

Photogrammetry and Volumetric Capture

By Timothy Porter, Underminer Studios
Graphics by Alex Porter



Current techniques of communication with video, augmented reality (AR), and virtual reality (VR) fall short in information exchange, and often require additional resources to help convey the message. This all costs more time and money, ultimately reducing platform effectiveness, so we need a better solution.

Volumetric capture and photogrammetry use images from cameras and sensors to create 3D meshes, which can be merged seamlessly into game engines, VR headsets, AR environments, and merged-reality (MR) worlds for a deep psychological influence

on users. Anywhere you would traditionally use a computer-generated (CG) asset is a great place to use a volumetric object for increased immersion. This article compares volumetric capture and photogrammetry, and takes a deeper dive into technical specifications, package sizes, capture options, computing needs, and cost analysis. It also looks at the benefits and complexity of each style and its use cases, as well as the engagement and retention in creating immersive realism for digital formats, including VR, AR, and MR.

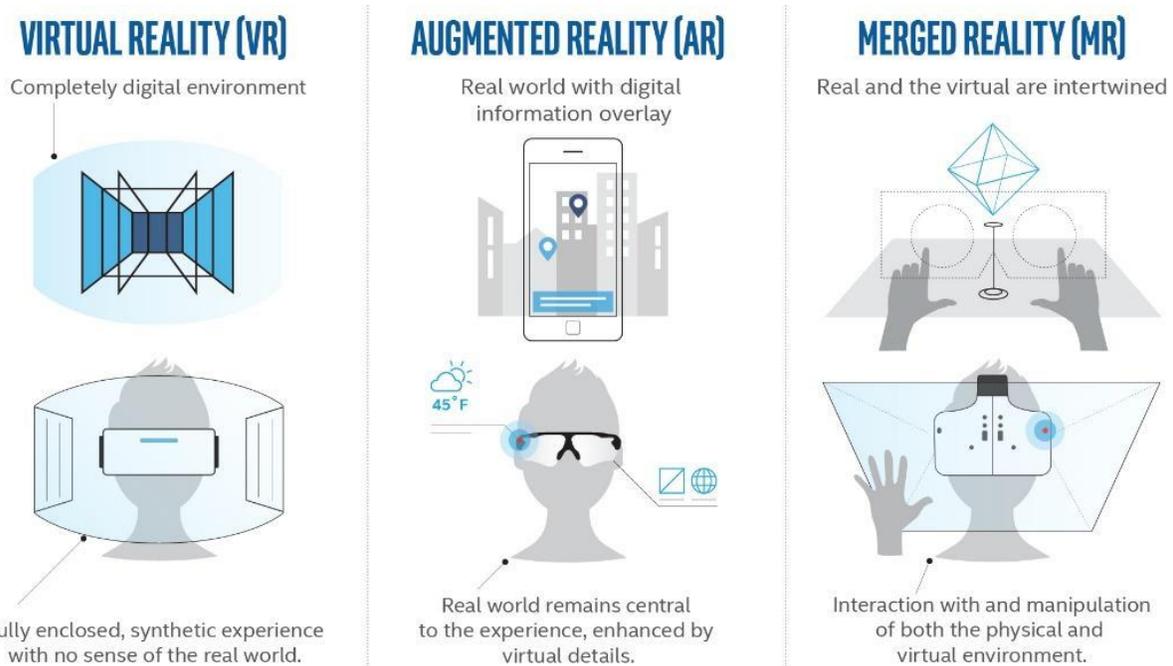


Figure 1: Virtual reality, augmented reality, and merged reality explained.

Photogrammetry and Volumetric Capture Techniques

Photogrammetry is defined as the use of photography in “surveying and mapping to ascertain measurements between objects.” It dates back to the mid-19th century, when researchers discovered that a minimum of two images could be used to identify *lines of sight* from the individual camera points to objects in the photographs, and thus extrapolate 3D data.

Computational photogrammetry has been utilized for decades and, as one would expect, has vastly increased the possible applications of the technique. Volumetric capture is a more recent innovation, which, as [techcrunch.com](https://www.techcrunch.com) explains, records footage of a real person from various viewpoints, “after which software analyzes, compresses, and recreates all the viewpoints of a fully volumetric 3D human.”

