Forge: A volumetric video processing platform for everyone

Rafael Pagés, Jan Ondřej, Konstantinos Amplianitis, Nicolás Moreno de Palma
Vologram Limited
Dublin, Ireland
rafa@vologram.com

Aljosa Smolic
Trinity College Dublin
Dublin, Ireland

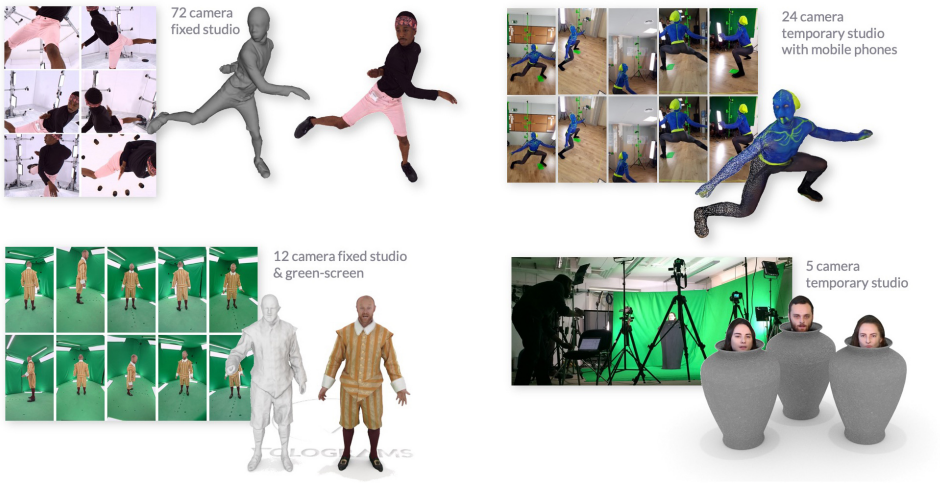

Figure 1: Volumetric video content created with Forge using different equipment and camera configurations.

ABSTRACT

We present Forge, a volumetric video processing platform running on the cloud that works with different camera configurations and adapts to the user needs. From high-end studios with dozens of synchronised cameras to scenes casually captured with smartphones, Forge allows creators to produce high-quality 3D human content autonomously with the equipment they already have, in an affordable way, unlocking new and accessible ways to create 3D human content at scale.

CCS CONCEPTS

• **Computing methodologies** → **3D imaging**; **Computational photography**.

KEYWORDS

Free-Viewpoint Video, Volumetric Video, Computer Vision, Augmented Reality, Virtual Reality

1 INTRODUCTION

Whether you turn on the TV, pick up your phone, or simply head out for a walk, a quick interaction with any medium shows the dominating prevalence of human content. Anyone in a news-worthy situation, i.e., a vlogger, a sports celebrity or even the general public, in other words, *humans*, are central to the content we consume, in any channel or form. Photography empowered content creators to bring real people into static visual media (from newspapers and street ads to Instagram). Video did its part for dynamic visual content (from TV and films to YouTube). However, we are missing a solution that suits new media such as extended reality (XR, which includes Virtual, Augmented and Mixed Reality), visual effects (VFX) and video games. These 3D realms are stuck in the *cartoon era*: instead of real humans, we are limited to artificial humans designed by 3D artists and developers. Existing solutions are not only uncannily artificial, but either of low quality (billboards, avatars, etc.) or very expensive and resource-intensive (3D digital design, photogrammetry combined with animation, etc.).

Volumetric video, also known as Free-Viewpoint Video [6], is a field that has the potential of shaping the way we communicate in the future, and some technology giants such as Microsoft and Intel have already developed their own solutions. Typically, these solutions involve complex and extremely expensive camera setups that include dozens (or even hundreds) of synchronised cameras and other specific hardware, making them accessible only to clients with massive budgets. In contrast, our patent-pending technology has been designed to work in challenging conditions (indoors or outdoors, with a reduced number of cameras and even with handheld smartphones) as our 3D reconstruction system uses state-of-the-art algorithms to estimate the shape and appearance of human performances. This opens up new possibilities for creators, artists and professionals who could not enter this field before, and are looking for a solution tailored to their needs. The next step towards making volumetric video mainstream is Forge: a scalable software-only, camera-agnostic cloud platform with high-end quality and low-end pricing, that makes our volumetric video technology available to content creators around the world. XR, VFX and video game studios will be able to feed their apps, productions and games with human

