. (INT8)	200 TOPS	32 TOPS	N/A
Dimensions	100mm x 87mm	100mm x 87mm	45mm x 70mm
TDP	15W-50W	30W	10W
Price	?	$899	$99

Source: AnandTech

While not being recent technologies, RTK (real-time kinematic) and PPK (post-processed kinematic) drones are being utilized more and more. Both methods correct the location of drone mapping data and eliminate the need for GCPs (ground control points), bringing absolute accuracy down to cm range. With 5G technology becoming mainstream, real-time streaming of high-resolution data to servers is a real possibility, with companies like Qualcomm integrating 5G into their SoCs, like Flight RB5.

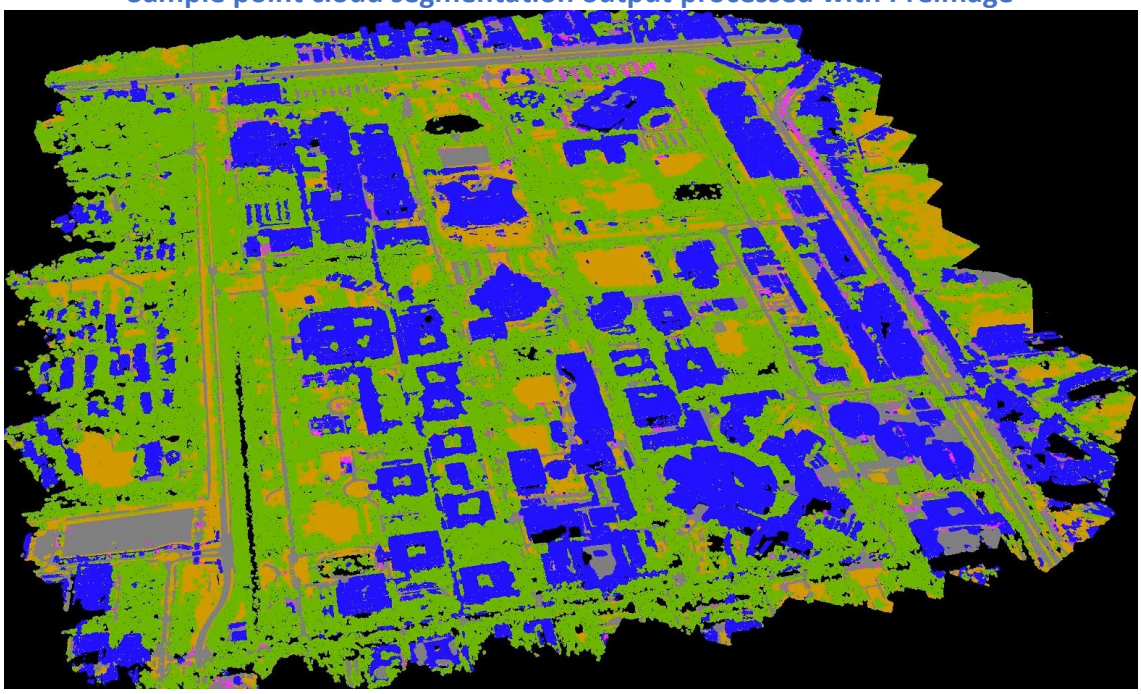
What is Preimage Building?

As evident from earlier sections, there are several problems with the rapidly growing photogrammetry software industry, mitigating which requires a cumulative effort from several individuals/organizations, to ensure that the offerings are ready for the incoming needs of the end-users in the future. Preimage is one of the several companies innovating in this domain, with the eventual goal of steering the industry towards a cloud-native, AI-driven, and highly scalable future.

Preimage leverages the best parts of both classical geometric CV techniques and the on-going AI paradigm shift. Using the latest advances in the field of computer vision and deep learning, Preimage's proprietary photogrammetry engine aims to deliver excellent qualitative outputs while not compromising on the quantitative accuracy of the outputs, in a fraction of the time required by most software.

Deep learning algorithms have shadowed classical algorithms for a lot of tasks including but not limited to multi-view stereo, feature matching & semantic segmentation. There are portions of the photogrammetry pipeline where classical methods still generalize better than deep learning ones, and Preimage solves the speed bottlenecks (if any) associated with them through its high-performance distributed computing infrastructure, including GPU native programming.