Photogrammetry and volumetric capture are quite different technologically, but they share a large amount of software overlap when it comes to image processing. Taking multiple images, the system detects points of interest within every image in a series. It then goes back through the images and matches points of interest on one image to other points of interest that occur on different images. When those points match up, it creates a point in 3D space that also contains color information. This is repeated for thousands—if not millions—of points, refining the mesh as it goes along, creating a sparse point cloud. This is by far the most important part of the process.

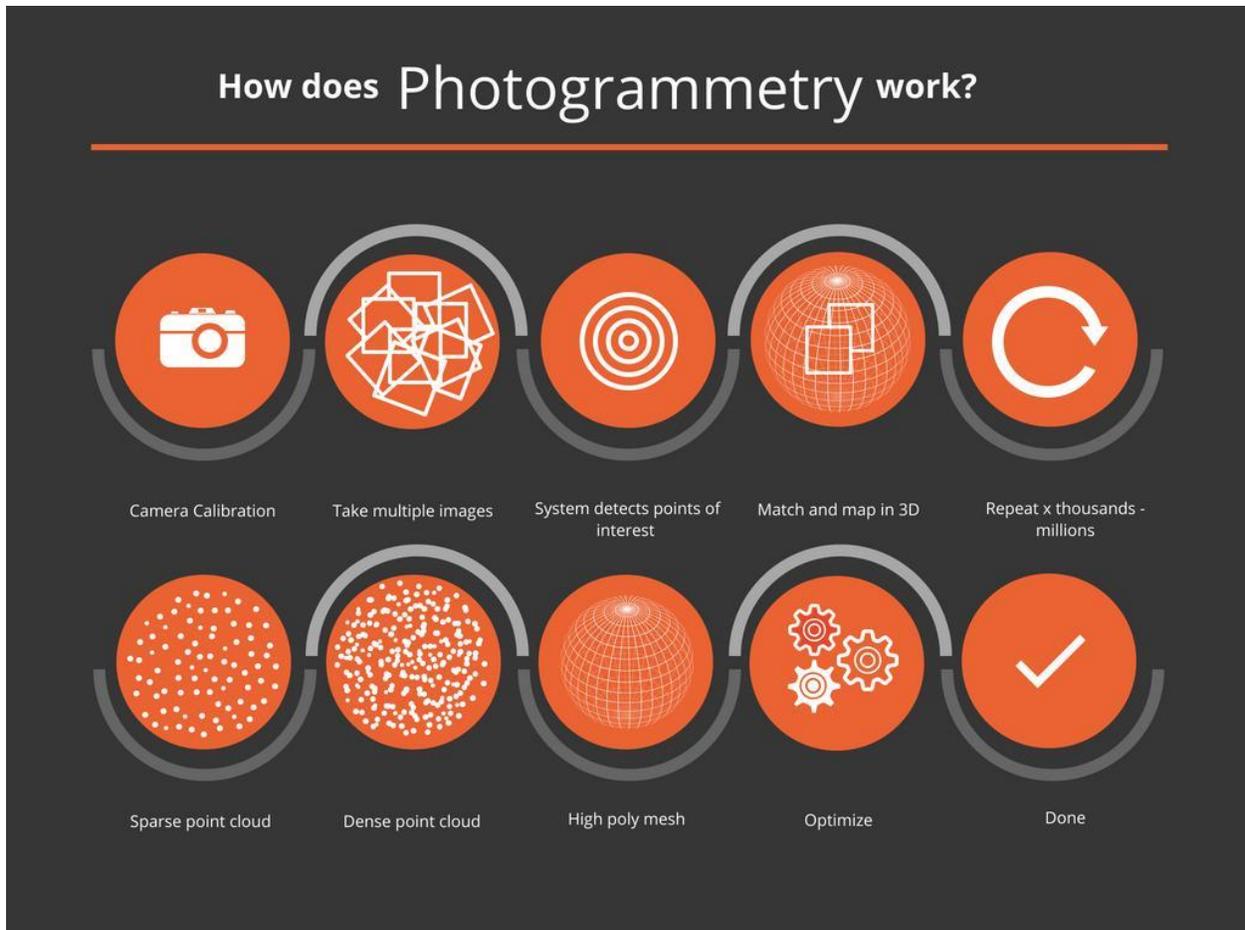


Figure 2: *How photogrammetry works.*

The sparse point cloud is the core of the rest of the processes. You can often trace a failed model back to a failure in the sparse point cloud generation process. This sparse point cloud is then sent through other systems that change it to a dense point cloud, and eventually a high poly mesh. This mesh might be optimized and smoothed to remove points that are not part of the main mesh. The result is a surface that is visually pleasing and is easier to work with.

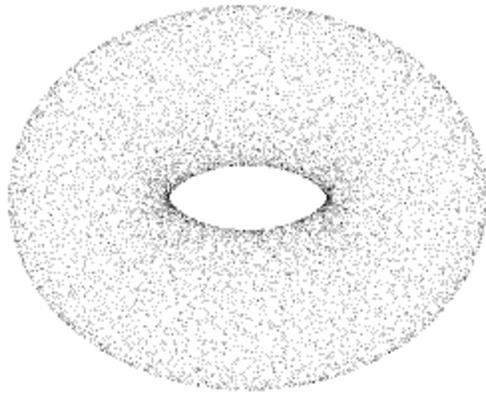
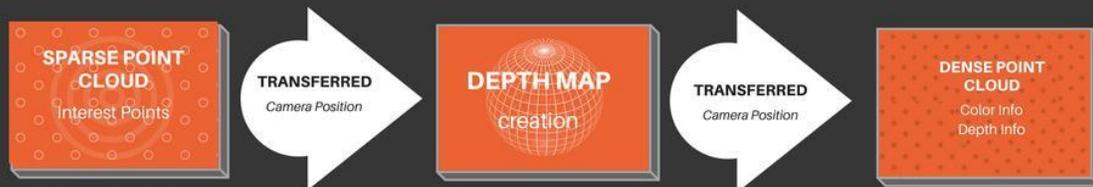


Figure 3: *The sparse point cloud is an integral part of creating both photogrammetry and volumetric capture.*

Volumetric capture a process of creating a point cloud from a frame or sequence of frames and volumetric video the playback of captured volumetrics into video format, both have had contradictory definitions in the past. These fields are changing at such a rapid pace that it is challenging to create a consensus amongst the professionals blazing the trails in these areas. For years, this process was relegated to using cameras with depth information (RGB-D cameras) which is no longer the case. Depth