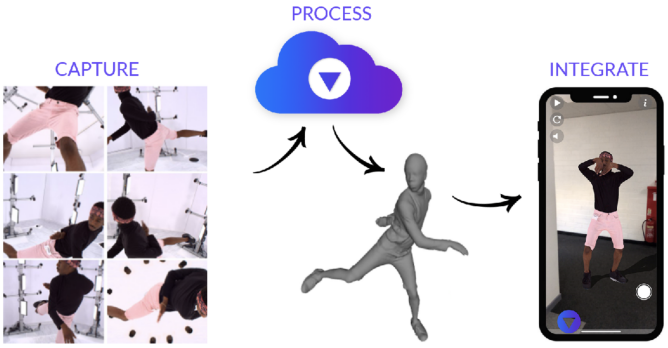

Figure 2: Diagram of how Forge integrates in 3D human content creation pipelines.

content directly from reality, using a camera setup that suits their needs, following this simple approach:

- **Setup**: users register on Forge and set up their equipment to record the performance from surrounding angles following our online guidelines.
- **Capture**: users video-shoot the performance and upload the resulting videos to Forge.
- **Process**: Forge processes the standard 2D videos into 3D volumetric video. The resulting 3D assets are made available to download, together with additional useful data (3D skeleton, camera calibration information, etc.).
- **Integrate**: users download the 3D assets and integrate them into their digital project.

2 TECHNOLOGY

2.1 Volumetric video pipeline

Forge's volumetric video reconstruction capabilities are based on the work introduced by Pagés et al. [4], which was designed with the goal of becoming a software-only solution for simplifying the creation of volumetric content. This is different from other technologies that were built for a specific hardware setup, such as the work from Guo at el. [3] with an enhanced light-stage or Collet et al. [2], with more than a hundred cameras, including infrared cameras and projectors. Our algorithms are robust and prepared to work in uncontrolled environments (outdoors, different camera/subject positions, etc.), with anything from a low number of consumer level smartphones to a large number of production video cameras. In summary, the key components of our volumetric video technology are the following:

- Automatic camera calibration.
- Automatic foreground-background segmentation.
- Scene volume estimation through a multi-source shape-from-silhouette algorithm.
- Enhanced depth map estimation, using denser sparse point clouds [1] and multi-view stereo.
- Automatic 3D skeleton estimation.
- Temporally coherent meshes, through automatic keyframe detection and mesh tracking.
- High-quality multi-view texture mapping [5]
- Sequence encoding based on video textures and keyframe meshes.

2.2 Cloud processing platform

First, Forge uses the power of the cloud to be accessible from virtually any location, deploying processing resources on-demand. Our platform processes requests for different customers in parallel and executes parts of a single request on many CPUs at the same time, enabling scalability and timeliness. Consequently, this minimises clients' initial investment, as they don't need their own servers.

Forge is built as a containerised cloud platform, splitting each component of the volumetric reconstruction pipeline into microservices that perform discrete self-contained tasks. The system is hosted on Amazon AWS and uses the Docker container specification for full infrastructure automation. This approach has a number of advantages over standard cloud infrastructure where full systems were built on a small number of individual virtual machines:

- Distributed functionality results in fewer failure bottlenecks.
- Automated resource allocation for minimal manual operation and processing at scale.
- Each microservice is completely self-contained and responsible for its own functionality, which enables
- Agile integration of new services: build and go into production in a matter of hours, not days.

REFERENCES

[1] D. Berjón, R. Pagés, and F. Morán. 2016. Fast feature matching for detailed point cloud generation. In *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. 1–6.

[2] Alvaro Collet, Ming Chuang, Pat Sweeney, Don Gillett, Dennis Evseev, David Calabrese, Hugues Hoppe, Adam Kirk, and Steve Sullivan. 2015. High-quality streamable free-viewpoint video. *ACM Transactions on Graphics (ToG)* 34, 4 (2015), 69.

[3] Kaiwen Guo, Peter Lincoln, Philip Davidson, Jay Busch, Xueming Yu, Matt Whalen, Geoff Harvey, Sergio Orts-Escolano, Rohit Pandey, Jason Dourgarian, et al. 2019. The relightables: volumetric performance capture of humans with realistic relighting. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–19.

[4] R. Pagés, K. Amplianitis, D. Monaghan, J. Ondřej, and A. Smolic. 2018. Affordable content creation for free-viewpoint video and VR/AR applications. *Journal of Visual Communication and Image Representation* 53 (2018), 192 – 201.

[5] R. Pagés, D. Berjón, F. Morán, and N. García. 2015. Seamless, Static Multi-Texturing of 3D Meshes. *Computer Graphics Forum* 34 (2015), 228–238.

[6] Aljoscha Smolic. 2011. 3D video and free viewpoint video—From capture to display. *Pattern recognition* 44, 9 (2011), 1958–1968.