Sample point cloud segmentation output processed with Preimage



Dataset source: Sensefly

Statistics on the above segmentation inference from the test set

Class	Precision	Recall	F-1 Score
Ground	77%	69%	72%
Road	73%	82%	77%
Vegetation	83%	92%	87%
Structure	92%	89%	91%
Vehicles	57%	49%	53%
Wt. Avg.	**82%**	**84%**	**83%**

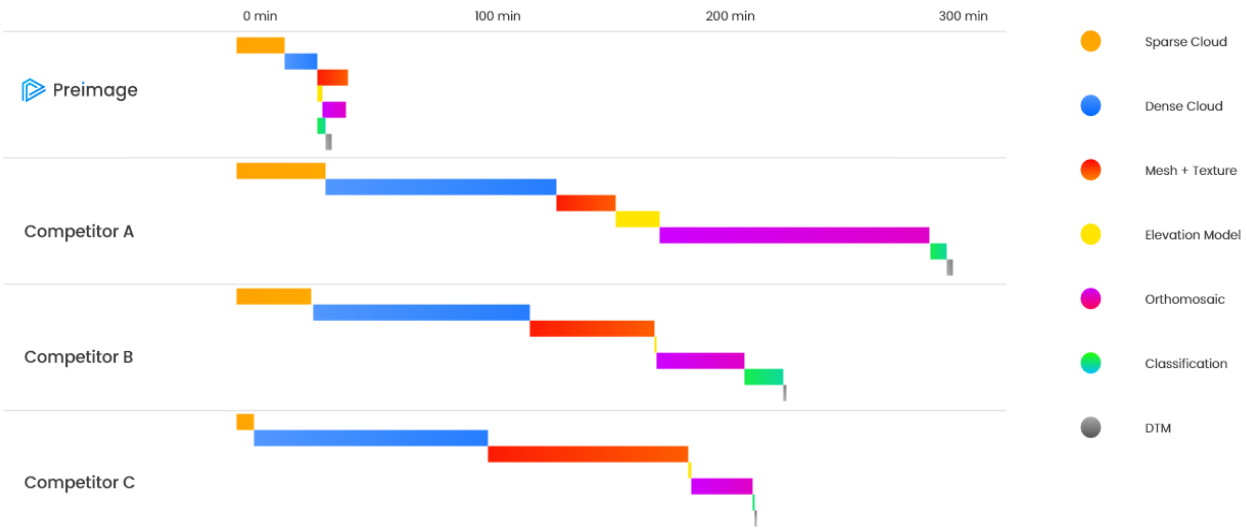
Class	Precision	Recall	F-1 Score
Ground	89%	81%	85%
Non-Ground	90%	94%	92%
Wt. Avg.	**90%**	**89%**	**90%**

Preimage's cloud infrastructure is designed to utilize the compute resources in the most efficient fashion. Being on the cloud allows an application developer to have foresight over the hardware resources that would be available to the application during run-time, and thus they can optimize for the same.

Different parts of the processing pipeline run on specific kinds of optimized instances by AWS Elastic Cloud Compute (EC2), thus the memory-heavy tasks can benefit from the R-series, while the I/O heavy tasks run best on the I-series instances. All of these services seamlessly communicate with each other, with efficient failure recovery mechanisms in place to actively monitor processing of each project.

Preimage also relies heavily on CUDA-based GPU programming, which if optimized properly and for the correct tasks, can yield orders of magnitude worth of speed-ups over CPU programming. GPUs also offer a great speedup when it comes to deep learning inference, and several libraries like TensorRT (by Nvidia) and accelerate (by Huggingface) are also extensively used to speed up the DL inference tasks. In test-settings, Preimage has performed 4-5x faster than competition when processing drone image datasets end-to-end.

Processing time for 1075 aerial images: Preimage vs Competition



All of this ensures a smooth, managed, and fast photogrammetry experience for the end-user, who neither has to purchase perpetual licenses for software, expensive machines to run them, nor constantly monitor the system resources during processing.

Conclusion

The benefits of generating 3D models from real-world is driving the adoption of photogrammetry in industries beyond drone-based surveying such as gaming, AR/VR, and e-commerce. This decade is increasingly going to be shaped with people consuming content much more in 3D than 2D. However, the current photogrammetry pipelines won't be able to handle the scale of data processing nor the aspirations of the end-users who will be creating or consuming such content. We need a rethink in terms of how photogrammetry software is designed, and the recent technological advances offer us a way to do just that.