cameras used secondary technology to create depth information in a host of different ways, from infrared (IR), to stereo disparagement, with lasers falling somewhere in the middle. Many other styles of technology can now be leveraged to create depth. If a depth can be distinguished from any technique or camera, it can be used to volumetrically capture an object and create a CG mesh. Depth to point cloud from capture is the first widely accepted methodology to define the term volumetric. The following discussion of volumetric video scanners aligns the above with the use multiple cameras.

What do Photogrammetry point clouds do?



What about with IR cameras?



Figure 4: What do photogrammetry point clouds do?

The output of these scanners is often broken down into three parts: images, depth maps, and sparse point cloud information. The depth map and sparse point cloud are different slices of the same data. Sparse point cloud output is a text file that contains RGB with 3D information for a given point, while depth maps only provide distance information based on the position relative to the camera and its view. The combination of information about an image and a depth map from the same view can create the same data and information found in a sparse point cloud. At this point in the generation process most pipelines merge into one, and the data that comes from RGB-D cameras and RGB cameras is indistinguishable in its processing. The next steps are cleanup, dense point cloud generation, mesh creation, smoothing, and optimization.



Figure 5: *Structure Sensor**: An RGB-D scanner is used for volumetric capture.

Types of Photogrammetry

Photogrammetry can be created in several different ways. The most common is a point-and-shoot style system, where a user takes photos of the world, or an object, from a single camera. Other types of photogrammetry are multi camera, and video-to-photogrammetry.

Point-and-Shoot Photogrammetry

When you use point-and-shoot photogrammetry, you move a spot at a time to sufficiently cover the entire area in order to get a mesh with no holes or texture issues. This setup is quite inexpensive, and can even be done on a mobile device for the capture side, while processing is still relegated to a PC. The big benefit of this system is that the meshes can be captured anywhere. When you fully understand the point-and-shoot system, you can get great mesh results in very little time. It's also portable and requires the least amount of gear.

To make the process of capturing easier, add a spinning base to the pipeline, as shown in Figure 6. Set the object on the spinning base and move it a few degrees for each new picture to ensure that the system will access the same light conditions and cover the entire model. Next, lift and lower the camera tripod, capturing the asset from multiple angles. This decreases the cost of lighting and time of setup while still producing consistent results.



Figure 6: Adding a spinning base makes repetitive camera shots easier.

The point-and-shoot technique requires you to physically move around the mesh in all ways possible. If you miss a single angle, the mesh might be useless from a particular viewpoint. Also, animated objects are almost impossible to capture with this technique. When taking a photo, even inanimate objects are affected by external forces, such as a leaf blown in the wind. Drawbacks aside, point-and-shoot is the best capturing technique for those starting to use photogrammetry. Learning how the mesh generation works and reacts to your input is key to success in mesh generation.



Figure 7: High-end cameras are required for point-and-shoot techniques.

Multi-Camera Photogrammetry

Multi-camera photogrammetry is a vastly more advanced type of photogrammetry. You take photos of an object from multiple locations, the same as with the point-and-shoot system. But multi-camera photogrammetry removes the repetitive nature of one person taking photos from multiple locations by mimicking the camera movement with extra cameras.

Multi-camera photogrammetry has many benefits:

- Allows for faster capture of an object
- Enables capturing a true slice of time
- Easier to capture animated objects
- Allows for generation of meshes in action

Because the cameras are static, the quality of the mesh and the sparse point cloud increases.

If a team captures a large number of images, multi-camera photogrammetry takes less time to calculate camera location and angle. Also, these calculations can be recorded and used in further setups, which massively reduces processing time.

Bear in mind, however, that multi-camera photogrammetry costs more, is more complex, and is less portable—most multi-camera systems require at least 40 cameras to create believable results. This makes capture more difficult and increases

the cost of setup. All of the shutters should be synced up exactly for best results. Forty or more cameras require at least 3,000 watts of power for a good setup which, in itself, decreases portability. This is the reason most systems are set up at single-purpose studios.



Figure 8: Multi-camera photogrammetry with videogrammetry capabilities.

Video-to-Photogrammetry

Another common technique is video-to-photogrammetry. It combines the two types of photogrammetry with unfavorable results, so is included here more as a warning. Video-to-photogrammetry is often captured on a cell phone or a GoPro* camera. You would physically move around an area from one location to another while the camera is recording. With current technology and software available, this produces subpar results and low-quality meshes. Even with improvements in technology, cameras designed with rolling shutters and low-quality lenses inherently cause quality issues (for example, using a low-end drone flyover for captures). Despite being easy to use, video-to-photogrammetry is not recommended.



Figure 9: Drone with small cameras used in video-to-photogrammetry.

Use Cases for Photogrammetry

Photogrammetry has many uses across different sectors. An archivist can use it to study, and safely archive, fragile or historically significant objects. In the entertainment industry, it is used to create games and movies. And now, forensic science can utilize photogrammetry to reconstruct crime scenes.

Uses and Techniques for Object Preservation

Meshes of delicate objects can live on forever, digitally. Students and scientists can study these assets at multiple educational facilities simultaneously, without fear of specimen decay or destruction. Scientific specimens that aren't easy to transport, such as trace fossils, are at much less risk than before. Before photogrammetry, if you wanted to study animal tracks, trails and burrows, you would have to either dig them out of the ground—greatly increasing the likelihood of damage—or fill the holes with plaster or another hardening agent, allow that to set, and bring the mold back to the lab. Using plaster casts in archaeological studies carries the risk of hindering more advanced methods. Objects that are digitally available can be re-examined as technology evolves and progresses.

The most common form of photogrammetry capture for the research community involves the use of a single camera. A moving base can be added to help capture all directions and increase consistency of results.

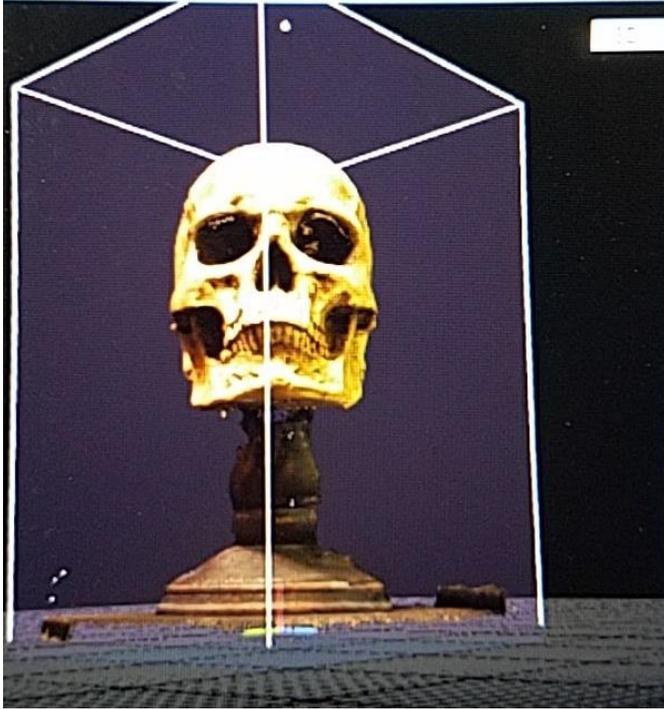


Figure 10: Human skull captured for medical research.

Capture Techniques for Entertainment

To prevent injuries to highly paid actors and actresses, or to work around unfortunate deaths before production wraps, video games and movies often utilize photogrammetrically captured data to create digital meshes. These meshes are then prepared for CG with rigging, cleanup, and relight, and then added to scenes. Digital doubles stand in for impossible, dangerous, or over-the-top sequences, such as the time slice, or bullet-time effect, in the movie *The Matrix*. To capture this effect, multiple cameras are placed in an arc around the actors performing the sequence.

The most common style of photogrammetry used in movies and games utilizes multi-camera photogrammetry along with motion capture. The character is digitally captured with the array and then rigged. (Rigging is the process of creating a structure within a mesh that animators can manipulate to move the mesh efficiently and accurately.) The character's movement is driven by motion capture. At times, the scenes are photogrammetrically captured as well, and most CG-heavy movies today utilize some form of photogrammetry.

Crime Reconstruction

Photogrammetry is used more and more to reconstruct crime scenes, and to preserve existing scenes and objects for future examination. Typically, several days after an incident, the only remaining artifacts are photos and evidence bags. When prosecutors or investigators need to piece the crime scene back together, without being able to physically go to a location, the use of VR headsets and photogrammetrically captured locations can reconstruct the situation as it was when the crime was committed. A fully captured location, with size markers, can give exact results and provide visual angles to view objects in a way that might not be possible, even if you could go to the physical location without disturbing evidence. Objects can also be rescaled and overlaid on other objects to help provide a fuller view of the possible events. Detectives can analyze scenes and objects without any contamination or destruction of evidence.

Types of Volumetric Capture

Compared with the long-established principle of photogrammetry, volumetric capture is the disruptive new kid on the block. The hardware options are vast and changing rapidly; low-cost options such as Intel® RealSense™ Depth Camera, Structure Sensor* from Occipital, and Kinect* from Microsoft (deprecated), have capabilities once restricted to very expensive, extremely high-end scanners.



Figure 11: Vive* VR setup, with goggles and controllers.

In January 2018, *Variety* reported on the opening of a new dedicated volumetric video capture facility named Intel Studios, reporting that “Intel wants to help Hollywood embrace the next generation of immersive media ... essentially producing high-end holographic content for VR, AR, and the likes.”



Figure 12: Inside the new Volumetric Capture Studio (Intel).

Narrow-Band and Wide-Band Laser Scanners

Lasers are the most commonly used volumetric capture technology. Given that lasers are an immense and broad topic, this is a general discussion focusing on just a minor subset of laser scanners on the market.

A laser can be narrow band or wide band. Narrow bands can be produced in strengths ranging from Class 1 to Class 4, with Class 1 being *inherently safe*, and Class 4 causing permanent damage to eyes and skin. Commercial laser scanners are mostly categorized as 2a, on the lower end of the Class 2 spectrum, meaning temporary viewing should not cause permanent damage (1,000 seconds of continuous viewing would be required to damage the retina, though safety precautions are always advised when using laser scanners). Narrow-band laser scanners produce the most accurate results, down to the micromillimeter scale from large distances. The downsides to this technology, however, are the cost, time of scan, and file formats.

A good quality scanner from one of the major manufacturers can easily cost hundreds of US dollars. The time to scan a single room with very accurate results is often more than 24 hours. If anything in that room moves while scanning is in progress, you can end up with ghosting, and other issues. Another factor in cost is file format. Most major companies create their own scanning data file type. As a result, you are often forced to use the company's proprietary viewing and conversion software, which increases costs and pipeline complexity. The IEEE is calling for standardized file

formats, and it has already made headway into the prosumer and academic research areas of use. Because the equipment can be difficult to set up and maintain, it is better suited to ultra high-end professional use.

A wide-band laser includes infrared (IR), which is below visible light. IR technology is used to create a wide-beamed laser that pulses multiple times per second with a pattern attached to it. When bounced off an object, this pattern is perturbed by the surface of the object. The IR camera detects the disruptions in the pattern, and from them calculates the depth very quickly. One of the first commercially available units in this category was the Microsoft Kinect. The Kinect had some issues, including secondary lights (especially sunlight), reflective surfaces, and problems detecting some users. Microsoft fixed many of the issues in later software and firmware updates, but the unit was eventually deprecated. Developers loved the inexpensive and readily available IR sensor, to the point that there are many offshoots of the most popular sensor connection software (OpenNI*) that are dedicated to its use and connectivity.

Even more advanced systems, such as the Intel RealSense Depth Camera D400-Series, can turn off their IR emitter and use the sun or any externally sourced IR. This cost-effective solution can overcome the major issue that has plagued low-end sensors for a very long time—sunlight. Most IR systems only create their own IR pattern; Intel's accepts the IR bounces off of any object and detects the object's position. The obvious upsides of this technology include more accurate depth detection, faster processing, and reduction of cameras needed. The downsides include often noisy results, with edges missing a considerable amount of detail, and increased technological complexity compared to the single-camera setups.



Figure 13: *The Intel® RealSense™ Depth Camera D400-Series is able to use sunlight as an IR source.*

Structured Light Imaging Scanners

The next most common type of volumetric capture utilizes structured light imaging. Examples of these scanners include the Occipital Structure Sensor. These scanners create high-intensity blasts of light at an ultra-fast rate that creates a pattern on the surface of any object. The system estimates the distance between a point on an object and the camera, based on the amount of time that a system takes to receive the light back, and the degree to which the pattern is perturbed.

Structured light imaging scanners use many of the same concepts as the IR scanner, but this technology is less affected by light issues (even though bright light causes a loss in tracking). These scanners have the same difficulty and ease of use as the IR scanners, and were developed in response to the IR issue with light sensitivity.

Volumetric Playback

Volumetric playback in video form combines IR and structured light technology, and takes the concept of photogrammetry one step further to make a fully digitized moving object. This technique creates a single mesh every frame. But many different issues arise when scaling to the volumetric video arena. Every camera (typically more than 40) needs to be in sync. Every single sensor (normally one-third to one-fourth of the camera numbers) needs to fire at the same rate, or a multiplicative of the same rate, as the cameras are then frame-dropped to match the camera rate. If lighting is not just right the meshes can be affected. For example, halogen lights produce a pulse in light, and while it's not perceptible by normal vision, this pulsing can interfere with computer vision and sensor technology.

Processing time is lengthy— hours, possibly days; even very high-end studios can process only six minutes of video in an entire day. This represents the pinnacle of difficulty, complexity, and cost. Studios can spend millions of US dollars simply to create a very small circle that a person can stand in, not much wider than an arm span. Then the challenge comes down to the playback side of the assets. The output of a single second of video from raw volumetric capture is at least 700 MB. This output cannot easily be played back in its current form, so it must be optimized—massively. Most volumetric videos are between 18 MB to 28 MB per second after these processes—still huge, but far easier to manipulate.



Figure 14: Volumetric video of author Timothy Porter (see <https://sketchfab.com/models/4c57f8a6bdfd4718845ef0265975133d>).

Photogrammetry Cost/Complexity Analysis

Each style of photogrammetry discussed here has its pros and cons. Consider the costs, benefits, and disadvantages of each in the following table:

Photogrammetry Comparison			
Style of Photogrammetry	Cost	Benefits	Disadvantages
Point and Shoot	USD 50--USD 5,000	Ease of setup Quick to shoot Portability	No animate objects Capture time can be quite long
Multi camera	USD 5,000--USD 200,000	Ability to capture animate objects Low capture time	Cost Portability

Video	USD 50–USD 5,000	Easiest setup Portability	Subpar quality, depending on the software the user needs to make frames from the video for import
-------	------------------	------------------------------	--

Table 1: Cost/benefit analysis of photogrammetry techniques.

Visualized another way, the following graphic illustrates different types of photogrammetry on the axes of computational needs and cost.

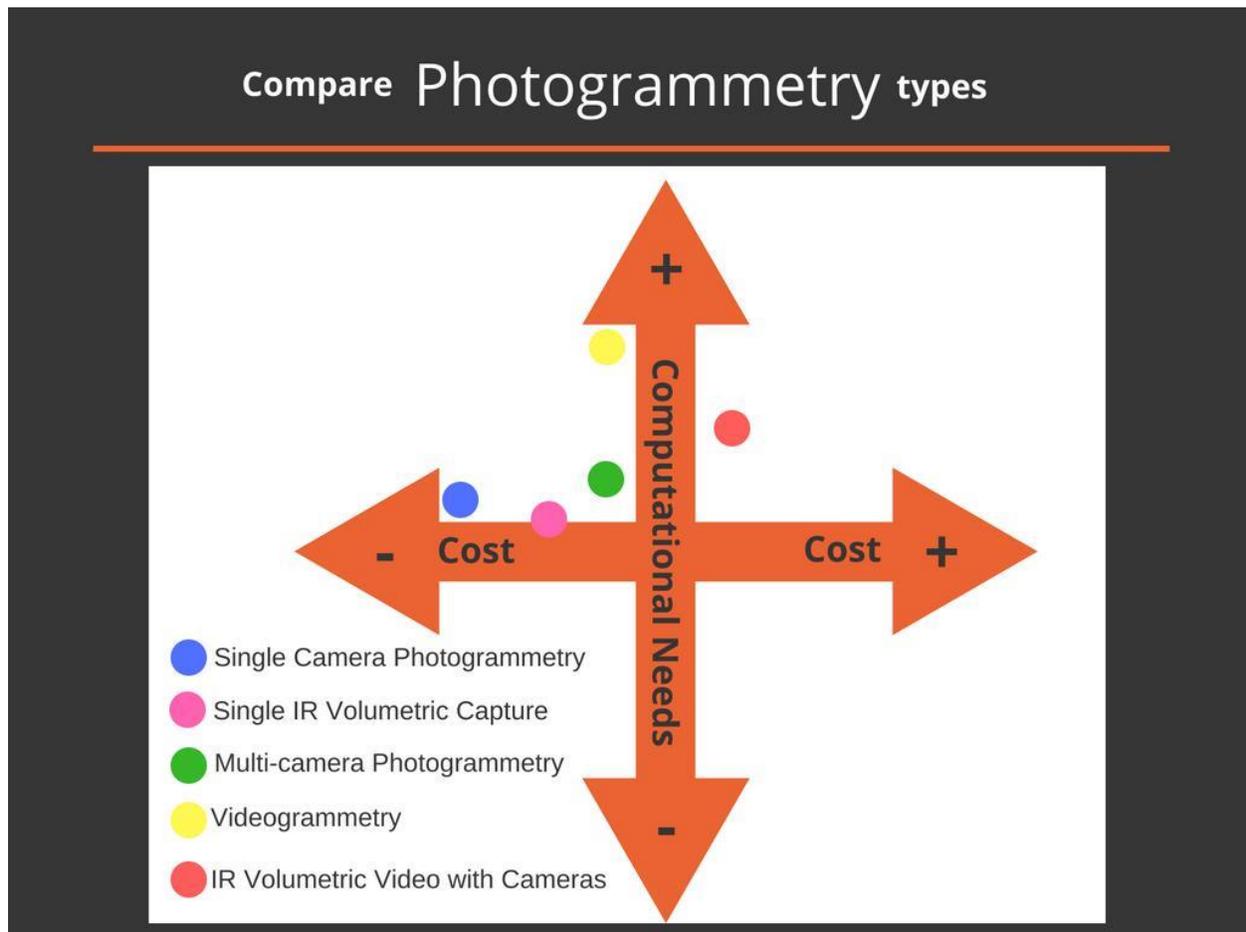


Figure 15: Comparison of photogrammetry types.

Future Trends Promise Higher Quality, Lower Costs

Volumetric video will revolutionize the way we consume and utilize content, just as TV and computers did before it. Technologies currently used for fast-paced and machine

interpreted needs, like navigating around large objects in self-driving cars, will become so high quality that they can entirely override the current technological stack required for capturing live performances. The evolution of this technology is moving at a rapid pace in both the consumer and commercial sectors. It is likely that current photogrammetry technology will be surpassed, if not entirely supplanted, in a few years. We can't know what will take its place, but taking into account the laws of physics, the creation and capture of a lightbeam will remain a fundamental necessity. **Some interesting options we are likely to see include wifi, radio, and the visual light spectrum.** Currently, our cultural future and advancement in this technological playscape are intertwined in a place where volumetric capture and photogrammetry are synonymous with viewing.

Photogrammetry on lower-end devices costing less than USD 5,000 often create a significantly better result when scanning objects for study, to archive, or for CG asset creation, than with a scanners comparable in price. Although several market leaders like Intel's RealSense D4000 are attempting to change the game with low-cost scanners that are designed to disrupt this marketplace, photogrammetry is still the best technology when you weigh cost and quality.

Summary

Technology related to photogrammetry and volumetric capture is rapidly changing, but understanding the different types of volumetric capture—including the costs, benefits, and complexity of each, with current use cases, as illustrated here—puts you well on the way to incorporating these technologies into your next project which you can post on [Intel Developer Mesh](#). You can find our work with [Volumation on DevMesh](#) and [Siggraph 2018 Introduction to Volumation](#).

About the Author

Timothy Porter is a pioneer in visual technology, and the CTO at Underminer Studios. He started his career in video games as a technical artist, before moving on to endeavors as a serial entrepreneur, and later as Pipeline Technical Director at Sony Pictures Imageworks. During his career, he has published more than 50 titles on multiple platforms. Driven by a mission to utilize entertainment paradigms in leading-edge technology, Tim is well-known for inspiring others to imagine the endless possibilities technology affords, and executing that vision. He was recognized by Intel as a 2017 Top Innovator.



Underminer Studios is using VR/AR to humanize the digital world, going to fantastic places within education, entertainment, health, and business for more engaging experiences. We design innovative products that will help define the future of human-computer interaction; combining the power of analytics and visualization technologies for compelling uses. Learn more about our work with volumetric capture at <http://underminerstudios.com/volumation>.

Notices

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

The products and services described may contain defects or errors known as errata which may cause deviations from published specifications. Current characterized errata are available on request.

Copies of documents which have an order number and are referenced in this document may be obtained by calling 1-800-548-4725 or by visiting www.intel.com/design/literature.htm.

Intel, the Intel logo, and Intel RealSense are trademarks of Intel Corporation in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others.

Microsoft, Windows, and the Windows logo are trademarks, or registered trademarks of Microsoft Corporation in the United States and/or other countries.

© 2018 Intel Corporation